

Keyifli Okumalar Dileriz

Binlerce kategoride milyonlarca PDF e-kitap, ders kitapları, dergiler ve romanlara platformumuz üzerinden ücretsiz ulaşabilirsiniz.

RESMİ WEB SİTEMİZ

<https://rantey.org>

Yedek ve Duyuru Kanallarımız

Sitemize erişim engeli gelmesi veya adres değişikliği durumunda aşağıdaki kanallardan güncel giriş adresine erişebilirsiniz:

- Telegram: <https://t.me/birkitapbir>
- WhatsApp Kanalımız : <https://whatsapp.com/channel/0029VbDHv8uE50Us4lvMoc0Y>

Uygulamalarla MİKRO & KONOMETRİ

EDİTÖRLER

PROF. DR. ŞENAY ÜÇDOĞRUK BİRECİKLİ - PROF. DR. SEDA ŞENGÜL



UYGULAMALARLA MİKROEKONOMETRİ

Editörler

Prof. Dr. Şenay Üçdoğruk Birecikli - Prof. Dr. Seda Şengül



UYGULAMALARLA MİKROEKONOMETRİ

Editörler: Prof. Dr. Şenay Üçdoğruk Birecikli - Prof. Dr. Seda Şengül

Yayın No. : 3308
İktisat/Ekonomi: 253
ISBN : 978-625-439-171-2
E-ISBN : 978-625-439-172-9
Basım Sayısı : 1. Basım, Şubat 2021

© Copyright 2021, NOBEL AKADEMİK YAYINCILIK EĞİTİM DANIŞMANLIK TİC. LTD. ŞTİ. SERTİFİKA NO.: 40340

Bu baskının bütün hakları Nobel Akademik Yayıncılık Eğitim Danışmanlık Tic. Ltd. Şti.ne aittir. Yayınevinin yazılı izni olmaksızın, kitabın tümünün veya bir kısmının elektronik, mekanik ya da fotokopi yoluyla basımı, yayımı, çoğaltımı ve dağıtımı yapılamaz.

Nobel Yayın Grubu, 1984 yılından itibaren ulusal ve 2011 yılından itibaren ise uluslararası düzeyde düzenli olarak faaliyet yürütmekte ve yayınladığı kitaplar, ulusal ve uluslararası düzeydeki yükseköğretim kurumları kataloglarında yer almaktadır.

Genel Yayın Yönetmeni : Nevzat Argun -nargun@nobelyayin.com-
Yayın Koordinatörü : Gülfem Dursun -gulfem@nobelyayin.com-

Redaksiyon : Dilek Gezgın -dilek@nobelyayin.com-
Sayfa Tasarım : Ahmet S. Baydar -ahmet@nobelyayin.com-
Kapak Tasarım : Mehtap Yürümez -mehtap@nobelyayin.com-
Baskı Sorumlusu : Yavuz Şahin -yavuz@nobelyayin.com-
Baskı ve Cilt : Genç Offset Sertifika No.: 32284
Süzgün Sokak No.:18/3 İskitler/Ankara

Kütüphane Bilgi Kartı

Üçdoğruk Birecikli, Şenay., Şengül, Seda.

Uygulamalarla Mikroekonometri / Editörler: Şenay Üçdoğruk Birecikli - Seda Şengül

1. Basım. xviii + 542 s. 19,5x27,5 cm. Kaynakça ve dizin var.

ISBN: 978-625-439-171-2

E-ISBN: 978-625-439-172-9

1. Yatay Kesit Veri Analizi 2. Uygulamalı Ekonometri 3. Stata Uygulamalı Mikroekonometri

Genel Dağıtım

ATLAS AKADEMİK BASIM YAYIN DAĞITIM TİC. LTD. ŞTİ.

Adres: Bahçekapı mh. 2465 sk. Oto Sanayi Sitesi No:7 Bodrum Kat Şaşmaz-ANKARA - siparis@nobelyayin.com-

Telefon: +90 312 278 50 77 - **Faks:** 0 312 278 21 65

E-Satış: www.nobelkitap.com - www.atlaskitap.com - **Bilgi:** esatis@nobelkitap.com - info@atlaskitap.com

Dağıtım ve Satış Noktaları: Alfa Basım Dağıtım, Arasta, Arkadaş Kitabevi, D&R Mağazaları, Dost Dağıtım, Ekip Dağıtım, Kida Dağıtım, Kitapsan, Nezh Kitapları, Pandora, Prefix, Remzi Kitapları

BÖLÜMLER VE YAZARLARI

Bölüm 1. Mikro Ekonometride Veri ve Verilerin Özellikleri

Bedriye Saraçoğlu

Hacı Bayram Veli Üniversitesi, İktisadi ve İdari Bilimler Fakültesi Ekonometri Bölümü, Emekli Öğretim Üyesi-Ankara

Bölüm 2. Doğrusal Olmayan Regresyon Modelleri ve En Çok Olabilirlik Tahmincisi (R ve Stata Uygulamaları)

Burak Alparslan Eroğlu

İstanbul Bilgi Üniversitesi, İşletme Fakültesi Ekonomi Bölümü, Eyyüpsultan-İstanbul

Bölüm 3. İki Durumlu Modeller ve Stata Uygulamaları

Hamdi Emeç

Dokuz Eylül Üniversitesi, İktisadi ve İdari Bilimler Fakültesi Ekonometri Bölümü, Buca-İzmir

Bölüm 4. Çok Durumlu Modeller ve Stata Uygulamaları

Ali Kemal Çelik

Ardahan Üniversitesi, İktisadi ve İdari Bilimler Fakültesi İşletme Bölümü, Ardahan

Erkan Oktay

Atatürk Üniversitesi, İktisadi ve İdari Bilimler Fakültesi Ekonometri Bölümü, Erzurum

Bölüm 5. Sıralı Modeller ve Stata Uygulamaları

Özlem Kiren Gürler

Dokuz Eylül Üniversitesi, İktisadi ve İdari Bilimler Fakültesi Ekonometri Bölümü, Buca-İzmir

Bölüm 6. Ardışık Logit Modeli ve Stata Uygulamaları

Özlem Kiren Gürler

Dokuz Eylül Üniversitesi, İktisadi ve İdari Bilimler Fakültesi Ekonometri Bölümü, Buca-İzmir

Hatice Özkoç

Muğla Sıtkı Koçman Üniversitesi, İktisadi ve İdari Bilimler Fakültesi İşletme Bölümü, Kötekli-Muğla

Bölüm 7. Yuvalanmış Logit Model ve Stata Uygulaması

Öznur Özdamar Giovanis

İzmir Bakırçay Üniversitesi, İktisadi ve İdari Bilimler Fakültesi İktisat Bölümü, Menemen-İzmir

Eleftherios Giovanis

Manchester Metropolitan University, Department of Economics, Policy And International Business,

Aydın Adnan Menderes Üniversitesi, Nazilli İktisadi ve İdari Birimler Fakültesi Maliye Bölümü, Nazilli-Aydın

Bölüm 8. Saymalı Modeller ve Stata Uygulamaları

Şenay Üçdoğru Birecikli

Dokuz Eylül Üniversitesi, İktisadi ve İdari Bilimler Fakültesi Ekonometri Bölümü, Buca-İzmir

Arzu Kökcen Eryavuz

Dokuz Eylül Üniversitesi, Sosyal Bilimler Enstitüsü, Buca-İzmir

Mehmet Ozan Özdemir

Dokuz Eylül Üniversitesi, İktisadi ve İdari Bilimler Fakültesi Ekonometri Bölümü, Buca-İzmir

Bölüm 9. Sınırlı Bağımlı Değişkenli Modeller ve Stata Uygulamaları

Seda Şengül

Çukurova Üniversitesi, İktisadi ve İdari Bilimler Fakültesi Ekonometri Bölümü, Balcalı-Adana

Bölüm 10. Switching Regresyon Modeli ve Stata Uygulaması

Seda Şengül

Çukurova Üniversitesi, İktisadi ve İdari Bilimler Fakültesi Ekonometri Bölümü, Balcalı-Adana

Salih Çam

Çukurova Üniversitesi, İktisadi ve İdari Bilimler Fakültesi Ekonometri Bölümü, Balcalı-Adana

Bölüm 11. Doğrusal Araç Değişkenli Regresyon Modeli (Iv, 2sls, Gmm) ve Stata Uygulaması

Merve Ertok Onurlu

Biga İktisadi ve İdari Bilimler Fakültesi Ekonometri Bölümü, Biga-Çanakkale

Bölüm 12. Panel İki Durumlu Modeller ve Stata Uygulamaları

Çiler Sigeze Güney

Çukurova Üniversitesi, İktisadi ve İdari Bilimler Fakültesi Ekonometri Bölümü, Balcalı-Adana

Bölüm 13. Panel Çok Durumlu Modeller ve Stata Uygulamaları

Çiler Sigeze Güney

Çukurova Üniversitesi, İktisadi ve İdari Bilimler Fakültesi Ekonometri Bölümü, Balcalı-Adana

Bölüm 14. Panel Sıralı Modeller ve Stata Uygulamaları

Canan Güneş

Biga İktisadi ve İdari Bilimler Fakültesi Ekonometri Bölümü, Biga-Çanakkale

Bölüm 15. Panel Saymalı Modeller ve Stata Uygulamaları

Egemen İpek

Tarsus Üniversitesi, Uygulamalı Bilimler Fakültesi Finans ve Bankacılık Bölümü, Tarsus-Mersin

Bölüm 16. Yaşam Analizleri ve Stata Uygulamaları

Arzu Kökcen Eryavuz

Bölüm 17. Politika Analizinde Mikro Ekonometrik Yöntemler ve Stata Uygulamaları: Farkların Farkı Yöntemi, Regresyon Süreksizliği Yöntemi, Eğilim Skoru Eşleştirme Yöntemi

Eleftherios Giovanis

Manchester Metropolitan University, Department of Economics, Policy and International Business,

Aydın Adnan Menderes Üniversitesi, Nazilli İktisadi ve İdari Birimler Fakültesi Maliye Bölümü, Nazilli-Aydın

Öznur Özdamar Giovanis

İzmir Bakırçay Üniversitesi, İktisadi ve İdari Bilimler Fakültesi İktisat Bölümü, Menemen-İzmir

Bölüm 18. Yapısal Eşitlik Modelleri ve Stata Uygulamaları

Hatice Özkoç

Muğla Sıtkı Koçman Üniversitesi, İktisadi ve İdari Bilimler Fakültesi İşletme Bölümü, Kötekli-Muğla

ÖN SÖZ

Ekonometrik araştırmalara uzunca bir dönem makroekonomik modellerin ve zaman serilerinin hâkim olduğu, mikroekonomik uygulamaların *Talep analizi*, *Beşeri Sermaye Teorisi*, *Üretim Fonksiyonu* analizi gibi bazı alanlarla sınırlı kaldığı söylenebilir. 1970'lerden itibaren mikroekonomi alanında hem metodolojik hem de uygulamalı araştırmaların öneminin artmasıyla bu araştırmalar ekonometri literatüründe *mikroekonometrik* çalışmalar olarak adlandırılmaya başlandı.

1950'li ve 1960'lı yıllardaki Ekonometri kitaplarında yer alan mikroekonometrik çalışmalar yatay kesit verileriyle yapılan doğrusal regresyon modelleriyle sınırlıydı. 1960'lı yıllardan sonra ekonomi alanında Nobel ödüllü (1981 yılı) James Tobin'in Tobit modeliyle (1958) öncülük ettiği, T. Amemiya ve G. S. Maddala tarafından geliştirilen mikroekonometrik çalışmaların ekonominin yanı sıra diğer sosyal bilim dallarında da kapsamlı bir şekilde uygulanmaya başlanmasıyla mikroekonometrik çalışmalar da ekonometri kitaplarında hak ettiği yeri almaya başladı. Aynı şekilde, mikroekonometrinin gelişiminde Nobel ödüllü (2000 yılı) James Heckman'ın Heckman seçim modelinin, Nobel ödüllü (2000 yılı) Daniel McFadden'ın kesikli seçim modellerini analiz etmek için geliştirdiği teori ve yöntemlerin, Nobel ödüllü (2005 yılı) Angus Deaton'ın yoksulluk, eşitsizlik, refah ve tüketimle ilgili çalışmalarının da katkısı göz ardı edilemeyecek kadar önemlidir.

Mikroekonometrik yöntemlerin gelişmesinde pek çok alanda mikro düzeyde büyük veri setlerindeki artışlar, panel veri setlerinin yaygınlaşması, bilgisayar teknolojisinin gelişmesiyle veri toplamada ve hesaplamadaki kolaylıklar da etkili olmuştur. Bunların yanında, birçok yeni ekonometrik yöntemin standart yazılım paket programlarının bir parçası haline gelmesinin ve kolay programlanmasını sağlayan yeni yazılımların geliştirilmiş (LIMDEP, R, STATA gibi) olmasının da katkısı olmuştur. Bu gelişmelerin sonucu olarak, uluslararası literatürde mikroekonometri alanında hem teorik hem de uygulamalı çok sayıda kitap ve makale olmasına rağmen Türkçede yazılmış kaynak sınırlı sayıdadır. Mikroekonometri alanında değerli çalışmalar yapan yazarlarımızın katkısıyla uygulamaya yönelik hazırlanan bu kitap; ekonomi, işletme bilim dallarının yanı sıra tüm sosyal bilim dallarındaki araştırmacıların, öğrencilerin bir başvuru kaynağı olarak yararlanabilecekleri bir referans kitap niteliğindedir.

18 bölümden oluşan bu kitapta, ilk olarak mikroekonometrik modellerin teorisi anlatılmış, daha sonra gerçek veri setleriyle Stata uygulamaları yapılarak, bu alanda duyulan Türkçe kaynak ihtiyacına katkı sunmaya çalışılmıştır. Kitaptaki bölümlerin veri setlerine (dta) ve komut (do dosyaları) dosyalarına ön söz sonunda verilen karekodla erişilerek her bölümün uygulamaları kolayca yapılabilir. Bu kitabın şüphesiz bazı eksiklikleri içermesi doğaldır. Gelecek baskılarda okuyucu-

lardan gelen dönüşlerle kitaptaki olası eksiklikler giderileceği gibi ileri Mikroekonometrik yöntemler de eklenecektir.

Son olarak, kitabın hazırlanmasında en büyük desteği veren öğretim üyesi yazarlarımıza değerli katkılarından dolayı teşekkür ederiz. Kitabımıza tanıtım yazısı yazma inceliğini gösteren, ulusal ve uluslararası literatürde çok değerli çalışmalarıyla ülkemizde bu alanın gelişmesine öncülük eden Orta Doğu Teknik Üniversitesi İktisat Bölümü emekli öğretim üyesi değerli hocamız *Prof. Dr. Aysıt Tansel'e* şükranlarımızı sunarız. Covid-19 Pandemisi nedeniyle, tüm aşamalarını ev ortamında gerçekleştirdiğimiz kitabımızın hazırlanma sürecinde en büyük özveriyi gösteren sevgili ailelerimize, kitabın yazımı için bizi teşvik eden Nobel Akademik Yayıncılık yetkilisi Sayın Gürbüz Ozan'a, baskı aşamasındaki özenli ve özverili çalışmasıyla Sayın Ahmet Baydar'a, kitabın basımı ve yayınlanmasında emeği geçen Nobel Akademik Yayıncılığın değerli çalışanlarına teşekkür ederiz.

Uygulamalarla Mikroekonometri kitabının öğrencilerimize ve araştırmacılara yararlı olmasına dileriz.

Şenay Üçdoğruk-Birecikli & Seda Şengül, Şubat 2021



*Kitaptaki bölümlerin veri setlerine (dta) ve komut (do dosyaları)
dosyalarına yukarıdaki karekodu taratarak ya da
https://www.nobelyayin.com/uygulamalarla-mikroekonometri_16871.html
adresini ziyaret ederek ulaşabilirsiniz.*

SUNUŞ

Prof. Dr. Aysıt Tansel

Ekonometrik yöntemler sürekli geliyor. Buna paralel olarak da bu yöntemleri kullanan profesyonel makalelerin sayıları hızla artmaktadır. Ekonometrik yöntemlere olan ilgi son yıllarda ülkemizde de artmıştır. Kısa bir inceleme, sadece 2020 yılı içerisinde beşin üzerinde Türkçe ekonometri kitabı basıldığını göstermektedir. Bunlar temel ekonometrik yöntemleri anlatan kitaplar yanında panel veri ekonometrisi ve büyük veri ekonometrisi gibi özel konuları da ele alan ve EVIEWS, STATA ve R gibi programlarla örnekler veren kitaplardır.

Elimizdeki kitap Türkiye’de ilk mikroekonometri kitabı olma özelliğini taşıyor. Mikroekonometrik yöntemleri konu alan ilk Türkçe kitap olması dolayısı ile Türkçe ekonometri yazınına önemli bir katkı sağlamaktadır. Kitap yöntemleri ve uygulamaları buluşturmayı amaçlamıştır. Önce mikroekonometrik yöntemler kısa, odaklı ve sezgisel olarak anlatılmış arkasından bu yöntemlerin STATA komutları ile anlaşılır bir biçimde gerçek veri setlerine uygulamaları gösterilmiştir. Kullanılan veri setlerinin büyük bir kısmı Türkiye’ye aittir. Böyle bir yaklaşım gereksinim duyulan bir yoldur. Mikroekonometri alanında yatay-kesit veri setleri veya panel veri setleri kullanılır. Bu veri setleri birimsel düzeyde bireyleri, haneleri, firmaları, şehirleri, bölgeleri veya ülkeleri içerebilir. Panel veri setlerinde aynı birimler zaman içerisinde izlenir. Mikroekonometri bu birimlerin ekonomik davranışlarını regresyon yöntemleri kullanarak inceler.

Son zamanlarda mikroekonometrik yöntem kullanan çalışmaların artmasında bir kaç neden önemli olmuştur. Birinci olarak, mikroekonometrik yöntemler kuramsal yönden hızla gelişmektedir. Buna paralel olarak yazılım paket programları kuramsal gelişmeleri izleyerek gelişmiştir. İkinci olarak, başlangıçta sadece gelişmiş ülkelerde toplanan yatay-kesit ve panel veri setlerinin son zamanlarda gelişmekte olan ülkelerde de toplanarak yaygınlaşması bir etken olmuştur. Son olarak da, bilgisayar teknolojisindeki gelişmeler bu yöntemleri uygulamada özellikle zaman maliyetini azaltmıştır.

Uygulamalı çalışmalar veri kullanarak politika geliştirdikleri ve nedensel sorunları yanıtlamaya çalıştıkları zaman en faydalı olurlar. Bu yönden faydalı olan ve son zamanlarda çok kullanılan yöntemler (araç değişkenler, genelleştirilmiş momentler yöntemi, farkların-farkı yöntemi ve regresyon süresizliği analizi) kitabın son bölümlerinde ele alınmıştır. Bu konuların ele alınması çok faydalı olmuştur. İleriki baskılarda daha ileri yöntemler de ele alınabilir.

Kitap regresyon yöntemi, istatistik ve olasılık kuramlarının arka plan bilgisi olarak bilindiğini varsaymaktadır. Kitap ekonometri ve ekonomi bölümleri, toplumsal bilimler ve sağlık bilimleri

bölümlerindeki ileri düzey lisans öğrencileri yanında, yüksek lisans ve doktora öğrencileri ile profesyonel araştırmacılara yol gösterici ve faydalı olacaktır.

Kitaptaki her bir bölüm alanında uzman değerli ekonometriciler tarafından hazırlanmıştır. Kitaba katkı veren değerli araştırmacı-yazarları tek tek kutluyorum. Kitaba öncülük eden, editörlüğünü yapan ve içinde bulunduğumuz Covid-19 pandemi döneminde özveri ile çalışan değerli meslektaşlarım Seda Şengül ile Şenay Üçdoğruk-Birecikli'yi candan kutluyorum.

Prof. Dr. Aysıt Tansel, Şubat 2021

İÇİNDEKİLER

1. MİKROEKONOMETRİ VE VERİLERİN YAPISI	1
1.1. Kukla Değişkenler	3
1.1.1. Bağımlı Değişkenin Kukla Olduğu Durumlar	3
1.1.2. Bağımsız Değişkenlerin Kukla Olduğu Durumlar ve Stata Uygulamaları	4
1.2. Kukla Değişkenlerin Farklı Kullanımları	10
1.2.1. Birden Fazla Kukla Değişken Durumu	10
1.2.2. Birden Fazla Nicel Açıklayıcı Değişken, Birden Fazla Kukla Değişken Durumu ..	11
1.2.3. Kukla Değişkenlerin Birbiri ile Etkileşime Girmeleri	11
1.2.4. Kukla Değişkenlerde Chow testi	14
1.3. Kaynaklar	18
2. EN ÇOK OLABİLİRLİK TAHMİNCİSİ, DOĞRUSAL OLMAYAN REGRESYON MODELİ VE STATA UYGULAMALARI	19
2.1. En Çok Olabilirlik Tahmin Edicisi	20
2.1.1. Olabilirlik Fonksiyonu	20
2.1.2. Numerik Optimizasyon Yöntemleri	21
2.1.3. En Çok Olabilirlik Tahmin Edicisinin Varyansı	23
2.1.3.1. Bilgi Matrisi ve Hessian	23
2.1.4. En Çok Olabilirlik Tahmin Edicisinin Limit Özellikleri	23
2.1.5. Değişmezlik İlkesi ve Delta Yöntemi	24
2.1.6. Sözde En Çok Olabilirlik Tahmini	25
2.1.7. Hipotez Testleri	25
2.1.7.1. Wald Testi	25
2.1.7.2. Olabilirlik Oranı Testi	26
2.1.7.3. Lagranj Çarpanı Testi	26
2.1.8. Tek Kısıtlı ve Tek Taraflı Hipotezlerin Test Edilmesi	26
2.2. Doğrusal Olmayan Regresyon Modelleri	27
2.3. En Çok Olabilirlik Tahmininin Stata Uygulamaları	28
2.3.1. Örnek 1: Basit Poisson Modeli	28
2.3.2. Örnek 2: Çok değişkenli Normal Olmayan Hata Terimine Sahip Modelin Tahmini	30
2.3.3. Örnek 3: Doğrusal Olmayan Regresyon Modelinin “ml” Fonksiyonuyla İncelenmesi	33
2.3.3.1. Doktor Ziyareti Modelinin Poisson Regresyon Tahmini	35

2.3.3.2. Doktor Ziyareti Modelinin Doğrusal Olmayan En Küçük Kareler ile Tahmini.....	35
2.3.3.3. Doktor Ziyareti Doğrusal Olmayan Modelinin En Çok Olabilirlik Tahmin Edicisiyle Tahmini	37
2.4. Sonuç	41
2.5. Kaynaklar	42
3. İKİ DURUMLU MODELLER VE STATA UYGULAMALARI.....	43
3.1. Doğrusal Olasılık Modeli (DOM)	43
3.1.1. Doğrusal Olasılık Modelindeki Sorunlar.....	44
3.1.1.1. Hata Terimlerinin Normal Dağılmaması Sorunu.....	44
3.1.1.2. Hata Terimlerinin Değişen Varyanslı Olma Sorunu	45
3.1.1.3. Doğrusal Olmama Sorunu.....	45
3.1.1.4. Belirlilik Katsayısının Küçük Olması Sorunu	46
3.1.2. Doğrusal Olasılık Modeli İçin Stata Uygulaması.....	46
3.2. Logit Model	49
3.2.1. Logit Modelin En Çok Olabilirlik Yöntemiyle (EÇOY) Tahmini	51
3.2.2. Logit Modeli İçin Stata Uygulaması	51
3.2.2.1. Logit Modelinin EÇOY İle Tahmini.....	52
3.2.2.2. Logit Modelinde Fark Oranlarının (Odds Ratio) ve Marjinal Etkilerin Bulunması ve Yorumlanması	53
3.3. Probit Model	57
3.3.1. Probit Modelin En Çok Olabilirlik Yöntemiyle (EÇOY) Tahmini	58
3.3.2. Probit Modeli İçin Stata Uygulaması	59
3.3.2.1. Probit Modelinin EÇOY ve Standart Hatalarının Dirençli (Robust) Tahmini.....	59
3.3.2.2. Probit Modelinde Marjinal Etkilerin Bulunması ve Yorumlanması	60
3.4. Logit ve Probit Modeller İçin İyi Uyum İstatistikleri ve Model Seçim Kriterleri	63
3.5. Kaynaklar	68
4. ÇOK DURUMLU MODELLER VE STATA UYGULAMALARI.....	71
4.1. Çok Durumlu Logit Modeli	72
4.1.1. İlişkisiz Alternatiflerin Bağımsızlığı Varsayımı	74
4.2. Çok Durumlu Logit Modeli	77
4.2.1. Çok Durumlu Logit Modeli Tahmin Sonuçları	79
4.3. Çok Durumlu Probit Modeli Stata Uygulaması	94
4.3.1. Çok Durumlu Probit Modeli Tahmin Sonuçları.....	95
4.4. Kaynaklar	100
5. SIRALI REGRESYON MODELLERİ ve STATA UYGULAMALARI.....	103
5.1. Sıralı Değişkenler için Gizli Değişken Modeli	103
5.2. Sıralı Modelin Tahmini	107
5.2.1. Sıralı Logit Model Tahmini.....	108
5.2.2. Sıralı Probit Model	109

5.3. Model Tahmin Sonuçlarının Yorumlanması	110
5.3.1. Fark Oranlarının Elde Edilmesi	110
5.3.2. Marjinal Etkilerin Yorumu	110
5.4. Paralel Eğriler Varsayımı	111
5.5. Sıralı Model Tahminlerinin Stata Uygulaması	112
5.5.1. Sıralı Logit Modelinin Stata Tahminleri	115
5.5.2. Sıralı Probit Modelinin Stata Tahminleri	123
5.5.3. Paralel Regresyon Varsayımı Testi	129
5.6. Genelleştirilmiş Sıralı Logit Model	133
5.7. Kaynaklar	138
6. ARDIŞIK LOGİT MODEL ve STATA UYGULAMALARI	141
6.1 Koşullu Geçişlerin Modellenmesi	142
6.2 Ardışık Logit Modellerin Yorumlanması	144
6.3 Ardışık Logit Modelinin Stata Uygulaması	145
6.4. Ardışık Model Sonuçları	148
6.5. Kaynaklar	159
7. YUVALANMIŞ (NESTED) LOGIT MODELİ VE STATA UYGULAMALARI	161
7.1. Yuvalanmış Logit Modelinin Teorik Çerçevesi	161
7.2. Yuvalanmış Logit Modeli: Stata Uygulaması	165
7.3. Yuvalanmış Logit Modelinde Kullanılan Stata Komutları	176
7.4. Kaynaklar	176
8. SAYMALI MODELLER VE STATA UYGULAMALARI	179
8.1. Poisson Regresyon Modeli	182
8.2. Negatif Binom Regresyon Modeli	192
8.3. Sıfır Yığılmalı Sayma Modelleri	200
8.3.1. Sıfır Yığılmalı Poisson Regresyon Modeli	201
8.3.2. Sıfır Yığılmalı Negatif Binom Regresyon Modeli	206
8.4. Saymalı Modellerin Karşılaştırılması	211
8.5. Kaynaklar	217
9. SINIRLI BAĞIMLI DEĞİŞKENLİ MODELLER VE STATA UYGULAMALARI ...	219
9.1. Budanmış (Truncated) ve Sansürlü (Censored) Veriler	219
9.1.1. Budanmış ve Sansürlenmiş Dağılımlar Üzerine Bazı Temel Bilgiler	220
9.1.1.1 Budanmış Normal Dağılım	220
9.1.1.2 Budanmış Regresyon Modeli	222
9.1.1.3 Sansürlü Normal Dağılım	222
9.2. Tobit Modeli	223
9.2.1. Tobit Modelinde Beklenen Değerler	224
9.2.2. Tobit Modelinde Marjinal Etkiler	225
9.2.3. Sansürlü Veriler Neden En küçük Kareler ile Tahmin Edilemez?	225
9.2.4. Tobit ve Probit Modelleri Arasındaki İlişki	226
9.2.5. Tobit Modelinde Değişen Varyans	226

9.2.6. Tobit Modelinin Stata Uygulaması	227
9.3. Heckman Seçim Modeli	233
9.3.1. Örneklem Seçim Modellerinin Tahmini.....	234
9.3.1.1. Heckman Seçim Modelinin En Çok Olabilirlik Tahmini	235
9.3.1.2. Heckman'ın İki Aşamalı Tahmin Süreci	235
9.3.2. Heckman Seçim Modellerinin Stata Uygulamaları	236
9.4 Çift Eşik (Double Hurdle) Modeli	243
9.4.1. Çift Eşik Modelinde Değişen Varyans	245
9.4.2. Çift Eşik Modelinde Marjinal Etkiler	245
9.5. Çift Eşik Modelinin Stata Uygulaması	246
9.6.Kaynaklar	250
10. SWITCHING REGRESYON MODELİ VE STATA UYGULAMASI	253
10.1. Rejim Kavramı	253
10.2. Switching Regresyon Tahmin Problemi	255
10.2.1. Dışsal (exogenous) Switching Regresyon Modeli	255
10.2.2. İçsel (Endogenous) Switching Regresyon Modeli	256
10.3. Switching Regresyon Modelinin Stata Uygulaması	258
10.4. Kaynaklar	267
11. DOĞRUSAL ARAÇ DEĞİŞKENLİ REGRESYON MODELİ (IV, 2SLS, GMM) VE STATA UYGULAMASI	269
11.1. İçsellik Kavramı	269
11.2. Araç Değişken (IV-Instrumental Variable)	270
11.3. Doğrusal Araç Değişkenli Regresyon Modeli Tahmini	271
11.3.1. Basit Doğrusal Regresyon Modeli Kapsamında Araç Değişkenler Yöntemi ve Stata Uygulaması.....	271
11.3.2. Zayıf Araç Değişken Problemi ve Stata Uygulaması	279
11.4. Çoklu Regresyon Kapsamında Araç Değişkenler Yöntemi	281
11.4.1. İki Aşamalı En Küçük Kareler Yöntemi (2AEKK)	282
11.4.2. İki Aşamalı En Küçük Kareler Yöntemi (2AEKK) Stata Uygulaması	282
11.5. Açıklayıcı Değişkenler İçin İçsellik Sınaması ve Stata Uygulaması	288
11.6. Zayıf Araç Değişkenler İçin İstatistiksel Testler ve Stata Uygulaması.....	290
11.7. Genelleştirilmiş Momentler Yöntemi (GMM) ve Stata Uygulaması	292
11.8. Belirlenme Kısıtlarına İlişkin Testler	294
11.8.1. Aşırı Belirlenme Durumunda Araçların Geçerliliğinin Test Edilmesi.....	295
11.8.2. Aşırı Belirlenme Durumunda Araçların Geçerliliğinin Test Edilmesi İçin Stata Uygulaması	295
11.9. Özet ve Sonuç	296
11.10. Kaynaklar	298
12. PANEL İKİ DURUMLU MODELLER VE STATA UYGULAMALARI	301
12.1. Panel İki Durumlu Modeller	301
12.2. Rassal Etkiler İki Durumlu Modeller	302

12.3. Sabit Etkiler İki Durumlu Modeller	304
12.4. Panel İki Durumlu Modeller Stata Uygulaması	305
12.5. Kaynaklar	313
13. PANEL ÇOK DURUMLU MODELLER VE STATA UYGULAMALARI	315
13.1. Panel Çok Durumlu Modeller	315
13.2. Rassal Etkiler Çok Durumlu Probit Model	316
13.3. Rassal Etkiler Çok Durumlu Logit Model	317
13.3.1. Rassal Etkiler Çok Durumlu Logit Model Stata Uygulaması	318
13.4. Kaynaklar	326
14. PANEL SIRALI MODELLER VE STATA UYGULAMALARI	329
14.1. Panel Sıralı Modeller	329
14.2. Rassal Etkiler Sıralı Probit Model	331
14.2.1. Rassal Etkiler Sıralı Probit Model Stata Uygulaması	333
14.3. Rassal Etkiler Sıralı Logit Model	340
14.3.1. Rassal Etkiler Sıralı Logit Model Stata Uygulaması	340
14.4. Kaynaklar	345
15. PANEL SAYMALI MODELLER VE STATA UYGULAMALARI	347
15.1. Panel Poisson Modelleri	348
15.1.1 Örneklem Ortalaması Panel Poisson Modeli	348
15.1.1.1. Örneklem Ortalaması Panel Poisson Model Stata Uygulaması	349
15.1.2. Sabit Etkiler Panel Poisson Modeli	367
15.1.2.1. Sabit Etkiler Panel Poisson Model Stata Uygulaması	369
15.1.3. Rassal Etkiler Panel Poisson Modeli	375
15.1.3.1 Rassal Etkiler Panel Poisson Model Stata Uygulaması	376
15.1.4 Model Seçimi	380
15.2 Panel Negatif Binom Modelleri	381
15.2.1 Örneklem Ortalaması Panel Negatif Binom Modeli	382
15.2.1.1 Örneklem Ortalaması Panel Negatif Binom Model Stata Uygulaması	382
15.2.2 Sabit Etkiler Panel Negatif Binom Modeli	392
15.2.2.1. Sabit Etkiler Panel Negatif Binom Model Stata Uygulaması	393
15.2.3 Sabit Etkiler Panel Negatif Binom Model Stata Uygulaması	398
15.2.3.1. Rassal Etkiler Panel Negatif Binom Model Stata Uygulaması	398
15.2.4 Rassal Etkiler Panel Negatif Binom Modeli	403
15.3 Kaynaklar	403
16. YAŞAM ANALİZLERİ VE STATA UYGULAMALARI	405
16.1. Yaşam Analizinde Kullanılan Sansürlü Veri Türleri	406
16.2. Yaşam Analizinde Kullanılan Fonksiyonlar	412
16.3. Yaşam Analizi Yöntemleri	414
16.3.1. Parametrik Olmayan Analiz Yöntemleri	414
16.3.1.1. Yaşam Tablosu (LT) Analizleri	414
16.3.1.2. Kaplan Meier (KM) Yöntemi	416

16.3.1.3. Cox Regresyon Modeli	422
16.4. Kaynaklar	429

17. POLİTİKA ETKİ ANALİZİNDE MİKROEKONOMETRİK YÖNTEMLER VE STATA UYGULAMALARI..... 433

17.1. Farkların Farkı Yöntemi	434
17.1.1. Farkların Farkı Yönteminin Teorik Çerçevesi	434
17.1.2. Farkların Farkı Yöntemi: STATA Uygulaması.....	437
17.1.2.1. Analizde Kullanılan Değişkenlerin Tanımları ve Tanımlayıcı İstatistikler	439
17.1.2.2. Doğrusal Olasılık Modeli (DOM) Temelinde Farkların Farkı (FF) Analizi	440
17.1.2.3. Logit Modeli Temelinde Farkların Farkı (FF) Analizi	442
17.1.2.4. Paralel Eğilim Varsayımı Testleri.....	444
17.1.2.4.1. LAGS (Gecikmeli Değerler) Testi.....	444
17.1.2.4.2. Grafikselsel Test	448
17.1.3. Farkların Farkı Analizinde Kullanılan Stata Komutları	449
17.2. Eğilim Skoru Eşleştirme Yöntemi.....	450
17.2.1. Eğilim Skoru Eşleştirme Yönteminin Teorik Çerçevesi	450
17.2.2. Eğilim Skoru Eşleştirme Yöntemi: Stata Uygulaması	455
17.2.2.1. Yerine Koymama Yoluyla En Yakın Komşu Eşleştirmesi	455
17.2.2.2. Yerine Koyma Yoluyla En Yakın Komşu Eşleştirmesi	461
17.2.2.3. Yarıçap Eşleştirmesi	464
17.2.2.4. Mahalanobis Metrik Eşleştirmesi.....	466
17.2.2.5. Kernel Eşleştirmesi	469
17.2.3. Eğilim Skoru Eşleştirme Yönteminde Kullanılan Stata Komutları.....	475
17.3. Regresyon Süreksizliği Analizi	476
17.3.1. Regresyon Süreksizliği Analizinin Teorik Çerçevesi.....	476
17.3.2. Regresyon Süreksizliği Analizi: Stata Uygulaması.....	481
17.3.2.1. Analizde Kullanılan Değişkenlerin Tanımları, Tanımlayıcı İstatistikler ve RSA'nın Grafikselsel Gösterimi	482
17.3.2.2. Dikdörtgen Kernel Fonksiyonu Temelinde Doğrusal RSA	487
17.3.2.3. Dikdörtgen Kernel Fonksiyonu Temelinde Kuadratik RSA.....	492
17.3.3. Regresyon Süreksizliği Analizinde Kullanılan Stata Komutları	497
17.4. Bölüm Değerlendirmesi.....	498
17.5. Kaynaklar	498

18. YAPISAL EŞİTLİK MODELLERİ VE STATA UYGULAMASI 501

18.1. Yapısal Eşitlik Modellerinde Yer Alan Temel Kavramlar.....	501
18.1.1. Ölçme Modeli	503
18.1.2. Yapısal Model	503
18.1.3. Toplam, Doğrudan ve Dolaylı Etkiler	504
18.2. Yapısal Eşitlik Modellerinin Tahmini	505
18.2.1. En Çok Olabilirlik Tahmin Yöntemi (EÇÖ)	506

18.2.2. Ağırlıksız En Küçük Kareler	506
18.2.3. Genelleştirilmiş En Küçük Kareler Yöntemi	507
18.2.4. Ağırlıklı En Küçük Kareler Yöntemi	507
18.2.5. Standartlaştırılmış ve Standartlaştırılmamış Katsayılar	508
18.3. Doğrulayıcı Faktör Analizi (DFA)	508
18.4. Model Kurma Aşamaları ve Uyum Ölçütleri	511
18.5. Uygulamalı Örnek-1	512
18.6. Uygulamalı Örnek-2	524
18.7. Özet.....	535
18.8. Kaynaklar	537

BÖLÜM 1

MİKROEKONOMETRİ VE VERİLERİN YAPISI

Bedriye Saraçoğlu¹

Ekonometrik bir çalışmada, regresyon modelinin kurulması ve regresyon modelinde yer alan değişkenlerin belirlenmesi, bu değişkenlerle ilgili verilerin elde edilmesi son derece önemlidir. Çalışmanın güvenilirliğinin verilerin güvenilirliğine bağlı olduğu söylenebilir. Ekonometride genellikle üç tür veri kullanılmaktadır. Bu veriler zaman serisi verileri, yatay kesit verileri ve panel verilerdir.

Zaman Serisi Verileri: Bir veya birden fazla değişkenin değerlerinin belirlenen bir zaman boyutu içerisindeki gidişini gösteren verilerdir. Örneğin, milli gelir, işsizlik, ihracat vb. değişkenlerle ilgili veriler zaman serisi verileridir. Bu veriler yıllık, aylık, dönemlik, haftalık, günlük, saatlik olabilir. Gözlemler genellikle zaman içerisinde birbirini izleyecek şekilde eşit aralıklarla elde edilir.

Yatay Kesit Verileri ya da kısaca Kesit Verileri: Tüketiciler, üreticiler, hanehalkları, firmalar vb. hakkında zamanın belirli bir noktasında toplanan verilerdir. Bu tür verilerde zaman sabittir, fakat sabit zamanda izlenen farklı birimler söz konusudur. Örneğin, bir ülkede belirli bir andaki üniversite sayısı, bölgelere göre bir üretim dalındaki iş yerlerinin sayısı, işsiz sayısı gibi. Bunlar genellikle herhangi bir anda yapılan anketlerle belirlenen verilerdir. Yatay kesit verilerine, Türkiye İstatistik Kurumu'nun (TÜİK) Hanehalkı Gelir ve Tüketim Harcamaları Anketi'nden elde edilen veriler, Hanehalkı İşgücü Anketinden elde edilen veriler, Gelir ve Yaşam Koşulları Araştırması Anketinden elde edilen veriler örnek olarak gösterilebilir.

Panel (Karma) Veri: Yukarıda ayrı ayrı söz edilen zaman serisi ve yatay kesit veri türlerinin birleşiminden oluşan verilere panel veri adı verilir. Bu tür veri hem zaman hem de yatay kesit birimlerine göre değişim gösterir. Örneğin, Türkiye'nin beş ilinde son 20 yıla ait pamuk arzı verileri gibi. Panel veri; aynı yatay kesit birimlerinin zamana göre değişimini sa panel veri, farklı birimler için aynı zaman dönemindeki değişken değerlerini içeriyorsa havuzlanmış veri olarak adlandırılmaktadır. Yukarıda verilen örnekte son 20 yılın pamuk arzı aynı beş il için veriliyorsa bu veri panel veri setine karşılık gelir. Havuzlanmış veriler ise daha çok önce ve sonranın analizi biçiminde politika analizi amaçlı çalışmalarda kullanılmaktadır. Örneğin bir bölgeye yapılan geri dönüşüm tesisinin bölgedeki ev fiyatları üzerindeki etkisi, belirli bir bölgede ya da ülkede doğurganlık örüntüsünde bir değişimin olup olmadığı vb. araş-

¹ Prof. Dr. Hacı Bayram Veli Üniversitesi, İktisadi ve İdari Bilimler Fakültesi, Ekonometri Bölümü Emekli Öğretim Üyesi. E-mail: bedriye.saracoglu@hbv.edu.tr

tırmalarda havuzlanmış veriler ön plana çıkar. Zaman serisi ekonometrisi zaman serisi verilerini kullanarak yapılan analizleri kapsarken, yatay kesit ve panel veriler ile yapılan analizler genel ekonometrinin ve mikroekonometrinin konusunu oluşturmaktadır. Özünde regresyon analizine dayanan ekonometri türlerinin kendine özgü ilave varsayımları ve özellikleri vardır. Örneğin zaman serilerinde verilerin durağanlık, otokorelasyon, mevsimsellik gibi özelliklerinin incelenmesi ön planda iken anket verilerinde değişen varyans, yani farklı yayılımın incelenmesi, test edilmesi ön plana çıkmaktadır.

Makroekonometri, zaman serisi verilerini kullanarak, mikroekonometri yatay kesit ve panel verileri kullanarak gerçek sorulara yanıt arayan araştırmaları kapsar. Mikroekonometri, belirli bir zaman döneminde ya da zamanın bir noktasında nicel değişkenler arasındaki ilişkileri incelerken, niteliksel faktörlerin etkilerinin hangi metodoloji ile yapılacağına analizi ile de ilgilenen ekonometri dalıdır. Mikroekonometride metodolojik konular, belirli bir konuda sayısal yanıt isteyen önemli gerçekler etrafında düzenlenir. Mikroekonometrik uygulamaların kapsamı oldukça geniştir. Firma, hanehalkı ve birey gibi birimler hakkındaki verilerin farklı ekonometrik yöntemler ile incelenmesi mikro ekonometrinin kapsamındadır. İşgücü arzı, göç, tüketim davranışları, meslek seçimleri, firmaların üretim ve istihdam kararları mikroekonometrinin ilgilendiği konulara örnek olarak verilebilir. Literatür incelendiğinde, bireysel talep analizlerinde, iş gücü ekonomisinde, finansal alanda, emlak piyasasında, sağlık sektörü, bölgesel ekonomi gibi birçok alanda mikroekonometrik analizlerin kullanıldığı görülmektedir. Bu analizlere ilginin giderek artması ise yeni yöntemlerin ve testlerin çok hızlı bir şekilde gelişmesine de neden olmaktadır.

Bir ekonometrik modelde kullanılan değişkenler nicel ve nitel türde olabilir. Nicel değişkenler, gelir, harcama, fiyat, boy, yaş gibi bir ölçü birimi ile ölçülebilen ve sayılabilen değişkenlerdir. Nitel değişkenler ise ekonomik birimlerin genellikle demografik özellikleri, medeni durum, cinsiyet, ırk, din, meslek, eğitim durumu gibi kişisel özellikleri, belirli bir konudaki tutumları ya da ekonomik bir olayın bölgesel farklılıkları, mevsimsel farklılıkları gibi ölçülemeyen ve sayısal olarak ifade edilemeyen değişkenlerdir. Ekonometrik modellerde önemli olduğu ve modelde yer alması gerektiği düşünülen bu tür nitel değişkenleri modele yansıtmak için kukla değişken adı verilen bir yardımcı ya da yapay değişken türünden yararlanılır. Bunlar istenilen özelliğin varlığı için 1 değerini, diğer durum için 0 değerini alan kesikli değişkenlerdir. Örneğin işteki durum için tanımlanan bir değişken, kişi ücretli işte çalışıyorsa 1, ücretli işte çalışmıyorsa (ücretsiz aile işçisi gibi) 0 değerini alan kukla değişkendir. Başka bir örnek vermek gerekirse yaş nicel bir değişkendir. Ancak örnekleme yaş değişkeni 40 yaş altı ve 40 yaş üstü olmak üzere gruplandırmak istenirse kukla değişken oluşturulmuş olunur. Bu tür nitel değişkenler zaman serisi modellerinde de yer alabilir. Örneğin bir ekonomik kararın etkilerini araştırmak için alınan karardan önceki ve sonraki durum karşılaştırılabilir. Benzer şekilde bir salgın ya da doğal afet öncesi ve sonrası, bir sektördeki gelişmeler yine nitel bir değişken ile araştırılabilir.

1.1. Kukla Değişkenler

Bir regresyon analizinde kukla değişkenler temelde iki görev yüklenir:

- i) Bağımlı değişken olarak kullanılan kukla değişkenler
- ii) Bağımsız değişken olarak kullanılan kukla değişkenler

Kukla değişkenlerin açıklayıcı değişken olarak kullanıldıkları modeller genel regresyon prensiplerinden fazla farklılık göstermeyip en küçük kareler yöntemi (EKKY) ile parametreleri tahmin edilebilir. Kukla değişkenlerin bağımlı değişken olarak kullanıldıkları durumdaki modeller ise genel regresyon prensiplerinden teknik farklılıklar gösterir. Bağımlı değişkenin kukla olduğu modeller mikroekonometrik modelleri kapsadığından detaylarıyla bu kitabın ilerleyen bölümlerinde anlatılacaktır.

1.1.1. Bağımlı Değişkenin Kukla Olduğu Durumlar

Bağımlı değişkenin kukla değişken olarak kullanıldığı logit, probit, tobit modeli ve bunların türevlerinden oluşan çoklu tercih modelleri, sıralı logit ve probit modelleri, yuvalanmış logit vb. modellerin parametreleri EKKY ile tahmin edilememektedir. Bunlar, mikroekonometrik modeller olarak da adlandırılmaktadır. Mikroekonometride modeller bağımlı değişkenin aldığı değerlere bağlı olarak değişiklik göstermektedir. Bu bölümde mikroekonometride kullanılan veri tipleri ve bağımlı değişkenin aldığı değerlere göre kullanılan mikroekonometrik modeller tanıtılacaktır:

i) Sayma veriler (Count data): Bağımlı değişkenin verileri 1, 2, 3 gibi sayma verilerden oluşabilir ($y=0, 1, 2, 3, 4, 5, 6, \dots$ gibi). Örneğin tiyatro festivali boyunca tiyatroya giden izleyici sayısı, bir futbol maçında gol olan penaltıların sayısı, bir ay içinde meydana gelen otomobil kazası sayısı vb. Bağımlı değişkenin sayma verilerden oluşması durumunda, poisson regresyon, negatif binominal regresyon vb modeller kullanılmaktadır.

ii) Bağımlı değişken iki değer alabilir (Discrete, Binary) ($y=0, 1, 0, 0, 1, 0, 0, 0, 1, 1, 1, \dots$ gibi). Genellikle evet-hayır gibi nitel seçimleri temsil eder. İşgücü piyasasına katılma durumu, bir seçimde oy kullanma durumu, bir organizasyona üye olma durumu gibi.

Bağımlı değişkenin iki değer alması durumunda probit, logit modelleri kullanılır.

iii) Bağımlı değişken çok seçimli (tercihli) değerler alabilir. Çok seçimli değişkenlerin değerlerde sınıflayıcı (nominal) ve sıralayıcı (ordinal) ölçekte olabilir. Anket yoluyla toplanan verilerden bir yargı ifadesine beşli ölçekte verilen cevaplardan oluşturulan değerler (kesinlikle katılıyorum($y=1$), katılıyorum($y=2$), hiç katılmıyorum($y=5$)); küçük, orta, büyük işletme büyüklükleri gibi sıralayıcı değerleri bağımlı değişkenin alması durumunda sıralı logit, sıralı probit gibi modeller kullanılırken, bağımlı değişkenin işteki durum (ücretli, yevmiyeli, kendi hesabına çalışıyor olması vb), meslekler, ulaşım araçlarının seçimi, satış tipleri, marka tercihleri gibi sınıflayıcı (nominal) değerleri alması durumunda da çok durumlu logit ve probit modelleri kullanılır.

iv) Bağımlı değişkenin verilerinin sansürlenmiş (censored) ya da kırılmış (budanmış-truncated) olması durumunda kullanılacak modeller, Tobit, Heckman, Çift Eşik modelleridir. Sansürlü ya da kırılmış değişkenler genellikle süreklidir. **Kırılmış veri tipinde** bağımlı ve bağımsız değiş-

kenlere ilişkin veri yoktur. Bu değişkenlere ait veriler belli bir noktada kırılmıştır. Yoksulluk sınırı altındaki aileler, büyük ölçekli işletmelerde çalışanlar vb. örnek olarak verilebilir. **Sansürlü veri tipinde** bağımlı değişkene ait verilerin bir kısmı gözlemlenmemiş olup açıklayıcı değişkenlere ait veriler eksiksiz olarak gözlemlenmiştir. Örneğin Türkiye’de ailelerin eğlence-kültür harcamasının belirleyicileri ile ilgili bir çalışma yapılmak istendiğini varsayalım. Veriler anket yoluyla sağlanmış olsun. Anket döneminde bazı aileler eğlence-kültür harcaması yapmamış ya da hiç eğlence-kültür harcaması yapmıyor olabilir. Bu durumda bağımlı değişkene ait gözlemler sıfır olarak kaydedilir. Ancak bu ailelerin geliri, çocuk sayısı, hanehalkı reisinin öğrenim durumu vb özelliklerine ilişkin veriler kaydedilmiş olabilir. Mikroekonometrik çalışmalarda bağımlı değişkeni belirlemek son derece önemlidir. Çünkü bağımlı değişkenin aldığı değerlere göre ekonometrik model belirlenir.

Mikro ekonometride panel veri modelleri de son yıllarda sıklıkla kullanılmaya başlanmıştır. Panel veri modellerinde bağımlı değişkeni nitel olan modellere nitel bağımlı değişkenli panel veri modelleri denir. Nitel bağımlı panel veri modelleri içerisinde bağımlı değişkenin 1 ve 0 değerini aldığı iki uçlu modellere iki durumlu panel, ikiden fazla kategorinin olduğu modellere çok durumlu panel tercih modelleri denir. Bazen de kategorik değişkenler kendi doğasından gelen sıralı bir yapıya sahiptir. Bağımlı değişkenin sıralı ölçeğe sahip olduğu panel veri modelleri panel sıralı modeller olarak adlandırılır. Bu tür modeller, daha çok pazarlama, işletme, psikoloji gibi sosyal bilimlerin hemen her alanında birimlerin (firmalar, haneler, fertler) rassal olarak seçildiği anket çalışmalarında kullanılır.

1.1.2. Bağımsız Değişkenlerin Kukla Olduğu Durumlar ve Stata Uygulamaları

Sadece 0 ve 1 olmak üzere iki değer alabilen değişkenler iki değerli, dummy, kukla, yapay, gölge, yardımcı değişken gibi isimlerle adlandırılmaktadır. Kukla değişkenler ekonometride genellikle ölçülemeyen faktörlerin etkileri için kullanılır. Eğer ele alınan ilişkide bağımlı değişkeni etkileyen cinsiyet, meslek, medeni durum, bölge, eğitim seviyesi, iş tecrübesi, din, ırk, uyruk, savaşlar, grevler, siyasi karışıklıklar darbeler, hükümetin iktisadi politikasında değişiklikler, yangın, deprem, sel ve benzeri doğal afetler gibi niteliksel enformasyon sağlayan faktörler varsa bunların ekonometrik modelde ifade edilmeleri gerekebilir. Sadece sınıflama düzeyinde ölçülebilen niteliksel faktörlerin etkisini nicel bir hâle getirmek için kukla değişken tekniğine başvurulur. Ayrıca açıklayıcı değişken olarak yer alan bazı niceliksel faktörlerin sınıflandırılmasında da kullanılır. Örneğin kişilerin yaşı ya da bitirdikleri eğitim basamağı yıl olarak ölçüldüğünde nicel birer açıklayıcı değişken iken, sınıflandırılarak nitel hâle getirilebilir ve daha sade bir biçimde kukla değişkenlerle ifade edilebilir.

Doğrusal veya doğrusal olmayan bir modelde açıklayıcı değişken olarak yer alan kukla değişkenler genellikle, istenen özellik sağlandığında “1” sağlanmadığında ise “0” değerini alırlar. Kukla değişkenler, yardımcı değişken olup başlıca aşağıda ifade edilen dört amaca hizmet eder:

- 1) Regresyon analizinde niteliksel faktörlerin etkisini yansıtmak
- 2) Bazı nicel değişkenleri sınıflara ayırarak kesikli değişkene dönüştürüp modele dâhil etmek

- 3) Niteliksel faktörlerin etkisiyle yapısal parametrelerde meydana gelecek değişiklikleri temsil etmek
- 4) Zaman serisi verileri ile yapılan analizlerde, mevsimsel farklılıkların, dönemsel farklılıkların kısaca periyodik dalgalanmaların yol açtığı farklılıkları araştırmaktır.

Ayrıca bağımsız değişkenin sadece kukla değişken olduğu modeller varyans analizi modelleri mahiyetinde iken, kukla değişkenin yanı sıra nicel açıklayıcı değişkenin de yer aldığı modeller kovaryans analizi tipindedir. Yukarıdaki dört durumda da kukla açıklayıcı değişkenler regresyon modeline aşağıdaki varsayımlar altında ilave edilir:

i) Niteliksel faktörler, sadece fonksiyonun kaymasına neden olup, eğim parametresini değiştirmemektedir. Modele eklenen kukla değişken sabit terimdeki değişimi yansıtmaktadır. Bu da bazen sabit terim kuklası ya da kesişim kuklası olarak adlandırılabilir.

ii) Önsel iktisadi bilgilerimize göre, niteliksel faktörlerin sadece fonksiyonun kaymasına değil, aynı zamanda eğiminin de değişmesine yol açtığını düşünmek için haklı nedenler varsa, sabit terim kuklasının yanı sıra bir de eğim değişmelerini belirlemek için kukla değişken tanımlamak gerekir. Bu da eğim kuklası olarak adlandırılabilir.

iii) Her iki varsayım altında birden fazla niteliksel faktörün etkisi söz konusu ise birden fazla kukla değişken kullanılır. Bu durumda sabit etki ve karşılıklı etki modelleri söz konusu olur. Bu durum aşağıda sırayla açıklanacaktır.

Modele alınacak kukla değişken sayısı, sabit terimin var olup olmamasına göre değişmektedir. Bir kukla değişkenin m tane sınıfı varsa ve eğer modelde sabit terim de varsa m sınıfın $m-1$ tanesi modele kukla değişken olarak eklenir. Söz konu m sınıftan hangisinin referans kategori (karşılaştırma seçeneği) olacağı araştırmacıya bağlıdır. Modelde sabit terim yoksa m sınıfın hepsi modele dâhil edilir. Aşağıda kukla değişkenlere hangi durumlarda karşılaşılabileceği daha detaylı sınıflandırılmıştır:

- Sabit terimleri farklı, eğimleri aynı
- Sabit terimleri aynı, eğimleri farklı
- Sabit terim ve eğimleri farklı olabilir.

1. Durum: Sabit terimlerin farklı, eğimlerin aynı olması

Parametrelerdeki değişikliğin sadece sabit terimde bulunduğu yolunda ön bilgi varsa sabit parametrelerin farklı, fakat eğim katsayılarının değişmediği durum söz konusudur. Örneğin cinsiyet değişkeninin erkek ve kadın olmak üzere iki sınıfı vardır. Çalışmanın amacına göre örneğin erkeğin modelde olması isteniyorsa kullanılan kukla değişken aşağıdaki şekilde tanımlanır:

$$Y_i = \beta_1 + \beta_2 D_i + \beta_3 X_i + u_i \quad (1.1)$$

(1.1) no'lu eşitlik sabit terimin farklı eğimin aynı olduğu modeldir. Bu iki sınıfı temsilen kukla değişken tuzağına düşmemek için bir kukla değişken sınıfının tanımlanması yeterlidir. Aşağıda

sırasıyla önce tek bir kukla değişkenin olduğu varyans analiz modeli, daha sonra da bir kukla ve nicel değişkenin olduğu model örneğine geçilecektir. Veri setinde 15-65 yaş aralığında bireylerin ücretine etki eden değişkenler bulunmaktadır.

Tablo 1.1. bagimsiz_kukla.dta² Veri Seti

describe

obs:	1,211			
vars:	20			
variable name	storage type	display format	value label	variable label
yas	int	%8.0g		YAS
is_baslama_yil	int	%8.0g		IS_BASLAMA_YIL
ucret	long	%8.0g		GELIR_GECENAY_K
erkek	byte	%8.0g		cinsiyet== 1.0000
kadin	byte	%8.0g		cinsiyet== 2.0000
okul_bmemis	byte	%8.0g		okul_biten_k== 0.0000
ilkokul	byte	%8.0g		okul_biten_k== 1.0000
orta	byte	%8.0g		okul_biten_k== 2.0000
yuksekokul_univ	byte	%8.0g		okul_biten_k== 4.0000
yl_dr	byte	%8.0g		okul_biten_k== 5.0000
genel_lise	byte	%8.0g		okul_biten_k== 31.0000
meslek_lise	byte	%8.0g		okul_biten_k== 32.0000
egitim	float	%9.0g		
bekar	byte	%8.0g		medeni_durum== 1.0000
deneyim	float	%9.0g		
erkek_deneyim	float	%9.0g		
erkek_bekar	float	%9.0g		
erkek_evli	float	%9.0g		
lnucret	float	%9.0g		
evli	float	%9.0g		

Ücretin bağımlı değişken olduğu modelde, kullanılacak bağımsız değişkenler bireyin yaşı ve deneyim nicel değişken olmak üzere, kukla değişkenler cinsiyet (iki sınıflı), eğitim düzeyi (yedi sınıflı), medeni durum (iki sınıflı) şeklindedir. Kukla değişken sınıflarından biri model kurulurken referans sınıf olarak seçilecektir. Tanımlayıcı istatistikler Tablo 1.2.'de verilmiştir.

² bagimsiz_kukla.dta veri setine Uygulamalarla Mikroekonometri kitabının ilk sayfasında yer alan karekod numarasıyla ulaşabilirsiniz.

Tablo 1.2. Tanımlayıcı İstatistikler

sum yas erkek kadın okul_bmemis ilkokul orta genel_lise meslek_lise yüksekokul_univ
yl_dr egitim bekar evli deneyim ucret lnucret

Variable	Obs	Mean	Std. Dev.	Min	Max
yas	1,211	38.53509	9.690042	15	65
erkek	1,211	.625929	.484082	0	1
kadın	1,211	.374071	.484082	0	1
okul_bmemis	1,211	.0057803	.0758399	0	1
ilkokul	1,211	.0693642	.2541773	0	1
orta	1,211	.0858794	.2803017	0	1
genel_lise	1,211	.0825764	.2753546	0	1
meslek_lise	1,211	.0809249	.2728323	0	1
yüksekokul~v	1,211	.5400495	.4985994	0	1
yl_dr	1,211	.1354253	.3423186	0	1
egitim	1,211	4.325351	1.456196	0	6
bekar	1,211	.2146986	.4107827	0	1
evli	1,211	.7853014	.4107827	0	1
deneyim	1,211	8.95706	9.176496	0	41
ucret	1,211	4455.8	2738.773	100	26500
lnucret	1,211	8.236767	.5919422	4.60517	10.1849

Ücret modelinde değişen varyans sorunuyla karşılaşıldığından dirençli tahmincilerle çalışılmıştır.

Tablo 1.3. Bir Kukla Değişkenin Ücret Üzerindeki Etkisi (Varyans Analizi Modeli)

regress ucret erkek, robust

Linear regression				Number of obs	=	1,211
				F(1, 1209)	=	12.10
				Prob > F	=	0.0005
				R-squared	=	0.0091
				Root MSE	=	2727.3
ucret	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]	
erkek	541.1417	155.5927	3.48	0.001	235.88	846.4034
_cons	4117.084	115.2005	35.74	0.000	3891.069	4343.099

Tablo 1.3'te sunulan model tahmin sonuçlarına göre erkek değişkeninin tahmincisi istatistiki açıdan anlamlıdır ($p > |t|$). Cinsiyete göre ücret farklıdır. Bu durumda, kadın ve erkek için farklı sabit katsayılar elde edilmiştir. Model, erkeklerin ücretinin kadınların ücretinden 541.1417 birim daha fazla olduğunu söylemektedir. Bu durumda kadınların ücreti 4117.084 birim iken erkeklerin ücreti $(4117.084 + 541.1417 = 4658.2257)$ 4658.2257 birimdir. Model aşağıdaki gibi de yazılabilir:

Burada $D_i = \begin{cases} 1 & \text{Erkek} \\ 0 & \text{Kadın} \end{cases}$

$$\hat{Ücret}_i = 4117,08 + 541,14 \text{erkek}_i$$

Tablo 1.4'te görüldüğü üzere bu kez modele nicel değişken eklenmiştir.

Tablo 1.4. Bir Kukla Değişken ve Bir Nicel Değişken Durumunda Ücret Denklemi (Kovaryans Modeli)

regress ucet erkek deneyim,robust

Linear regression				Number of obs	=	1,211
				F(2, 1208)	=	108.67
				Prob > F	=	0.0000
				R-squared	=	0.1777
				Root MSE	=	2485.6

ucet		Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]

erkek		635.1863	140.3871	4.52	0.000	359.7566 910.616
deneyim		122.6293	8.811219	13.92	0.000	105.3424 139.9163
_cons		2959.82	117.7782	25.13	0.000	2728.748 3190.893

Model aşağıdaki gibi de yazılabilir.

$$\hat{Ücret}_i = 2959,82 + 635,19 \text{Erkek} + 122,63 \text{Deneyim}_i$$

Tablo 1.4'te verilen modelde iki katsayı da istatistiksel olarak anlamlıdır. Her iki katsayı içinde H_0 hipotezi reddedilmektedir. Model yorumlanabilir. Teorik kısımda anlatıldığı gibi, kadın ve erkek modellerinin sabit terimleri farklı, deneyim tahmincisi (eğim) aynıdır. Buna göre, aynı deneyime sahip iki çalışan, cinsiyet farkı nedeniyle 635.18 TL/ay farklı ücret almaktadır. Bu fark, katsayı işareti pozitif olduğu için erkeklerin lehinedir. Ücretin deneyime göre türevi, fazladan bir yıl daha çalışan bireyin, ücretinde meydana getireceği değişikliği göstermektedir. Bir yıl fazla deneyimin ücrete yaptığı etki ortalama 122.6293 TL/ay kadardır. Bu değer (eğim), kadın ve erkek için aynıdır. Yani kadın ve erkek bir yıl fazla çalıştığında 122.6293 TL/ay ücret artışı elde etmektedir.

2. Durum: Sabit terimlerin aynı, eğimlerin farklı olması

Bu durum sadece eğim katsayılarında farklılık olduğu yolunda ön bilgi varsa ortaya çıkar. Kukla değişkenin bulunduğu bir regresyon modelinde, eğimin değişmesi kukla değişkenin modeldeki nicel açıklayıcı değişkenle ya da değişkenlerle etkileşime girmesi anlamına gelmektedir. Eğimin değişmesine sebep olduğu düşünülen açıklayıcı değişkenle kukla değişken çarpılarak $D_i X_i$ etkileşim terimi modele eklenir. Literatürde bu değişkenin parametresi etkileşim parametresi olarak da anılmaktadır.

$$Y_i = \beta_1 + \beta_2 D_i X_i + \beta_3 X_i + u_i \quad (1.2)$$

(1.2) modeli yine EKKY ile tahmin edilir. Burada β_2 katsayısının anlamlılığına (diğ er bir ifadeyle sıfırdan farklılığına) t testiyle bakılır ve anlamlı çıkarsa, sınıflar arasında eğ im farklılığının olduğ una karar verilir. Bu da Y bağımlı değışkeninin X açıklayıcı değışkene göre eğ iminin iki sınıfta farklı olduğ u anlamına gelir. Eğı er yapısal değışim araştırılıyorsa yapısal değışimin (varsa diğ er etmenlerin yanı sıra) eğ im katsayılarından kaynaklandığı kabul edilir. Aşağıdaki Stata çıktısı (1.2) eşitliğinden yararlanarak oluşturulan bir modeldir. Erkek kuklası ile deneyim değışkeni çarpılarak etkileşim terimi oluşturulmuştur.

Tablo 1.5. Sabit Terim Aynı Eğ imin Farklı Olduğ u Ücret Denklemi

```
gen erkek_deneyim=erkek*deneyim
regress ucet deneyim erkek_deneyim,robust
```

Linear regression				Number of obs	=	1,211
				F(2, 1208)	=	99.50
				Prob > F	=	0.0000
				R-squared	=	0.1725
				Root MSE	=	2493.4

	ucret	Robust Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
	deneyim	99.13943	12.34932	8.03	0.000	74.91093 123.3679
erkek_	deneyim	37.55327	13.44287	2.79	0.005	11.17931 63.92723
	_cons	3364.004	95.40452	35.26	0.000	3176.827 3551.181

Tablo 1.5'teki modelde tahmincilerin tamamı %1 anlamlılık düzeyinde sıfırdan farklıdır. Görüldüğü gibi sabit terimler aynı, eğ im farklıdır. Buna göre, kadınların çalıştıkları fazladan her yıl 99.139TL/ay ücret farkına yol açarken; erkeklerin çalıştıkları fazladan her yıl, 136.692 TL/ay ücret farkına (eğ im farkı 37.55327 TL/ay) neden olmaktadır. Model aşağıdaki gibi de yazılabilir.

$$\hat{U}_{cret_i} = 3364 + 99,14Deneyim_i + 37,55(Erkek * Deneyim)$$

3. Durum: Sabit terim ve eğ imin farklı olması durumu

Modelde sabit terim ve eğ imlerin farklı olması hâlinde (1.3) no'lu eşitlik yazılabilir.

$$Y_i = \beta_1 + \beta_2 D_i + \beta_3 (D_i X_i) + \beta_4 X_i + u_i \quad (1.3)$$

Sabit terim farkı β_2 katsayısının t istatistiğiyle, eğ im farkı β_3 katsayısının t istatistiğiyle araştırılır. Eğı er belli bir anlamlılık düzeyinde sıfırdan farklı iseler, iki sınıfın fonksiyonları arasında bir fark vardır denir. (1.3) eşitliği dikkate alınarak aşağıdaki ücret modeli kurulmuş, tahmin sonuçları Tablo 1.6'da verilmiştir:

Tablo 1.6. Sabit Terim ve Eğimin Farklı Olduğu Ücret Denklemi**regress ucet erkek deneyim erkek_deneyim, robust**

Linear regression

Number of obs = 1,211

F(3, 1207) = 71.98

Prob > F = 0.0000

R-squared = 0.1778

Root MSE = 2486.5

ucet	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]	
erkek	576.9897	190.3443	3.03	0.002	203.5474	950.4321
deneyim	118.7511	14.74755	8.05	0.000	89.81736	147.6847
erkek_deneyim	6.369224	18.38905	0.35	0.729	-29.70883	42.44728
_cons	2996.42	144.3142	20.76	0.000	2713.285	3279.555

Model aşağıdaki gibi de yazılabilir.

$$\text{Ücret}_i = 2996,42 + 576,99\text{Erkek} + 118,75\text{Deneyim}_i + 6,37(\text{Erkek} * \text{Deneyim}_i)$$

Tablo 1.6'daki modelde etkileşim teriminin katsayısı istatistiksel olarak anlamsız çıkmıştır (t istatistiği=0.35). Ancak etkileşim teriminin katsayısının anlamlı çıktığı varsayılarak yorum yapıldığında, kadınlar ilk işe başladıklarında, hiç deneyimleri yokken 2996.42 TL/ay ücret almaktadır. Çalıştıkları her fazladan yıl için kadınlara 118.751 TL/ay ücret artışı yapılmaktadır. Erkekler ilk işe başladıklarında, hiç deneyimleri yokken 3573.4097 TL/ ücret (sabit terim farkı=576.9887) almaktadır. Çalıştıkları her fazladan yıl için, erkeklere 125.1203 TL/ay ücret artışı (eğim farkı=6.369294) yapılmaktadır. erkek_deneyim etkileşim teriminin katsayısı da istatistiki olarak anlamsızdır. Araştırmanın amacına bağlı olarak modelden etkileşim terimi çıkartıldığında ise yukarıdaki Tablo 1.6'daki çıktı, sabit terimin farklı eğimin aynı olduğu modele dönüşür. Ancak bu durum iktisat teorisi açısından irdelenmesini ve yorumlanmasını araştırmacıya bırakılmıştır.

1.2. Kukla Değişkenlerin Farklı Kullanımları

Aşağıda kukla değişkenlerin farklı kullanımlarıyla ilgili birden fazla kukla değişken durumu, birden fazla nicel açıklayıcı değişken, kukla değişkenlerin birbiri ile etkileşime girmeleri, kukla değişkenlerde Chow testi ve mevsimlik etkinin ölçülmesi incelenecektir.

1.2.1. Birden Fazla Kukla Değişken Durumu

İkiden fazla sınıf olması anlamına gelmektedir. Sınıflardan biri dışlanır buna referans sınıf veya temel sınıf denir. Dışlanan sınıfın etkisi sabit terime yansır. Her bir kukla değişkenin katsayısı, kendi temsil ettiği sınıf ile dışlanan sınıf arasındaki farkı gösterir. Örneğin, sağlık harcamalarında nicel açıklayıcı değişken olan gelirin yanı sıra kişilerin eğitimlerinin de etkisi olduğu düşünülürse aşağıdaki (1.4) eşitliğindeki gibi bir model oluşturulabilir:

$$Y_i = \beta_1 + \beta_2 X_i + a_1 D_1 + a_2 D_2 + a_3 D_3 + u_i \quad (1.4)$$

Burada,

Y_i : kişisel bakım ve sağlık harcamaları

X_i : harcanabilir gelir

D_1 1: *ilköğretim*, 0: *diğerleri*

D_2 1: *orta öğretim* 0: *diğerleri* D_3 1: *yüksek öğretim*, 0: *diğerleri*

Çoklu bağlantı sorunu nedeniyle sınıflardan biri referans sınıf olarak kabul edilir ve model (1.5)'deki gibi yazılır. Örneğin ilköğretim (8 yıllık eğitim) sınıfı dışlansın:

$$Y_i = \beta_1 + \beta_2 X_i + a_2 D_2 + a_3 D_3 + u_i \quad (1.5)$$

Katsayıların yorumu şöyle olacaktır: β_1 : regresyonun genel prensibine göre diğer değişkenlerin değeri sıfırken Y 'nin alacağı ortalama değeri vereceğinden, gelir sıfırken dışlanan grubu yani ilköğrenim mezunları için ortalama sağlık harcamaları düzeyini gösterecektir. β_2 : yine regresyonun genel prensibine göre (ceteris paribus), niteliksel faktörün bütün düzeyleri için, eğitimin tüm düzeyleri için marjinal harcama eğilimini, yani gelirden meydana gelecek bir birimlik değişikliğin sağlık harcamalarında yaratacağı değişikliği (artışı) gösterir. α_2 : aynı gelir düzeyinde, ortaöğrenim (lise) mezunlarının ilköğrenim mezunlarına göre sabit terim farkını diğer bir ifadeyle sağlık harcamaları farkını verir. α_3 : aynı gelir düzeyinde, yükseköğrenim yapmış olanların ilköğrenim mezunlarına göre yaptıkları sağlık harcamalarını verir.

1.2.2. Birden Fazla Nicel Açıklayıcı Değişken, Birden Fazla Kukla Değişken Durumu

Birden fazla nicel açıklayıcı değişken birden fazla kukla değişken bulunduğu niteliksel faktörü temsilen kukla değişkenler modele yukarıda açıklanan genel prensipler göz önüne alınarak dâhil edilirler. Bu kukla değişkenlerin iki sınıfı olabileceği gibi ikiden fazla da olabilir ve kukla değişkenler modeldeki sürekli açıklayıcı değişkenlerle etkileşim içerisinde bulunabilir. Bu durum eğitim değişimlerinden başka bir şey değildir. Ayrıca kukla değişkenler kendi aralarında da etkileşimde bulunabilir. Buna da karşılıklı etki durumu denir. Karşılıklı etki durumu daha sonra ele alınacaktır.

$$Y_i = \beta_1 + \beta_2 X_{1i} + \beta_3 X_{2i} + a_1 D_1 + a_2 D_2 + \gamma_1 D_1 X_{1i} + \gamma_2 D_1 X_{2i} + \gamma_3 D_2 X_{1i} + \gamma_4 D_2 X_{2i} + u_i \quad (1.6)$$

1.2.3. Kukla Değişkenlerin Birbiri ile Etkileşime Girmeleri

Şimdiye kadarki anlatımlarda niteliksel faktörlerin birbirleri ile etkileşime girmedikleri varsayımı bulunuyordu. Birçok durumda bu varsayım geçerli olamayabilir.

$$Y_i = \beta_1 + \beta_2 X_i + a_2 D_2 + a_3 D_3 + u_i \quad (1.7)$$

(1.7) eşitliğinde Y : giyim harcamalarını, X : geliri, D_2 : cinsiyeti ve D_3 : bölgesel farklılığı temsil etsin. Kukla değişkenler aşağıdaki gibi tanımlansın.

$$D_2 \text{ 1: kadın, 0: erkek } D_3 \text{ 1: büyük kentler, 0: küçük kentler}$$

Burada örtük varsayım cinsiyetin etkisinin bölgeye göre değişmediğidir. Yani kadın olmanın giyim harcamaları üzerinde yaratacağı etkinin hem büyük şehirlerde hem de küçük yerleşim yerlerinde aynı olacağı kabul ediliyordu. Ya da büyük kentli olmanın giyim harcamaları üzerinde yaratacağı etkinin hem kadınlarda hem de erkeklerde aynı olduğu kabul ediliyordu. Oysaki bu-

yük kentlerdeki kadınların giyim harcamaları farklı olabilir. Bu durumda her iki kukla değişkenin çarpımından oluşan etkileşim terimi (D_2D_3)'nin modele dâhil edilmesi gerekir. Yeni model

$$Y_i = \beta_1 + \beta_2 X_i + a_1 D_2 + a_2 D_3 + \gamma(D_2 D_3) + u_i \quad (1.8)$$

hâlini alır.

Tablo 1.7'de, eşitlik (1.9)'da iki kukla değişkenin bulunduğu etkileşim terimli model örneğinin tahmini verilmiştir.

$$Y_i = \beta_1 + a_1 D_2 + a_2 D_3 + \gamma(D_2 D_3) + u_i \quad (1.9)$$

Burada Y ücreti(TL), D_2 birey erkek ise 1, değilse 0 değerini, D_3 birey bekâr ise 1 değilse 0 değerini alan kukla değişkenleri göstermektedir. Modelde değişen varyans olduğu için dirençli tahmincilerle çalışılmıştır.

Tablo 1.7. İki Kukla Değişkenin Bulunduğu Etkileşim Terimli Ücret Modeli

gen erkek_bekar=erkek*bekar

regress ucet erkek bekar erkek_bekar, robust

Linear regression

Number of obs = 1,211

F(3, 1207) = 18.12

Prob > F = 0.0000

R-squared = 0.0375

Root MSE = 2690.3

ucet	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]	
erkek	697.3277	181.1964	3.85	0.000	341.8329	1052.822
bekar	-662.9103	244.9614	-2.71	0.007	-1143.508	-182.3129
erkek_bekar	-649.2273	326.1938	-1.99	0.047	-1289.197	-9.257354
_cons	4251.715	133.9875	31.73	0.000	3988.84	4514.589

Ücret, erkek ve bekâr kukla değişkenlerin ortak etkileşim terimiyle (erkek_bekar) tahmin edildiğinde, modelin tüm katsayıları %5 anlamlılık düzeyinde istatistiksel olarak sıfırdan farklıdır. Erkeklerin, kadınlara göre ve evlilerin bekârlara göre ortalama olarak daha fazla ücret aldığı söylenebilir. Daha detaylı yorumlanmak istenirse, ücret denkleminde $D_2=0$, $D_3=0$ olduğunda evli kadınların alacağı ortalama ücret 4251,71 TL olarak bulunur. Benzer şekilde bekâr kadınların alacağı ortalama aylık ücret 3588.80 TL ($4251,715-662,91=3588,80$ TL) bulunur. Buradan bekâr ve evli kadın ücret farkı ($4251,715-3588,80 = -662,91$ TL) 662,91 TL'dir. Evli kadınlar bekâr kadınlardan ortalama olarak 662,91 TL daha fazla ücret alıyor. Bu fark ta D_3 bekâr değişkeninin katsayısıdır.

Yine aynı denklemde $D_2=1, D_3=1$ konursa bekâr erkeklerin alacağı ortalama aylık ücret 3636.85 TL ($4251,715+697,3277-662,9103-649,2273=3636,85$) olarak bulunur. Evli erkeklerin alacağı ortalama aylık ücret de 4949,042 TL'dir. $4949,042-3636,85=1312,19$ TL evli erkek ve bekâr erkek ücreti arasındaki farktır. Bekâr erkek 1312,19 TL daha az ücret almaktadır. O hâlde, 697,32 katsayısı evliler için erkek-kadın ücret farkını gösterir. Evli erkek evli kadından 697,32

TL daha fazla ücret almaktadır. -662,91 katsayısı kadınlar için evli-bekar farkını gösterir. Bekâr kadın evli kadından ortalama olarak 662.91 TL daha az ücret alıyor. $1312,19 - 662,91 = 649,28$ ise etkileşim teriminin katsayısıdır.(yuvarlama farkları olabilir)

$\gamma = -649.23$; etkileşim parametresi (interaction terim) olup her iki kukla değişkenin 1 değerini aldığı gruplar için etki farkıdır. -649.23 katsayısı bekâr ve erkek olmanın yarattığı fark etkisidir. Daha açık bir ifadeyle: bekâr olmak kadınlarda ve erkeklerde daha az ücret almaya yol açmaktadır. Ancak erkeklerde bekâr olmanın yarattığı fark etkisi kadınlarda yarattığı fark etkisine göre (649.23) bu kadar daha fazladır.

$$\hat{Y}_i = 4251,715 + 697,3277D_2 - 662,9103D_3 - 649.2273(D_2D_3)$$

ya da

$$\hat{Ücret}_i = 4251,715 + 697,3277Erkek - 662,9103Bekar - 649,2273(Bekar \times erkek)$$

Tablo 1.8. Kukla Değişken Kategorilerinin İki'den Fazla Olması Durumunda Ücret Denklemi

regress lnucret yas ilkokul orta genel_lise meslek_lise yüksekokul_univ yl_dr bekar deneyim

Linear regression						
Source	SS	df	MS	Number of obs	=	1,211
Model	226.131231	10	22.6131231	F(10, 1200)	=	137.15
Residual	197.847428	1,200	.164872857	Prob > F	=	0.0000
				R-squared	=	0.5334
				Adj R-squared	=	0.5295
Total	423.978659	1,210	.350395586	Root MSE	=	.40605
lnucret	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
erkek	.2573416	.024579	10.47	0.000	.2091189	.3055642
yas	-.0016657	.0016493	-1.01	0.313	-.0049016	.0015702
ilkokul	-.0929424	.1605605	-0.58	0.563	-.4079529	.2220682
orta	-.0063775	.1591291	-0.04	0.968	-.3185797	.3058247
genel_lise	.1368064	.1591911	0.86	0.390	-.1755174	.4491303
meslek_lise	.2393276	.1593708	1.50	0.133	-.0733489	.5520041
yüksekokul_univ	.6972986	.1552583	4.49	0.000	.3926907	1.001907
yl_dr	1.014311	.1574252	6.44	0.000	.7054521	1.323171
bekar	-.2155088	.0322002	-6.69	0.000	-.2786838	-.1523338
deneyim	.0170473	.0016966	10.05	0.000	.0137187	.0203759
_cons	7.495844	.1633949	45.88	0.000	7.175272	7.816415

Tablo 1.8'deki yaş ve deneyim değişkenleri nicel değişkenlerdir. Yani açıklayıcı değişkendeki bir birim değişmeye karşılık ücret oranındaki nisbi (ya da yüzde) değişimi vermektedir. Diğer değişkenlerin etkisi sabit tutulduğunda, işte çalışılan süre arttıkça, bir yıl daha fazla deneyimli olanın ücreti yüzde bir artmaktadır. Bireyin yaşının ücret üzerindeki marjinal katkısı ise yaş ilerledikçe azalmaktadır. t istatistiği çalışılan örneklem için anlamsız çıkmıştır. Erkeklerin ücreti ile kadınların ücreti arasındaki yüzde değişim bulunmak istendiğinde kukla değişken katsayısının ters logaritması alınmalıdır. Bu durumda, erkeğin ücreti kadının ücretinden

$(antilog_e(0.2573) = e^{0.2573} = 1.2934, (1.2934-1=0.2934))$ %29 daha fazladır. Eğitim düzeylerine bakıldığında toplamda altı sınıf vardır. Okuryazar olmayan birey, kukla değişen tuzağına düşmemek için dışarıda tutulmuştur ve referans sınıf olarak alınmıştır. Bireyin ilkökul, orta, genel lise ve meslek lisesi sınıflarındaki değişkenlerin katsayılarının t istatistik değerlerine bakıldığında istatistiki olarak anlamsız, yüksekokul, üniversite, yüksek lisans ve doktora kategorilerinin katsayıları istatistiki olarak anlamlı çıkmıştır. Üniversite, yüksek lisans ve doktora yapmış olanların okul bitirmemiş olana göre ücretinde artış vardır. Bu durumda, üniversite mezunu olan bir bireyin ücreti, okul bitirmemiş bir bireyin ücretinden yaklaşık %1.008 daha fazladır $(antilog_e(0.6972) = e^{0.6972} = 2.00812; (2.00812-1=1.00812))$. Yüksek lisans ve doktora yapmış olanın bir birey okul bitirmemiş bir bireyden yaklaşık %1.75 daha fazla ücret almaktadır $(antilog_e(1.014311) = e^{1.014311} = 2.7574, (2.7574-1=1.7574))$.

1.2.4. Kukla Değişkenlerde Chow testi

Chow testi ile sınanan hipotez kesit verilerinde kukla değişkenler eklenerek de araştırılabilir. Birinci yol: Aşağıdaki örnekte olduğu gibi kazanılan ücrette cinsiyete göre değişiklik olup olmadığı sınanmak istensin. İlk olarak örneklemden sadece kadınlar, daha sonra erkekler çekilerek kadınlar ve erkekler örneklemlerinden ve kadın erkek ayrımı yapmadan oluşturulan örneklemden üç farklı regresyon denklemi tahmin edilir. Tahminlenen üç regresyon modelinin hata kareler toplamaları hesaplanır. Daha sonra Temel Ekonometride bilinen Chow F testi formülünden yararlanarak F hesap istatistiği hesaplanır. Hipotez, kadın ve erkek bireylerin regresyon denkleminde katsayıları arasında bir farklılık yoktur (cinsiyete göre ücretlerde farklılık yok) şeklindedir. Eğer sıfır hipotezi reddedilmezse modele cinsiyet kuklası eklenmeyebilir. İkinci yol ise modele cinsiyet kuklası ekleyerek Tablo 1.9'daki 4 no'lu modelden t istatistiğini incelemek ve karar vermektir. Aşağıda iki yola göre modeller incelenmiştir:

Birinci yol:

Tablo 1.9. Ücret³ Denkleminin Chow testiyle İncelenmesi

1. no'lu denklem

regress lnucret yas egitim if erkek==1 (Erkeklerle ait ücret denklemi)

Source	SS	df	MS	Number of obs	=	758
Model	104.925765	2	52.4628826	F(2, 755)	=	264.68
Residual	149.651351	755	.19821371	Prob > F	=	0.0000
				R-squared	=	0.4122
				Adj R-squared	=	0.4106
Total	254.577116	757	.336297379	Root MSE	=	.44521
lnucret	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
yas	.0085094	.001624	5.24	0.000	.0053213	.0116975
egitim	.2457298	.0109379	22.47	0.000	.2242575	.267202
_cons	6.930636	.0795471	87.13	0.000	6.774476	7.086796

³ Bağımsız_kukla.dta verisini Uygulamalarla Mikro Ekonometri kitabının linkinden indirebilirsiniz.

2. no'lu denklem

regress lnucret yas egitim if erkek==0(Kadınlara ait regresyon denklemi)

Source	SS	df	MS	Number of obs	=	453
Model	86.1738858	2	43.0869429	F(2, 450)	=	245.90
Residual	78.8505283	450	.175223396	Prob > F	=	0.0000
Total	165.024414	452	.365098261	R-squared	=	0.5222
				Adj R-squared	=	0.5201
				Root MSE	=	.4186

lnucret	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
yas	.0202322	.0021398	9.46	0.000	.0160269 .0244375
egitim	.2956053	.0143078	20.66	0.000	.2674869 .3237238
_cons	6.023839	.1106065	54.46	0.000	5.80647 6.241208

3. no'lu denklem denklem

regress lnucret yas eğitim (Erkek-Kadın birarada)

Source	SS	df	MS	Number of obs	=	1,211
Model	175.697834	2	87.8489172	F(2, 1208)	=	427.43
Residual	248.280825	1,208	.205530484	Prob > F	=	0.0000
Total	423.978659	1,210	.350395586	R-squared	=	0.4144
				Adj R-squared	=	0.4134
				Root MSE	=	.45335

lnucret	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
yas	.0123213	.0013456	9.16	0.000	.0096812 .0149613
egitim	.2510611	.0089544	28.04	0.000	.2334933 .2686289
_cons	6.676038	.0669576	99.71	0.000	6.544672 6.807404

$$F = \frac{(248.2808 - (149.6513 + 78.8505))/3}{(149.6513 + 78.8505)/(758 + 453 - 2.3)} = 34.7$$

di invF(3,1205,0.95)
2.6122878

di Ftail(3,1205,24.224)
3.185e-15

Stata'dan veya F tablosundan kritik tablo değerine bakıldığında sıfır hipotezi reddedilir (F test istatistiği 34.7), iki denklem birbirinden farklıdır. Diğer bir ifadeyle erkeklerin ve kadınların ücretlerini etkileyen değişkenlerde yapısal farklılık vardır.

İkinci yol:

4. no'lu denklem

regress lnucret yas egitim erkek

Source	SS	df	MS	Number of obs	=	1,211
Model	190.795532	3	63.5985106	F(3, 1207)	=	329.20
Residual	233.183128	1,207	.193192318	Prob > F	=	0.0000
				R-squared	=	0.4500
				Adj R-squared	=	0.4486
Total	423.978659	1,210	.350395586	Root MSE	=	.43954

lnucret	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
yas	.0123785	.0013046	9.49	0.000	.0098189 .0149381
egitim	.2617526	.0087653	29.86	0.000	.2445558 .2789495
erkek	.2329795	.0263547	8.84	0.000	.1812735 .2846856
_cons	6.481759	.0685358	94.57	0.000	6.347297 6.616222

İkinci yolda kukla değişken ekleyerek cinsiyete göre ücretlerin farklı olup olmadığı incelenir. Erkek değişkeninin katsayısı istatistiksel olarak anlamlıdır. Yine cinsiyete göre ücret değişmektedir yorumu yapılabilir. Sonuç olarak üç modele ayrı ayrı regresyon uygulamaktansa, ikinci yolda olduğu gibi modelde kukla değişkenin katsayısının anlamlılığına bakarak bir farklılık olup olmadığı yorumu daha kolay yapılır.

Kukla değişkenlerle çalışarak yıllık, üç aylık, aylık etkiler ölçülebilir. Ekonomik resesyon veya kırılma dönemlerindeki etkiler ölçülmek istenirse, örneğin 2008 kriz öncesi dönemine sıfır, sonrasına bir değeri verilerek kukla değişkenler oluşturulabilir. Çalışmanın amacına ve serisine bağlı olarak kukla sınıfı veya sınıfları modele dâhil edilir. Gözlem sayısının yeterli büyüklükte olmasının serbestlik derecesi kaybına ve modelin anlamsız olmasına neden olabileceği unutulmamalıdır. Aşağıdaki örnekte otomobil satışlarının hangi üç aylık seriden etkilendiğini görebilmek için dört tane mevsim kuklası (q_1 , q_2 , q_3 ve q_4) oluşturulmuştur. Dördüncü üç aylık dönem (q_4) referans sınıftır ve rastlantısal belirlenmiştir. 1995.1-2014.4 yıllarında ($n=64$) bir ülkedeki otomobil satış adedine etki eden değişkenler aşağıda listelenmiştir. log-log model ile çalışılmış ve hatalarda değişen varyans sorununa rastlanmamıştır.

Tablo 1.10. mevsimlik_kukla.dta⁴ Veri Seti

describe

obs: 64
vars: 4

variable name	storage type	display format	value label	variable label
yıl	float	%8.0g		
Satısadedi	int	%8.0g		Satıs adedi

⁴ mevsimlik_kukla.dta veri setine Uygulamalarla Mikro Ekonometri kitabının linkinden ulaşabilirsiniz.

gelirsabitfiy	float	%8.0g	gelir(sabit fiy)
stoksay	float	%8.0g	sotomobil stok sayısı
nufus	float	%8.0g	
q1	byte	%8.0g	Q1
q2	byte	%8.0g	Q2
q3	byte	%8.0g	Q3
q4	byte	%8.0g	Q4
otofiyat_indeks	float	%8.0g	otomobil fiyat indeksi
lnsatis	float	%9.0g	
lngelir	float	%9.0g	
lnnufus	float	%9.0g	
lnfiyat	float	%9.0g	

Tablo 1.11. Değişkenlerin Tanımlayıcı İstatistikleri

Variable	Obs	Mean	Std. Dev.	Min	Max
sum					
yıl	64	2005.297	5.552577	1995.1	2014.4
Satisadedi	64	2488.594	332.922	1754	3337
gelirsabit~y	64	10.52136	.8456622	8.985	11.93
stoksay	64	109.7724	8.928266	93.145	123.3
nufus	64	233.7733	10.48944	215.973	251.966
q1	64	.25	.4364358	0	1
q2	64	.25	.4364358	0	1
q3	64	.25	.4364358	0	1
q4	64	.25	.4364358	0	1
otofiyat_i~s	64	95.2125	18.94685	60.2	121.4
lnsatis	64	7.810598	.1347831	7.469654	8.112827
lngelir	64	2.350251	.0799537	2.195556	2.479056
lnnufus	64	5.453359	.0449592	5.375154	5.529294
lnfiyat	64	4.535015	.2113652	4.097672	4.799091

Tablo 1.12. Otomobil Satışlarına Ait Model

regress lnsatis lnfiyat lngelir q1 q2 q3

Source	SS	df	MS	Number of obs	=	64
Model	.490326931	5	.098065386	F(5, 58)	=	8.69
Residual	.654161896	58	.011278653	Prob > F	=	0.0000
Total	1.14448883	63	.018166489	R-squared	=	0.4284
				Adj R-squared	=	0.3792
				Root MSE	=	.1062

lnsatis	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
lnfiyat	-.8193509	.1687627	-4.86	0.000	-1.157166 - .4815357
lngelir	2.38926	.446017	5.36	0.000	1.49646 3.28206
q1	.0239058	.0376059	0.64	0.527	-.0513706 .0991822
q2	.1338949	.0375757	3.56	0.001	.058679 .2091109
q3	.0537993	.0375624	1.43	0.157	-.0213899 .1289885
_cons	5.858107	.445765	13.14	0.000	4.965811 6.750402

Tablo 1.12'deki görüldüğü üzere, t istatistik değerlerine bakıldığında birinci üç aylık dönemin ve üçüncü üç aylık dönemin otomobil satışları üzerinde bir etkisi olmadığı belirlenmiştir. Ancak ikinci üç aylık dönemin satışlar üzerinde etkisi vardır. t istatistiği anlamlıdır (prob değeri=0.001<0.01). Detaylı incelendiğinde, ikinci üç aylık dönemin satışları dördüncü üç aylık dönemin satışlarından ($\text{antilog}_e(0.1338949) = e^{0.1338949} = 1.14327$) % 14 daha fazladır. Nicel değişkenlerden gelir değişkeninin etkisine bakıldığında gelirdeki %1'lik artış otomobil satışlarını %2.38 artırmaktadır. Bireylerin gelirinin otomobil satışları üzerinde istatistiksel olarak anlamlı bir etkisi vardır. Otomobil fiyat endeksindeki yüzde bir artış, otomobil satışlarını yüzde 0.81 oranında düşürmektedir. Otomobil fiyat endeksinin satışlar üzerinde istatistiksel olarak anlamlı bir etkisi vardır.

1. Bölümde kullanılan birinci örnek komutları(bağımsız_kukla.dta)

```
use "C:\bagimsiz_kukla.dta"*
describe
sum yas erkek kadın okul_bmemis ilkokul orta genel_lise meslek_lise yüksekokul_univ yl_dr
egitim evlenmdi evli deneyim ucet lnucet
regress ucet erkek, robust
regress ucet erkek deneyim, robust
gen erkek_deneyim=erkek*deneyim
regress ucet deneyim erkek_deneyim, robust
regress ucet erkek deneyim erkek_deneyim, robust
gen erkek_bekar=erkek* evlenmdi
regress ucet erkek evlenmdi erkek_bekar, robust
regress lnucet yas ilkokul orta genel_lise meslek_lise yüksekokul_univ yl_dr evlenmdi deneyim
regress lnucet yas egitim if erkek==1
regress lnucet yas egitim if erkek==0
regress lnucet yas egitim
di invF(3,1205,0.95)
di Ftail(3,1205,24.224)
regress lnucet yas egitim erkek
```

1. Bölümde kullanılan ikinci örnek komutları(mevsimlik_kukla.dta)

```
use "C:\mevsimlik_kukla.dta"
describe
sum
regress lnsatis lnfiyat lngelir q1 q2 q3
```

KAYNAKLAR

Long J S, 1997. Regression Models for Categorical and Limited Dependent Variables. Sage Publications. Stock, J H., Watson M W., 2011. Ekonometriye Giriş, (Çeviren: Bedriye Saraçoğlu) Efil Yayınevi. Gujarati D. 2016. Örneklerle Ekonometri, (Çeviren: Nasip Bolatoğlu) İkinci Edisyon, Tarcan Matbaası. Wooldridge, J.M. (2013), Ekonometriye Giriş, (Çeviren: Ebru Çağlayan), Cilt 1, 4.Basım, Nobel Akademik Yayıncılık.

BÖLÜM 2

EN ÇOK OLABİLİRLİK TAHMİNCİSİ, DOĞRUSAL OLMAYAN REGRESYON MODELİ VE STATA UYGULAMALARI

Burak Alparslan Eroğlu¹

Kitabın bu bölümünde ekonometrik/istatistiki tahmin yöntemlerinden *En Çok Olabilirlik (EÇO)* yöntemi üzerinde durulacaktır. EÇO tahmin yöntemi, istatistik ve ekonometri yazınının parametrik tahmin ediciler sınıfında olup, sosyal ve fen bilimleri alanında birçok modelin incelenmesinde sıklıkla kullanılmaktadır. Bu yöntem, uygulayıcıların elindeki veri setini parametrik bir dağılım fonksiyonuyla ilişkilendirmek suretiyle, modeldeki bilinmeyen anakütle parametrelerinin belirlenmesi işlemi şeklinde ortaya çıkar. Dolayısıyla, EÇO yönteminin uygulamasında elimizdeki veri setinin dağılımı hakkında ön bilgi ya da varsayım ile model parametrelerinin bu dağılımla nasıl etkileştiği bilgisine ihtiyaç duyulur.

EÇO tahmin yönteminin mantığını anlatmak için basit bir örnek kullanılabilir. Analizde incelenecek bağımlı değişken n adet gözlemi bulunan $\{y_i\}_{i=1}^n$ serisi olsun. Bu değişkenin θ parametre kümesiyle ilişkili D dağılımından geldiği düşünölsün. Dolayısıyla, her bir $i \in \{1, \dots, n\}$ için $y_i \sim D(\theta)$ şeklinde yazılabilir. Bu yapıda D dağılımı bilinse dahi, gerçek anakütle parametre kümesi θ 'nın bilinmediği varsayılacaktır. Eldeki verileri yukarıda tanımlanan dağılımdan gelen sadece bir gerçekleşme olarak düşünölebilir. Yani, birçok alternatif içinden eldeki $\{y_i\}_{i=1}^n$ serisi ortaya çıkmış olarak kabul edilir. Burada, eldeki serilerin ortaya çıkması ihtimalinin bilinmeyen anakütle parametre kümesine bağılı olduğunu hatırlanmalıdır. Böylece, şu soru sorulabilir: Eldeki veri setinin gerçekleşme *ihhtimalini* en üst düzeye çıkaran θ değeri nedir? Bu soruyu sormaktaki neden, “eldeki gerçekleşen veri seti, ortaya çıkma ihtimali yüksek olduğu için karşımıza çıkmıştır” öngörüsü olduğu söylenebilir. Bu sorunun cevabının matematiksel olarak verilmesi EÇO tahmin edicisini ortaya çıkaracaktır.

EÇO tahmin edicisinin önemli bir özelliğı, hem doğrusal hem de doğrusal olmayan modellerde kullanılabilmesidir. EÇO tahmin edicisinin modelin yapısı üzerine bir sınırlaması yoktur. Bu yöntemin uygulanabilmesi için tek kısıt, ekonometrik veya istatistiki modelin dağılım özelliklerinin bilinmesidir. Böylece, bu yöntem, gayet esnek bir kullanım alanı olan tahmin edici sınıfı

¹ Dr. Öğretim Üyesi, İstanbul Bilgi Üniversitesi, Ekonomi Bölümü. Eski Silahtarağı Elektrik Santrali Kazım Karabekir Cad. No: 2/13 1.1 203 34060 Eyöpsultan İstanbul, email: burak.eroglu@bilgi.edu.tr.

olarak karşımıza çıkmaktadır. Özellikle, mikro-ekonometri alanında birçok model, EÇO tahmin edicisi yardımıyla analiz edilmektedir. Mesela, seçim modellerinden probit ve logit modelleri, ölçek tabanlı faktör modellemesinde kullanılan yapısal eşitlik modelleri ve saymalı değişkenlerin analizinde kullanılan Poisson ve negatif binom regresyon modelleri, EÇO tahmin edicisi kullanılarak incelenmektedir. Dolayısıyla, bu yöntemin doğasını ve özelliklerini anlamak özellikle önem arz etmektedir. Kitabın bu kısmında yukarıdaki konulara değinilerek, EÇO yönteminin önemi vurgulanmaya çalışılacaktır.

Bu bölümde EÇO prensibinin matematiksel bir tanımlaması yapıp, EÇO prensibiyle elde edilen tahmin edicilerin özellikleri üzerinde durulacaktır. Sonrasında bu yöntemin Stata yazılımı kullanılarak nasıl uygulanabileceği üç farklı örnek üzerinde tartışılacaktır. Bölümün birinci kısmında, EÇO yönteminin formal bir tanımlaması yapılacaktır. İkinci kısımda, doğrusal olmayan ekonometrik/istatistiki modeller tanıtılacaktır. Üçüncü kısımda, EÇO tahmin edicisinin Stata yazılımında nasıl uygulanabileceği konusu takip edecektir. Dördüncü kısımda ise sonuca yer verilecektir. Çalışmada kullanılacak belli başlı notasyon şöyle listelenebilir. T , bir vektör veya matrisin transpozunu, $\cdot$ fonksiyon veya matematiksel objelerin bir tanımla eşleştiğini, plim ihtimal limitini ve $[x]$ x 'e soldan en yakın tamsayıyı ifade edecektir. Ayrıca, bir fonksiyonun birinci türevi $(\cdot)$ işaretiyle gösterilecektir. Bu bölümdeki bütün limitler gözlem sayısı sonsuza giderken hesaplanacaktır.

2.1. EN ÇOK OLABİLİRLİK TAHMİN EDİCİSİ

En çok olabilirlik tahmin edicisinin tanımlanmasında Cameron ve Triverdi'nin (2005) çalışmasından yararlanılacaktır. Öncelikle bu çalışmada doğrusal veya doğrusal olmayan regresyon modellerinin tahmini üzerinde yoğunlaşılacaktır. Dolayısıyla, girişte verilen örneğe ek olarak, bağımlı değişkeni açıklayacak eşdeğişken vektörü $\{X_i\}_{i=1}^n$ şeklinde tanımlanacaktır. Ayrıca, bağımlı ve bağımsız değişken arasındaki ilişki bağlantı fonksiyonu adı verilen $y_i = g(X_i, \theta)$ ilişkisiyle tanımlanacaktır. Bu ilişki uygulamacılara farklı regresyon modelleriyle çalışma imkânını verecektir.

2.1.1. Olabilirlik Fonksiyonu

İncelenen veri setindeki değişkenler vektörel olarak $\mathbf{y} = [y_1, \dots, y_n]^T$ ve $\mathbf{X} = [X_1^T, \dots, X_n^T]^T$ şeklinde tanımlansın. $f(\mathbf{y}, \mathbf{X} | \theta)$, p boyutlu bilinmeyen anakütle parametre kümesi θ veri iken $(\mathbf{y}, \mathbf{X})$ değişkenlerinin birleşik olasılık yoğunluk fonksiyonu olsun. Bu fonksiyon, aynı zamanda, gerçekleşen $(\mathbf{y}, \mathbf{X})$ verisinin θ 'ya göre olabilirliğini vermektedir. Ancak, bu bölümde, Cameron ve Triverdi'de (2005) olduğu gibi $\mathbf{y}$ 'nin $\mathbf{X}$ göre şartlı olasılık yoğunluk fonksiyonuna, diğer bir deyişle olabilirliğine yoğunlaşılacaktır. Bu fonksiyon $L_n(\theta) = f(\mathbf{y} | \mathbf{X}, \theta)$ şeklinde ifade edilecektir. Aşağıdaki tabloda bu dağılım fonksiyonlarının bazı örnekleri görülebilir.

Tablo 2.1. Belirli İhtimal Dağılım Fonksiyonları

Dağılım Adı	Parametre Kümesi	Dağılım Fonksiyonu $f(\mathbf{y} \mathbf{X}, \theta)$	Dağılım Gösterimi
Normal	$\theta = [\beta^T, \sigma]$	$(2\pi\sigma^2)^{-0.5} e^{-(y-X\beta)/(2\sigma^2)}$	$y \sim \mathcal{N}(X\beta, \sigma^2)$
Poisson	$\theta = [\beta]$	$e^{-(X\beta)} (X\beta)^y / y!$	$y \sim \text{Poi}(X\beta)$
Negatif Binomial	$\theta = [\beta^T, \vartheta]$	$\frac{\Gamma(y + \vartheta)}{\Gamma(y + 1)\Gamma(\vartheta)} \left(\frac{\vartheta}{\vartheta + X\beta}\right)^\vartheta \left(\frac{X\beta}{\vartheta + X\beta}\right)^y$	$y \sim \text{nbinom}(X\beta, \vartheta)$

Yukarıdaki yapıyı eldeki veri setine uygulamak için, öncelikle $\{(y_i, X_i)\}_{i=1}^n$ serilerinin gözlemlendiği düşünölsün. Ayrıca, bütün $i \in \{1, \dots, n\}$ için y_i 'nin X_i ve θ ya göre şartlı dağılımının birbirine özdeş ve birbirinden bağımsız olduđu varsayölsün. Bu varsayım (bağımsızlık varsayımı) bize veri vektörü y 'nin şartlı olabilirliğinin $L_n(\theta) = f(y|X, \theta) = \prod_{i=1}^n f(y_i|X_i, \theta)$ şeklinde yazılabilmesine imkân verecektir.

Örneğin, $y_i = X_i^\top \beta + \epsilon_i$ ve $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$ olmak üzere çok değişkenli doğrusal regresyon modelini ele alönsün. Söz konusu modelde basit bir cebir ile $y_i \sim \mathcal{N}(X_i^\top \beta, \sigma^2)$ dağılımına ulaşılabilir. Böylece, Tablo 2.1'deki sonucu kullanarak y 'nin şartlı olabilirliği $\theta = [\beta', \sigma]$ iken $L_n(\theta) = \prod_{i=1}^n (2\pi\sigma^2)^{-0.5} e^{\{-(y_i - X_i^\top \beta)/(2\sigma^2)\}} = (2\pi\sigma^2)^{-0.5n} e^{\{\sum_{i=1}^n (y_i - X_i^\top \beta)/(2\sigma^2)\}}$ şeklinde yazılabilir. Bu şartlı olasılık, θ parametre kümesini çözümlemek için bir objektif fonksiyonu olarak kullanılabilir. Yani, $L_n(\theta)$ fonksiyonu θ 'ya göre optimize edilecektir. Amacımız, en yüksek olabilirliği elde etmek olduđu için, bu optimizasyon işlemi maksimizasyon olarak tasarlanacaktır. $L_n(\theta)$ fonksiyonunun en yüksek değeri veren θ vektörüne en çok olabilirlik tahmini ismi verilecektir. Yukarıdaki örnekte göröleceğı üzere, olabilirlik fonksiyonu yüksek seviyede doğrusal olmama özelliğı taşıyacaktır. Bu durumdan kurtulmak için tekdüze bir dönüşüm olan doğal logaritma kullanılabilir. $\mathcal{L}_n(\theta) = (1/n) \log[L_n(\theta)] = (1/n) \sum_{i=1}^n \log[f(y_i|X_i, \theta)]$ fonksiyonun (ortalama²) log-olabilirlik fonksiyonu olarak tanımlanabilir (Cameron ve Trivedi, 2005). Log-olabilirlik fonksiyonun optimal değeri (en büyük değeri) olabilirlik fonksiyonunun optimal değeri verecektir. Optimal değer, diğeri bir deyişle EÇO tahmin edicisini, θ parametre kümesi θ 'nın alabileceğı değerler olarak düşünölsürse $\hat{\theta}_{EÇO} = \argmax_{\theta \in \Theta} \mathcal{L}_n(\theta)$ şeklinde yazılabilir.

Yukarıda tanımlanan maksimizasyon işlemi gerçekleştirerek için, $\mathcal{L}_n(\theta)$ fonksiyonunun θ 'ya göre birinci türevini alıp sıfıra eşitlemek gerekecektir:

$$G(\theta) := \frac{\partial \mathcal{L}_n(\theta)}{\partial \theta} = \frac{1}{n} \sum_{i=1}^n \frac{\partial \log[f(y_i|X_i, \theta)]}{\partial \theta} = 0. \quad (2.1)$$

Denklem (2.1) skor (gradyan) veya ilk derece şartı olarak adlandırılır. Yukarıdaki normal dağılım örneğıne dönölsürse, β parametresine göre skor fonksiyonu,

$$\frac{\partial \mathcal{L}_n(\theta)}{\partial \beta} = \frac{1}{n} \sum_{i=1}^n \log[f(y_i|X_i, \theta)] = - \sum_{i=1}^n 2(y_i - X_i^\top \beta) \beta / (2\sigma^2) = 0, \quad (2.2)$$

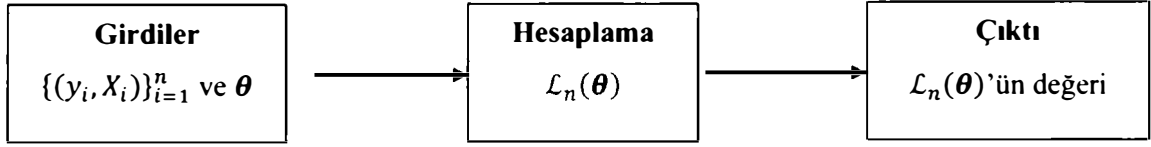
şeklinde yazılıp çözüldüğünde, EÇO tahmin edicisi $\hat{\beta}_{EÇO} = (X^\top X)^{-1} X^\top y$ olarak bulunabilir. Dikkat edilirse elde edilen çözümün, En küçük kareler tahmin edicisiyle denk olduđu görölebilir. Ancak, ekonometri ve istatistik alanında nadiren denklem (2.1)'in kapalı çözümü elde edilebilir. $\mathcal{L}_n(\theta)$ fonksiyonunun numerik optimizasyon yöntemiyle maksimizasyon işlemi gerçekleştirerek gerekecektir.

2.1.2. Numerik Optimizasyon Yöntemleri

Numerik optimizasyon yöntemleri Hamilton'da (1994) olduđu gibi kara kutu metaforuyla açıklanabilir.

² Ortalama ifadesi log-olabilirlik fonksiyonun gözlem sayısına bölünmesinden dolayı ortaya çıkmaktadır. Bu işlem EÇO tahmin edicisinin analitik çözümlemesinde kolaylık sağlamak için gereklidir. Ayrıca, bu işlem EÇO tahmin edicisinin değeri değiştirmez.

Şekil 2.1. Numerik Optimizasyon (Kara Kutu Metaforu)



Şekil 2.1.'e göre gözlemlenen veriler ve θ 'nin herhangi bir değeri kara kutu prosedürüne sokulduğunda çıktı olarak log-olabilirlik değeri alınacaktır. Yöntem farklı θ değerleri için uygulandığında log-olabilirliğin farklı numerik değerleri ortaya çıkacaktır. Numerik optimizasyondaki temel amaç, yeterince farklı θ değerine karşılık gelen log-olabilirlik fonksiyonlarının kıyaslanması sonucu en yüksek log-olabilirlik değerini veren θ değerini elde etmeye çalışmak olacaktır. Tablo 2.2.'de sıklıkla kullanılan numerik optimizasyon yöntemlerinin özetleri bulunabilir.

Tablo 2.2. Numerik Optimizasyon Yöntemleri

Optimizasyon Yöntemi	Tanım	Notlar
Grid taraması	Belirli bir aralıkta θ 'nin farklı değerleri için $L_n(\theta)$ 'nin hesaplanması ve bu değerlerin kıyaslanması şeklindedir	Genellikle bir parametrelili optimizasyonlarda kullanılması daha uygundur. En basit numerik optimizasyon yöntemi olarak bilinir.
En dik iniş yöntemi	Bu yöntemde, $\theta^{(0)}$ değerinden başlayarak, parametre değerinin log-olabilirlik fonksiyonunun türevi doğrultusunda güncellenmesi şeklinde ortaya çıkar. Her bir aşama i 'de $\theta^{(i)} = \theta^{(i-1)} - \epsilon G(\theta^{(i-1)})$ $\epsilon > 0$ sabit olmak üzere parametre değeri güncellenir. Ardışık iki parametre değeri değişmemeye başladığında optimal nokta bulunur.	Bu yöntem genellikle tek global optimal noktaya sahip ve sürekli fonksiyonların optimizasyonunda daha etkilidir.
Newton ve Raphson yöntemi ve türevleri	En dik iniş yönteminde ϵ sabiti yerine $L_n(\theta^{(i-1)})$ 'nin ikinci türevinin (Hessian matrisi) kullanılmasıyla oluşturulur. Eğer Hessian yerine türevin dış çarpımı kullanılırsa Berndt–Hall–Hall–Hausman (BHHH) yöntemi elde edilir.	En dik iniş yöntemine göre daha hızlı çalışmaktadır (Hamilton, 1994).
Höristik Yöntemler	Bu yöntemler, Newton ve Raphson yöntemi gibi optimize edilen fonksiyonun analitik özelliklerini kullanmaz. Bu türün en önemli örnekleri, Kennedy ve Eberhart'ın (1995) Particle Swarm optimizasyonu, Holland'ın (1975) Genetik algoritması ve Pincus'un (1970) benzetilmiş tavlama algoritması olarak sıralanabilir	Höristik yöntemler, analitik optimizasyon yöntemlerine göre daha esnek şartlarda çalışabilir. Örneğin, optimize edilen fonksiyonların sürekli olmaması gibi durumlarda bu yöntemler daha başarılı olabilir.

Şekil 2.1'deki yöntemler uygulanarak eldeki log-olabilirlik fonksiyonun en yüksek değerini veren katsayılar bulunabilir. Ancak, numerik optimizasyon yöntemlerinde belirli problemlerle karşılaşılabilir. Problemlerden bazıları, log-olabilirlik fonksiyonun birden fazla lokal en yüksek bölgesinin olması, log-olabilirlik fonksiyonun sürekli olmaması ve belirli bir aralıkta düze yakın bir seyir izlemesi olarak sıralanabilir. Ayrıca, bazı optimizasyonlarda başlangıç parametrelerinin değeri optimizasyonu etkileyebilir. Söz konusu durumdan şüpheleniliyorsa, farklı başlangıç değerleriyle optimizasyon yeniden çalıştırılıp en iyi sonuç elde edilebilir.

2.1.3. En Çok Olabilirlik Tahmin Edicisinin Varyansı

EÇO tahmin edicisi, diğer istatistiki tahmin yöntemleri gibi hipotez testlerinde kullanılabilir. Ancak, hipotez testlerini gerçekleştirmek için elde edilen tahmin edicilerin varyansına ihtiyaç duyulmaktadır. Aşağıda bu varyansın nasıl elde edileceği görülecektir.

2.1.3.1. Bilgi Matrisi ve Hessian

EÇO tahmin edicisinin varyansını elde edebilmek için iki matematiksel obje tanımlanabilir. Birinci obje olan $p \times p$ Bilgi matrisi $I(\theta) = E_{\theta}[G(\theta)^T G(\theta)]$ log-olabilirlik fonksiyona ait gradyanın θ ya göre varyans-kovaryans matrisi olarak tanımlanabilir. İkinci obje ise, $p \times p$ boyutlu $H(\theta) = \partial^2 \mathcal{L}_n(\theta) / \partial \theta^2$ şeklinde tanımlanan log-olabilirlik fonksiyonun ikinci türevi olan Hessian matrisi olarak ifade edilebilir. Davidson ve Mackinnon'ın (2004) gösterdiği gibi bu iki obje limitte $\text{plim}_{n \rightarrow \infty} (1/n) I(\theta) = -\text{plim}_{n \rightarrow \infty} (1/n) H(\theta)$ ilişkisiyle bağlıdır. Objelerin denkliği ve bölümünde üzerinde durulmayacak olan cebirsel detaylar sonucunda, EÇO tahmin edicisinin varyansı eşitlik 2.3'deki şekilde bulunabilir:

$$\text{var}(\hat{\theta}_{E\text{ÇO}}) = -H(\theta)^{-1} = I(\theta)^{-1}. \quad (2.3)$$

Bu varyans-kovaryans matrisi kullanılarak standart çıkarımlar yapılabilir. Ayrıca, örnek uzayda hesaplanan Hessian ve bilgi matrisleri $H(\hat{\theta})$ ve $I(\hat{\theta})$ şeklinde de gösterilebilir.

2.1.4. En Çok Olabilirlik Tahmin Edicisinin Limit Özellikleri

Bu kısımda, EÇO tahmin edicilerinin limit özellikleri incelenecektir. Öncelikle, limit özellikleri için gerekli varsayımları Cameron ve Triverdi'nin (2005) 5.5 numaralı önermesinden faydalanaarak aşağıdaki şekilde sıralanacaktır:

Varsayım 2.1. (a). İncelenen veri seti $\{(y_i, X_i)\}_{i=1}^n$, X veri iken $f(y_i|X_i, \theta)$ şartlı yoğunluk fonksiyonuyla oluşturulmuştur.

(b). Şartlı yoğunluk fonksiyonu tekil şekilde belirlenmiştir, yani, $f(y_i|X_i, \theta^{(1)}) = f(y_i|X_i, \theta^{(2)})$ eşitliği sadece $\theta^{(1)}, \theta^{(2)'}ye$ eşit olduğunda geçerlidir.

(c). Asimptotik Hessian matrisi $A_0 = -\text{plim}_{n \rightarrow \infty} (1/n) H(\theta)$ tanımlıdır ve tekil değildir.

(d). Log-olabilirlik fonksiyonunun türev ve integral dereceleri tersine çevrilebilir.

(e). Gerçek anakütle parametresi θ tanımlı olduğu reel kümenin bir iç noktası olmalıdır (Andrews, 1999).

Yukarıdaki varsayımlardan (a)'ya göre tahmin edilen modelde yanlış belirleme sorunu olmadığı ifade edilmektedir. (a) ve (b) birlikte, tahmin edilmek istenen parametrelerin tekil olarak tanımlanmış olduğunu varsaymaktadır. (c) maddesi ise asimptotik varyans-kovaryans matrisinin tanımlı ve tersine-çevrilebilir olmasıyla ilgilidir. Cameron ve Triverdi'ye göre (d) maddesi düzenleyici koşulların ortaya çıkması için gerekli üst düzey bir varsayımdır. Son olarak, Andrews (1999) EÇÖ'nün üyesi olduğu ekstremum tahmin edicilerin limit dağılımlarının standart şartlara uyabilmesi için, gerçek parametre değerinin, parametrenin alabileceği değerler kümesinin sınır noktalarında olmaması gerektiğini ifade etmektedir.

Önerme 2.1. Varsayım 2.1 altında, EÇÖ tahmin edicisi $\hat{\theta}_{E\check{O}}$ tutarlı ve $\sqrt{n}(\hat{\theta}_{E\check{O}} - \theta)$ limitte $\mathcal{N}(0, A_0^{-1})$ dağılımına sahiptir.

Önerme 2.1.'e ek olarak EÇÖ tahmin edicileri, sapmasız tahmin ediciler arasında en düşük varyansa sahip olma, Cremor-Rao alt sınırına ulaşma özelliğine sahiptir (Cameron ve Triverdi, 2005).

2.1.5. Değişmezlik İlkesi ve Delta Yöntemi

Ekonometri ve istatistikte, uygulamacılar sıklıkla model parametrelerinin doğrusal olmayan fonksiyonları üzerinden çıkarım yapma ihtiyacı duymaktadır (Davidson ve Mackinnon, 2004). EÇÖ tahmin yöntemi, değişmezlik ilkesini taşıması itibarıyla, uygulamacılara çok büyük bir kolaylık sağlamaktadır. Değişmezlik ilkesi, Aşağıda örnekle açıklanabilir. İncelenen modelde bilinmeyen anakütle parametresi(leri) θ olsun ve $\gamma = h(\theta)$ ise parametrenin(lerin) bir dönüşümü (doğrusal ya da doğrusal olmayan) olarak tanımlansın. Eğer $\hat{\theta}$, θ 'nın EÇÖ tahmin edicisi ise, değişmezlik ilkesi altında, $\hat{\gamma} = h(\hat{\theta})$ de γ 'nın tahmin edicisi olacaktır. Bu durumda, $\hat{\theta}$ elde edildikten sonra $\hat{\gamma}$ için yeniden tahmin işlemini gerçekleştirmeye gerek olmayacaktır. Özelliğin kullanımı kısım 2.3.1'de incelenecektir.

Değişmezlik ilkesi, orijinal parametrenin bir dönüşümünü tahmin etmede kolaylık sağlasa da yeni parametrenin dağılım özellikleri için birkaç detaya daha dikkat etmek gerekir. Çünkü değişmezlik ilkesi parametre tahmini için geçerli iken, parametrelerin standart sapma hesaplamalarında geçerliliğini kısmen yitirecektir. Yukarıdaki örneğe dönülürse, $var(\hat{\theta})$ $\hat{\theta}$ 'in varyansı olsun. Genel olarak $var(\hat{\gamma}) \neq h[var(\hat{\theta})]$ olduğu söylenebilir. Ancak, dönüştürülmüş parametre tahminine ait varyansın hesaplamasında bir engel oluşmayacaktır. Şöyle ki, $h(\cdot)$ birinci türevi olan ve türevinin $\hat{\theta}$ 'deki değeri sıfır olmayan bir fonksiyon olsun. Birinci derece Taylor genişlemesi yardımıyla $var(\hat{\gamma}) = h'(\hat{\theta})var(\hat{\theta})h'(\hat{\theta})^T$ sonucuna ulaşılabilir (Davidson ve MacKinnon). Ek olarak, Önerme 2.1 şu şekilde genişletilebilir.

Önerme 2.2. Varsayım 2.1'in şartları geçerli ve $h(\cdot)$ birinci türevi olan ve bu türevin EÇÖ tahminindeki değeri sıfır olmayan bir fonksiyon olsun. Öyleyse $\hat{\gamma}$ 'in limit dağılımı şöyle gösterilebilir: $\sqrt{n}(\hat{\gamma}_{E\check{O}} - \gamma) \xrightarrow{d} \mathcal{N}(0, h'(\hat{\theta})A_0^{-1}h'(\hat{\theta})^T)$.

Önerme 2.2. yardımıyla sonraki kısımlarda anlatılacak hipotez testleri, EÇÖ parametrelerinin doğrusal olmayan dönüşümlerine de uygulanabilir.

2.1.6. Sözde En Çok Olabilirlik Tahmini

Önceki bölümlerde, bazı düzenleyici varsayımlar altında EÇÖ tahmin edicilerin ideal özelliklerine değinildi. Ancak, varsayımlardan bir veya daha fazlası ihlal edilirse, EÇÖ tahmin edicilerinin ideal özellikleri ortadan kalkmaktadır. Varsayım 2.1-(a)'nın eksikliğinde EÇÖ tahmin edicileri irdelenecektir. Öncelikle, varsayımın ortadan kalkması, analiz edilen modelin yanlış bir dağılımla tahmin edilmeye çalışıldığı anlamına gelecektir. Ancak, EÇÖ'nun kısım 2.4'te belirtilen özelliklerini çok önemli şekilde etkilemeyecektir.

Cameron ve Triverdi'ye (2005) göre, yanlış olabilirlik fonksiyonu altında tahmin edilen EÇÖ katsayıları limitte sözde gerçek parametre değeri θ^* 'a yakınsayacaktır ve eğer θ^* değeri gerçek parametre değeri θ 'ya eşit değilse, EÇÖ tahmin edicisi limitte tutarsız olacaktır. Fakat, söz konusu durum asimptotik çıkarım odaklı analizleri etkilemeyecektir. Şöyle ki, Huber (1967) ve White'a (1982) göre EÇÖ tahmin edicileri sözde gerçek parametre değeri etrafında normal dağılıma sahiptir. Öyleyse, $\sqrt{n}(\hat{\theta}_{E\hat{C}O} - \theta^*)$ limitte sıfır ortalamalı ve $(1/n)I^*(\theta)^{-1}H^*(\theta)I^*(\theta)^{-1}$ varyans-kovaryans matrisli normal dağılıma sahiptir. Dolayısıyla, standart çıkarımlar EÇÖ tahmin edicisi kullanılarak yerine getirilebilir.

Ek olarak, White'a (1982) göre, sözde EÇÖ yönteminin aslında gerçek yoğunluk fonksiyonu ile sözde gerçek yoğunluk fonksiyonu arasındaki Kullback–Leibler Bilgi Kriteri bazlı mesafeyi minimize etmektedir.

2.1.7. Hipotez Testleri

Davidson ve Mackinnon'a (2004) göre, EÇÖ tahmin edicisi kullanılarak üç farklı yöntemle istatistiki çıkarımda bulunulabilir. Bu yöntemler, olabilirlik oranı, Wald ve Lagranj çarpanı testleri olarak sıralanabilir. Farz edilsin ki $r(\theta)$, θ 'nın pürüzsüz $h < p$ boyutlu bir fonksiyonu olsun ve $H_0: r(\theta) = 0$ boş hipotezini $H_a: r(\theta) \neq 0$ alternatif hipotezine karşı test etmek istensin. Testin, her üç yöntemle nasıl yapılacağı sırasıyla aşağıdaki kısımlarda açıklanacaktır.

2.1.7.1. Wald Testi

Genel bir hipotez test etme yöntemi olan Wald testi EÇÖ tahmininde de kullanılabilir. Yukarıda belirtilen $H_0: r(\theta) = 0$ hipotezi için aşağıdaki ikinci dereceden test istatistiği kullanılabilir:

$$Wald_{r(\theta)} = r^T(\hat{\theta}_{E\hat{C}O})[R(\hat{\theta}_{E\hat{C}O})var(\hat{\theta}_{E\hat{C}O})R^T(\hat{\theta}_{E\hat{C}O})]^{-1}r(\hat{\theta}_{E\hat{C}O}). \quad (2.4)$$

(2.4)'deki gösterimde $var(\hat{\theta}_{E\hat{C}O})$ denklem (2.3)'te belirtilmiş ve $R(\hat{\theta}_{E\hat{C}O})$ ise i 'ye j 'inci elemanı $r(\theta)$ 'nın i 'inci elemanın θ_j 'ye göre türevi $R_{i,j}(\hat{\theta}_{E\hat{C}O}) = \partial r_i(\theta)/\partial \theta_j$ olacak şekilde tanımlanmış bir $h \times p$ matristir. Davidson and MacKinnon (2004), sıfır hipotez altında test istatistiğinin $\chi^2(h)$ olarak dağıldığını söylemektedir. η anlamlılık düzeyi iken, $Q_{1-\eta}[\chi^2(h)]$ 'yı $\chi^2(h)$ dağılımın $(1 - \eta)$ 'inci kantili olarak tanımlansın. Eğer hesaplanan test istatistiği $Wald_{r(\theta)}$, kritik değer $Q_{1-\eta}[\chi^2(h)]$ 'dan daha büyükse bu sıfır hipotez reddedilecektir.

2.1.7.2. Olabilirlik Oranı Testi

Olabilirlik oran testi iki farklı EÇÖ tahminini gerektirmektedir. Bunlardan ilki kısıtsız olarak tanımlanan orijinal model olacaktır. Modelin EÇÖ tahmini $\hat{\theta}_{EÇÖ,U}$ olarak gösterilecek ve modele ait hesaplanan log-olabilirliği $\mathcal{L}_n(\hat{\theta}_{EÇÖ,U})$ olarak tanımlanacaktır. Kısıtlı model olarak adlandırılan ikinci model ise sıfır hipotez altındaki kısıt $r(\theta) = 0$ denkliği modelde yerine koyularak EÇÖ tahmini gerçekleştirilecektir. Tahmin sonucu elde edilen değer $\hat{\theta}_{EÇÖ,R}$ ve tahmin değeri kullanılarak hesaplanan log-olabilirlik fonksiyonu $\mathcal{L}_n(\hat{\theta}_{EÇÖ,R})$ şeklinde gösterilecektir. Böylece, kısıtlı ve kısıtsız modellerden oluşturulan log-olabilirlik değerlerini kullanarak olabilirlik oranı test istatistiği aşağıdaki şekilde elde edilecektir:

$$LR_{r(\theta)} = 2[\mathcal{L}_n(\hat{\theta}_{EÇÖ,U}) - \mathcal{L}_n(\hat{\theta}_{EÇÖ,R})]. \quad (2.5)$$

Test istatistiğinin değerinin her zaman pozitif olacağı söylenebilir. Bu durum en küçük kareler yönteminde kısıtlı modelin r-karesinin daha küçük olması durumuyla benzer bir özellik göstermektedir. Diğer bir deyişle, modele yeni parametre ekledikçe modelin log-olabilirliği artacaktır. Ayrıca, sıfır hipotez doğruysa denklem (2.5)'teki test istatistiği $\chi^2(h)$ dağılımına sahip olacaktır (Davidson ve Mackinnon, 2004). Dolayısıyla, Wald testine benzer bir şekilde çıkarım yapılabilir.

2.1.7.3. Lagranj Çarpanı Testi

Test türünün ismi Lagranj yöntemiyle optimizasyonu çağrışırsa da test istatistiğini oluştururken doğrudan Lagranj yöntemi kullanılmayacaktır. Hatta, bu test istatistiğini türetmek için sadece sıfır hipotezde belirtilen kısıtlı modeli tahmin etmek yeterli olacaktır. Öncelikle, denklem (2.1)'de tanımlanmış gradyanı $\hat{\theta}_{EÇÖ,R}$ değerinde hesaplayıp $G(\hat{\theta}_{EÇÖ,R})$ objesi elde edilecektir. Aynı şekilde, log-olabilirlik fonksiyonunun Bilgi matrisi veya Hessian'ını $\hat{\theta}_{EÇÖ,R}$ 'da hesaplayıp $I(\hat{\theta}_{EÇÖ,R})$ objesi elde edilecektir. $\hat{\theta}_{EÇÖ,R}$ ve $I(\hat{\theta}_{EÇÖ,R})$ objelerini kullanarak, Lagranj çarpanı test istatistiği,

$$LM_{r(\theta)} = G^T(\hat{\theta}_{EÇÖ,R})I(\hat{\theta}_{EÇÖ,R})^{-1}G(\hat{\theta}_{EÇÖ,R}), \quad (2.6)$$

şeklinde hesaplanacaktır. Diğer iki test istatistiği gibi Lagranj çarpanı testi de $\chi^2(h)$ dağılımına sahip olacaktır. Bundan dolayı, Lagranj çarpanı testinde de Wald ve olabilirlik oranı testlerindeki çıkarım kurallarının aynısı kullanılabilir. Ayrıca, Buse (1982), log-olabilirlik fonksiyonunun ikinci dereceden olması durumunda Wald, Lagranj çarpanı ve olabilirlik oranı testleri numerik olarak özdeş olacaktır.

2.1.8. Tek Kısıtlı ve Tek Taraflı Hipotezlerin Test Edilmesi

Yukarıda değinilen üç farklı yöntemde de alternatif hipotez çift taraflı olarak tanımlanmıştır. Ancak, doğrusal regresyon modellerinde olduğu gibi bazı durumlarda tek taraflı alternatifleri de göz önünde bulundurmamak gerekebilir. Önerme 2.1'in sonuçlarını kullanarak hipotez testleri gerçekleştirilebilir. Örneğin, test etmek istenilen hipotez herhangi bir $j \in \{1, \dots, p\}$ ve bilinen θ_{0j} hipoteze edilmiş değeri için $H_0: \theta_j \geq \theta_{0j}$ 'a karşı $H_a: \theta_j < \theta_{0j}$ olsun. Hipotezde tek kısıt olduğundan dolayı z istatistiğiyle test edilebilir. z-istatistiği aşağıdaki şekilde hesaplanabilir:

$$z_{\theta_j} = \frac{\sqrt{n}(\hat{\theta}_{j,E\check{O}} - \theta_{0j})}{se(\hat{\theta}_{j,E\check{O}})} \quad (2.7)$$

Test istatistiğini elde etmek için gerekli $se(\hat{\theta}_{j,E\check{O}})$ ifadesini de $var(\hat{\theta}_{E\check{O}})$ objesinin j 'inci diagonal elamanının karekökü alınarak hesaplanabilir. Önerme 2.1'e göre bu test istatistiği, EKK yönteminin aksine, standart normal dağılıma sahip olacaktır. Dolayısıyla, yukarıdaki hipotezi reddetmek için hesaplanan z_{θ_j} 'nin standart normal dağılımın η 'inci kantilinden küçük olup olmadığına bakmak gerekecektir. Ayrıca, birden fazla parametre içeren tek kısıtlı hipotezler de benzer şekilde Önerme 2.1'i kullanarak çözümlenebilir.

2.2. DOĞRUSAL OLMAYAN REGRESYON MODELLERİ

İkinci kısımda belirtildiği gibi EÇÖ tahmin yöntemi, bağımlı ve açıklayıcı değişkenler arasındaki ilişki konusunda bir kısıt koymamaktadır. Mesela, sıradan en küçük kareler yöntemiyle tahmin edilen modelin parametreleri (katsayılar) bağımlı değişkene göre doğrusal olmalıdır. Bu duruma parametreye göre doğrusallık adı verilir. Eğer bir model parametrelere (katsayılar) göre doğrusal değilse doğrusal olmayan en küçük kareler (EKK), genelleştirilmiş moment yöntem veya EÇÖ yöntemi kullanılabilir. Üç yöntem arasındaki temel farklılık, EÇÖ metodunda verinin ihtimal dağılımı bilgisinin de tahmin sürecinde kullanılmasıdır. Kitabın bu bölümünde EÇÖ tahmin edicilerine odaklanıldığı için doğrusal olmayan EKK ve genelleştirilmiş moment yönteminden bahsedilmeyecektir. Ancak, okuyucular yöntemler hakkında detaylı bilgiyi Davidson ve MacKinnon'da (2004) bulabilir.

Doğrusal olmayan modeller, genel itibarıyla, $g(\cdot)$ doğrusal olmayan bir fonksiyon, X_i açıklayıcı değişkenler, θ bilinmeyen parametre vektörü ve ϵ_i ise hata terimi olmak üzere, $y_i = g(X_i, \theta, \epsilon_i)$ şeklinde ifade edilebilir. Bazı özel durumlarda, $y_i = g(X_i, \theta) + \epsilon_i$ modelinde olduğu gibi, hata terimini toplanır olarak tanımlanabilir. Örneğin, $y_i = X_i^\beta + \epsilon_i$ modeli, toplanır doğrusal olmayan bir regresyon modelidir.

Davidson ve Mackinnon (2004) doğrusal olmayan regresyon modellerini toplanır model olarak tanımlasa da bu tanım doğrusal olmayan en küçük kareler yönteminin motivasyonu için gereklidir. Hâlbuki mikro-ekonometride kullanılan ve doğrusal bir yapıya sahip olmayan probit, logit vb. modellerde doğrudan bir hata terimi tanımlaması yoktur. Bunun yerine verinin şartlı dağılımı üzerine varsayımlar yapılır.³ Doğrusal olmayan en küçük kareler yöntemi, probit, logit vb. modellerdeki kullanımı çok kısıtlı olsa dahi, EÇÖ toplanır ve toplanır olmayan modellerde daha etkin bir şekilde çalışmaktadır (Davidson ve Mackinnon, 2004).

Bunlara ek olarak, Mikro-ekonometri alanındaki birçok modellemede doğrusal olmama durumuyla sıklıkla karşılaşmaktadır. Modellerin bir kısmında bazı dağılımsal varsayımlarda bulunulur. Bundan önce sıklıkla dile getirilmiş olan probit, logit modelleri ve bu modellerin türevlerinin doğrusal olmayan modeller sınıfına girdiği söylenebilir. Örneğin, y_i çiftli değer alan seçim değişkeni, X_i bu seçimi açıklamada kullanılan bağımsız değişkenler olsun. Bu durumda, $\Phi(\cdot)$ standart normal dağılımın toplam dağılım fonksiyonu olmak üzere, regresyon modeli $E[y_i|X_i, \beta] = \Phi(X_i^\top \beta)$ şeklinde yazılabilir. Bu durumda, $\Phi(\cdot)$ doğrusal olmayan bir fonksiyon olduğu için

³ probit ve logit gibi modellerde gözlemlenmeyen bir değişken varsayılarak, hata terimi modele eklenilebilir. Ancak, bu durumda modelin dolaylı yoldan bir hata terimi içerdiği söylenebilir.

bağımlı değişkenin şartlı ortalamasının parametrelere göre doğrusal olmayan bir yapıda olduğu görülecektir.

Doğrusal olmayan modellerin, doğrusal modellerden en önemli farkı, açıklayıcı değişkenler üzerine yapılan çıkarımlardır. Temel olarak, söz konusu modellerdeki marjinal değerler, doğrusal modellerle aynı yaklaşımla elde edilse de, yapıları itibarıyla büyük farklılıklar göstermektedir. Farklılık, doğrusal olmayan modellerin marjinal etkilerinin de doğrusal olmasıyla ortaya çıkar. Örneğin, basit regresyon modelinde j 'inci açıklayıcı değişkenin marjinal (kısmi) etkisi $\partial E[y_i|X_i, \beta]/\partial \beta_j = \beta_j$ yine bilinmeyen parametrenin doğrusal bir fonksiyonu iken, doğrusal olmayan bir modelde $\partial E[y_i|X_i, \beta]/\partial \beta_j = X_j g'(X_i, \beta)$ şeklinde gösterilip genellikle doğrusal olmama özelliği göstermektedir. Örneğin, probit modelinde marjinal etki, $\phi(\cdot)$ normal dağılımın ihtimal yoğunluk fonksiyonu olmak üzere, $\partial E[y_i|X_i, \beta]/\partial \beta_j = X_j \phi(X_i^T \beta)$ şeklinde bulunabilir. Marjinal analiz konusunu 4.3'teki örnekte daha detaylı olarak incelenecektir.

2.3. En Çok Olabilirlik Tahmininin Stata Uygulamaları

Daha önce bahsedildiği gibi, EÇO tahmin edicileri istatistiki ve ekonometrik çalışmalarda sıklıkla kullanılmaktadır. Stata programı da farklı istatistiki veya ekonometrik modellerde hazır işlev olarak EÇO tahmin edicisini kullanmaktadır. Örneğin, kitabın diğer bölümlerinde işlenecek olan probit ve logit modelleri, poisson ve negatif binom saymalı modelleri ve yapısal eşitlik modelleri gibi yöntemlerde Stata tahmin edici olarak EÇO yöntemi kullanılmaktadır. Söz konusu modeller için örnekler ve Stata uygulaması diğer bölümlerde verileceği için, aşağıda kullanıcı tarafından belirlenen olabilirlik fonksiyonların incelenmesi üzerine yoğunlaşılacaktır.

2.3.1. Örnek 1: Basit Poisson Modeli

Aşağıdaki örnekte öncelikle basit bir poisson rastsal değişkeni oluşturulacaktır. Belirli bir zaman aralığında bir olayın kaç defa olduğunu gösteren rastsal değişkeni $\{u_i\}_{i=1}^n$ şeklinde gösterilirse, her bir $i \in \{1, \dots, n\}$ için $u_i \sim \text{Poisson}(\lambda)$ ile dağılacaktır. Poisson dağılımında λ , belirli bir zaman aralığında ortalamada kaç olay olacağını gösteren parametre olarak tanımlanacak ve EÇO yöntemiyle parametre tahmin edilmeye çalışılacaktır. Öncelikle, aşağıdaki Stata kodunu kullanarak $n = 20$ ve $\lambda = 3$ değerleriyle rastsal değişken oluşturulacaktır.⁴

Tablo 2.3. Poisson(3) Dağılımdan Rastgele Oluşturulmuş Gözlemler

```
clear /* Stata hafızasındaki değişkenleri silecek */
set obs 20 /* gözlem sayısını 20'ye eşitleyecek */
gen double u=rpoisson(3) /* Poisson(3) dağılımında rastgele 20 adet değişken oluşturacak */
```

2 2 2 4 6 3 2 3 2 2 3 1 3 5 3 2 5 0 6 3

Tablo 2.3'teki değerleri kullanarak, λ parametresi EÇO yöntemiyle tahmin edilebilir. Bunun için Poisson dağılımının log-olabilirlik fonksiyonunu yazmak gerekecektir. İlk olarak, Poisson dağı-

⁴ /* */ arasında kalan ifadeler Stata tarafından yorum olarak algılanacak ve bu ifadeler için herhangi bir işlem yapılmayacaktır. Bu kısım sadece bilgilendirme amaçlı kullanılacaktır.

lının ihtimal yoğunluk fonksiyonu $f(u_i|\lambda) = (\lambda^{u_i} e^{-\lambda}) / (u_i!)$ şeklinde yazılmalıdır. Tablo 2.3'teki gözlemler için log-olabilirlik fonksiyonu (2.8)'deki gibi gösterilebilir:

$$\mathcal{L}_n(\lambda) = \sum_{i=1}^n u_i \ln(\lambda) - \lambda - \ln(u_i!). \quad (2.8)$$

Denklem (2.8)'deki log-olabilirlik fonksiyonunun λ 'ya göre türevi alıp sıfıra eşitlenirse, basit cebirsel işlemlerle $\hat{\lambda}_{E\check{O}} = \bar{u}$ denklğine ulaşılabilir. Tablo 2.3'teki değerleri göz önünde bulundurularsa $\hat{\lambda}_{E\check{O}} = 2,95$ sonucuna ulaşılır. Ancak, Tablo 2.3'teki örnek, numerik optimizasyon yöntemiyle de çözümlenebilir. Diğer bir deyişle Stata'da denklem (2.8)'i ve Tablo 2.3'teki değerleri kullanarak EÇÖ tahmin edicisi elde edilebilir. Bunu gerçekleştirmek için Stata'nın "mlexp" fonksiyonu kullanılabilir. Fonksiyonun kullanım şekli "mlexp (log-olabilirlik fonksiyonun matematiksel gösterimi)" gibi olacaktır. İncelenen örnekte aşağıdaki kod kullanılacaktır:

```
mlexp (ln(({lambda}^u)*exp(-{lambda}))/ (round(exp(lnfactorial(u)),1))))
```

Yukarıdaki kodun birkaç detayından bahsetmek gerekecektir. İlk olarak, log-olabilirlik fonksiyonundaki parametreler tırnaklı parantez içerisinde yazılacaktır, mesela **{lambda}**. Fonksiyonda gözlemlenen değişkenleri Stata hafızasındaki sembollerıyla göstermek gerekmektedir. Son olarak, **(round(exp(lnfactorial(u)),1))**, u değişkeninin faktöriyelini almak için kullanılmıştır. Burada Stata 16'da hazır faktöriyel fonksiyonu olmadığı için değişkenin öncelikle logaritmik faktöriyelini alıp sonrasında üstel fonksiyonunu alarak değişkenin faktöriyeli elde edilebilir. Kod çalıştırıldığında EÇÖ tahminin sonuçları Tablo 2.4'te görülebilir.

Tablo 2.4. Basit Poisson Modelinin EÇÖ Tahmin Edicisi

```
mlexp (ln(({lambda}^u)*exp(-{lambda}))/ (round(exp(lnfactorial(u)),1))))
```

```
initial:      log likelihood =      -<inf>   (could not be evaluated)
feasible:      log likelihood =   -92,40981
rescale:       log likelihood =  -39,72276
Iteration 0:    log likelihood =  -39,72276
Iteration 1:    log likelihood = -37,266504
Iteration 2:    log likelihood = -36,687971
Iteration 3:    log likelihood = -36,687622
Iteration 4:    log likelihood = -36,687622
```

Maximum likelihood estimation

Log likelihood = -36,687622 Number of obs =20

	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
/lambda	2,95	,3840573	7,68	0,000	2,197262 3,702738

Tablo 2.4.'te, "Log likelihood" tahmin edilen parametre değerinde modelin log-olabilirliğini, "Number of obs" gözlem sayısını, "Coef." EÇÖ tahminini, "Std. Err.", EÇÖ tahminin standart sapmasını, "z" tahmin edilen parametrenin anlamlılık testinin z istatistiğini, "P>|z|" bu teste ait p-değerini ve en sağdaki iki element ise alt ve üst güven aralığını vermektedir. Bu sonuçlara göre

numerik optimizasyon ve analitik sonuç aynı değeri vermekte ve gerçek parametre değerine gayet yakın çıkmaktadır. Ayrıca, Tablo 2.4.'teki bilgileri kullanarak bazı hipotezler test edilebilir. Örneğin, $H_0: \lambda = 3$ hipotezi çift taraflı olarak test edilmek istensin. Bu test için kısım 2.1.7.4'te değinilen z -istatistiği kullanılabilir. Öyleyse, z değeri basit birkaç işlem sonrasında $z_\lambda = (2,95 - 3)/0,3840573 = -0,1316$ olarak hesaplanabilir. z istatistiğinin p -değeri ise 0,447650343 olduğu için %10 anlamlılık düzeyinde sıfır hipotezi reddedilemeyecektir. Dolayısıyla, istatistiki olarak λ 'nın 3'e eşit olduğu söylenebilir.

Örnekteki dağılımdan gelişi güzel oluşturulmuş bir poisson rastsal değişkeninin, beklenen değerden büyük olma ihtimali nedir?"- sorusuna cevap aransın. Soru matematiksel dile çevrilirse, $u \sim \text{Poisson}(\lambda)$ iken bulmaya çalışılan değer $P(u > \lambda)$ olacaktır. İhtimal, $h(\lambda) = P(u > \lambda) = 1 - \sum_{j=0}^{\lfloor \lambda \rfloor} e^{-\lambda} (\lambda^j / j!)$ şeklinde yazılabilir (Nowak, 2011). Değişmezlik prensibi bu fonksiyona uygulanır ve $\lfloor \lambda \rfloor = \lfloor 2,95 \rfloor = 3$ denkliği kullanılırsa $h(\hat{\lambda}) = 1 - \sum_{j=0}^3 e^{-2,5} (2,5^j / j!) = 0,4561$ sonucuna ulaşılabilecektir. Böylece, Poisson dağılımdan gelen bir rastsal değişkenin dağılımın ortalamasından %45,61 oranında büyük olacağı sonucu EÇÖ tahmin edicisi yardımıyla çıkarılabilir.

2.3.2. Örnek 2: Çok Değişkenli Normal Olmayan Hata Terimine Sahip Modelin Tahmini

Şimdi, bir önceki örneğe göre daha kapsamlı bir model üzerine çalışılacaktır. İncelenen model hem doğrusallık hem de normal dağılıma sahip olma durumlarını ihlal edecektir. Veri seti olarak Riphahn vd. (2003) makalesindeki gelir, yaş, eğitim ve cinsiyet değişkenleri kullanılacaktır. Veri seti, Almanya'da sağlık sektörünü incelemek amacıyla toplanmış bir panel verinin en son yılı olan 1994'e ait yatay-kesit olarak oluşturulacaktır. Yaş, eğitim ve cinsiyetin gelir üzerine etkisi, doğrusal olmayan bir regresyon ile açıklanmaya çalışılacaktır⁵. Greene'de (2003) tanımlanmış ilgili model denklem (2.9)'da görülebilir.

$$E[\text{gelir}_i | \text{yaş}_i, \text{eğitim}_i, \text{cinsiyet}_i] = \exp(\beta_0 + \beta_1 \text{yaş}_i + \beta_2 \text{eğitim}_i + \beta_3 \text{cinsiyet}_i) \quad (2.9)$$

$$E[\text{gelir}_i | \text{yaş}_i, \text{eğitim}_i, \text{cinsiyet}_i] = \mu(X_i, \beta)$$

Modeli tahmin etmek için Greene (2003) gibi aşağıdaki gamma dağılımı kullanılacaktır:

$$f(y_i | X_i, \delta) = \frac{1}{\Gamma(\zeta) \mu(X_i, \beta)^\zeta} \exp[-y_i / \mu(X_i, \beta)] y_i^{\zeta-1}. \quad (2.10)$$

Gösterimde, $\delta = [\zeta, \beta_0, \beta_1, \beta_2, \beta_3]$, $y_i := \text{income}$ (gelir) ve $X_i = [\text{age}_i, \text{educ}_i, \text{Female}_i]$ (yaş, eğitim ve cinsiyet) değişkenleri olarak tanımlanacaktır. Ayrıca, gelirin şartlı dağılımının denklem (2.10)'dan geldiği farz edilirse, gelirin yaş, eğitim ve cinsiyet veri iken beklenen değerinin denklem (2.9) tarafından belirleneceği basit bir cebir ile gösterilebilir. Denklem (2.10)'u baz alarak oluşturulacak EÇÖ tahmini aslında, doğrusal olmayan bir denklem olan (2.9)'u gamma dağılımı yardımıyla tahmin edecektir.

EÇÖ tahminini gerçekleştirmek üzere ilk önce log-olabilirlik fonksiyonu yazılacaktır.

$$\mathcal{L}_n(\delta) = -n \ln(\Gamma(\zeta)) - \zeta \sum_{i=1}^n \ln[\mu(X_i, \beta)] - \sum_{i=1}^n [y_i / \mu(X_i, \beta) - (\zeta - 1) \ln(y_i)]. \quad (2.11)$$

⁵ Bu veri setine "income.dta" dosyasından ulaşılabilir.

Log-olabilirlik fonksiyonuyla EÇO tahminini gerçekleştirmek için Stata'da aşağıdaki komutlar kullanılacaktır.

```
mlexp ("log-olabilirlik"), vce(opg)
"log-olabilirlik" := ( -_N*lngamma({zeta}) {zeta}*({beta0}+{beta1}*age+
{beta2}*educ+{beta3}*Female)-income/exp({beta0}+ {beta1}*age+{beta2}*educ+
{beta3}*Female)+({zeta}-1)*ln(income))
```

Yukarıdaki komutta **vce(opg)** opsiyonu, EÇO tahminlerinin varyans-kovaryans matrisini gradyanın dış çarpımı şeklinde hesaplanan bilgi matrisi kullanılarak oluşturmak için gereklidir. Bu opsiyonun diğer alternatifleri ise, "**oim**" gözlemlenen bilgi matrisi, "**robust**" Huber/White heteroskedastisite ve özilinti düzeltilmiş varyans kovaryans matrisi vb. şeklinde listelenmiştir. Ayrıca, log-olabilirlik fonksiyonundaki "**_N**" stata dilinde toplam gözlem sayısını ve **lngamma()** fonksiyonu ise ilgili argümanının logaritmik gamma fonksiyonunu vermektedir. Tablo 2.5'te Stata komutunun çalıştırılmasından elde edilen çıktı görülecektir.

Tablo 2.5. Doğrusal Olmayan Gamma Regresyonu EÇO Tahmini

```
mlexp ("log-olabilirlik"), vce(opg)
"log-olabilirlik" := ( -_N*lngamma({zeta}) {zeta}*({beta0}+{beta1}*age+
{beta2}*educ+{beta3}*Female)-income/exp({beta0}+ {beta1}*age+{beta2}*educ+
{beta3}*Female)+({zeta}-1)*ln(income))
```

Maximum likelihood estimation

Log	likelihood	=				
1353886,6					Number of Obs. = 3377	
	Coef.	Std. Err.	z	P>z	[95% Conf. Interval]	
/zeta	1,4617	0,0915	15,97	0,000	1,2824	1,6411
/beta0	7,3021	0,1950	37,45	0,000	6,9199	7,6843
/beta1	0,0026	0,0024	1,08	0,278	-0,0021	0,0072
/beta2	0,0521	0,0119	4,37	0,000	0,0287	0,0754
/beta3	0,0021	0,0516	0,04	0,968	-0,0990	0,1032

Bu tabloya göre, gelir üzerine etkilere bakıldığında cinsiyet ve yaş değişkenlerinin katsayılarının istatistiki olarak anlamsız, fakat eğitim düzeyinin katsayısı ise istatistiki olarak anlamlı ve pozitif olarak karşımıza çıkmaktadır. Ayrıca, gamma dağılımının skala parametresi ζ da istatistiki olarak anlamlı çıkmıştır.

Örnek üzerinde, 2.1.7'de açıklanan hipotez testleri Stata'nın olabilirlik oranı ve Wald testlerini kolaylıkla yapabilmesine olanak sağlayan fonksiyonları yardımıyla uygulanacaktır. Örneğin, $H_0: \beta_1 = \beta_2 = 0$ sıfır hipotezi çift taraflı alternatife karşı test etmek istensen. Olabilirlik oranı testini kullanabilmek için, hipotezin kısıtları altında oluşturulmuş modeli tahmin etmek ve öncesinde kısıtsız modelimizin sonuçlarını sisteme kaydetmek gerekmektedir. "**estimates store full**" konutunu kullanarak kısıtsız modelin çıktıları kaydedilebilir ("**full**" yerine istenilen başka bir isim de yazılabilir). Daha sonra,

komutu yazılarak Wald testi çalıştırılabilir. Söz konusu komutta birden fazla kısıt olduğu için her bir kısıt ayrı bir parantezde yazılmalıdır. Ayrıca, örnekteki parametreler doğrudan bir değişkenle ilişkilendirilmediği için bu parametreleri çağırırken “_b[*parametre ismi*:_cons]” yapısı kullanılacaktır. Parametre isimleri, tahmin çıktı tablosunda değişken ismi “/” olmadan hâli olarak düşünülebilir. Bu komutu kullanarak test çalıştırıldığında Wald test istatistiği 1,19 ve bu testin p-değerini 0,5504 olarak hesaplanmaktadır. Bu durumda, yine sıfır (H_0) hipotez reddedilememektedir.

Bu örnekte son olarak doğrusal olmayan bir kısıt, Wald istatistiğiyle test edilecektir. İlgilendiğimiz hipotez $H_0: \beta_1\beta_2\beta_3 = 0$ olsun, yani yaş, eğitim durumu veya cinsiyet değişkenlerinden en az bir tanesinin anlamsız olması (bu değişkenlerin katsayılarının hepsinin sıfıra eşit olduğu) test edilsin. Bu testi gerçekleştirmek için aşağıdaki komut kullanılacaktır:

```
testnl (_b[beta1:_cons]*_b[beta2:_cons]*_b[beta3:_cons]=0 )
```

Elde edilen sonuçlara göre, Wald istatistiği 0,00165041 ve p-değeri ise 0,9676 çıkmaktadır. Bu durumda istatistiki olarak değişkenlerden en az bir değişkenin katsayısının sıfır olduğu yargısına varılabilir.

2.3.3. Örnek 3: Doğrusal Olmayan Regresyon Modelinin “ml” Fonksiyonuyla İncelenmesi

Bu örnekte Cameron ve Triverdi’nin (2009) ABD’deki bireylerin doktor muayenelerini ziyaret sayısını modellediği veri seti kullanılacaktır (“mus10.dta”⁶). Cameron ve Triverdi’den (2009) biraz ayrılmak için bu analizde 2002 yerine 2001 verisiyle devam edilecektir. Bu veri setinde değişken olarak hastaların doktor muayenehanelerini ziyaret sayısını veren “docvis” değişkenini; açıklayıcı değişken olarak da hastanın özel sağlık sigortası durumunu gösteren kukla değişken “private” (eğer hastanın özel sağlık sigortası mevcutsa private=1), hastanın kronik rahatsızlığı olup olmadığını gösteren kukla değişken “chronic” (eğer hastanın kronik rahatsızlığı var ise chronic=1), hastanın cinsiyetini belirten kukla değişken “female” (eğer hasta kadın ise female=1) ve hastanın gelir durumunu gösteren “income” değişkeni kullanılacaktır. Bu örneği analiz ederken 2.2’inci kısımda açıklanan doğrusal olmayan model kullanılacaktır. Bu modelin, anakütle regresyon fonksiyonu, $X_i = [1, private_i, chronic_i, female_i, income_i]'$ ve $\beta = [\beta_0, \dots, \beta_4]'$ olmak üzere

$$E[docvis_i|X_i] = \exp(X_i'\beta), \quad (2.12)$$

şeklinde tanımlanacaktır. Modelleme ile ortalama doktor ziyareti sayısının pozitif olması gibi doğrusal regresyon modelinde elde edilemeyen bir özelliği kullanılabilir. Ancak, modelin yorumlaması doğrusal modellere göre biraz daha karmaşık olacaktır.

Analizde farklı yöntemleri karşılaştırmak için, denklem (2.12)’deki regresyon fonksiyonunu veren üç model üzerinde çalışılacaktır. Bu modeller sırasıyla Poisson regresyon modeli, doğrusal olmayan en küçük kareler tahmini ve normal dağılım altında EÇÖ tahmin edicisi şeklinde sıralanacaktır.

⁶ “mus10.dta” verisine Uygulamalarla Mikro Ekonometri Kitabının web sayfasından ulaşabilirsiniz.

Tahmin yöntemlerine geçmeden önce kullanılacak veri setinin temel özellikleri incelenecektir. Bunun için Stata'da aşağıdaki komutların kullanılması gerekecektir.

Tablo 2.8. Analizde Kullanılacak Değişkenlerin Tanımlanması

```
use mus10data.dta, clear /* hedef dizinde bulunan veri seti yükleyecek */

keep if year01==1 /* year01 2001 yılına ait verilerin kukla değişkeni olduğu
için sadece bu yıla ait gözlemleri kullanmak için keep fonksiyonunu kul-
lanılacaktır. */

(23,940 observations deleted)

describe docvis private chronic female income /* Regresyonda kullanılacak de-
ğişkenlerin özelliklerini özetliyecektir */

Variable  storage  display  value
  name      type    format   label  variable label
docvis      int     %8.0g          number of doctor visits
private     byte     %8.0g          = 1 if private insurance
chronic      byte     %8.0g          = 1 if a chronic condition
female      byte     %8.0g          = 1 if female
income      float    %9.0g          Income in $ / 1000
```

Değişken tiplerimizi inceledikten sonra değişkenlerin temel istatistiki bilgilerini özetlemek için aşağıdaki komut kullanılacaktır:

Tablo 2.9. Analizde Kullanılacak Değişkenlerin Temel İstatistiki Göstergeleri

```
summarize docvis private chronic female income /* Regresyonda kullanılan de-
ğişkenlerin temel istatistiklerini özetleyecektir*/
```

Variable	Obs	Mean	Std. Dev.	Min	Max
docvis	6.184	3,896345	7,873603	0	152
private	6.184	0,824709	0,380247	0	1
chronic	6.184	0,341203	0,474152	0	1
female	6.184	0,471863	0,499248	0	1
income	6.184	34,53861	28,15369	-41,952	238,352

Yukarıdaki Stata çıktısına göre toplamda 2001 yılında 6.184 hasta üzerinde analiz gerçekleştirilmektedir. Örnek uzaydaki hastalar yılda ortalama 3,89 kez doktorlara muayene olmaktadır. Hastalarda hiç muayene olmayan olduğu gibi toplamda 152 kez doktorları ziyaret eden bireyler de vardır. Ayrıca, yukarıdaki çıktıya göre 2001 yılında hastaların 82%'sinin özel sigortası varken, hastaların 34%'ünün kronik bir rahatsızlığı mevcuttur. Son olarak, hastaların 47% sini kadınlar oluşturmaktadır.

2.3.3.1. Doktor Ziyareti Modelinin Poisson Regresyon Tahmini

İlk model olan poisson regresyonu diğer bölümlerde daha detaylı anlatılacağı için bu kısımda detaylara yer verilmeyecek. Denklem (2.12)'yi poisson regresyonunu sağlam (dirençli) standart sapma (**vec(robust)**) komutu ile tahmin etmek için aşağıdaki komut kullanılır:

Tablo 2.10. Doktor Ziyareti Modelinin Poisson Regresyon Tahmin Sonuçları

```
poisson docvis private chronic female income , vce (robust)
```

Iteration 0: log pseudolikelihood = -25839,148

Iteration 1: log pseudolikelihood = -25838,554

Iteration 2: log pseudolikelihood = -25838,554

Poisson regression	Number of obs = 6.184
	Wald chi2(4) = 709,97
	Prob > chi2 = 0,0000
Log pseudolikelihood=-25838,554	Pseudo R2 = 0,1714

		Robust				
docvis	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
private	0,724498	0,091021	7,96	0,000	0,546100	0,902896
chronic	1,08513	0,048772	22,25	0,000	0,989539	1,180722
female	0,449829	0,051972	8,66	0,000	0,347966	0,551692
income	0,002339	0,000744	3,15	0,002	0,000882	0,003796
_cons	-0,13209	0,091969	-1,44	0,151	0,312345	0,048166

Tablo 2.10 çıktısını tanımlamak gerekirse, “Log pseudolikelihood” tahmin edilen katsayı değerleri için sözde log-olabilirlik değerini, “Wald chi2(4)” modelin genel anlamlılık düzeyini test ($H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$ hipotezinin testi) etmek için kullanılan Wald istatistiğini, “Prob > chi2” bu testin p-değerini, “Pseudo R2” Stata tarafından rapor edilen modelin uyum iyiliği ($(\mathcal{L}_n(\hat{\theta}_U))$ tahmin edilen modelin log-olabilirliği ve $\mathcal{L}_n(\hat{\theta}_R)$ ise $\beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$ kısıtı altında tahmin edilen modelin log-olabilirliği olmak üzere $PseudoR2 = (1 - \mathcal{L}_n(\hat{\theta}_U)/\mathcal{L}_n(\hat{\theta}_R))$ şeklinde tanımlanabilir), “Coef.” satır isminde belirtilen değişkenin EÇO tahmini, “Robust Std. Err.” Aynı satırdaki EÇO tahminlerinin sağlam standart sapmaları, “z” EÇO tahminlerinin bireysel anlamlılık test istatistikleri, “P>|z|” bu testin p-değeri ve “[95% Conf. Interval]” ise EÇO tahminin 95% güven aralığı olduğu söylenebilir.

Tablo 2.10'daki Pseudo-R2'ye göre, bağımsız değişkenler bağımlı değişkendeki varyasyonun %17'sini açıklarken, Wald testi modelin genel olarak anlamlı olduğunu göstermektedir. Ayrıca, sabit terim haricinde bütün katsayılar z istatistiği göz önünde bulundurulduğunda %1'lik anlamlılık düzeyinde istatistiki olarak sıfırdan farklı çıkmakta ve artı değer almaktadır.

2.3.3.2. Doktor Ziyareti Modelinin Doğrusal Olmayan En Küçük Kareler ile Tahmini

Poisson regresyon modelinde doktor ziyareti değişkenin şartlı dağılımın Poisson olduğu kabul edilmiştir. Ancak, böyle bir dağılım varsayımı yapmadan da denklem (2.12) tahmin edilebilir.

Tahmin için, u_i hata terimi olmak üzere: $docvis_i = \exp(X_i'\beta) + u_i$ denklemi kullanılabilir. Söz konusu modelin hata teriminin karelerinin toplamını $\sum_{i=1}^n [docvis_i - \exp(X_i'\beta)]^2$ katsayı değerleri β' 'ya göre minimize edilerek bulunabilir. Minimizasyonun EKK'nin aksine kapalı form sonucu yoktur ve sonuçları aynı EÇO gibi numerik optimizasyon yapılarak elde edilir. Bu bölümün kapsamı dışında olduğu için detaylara değinilmeyecektir, ancak, doğrusal olmayan EKK için detaylı bilgi Cameron ve Triverdi (2009) ve Davidson ve Mackinnon'da (2005) bulunabilir. Yöntemi çalıştırabilmek için aşağıdaki Stata komutlarına başvurulabilir:

Tablo 2.11. Doktor Ziyareti Modelinin Doğrusal Olmayan En Küçük Kareler ile Tahmin Sonuçları

```
nl (docvis = exp({private}*private + {chronic}*chronic + {female}*female
+{income}*income + {intercept} ) )
(obs = 6,184)
```

```
Iteration 0: residual SS = 355282,5
Iteration 1: residual SS = 341294,5
Iteration 2: residual SS = 339783,2
Iteration 3: residual SS = 339779,1
Iteration 4: residual SS = 339779,1
Iteration 5: residual SS = 339779,1
Iteration 6: residual SS = 339779,1
Iteration 7: residual SS = 339779,1
```

Source	SS	df	MS	Number of obs	=	6.184
Model	137409,9	5	27481,98	R-squared	=	0,288
Residual	339779,1	6179	54,98934	Adj R-squared	=	0,2874
Total	477189	6184	77,1651	Root MSE	=	7,41548
				Res. dev.	=	42324,58

docvis	Coef.	Std. Err.	t	P>t	[95% Conf. Interval]	
/private	0,604398	0,100788	6,00	0,000	0,4068175	0,8019780
/chronic	1,055272	0,053838	19,60	0,000	0,9497318	1,1608130
/female	0,416087	0,045207	9,20	0,000	0,3274664	0,5047081
/income	0,002073	0,000671	3,09	0,002	0,0007572	0,0033885
/intercept	0,035175	0,110597	0,32	0,750	-0,1816348	0,2519837

Tablo 2.11'de, poisson regresyon tablosuna göre birkaç farklı istatistik bulunmaktadır. "R-squared" ve "Adj R-squared" EKK'den bilinen R-kare ve düzeltilmiş R-kareye denk gelmektedir. Ayrıca "Source" başlığı altında yer alan değerlerden "Model SS" tahmin edilen modelin varyasyonu, "Model df" serbestlik derecesi, "Model MS" ise SS/df değerini verirken, bir alt satırda söz konusu değerler artık (hata terimi-kalıntı) terim için hesaplanmıştır. Bölümün en alt satırında ise bağımlı değişkenin ilgili istatistiklerine "Total" başlığı altında yer verilmiştir. Tablodaki diğer bir farklılık ise, anlamlılık düzeyini test etmek için z istatistiği yerine t istatistiği kullanılmıştır. Söz konusu durum EKK ile özdeştir.

Sonuçlara bakıldığında, regresyondan elde edilen sonuçlar hem katsayı değerleri hem de anlamlılık testlerine göre poisson regresyona çok benzemektedir. Sadece, istatistiki olarak anlamsız olan sabit katsayısının işareti değişmekte, artı değer almaktadır.

2.3.3.3. Doktor Ziyareti Doğrusal Olmayan Modelinin En Çok Olabilirlik Tahmin Edicisiyle Tahmini

Şimdi incelenen kısımda ise, “ml” komutu yardımıyla EÇO tahmin edicisini nasıl elde edilebileceği üzerine durulacaktır. Denklem (2.12)’yi tahmin etmek için öncelikle hata terimi u_i için bir dağılım varsayılması gerekmektedir. Gösterim amaçlı olarak, $u_i \sim \mathcal{N}(0, \sigma^2)$, sıfır ortalamalı ve σ^2 varyanslı normal bir rastsal değişken olarak kabul edilecektir. Dolayısıyla, regresyon modelinin log-olabilirliği 2.13’teki şekilde yazılabilir:

$$\mathcal{L}_n(\beta, \sigma^2) = -0,5n\ln(2\pi) - 0,5n\ln(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n [\text{docvis}_i - \exp(X_i'\beta)]^2. \quad (2.13)$$

EÇO tahmin edicisini “ml” komutuyla yazabilmek için öncelikle log-olabilirliği yani objektif fonksiyonunun Stata’ya tanıtılması gerekmektedir. Bunu yapmak için aşağıdaki komut “program ... end” sözdizimi kullanılabilir.

```
program lfnlnorm /*lfnlnorm log-olabilirlik fonksiyonunun belirlediğimiz ismi */
  args lnf Xbeta Sigma /* "args" ile komut içerisinde kullanacağımız değişken-
  leri tanımlıyoruz. lnf: log=olabilirlik fonksiyonu, Xbeta: Xi*beta, Sigma:
  \sigma^2 */
  quietly replace `lnf' = ln(normalden($ML_y1 - exp(`Xbeta'), `Sigma')) /* BU ko-
  mut log=olabilirliği herhangi bir çıktı yazdırmadan (quietly) docvis-
  exp(`Xbeta') ortalamalı ve Sigma varyanslı normal dağılımın logaritması ola-
  rak dolduracak */
end
```

Yukarıdaki komut hakkında birkaç açıklama yapılacaktır. İlk olarak \$ML_y1 ifadesi evrensel bir ifade olup bağımlı değişkeni (burada docvis) ifade edecektir. Ayrıca, log-olabilirlik içerisinde tanımlanan ögeler “`öge’ ” şeklinde yazılması gerekmektedir. Bu ögeler EÇO tahmininde kullanıcı tarafından Stata’ya tanıtılacaktır. Log-olabilirlik fonksiyonu toplam hâlinde görülmemektedir. Stata toplama işini kendiliğinden yapacaktır.

Tanımlanan log-olabilirlik fonksiyonu ile tahmin yapabilmek için, EÇO problemini de tanımlamak gerekecektir. Tanımlama aşağıdaki komutla yapılacaktır:

```
ml model lf lfnlnorm (Xbeta: docvis = private chronic female income)
(Sigma:), vce(robust)
```

Bu komutta ilk üç öge standart olup “lf” (Stata’daki en basit olabilirlik yöntemi) yöntemiyle model tahmin edilebilecektir. Söz konusu ögelerden sonra yukarıda tanımlanan log-olabilirlik fonksiyonunun ismi yazılmalıdır. İsmi ise, parantez içerisinde model takip etmektedir. Burada birkaç detaya değinmekte fayda olacaktır. Öncelikle, iki parantez grubu arka arkaya gelmektedir. Her iki parantez içerisinde ilk önce denklem ismi yazılır ve iki nokta üstü üste işareti konulup denklem yazımına geçilir. İlk grup, yukarıdaki tanımlamada “lfnlnorm” programındaki “Xbeta”ya

denk gelirken, ikinci parantez ise “**Sigma**” objesine denk gelecektir. Eğer daha fazla obje tanımlanmak istenirse, tanımlanan obje sayısı kadar parantez kullanmak gerekecektir. “**Xbeta**”nın içeriğini tanımlayan ilk parantezdeki format ile “**lfnlnorm**” fonksiyonu karşılaştırılırsa, ikili arasındaki ilişkinin ($\$ML\_y1 = \exp(\text{“Xbeta”})$) içerisindeki değişkenler) yapısından geldiği götürülecektir. İkinci parantezin denklem ismi haricinde boş olması ise paranteze denk gelen “**Sigma**” objesinin sadece bir parametre olduğunu göstermektedir. Son olarak, virgülden sonraki **vce()** opsiyonu EÇO tahmin edicilerin standart sapmalarının nasıl tahmin edileceğini vermektedir. Opsiyonlar Tablo 2.8’de görülebilir. Aşağıdaki örnekte, Huber/White sandviç tahmin edicisi kullanılacaktır.

Tablo 2.12. En Çok Olabilirlik Tahmin Edicilerin Standart Sapma Tahmin Edici Opsiyonları

Olabilirlik tabanlı	Açıklama
<i>oim</i>	Gözlemlenen Bilgi Matrisi
<i>opg</i>	Gradyan vektörlerin dış çarpımı
Sandviç Tahmin edicisi	
<i>robust</i>	Huber/White sandviç tahmin edicisi
<i>cluster clustvar</i>	Kümelenmiş sandviç tahmin edicisi
Tekrarlama tabanlı	
<i>bootstrap</i>	Bootstrap tahmini
<i>jackknife</i>	Jackknife tahmini

Tahmin yöntemini çalıştırmak için gerekli komutu ve tahmin sonuçları şu şekilde verilmektedir:

Tablo 2.13. Doktor Ziyareti Modelinin En Çok Olabilirlik Tahmin Sonuçları

ml maximize

initial:log pseudolikelihood = -<inf>(could not be evaluated)

feasible:log pseudolikelihood = -671053,84

rescale:log pseudolikelihood = -30331,502

rescale eq:log pseudolikelihood = -22179,439

Iteration 0:log pseudolikelihood = -22179,439(not concave)

Iteration 1:log pseudolikelihood = -22066,685

Iteration 2:log pseudolikelihood = -21387,628

Iteration 3:log pseudolikelihood = -21165,67

Iteration 4:log pseudolikelihood = -21162,329

Iteration 5:log pseudolikelihood = -21162,29

Iteration 6:log pseudolikelihood = -21162,29

	Number of obs	=	6.184
	Wald chi2(4)	=	618,33
Log pseudolikelihood = -21162,29	Prob > chi2	=	0

Robust					
docvis	Coef.	Std. Err.	z	P>z	[95% Conf. Interval]

Xbeta						
private	0,604398	0,095436	6,33	0,000	0,417347	0,791448
chronic	1,055272	0,047629	22,16	0,000	0,96192	1,148624
female	0,416087	0,057503	7,24	0,000	0,303385	0,52879
income	0,002073	0,000783	2,65	0,008	0,000538	0,003608
_cons	0,035174	0,100365	0,35	0,726	-0,16154	0,231887
Sigma						
_cons	7,412482	0,442208	16,76	0,000	6,54577	8,279194

Tablo 2.13'teki çıktıları yorumlamadan önce, önemli bir ayrıntı göze çarpabilir. Doğrusal olmayan EKK tahminleri ile modelin tahminlerinin birbirine özdeş olduğu görülmektedir. Aslında söz konusu durum matematiksel olarak karşımıza çıkmaktadır. Her iki yöntem de açıklayıcı değişkenlerin katsayılarını $\sum_{i=1}^n [docvis_i - \exp(X_i'\beta)]^2$ ifadesini minimize ederek bulmaktadır. Ancak, standart sapmalar ve dolayısıyla a test istatistikleri birbirlerine yakın olsa da aynı değildir. Fark olabilirlik fonksiyonundan ve de farklı `vce()` opsiyonu kullanmaktan kaynaklanmaktadır. Bunlar dışında, kısım 2.3.3.2'de yapılan bütün analizler aynen bu tahmin çıktıları içinde geçerlidir.

Tablo 2.13'teki örnekte (gamma regresyon modelinde) hipotez testlerine değinildiği için şimdiki kısımda bu testlere yer verilmemiştir. Ancak, Stata, aynı "`mlexp`" komutunda olduğu gibi "`ml`" komutuyla yapılan tahminlerde Wald ve Olabilirlik oranı testlerine izin vermektedir. Söz konusu testler yerine doğrusal olmayan modellerde önemli bir analiz olan marjinal etkilerin tahminine yer verilecektir. Marjinal etki (kısmi etki), üçüncü kısımda anlatıldığı gibi, bir açıklayıcı değişkendeki küçük değişimin, diğer açıklayıcı değişkenler sabit iken, bağımlı değişkenin ortalama değerini ne kadar değiştirdiği şeklinde tanımlanabilir (Cameron and Triverdi, 2009). Matematiksel olarak söz konusu etki, Tablo 2.13'teki örnek üzerinde, j 'inci sürekli (ayrık) değişken için $ME_j = \partial E[y|X = X^*] / \partial X_j$ ($ME_j = E[y|X = X^*, X_j = x_{j,1}] - E[y|X = X^*, X_j = x_{j,0}]$) şeklinde tanımlanabilir. Stata'da "`margins`" komutunu kullanarak hem marjinal değerleri hem de marjinal değerlerin standart sapmaları elde edilebilir. Tablo 2.13'teki örnekten yola çıkarak iki farklı marjinal etki tanımlaması incelenecektir.. Birincisi ortalama marjinal etki olup, şu şekilde elde edilecektir:

Tablo 2.14. Ortalama Marjinal Etki Sonuçları

<code>margins, dydx(*) expression(exp(predict(equation(Xbeta))))</code>						
Average marginal effects			Number of obs = 6.184			
Model VCE : Robust						
Expression : <code>exp(predict(equation(Xbeta)))</code>						
dy/dx w.r.t. : private chronic female income						
	dy/dx	Delta-method Std. Err.	z	P> z	[95% Conf. Interval]	
private	2,373612	0,380162	6,24	0,000	1,628509	3,118715
chronic	4,144322	0,223525	18,54	0,000	3,706221	4,582423
female	1,63408	0,226605	7,21	0,000	1,189942	2,078217
income	0,008141	0,003066	2,66	0,008	0,002131	0,014150

Tablo 2.14’de görüldüğü gibi “**margins**” komutu ön ayarlı opsiyonlarla kullanılırsa bütün gözlemler (gözlemlerdeki açıklayıcı değişken değerinde) için marjinal etkileri hesaplayıp, bunların ortalamasını (ortalama marjinal etki) hesaplayacaktır. “**dydx(*)**” ifadesi marjinal etkileri hesaplamak için, yıldız ise bütün değişkenler için söz konusu işlemi yapmak için gereklidir. Ek olarak, “**expression**” komutu bağımlı ve açıklayıcı değişkenler arasındaki ilişkiyi kurgulamaktadır. Bu komutta, “**predict(equation(“denklem ismi”))**”, yukarıda tanımlana “**Xbeta**” denkleminin tahminini vermektedir. Tahminin üstel fonksiyonu olarak, bağımlı değişkenle açıklayıcı değişkenler arasındaki doğrusal olmayan ilişki Stata’ya tanıtılabilir. Eğer, kullanıcı tarafından yazılan bir olasılırlık üzerine çalışılıyor ve bu opsiyon es geçilmişse, ilişki doğrusal olarak kabul edilecek ve marjinal etkiler EÇO tahmin edicisine eşitlenecektir. Dolayısıyla, örnek için doğrusal tahminin üstelini alarak söz konusu durum düzeltiler. Yukarıdaki sonuçlara göre bütün marjinal etkiler 1%’lik anlamlılık düzeyinde sıfırdan farklıdır. En yüksek marjinal etki kronik hastalığı olanlarda bulunmaktadır. Gelirin marjinal etkisi nispeten küçüktür. Ayrıca, kadınların erkeklerden ortalamada daha fazla doktora gittiği söylenebilir. Bir taraftan, özel sigortaya sahip olanlar olmayanlara göre ortalamada 2,37 kez daha fazla doktora görünmektedir. Aynı şekilde, kronik hastalığı bulunanlar bulunmayanlara göre 4,14 kez daha fazla doktor ziyaretinde bulunmaktadır. Diğer taraftan, gelir düzeyindeki 1 birimlik (1/1000 Dolar) artış, doktor a gitme sayısını ortalamada 0,008 kez artırmaktadır.

İkinci incelenecek marjinal etki tipi ise Ortalamada Marjinal etki olup, bu etki sadece açıklayıcı değişkenlerin ortalama değeri için hesaplanmaktadır. Bu etkiyi elde etmek için bir önceki komuta “**atmeans**” opsiyonu eklenerek şu komut yazılmalıdır:

Tablo 2.15. Ortalamada Marjinal Etki Sonuçları

margins, dydx(*) atmeans expression(exp(predict(equation(Xbeta))))

		Number of obs		=	
Conditional marginal effects		6.184			
Model VCE : Robust					
Expression : exp(predict(equation(Xbeta)))					
dy/dx w.r.t. : private chronic female income					
At	private	=	,8247089	(mean)	
	chronic	=	,3412031	(mean)	
	female	=	,4718629	(mean)	
	income	=	34,53861	(mean)	
Delta-method Std.					
	dy/dx	Err.	z	P> z	[95% Conf.Interval]
private	1,931104	0,296677	6,51	0,000	1,349628 2,512579
chronic	3,371703	0,151142	22,3	0,000	3,075470 3,667936
Female	1,329441	0,174460	7,62	0,000	0,987505 1,671377
income	0,006623	0,002484	2,67	0,008	0,001755 0,011491

Tablo 2.15’teki çıktılar ortalama marjinal etki değerlerine hem marjinal etki hem de bu etkilerin standart sapmaları bağlamında benzemektedir. Bu tablolara bakarak, marjinal etki analizinin

sağlam olduğunu ve marjinal etki kurgusuna çok bağımlı olmadığı söylenebilir. Son olarak, marjinal etkileri belirli açıklayıcı değişken değerlerinde hesaplanacaktır. Mesela şu sorunun cevabını aransın: “Özel sigortası olmayan ve kronik rahatsızlığı olan ve \$20,000 geliri olan bir kadının gelir durumu artarsa, bu durumun doktor ziyaret sayısına etkisi nedir?” Bu soruyu şu komutla analiz edilebilir:

Tablo 2.16. Belirli Değerlerde Marjinal Etki Sonuçları

```
margins, dydx(income) at(private=0 chronic=1 female=1 income=20) expres-
sion(exp(predict(equation(Xbeta))))
```

```
Conditional marginal effects      Number of obs =      6.184
Model VCE      : Robust
Expression: exp(predict(equation(Xbeta)))
dy/dx w.r.t. : income
at :   private =      0
      chronic  =      1
      female   =      1
      income   =     20
```

Delta-method						
	dy/dx	Std. Err.	z	P> z	[95% Conf. Interval]	
income	0,009747	0,003906	2,5	0,013	0,002091	0,017402

Bu sonucu elde etmek için “at()” opsiyonu içinde hangi değişkenin hangi değerde olması isteniyorsa yazılabilir. Burada, bütün değişkenlere değer girmek gibi bir zorunluluk yoktur. Değer girilmemesi durumunda Stata ortalama marjinal etkiyi hesaplayacaktır. Yani, Stata bütün gözlemler için eksik bırakılan değişkenlerin değerlerinde marjinal etkiyi bulup bunları toplayacaktır. Ayrıca “dxdy()” içerisinde sadece “income” yazılması, sadece ilgili değişken için marjinal etkilerin hesaplanmasına neden olacaktır. Tablo 2.16’ya göre “Özel sigortası olmayan ve kronik rahatsızlığı olan ve \$20,000 geliri olan bir kadının geliri marjinal olarak artarsa, kadının doktor ziyaret sayısı ortalamada 0,009 kadar artacaktır. Buna ek olarak, bu etki 5% anlamlılık düzeyinde istatistiki olarak anlamlıdır.

2.4. SONUÇ

Kitabın bu bölümünde sosyal bilimler alanında sıklıkla kullanılan EÇO tahmin edicisi tanıtılmaya ve bu tahmin edicinin özellikleri anlatılmaya çalışılmıştır. Tahmin edicinin verinin belirli dağılım özellikleri kullanılarak parametrik yöntemlerle nasıl analiz edileceği gösterilmiştir. EÇO tahmin edicilerinin belirli şartlar altında hem doğrusal hem de doğrusal olmayan modellerde, tutarlılık, asimptotik normallik ve etkinlik özelliklerini sağladığından bahsedilmiştir.

Bölümde doğrusal olmayan modellerin EÇO yardımıyla tahmininin Stata programı yardımıyla nasıl yapılacağı konusuna bir giriş yapılmıştır. Bu girişte parametre ve parametrelerin standart hata tahminleri, parametreler üzerinde yapılabilecek doğrusal ve doğrusal olmayan hipotez testlerinden ve doğrusal olmayan modellerde marjinal etkilerden bahsedilmiştir. Yapılan tahminler-

de temel Stata komutları kullanılmıştır. Ancak, kitabın diğer bölümlerinde kullanılacak bazı EÇO tabanlı tahmin gerektiren modeller için Stata'nın hazır modülleri olduğunu belirtmek isteriz. Örneğin, Stata'nın "Statistics" modülü altında "(İkili sonuçlar (Binary outcomes))", "Sıralanmış sonuçlar (Ordered outcomes)", "Saymalı sonuçlar (Count outcomes)" vb. başlıklı modellerde Stata'nın EÇO kullandığı söylenebilir. Ancak, anlatılan bölümde kullanıcı tarafından olabilirlik fonksiyonu düzenlenebilen modeller üzerinde durmaya çalışılmıştır.

2.5. KAYNAKLAR

- Andrews, D. W. (1999). Estimation when a parameter is on a boundary. *Econometrica*, 67(6), 1341-1383.
- Buse, A. (1982). The likelihood ratio, Wald, and Lagrange multiplier tests: An expository note. *The American Statistician*, 36(3a), 153-157.
- Cameron, A. C., & Trivedi, P. K. (2005). *Microeconometrics: methods and applications*. Cambridge university press.
- Cameron, A. C., & Trivedi, P. K. (2009). *Microeconometrics using stata* (Vol. 5, p. 706). College Station, TX: Stata press.
- Davidson, R., & MacKinnon, J. G. (2004). *Econometric theory and methods* (Vol. 5). New York: Oxford University Press.
- Greene, W. H. (2003). *Econometric analysis*. Pearson Education India.
- Hamilton, J. D. (1994). *Time Series Analysis*. Princeton University Press, Princeton
- Holland, J. (1975). *Adaptation In Natural and Artificial Systems*. University of Michigan Press, Ann Arbor.
- Huber, P. J. (1967), "The Behavior of Maximum Likelihood Estimates under Nonstandard Conditions," in *Proceedings of the Fifth Berkeley Symposium*, J. Neyman (Ed.), 1, 221-233, Berkeley, CA, University of California Press
- Kennedy, J., & Eberhart, R. (1995). Particle swarm optimization. In *Proceedings of ICNN'95-International Conference on Neural Networks* (Vol. 4, pp. 1942-1948). IEEE.
- Nowak, R. D., (2011). ECE 830 Fall 2011 Statistical Signal Processing [Ders notları]. University of Wisconsin-Madison.
https://nowak.ece.wisc.edu/ece830/ece830_fall11_lecture15.pdf
- Pincus, M. (1970). Letter to the editör-a Monte Carlo method for the approximate solution of certain types of constrained optimization problems. *Operations research*, 18(6), 1225-1228.
- Riphahn, R. T., Wambach, A., and Million, A., (2003) "Incentive Effects in the Demand for Health Care: A Bivariate Panel Count Data Estimation", *Journal of Applied Econometrics*, Vol. 18, No. 4, , pp. 387-405.
- White, H. (1982), "Maximum Likelihood Estimation of Misspecified Models," *Econometrica*, 50, 1-25.

BÖLÜM 3

İKİ DURUMLU MODELLER VE STATA UYGULAMALARI

Hamdi Emeç *

Bağımlı değişken iki değer alan nitel değişkenlerden oluşabilmektedir. Bu bölümde bağımlı değişken nitel(sürekli) olarak olumlu-olumsuz, evet-hayır, başarılı-başarısız, kabul-red... gibi iki değere sahip olduğu modeller tanıtılacaktır. Bu modellere iki durumlu modeller denilmektedir. En çok kullanılan iki durumlu modeller; doğrusal olasılık, logit ve probit modelleridir.

3.1. Doğrusal Olasılık Modeli (DOM)

Doğrusal regresyon modelinde hata paylarına ilişkin bir kısıtlama getirilmediği gibi, normal dağılımları varsayılmakta ve bağımlı değişkenin $(-\infty, +\infty)$ aralığında tanımlanmasını gerektirmektedir. Bağımlı değişken sürekli olduğunda normal dağılım gösterdiği varsayılmasına rağmen, bağımlı değişkeni iki veya daha fazla değer alan modellerde amaç seçimin olasılığının saptanmasıdır. Çünkü bağımlı değişken sürekli değildir ve bunun için ölçme değil sayma söz konusu olmaktadır. Olasılık değeri 0 ile 1 arasında bir değer aldığından iki değerli bağımlı değişkeni kullanan doğrusal regresyon modeli “doğrusal olasılık modeli(DOM)” olarak adlandırılır (Goldberger, 1964: 248; Amemiya, 1981: 1483; Kmenta, 1971: 409; Long, 1997: 34-35).

Aşağıda (3.1)’deki basit model incelensin:

$$Y_i = \beta_0 + \beta_1 X_i + u_i \quad (3.1)$$

Bağımsız değişkenlerin doğrusal fonksiyonları olarak açıklanan DOM’nde bağımsız değişkenlerin değerleri veri iken koşullu bağımlı değişkenin beklenen değerleri, meydana gelen olayın koşullu olasılıkları olarak açıklanmaktadır (Maddala, 1992: 332). Burada X sürekli bir değişken, Y_i ise 1 ve 0 değerlerinden oluşan kategorik bir değişken olsun. Bu model klasik bir regresyon modelidir, fakat bağımlı değişken iki değerli olduğundan DOM’dır. Çünkü X_i veriyken Y_i ’nin koşullu olasılığı olarak yorumlanabilir, yani $P(Y_i = 1 | X_i)$ olmaktadır (Gujarati ve Porter, 2012: 543). Burada temel olan Y_i sıfır ve bir değerlerini alan bir ikili değişken olduğunda, $P(Y_i = 1 | X_i) = E(Y_i | X_i)$ eşitliğinin her zaman doğru olmasıdır. Böylece “başarı olasılığı”, bir başka değişle $Y_i = 1$ olasılığı, Y_i ’nin beklenen değeriyle aynıdır (Wooldridge, 2013:247). Y_i bernoulli olasılık dağılımına uyar ve beklenen değer tanımından aşağıdaki eşitlik bulunur:

* Doç. Dr., Dokuz Eylül Üniversitesi, İktisadi ve İdari Bilimler Fakültesi, Ekonometri Bölümü.
E-mail: hamdi.emec@deu.edu.tr

$$E(Y_i) = 0(1 - P_i) + 1(P_i) = P_i \quad (3.2)$$

Buradan da

$$E(Y_i|X_i) = \beta_0 + \beta_1 X_i = P_i \quad (3.3)$$

elde edilir.

(3.1) modelinin koşullu beklenen değeri, aslında Y_i 'nin koşullu olasılığı olarak yorumlanmaktadır (Gujarati ve Porter, 2012: 543). β_0 her bir bağımsız değişkenin (X_i) sıfır olarak belirlendiği durumlarda öngörülen başarı olasılığıdır. Eğim katsayısı (β_1) ise bağımsız değişkenin (X_i) bir birim arttığı durumda başarı olasılığındaki tahmin edilen değişikliği ölçer (Wooldridge, 2013:247). Regresyon katsayıları EKKY ile tahmin edilir ve EKK standart hataları, güven aralıkları ve hipotez sınamaları için kullanılır (Stock ve Watson, 2010: 391).

3.1.1. DOM'nde Sorunlar

DOM'de hataların (u_i) binom dağımlı olması, değişen varyanslı olması, olasılıkların (0,1) aralığının dışında kalabilmesi ve çoğu durumda belirlilik katsayısı değerinin en iyi modelin bulunmasında iyi bir ölçüt olmaması gibi sorunlarla karşılaşmaktadır.

3.1.1.1. Hata Terimlerinin Normal Dağılmaması Sorunu

DOM'nde hataların (u_i) normal dağıldığı varsayımının yerine getirilmesi olanaklı değildir. Çünkü u_i de aynı Y_i bağımlı değişkeni gibi iki değer almaktadır. (3.1) eşitliği aşağıdaki gibi yazıldığında:

$$u_i = Y_i - \beta_0 - \beta_1 X_i \quad (3.4)$$

elde edilir. Böylece u_i 'lerin olasılık dağılımı Tablo 3.1'de gösterildiği gibi olacaktır:

Tablo 3.1. u_i 'lerin Olasılık Dağılımı

Y_i	u_i	Olasılık
0	$-\beta_0 - \beta_1 X_i$	$1 - P_i$
1	$1 - \beta_0 - \beta_1 X_i$	P_i

Tablo 3.1'den anlaşılaçağı üzere DOM'nin hata terimleri normal dağılmamaktadır (Gujarati ve Porter, 2012: 544). Fakat hata terimlerinin normal dağılmaması EKKY'nin uygulanması için bir sakınca oluşturmamaktadır. Merkezi Limit Teoremine göre büyük örneklerde hata terimleri limitte normal dağılmaktadır (Amemiya, 1985: 92; Malirivaud, 1970: 251). Bu durumda büyük örneklemelerde DOM'ndeki hata terimlerinin normal dağılmaması bir sorun olarak araştırmacıların karşısına çıkmayacaktır.

3.1.1.2. Hata Terimlerinin Değişen Varyanslı Olma Sorunu

Hata terimleri binom dağılımına sahip olduğundan $Var(u_i)$ 'ler eşit olmayacak ve değişen varyans sorunu ortaya çıkacaktır. Tablo 3.1.'de hatalara ait olasılık değerleri verilmişti. Buradan varsayım gereği; $var(u_i)=E[u_i-E(u_i)]^2=E(u_i^2)$ ve $E(u_i)=0$ olduğundan ve ayrıca u_i 'nin Tablo 3.1.'deki olasılık dağılımı kullanılarak:

$$\begin{aligned} Var(u_i) &= E(u_i^2) = (-\beta_0 - \beta_1 X_i)^2 (1-P_i) + (1 - \beta_0 - \beta_1 X_i)^2 (P_i) \\ &= (-\beta_0 - \beta_1 X_i)^2 (1 - \beta_0 - \beta_1 X_i) + (1 - \beta_0 - \beta_1 X_i)^2 (\beta_0 + \beta_1 X_i) \\ &= (\beta_0 + \beta_1 X_i)(1 - \beta_0 - \beta_1 X_i) \end{aligned} \quad (3.5)$$

veya

$$Var(u_i) = E(Y_i | X_i) [1 - E(Y_i | X_i)] = P_i (1 - P_i) \quad (3.6)$$

elde edilir. Modelde u_i , Y_i 'nin şartlı beklenti değerine bağlı olup değişen varyansa sahiptir. Değişen varyans durumunda, EKKY'nden ele edilen tahmin ediciler hem sapmasız hem de en küçük varyanslı (etkin) olmayacaktır (Gujarati, 2001: 544; Long, 1997: 38). P_i 'nin 0 veya 1'e yakın değerleri için düşük varyanslar söz konusu olurken, P_i 'nin 0.5'e yakın değerlerinde varyans daha yüksek olarak elde edilecektir (Pindyck-Rubinfeld, 1985: 276).

Ancak yine de değişen varyans sorununu aşmak için Goldberger ve Zellner ile Lee, ağırlıklı en küçük kareler yöntemini önermişlerdir. Goldberger'in (1964) önerdiği iki aşamalı EKKY'ne göre önce ($\hat{Y}_i$) değerleri bulunur, ardından bulunan tahmini değerler 1 değerinden çıkarılarak kendisi ile çarpılır ($w_i = \hat{Y}_i(1 - \hat{Y}_i)$) ve elde edilen sonucun karekökü aşağıdaki (3.7) numaralı eşitlik yardımıyla bulunur:

$$\sqrt{E(Y_i | X_i) [1 - E(Y_i | X_i)]} = \sqrt{P_i (1 - P_i)} = \sqrt{w_i} \quad (3.7)$$

Böylece elde edilen bu ağırlık değeri, (3.1) nolu modelin her iki tarafı için, aşağıdaki (3.8) nolu eşitlikte gösterildiği gibi bölünerek değişen varyans sorunu çözülür:

$$\frac{Y_i}{\sqrt{w_i}} = \frac{\beta_0}{\sqrt{w_i}} + \beta_1 \frac{X_i}{\sqrt{w_i}} + \frac{u_i}{\sqrt{w_i}} \quad (3.8)$$

Böyle bir durumda hata terimi sabit varyansa sahiptir (Gujarati ve Porter, 2012: 545; Long, 1997: 39).

3.1.1.3. Doğrusal Olmama Sorunu

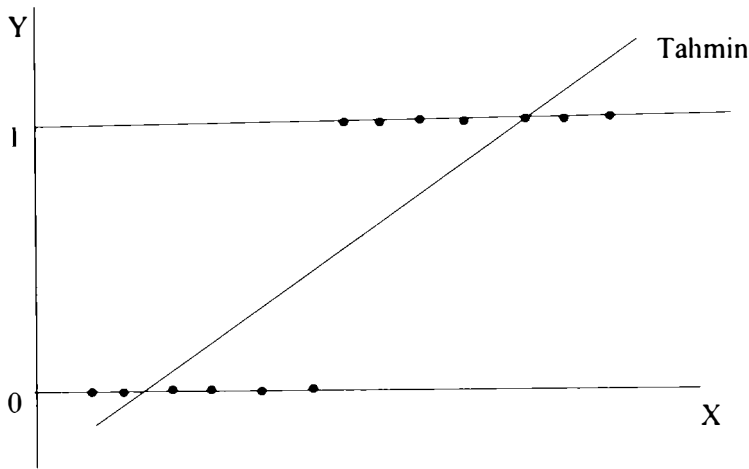
Doğrusal olasılık modellerinde, $E(Y_i | X_i)$ çoğu durumda 0 ile 1 sınırlarının dışında kalabilir. Bu aralık dışında elde edilen bir değeri yorumlamak anlamlı değildir. Bu sorunu aşmak için elde edilen $\hat{Y}_i$ değerlerinin 0 ile 1 arasında olduğuna bakılır. Eğer 0'dan küçük değerler var ise bunlara sıfır; 1'den büyük değerler var ise bunlara 1 değeri verilerek sorun aşılabılır. Ancak bu çözüm pek gerçekçi değildir, çünkü gerçekleşmeyen bir olaya 1 olasılığı, gerçekleşen bir olaya da 0 olasılığı karşılık gelebilir.

Sorunun önüne geçmek için bir diğer yol ise $\hat{Y}_i$ koşullu olasılıklarının 0 ile 1 arasında olmalarını sağlayan bir tahmin tekniği geliştirmektir. Geliştirilen Logit ve Probit modelleri tahmin edilen olasılıkların gerçekten 0 ile 1 mantıksal sınırları içinde kalmasını sağlar (Gujarati ve Porter, 2012: 545; Long, 1997: 39, Pindyck ve Rubinfeld, 1985: 277).

3.1.1.4. Belirlilik Katsayısının Küçük Olması Sorunu

Bağımlı değişken 0-1 şeklinde bir kukla değişken olarak tanımlanır ve bağımsız değişkenler ile model kurulur. Bağımlı değişken değerleri Şekil 3.1'de görüldüğü gibi 0 ile 1 arasında kalmalıdır. Bu sonuç bağımlı değişkenin öngörülen değerlerinin bireyin özellikleri verildiğinde olayın gerçekleşme olasılığını ifade etmektedir. Şekil 3.1'de noktalar örnek gözlemlerini göstermektedir ve bağımsız değişken X 'in yüksek değerleri, bağımlı kukla değişkenin değeri 1'e karşılık gelirken, X 'in küçük değerleri yine bağımlı kukla değişkenin 0 değerine karşılık gelmektedir. X 'in son derece küçük değerlerinde regresyon doğrusu olayın gerçekleşmesi durumu için negatif olasılık değeri verirken, benzer şekilde X 'in çok büyük değerleri için ise regresyon doğrusu 1'den büyük olasılık tahminine neden olacaktır. Bu yüzden belirlilik katsayısı değeri çok küçük olacak ve bir tahmin kriteri olarak kullanılmaması gerekliliği ortaya çıkacaktır (Kennedy, 2000:183; Gujarati, 2001: 546).

Şekil 3.1. Örnek Gözlemler



(Kaynak: Kennedy, 2000: 183)

3.1.2. DOM İçin Stata Uygulaması

DOM ile ilgili uygulamada TÜİK'in 1999 yılında yapmış olduğu iş gücü anketinin verileri kullanılmıştır. Veriler Türkiye genelinde hane bireylerinin cinsiyet, yaş, iş arayıp aramadığı, medeni durumu ile kırsal veya kentsel kesimdeki yerleşim yerleri göz önüne alınarak değerlendirilmiştir. Kurulan modelin amacı iş aramada hangi değişkenlerin etkili olduğunun belirlenmesidir. Bu nedenle bağımlı değişken kişinin iş arayıp aramadığı şeklinde alınmıştır. Veriler elde edilirken Türkiye'nin iş gücü potansiyeli dikkate alınarak modele sadece 12-65 yaş aralığındaki kişiler

dâhil edilmiştir. Gözlem sayısı 17.233 kişidir. Aşağıda çalışmada kullanılan değişkenler görülmektedir¹:

Tablo 3.2. Değişkenlerin Tanımları

Değişken	Değişken tanımları
isarama	Birey iş arıyorsa 1, iş aramıyorsa 0 değerini alan kukla değişken
cinsiyet	Birey erkek ise 1, değilse 0 değerini alan kukla değişken
medeni_durum	Birey evli ise 1, değilse 0 değerini alan kukla değişken
yerleşim_yeri	Birey kentte yaşıyorsa 1, kırdan yaşıyorsa 0 değerini alan kukla değişken
yas	Bireyin yaşını gösteren sürekli değişken

Stata paket programında değişkenlerin tanımlarını listelemek için aşağıda verilen “describe” komutu kullanılır.

describe

Contains data from C:\Users\Hamdi EMEÇ\Desktop\kitap\bolum_3.dta

obs: 17,233

vars: 5

22 Aug 2020 16:33

variable name	storage type	display format	value label	variable label
yas	float	%9.0g		
isarama	float	%9.0g		
cinsiyet	float	%9.0g		
medeni_durum	float	%9.0g		
yerlesim_yeri	float	%9.0g		

Stata’da tanımlayıcı istatistikler aşağıda verilen “sum” komutu kullanılarak elde edilmektedir.

Tablo 3.3. Değişkenlerin Tanımlayıcı İstatistikleri

sum

Variable	Obs	Mean	Std. Dev.	Min	Max
yas	17,233	29.6162	15.38353	12	65
isarama	17,233	.0796147	.2707036	0	1
cinsiyet	17,233	.3422503	.4744767	0	1
medeni_durum	17,233	.4992746	.500014	0	1
yerlesim_yeri	17,233	.7826264	.4124709	0	1

¹ bolum_3.dta veri setine Uygulamalarla Mikro Ekonometri kitabının linkinden ulaşabilirsiniz.

Tablo 3.3'e göre toplam gözlem sayısı 17233'tür ve tüm değişkenlere ait ortalama ve standart sapma değerleri de tabloda yer almaktadır. Bunun yanında tüm değişkenler için frekans, yüzde ve birikimli yüzdeler Stata'da "tab" komutu ile elde edilebilir.

Tablo 3.3'e göre Türkiye genelinde 12-65 yaş aralığında olan toplam 17233 birey ile çalışılmıştır. Bireylerin %7.96'sı iş aramakta, %92.04'ü ise iş aramamaktadır. Anket yapılan bireylerin yaş ortalaması 29.61 ve standart sapması 15.38 iken bu bireylerin %65.78'sinin kadın, %34.22'sinin ise erkek olduğu görülmektedir. Bireylerin %49.93'ü evli, %50.07'si ise bekârdır. Ayrıca kentte oturanların oranının %78.26, kırdaki oturanların oranının ise %21.74 olduğu görülmektedir.

Stata'da DOM için EKKY tahmin sonuçları aşağıda verilen "reg" komutu kullanılarak elde edilmektedir.

Tablo 3.4. DOM İçin EKKY Tahmin Sonuçları

reg isarama yas cinsiyet medeni_durum yerlesim_yeri

Source	SS	df	MS	Number of obs	=	17,233
-----+-----				F(4, 17228)	=	248.35
Model	68.8430977	4	17.2107744	Prob > F	=	0.0000
Residual	1193.92554	17,228	.069301459	R-squared	=	0.0545
-----+-----				Adj R-squared	=	0.0543
Total	1262.76864	17,232	.073280446	Root MSE	=	.26325

isarama	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
-----+-----						
yas	-.000866	.0001786	-4.85	0.000	-.001216	-.000516
cinsiyet	.1350886	.0043918	30.76	0.000	.1264802	.143697
medeni_durum	.024689	.005634	4.38	0.000	.0136457	.0357322
yerlesim_yeri	.0250721	.0048907	5.13	0.000	.0154858	.0346583
_cons	.0270809	.0058477	4.63	0.000	.0156189	.0385428

Tablo 3.4 model sonuçlarına göre değişkenlere ilişkin katsayıların hem tümünün (F istatistik değeri 248.35'dir) hem de ayrı ayrı (p değerleri $\alpha=0.05$ değerinden küçük olduğundan) her birinin %5 önem düzeyine göre anlamlı olduğu sonucuna varılır. Diğer tüm değişkenler sabit tutulduğunda bireylerin yaşı ilerledikçe iş arama olasılığı 0.00086 oranında azalmaktadır. Aynı şekilde diğer tüm değişkenler sabit tutulduğunda erkeklerin referans sınıf olan kadınlara göre iş arama olasılığı 0.135 kadar daha fazladır. Benzer biçimde diğer tüm değişkenler sabit tutulduğunda evli olan bireylerin referans sınıf olan evli olmayan bireylere göre iş arama olasılığı 0.024 daha fazladır. Son olarak yine diğer tüm değişkenler sabit tutulduğunda kentte yaşayan bireylerin, kırdaki yaşayan bireylere göre iş arama olasılığı 0.025 daha fazladır.

Modelde değişen varyans olup olmadığı Breusch-Pagan/Cook-Weisberg istatistiği ile hesaplanarak Tablo 3.5'te gösterilmiştir.

Tablo 3.5. DOM İçin Değişen Varyansın Breusch-Pagan / Cook-Weisberg ile Test Edilmesi

hettest

```

Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
Ho: Constant variance
Variables: fitted values of isarama
chi2(1)      = 4352.88
Prob > chi2  = 0.0000

```

Tablo 3.5'te, χ^2 değerinin 4352.88 bulunarak H_0 hipotezinin red edildiği ($\text{prob} > \chi^2 = 0.0000$) ve modelde değişen varyans sorunu olduğu görülmektedir. DOM modelinde hata teriminin değişen varyanslı olması durumunda tahminçiler tutarsız oldukları için STATA'da dirençli standart hatalar Tablo 3.4'te verilen komutun sonuna, virgülden sonra "robust" yazılarak hesaplanmaktadır. Stata uygulaması Tablo 3.6'da verilmiştir.

Tablo 3.6. DOM Modeli İçin Dirençli (Robust) Tahmini

reg isarama yas cinsiyet medeni_durum yerlesim_yeri, robust

Linear regression				Number of obs	=	17,233
				F(4, 17228)	=	174.15
				Prob > F	=	0.0000
				R-squared	=	0.0545
				Root MSE	=	.26325

		Robust				
isarama		Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
-----+-----						
yas		-.000866	.0001467	-5.91	0.000	-.0011535 - .0005786
cinsiyet		.1350886	.0056419	23.94	0.000	.1240298 .1461474
medeni_durum		.024689	.0062937	3.92	0.000	.0123527 .0370253
yerlesim_yeri		.0250721	.0046305	5.41	0.000	.0159958 .0341484
_cons		.0270809	.0049106	5.51	0.000	.0174555 .0367062

Tablo 3.4 ile Tablo 3.6'daki Stata çıktıları incelendiğinde standart hatalarda önemli farklılıklar gözlenmemekte ve katsayılar Tablo 3.4'teki gibi yorumlanmaktadır.

3.2. Logit Model

DOM'ndeki sorunların ortadan kaldırılması için kullanılan logit modeli, lojistik birikimli dağılımdan türetilen bir modeldir. Bazı çalışmalarda logit model yerine lojistik regresyon modeli şeklinde de ifade edilmektedir. Bağımlı değişken iki durumlu olduğunda genellikle logit modelinin tercih edilmesinin nedenlerinin başında matematiksel olarak Lojistik fonksiyonun kullanımının daha esnek ve kolay olması gelmektedir. Ayrıca açıklayıcı değişkenlerin daha anlamlı bir şekilde yorumlanabilmesi de logit modelinin tercih edilme sebebi olmaktadır.

Aşağıda iki durumlu bağımlı bir kukla değişkenin sürekli bağımsız değişken ile olan ilişkisini gösteren doğrusal regresyon modeli verilmiştir:

$$P_i = E(Y = 1 | X_i) = \beta_0 + \beta_1 X_i \quad (3.9)$$

Burada X_i sürekli bir değişken iken $Y=1$, gerçekleşmesi istenen olay olsun. Gerçekleşmesi istenen olayın olasılığı:

$$P_i = E(Y_i = 1 | X_i) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_i)}} \quad (3.10)$$

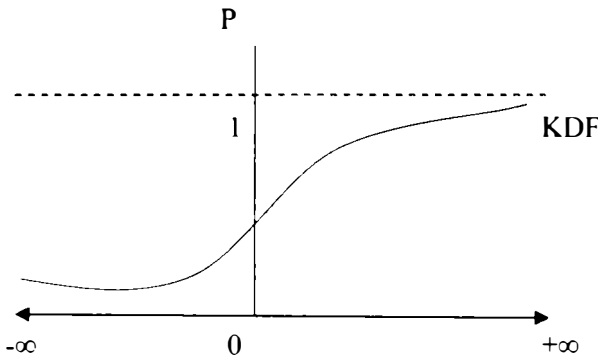
şeklinde dir. Burada $Z = \beta_0 + \beta_1 X_i$ dersek;

$$P_i = \frac{1}{1 + e^{-Z_i}} \quad (3.11)$$

olacaktır.

Genel olarak (3.11) eşitliğine “lojistik dağılım fonksiyonu” denir. Birikimli lojistik dağılım fonksiyonu Şekil 3.2’de gösterildiği biçimdedir. Fonksiyon alt sınırı “0” üst sınırı “1” olan “S” şeklinde bir eğriyi göstermektedir:

Şekil 3.2. Birikimli Lojistik Dağılım Fonksiyonu



(Kaynak: Gujarati, 2006: 553)

$Z_i \rightarrow -\infty$ 'dan $Z_i \rightarrow +\infty$ 'a kadar olan aralıkta değişirken P_i ise 0 ile 1 arasında değerler alır. P_i 'nin Z_i ya da X_i ile ilişkisi doğrusal olmadığı gibi β 'larla olan ilişkisi de doğrusal değildir ve bu nedenle anakütle katsayılarını tahmin ederken EKKY kullanılamaz. Bunun nedeni, bağımlı kukla değişkenli modeli tahmin edilirken EKKY varsayımlarının çoğunun ihlal edilmesindendir.

EKK sürecinin bu yapıda kullanılmamasına rağmen aslında aşağıda gösterildiği gibi özünde doğrusal olduğuna dikkat ediniz:

Bir olayın gerçekleşme olasılığı P_i ise, gerçekleşmeme olasılığı $(1-P_i)$ olacaktır:

$$(1-P_i) = \frac{1}{1 + e^{Z_i}} \quad (3.12)$$

Dolayısıyla:

$$\frac{P_i}{(1-P_i)} = \frac{1 + e^{Z_i}}{1 + e^{-Z_i}} = e^{Z_i} \quad (3.13)$$

Denklem (3.13)'ün doğal logaritması alındığında:

$$L_i = \ln\left(\frac{P_i}{(1-P_i)}\right) = Z_i = \beta_0 + \beta_1 X_i \quad (3.14)$$

elde edilir. Değişken katsayısı β_1 , X_i 'deki bir birimlik değişiminin Z_i 'deki değişmesini ölçer. Sabit kesme β_0 'ın bir anlamı yoktur.

Bu durumda $\frac{P_i}{(1-P_i)}$, bir olayın gerçekleşme olasılığının gerçekleşmeme olasılığına oranıdır, yani fark oranıdır (odds ratio). Fark oranının logaritması L , yalnız X 'e göre değil anakütle katsayılarına göre de doğrusaldır. L 'ye **logit** denir ve denklem (3.14)'teki gibi olan modellere de **Logit Modeli** adı verilmektedir. Görüldüğü üzere L_i , X 'e göre doğrusal olmakla birlikte, olasılıkların kendisi böyle değildir. Ayrıca olasılıklar 0 ile 1 arasında sınırlı kalırken, logitte bir sınırlama yoktur (Ramanathan, 1995: 279-280; Greene, 1993: 638; Maddala, 1992: 328; Gujarati ve Porter, 2012: 553-555; Gujarati, 2016: 233-236; Cox ve Snell, 1989: 21).

3.2.1. Logit Modelin En Çok Olabilirlik Yöntemiyle (EÇÖY) Tahmini

(3.14) modelini tahmin edebilmek için, X_i 'den başka logit L değerlerini de bilmek gerekmektedir. Burada, bazı güçlüklerle karşılaşılır. Çünkü istenen durumda $P_i = 1$ (örneğimizde istenen durum kişi iş arıyorsa şeklindedir), değilse $P_i = 0$ 'dır. Bu değerler doğrudan L 'de yerine koyulursa:

$$\begin{aligned} L_i &= \ln\left(\frac{1}{0}\right) && \text{istenen koşul sağlanırsa} \\ L_i &= \ln\left(\frac{0}{1}\right) && \text{istenen koşul sağlanmaz ise} \end{aligned} \quad (3.15)$$

şeklinde bulunur. Bu ifadeler anlamsız olduğundan anakütle katsayılarını tahmin edebilmek için en çok olabilirlik yöntemi (EÇÖY) kullanılır (Gujarati ve Porter, 2012: 556). Logit modelinde bağımsız değişkenler ve parametre tahminleri için olasılık değerleri lojistik dağılım kullanılarak elde edilir (Stock ve Watson, 2011: 402; Agresti, 1996 :84).

3.2.2. Logit Modeli İçin Stata Uygulaması

Logit Modelinin Stata'da yapılan uygulaması için de TÜİK'in 1999 yılında yapmış olduğu iş gücü anketinin verileri kullanılmıştır. 3.1.2. kısmında veri seti ve ilgili değişkenlere ait tanımlamalar ve tanımlayıcı istatistikler gösterilmişti. Bu kısımda Stata'da sırayla logit modeli için EÇÖY ile tahmini, fark oranları ile marjinal etkilerinin bulunması ve yorumlanması gösterilecektir.

3.2.2.1. Logit Modelinin EÇÖY ile Tahmini

Stata’da logit modelinin EÇÖY ile uygulaması için “logit” komutu kullanılır. Tablo 3.7’de Logit Modelinin EÇÖY ile Tahmini verilmiştir.

Tablo 3.7. Logit Modelinin EÇÖY İle Tahmini

logit isarama yas cinsiyet medeni_durum yerlesim_yeri

Iteration 0: log likelihood = -4787.798
Iteration 1: log likelihood = -4409.2004
Iteration 2: log likelihood = -4337.9606
Iteration 3: log likelihood = -4337.9023
Iteration 4: log likelihood = -4337.9023

Logistic regression

Number of obs = 17,233
LR chi2(4) = 899.79
Prob > chi2 = 0.0000
Pseudo R2 = 0.0940

Log likelihood = -4337.9023

isarama	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
yas	-.0168647	.0029707	-5.68	0.000	-.0226872	-.0110423
cinsiyet	1.822261	.0673845	27.04	0.000	1.69019	1.954332
medeni_durum	.5375729	.0990602	5.43	0.000	.3434184	.7317273
yerlesim_yeri	.3808126	.0753081	5.06	0.000	.2332114	.5284138
_cons	-3.450976	.094425	-36.55	0.000	-3.636046	-3.265907

Not: Eğer modellerde, model çıktısından önce iterasyonların görülmesi istenmiyorsa komuta aşağıdaki şekilde virgülden sonra “nolog” ifadesi yazılmalıdır:

logit isarama yas cinsiyet medeni_durum yerlesim_yeri, nolog

Olabilirlik oran (LR) testi (899.79) modelin bir bütün olarak anlamlı olduğunu göstermektedir (prob>chi2=0.0000). Her bir değişkenin katsayısının sıfırdan farklı olup olmadığına bakıldığında, katsayılar %5 anlamlılık düzeyinde istatistiki olarak anlamlıdır (p değerleri $\alpha = 0.05$ değerinden küçük olduğundan). Logit modelde katsayıların yorumlanması mümkün değildir, fakat katsayıların aldıkları işarete göre bazı yorumlamalar yapılabilir. Şöyle ki “iş arama” örneği için katsayıların işaretleri yorumlanmak istendiğinde; yaş arttıkça kişilerin daha az iş aradıkları, erkeklerin kadınlara göre, evli olanların evli olmayanlara göre ve kentte oturanların kırdaki oturanlara göre daha fazla iş aradıkları söylenebilir.

3.2.2.2. Logit Modelinde Fark Oranlarının (Odds Ratio) ve Marjinal Etkilerin Bulunması ve Yorumlanması

Logit modelde değişkenlerin katsayıları yorumlanmak istenirse, bu yorumlar ya fark oranlarından² ya da marjinal etkilerden yapılır. İki durumlu logit modelde, belirli bir değişkenin katsayısının işareti bağımlı değişkenin fark oranlarının logaritmalarını artırıyor veya azalıyor olduğunu göstermektedir. Fark oranlarının logaritmaları ile bağımlı değişken kategorilerinin olasılıkları aynı yönde hareket etmektedir (Borooah, 2002: 51). Logit modeli için fark oranları ve marjinal etkilerin Stata'da uygulaması aşağıdaki Tablo 3.8 ve Tablo 3.9'da gösterilmiştir.

² Fark oranı (odds ratio), bir durumun olma olasılığı ve olmama olasılığı arasındaki ilişkiyi göstermektedir. Aşağıda cinsiyet değişkeni katsayısının fark oranı değeri ilk olarak STATA'da işaretleme ile cinsiyet değişkeni arasındaki çapraz tablodan yararlanarak işlem yoluyla, sonrada logit modeliyle bulunmuştur:

tab cinsiyet isarama

cinsiyet	isarama		Total
	0	1	
0	10,936	399	11,335
1	4,925	973	5,898
Total	15,861	1,372	17,233

Elde edilen çapraz tablodan sırasıyla aşağıdaki olasılıklar elde edilir:

Kadınların iş aramama olasılığı: $\frac{10936}{11335} = 0.96479929$

Kadınların iş arama olasılığı: $\frac{399}{11335} = 0.03520071$

Erkeklerin iş aramama olasılığı: $\frac{4925}{5898} = 0.83502882$

Erkeklerin iş arama olasılığı: $\frac{973}{5898} = 0.16497118$

Elde edilen bu olasılık değerleri, birbirine oranlarsa :

Kadınların iş arama göreli olasılığı: $\frac{0.03520071}{0.96479929} = 0.03648501$

Erkeklerin iş arama göreli olasılığı: $\frac{0.16497118}{0.83502882} = 0.19756346$

bulunur. Son olarak fark oranı için iki olasılığın oranı: $\frac{0.19756346}{0.03648501} = 5.41492136$ değeri elde edilir. Böylece erkeklerin iş arama göreli olasılığı, kadınların iş arama göreli olasılığından yaklaşık 5.41 kat daha fazladır denir. İkinci olarak **logit isarama cinsiyet,or** komutu çalıştırıldığında yine, e^{β} değeri 5.4149 olarak bulunur. Modelin çalıştırılması okuyucuya bırakılmıştır.

Tablo 3.8. Logit Modelinde Fark Oranlarının (Odds Ratio) Elde Edilmesi**logit isarama yas cinsiyet medeni_durum yerlesim_yeri,or**

Logistic regression	Number of obs	=	17,233
	LR chi2(4)	=	899.79
	Prob > chi2	=	0.0000
Log likelihood = -4337.9023	Pseudo R2	=	0.0940

	isarama	Odds Ratio	Std. Err.	z	P> z	[95% Conf. Interval]
yas		.9832767	.002921	-5.68	0.000	.9775682 .9890185
cinsiyet		6.185829	.4168288	27.04	0.000	5.42051 7.059203
medeni_durum		1.711847	.1695759	5.43	0.000	1.409758 2.078668
yerlesim_yeri		1.463473	.1102114	5.06	0.000	1.262648 1.69624
_cons		.0317147	.0029947	-36.55	0.000	.0263564 .0381623

Note: _cons estimates baseline odds.

Not: Bu tablonun aynısı aşağıdaki komut ile de elde edilir:

logistic isarama yas i.cinsiyet i.medeni_durum i.yerlesim_yeri

Fakat bu komutu yazmadan önce menüden bazı işlemler yapmak gereklidir. Bu işlemler aşağıda “iş arama” örneği için sırası ile verilmiştir:

- Menüden Statistics→Binary outcomes→Logistic regressions
- Çıkan pencerede Model kısmı için “Dependent variable”dan “isarama” değişkeni modelin bağımlı değişkeni olarak seçilir.
- Bağımsız değişkenler için ise “Independent variables” kısmından yandaki üç nokta tıklanır.
- “Create varlist with factor or time-series variables” penceresinde tek tek bağımsız değişkenler tanıtılır. “Type of variable”da “Factor variable” işaretli kalır. “Spesification”da “Main effect” seçili iken tek tek değişkenler burada, “Add to varlist” seçilerek eklenir. Dikkat edilecek tek nokta sürekli değişkenler için “Variable” kısmının yanındaki sürekli kutucuğunu “c.” işaretlemek olacaktır.

Logit modelin EÇÖY ile tahminde parametrelerin işaretleri olasılıktaki değişimin yönünü göstermişti ve yorumlar buna göre gerçekleştirilmişti. Logit model için hesaplanan fark oranlarında ise olasılıktaki değişimin yönünü dikkate alarak, referans sınıfa göre bu değişimin büyüklüğü hakkında yorumlamalarda bulunulur. Fark oranı 1’den büyükse bağımsız değişkendeki artış bağımlı değişkeni artırır; fark oranı 1’den küçükse bağımsız değişkendeki artış bağımlı değişkeni azaltır. Fark oranları, Tablo 3.7’den elde edilen logit modelindeki bağımsız değişkenlerin katsayıları e^{β} şeklinde hesaplanarak elde edilir. Bir örnek ile anlatılırsa, yaş değişkeninin katsayısı -0.0168647’dir. Bu değer e^{β} ’da yerine koyulursa; $e^{-0.0168647} = 0.9832767$ değeri Tablo 3.8’deki yaş değişkenine karşılık gelen fark oranı olarak bulunur. Yüzde değişimden $(0.9832-1)*100 = -1.68$ değeri bulunur. Kişi bir yıl daha yaşlandığında iş arama olasılığı %1.68 azalır yorumu yapılır. Erkeklerin iş arama olasılığı kadınların iş arama olasılığından görece olarak yaklaşık 6.18 kat daha fazladır. Evlilerin iş arama olasılığı evli olmayanların iş arama olasılığından 1.71 kat

daha fazladır. Kentte oturanların iş arama olasılığı kırdaki oturanların iş arama olasılığından 1.46 kat daha fazladır.

Eğer tahminlenen olasılıklara ihtiyaç varsa Stata’da komut penceresine

predict p

yazılır. Bu komut sonrasında “(option pr assumed; Pr(isarama))” şeklinde bir uyarı ile “Data Editor” penceresinde “p” isimli yeni bir değişken eklenir ve tüm olasılıklar görülür.

Eğer elde edilen logit modelden hareketle marjinal etkiler veya esneklikler hesaplanmak istenirse logit modelinin tahmininin ardından “mfx” komutu yazılır ve Tablo 3.9’daki çıktı elde edilir.

Tablo 3.9. Logit Modelinde Marjinal Etkiler

mfx

Marginal effects after logit

y = Pr(isarama) (predict)
= .059505

variable	dy/dx	Std. Err.	z	P> z	[95% C.I.]	X
-----+-----						
yas	-.0009438	.00016	-5.75	0.000	-.001266 -.000622	29.6162
cinsiyet*	.1405963	.00571	24.61	0.000	.1294 .151792	.34225
medeni~m*	.0303354	.0056	5.41	0.000	.01935 .041321	.499275
yerles~i*	.019453	.0035	5.56	0.000	.012591 .026315	.782626
-----+-----						

(*) dy/dx is for discrete change of dummy variable from 0 to 1

Not: Bu tablonun benzeri aşağıdaki komut ile de elde edilir. Yukarıdaki çıktıdan farkı modeldeki bağımsız değişkenler için ortalama değerlerini vermesidir.

margins, dydx(*) atmeans

Logit modelde marjinal etkiler olasılık olarak yorumlanır. Bu nedenle Tablo 3.9’da elde edilen Stata çıktısındaki marjinal etkilerin yorumları şu şekilde olmaktadır: diğer değişkenler sabit tutulduğunda ortalama yaş bir yıl arttığında iş arama olasılığı yaklaşık %0.09 kadar azalır. Benzer şekilde erkeklerin referans sınıf olan kadınlara göre iş arama olasılığı %14 daha fazladır. Medeni durumda ise evli olanların evli olmayanlara göre iş arama olasılığı %3 daha fazla ve son olarak kentte oturanların kırdaki oturanlara göre iş arama olasılığı yaklaşık %2 daha fazladır.

Tablo 3.10 ve Tablo 3.11’de tüm kategorilerde bağımsız kukla ve nicel değişken yaş için olasılık tahminleri elde edilmiştir. Burada önce aşağıdaki komut Stata’da yazılır:

logistic isarama yas i.cinsiyet i.medeni_durum i.yerlesim_yeri

Komut sonrasında model elde edildikten sonra “margins” komutu ile bağımsız kukla değişkenler komut satırına yazılıp aşağıdaki Tablo 3.10 elde edilir:

Tablo 3.11. Bağımsız Sürekli Değişken İçin Olasılık Tahminleri**margins, at (yas=(12(10)65))**

```

Predictive margins                                Number of obs      =      17,233
Model VCE      : OIM
Expression     : Pr(isarama), predict()
1._at         : yas              =          12
2._at         : yas              =          22
3._at         : yas              =          32
4._at         : yas              =          42
5._at         : yas              =          52
6._at         : yas              =          62

```

		Delta-method				
		Margin	Std. Err.	z	P> z	[95% Conf. Interval]
_at						
1		.1028227	.0052019	19.77	0.000	.0926271 .1130182
2		.0892346	.0028797	30.99	0.000	.0835905 .0948787
3		.0771908	.0019973	38.65	0.000	.0732762 .0811055
4		.0665755	.0026745	24.89	0.000	.0613336 .0718174
5		.0572672	.0036311	15.77	0.000	.0501503 .0643841
6		.0491432	.0043983	11.17	0.000	.0405226 .0577637

Tablo 3.11'e göre; 12 yaşında bir kişinin iş arama olasılığı %10, 22 yaşında bir kişinin iş arama olasılığı %8.9, 32 yaşında bir kişinin iş arama olasılığı %7.7, 42 yaşında bir kişinin iş arama olasılığı %6.6, 52 yaşında bir kişinin iş arama olasılığı %5.7 ve 62 yaşında bir kişinin iş arama olasılığı %4.9'dur. Tüm olasılıklar %5 önem seviyesinde anlamlıdır.

3.3. Probit Model

Probit model, iki durumlu modellerde yaygın olarak kullanılan ve parametreleri doğrusal olmayan bir modeldir. Uygun seçilmiş bir birikimli dağılım fonksiyonu kullanarak iki kategorili bir bağımlı değişkenin davranışı açıklanabilir. Eğer bir Z değişkeni, μ ortalama, σ^2 varyansla normal dağılıyorsa olasılık yoğunluk fonksiyonu (OYF):

$$f(Z) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(Z-\mu)^2}{2\sigma^2}} \quad (3.16)$$

biçimindedir. KDF ise:

$$F(Z) = \int_{-\infty}^{Z_0} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(Z-\mu)^2}{2\sigma^2}} \quad (3.17)$$

şeklindeir. Burada Z_0 , Z 'nin belli bir değeridir. Normal birikimli dağılım fonksiyonu olan (3.17)'den türeyen eşitlik probit modelidir. Probit modeli, olasılığın 0 ile 1 arasında değiştiği, i. bireyin kararının gözlenemeyen bir fayda indeksine bağlı olduğunu varsayar. Z değerleri ile olasılıklar arasında doğrusal olmayan ilişki vardır ve olasılık değerleri 0 veya 1 gibi uç değerlere yaklaştıkça değerlerindeki bir birimlik değişim, fonksiyonda daha az bir değişime neden

olacaktır (Gujarati, 2001: 564; Liao, 21: 1994). (3.17) formülünün amacı, Probit eşitliğindeki doğrusal olarak tanımlanan Z 'yi doğrusal olmayan şekilde tanımlanan olasılık değerlerine geri dönüştürmektir (Pampel, 2000: 59).

Probit model fayda teorisi veya rasyonel tercih davranışı üzerine oturtulduğundan kişinin iş arama davranışı, i'inci bireyin gözlenemeyen bir endekse bağlı olarak iş arayıp aramamasını ifade eder. Açıklayıcı değişken X_i 'ye bağlı olarak belirlenen I_i fayda endeksi ne kadar büyük olursa, iş arama olasılığı da o kadar yüksek olur. Bu I_i endeksi aşağıdaki şekilde yazılır:

$$I_i = \beta_1 + \beta_2 X_i \quad (3.18)$$

Burada X_i , i. Bireyin yaşı olsun. Birey iş arıyorsa $Y_i = 1$, iş aramıyorsa $Y_i = 0$ 'dır. Aynı durum endeks için kullanılırsa, her bireyin bir kritik veya eşik endeks değerinin I_i^* ile belirlenebileceğini varsaymak uygun olacaktır. Yani, $I_i > I_i^*$ ise birey iş arıyor olacak, $I_i < I_i^*$ durumunda ise iş aramıyor olacaktır. Birey ya iş arıyordur, ya da aramıyordur. Doğrudan gözlenemeyen I_i^* eşik değeri, aynı ortalama ve varyans ile normal dağılımlı kabul edilirse (3.18) denklemindeki parametreler hesaplanabilir.

Normallik varsayımı altında, $I_i^* \leq I_i$ standart normal birikimli dağılım fonksiyonu hesaplanabilir:

$$P_i = \Pr(Y = 1) = \Pr(I_i^* \leq I_i) = F(I_i) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{I_i} e^{-t^2/2} dt = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\beta_0 + \beta_1 X_i} e^{-t^2/2} dt \quad (3.19)$$

Burada t , standartlaştırılmış normal değişkendir ($t \sim N(0,1)$) (Gujarati, 2001: 564).

Probit modelin tahmincilerinin yaklaşık 1.6 ile 1.8 katı, aynı modelin logit tahmincilerini verir (Long, 1997: 129). Logit katsayılar, bu sayılara bölünerek birimsel karşılaştırmalar yapılabilir. İki durumlu bağımlı değişkenli probit modellerin bazı varsayımları vardır. Bunlardan ilki bağımlı değişkenin, örneklemdeki tüm gözlemler için 0 veya 1 değerini almasıdır. Diğer bir varsayım örneklemdeki birimlerin birikimli normal dağılım fonksiyonu şeklinde ifade edilmesidir. Bunun yanında örneklemdeki her bir birime ait bağımlı değişkenin değeri istatistiksel olarak bağımsız olmalıdır ve tüm bağımsız değişkenler arasında tam veya yaklaşık doğrusal bağımlılık olmalıdır (Aldrich ve Nelson, 1984: 49).

3.3.1. Probit Modelin En Çok Olabilirlik Yöntemiyle (EÇÖY) Tahmini

Probit modelinin parametre değerlerinin tahmininde logit modelde olduğu gibi EÇÖY kullanılmaktadır. EÇÖY'nde probit olabilirlik değerini maksimum yapacak parametre değerlerini tahmin etmek amaçlanmaktadır. Probit modelinin logit modelinden farkı ise bağımsız değişkenler ve parametre tahminleri için olasılık değerlerini lojistik dağılım yerine birikimli standart normal dağılımı kullanarak elde edilmesidir (Agresti, 1996: 84).

3.3.2. Probit Modeli İçin Stata Uygulaması

Probit Modelinin Stata’da yapılan uygulaması için de TÜİK’in 1999 yılında yapmış olduğu iş gücü anketinin “bolum_3.dta” verileri kullanılmıştır. 3.1.2. kısmında veri seti ve ilgili değişkenlere ait tanımlamalar ve tanımlayıcı istatistikler gösterilmişti. Bu kısımda Stata’da sırayla probit modeli için EÇÖY ile standart hatalarının dirençli tahmini ve marjinal etkilerinin bulunması ve yorumlanması gösterilecektir.

3.3.2.1. Probit Modelinin EÇÖY ve Standart Hatalarının Dirençli Tahmini

Stata’da logit modelinin EÇÖY ile uygulaması için “probit” komutu kullanılır. Tablo 3.12’de Probit modelinin en çok olabilirlik yöntemi ile tahmini verilmiştir.

Tablo 3.12. Probit Modelinin EÇÖY İle Tahmini

probit isarama yas cinsiyet medeni_durum yerlesim_yeri

Probit regression				Number of obs	=	17,233
				LR chi2(4)	=	883.95
				Prob > chi2	=	0.0000
Log likelihood = -4345.8208				Pseudo R2	=	0.0923

isarama		Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
-----+-----						
yas		-.0064633	.0014436	-4.48	0.000	-.0092926 -.0036339
cinsiyet		.8640123	.0308723	27.99	0.000	.8035038 .9245208
medeni_durum		.137459	.0444453	3.09	0.002	.0503477 .2245703
yerlesim_yeri		.194079	.0375574	5.17	0.000	.1204678 .2676902
_cons		-1.855228	.0462635	-40.10	0.000	-1.945903 -1.764554

Olabilirlik oran testi (LR) (883.95) modelin bir bütün olarak anlamlı olduğunu göstermektedir (prob>chi2=0.0000). Her bir değişkenin katsayısının sıfırdan farklı olup olmadığına bakıldığında, katsayılar %5 anlamlılık düzeyinde istatistiki olarak anlamlıdır (p değerleri $\alpha = 0.05$ değerinden küçük olduğundan). Probit modelinde de logit de olduğu gibi katsayıların doğrudan yorumlanması mümkün değildir, fakat katsayıların aldıkları işarete göre bazı yorumlamalar yapılabilir. Şöyle ki “iş arama” örneği için katsayıların işaretleri yorumlanmak istendiğinde; yaş arttıkça kişilerin daha az iş aradıkları, erkeklerin kadınlara göre, evli olanların evli olmayanlara göre ve kentte oturanların kırdaki oturanlara göre daha fazla iş aradıkları söylenir.

Probit modelde hata terimi değişen varyanslıysa tahminciler tutarsız olduğundan Stata’da dirençli standart hatalar hesaplanmaktadır. Stata uygulaması Tablo 3.13’te verilmiştir. Tablo 3.12 ile Tablo 3.13’teki Stata çıktıları incelendiğinde standart hatalarda önemli farklılıklar gözlenmemektedir. Ancak bağımsız değişkenlerle hata teriminin önemli ölçüde ilişkili olması durumunda standart hatalar değişecektir.

durumda ise evli olanların evli olmayanlara göre iş arama olasılığı %1.7 daha fazla ve son olarak kentte oturanların kırdı oturanlara göre iş arama olasılığı yaklaşık %2 daha fazladır.

Tablo 3.15 ve Tablo 3.16’da tüm kategorilerde bağımsız kukla ve sürekli değişken olan yaş için tahmini olasılıklar elde edilmiştir. Burada daha önce logit modelde yapılan işlem Stata menüsünden probit model için de tekrar edilir. Bu işlemler aşağıda “iş arama” örneği için sırası ile verilmiştir:

- Menüden Statistics→Binary outcomes→Probit regressions
- Çıkan pencerede Model kısmı için “Dependent variable”dan “isarama” değişkeni modelin bağımlı değişkeni olarak seçilir.
- Bağımsız değişkenler için ise “Independent variables” kısmından yandaki üç nokta tıklanır.
- “Create varlist with factor or time-series variables” penceresinde tek tek bağımsız değişkenler tanıtılır. “Type of variable”da “Factor variable” işaretli kalır. “Specification”da “Main effect” seçili iken tek tek değişkenler burada “Add to varlist” seçilerek eklenir. Dikkat edilecek tek nokta sürekli değişkenler için “Variable” kısmının yanındaki sürekli kutucuğunu “c.” işaretlemek olacaktır.

İşlemler gerçekleştirildikten sonra çıkan pencerede OK tuşuna basıldığında aşağıdaki komut ve model çıktısı elde edilir:

probit isarama yas i.cinsiyet i.medeni_durum i.yerlesim_yeri

Model bulunduktan sonra “margins” komutu ile bağımsız kukla değişkenler komut satırına yazılıp aşağıdaki Tablo 3.15 elde edilir.

Tablo 3.15. Bağımsız Kukla Değişkenler İçin Olasılık Tahminleri

margins cinsiyet medeni_durum yerlesim_yeri,atmeans

Adjusted predictions			Number of obs		=	17,233	
Model VCE	: OIM						
Expression	: Pr(isarama), predict()						
at	: yas	=	29.6162 (mean)				
	0.cinsiyet	=	.6577497 (mean)				
	1.cinsiyet	=	.3422503 (mean)				
	0.medeni_d~m	=	.5007254 (mean)				
	1.medeni_d~m	=	.4992746 (mean)				
	0.yerlesim~i	=	.2173736 (mean)				
	1.yerlesim~i	=	.7826264 (mean)				

			Delta-method				
			Margin	Std. Err.	z	P> z	[95% Conf. Interval]

cinsiyet							
	0		.0339158	.0017084	19.85	0.000	.0305674 .0372641
	1		.1679967	.0050788	33.08	0.000	.1580423 .177951

medeni_durum							
0		.0549052	.0031008	17.71	0.000	.0488278	.0609827
1		.0719273	.0037207	19.33	0.000	.0646348	.0792197
yerlesim_yeri							
0		.0462546	.0033763	13.70	0.000	.0396372	.0528721
1		.0683454	.0023088	29.60	0.000	.0638203	.0728705

Buna göre Tablo 3.15'te cinsiyete bakıldığında 0 kadına ve 1 erkeğe karşılık geldiğinden; kadınların iş arama olasılığı %3, erkeklerin iş arama olasılığı %16.7'dir. Medeni durumda 0 evli olmayanlara ve 1 evli olanlara karşılık geldiğinden; evli olmayanların iş arama olasılığı %5.4, evli olanların ise %7'dir. Yerleşim yeri değişkeninde ise 0 kırdaki oturanlara ve 1 kentte oturanlara karşılık geldiğinden; kırdaki oturanların iş arama olasılığı %4.6, kentte oturanların iş arama olasılığı ise %6.8'dir. Tüm olasılıklar %5 önem seviyesinde istatistiksel olarak anlamlıdır.

Aşağıda Tablo 3.16'da yaş değişkeni için iş arama olasılıkları (tahmini olasılıklar) bulunmuştur. Araştırmacı, araştırmanın yapısına göre belli aralıklardaki yaşlara göre istediği olasılıkları hesaplayabilir. "iş arama" örneğinde "12-65" yaş arası olduğundan, bu yaş aralığı için 10'ar arttırarak olasılıklar hesaplanmıştır. Stata'da marjinal etkiler için "margins" komutunun sonuna sürekli değişken olan "yas"ın yanına önce en düşük yaş değeri olan 12, ardından 10'ar arttırıldığı için 10 değeri ve son olarak en yüksek yaş değeri 65 yazılır. Tablo 3.16'da bağımsız sürekli değişkenler için marjinal etkiler tablosunda Stata'da elde edilen komut ve çıktısı görülmektedir.

Tablo 3.16. Bağımsız Sürekli Değişken İçin Olasılıkların Tahmini

margins, at (yas=(12(10)65))

Predictive margins				Number of obs	=	17,233
Model VCE	:	OIM				
Expression	:	Pr(isarama), predict()				
1._at	:	yas	=	12		
2._at	:	yas	=	22		
3._at	:	yas	=	32		
4._at	:	yas	=	42		
5._at	:	yas	=	52		
6._at	:	yas	=	62		
		Delta-method				
		Margin	Std. Err.	z	P> z	[95% Conf. Interval]
_at						
1		.095782	.0045279	21.15	0.000	.0869075 .1046566
2		.0861027	.002626	32.79	0.000	.0809559 .0912496
3		.0771725	.0020339	37.94	0.000	.0731862 .0811589
4		.068962	.0028606	24.11	0.000	.0633553 .0745688
5		.0614395	.0039889	15.40	0.000	.0536214 .0692576
6		.0545713	.0050013	10.91	0.000	.0447689 .0643736

Tablo 3.16 yorumlanacak olursa; 12 yaşında bir kişinin iş arama olasılığı %9.5, 22 yaşında bir kişinin iş arama olasılığı %8.6, 32 yaşında bir kişinin iş arama olasılığı %7.7, 42 yaşında bir kişinin iş arama olasılığı %6.8, 52 yaşında bir kişinin iş arama olasılığı %6.1 ve 62 yaşında bir kişinin iş arama olasılığı %5.4'tür. Tüm olasılıklar %5 önem seviyesinde istatistiksel olarak anlamlıdır.

3.4. Logit ve Probit Modeller İçin İyi Uyum İstatistikleri ve Model Seçim Kriterleri

Bağımlı kukla değişkenli modellerde uygunluğun birçok sayısal ölçüsü doğrusal regresyon modelindeki belirlilik katsayısından yapılandırılarak elde edilir. Aldrich ve Nelson (1984) geliştirdikleri iyi uyum ölçüsü olan *Pseudo R²* değerini model için hesaplanan χ^2 istatistik değerini kullanarak aşağıdaki gibi hesaplamışlardır (Aldrich ve Nelson, 1984: 57):

$$R^2 = \frac{\chi^2}{\chi^2 + n} \quad (3.20)$$

Bir başka uyum iyiliği ölçüsü ise hesaplanan hatalara dayanan özet istatistik olan *Pearson χ^2* istatistiğidir ve formülasyonu aşağıda gösterilmiştir (Long, 1997: 98):

$$Pearson \chi^2 = \sum_{i=1}^n e_i^2 \quad (3.21)$$

Buradaki hatalar, iki durumlu bağımlı değişkenli modelden elde edilen hatalardır ve bu hataların büyük değerleri eldeki gözlemlerin modele uygun olmadığını ifade eder.

Diğer bir uyum iyiliği testi olan Hosmer-Lemeshow (2000) tahmin edilen olasılık değerlerini gruplandırarak test etmeyi önermişlerdir. Burada iki farklı grupta stratejisi kullanılır: Bunlardan ilki tahmini olasılık değerlerinin yüzdeliklerine dayalı olarak tablollaştırılması ve ikincisi tahmini olasılık değerlerinin kesin değerlerine uygun olarak tablollaştırılmasıdır. (Hosmer ve Lemeshow, 2000: 156)³.

Stata'da model seçim kriterlerinde ise hem logit hem de probit model için "fitstat" komutuyla elde edilen kriterler ve istenen özellikler yanlarında verilerek aşağıda gösterilmiştir:⁴

Log-Lik Intercept Only (Logartitmik Olabilirlik Değeri): İyi bir modelde büyük olması gerekir.

Log-Lik FullModel: (Logartitmik Olabilirlik Değeri): İyi bir modelde büyük olması gerekir.

D(1707) (Deviance istatistiği): İyi bir modelde büyük olması gerekir.

LR (Olabilirlik Oran İstatistiği): İyi bir modelde büyük olması ve anlamlı olması ((Prob>LR)<0.05) gerekir.

McFadden's R²: İyi bir modelde büyük olması gerekir.

McFadden's Adj R²: İyi bir modelde büyük olması gerekir.

³ Ayrıntılı bilgi için bkz. <https://www.stata.com/manuals/restatgof.pdf>

⁴ Ayrıntılı bilgi için bkz. <https://stats.idre.ucla.edu/stata/webbooks/logistic/chapter3/lesson-3-logistic-regression-diagnostics/>

Cragg & Uhler's R^2 : İyi bir modelde büyük olması gerekir.

McKelvey and Zavoina's R^2 : İyi bir modelde büyük olması gerekir.

Efron's Variance of y^* : İyi bir modelde küçük olması gerekir.

Variance of error: İyi bir modelde küçük olması gerekir.

Count R^2 : İyi bir modelde büyük olması gerekir.

Adj Count R^2 : İyi bir modelde büyük olması gerekir.

AIC (akaike bilgi kriteri): İyi bir modelde küçük olması gerekir.

AIC*n: İyi bir modelde küçük olması gerekir.

BIC (schwars/bayesian bilgi kriteri): İyi bir modelde küçük olması gerekir.

BIC': İyi bir modelde küçük olması gerekir.

Modelin uygunluğu için bulunan değerler, modelin geçerli olup olmadığı ile ilgili olarak bilgi verse de bu değerler ancak araştırmacıya model hakkında yüzeysel bir fikir verir.

Aşağıda önce logit ve sonrasında ise probit model için sırayla χ^2 ve Hosmer-Lemeshow Uyum İyiliği İstatistikleri, model seçim kriterleri ve tahmin başarı tabloları verilmiş ve gerekli yorumlar yapılmıştır.

Tablo 3.17. Logit Model İçin χ^2 ve Hosmer-Lemeshow Uyum İyiliği İstatistikleri

**quietly logit isarama yas cinsiyet medeni_durum yerlesim_yeri
estat gof**

Logistic model for isarama, goodness-of-fit test

number of observations =	17233
number of covariate patterns =	389
Pearson chi2(384) =	3375.60
Prob > chi2 =	0.0000

estat gof, group(10)

Logistic model for isarama, goodness-of-fit test

(Table collapsed on quantiles of estimated probabilities)

number of observations =	17233
number of groups =	10
Hosmer-Lemeshow chi2(8) =	238.20
Prob > chi2 =	0.0000

Stata'da önce logit model elde edilmektedir. Fakat model çıktısı elde edilmek istenmiyorsa komutun önüne "quietly" ifadesi yazılır. "estat gof" komutu ile *Pearson* χ^2 , "estat gof,group(10)" komutu ile Hosmer-Lemeshow istatistiği elde edilir. Her iki istatistikte de önem düzeyleri ($\text{Prob} > \chi^2 = 0.0000$) %5'ten küçük olduğundan modelde tanımlama hatası olduğu söylenir.

Tablo 3.18. Logit Model İçin Model Seçim Kriterleri

fitstat

Measures of Fit for logit of isarama

Log-Lik Intercept Only:	-4787.798	Log-Lik Full Model:	-4337.902
D(17228):	8675.805	LR(4):	899.791
		Prob > LR:	0.000
McFadden's R2:	0.094	McFadden's Adj R2:	0.093
Maximum Likelihood R2:	0.051	Cragg & Uhler's R2:	0.119
McKelvey and Zavoina's R2:	0.180	Efron's R2:	0.065
Variance of y*:	4.014	Variance of error:	3.290
Count R2:	0.920	Adj Count R2:	0.000
AIC:	0.504	AIC*n:	8685.805
BIC:	-159376.124	BIC':	-860.773

Stata'da logit model için seçim kriterli model elde edildikten sonra "fitstat" komutu bulunur. Logit model için seçim kriterleri, probit model de elde edildikten sonra yorumlanacaktır. Logit modelde tahmin başarı tablosu ise Stata'da komut satırına "estat classification" yazılarak elde edilir. Fakat önemle tekrar etmek gerekecek olursa, hem model seçim kriterleri hem de tahmin başarı tablosunun bulunabilmesi için önce mutlaka ilgili logit model bulunmalıdır.

Tablo 3.19. Logit Model İçin Tahmin Başarı Tablosu

estat classification

Logistic model for isarama

----- True -----			
Classified	D	~D	Total
-----+-----			
+	0	0	0
-	1372	15861	17233
-----+-----			
Total	1372	15861	17233

Classified + if predicted $\text{Pr}(D) \geq .5$

True D defined as isarama != 0

Sensitivity	Pr(+ D)	0.00%
Specificity	Pr(- ~D)	100.00%
Positive predictive value	Pr(D +)	.%
Negative predictive value	Pr(~D -)	92.04%

False + rate for true ~D	Pr(+ ~D)	0.00%
False - rate for true D	Pr(- D)	100.00%
False + rate for classified +	Pr(~D +)	.%
False - rate for classified -	Pr(D -)	7.96%

Correctly classified		92.04%

Logit model için tahmin başarı tablosuna bakıldığında logit modelin doğru tahmin etme başarısı %92.04'dür. Veri setinde "iş arama" değişkeninin 0 ve 1 kategorilerinin frekans ve yüzdesine bakıldığında 0 değerlerinin, 1 değerlerine göre çok daha fazla olduğu görülmektedir. Modelin 0 değerlerini tahmin etme başarısı iyi iken, 1 değerlerini tahmin etme başarısı çok zayıftır.

Tablo 3.20. Probit Model İçin χ^2 ve Hosmer-Lemeshow Uyum İyiliği İstatistikleri

**quietly probit isarama yas cinsiyet medeni_durum yerlesim_yeri
estat gof**

```
Probit model for isarama, goodness-of-fit test
      number of observations =      17233
      number of covariate patterns =      389
      Pearson chi2(384) =      3311.14
      Prob > chi2 =      0.0000
```

estat gof, group(10)

```
Probit model for isarama, goodness-of-fit test
      (Table collapsed on quantiles of estimated probabilities)
      number of observations =      17233
      number of groups =      10
      Hosmer-Lemeshow chi2(8) =      248.50
      Prob > chi2 =      0.0000
```

Stata'da önce probit model elde edilmektedir. Fakat model çıktının görünmesi istenmiyorsa komutun önüne "quietly" ifadesi yazılır. "estat gof" komutu ile $Pearson \chi^2$, "estat gof,group(10)" komutu ile Hosmer-Lemeshow istatistiği elde edilir. Her iki istatistikte de önem düzeyleri ($Prob > \chi^2 = 0.0000$) %5'ten küçük olduğundan modelde tanımlama hatası olduğu söylenir. Stata'da probit model için seçim kriterleri de aynı logit modelde olduğu gibi "fitstat" komutuyla bulunur.

Tablo 3.21. Probit Model İçin Model Seçim Kriterleri

fitstat

```
Measures of Fit for probit of isarama
Log-Lik Intercept Only:  -4787.798      Log-Lik Full Model:      -4345.821
D(17228):                8691.642      LR(4):                  883.954
                          Prob > LR:      0.000
McFadden's R2:           0.092          McFadden's Adj R2:      0.091
Maximum Likelihood R2:    0.050          Cragg & Uhler's R2:     0.117
```

McKelvey and Zavoina's R2:	0.144	Efron's R2:	0.058
Variance of y*:	1.168	Variance of error:	1.000
Count R2:	0.920	Adj Count R2:	0.000
AIC:	0.505	AIC*n:	8701.642
BIC:	-159360.287	BIC':	-844.936

Probit modelde tahmin başarı tablosu Stata'da komut satırına “estat classification” yazılarak elde edilir. Fakat önemle tekrar etmek gerekecek olursa, hem model seçim kriterleri hem de tahmin başarı tablosunun bulunabilmesi için önce probit model elde edilmelidir. Probit model için tahmin başarısı aşağıda Tablo 3.22’de gösterilmiştir:

Tablo 3.22. Probit Model İçin Tahmin Başarı Tablosu

estat classification

Probit model for isarama			
----- True -----			
Classified	D	~D	Total
-----+-----	-----	-----	-----
+	0	0	0
-	1372	15861	17233
-----+-----	-----	-----	-----
Total	1372	15861	17233
Classified + if predicted Pr(D) >= .5			
True D defined as isarama != 0			

Sensitivity	Pr(+ D)	0.00%	
Specificity	Pr(- ~D)	100.00%	
Positive predictive value	Pr(D +)	.%	
Negative predictive value	Pr(~D -)	92.04%	

False + rate for true ~D	Pr(+ ~D)	0.00%	
False - rate for true D	Pr(- D)	100.00%	
False + rate for classified +	Pr(~D +)	.%	
False - rate for classified -	Pr(D -)	7.96%	

Correctly classified		92.04%	

Probit model için tahmin başarı tablosuna bakıldığında probit modelini doğru tahmin etme başarısı %92.04’tür. Tablo 3.22’de “iş arama” değişkeninin 0 ve 1 kategorilerinin frekans ve yüzdesine bakıldığında 0 değerlerinin, 1 değerlerine göre çok daha fazla olduğu görülmektedir. Modelin 0 değerlerindeki tahmin etme başarısı iyi iken, 1 değerlerini tahmin etme başarısı çok zayıftır. Tablo 3.21, Tablo 3.18 ile karşılaştırıldığında farklı kriterlerin, farklı modeli tercih ettiği görülmektedir. Bu durumda en sağlıklı olan seçim AIC ve BIC değerlerinin her ikisine de bakarak karar vermektir. İki model karşılaştırıldığında logit modelden elde edilen AIC ve BIC değerlerin probit modelden elde edilen değerlerden daha küçük olduğu görülür ve logit modelinin verilere daha iyi uyum sağladığı söylenebilir. O nedenle seçilen logit model için yapılan katsayı yorum-

lamaları geçerli olacaktır. Fakat tüm bu kriterler dikkate alınmasına rağmen önemli olan, tahmincilerin beklenen işaretlere sahip olması, istatistiki anlamlılıkları ve yorumlanmalarıdır.

Bölümde Kullanılan Stata Komutlarının Toplu Gösterimi

```
use "C:\Users\.....\bolum_3.dta"
describe
sum
reg isarama yas cinsiyet medeni_durum yerlesim_yeri
hettest
reg isarama yas cinsiyet medeni_durum yerlesim_yeri, robust
logit isarama yas cinsiyet medeni_durum yerlesim_yeri
logit isarama yas cinsiyet medeni_durum yerlesim_yeri, nolog
logit isarama yas cinsiyet medeni_durum yerlesim_yeri, or
logistic isarama yas i.cinsiyet i.medeni_durum i.yerlesim_yeri
predict p
mfx
margins, dydx(*) atmeans
margins cinsiyet medeni_durum yerlesim_yeri, atmeans
margins, at (yas=(12(10)65))
probit isarama yas cinsiyet medeni_durum yerlesim_yeri
probit isarama yas cinsiyet medeni_durum yerlesim_yeri, robust nolog
quietly logit isarama yas cinsiyet medeni_durum yerlesim_yeri
estat gof
estat gof, group(10)
fitstat
estat classification
quietly probit isarama yas cinsiyet medeni_durum yerlesim_yeri
```

KAYNAKÇA

- AMEMİYA, T. (1981). Qualitative Response Models: A Survey, *Journal of Econometric Literature*, 19, 1483-1536.
- AMEMİYA, T. (1985). *Advanced Econometrics*, Cambridge, MA; Harvard University Press.
- AGRESTI, Alan (1996). *An Introduction to Categorical Data Analysis*, John Wiley & Sons.
- ALDRICH, John H., NELSON (1984). Forrest D. *Linear Probability Logit and Probit Models*, USA, Sage Publication
- BOROOAH, Vani K. (2002). *Logit and Probit Ordered and Multinomial Models*, Sage Publications
- COX, D. R., SNELL (1989). E. J. *Analysis of Binary Data*, Second Edition, Chapman & Hall.
- GOLDBERGER, A.S. (1964). *Econometric Theory*, New York: John Wiley.
- GREENE W. H. (1993). *Econometric Analysis*. 2. Edition, McMillian Pub. Comp., New York.
- GUJARATI, Damodar N. (2001). *Temel Ekonometri*, (Çev: Ümit Şenesen, Gülay Günlük Şenesen), Literatür Yayıncılık, 2. Baskı, İstanbul.
- GUJARATI, D.N. (2006). *Essentials of Econometrics*. 3rd Edition, McGraw-Hill.

- GUJARATI, Damodar N., Porter D. C. (2012). Temel Ekonometri, 5. Basımdan çeviri (Çev: Ümit Şenesen, Gülay Günlük Şenesen), Literatür Yayıncılık, İstanbul
- GUJARATI, Damodar N. (2016). Örneklerle Ekonometri, BB101 Yayınları, Ankara
- HOSMER, David W., LEMESHOW, Stanley (2000). Applied Logistic Regression, 2nd Edition, USA, John Wiley & Sons
- KMENTA, J. (1971). Elements of Econometrics. New York: Mac Millan.
- KENNEDY, Peter (2000). Ekonometri El Kitabı, Gazi Kitabevi, Ankara.
- LIAO, Tim Futing (1994). Interpreting Probability Models Logit Probit and Other Generalized Linear Models, Sage University Paper
- LONG, J. Scott (1997). Regression Models for Categorical and Limited Dependent Variables, USA, Sage Publication.
- MADDALA, G.S.(1992). Econometrics, Macmillian Publishing Company, New York.
- MALIRIVAUD E. (1970). Statistical Methods of Econometrics, North-Holland, Amsterdam.
- PAMPEL, Fred C. (2000). Logistic Regression A Primer, USA, Sage Publication.
- PINDYCK R. S, Rubinfeld D. L (1985). Econometric Models and Economic Forecasts, McGraw Hill Inc.
- RAMANATHAN, Ramu (1995). Introductory Econometrics, Third Edition, The Dryden Press.
- STOCK, J. H., & WATSON, M. W. (2010). Introduction to Econometrics. Addison- Wesley.
- WOOLDRIDGE, J. M. (2013). Introductory Econometrics: A Modern Approach, Fifth Edition. USA.

BÖLÜM 4

ÇOK DURUMLU MODELLER VE STATA UYGULAMALARI

Ali Kemal Çelik¹

Erkan Oktay²

Karar vericinin davranışını belirlemede tercih modellerinden yararlanılmaktadır. Bireyin iş gücünde olup olmadığı iki durumlu bir değişken iken; bireyin inşaat, hizmet, maden, imalat ve tarım sektörlerinden hangisinde çalıştığı, yine bireyin A, B, C, D veya E partilerinden hangisine oy vereceği; iş gücünde olan bir bireyin işine giderken tren, otobüs, özel araç, feribot ya da bisiklet gibi araçlardan hangi ulaşım aracını kullandığı sıralı olmayan kategorik değişkene örnek gösterilebilir ve araştırmacı bu tercihlere etki eden değişkenlerin neler olduğunu tahmin etmek istediğinde çok durumlu logit ve probit modellerinden yararlanmalıdır. Bireyler, sonuçta yukarıdaki örneklerden de anlaşılacağı üzere alternatifler kümesinden birini seçmek zorundadır.

Çok durumlu kesikli tercih modelleri, karar vericilerin belirli alternatifler arasından yaptıkları tercihleri tarif etmektedir. Söz konusu karar vericiler; insanlar, hane halkları, firmalar ya da herhangi bir karar verme birimi olabileceği gibi alternatifler ise hakkında karar verilmek zorunda olan herhangi bir seçenek ya da öge olabilmektedir. Kesikli tercih çatısında herhangi bir model kurabilmek için tercih kümesi olarak adlandırılan alternatifler kümesinin üç farklı niteliğe sahip olması gerekmektedir. Birinci olarak karar vericinin bakış açısına göre alternatiflerin birbirlerinden bağımsız olması gerekmektedir. Diğer bir ifade ile bir alternatifin tercih edilmesi, diğer alternatiflerden herhangi birinin tercih edilmesine ihtiyaç duyulmadığı anlamına gelmektedir. Karar verici, tercih kümesinden yalnızca bir alternatifi tercih etmektedir. İkinci olarak tercih kümesinin tüm olası alternatifleri kapsayacak şekilde yeterince büyük olması gerekmektedir. Üçüncü olarak ise alternatif sayısının sonlu bir değer olması gerekmektedir. Karar verici, alternatifleri sayabilmeli ve neticede alternatifleri saymayı bitirebilmelidir. Söz konusu üç nitelik koşulu arasından birinci ve ikinci koşulun sınırlayıcı olmadığı kabul edilmektedir (Train, 2003).

Çok durumlu modeller, bağımlı değişkene ait ikiden fazla alternatifin olması durumunda kullanılmakta ve bağımlı değişken kategorilerinin özelliğine göre sıralı ve sıralı olmayan modeller olarak ayrılmaktadır. Sıralı modeller, tanım gereği bağımlı değişkene ait kategorilerin doğal bir sıralamaya sahip olduğu modelleri ve sıralı olmayan modeller ise bağımlı değişkene ait kategorilerin doğal bir sıralamaya sahip olmadığı modelleri tanımlamaktadır. Çok durumlu modellerde

¹ Doç. Dr. Ardahan Üniversitesi, İktisadi ve İdari Bilimler Fakültesi, İşletme Bölümü.
E-mail: alikemalcelik@ardahan.edu.tr

² Prof.Dr. Atatürk Üniversitesi, İktisadi ve İdari Bilimler Fakültesi, Ekonometri Bölümü.
E-mail: erkanoktay@atauni.edu.tr

genel amaç, durumların bahis değerlerini ortak değişkenlerin bir fonksiyonu olarak modellemek ve sonuçları farklı durumların tercihleri için bahis değerleri açısından açıklayabilmektir (Hosmer ve Lemeshow, 2000). Tesadüfi fayda teorisi, tüketicilerin belirli alternatifler arasından kendilerine en yüksek faydayı sağlayacak alternatifini tercih edeceği varsayımına dayanmakta ve çok durumlu kesikli tercih modellerini tesadüfi fayda teorisi temelinde değerlendirmek mümkün olabilmektedir.

Gerçek hayatta birçok tercih kümesi, kendi arasında doğal bir sıralamaya sahip olmayan ikiden çok alternatifte sahiptir. Örneğin; hanehalklarının yolculuk yapma amaçları (eğitim, iş, alışveriş, eğlence, vb.), bireylerin ulaştırma tercihleri (özel araç, uçak, otobüs, metro, tren, vb.), hanehalklarının konut tercihleri (apartman, müstakil ev, lojman, diğer, vb.), bireylerin dinlediği favori müzik türleri (klasik, popüler, rock, vb.), hanehalklarının tasarruf tercihleri (gayrimenkul, altın, döviz, hisse senedi, vb.) çok durumlu tercih kümeleri arasında gösterilebilmektedir. Çok durumlu logit ve probit modeller, bağımlı değişken kategorilerinin ikiden fazla olması ve sözü edilen kategorilerin kendi arasında doğal bir sıralamaya sahip olmaması durumunda kullanılmaktadır. Çok durumlu modeller, uygulamada araştırmacılar tarafından sıklıkla kullanılan kesikli tercih modelleri arasında yer almaktadır. Bu bölümde çok durumlu probit ve çok durumlu logit modeller tanıttıldıktan sonra sözü edilen iki model Stata uygulamalarıyla anlatılacaktır.

4.1. Çok Durumlu Logit Modeli

Bir bireyin x_j nitelik vektörü ile ifade edilen J adet alternatif ile ($j = 1, \dots, J$) karşı karşıya kaldığı varsayıldığında, söz konusu bireye ait fayda fonksiyonu,

$$U = V(s, x) + \varepsilon(s, x) \quad (4.1)$$

eşitliği ile gösterilmektedir. Eşitlik (4.1)'de, V değişkeni tesadüfi olmayan bir değişkendir ve toplumun "temsili" beğenilerini göstermektedir. Bununla birlikte, s değişkeni ölçülen nitelikleri temsil etmekte ve ε değişkeni tesadüfi bir değişken olup sözü edilen bireyin kişisel beğenilerini yansıtmaktadır. Söz konusu birey, en yüksek faydayı sağlayan alternatifini tercih etmektedir. Bireyin davranış ilkesinin h_ε ve $B = \{x_1, \dots, x_J\}$ olduğu varsayıldığında, s özelliğine ve B alternatif tercihler kümesine sahip ana kütleden tesadüfi olarak seçilen bir bireyin x_i tercihinde bulunma olasılığı,

$$P_i \equiv P(x_i | s, B) = \pi[\{h_\varepsilon \in H | h_\varepsilon(s, B) = x_i\}] \quad (4.2)$$

$$P_i = P[\varepsilon(s, x_j) - \varepsilon(s, x_i) < V(s, x_i) - V(s, x_j)] \text{ her } j \neq i \text{ için} \quad (4.3)$$

şeklinde ifade edilmektedir. Eşitlik (4.2) ve Eşitlik (4.3)'teki π olasılığı, $j = 1, \dots, J$ için $\varepsilon_j = \varepsilon(s, x_j)$ değerleri üzerinde $F(\varepsilon_1, \dots, \varepsilon_J)$ ya da,

$$F(\varepsilon_1, \dots, \varepsilon_J) = \pi[\{h_\varepsilon \in H | \varepsilon(s, x_j) \leq \varepsilon_j, j = 1, \dots, J \text{ için}\}] \quad (4.4)$$

şeklinde bir bütünleşik birikimli dağılım fonksiyonu oluşturmaktadır. Ayrıca, F_i 'nin i 'nci değişkene göre kısmi türev ve $V_i = V(s, x_j)$ olduğu varsayıldığında Eşitlik (4.4),

$$P_i = \int_{\delta=-\infty}^{+\infty} F_i(\varepsilon + V_i - V_i, \dots, V_i - V_j) d\varepsilon \quad (4.5)$$

şeklinde yeniden yazılabilmektedir (McFadden, 1974).

Tesadüfi olarak çekilen en yüksek değerlerin dağılımları, aykırı değer dağılımları olarak adlandırılmakta ve birinci, ikinci ve üçüncü tip dağılımlar olmak üzere üç sınıfa ayrılmaktadır. Bu dağılım tipleri arasında en yaygın kullanılan dağılım Gumbel dağılımı olarak da adlandırılan birinci tip dağılımdır. Aykırı değer birinci tip dağılımda tesadüfi olarak çekilen en yüksek değerler de aykırı değer birinci tip dağılım göstermektedir. Aykırı değer birinci tip dağılımın olasılık yoğunluk fonksiyonu,

$$f(\varepsilon) = \eta e^{[-\eta(\varepsilon-\omega)]} e^{-e^{[-\eta(\varepsilon-\omega)]}} \quad (4.6)$$

biçiminde gösterilmektedir. Eşitlik (4.6)'da η pozitif ölçüm parametresini ve ω konum parametresini temsil etmektedir. Aykırı değer birinci tip dağılımı üzerinde tahmin edilebilir bir model türetmek için Eşitlik (4.6) yeniden gözden geçirilirse,

$$P_n(i) = P\left(\beta_i X_{in} + \varepsilon_{in} \geq \max_{\forall l \neq i} (\beta_l X_{ln} + \varepsilon_{ln})\right) \quad (4.7)$$

elde edilmektedir. Aykırı değer birinci tip dağılım için eğer tüm ε_{ln} hata terimleri, ω_{ln} konum parametreleri ve ortak bir η ölçüm parametresi (bir başka deyişle eşit varyansa) ile özdeş dağılıma sahip tesadüfi değişkenler ise bu durumda $\beta_l X_{ln} + \varepsilon_{ln}$ ifadesinin en yüksek değeri,

$$\frac{1}{\eta} \ln \sum_{\forall l \neq i} e^{(\eta \beta_l X_{ln})} \quad (4.8)$$

konum parametresi ile aykırı değer birinci tip dağılıma sahip olmaktadır. Eğer;

ε'_i : i değerine eşit olmayan tüm olası kesikli sonuçların en yüksek değeri ile ilişkili hata terimini,

$\beta' X'_n$: i değerine eşit olmayan tüm olası kesikli sonuçların en yüksek değeri ile ilişkili bağımsız değişken sonuçlarını temsil ederse,

$$\beta' X'_n = \frac{1}{\eta} \ln \sum_{\forall l \neq i} e^{(\eta \beta_l X_{ln})} \quad (4.9)$$

olduğu gösterilebilmektedir. Bu sonuçları kullanarak Eşitlik (4.7)

$$P_n(i) = P(\beta_i X_{in} + \varepsilon_{in} \geq \beta' X'_n + \varepsilon'_n) \quad (4.10)$$

ya da

$$P_n(i) = P(\beta' X'_n + \varepsilon'_n - \beta_i X_{in} + \varepsilon_{in} \leq 0) \quad (4.11)$$

biçimlerinde yeniden yazılmaktadır. Bununla birlikte, ortak ölçüm parametresi η olan iki bağımsız dağılmış aykırı değer birinci tip bağımsız değişken arasındaki fark, lojistik dağılımlı olduğundan

$$P_n(i) = \frac{1}{1 + e^{\eta(\beta' X_n' - \beta_i X_{in})}} \quad (4.12)$$

ya da terimler yeniden düzenlenerek

$$P_n(i) = \frac{e^{\eta(\beta_i X_{in})}}{e^{\eta(\beta_i X_{in})} + e^{\eta(\beta' X_n')}} \quad (4.13)$$

biçiminde ifade edilmektedir. Eşitlik (4.13)'te η parametresi 1'e eşitlenerek,

$$P_n(i) = \frac{e^{(\beta_i X_{in})}}{e^{(\beta_i X_{in})} + e^{[\ln \sum_{l \neq i} e^{\beta_l X_{ln}}]}} \quad (4.14)$$

ya da daha basit bir ifade ile

$$P_n(i) = \frac{e^{(\beta_i X_{in})}}{\sum_{l \neq i} e^{\beta_l X_{ln}}} \quad (4.15)$$

nihai standart çok durumlu logit modeli eşitliği elde edilmektedir. Parametre vektörleri β değerlerinin en çok olabilirlik yöntemi ile tahmini için log-olabilirlik fonksiyonu ise

$$LL = \sum_{n=1}^N \left(\sum_{i=1}^I \delta_{in} \left[\beta_i X_{in} - \ln \sum_{l \neq i} e^{\beta_l X_{ln}} \right] \right) \quad (4.16)$$

biçiminde ifade edilmektedir. Eşitlik (4.16)'da I sonuç değişkeni sayısını temsil etmektedir. δ_{in} değişkeninin n gözlemi için i sonucunu vermesi hâlinde 1'e, aksi bir durumda sıfıra eşit olduğu kabul edilmektedir (Washington ve diğ., 2011).

4.1.1. İlişkisiz Alternatiflerin Bağımsızlığı Varsayımı

Bir tercih modelinde, U olası tüm alternatifleri kapsayan bir evrensel kümeyi, T alt kümesi U evrensel kümesinin bir alt kümesini, x değişkeni T alt kümesinin bir elemanını, $P_T(x)$ ifadesi x elemanının tercih edilme olasılığını temsil etsin. Bununla birlikte, S alt kümesinin T alt kümesinin başka bir alt kümesini ve $P_T(S)$ ifadesinin ise S alt kümesinden bir eleman seçilme olasılığını temsil ettiği varsayalım. T kümesinden bir tercihte bulunduğu; eğer tüm $x, y \in T$ için $P(x, y) \neq 0$ ise, $R \subset S \subset T$ için $P_T(R) = P_S(R)P_T(S)$;

eğer bazı $x, y \in T$ için $P(x, y) = 0$ ise, her bir $S \subset T$ için $P_T(S) = P_{T-\{x\}}(S - \{x\})$

aksiyomları öne sürülmektedir. Bu aksiyom T alt kümesinin tüm alt kümeleri için de geçerli olmaktadır. Eğer tüm $x, y \in T$ için $P(x, y) \neq 0$ ise, herhangi bir $S \subset T$ için $x, y \in S$ olacak şekilde,

$$\frac{P(x, y)}{P(y, x)} = \frac{P_S(x)}{P_S(y)} \quad (4.17)$$

eşitliğine ulaşılabilmektedir. Eşitlik (4.17), en basit ifadeyle tercihler kümesindeki tüm olası alternatifler arasından bir alternatifin tercih edilme olasılığının diğer bir alternatifin tercih edilme

olasılığına bağlı olmaması gerektiğini göstermekte ve ilişkisiz alternatiflerin bağımsızlığı olarak adlandırılmaktadır (Luce, 1959).

Çok durumlu logit modeli, uygulamada sıklıkla kullanılan kesikli tercih modellerinden biridir. Çok durumlu logit modeli, iki durumlu logit modelinin genelleştirilmiş bir durumudur ve gruplandırılmış veri setleri ile log-doğrusal modele eş değerdir (Powers and Xie, 2000; Üçdoğruk, 2002). Çok durumlu logit modeli, normallik, doğrusallık, eşit varyans gibi koşulları gerektirmektedir. Bununla birlikte, birçok araştırmacı tarafından modelin en önemli kısıtlamalarından birinin ilişkisiz alternatiflerin bağımsızlığı varsayımı olduğu ileri sürülmektedir. İlişkisiz alternatiflerin bağımsızlığı varsayımı, aynı koşullar altında iki alternatif arasından bir bireyin yaptığı tercihin uygun diğer tercihlerden etkilenmeyeceğini öne sürmekte ve öne sürülen sıfır hipotezi;

H_0 : İki alternatif için olasılık değerleri diğer alternatiflerden bağımsızdır

şeklinde tanımlanabilmektedir.

İlişkisiz alternatifler varsayımının ihlal edilip edilmediğini sınamak için tercih kümesi paylaşımı ve modele dayalı sınamalar olmak üzere iki temel sınama yöntemi kullanılabilmektedir (Cheng ve Long, 2007). İlişkisiz alternatiflerin bağımsızlığı varsayımını sınamak için geliştirilen birinci test, McFadden ve diğ. (1981) tarafından öne sürülmüştür. Olabilirlik oranı temelinde geliştirilen bu test, sınırlı model tahmininden elde edilen log-olabilirlik eşitliğinin değerini tüm model tahminlerini sınırlı model tahminlerinin yerine koyarak elde edilen değerler ile karşılaştırmaktadır. İlerleyen yıllarda, Small ve Hsiao (1985) bu olabilirlik testinin asimtotik sapmalı olduğu ispatlamışlar ve bu sapmalı tahminleri gideren alternatif bir olabilirlik oranı testi geliştirmişlerdir. Hausman ve McFadden (1984) ise sınırlı ve tüm modelleri karşılaştıran üçüncü bir test öne sürmüşlerdir (Cheng ve Long, 2007).

Çok durumlu bir logit modeli, iki durumlu logit değerlerinin bağımlı değişken kategorilerinin tüm çiftleri için eşanlı olarak tahmin edilmesi olarak görülebilmektedir. Söz konusu modelin etkin bir şekilde tahmin edilmesi, parametreler üzerinde birtakım belirli mantıksal kısıtlamaların olmasını gerektirmektedir. Örneğin, bağımlı değişkenin üç kategoriye sahip olduğu çok durumlu bir logit modeli iki durumlu logit değerleri göz önünde bulundurularak tahmin edilebilmektedir. Bu durumda, birinci logit değeri birinci kategoriye ikinci kategori ile ve ikinci logit değeri ise birinci kategoriye üçüncü kategori ile karşılaştırmaktadır. Buna göre, bir y değişkeni, 1'den J 'ye kadar numaralandırılmış J kategoriye sahip bir bağımlı değişkeni; x vektörü, K bağımsız değişkeni ve kesim noktasını gösteren bir sabit terimi kapsayan vektörü temsil etmek üzere, verilen x vektörü için m kategorisinin gözlenme olasılığını gösteren çok durumlu bir logit modeli,

$$P(y = m|x) = \frac{\exp(x\beta_m)}{\sum_{j=1}^J \exp(x\beta_j)} \quad m = 1, \dots, J \text{ için} \quad (4.18)$$

biçiminde tarif edilmektedir. Eşitlik (4.18)'de $\beta_m = (\beta_{0m} \dots \beta_{km} \dots \beta_{Km})'$ vektörü, β_{0m} kesim noktası ve x_k bağımsız değişkeninin m kategorisindeki etkisini gösteren β_{km} katsayılarını kapsamaktadır. Eşitlik (4.18)'de tanımlanan tüm model için tahmin değerleri $\hat{\beta}_m^f$ için f simgesi, tahminlerin tüm sonuçları içeren modelden elde edildiğini temsil etmektedir. Sınırlandırılmış tahmin modeli, tüm modelden J kategorisinin dışlanması ile

$$P(y = m|\mathbf{x}) = \frac{\exp(\mathbf{x}\beta_m)}{\sum_{j=1}^{J-1} \exp(\mathbf{x}\beta_m)} \quad m = 1, \dots, J \text{ için} \quad (4.19)$$

biçiminde oluşturulabilmektedir. Eşitlik (4.19)'da gösterilen bu sınırlandırılmış modelde $\beta_1 = 0$ olduğu varsayılmakta ve modelden J kategorisi dışında herhangi keyfi bir kategori de dışlanabilmektedir. İlişkisiz alternatiflerin bağımsızlığı varsayımı altında, $\hat{\beta}_m^f$ model tahminleri hem tutarlı hem de etkin tahminler olmaktadır. Bununla birlikte, sınırlandırılmış tercih kümesinden elde edilen $\hat{\beta}_m^r$ sınırlandırılmış tahminleri tutarlı olmasına rağmen etkin tahminler olmamaktadır. İlişkisiz alternatiflerin bağımsızlığı varsayımını sınavan birçok test de benzer şekilde tüm model tahminleri ile sınırlandırılmış model tahminleri üzerine yoğunlaşmıştır. Sınırlandırılmış tercih kümesinden elde edilen tahmin katsayılarının vektörü $\hat{\beta}^r = (\hat{\beta}_2^r \dots \hat{\beta}_{2J-1}^r)'$ ve tüm modelden tahmin katsayılarının vektörü ise $\hat{\beta}^f = (\hat{\beta}_2^f \dots \hat{\beta}_{2J-1}^f)'$ ile temsil edilmektedir. Bu noktada, $\hat{\beta}^f$ vektörünün içerisinde, sınırlandırılmış modelde tahmin edilmediği için $\hat{\beta}_j^f$ katsayısının bulunmadığı görülmektedir (Cheng ve Long, 2007).

İlişkisiz alternatiflerin bağımsızlığını sınamak için McFadden ve diğ. (1981) tarafından öne sürülen yaklaşık olabilirlik oranı testi,

$$MTT = -2[L_r(\hat{\beta}^f) - L_r(\hat{\beta}^r)] \quad (4.20)$$

biçiminde tanımlanmaktadır. Eşitlik (4.20)'de L_r sınırlandırılmış tahmin için bir log-olabilirlik fonksiyonunu temsil etmektedir. İlişkisiz alternatiflerin bağımsızlığı varsayımı sağlandığında, MTT ifadesinin $\hat{\beta}_r$ tahmin vektörüne ait sıra değerlerine eşit serbestlik derecesine sahip bir ki-kare dağılımı gösterdiği kabul edilmektedir (Cheng ve Long, 2007). Small ve Hsiao (1985), MTT testinin sıfır hipotezini kabul etme noktasında asimptotik sapmalı olduğunu ispatlamış ve bu sorunu düzeltmek için düzeltilmiş bir MTT testi öne sürmüşlerdir. İlk aşamada örnek, tesadüfi olarak kabaca eşit büyüklükte A ve B gibi iki alt örneğe ayrılmıştır. Sonraki aşamada, model tahmini, $\hat{\beta}_A^f$ ve $\hat{\beta}_B^f$ tahmin katsayıları vasıtasıyla gerçekleştirilmektedir. Söz konusu iki alt örnekten elde edilen katsayıların ağırlıklı ortalaması,

$$\hat{\beta}_{AB}^f = \left(\frac{1}{\sqrt{2}}\right)\hat{\beta}_A^f + \left[1 - \left(\frac{1}{\sqrt{2}}\right)\right]\hat{\beta}_B^f \quad (4.21)$$

biçiminde olmaktadır. B alt örneğinden bağımlı değişkenin belirli bir kategorisine ait bütün durumların dışlandığı bir sınırlandırılmış alt örnek türetilmektedir. Keyfi olarak J kategorisi sınırlandırılmış alt örnekten dışlandığında sınırlandırılmış tercih kümesi, L_r olabilirlik oranı fonksiyonu ve $\hat{\beta}_B^r$ tahminlerine sahip sınırlandırılmış alt örnek kullanılarak tahmin edilmektedir. Böylece, sınırlandırılmış tercih kümesindeki parametre sayısına eşit serbestlik dereceli asimptotik ki-kare dağılımına sahip Small-Hsiao test istatistiği,

$$SH = -2[L_r(\hat{\beta}_{AB}^f) - L_r(\hat{\beta}_B^r)] \quad (4.22)$$

biçiminde tanımlanmaktadır (Cheng ve Long, 2007).

Hausman ve McFadden (1984), sıfır hipotezinin doğru olması durumunda tutarlı ve etkin $\hat{\beta}^f$ tahminlerini, tutarlı olmasına rağmen etkin olmayan $\hat{\beta}^r$ tahminleri ile karşılaştırmak için bir Hausman (1978) testi öne sürmüşlerdir. Buna göre test istatistiği,

$$HM = (\hat{\beta}^r - \hat{\beta}^f)' [\widehat{Var}(\hat{\beta}^r) - \widehat{Var}(\hat{\beta}^f)]^{-1} (\hat{\beta}^r - \hat{\beta}^f) \quad (4.23)$$

biçiminde yazılmaktadır. Eşitlik (4.23)'te $\widehat{Var}(\hat{\beta}^r)$ ve $\widehat{Var}(\hat{\beta}^f)$ tahmin edilmiş eşit varyans matrislerini temsil etmektedir. İlişkisiz alternatiflerin bağımsızlığı varsayımı sağlandığında HM test istatistiği, $\hat{\beta}^r$ sıra değerlerine eşit serbestlik dereceli asimtotik bir ki-kare dağılımına sahip olmaktadır. HM test istatistiğinin istatistiksel olarak anlamlı değerleri, ilişkisiz alternatiflerin bağımsızlığı varsayımının ihlal edildiğine işaret etmektedir. Hausman ve McFadden (1984), HM test istatistiğinin negatif olabileceği ve bu durumda ilişkisiz alternatiflerin bağımsızlığı varsayımının ihlal edilmeyeceği sonucuna ulaşmışlardır (Cheng ve Long, 2007; Çelik, 2016).

4.2. Çok Durumlu Logit Modelinin Stata Uygulaması

Çok durumlu logit modelinin Stata/MP 16.0 paket programı yardımıyla ekonometrik analizi gerçekleştirilmiştir. 2018 yılı Türkiye İstatistik Kurumu Hanehalkı (TÜİK) Bütçe Araştırması meta veri setinden faydalanılarak (Türkiye İstatistik Kurumu, 2018a), Türkiye'de hanehalklarının ısınmak amacıyla kullandıkları birinci yakıt türü tercihlerini etkileyen faktörler tespit edilmeye çalışılmaktadır. 2018 TÜİK Hanehalkı Bütçe Araştırması, 11.828 hane ve 40.688 ferde uygulanmıştır. TÜİK Hanehalkı Bütçe Araştırması, 1987 yılından başlayarak ve 2002 yılından itibaren ise her yıl düzenli olarak devam eden, geniş kapsamlı tekrarlı yatay kesit verilerinden oluşan bir araştırma olup (Türkiye İstatistik Kurumu, 2018b), tabakalı iki aşamalı küme örnekleme kullanılmaktadır (Türkiye İstatistik Kurumu, 2018b). Hanehalklarının ısınmak amacıyla tercih ettikleri yakıt türleri, en yüksek frekans değerlerine sahip dört kategori üzerinden değerlendirilmiştir. Tablo 4.1'de çok durumlu modellerin tahmini amacıyla kullanılan veri setindeki değişkenlere ait tanımlayıcı istatistikler sunulmaktadır¹. Tablo 4.1'deki frekans ve yüzde değerlerini Stata paket programında elde etmek için “*tabulate*” komutundan faydalanılmakta ve her bir değişkene ait söz konusu komut tekrarlanmaktadır.

Tablo 4.1. Değişkenlere ait Açıklamalar

Değişkenler	Etiket	Frekans	Yüzde
<i>Bağımlı değişken</i>			
Isınmak Amacıyla Kullanılan Birinci Yakıt Türü	Yakıt_Türü		
Kömür ve Kömür Türevleri*		4.155	35,13
Odun		1.739	14,70
Doğalgaz		5.052	43,71
Diğer yakıt türleri		882	7,46
<i>Bağımsız değişkenler (kesikli değişkenler)</i>			
Hanehalkı tipi			
Tek kişilik hanehalkı	Tek_Kişi_Hane	1.149	9,71
Çekirdek aile	Çekirdek_Aile	8.949	75,66
Diğer hanehalkı tipleri*	Diğer_Hane_Tipi	1.730	14,63

¹ mlogit.dta veri setine Uygulamalarla Mikro Ekonometri kitabının linkinden ulaşabilirsiniz.

Konutun büyüklüğü (m ²)			
100 m ² 'den küçük*	Küçük_Konut	4.664	39,43
100 – 149 m ²	Orta_Konut	5.923	50,08
150 m ² ve daha büyük	Büyük_Konut	1.241	10,49
Konutun ısıtma sistemi			
Kat kaloriferi sistemi (Kombi vb.)*	Kat_Kaloriferi	4.538	38,37
Soba (odun, kömür, doğalgaz, elektrik, tüp, vb.)	Soba	5.753	48,64
Bina içi merkezi kalorifer sistemi	Merkezi_Sistem	1.134	9,59
Diğer ısıtma sistemleri	Diğer_Isıtma	403	3,41
Otomobil sahipliği	Oto_Sahipliği		
Evet		5.110	43,20
Hayır*		6.718	56,80
<i>Bağımsız değişkenler (sürekli değişkenler)</i>		Ortalama	Std. Sap.
Konutun oda sayısı	Oda_Sayısı	3,56	0,85
Yıllık hanehalkı kullanılabilir gelirin logaritması	Log_YıllıkGelir	4,65	0,28

*Referans kategori

Stata paket programında **“describe”** komutu kullanılarak değişken tanımları listelenebilmekte ve Tablo 4.1’de özetlenen değişken tanımları, **“describe”** komutu ile Tablo 4.2’de aşağıdaki gibi gösterilebilmektedir.

Tablo 4.2. Değişkenlerin Stata Paket Programında Özetlenmesi
describe (Tablo 4.2. sayfa 101’de verilmiştir)

Stata paket programında; bağımlı değişken ve bağımsız değişkenlere ait özet istatistikleri için **“sum”** komutu kullanılmaktadır. Tablo 4.3’te görüldüğü üzere uygulamada kullanılan birçok değişken, değişken kategorilerinin bağımlı değişkene etkisini yansıtabilmek için kukla değişken olarak tanımlanmıştır. Bununla birlikte; konut oda sayısı, otomobil sahipliği ve hanehalkı yıllık kullanılabilir gelir değişkenleri sürekli değişken olarak kullanılmıştır.

Tablo 4.3. Değişkenlere ait Tanımlayıcı İstatistikler

sum

Variable	Obs	Mean	Std. Dev.	Min	Max
Yakıt_Türü	11,828	2.224975	1.012993	1	4
Tek_Kişi_Hane	11,828	.0971424	.2961641	0	1
Çekirdek_Aile	11,828	.7565945	.4291559	0	1
Diğer_Hane_Tipi	11,828	.1462631	.3533847	0	1
Oda_Sayısı	11,828	3.556392	.8495214	1	14
Küçük_Konut	11,828	.3943186	.4887245	0	1
Orta_Konut	11,828	.5007609	.5000206	0	1
Büyük_Konut	11,828	.1049205	.3064639	0	1
Kat_Kaloriferi	11,828	.3836659	.4862986	0	1
Soba	11,828	.4863882	.4998358	0	1

Merkezi_Sistem	11,828	.0958742	.2944311	0	1
Diğer_Isıtma	11,828	.0340717	.181421	0	1
Oto_Sahipliği	11,828	.4320257	.4953789	0	1
Log_YıllıkGelir	11,828	4.654381	.2826901	3.274095	6.191121

Çok durumlu modelleri tahmin etmeden önce bağımsız değişkenler arasında çoklu doğrusal bağlantı sorunu olup olmadığının test edilmesi gerekmektedir. Varyans şişkinlik faktörü (VIF), çoklu doğrusal sorununu test etmek için uygulamada sıklıkla kullanılmaktadır. Genel olarak VIF değerinin 5'ten küçük olması, bağımsız değişkenler arasında çoklu doğrusal bağlantı sorunu olmadığına işaret etmektedir. Çoklu doğrusal bağlantı testi için öncelikle **"regress"** komutu ile standart bir doğrusal regresyon modeli tahmin edilmekte ve ardından **"estat vif"** komutu ile bağımsız değişkenler arasında çoklu doğrusal bağlantı sorunu olup olmadığı test edilmektedir. Tablo 4.4'te uygulamada kullanılan bağımsız değişkenlere ait VIF değerleri sunulmaktadır. Tablo 4.4'te görüldüğü gibi tüm bağımsız değişkenlere ait VIF değerleri tavsiye edilen değer olan 5'in altında (Gujarati, 2011) olduğu için kullanılan bağımsız değişkenler arasında tahmin sonuçlarını etkileyecek bir çoklu doğrusal bağlantı sorunu olmadığı tespit edilmiştir.

Tablo 4.4. Çoklu Doğrusal Bağlantı Testi

estat vif

Variable	VIF	1/VIF
Oda_Sayısı	1.85	0.541023
Büyük_Konut	1.84	0.544543
Tek_Kişi_Hane	1.69	0.590471
Orta_Konut	1.61	0.621674
Çekirdek_Aile	1.56	0.642228
Log_YıllıkGelir	1.55	0.644608
Soba	1.41	0.709261
Oto_Sahipliği	1.20	0.830662
Merkezi_Sistem	1.15	0.870630
Diğer_Isıtma	1.06	0.946657
Mean VIF	1.49	

4.2.1. Çok Durumlu Logit Modelinin Tahmin Sonuçları

Tablo 4.5'te çok durumlu logit modelinin en çok olabilirlik tahmin sonuçları sunulmaktadır. Stata paket programında çok durumlu logit modelini tahmin edebilmek için **"mlogit"** komutu kullanılmaktadır. Tahmin yapabilmek için bağımsız değişken kategorilerden bir tanesinin referans kategorisi olarak seçilmesi gerekmektedir. Referans kategorisi seçilmediği takdirde, tam çoklu doğrusallık sorununa yol açmamak için Stata programı her bir bağımsız değişken kategorisinden bir tanesini kendiliğinden modelden dışlamaktadır. **"mlogit"** komutunda bağımlı değişken kategorilerinden bir tanesinin de referans kategorisi olarak seçilmesi gerekmektedir. Bağımlı değişken için referans kategorisi tanımlayabilmek için **"mlogit"** komutundan sonra değişkenler yazıl-

dıktan sonra “base” yazılarak referans kategorisinin bağımlı değişken içerisindeki sırası yazılmaktadır. Tablo 4.5’te görüldüğü gibi komut satırında “base(1)” yazılması uygulamadaki referans bağımlı değişken kategorisinin “kömür ve kömür türevleri” olduğunu işaret etmektedir. Stata paket programı ile çok durumlu logit modeli tahmininde “mlogit” komutu “base” çalıştırılmadan da tahmin edilebilmekte ve bu durumda tahmin amacıyla kullanılacak referans bağımlı değişken kategorisini program belirlemektedir. Benzer şekilde “level” komutu ile hangi önem seviyesinde katsayıların tahmin edileceği de belirlenebilmektedir. Sözü edilen komut yazılmadığında çok durumlu logit modeli, %5 önem seviyesinde tahmin edilmektedir.

Tablo 4.5. Çok Durumlu Logit Modeli Tahmin Sonuçları

mlogit Yakıt_Türü Tek_Kişi_Hane Çekirdek_Aile Oda_Sayısı Orta_Konut Büyük_Konut Soba Merkezi_Sistem Diğer_Isıtma Oto_Sahipliği Log_YıllıkGelir, base(1)

Iteration 0: log likelihood = -14268.087

Iteration 8: log likelihood = -7519.625

Multinomial logistic regression	Number of obs	=	11,828
	LR chi2(30)	=	13496.92
	Prob > chi2	=	0.0000
Log likelihood = -7519.625	Pseudo R2	=	0.4730

Yakıt_Türü	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
Kömür	(base outcome)					
Odun						
Tek_Kişi_Hane	.4748739	.1245612	3.81	0.000	.2307383	.7190094
Çekirdek_Aile	.1902857	.0799989	2.38	0.017	.0334908	.3470807
Oda_Sayısı	-.062912	.043248	-1.45	0.146	-.1476766	.0218525
Orta_Konut	-.1318186	.0722658	-1.82	0.068	-.273457	.0098198
Büyük_Konut	.0138219	.1370834	0.10	0.920	-.2548567	.2825005
Soba	2.01503	.2893484	6.96	0.000	1.447918	2.582143
Merkezi_Sistem	-2.542722	.6462939	-3.93	0.000	-3.809435	-1.276009
Diğer_Isıtma	.9377562	.5652595	1.66	0.097	-.1701321	2.045644
Oto_Sahipliği	.124828	.0657544	1.90	0.058	-.0040483	.2537044
Log_YıllıkGelir	.3111147	.1332981	2.33	0.020	.0498553	.5723741
_cons	-4.070115	.6862224	-5.93	0.000	-5.415086	-2.725143
Doğalgaz						
Tek_Kişi_Hane	1.171842	.1846451	6.35	0.000	.8099445	1.53374
Çekirdek_Aile	.7952252	.1258045	6.32	0.000	.5486529	1.041797
Oda_Sayısı	-.1124855	.0699472	-1.61	0.108	-.2495795	.0246084
Orta_Konut	-.6343917	.113282	-5.60	0.000	-.8564204	-.4123631
Büyük_Konut	-.3577066	.1753656	-2.04	0.041	-.7014168	-.0139963
Soba	-6.3042	.1230036	-51.25	0.000	-6.545283	-6.063117
Merkezi_Sistem	-3.222328	.1017265	-31.68	0.000	-3.421708	-3.022948

Diger_Isıtma	-3.965354	.3352399	-11.83	0.000	-4.622412	-3.308295
Oto_Sahipliği	-.5262235	.0915292	-5.75	0.000	-.7056174	-.3468296
Log_YıllıkGelir	2.079723	.1807577	11.51	0.000	1.725445	2.434002
_cons	-6.272737	.8516753	-7.37	0.000	-7.94199	-4.603485

Diger						
Tek_Kişi_Hane	.5347685	.1788834	2.99	0.003	.1841634	.8853736
Çekirdek_Aile	-.0295929	.1201284	-0.25	0.805	-.2650403	.2058545
Oda_Sayısı	-.3290452	.0690368	-4.77	0.000	-.4643548	-.1937356
Orta_Konut	.2616962	.1137578	2.30	0.021	.0387349	.4846574
Büyük_Konut	.8273366	.1927227	4.29	0.000	.4496071	1.205066
Soba	.3889268	.2545876	1.53	0.127	-.1100557	.8879093
Merkezi_Sistem	-.9323458	.3302334	-2.82	0.005	-1.579591	-.2851002
Diger_Isıtma	4.888562	.3186928	15.34	0.000	4.263936	5.513189
Oto_Sahipliği	-.3432282	.1049583	-3.27	0.001	-.5489426	-.1375138
Log_YıllıkGelir	.8027702	.2018235	3.98	0.000	.4072035	1.198337
_cons	-5.001033	.9749984	-5.13	0.000	-6.911994	-3.090071

Tablo 4.5'te görüldüğü gibi tahmin edilen çok durumlu logit modelinin %1 önem seviyesinde istatistiksel olarak anlamlı (LR ki-kare: 13.496,92; $p = 0.0000$) olduğu tespit edilmektedir. Tahmin edilen çok durumlu logit modeli, 11.828 gözlem ile 8 yinelemede (log-olabilirlik: -7.519,625) yakınsamıştır. Tahmin edilen çok durumlu logit modeli için hesaplanan Pseudo- R^2 değeri 0,4730 olarak bulunmuş ve mevcut literatürde kesikli tercih modelleri için tavsiye edilen Pseudo- R^2 değerlerine göre (Louviere ve diğ., 2000) istatistiksel olarak oldukça iyi bir uyum gösterdiği söylenebilmektedir. Çok durumlu logit modeli tahmin sonuçları, her bir bağımlı değişken kategorisi için bağımsız değişkenlere ait ayrı ayrı hesaplanmaktadır. Tablo 4.5'teki Stata çıktısındaki “_cons” değeri ise modelin sabit terimine ait tahmin sonuçlarını göstermektedir. Çok durumlu logit modeline ait tahmin sonuçları, bağımlı değişken ile bağımsız değişkenler arasındaki ilişkinin yönü hakkında bilgi vermektedir. Sözü edilen ilişkinin büyüklüğü hakkında bilgi edinmek için göreceli risk oranı, marjinal etkiler ve esneklik değerleri gibi istatistikler hesaplanmaktadır.

Çok durumlu logit modelinin en önemli varsayımı, ilişkisiz alternatiflerin bağımsızlığı varsayımıdır. Bu varsayımı test etmek için kullanılan en önemli testler Hausman-McFadden (1984) ve Small-Hsiao (1985) testleridir. Stata paket programında Hausman-McFadden testi için “**mlogtest, hausman**” ve Small-Hsiao testi için “**mlogtest, smhsiao**” komutları kullanılmaktadır. Tablo 4.6'da tahmin edilen çok durumlu logit modeli için Hausman-McFadden ve Small-Hsiao test sonuçlarını göstermektedir. Tablo 4.6'da elde edilen çıktılardaki en sağ sütunda tahmin edilen modele ait bağımlı değişken kategorileri için istatistiksel anlamlılık sonuçları gösterilmektedir. Tahmin edilen çok durumlu logit modelinin ilişkisiz alternatiflerin bağımsızlığı varsayımını ihlal etmediğini istatistiksel olarak ifade edebilmek için sözü edilen anlamlılık sonuçlarının 0.10 değerinden yüksek olması gerekmektedir. Tablo 4.6'da görüldüğü gibi uygulamada tahmin edilen çok durumlu logit modeline ait hesaplanan bağımlı değişken istatistiksel anlamlılık değerleri 0.10'dan yüksek olduğu için “olasılıkların diğer alternatifler için bağımsız olduğunu” öne süren H_0 hipotezi kabul edilmektedir. Böylece, tahmin edilen çok durumlu logit modelinin Hausman-McFadden ve Small-Hsiao testlerine göre ilişkisiz alternatiflerin bağımsızlığı varsayımını sağla-

dığı söylenebilmektedir. Stata paket programında, “**mlogtest, iia**” komutu kullanılarak tüm ilişkisiz alternatiflerin bağımsızlığı testi sonuçları elde edilebilmektedir. Long ve Freeze (2001), çok durumlu logit modeline ait farklı ilişkisiz alternatiflerin bağımsızlığı testlerinin farklı sonuçlar verebildiğini öne sürmektedir.

Tablo 4.6. İlişkisiz Alternatiflerin Bağımsızlığı Varsayımı Testleri
mlogtest, hausman

Hausman tests of IIA assumption (N=11828)

Ho: Odds(Outcome-J vs Outcome-K) are independent of other alternatives

	chi2	df	P>chi2
Kömür	14.057	22	0.899
Odun	0.863	22	1.000
Doğalgaz	1.649	22	1.000
Diğer	0.682	22	1.000

Note: A significant test is evidence against Ho.

mlogtest, smhsiao

Small-Hsiao tests of IIA assumption (N=11828)

Ho: Odds(Outcome-J vs Outcome-K) are independent of other alternatives

	lnL(full)	lnL(omit)	chi2	df	
Kömür	-1011.363	-996.418	29.889	22	0.121
Odun	-1900.970	-1893.254	15.432	22	0.843
Doğalgaz	-2632.974	-2622.814	20.321	22	0.563
Diğer	-2675.973	-2665.449	21.047	22	0.518

Note: A significant test is evidence against Ho.

Çok durumlu logit modeline ait tahmin sonrası olabilirlik oranı testi ve Wald testleri de hesaplanmaktadır. Birinci olabilirlik oranı ve Wald testlerine ait sıfır hipotezi (H_0), bağımsız değişken katsayılarının tüm çok durumlu logit modeli eşitlikleri için sıfır olduğunu öne sürmektedir. Öte yandan, ikinci olabilirlik oranı ve Wald testlerinde ise herhangi bir bağımlı değişken kategorisi çiftinin birleştirilip birleştirilemeyeceğini test etmektedir. Tablo 4.7’de bağımsız değişkenler için hesaplanan olabilirlik oranı ve Wald testleri sunulmaktadır. Olabilirlik oranı ve Wald testlerini ayrı ayrı hesaplayabilmek için sırasıyla “**mlogtest, lr**” ve “**mlogtest, wald**” komutları kullanılmaktadır. Tablo 4.7’de görüldüğü gibi hem olabilirlik oranı hem de Wald testleri için hesaplanan ki-kare değerleri %1 önem seviyesinde istatistiksel olarak anlamlı bulunmuştur. Böylece, “bağımsız değişkenler ile ilişkili tüm katsayıların sıfır olduğunu” öne süren H_0 hipotezi reddedilmektedir.

Tablo 4.7. Bağımsız Değişkenlere ait Olabilirlik Oranı ve Wald Testleri**mlogtest, lr**

LR tests for independent variables (N=11828)

Ho: All coefficients associated with given variable(s) are 0

	chi2	df	P>chi2
Tek_Kişi_Hane	54.443	3	0.000
Çekirdek_Aile	44.930	3	0.000
Oda_Sayısı	24.870	3	0.000
Orta_Konut	44.753	3	0.000
Büyük_Konut	24.239	3	0.000
Soba	9103.764	3	0.000
Merkezi_Sistem	1244.877	3	0.000
Diğer_Isıtma	2266.475	3	0.000
Oto_Sahipliği	49.612	3	0.000
Log_YıllıkGelir	148.642	3	0.000

mlogtest, wald

Wald tests for independent variables (N=11828)

Ho: All coefficients associated with given variable(s) are 0

	chi2	df	P>chi2
Tek_Kişi_Hane	53.650	3	0.000
Çekirdek_Aile	45.193	3	0.000
Oda_Sayısı	24.244	3	0.000
Orta_Konut	43.388	3	0.000
Büyük_Konut	25.297	3	0.000
Soba	2976.877	3	0.000
Merkezi_Sistem	1021.615	3	0.000
Diğer_Isıtma	618.917	3	0.000
Oto_Sahipliği	48.467	3	0.000
Log_YıllıkGelir	139.898	3	0.000

Tablo 4.8'de bağımlı değişken kategorilerinin birleştirilip birleştirilemeyeceğini öne süren olabilirlik oranı ve Wald testleri sunulmaktadır. Tablo 4.8'de görüldüğü gibi hem olabilirlik oranı hem de Wald testlerine ait 10 serbestlik derecesinde hesaplanan ki-kare değerleri %1 önem seviyesinde istatistiksel olarak anlamlı bulunmuştur. Böylece, herhangi bir alternatif çifti ile ilgili tüm katsayılarının sıfır olduğunu öne süren H_0 hipotezi reddedilerek tahmin edilen çok durumlu logit modeline ait bağımlı değişken çiftlerinin birleştirilemeyeceği öne sürülebilmektedir.

Tablo 4.8. Bağımlı Değişkenlere ait Olabilirlik Oranı ve Wald Testleri**mlogtest, wald**

Wald tests for combining alternatives (N=11828)

Ho: All coefficients except intercepts associated with a given pair of alternatives are 0 (i.e., alternatives can be combined)

	chi2	df	P>chi2
Kömür & Odun	142.964	10	0.000
Kömür & Doğalgaz	2981.432	10	0.000
Kömür & Diğer	570.999	10	0.000
Odun & Doğalgaz	1051.136	10	0.000
Odun & Diğer	231.508	10	0.000
Doğalgaz & Diğer	1105.599	10	0.000

mlogtest, lr

LR tests for combining alternatives (N=11828)

Ho: All coefficients except intercepts associated with a given pair of alternatives are 0 (i.e., alternatives can be collapsed)

	chi2	df	P>chi2
Kömür & Odun	484.214	10	0.000
Kömür & Doğalgaz	8541.436	10	0.000
Kömür & Diğer	1298.145	10	0.000
Odun & Doğalgaz	6598.361	10	0.000
Odun & Diğer	953.585	10	0.000
Doğalgaz & Diğer	3853.023	10	0.000

Tahmin edilen çok durumlu logit modelini çok durumlu probit sonuçları ile karşılaştırmak için “**est store mlogit**” komutu kullanılabilmektedir. Sözü edilen komut çalıştırıldığında Stata programı çok durumlu logit modeli tahmin sonuçlarını kaydetmektedir. Görelî risk oranı değerlerini hesaplamak için “**mlogit**” komutu ile birlikte “**rrr**” komutu kullanılmaktadır. Görelî risk oranı değerleri, bağımsız değişkenlerin bağımlı değişken üzerindeki etkisini kat olarak yansıtmaktadır. Görelî risk oranının 1’den küçük olması ilgili bağımsız değişken ile bağımlı değişken arasında negatif, 1’den büyük olması ise pozitif bir ilişki olduğunu işaret etmektedir. Görelî risk oranı 1’den küçük olduğunda, referans kategorinin ilgili bağımsız değişken kategorisi üzerindeki etkisi 1’in hesaplanan katsayıya bölünmesi ile hesaplanabilmektedir. Otomobil sahipliği, değişkenler tablosundan tesadüfî olarak seçilmiş ve hanehalkının kullandığı yakıt türü ile hanehalkının otomobil sahipliği kontenjans tablosu aşağıda verilmiştir. Tablo 4.9 incelendiğinde, Pearson ki-kare testi istatistiksel olarak anlamlıdır.

Tablo 4.9. Yakıt Türü ile Otomobil Sahipliği Kontenjans Tablosu**tab Yakıt_Türü Oto_Sahipliği**

1 =				
Kömür, 2		Otomobil sahipliği		
= Odun, 3				
=				
Doğalgaz,		0	1	
4 = Diğer		Hayır	Evet	Total
		-----+-----+-----		
1	Kömür	2,537	1,618	4,155
		37.76	31.66	35.13
		-----+-----+-----		
2	Odun	1,096	643	1,739
		16.31	12.58	14.70
		-----+-----+-----		
3	Doğalgaz	2,520	2,532	5,052
		37.51	49.55	42.71
		-----+-----+-----		
4	Diğer	565	317	882
		8.41	6.20	7.46
		-----+-----+-----		
Total		6,718	5,110	11,828
		100.00	100.00	100.00

Pearson chi2(3) = 175.6701 Pr = 0.000

Yakıt türü bağımlı değişkeni dört kategorili olup otomobil sahipliği (x) bağımsız değişkeni göstermektedir. 4 kategorili çok durumlu logit modeli örneğinde, belirli özelliklere sahip bir gözlemin kategorilerden her birine karşılık gelme olasılığının nasıl hesaplanacağı ile ilgili denklemler ve tahmin edilecek katsayılar $\beta^{(1)}$, $\beta^{(2)}$, $\beta^{(3)}$ ve $\beta^{(4)}$ 'dür.

$$P(\text{Kömür}) = \frac{e^{x\beta^{(1)}}}{e^{x\beta^{(1)}} + e^{x\beta^{(2)}} + e^{x\beta^{(3)}} + e^{x\beta^{(4)}}} \quad (4.24)$$

$$P(\text{Odun}) = \frac{e^{x\beta^{(2)}}}{e^{x\beta^{(1)}} + e^{x\beta^{(2)}} + e^{x\beta^{(3)}} + e^{x\beta^{(4)}}} \quad (4.25)$$

$$P(\text{Doğalgaz}) = \frac{e^{x\beta^{(3)}}}{e^{x\beta^{(1)}} + e^{x\beta^{(2)}} + e^{x\beta^{(3)}} + e^{x\beta^{(4)}}} \quad (4.26)$$

$$P(\text{Diğer}) = \frac{e^{x\beta^{(4)}}}{e^{x\beta^{(1)}} + e^{x\beta^{(2)}} + e^{x\beta^{(3)}} + e^{x\beta^{(4)}}} \quad (4.27)$$

Keyfi olarak dördüncü durum olan “diğer” seçeneği referans sınıf seçildiğinde kategorilerin olasılıklarındaki değişiklikler aşağıdaki gibi olmaktadır ($\beta^{(4)} = 0$). Olasılıkların toplamı bire eşit olduğundan referans kategoride bulunma olasılığı 1’den çıkarılarak da bulunur.

$$P(\text{Kömür}) = \frac{e^{x\beta^{(1)}}}{1 + e^{x\beta^{(1)}} + e^{x\beta^{(2)}} + e^{x\beta^{(3)}}} \quad (4.28)$$

$$P(\text{Odun}) = \frac{e^{x\beta^{(2)}}}{1 + e^{x\beta^{(1)}} + e^{x\beta^{(2)}} + e^{x\beta^{(3)}}} \quad (4.29)$$

$$P(\text{Doğalgaz}) = \frac{e^{x\beta^{(3)}}}{1 + e^{x\beta^{(1)}} + e^{x\beta^{(2)}} + e^{x\beta^{(3)}}} \quad (4.30)$$

$$Pr(\text{Diğer}) = \frac{1}{1 + e^{x\beta^{(1)}} + e^{x\beta^{(2)}} + e^{x\beta^{(3)}}} \quad (4.31)$$

Çok durumlu logit modelde göreceli risk oranı³ hesaplamasının kolay yapılabilmesi için sadece hanehalkının otomobile sahip olup olmadığını ifade eden kukla değişkenle çalışılmıştır. Çok durumlu logit modelin olabilirlik oranı istatistiğiyle kontenjans tablosunda bulunan Pearson ki-kare testinin sonuçları birbirleriyle uyumludur. Komutta, **base(4)** modelde “diğer” seçeneğinin referans alındığını göstermektedir.

Tablo 4.10. Örnek Çok Durumlu Logit Modeli

mlogit Yakıt_Türü Oto_Sahipliği, base(4)

Multinomial logistic regression	Number of obs	=	11,828
	LR chi2(3)	=	175.76
	Prob > chi2	=	0.0000
Log likelihood = -14180.208	Pseudo R2	=	0.0062

Yakıt_Türü	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
Kömür						
Oto_Sahipliği	.1281325	.07705	1.66	0.096	-.0228827	.2791477
_cons	1.501912	.0465197	32.29	0.000	1.410735	1.593089
Odun						
Oto_Sahipliği	.0446462	.0859773	0.52	0.604	-.1238663	.2131587
_cons	.6625967	.0517911	12.79	0.000	.561088	.7641055
Doğalgaz						
Oto_Sahipliği	.5826746	.0756059	7.71	0.000	.4344897	.7308594
_cons	1.495188	.0465482	32.12	0.000	1.403956	1.586421
Diğer	(base outcome)					

³ Eğer kategori sayısı iki olsaydı iki durumlu logit modeli tahmin edilecek ve fark oranları yorumlanacaktı. Fark oranlarının yorum ve hesaplamaları için 3. Bölüme bkznz.

Stata çıktısı görel risk oranlarını **rrr** seçeneğiyle otomatik olarak vermektedir. Ancak aşağıda olasılık hesabının nasıl yapıldığı gösterilecektir. Örneğin; otomobil sahibi olanların olmayanlara göre hanehalkının ısınmak amacıyla kömürü tercih etme olasılığının diğer kategorisindeki ısınma alternatifine göre olasılığı hesaplayabilmek için sırasıyla aşağıdaki olasılık hesaplamaları yapılmalıdır (aşağıda hesaplanan olasılık değerleri, kontenjan tablosunda frekansların altındaki yüzdelere de bulunmaktadır).

$$P(\text{Yakıt türü} = \text{Kömür} \mid \text{otomobil sahibi}) = \frac{e^{1,5019+0,1281}}{1 + e^{1,5019+0,1281} + e^{0,6625+0,04464} + e^{1,4951+0,5826}} \quad (4.32)$$

$$P(\text{Yakıt türü} = \text{Kömür} \mid \text{otomobil sahibi}) = 0,316 \quad (4.33)$$

$$P(\text{Yakıt türü} = \text{Diğer} \mid \text{otomobil sahibi}) = \frac{1}{1 + e^{1,5019+0,1281} + e^{0,6625+0,04464} + e^{1,4951+0,5826}} \quad (4.34)$$

$$P(\text{Yakıt türü} = \text{Diğer} \mid \text{otomobil sahibi}) = 0,062 \quad (4.35)$$

$$P(\text{Yakıt türü} = \text{Kömür} \mid \text{otomobil sahibi değil}) = \frac{e^{1,5019}}{1 + e^{1,5019} + e^{0,6625} + e^{1,4951}} \quad (4.36)$$

$$P(\text{Yakıt türü} = \text{Kömür} \mid \text{otomobil sahibi değil}) = 0,377 \quad (4.37)$$

$$P(\text{Yakıt türü} = \text{Diğer} \mid \text{otomobil sahibi değil}) = \frac{1}{1 + e^{1,5019} + e^{0,6625} + e^{1,4951}} \quad (4.38)$$

$$P(\text{Yakıt türü} = \text{Diğer} \mid \text{otomobil sahibi değil}) = 0,0841 \quad (4.39)$$

$$GRO = \frac{\frac{P(\text{Yakıt türü} = \text{kömür} \mid \text{otomobil sahibi})}{P(\text{Yakıt türü} = \text{diğer} \mid \text{otomobil sahibi})}}{\frac{P(\text{Yakıt türü} = \text{kömür} \mid \text{otomobil sahibi değil})}{P(\text{Yakıt türü} = \text{diğer} \mid \text{otomobil sahibi değil})}} = \frac{0,316}{0,062} = 1,136 \quad (4.40)$$

Otomobil sahibi olanların olmayanlara göre hanehalkının ısınmak amacıyla kömürü tercih etme olasılığı diğer kategorisindeki ısınma alternatifine göre 1,136 kat daha fazla olduğu yorumu yapılır. Tablo 4.11'deki Stata çıktısındaki sonuç yukarıdaki hesaplamadan gelmektedir. Benzer şekilde diğer olasılıklar da hesaplanabilir.

Tablo 4.11. Örnek Çok Durumlu Logit Modeli**mlogit Yakıt_Türü Oto_Sahipliği, rrr base(4)**

Multinomial logistic regression	Number of obs	=	11,828
	LR chi2(3)	=	175.76
	Prob > chi2	=	0.0000
Log likelihood = -14180.208	Pseudo R2	=	0.0062

Yakıt_Türü	RRR	Std. Err.	z	P> z	[95% Conf. Interval]	
Kömür						
Oto_Sahipliği	1.136704	.087583	1.66	0.096	.9773771	1.322003
_cons	4.490265	.2088856	32.29	0.000	4.098967	4.918918
Odun						
Oto_Sahipliği	1.045658	.0899029	0.52	0.604	.883498	1.237581
_cons	1.939823	.1004656	12.79	0.000	1.752578	2.147073
Doğalgaz						
Oto_Sahipliği	1.790822	.1353967	7.71	0.000	1.544175	2.076865
_cons	4.460177	.2076133	32.12	0.000	4.071273	4.886231
Diğer	(base outcome)					

Note: _cons estimates baseline relative risk for each outcome.

Tablo 4.12’de en çok olabilirlik yöntemi ile tüm değişkenler için tahmin edilen çok durumlu logit modeline ait görel risk oranları değerleri sunulmaktadır. Tablo 4.12’deki analiz sonuçlarına göre; odun bağımlı değişken kategorisi için tek kişilik hanehalkı değişkeninin katsayısı için hesaplanan görel risk oranı yaklaşık 1,61 bulunmuştur. Bu durumda, tek kişilik hanehalklarının ısınmak amacıyla odunu tercih etme olasılığının diğer hanehalklarına göre yaklaşık 2 kat (RRR: 1,61; $p < 0,01$) daha fazla olduğu söylenmektedir. Benzer şekilde, tek kişilik hanehalklarının doğalgaz ve diğer yakıt türlerini tercih etme olasılığının sırasıyla yaklaşık 3 kat (RRR: 3,23; $p < 0,01$) ve 2 kat (RRR: 1,71; $p < 0,01$) daha fazla olduğu tespit edilmiştir. Çekirdek ailelerinin diğer hanehalklarına göre ısınmak amacıyla odunu tercih etme olasılığının ise yaklaşık 1,21 kat (RRR: 1,21; $p < 0,05$) ve doğalgazı tercih etme olasılığının yaklaşık 2,21 kat (RRR: 2,21; $p < 0,01$) daha fazla olduğu tespit edilmiştir. Hanehalklarının ikamet ettikleri oda sayısının artması durumunda kömür tercih etme olasılığı diğer yakıt türlerini tercih etme olasılığına göre yaklaşık 1,39 kat (1/0,72; RRR: 0,72; $p < 0,01$) daha fazla olduğu görülmektedir.

Tablo 4.12. Çok Durumlu Logit Modeli Göreli Risk Oranı Değerleri

mlogit Yakıt_Türü Tek_Kişi_Hane Çekirdek_Aile Oda_Sayısı Orta_Konut Büyük_Konut Soba Merkezi_Sistem Diğer_Isıtma Oto_Sahipliği Log_YıllıkGelir, base(1) rrr

Iteration 0: log likelihood = -14268.087

Iteration 8: log likelihood = -7519.625

Multinomial logistic regression	Number of obs	=	11,828
	LR chi2(30)	=	13496.92
	Prob > chi2	=	0.0000
Log likelihood = -7519.625	Pseudo R2	=	0.4730

Yakıt_Türü	RRR	Std. Err.	z	P> z	[95% Conf. Interval]	
Kömür	(base outcome)					
Odu						
Tek_Kişi_Hane	1.607811	.200271	3.81	0.000	1.25953	2.052399
Çekirdek_Aile	1.209595	.0967663	2.38	0.017	1.034058	1.414931
Oda_Sayısı	.9390261	.040611	-1.45	0.146	.8627101	1.022093
Orta_Konut	.8765	.063341	-1.82	0.068	.7607451	1.009868
Büyük_Konut	1.013918	.1389914	0.10	0.920	.7750276	1.326442
Soba	7.500954	2.170389	6.96	0.000	4.254247	13.22545
Merkezi_Sistem	.078652	.0508323	-3.93	0.000	.0221607	.2791491
Diğer_Isıtma	2.554244	1.443811	1.66	0.097	.8435534	7.734141
Oto_Sahipliği	1.132954	.0744967	1.90	0.058	.9959599	1.288791
Log_YıllıkGelir	1.364946	.1819446	2.33	0.020	1.051119	1.77247
_cons	.0170754	.0117175	-5.93	0.000	.004449	.0655368
Doğalgaz						
Tek_Kişi_Hane	3.227933	.5960219	6.35	0.000	2.247783	4.63548
Çekirdek_Aile	2.21494	.2786494	6.32	0.000	1.73092	2.834307
Oda_Sayısı	.8936103	.0625055	-1.61	0.108	.7791284	1.024914
Orta_Konut	.5302579	.0600687	-5.60	0.000	.4246796	.6620838
Büyük_Konut	.6992782	.1226293	-2.04	0.041	.4958822	.9861012
Soba	.0018286	.0002249	-51.25	0.000	.0014369	.0023271
Merkezi_Sistem	.0398621	.004055	-31.68	0.000	.0326566	.0486576
Diğer_Isıtma	.0189613	.0063566	-11.83	0.000	.0098291	.0365785
Oto_Sahipliği	.5908321	.0540784	-5.75	0.000	.4938036	.7069258
Log_YıllıkGelir	8.002252	1.446468	11.51	0.000	5.615016	11.40443
_cons	.0018871	.0016072	-7.37	0.000	.0003555	.0100169
Diğer						
Tek_Kişi_Hane	1.707053	.3053635	2.99	0.003	1.202212	2.42389
Çekirdek_Aile	.9708407	.1166256	-0.25	0.805	.767175	1.228574
Oda_Sayısı	.7196105	.0496796	-4.77	0.000	.6285405	.8238757
Orta_Konut	1.299132	.1477864	2.30	0.021	1.039495	1.623619
Büyük_Konut	2.287219	.4407989	4.29	0.000	1.567696	3.336979

Soba		1.475397	.3756177	1.53	0.127	.8957842	2.430044
Merkezi_Sistem		.3936292	.1299895	-2.82	0.005	.2060593	.7519389
Diğer_Isıtma		132.7626	42.31048	15.34	0.000	71.08922	247.9405
Oto_Sahipliği		.7094763	.0744654	-3.27	0.001	.5775602	.8715223
Log_YıllıkGelir		2.231715	.4504124	3.98	0.000	1.50261	3.3146
_cons		.006731	.0065627	-5.13	0.000	.0009958	.0454987

Note: _cons estimates baseline relative risk for each outcome.

Analiz sonuçlarına göre, hanehalklarının ikamet ettiği konut büyüklüğünün 100 – 149 m² arasında olması durumunda ısınmak amacıyla odun ve doğalgaz tercih etme eğiliminin sırasıyla yaklaşık 0,88 kat (RRR: 0,877; $p < 0,10$) ve 0,53 kat (RRR: 0,530; $p < 0,01$) daha az ve diğer yakıt türlerini tercih etme eğiliminin ise yaklaşık 1,30 kat (RRR: 1,300; $p < 0,01$) daha fazla olduğu görülmektedir. Göreli olarak daha büyük (150 m²'den büyük) konutlarda ikamet eden hanehalklarının 100 m² ve daha küçük konutlarda ikamet eden hanehalklarına göre doğalgaz ve doğalgaz tercih olasılıklarının ise sırasıyla yaklaşık 0,70 kat (RRR: 0,699; $p < 0,05$) daha az ve yaklaşık 2,29 kat (RRR: 2,287; $p < 0,01$) daha fazla olduğu bulunmuştur.

Hanehalklarının ısıtma sistemi tipleri ile ilgili göreli risk oranı değerleri yorumlandığında, ikamet ettikleri konutlarda soba kullanan hanehalklarının ısınmak amacıyla odun tercih etme olasılığının yaklaşık 7,50 kat (RRR: 7,500; $p < 0,01$) daha fazla ve doğalgaz tercih etme olasılığının ise yaklaşık 0,002 kat (RRR: 0,002; $p < 0,01$) daha az olduğu görülmektedir. Odun kategorisi için bina içi merkezi ısıtma sistemine ait göreli risk oranı değerinin yaklaşık olarak 0,08 (RRR: 0,079; $p < 0,01$) olduğu görülmektedir. Böylece, referans kategorisi olan kat kaloriferi ısıtma sistemi kullanan hanehalklarının bina içi merkezi ısıtma sistemi kullanan hanehalklarına göre odun tercih etme olasılığının 12,5 kat (1/0,079) daha fazla olduğu tespit edilmektedir. Benzer şekilde, kat kaloriferi ısıtma sistemi kullanan hanehalklarının doğalgaz ve diğer yakıt türlerini tercih etme olasılığının sırasıyla 25 kat (1/0,040; RRR: 0,040; $p < 0,01$) ve yaklaşık 2,54 kat (1/0,394; RRR: 0,394; $p < 0,01$) daha fazla olduğu öne sürülebilmektedir. Diğer ısıtma sistemlerini tercih eden hanehalklarının ısınmak amacıyla odun tercih etme olasılığının yaklaşık 2,55 kat (RRR: 2,554; $p < 0,10$) ve diğer yakıt türlerini tercih etme olasılığının ise yaklaşık 133 kat (RRR: 132,763; $p < 0,01$) daha fazla olduğu görülmektedir. Kat kaloriferi sistemini kullanan hanehalklarının doğalgaz tercih etme olasılığı, diğer ısıtma sistemini kullanan hanehalklarına göre yaklaşık 53 kat (1/0,019; RRR: 0,019; $p < 0,01$) daha fazladır.

Otomobil sahibi olan hanehalklarının odun tercih etme olasılığının yaklaşık 1,13 kat (RRR: 1,133; $p < 0,10$) daha fazla olduğu görülmektedir. Otomobil sahibi olan hanehalklarının doğalgaz ve diğer yakıt türlerini tercih etme olasılığının ise sırasıyla 0,59 kat (RRR: 0,591; $p < 0,01$) ve yaklaşık 0,71 kat (RRR: 0,709; $p < 0,01$) daha az olduğu tespit edilmiştir. Yıllık hanehalkı kullanılabilir gelirin artması durumunda hanehalklarının ısınmak amacıyla odun tercih etme olasılığının yaklaşık 1,36 kat (RRR: 1,365; $p < 0,05$) daha fazla olduğu görülmektedir. Yıllık hanehalkı kullanılabilir gelirin artması durumunda hanehalklarının ısınmak amacıyla doğalgaz tercih etme olasılığı, kömür tercih etme olasılığına göre yaklaşık 8 kat (RRR = 8,00; $p < 0,01$) daha fazla olma eğilimindedir. Son olarak yıllık hanehalkı kullanılabilir gelirin artması durumunda hanehalklarının diğer yakıt türlerini tercih etme olasılığının kömür ve kömür türevlerine göre yaklaşık 2 kat (RRR: 2,232; $p < 0,01$) daha fazla olduğu söylenebilmektedir.

Tablo 4.13'te gösterildiği gibi çekirdek ailelerin diğer hanehalkı tiplerine göre kömür ve türevlerini tercih etme olasılıklarının yaklaşık %5 ($-0,050$; $p < 0,01$) daha az olduğu tespit edilmiştir. Öte yandan, çekirdek aileler diğer hanehalkı tiplerine göre ısınmak amacıyla odun ve doğalgazı sırasıyla yaklaşık %2 ($0,016$; $p < 0,05$) ve %4 ($0,041$; $p < 0,01$) daha fazla tercih etme eğilimindedir. Hanehalkının ikamet ettiği oda sayısı arttığında hanehalklarının ısınmak amacıyla kömür ve kömür türevlerini tercih etme olasılıklarının yaklaşık %2 ($0,019$; $p < 0,01$) daha fazla ve diğer yakıt türlerini tercih etme olasılıklarının ise yaklaşık %1 ($-0,013$; $p < 0,01$) daha az olduğu görülmektedir. Analiz sonuçlarına göre, hanehalklarının ikamet ettiği konut büyüklüğünün $100 - 149 \text{ m}^2$ arasında olması durumunda kömür ve kömür türevlerini tercih etme eğiliminin yaklaşık %3 ($0,031$; $p < 0,01$) ve diğer yakıt türlerini tercih etme eğiliminin ise yaklaşık %2 ($0,016$; $p < 0,01$) olduğu görülmektedir. Bu sonuçların aksine, görece olarak orta büyüklükte ($100 - 149 \text{ m}^2$) konutlarda ikamet eden hanehalklarının 100 m^2 ve daha küçük konutlarda ikamet eden hanehalklarına göre odun ve doğalgaz tercih olasılıklarının ise yaklaşık %1 ($-0,014$; $p < 0,10$) ve yaklaşık %3 ($-0,034$; $p < 0,01$) daha az olduğu bulunmuştur. 150 m^2 ve daha büyük konutlarda ikamet eden hanehalklarının görece olarak küçük konutlarda ikamet eden hanehalklarına göre doğalgazı tercih etme olasılığının yaklaşık %2 ($-0,023$; $p < 0,05$) daha az ve diğer yakıt türlerini tercih etme olasılığının ise yaklaşık %4 ($0,038$; $p < 0,01$) daha fazla olduğu tespit edilmiştir.

Çok durumlu logit tahmin sonuçlarına göre; ısıtma sistemi türü olarak soba kullanan hanehalkları, kat kaloriferi sistemi olan konutlarda ikamet eden hanehalklarına göre kömür ve kömür türevlerini yaklaşık %9 ($0,087$; $p < 0,01$), odunu yaklaşık %23 ($0,234$; $p < 0,01$) ve diğer yakıt türlerini ise yaklaşık %2 ($0,020$; $p < 0,05$) daha fazla tercih etme eğilimindedir. Isıtma sistemi türü olarak soba kullanan hanehalklarının kat kaloriferi sistemi kullanan hanehalklarına göre doğalgazı tercih etme olasılığı ise yaklaşık %34 ($-0,341$; $p < 0,01$) daha azdır. Bina içi merkezi kalorifer sistemi kullanan hanehalklarının kömür ve kömür türevlerini tercih etme olasılıklarının yaklaşık %38 ($0,382$; $p < 0,01$) daha fazla olduğu; odun ve doğalgazı tercih etme olasılıklarının ise yaklaşık %23 ($-0,235$; $p < 0,01$) ve yaklaşık %15 ($-0,152$; $p < 0,01$) daha az olduğu görülmektedir. Diğer ısıtma sistemlerini tercih eden hanehalklarının kat kaloriferi sistemi tercih eden hanehalklarına göre doğalgazı tercih etme eğiliminin yaklaşık %23 ($-0,234$; $p < 0,01$) daha az ve diğer yakıt türlerini tercih etme olasılıklarının ise yaklaşık %22 ($0,222$; $p < 0,01$) daha fazla olduğu tespit edilmiştir.

Otomobil sahibi olan hanehalklarının otomobil sahibi olmayan hanehalklarına göre ısınmak amacıyla kömür ve kömür türevleri ($0,021$; $p < 0,05$) ile odunu ($0,020$; $p < 0,01$) tercih etme olasılıklarının yaklaşık %2 daha fazla olduğu görülmektedir. Otomobil sahibi olan hanehalklarının sahip olmayanlara göre doğalgazı ve diğer yakıt türlerini tercih etme eğilimlerinin ise yaklaşık %3 ($-0,027$; $p < 0,01$) ve yaklaşık %1 ($-0,014$; $p < 0,01$) daha az olduğu söylenebilmektedir. Yıllık hanehalkı kullanılabilir gelirin artması durumunda hanehalklarının ısınmak amacıyla kömür ve kömür türevlerini tercih etme olasılıklarının yaklaşık %14 ($-0,138$; $p < 0,01$) daha az, doğalgazı ve diğer yakıt türlerini tercih etme olasılıklarının ise sırasıyla yaklaşık %10 ($0,104$; $p < 0,01$) ve yaklaşık %2 ($0,022$; $p < 0,05$) daha fazla olduğu tespit edilmiştir.

4.3. Çok Durumlu Probit Modeli

Çok durumlu probit modeli, hata terimlerinin birleşik çok değişkenli normal dağılıma sahip olduğu bir tesadüfi fayda modelidir. Söz konusu modelinin hata terimleri keyfi bir varyans/eşit varyans matrisine sahiptir ve ortalaması sıfırdır. Böylece, çok durumlu probit modelinin hata terimleri varyansları değişebilmekte ve hata terimleri ilişkili olabilmektedir. Bununla birlikte, çok durumlu probit modelinde tercih fonksiyonu kapalı bir yapıda kolaylıkla yazılamamakta ancak iki alternatif durumu için sayısal olarak değerlendirilebilmektedir. Tesadüfi fayda tercihleri alt yapısı kullanılarak iki durumlu bir tesadüfi fayda probit modeli,

$$U_1(\theta, a) = V_1(\theta, a) + \varepsilon_1(\theta, a), \quad (4.41)$$

$$U_2(\theta, a) = V_2(\theta, a) + \varepsilon_2(\theta, a) \quad (4.42)$$

biçiminde yazılmaktadır. Söz konusu eşitliklerde, a bir nitelik vektörünü temsil etmekte, $\varepsilon(\theta, a)$ hata terimleri vektörü, $(0, 0)$ ortalama, $\sum_{\varepsilon}(0, a)$ eşit varyans matrisine ve iki değişkenli normal bir dağılıma sahip olmaktadır. Alternatif 1'in tercih edilme olasılığı

$$P_1(\theta, a) = P\{V_1(\theta, a) + \varepsilon_1(\theta, a) > V_2(\theta, a) + \varepsilon_2(\theta, a)\} \quad (4.43)$$

$$P_1(\theta, a) = P\{\varepsilon_2(\theta, a) - \varepsilon_1(\theta, a) < V_1(\theta, a) - V_2(\theta, a)\} \quad (4.44)$$

biçiminde tanımlanmaktadır. Sıfır ortalama ve $\sigma^2(\theta, a)$ varyansa sahip $\varepsilon_2(\theta, a) - \varepsilon_1(\theta, a)$ ifadesi normal dağıldığı için söz konusu olasılık,

$$P_1(\theta, a) = \phi\left(\frac{V_1(\theta, a) - V_2(\theta, a)}{\sigma(\theta, a)}\right) \quad (4.45)$$

biçiminde de yazılabilmektedir ve $\phi(\cdot)$ standart birikimli normal dağılım fonksiyonunu temsil etmektedir. Eşitlik (4.45)'in her iki tarafına probit dönüşümü uygulanarak,

$$\phi^{-1}[P_1(\theta, a)] = V(\theta, a) \quad (4.46)$$

nihai eşitliği elde edilmektedir (Daganzo, 1979). $\sum_{\varepsilon}(0, a)$ toplamı içerisindeki eşit varyans terimlerinin söz konusu modele etkisini ortaya koymak için, farklı θ_1 ve θ_2 değerleri ile üçüncü alternatifin tercih edilme olasılığı,

$$V_1(\theta, a) = V_2(\theta, a) = V_3(\theta, a) = 0 \quad (4.47)$$

$$\sum_{\varepsilon}(\theta, a) = \begin{bmatrix} 1 & \theta_2 & 0 \\ \theta_2 & 1 & 0 \\ 0 & 0 & \theta_1 \end{bmatrix} \quad (4.48)$$

eklinde gösterilmektedir (Daganzo, Bouthelier ve Sheffi, 1977a, 1977b; Daganzo, 1979). Bir ana kütledeki bireylerin bağımsız hareket edeceği varsayımı ile çok durumlu model en çok olabilirlik yöntemi ile tahmin edildiğinde olabilirlik fonksiyonu,

$$L(\theta) = \prod_{n=1}^N P_{c(n)}(\theta, a_{(n)})F(a_{(n)}) \quad (4.49)$$

biçiminde yazılabilmektedir. Eşitlik (4.49)'da $a_{(n)}$ parametresi n 'inci gözlem için bir nitelik değişkenini, $c_{(n)}$ parametresi n 'inci gözlemin tercih edilmesini ve N ise veri setindeki gözlem sayısını temsil etmektedir. $F(a_{(n)})$ fonksiyonunun terimleri, θ değerinin bir fonksiyonu olmadığı için bu terimlerin değerleri en çok olabilirlik tahminini etkilememekte ve Eşitlik (4.49)'daki olabilirlik fonksiyonundan dışlanabilmektedir. Eşitlik (4.49)'un en yüksek değerine,

$$\log L(\theta) = \sum_{n=1}^N \log P_{c(n)}(\theta, a_{(n)}) \quad (4.50)$$

biçimindeki bir log-olabilirlik fonksiyonu ile ulaşılabilir (Daganzo, 1979).

4.3.1. Çok Durumlu Probit Modelinin Stata Uygulaması

Çok durumlu probit modeli, çok durumlu logit modelinin kısıtlayıcı ilişkisiz alternatiflerin bağımsızlığı varsayımını ihlal etmesi durumunda sıklıkla tercih edilen çok durumlu bir kesikli tercih modelidir. Tablo 4.14'te çok durumlu probit modeli en çok olabilirlik tahmin sonuçları sunulmaktadır. Stata paket programında çok durumlu probit modeli “**mprobit**” komutu ile tahmin edilmektedir. Çok durumlu probit modelinde faydalanan “**base**” ve “**level**” komutları, “**mprobit**” komutu ile birlikte kullanılabilir. Çok durumlu probit modeli, 11.828 gözlem ve 4 yineleme ile tahmin edilmiştir. Tahmin edilen çok durumlu probit modeli, %1 önem seviyesinde istatistiksel olarak anlamlı (Wald ki-kare: 6.512,05; $p = 0.0000$) bulunmuştur. Çok durumlu logit modeli tahmininde olduğu çok durumlu probit modeli de her bir bağımlı değişken kategorisi için ayrı ayrı tahmin katsayıları hesaplamaktadır.

Tablo 4.14. Çok Durumlu Probit Modelinin Tahmin Sonuçları

mprobit Yakıt_Türü Tek_Kişi_Hane Çekirdek_Aile Oda_Sayısı Orta_Konut Büyük_Konut Soba Merkezi_Sistem Diğer_Isıtma Oto_Sahipliği Log_YıllıkGelir, base(1)

```
Iteration 0:  log likelihood = -7598.7652
Iteration 1:  log likelihood = -7529.6785
Iteration 2:  log likelihood = -7521.0046
Iteration 3:  log likelihood = -7520.9436
Iteration 4:  log likelihood = -7520.9436
```

```
Multinomial probit regression      Number of obs      =      11,828
                                Wald chi2(30)           =      6512.05
Log likelihood = -7520.9436        Prob > chi2          =      0.0000
```

Yakıt_Türü	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
Kömür	(base outcome)					
Odun						
Tek_Kişi_Hane	.4007723	.100354	3.99	0.000	.2040821	.5974625
Çekirdek_Aile	.1701974	.0638744	2.66	0.008	.0450058	.295389

Oda_Sayısı	-.0505032	.034342	-1.47	0.141	-.1178123	.016806
Orta_Konut	-.1227165	.0580432	-2.11	0.034	-.236479	-.008954
Büyük_Konut	-.008451	.1092014	-0.08	0.938	-.2224818	.2055798
Soba	.775427	.1305923	5.94	0.000	.5194707	1.031383
Merkezi_Sistem	-1.672183	.2707227	-6.18	0.000	-2.20279	-1.141577
Diğer_Isıtma	.3768196	.2752064	1.37	0.171	-.162575	.9162142
Oto_Sahipliği	.0868569	.052977	1.64	0.101	-.0169762	.19069
Log_YıllıkGelir	.2789875	.1065245	2.62	0.009	.0702033	.4877717
_cons	-2.584211	.5148427	-5.02	0.000	-3.593284	-1.575138

Doğalgaz						
Tek_Kişi_Hane	.8210887	.1229012	6.68	0.000	.5802068	1.061971
Çekirdek_Aile	.4986339	.0818561	6.09	0.000	.3381988	.659069
Oda_Sayısı	-.1066429	.0452577	-2.36	0.018	-.1953463	-.0179395
Orta_Konut	-.3756462	.0729919	-5.15	0.000	-.5187076	-.2325848
Büyük_Konut	-.2073156	.1187681	-1.75	0.081	-.4400968	.0254656
Soba	-4.534972	.072688	-62.39	0.000	-4.677437	-4.392506
Merkezi_Sistem	-2.46696	.0737384	-33.46	0.000	-2.611485	-2.322436
Diğer_Isıtma	-2.92116	.1936022	-15.09	0.000	-3.300613	-2.541707
Oto_Sahipliği	-.3690458	.061605	-5.99	0.000	-.4897894	-.2483022
Log_YıllıkGelir	1.443311	.1221166	11.82	0.000	1.203966	1.682655
_cons	-4.00002	.5765674	-6.94	0.000	-5.130071	-2.869969

Diğer						
Tek_Kişi_Hane	.4252549	.1167009	3.64	0.000	.1965253	.6539846
Çekirdek_Aile	.0388038	.077781	0.50	0.618	-.1136441	.1912517
Oda_Sayısı	-.2156847	.0447687	-4.82	0.000	-.3034298	-.1279396
Orta_Konut	.1317241	.0733266	1.80	0.072	-.0119935	.2754417
Büyük_Konut	.5032332	.125335	4.02	0.000	.2575811	.7488853
Soba	-.265817	.1194269	-2.23	0.026	-.4998894	-.0317446
Merkezi_Sistem	-.9502508	.1596629	-5.95	0.000	-1.263184	-.6373172
Diğer_Isıtma	3.143828	.1739758	18.07	0.000	2.802842	3.484815
Oto_Sahipliği	-.223667	.0662024	-3.38	0.001	-.3534213	-.0939127
Log_YıllıkGelir	.6175689	.126782	4.87	0.000	.3690807	.866057
_cons	-3.357223	.6034775	-5.56	0.000	-4.540017	-2.174429

Tablo 4.14'te gösterilen çok durumlu probit tahmin sonuçlarını kaydetmek için “**est store mprobit**” komutu kullanılmaktadır. Tahmin edilen çok durumlu probit modeline ait özet istatistik değerlerini elde etmek için ise model tahmin edildikten sonra “**fitstat**” komutu kullanılmaktadır. Tablo 4.15'te görüldüğü gibi çok durumlu probit modeline ait düzeltilmiş Pseudo- R^2 değeri 0.547 olarak bulunmuş ve bu değer oldukça iyi bir uyum iyiliğini işaret etmektedir (Louviere ve diğ., 2000).

Tablo 4.15. Çok Durumlu Probit Modeline ait Özet İstatistikler

fitstat		
		mprobit
Log-likelihood	Model	-7520.944
Chi-square		
Deviance(df=11795)		15041.887
Wald(df=30)		6512.047
p-value		0.000
R2		
	Count	0.740
	Count(adjusted)	0.547
IC		
	AIC	15107.887
	AIC divided by N	1.277
	BIC(df=33)	15351.369

Tablo 4.16'da tahmin edilen çok durumlu probit modeline ait marjinal etki değerleri her bir bağımlı değişken kategorisi için ayrı ayrı sunulmaktadır. Çok durumlu probit modeli marjinal etki değeri sonuçlarının çok durumlu logit modeline yakın olduğu görülmektedir. Çok durumlu probit modeli tahmin sonuçları, çok durumlu logit modelinde yapılan yorumlara benzer şekilde değerlendirilebilmektedir.

Tablo 4.16. Çok Durumlu Probit Modeline ait Marjinal Etki Değerleri

margins, dydx(*)

Average marginal effects	Number of obs	=	11,828
Model VCE : OIM			

```

dy/dx w.r.t. : Tek_Kişi_Hane Çekirdek_Aile Oda_Sayısı Orta_Konut Büyük_Konut Soba
Merkezi_Sistem Diğer_Isıtma Oto_Sahipliği Log_YıllıkGelir
1._predict : Pr(Yakıt_Türü==Kömür), predict(pr outcome(1))
2._predict : Pr(Yakıt_Türü==Odun), predict(pr outcome(2))
3._predict : Pr(Yakıt_Türü==Doğalgaz), predict(pr outcome(3))
4._predict : Pr(Yakıt_Türü==Diğer), predict(pr outcome(4))

```

	Delta-method					
	dy/dx	Std. Err.	z	P> z	[95% Conf. Interval]	
Tek_Kişi_Hane						
_predict						
1	-.1047775	.0155698	-6.73	0.000	-.1352938	-.0742613
2	.0352266	.0124267	2.83	0.005	.0108708	.0595824

Bu bölümde kullanılan Stata komutları aşağıda toplu olarak yer almaktadır.

```
. use "C:\Users\Hp\Desktop\mlogit.dta", clear

. describe

. sum

. estat vif

. mlogit Yakıt_Türü Tek_Kişi_Hane Çekirdek_Aile Oda_Sayısı Orta_Konut Büyük_Konut
Soba_Merkezi_Sistem Diğer_Isıtma Oto_Sahipliği Log_YıllıkGelir, base(1)

. mlogtest, hausman

. mlogtest, smhsiao

. mlogtest, iia

. mlogtest, all

. mlogit Yakıt_Türü Tek_Kişi_Hane Çekirdek_Aile Oda_Sayısı Orta_Konut Büyük_Konut
Soba_Merkezi_Sistem Diğer_Isıtma Oto_Sahipliği Log_YıllıkGelir, base(1) rrr

. margins, dydx(*)

. mlogit Yakıt_Türü Tek_Kişi_Hane Çekirdek_Aile Oda_Sayısı Orta_Konut Büyük_Konut
Soba_Merkezi_Sistem Diğer_Isıtma Oto_Sahipliği Log_YıllıkGelir, base(1)

. est store mlogit

. est store mprobit

. lrtest mlogit mprobit, stats force
```

4.4. KAYNAKLAR

- Cheng, S., & Long, J. S. (2007). Testing for IIA in the multinomial logit model. *Sociological Methods & Research*, 35(4), 583-600.
- Çelik, A. (2016). Sıralı ve sıralı olmayan kesikli tercih modellerinin karşılaştırılması: Türkiye’de hanehalklarının yakıt türü tercihleri üzerine bir uygulama. Doktora Tezi, Atatürk Üniversitesi, Sosyal Bilimler Enstitüsü, Erzurum.
- Daganzo, C. (1979). *Multinomial Probit: The Theory and its Application to Demand Forecasting*. New York: Academic Press.
- Daganzo, C., Bouthelie, F. & Sheffi, Y. (1977b). Multinomial probit and qualitative choice: A computationally efficient algorithm. *Transportation Science*, 11(4), 338-358.
- Gujarati, D. (2011). *Basic Econometrics*. New York: McGraw-Hill.
- Hausman, J. A. (1978). Specification tests in econometrics. *Econometrica*, 46(6), 1251-1271.
- Hosmer, D. W. & Lemeshow, S. (2000). *Applied Logistic Regression*. New York: Wiley.

- Long, J. S., & Freese, J. (2001). *Regression Models for Categorical Dependent Variables using Stata*. College Station, Texas: Stata Press.
- Louviere, J. J., Hensher, D. A., & Swait, J. D. (2000). *Stated Choice Methods: Analysis and Applications*. Cambridge: Cambridge University Press.
- Luce, R. D. (1959). *Individual Choice Behavior: A Theoretical Analysis*. New York: John Wiley & Sons, Inc.
- McFadden, D. (1974). "Conditional Logit Analysis of Qualitative Choice Behavior". P. Zarembka (Ed.). *Frontiers in Econometrics* (ss. 105-142). New York: Academic Press.
- Powers, D. A. & Xie, Y. (2000). *Statistical Methods for Categorical Data Analysis*. New York: Academic Press.
- Train, K. E. (2003). *Discrete Choice Methods with Simulation*. Cambridge: Cambridge University Press.
- Türkiye İstatistik Kurumu. (2018a). *2018 Hanehalkı Bütçe Araştırması Meta Veri Seti*. Ankara: Türkiye İstatistik Kurumu.
- Türkiye İstatistik Kurumu. (2018b). *Hanehalkı Bütçe Araştırması 2018, Hanehalkı Bütçe Anketterinin Uygulama Yöntemi, Tanım ve Kavramlar*. Ankara: Türkiye İstatistik Kurumu.
- Small, K. A. & Hsiao, C. (1985). Multinomial logit specification tests. *International Economic Review*, 26(3), 619-627.
- Üçdoğruk, Ş. (2002). *Multinomial Logit Model: Bölgelerarası İşgücü Tercihleri Analizi*. İzmir: Dokuz Eylül Üniversitesi Yayınları.
- Washington, S., Karlaftis, M., & Mannering, F. (2011). *Statistical and Econometric Methods for Transportation Data Analysis*. Boca Raton, FL: CRC Press.

Tablo 4.2 Değişkenlerin Stata Paket Programında Özetlenmesi
describe

Contains data from C:\Users\Hp\Desktop\mlogit.dta

obs:	11,828			
vars:	14			12 Sep 2020 21:35
variable name	storage type	display format	value label	variable label
Yakit_Turu	float	%29.0g	yakit_turu	1 = Kömür, 2 = Odun, 3 = Doğalgaz, 4 = Diğer
Tek_Kişi_Hane	byte	%23.0g	hhtip1	1 = Tek kişilik hanehalkı, 0 = Değilse
Çekirdek_Aile	byte	%39.0g	hhtip2	1 = Çekirdek aile, 0 = Değilse
Diğer_Hane_Tipi	byte	%45.0g	hhtip3	1 = Diğer hanehalkı tipleri, 0 = Değilse
Oda_Sayısı	byte	%8.0g		Konutun oda sayısı
Küçük_Konut	byte	%19.0g	alan1	1 = 100 m2'den küçük, 0 = Değilse
Orta_Konut	byte	%10.0g	alan2	1 = 100 - 149 m2, 0 = Değilse
Büyük_Konut	byte	%22.0g	alan3	1 = 150 m2 ve daha büyük, 0 = Değilse
Kat_Kaloriferi	byte	%34.0g	isitma1	1 = Kat kaloriferi sistemi, 0 = Değilse
Soba	byte	%51.0g	isitma2	1 = Soba, 0 = Değilse
Merkezi_Sistem	byte	%35.0g	isitma3	1 = Merkezi kalorifer sistemi, 0 = Değilse
Diğer_Isıtma	byte	%26.0g	isitma4	1 = Diğer ısıtma sistemleri, 0 = Değilse
Oto_Sahipliği	byte	%9.0g	oto_sah	1 = Evet, 0 = Hayır
Log_YıllıkGelir	float	%8.0g		Hanehalkı yıllık kullanılabilir gelir

BÖLÜM 5

SIRALI REGRESYON MODELLERİ VE STATA UYGULAMALARI

Özlem Kiren Gürler¹

Üzerinde araştırma yapılan birçok sosyal olay sıralı kategorik değişkenlerle ölçülebilmektedir. Genellikle, sıralı kategorik değişkenler belirli bir niteliğin sırasını belirtmek için kullanılır. Gelir düzeylerini düşük, orta, yüksek şeklinde oluşturmak ya da eğitim seviyelerini ilköğretim, orta öğretim, lise, üniversite ve yüksek lisans dereceleri şeklinde sıralamak gibi örnekler verilebilir (Long, 1997: 137; Bender ve Grouven, 1997: 546). Yine, anket araştırmasında belirtilen görüşler kesinlikle katılıyorum ve katılıyorum arasındaki uzaklık ile katılıyorum ve katılmıyorum arasındaki uzaklığa bakılmadan da bir sıralama yapılabilir. (Powers ve Xie, 1999: 201, Katchova, 2013). Genel bir ifadeyle, sıra değerlerinin eşit aralıklı olması istenmekle birlikte kategorilerin temsil ettiği değerlerin aralıkları eşit olmayabilir. Ayrıca, sıra değerleri, kategoriler arasındaki sayısal üstünlüğün derecesini de göstermez. Sonuçları sıralı ele almak, araştırmanın hedefine bağlı olarak yapılması gereken bir seçimdir.

Sıralı değişken kategorileri genellikle 1'den başlayıp ardışık tamsayı değerleriyle kodlanmaktadır. Bu kodlama nedeniyle sıralı sonuçları doğrusal regresyon modeli ile analiz etmenin uygun olduğunu, ifade eden çalışmalar olmakla birlikte McKelvey ve Zavonia (1975) ve Winship ve Mare (1984) yaptıkları çalışmalarında, doğrusal regresyon model varsayımının ihlal edildiğini ve tahmin sonuçlarının yanlış olduğunu göstermişlerdir. Standart logit ya da probit modeli sadece iki (evet/hayır) durumlu değişkenler için geçerli iken, ikiden fazla kategori içeren değişkenlerin analizi için karmaşık yöntemler gerekmektedir. Sıralı kategorilere sahip bir cevap değişkeni için en uygun analiz yöntemi sıralı regresyon (SRM) modelidir (Long ve Freese, 2014: 309; Bender ve Grouven, 1997: 546-547). Bu kısımda sıralı regresyon modellerinden bahsedilecektir.

5.1. Sıralı Değişkenler İçin Gizli Değişken Modeli

Sıralı regresyon modelleri (SRM) çoğunlukla gizli değişkenli ifade edilmektedir. y^* , gizli değişkeni $-\infty, +\infty$ aralığında değişen değerler almaktadır. Sıralı regresyonda yapısal model eşitlik (5.1)'deki gibi verilebilir.

¹ Dr.Öğr.Üyesi, Dokuz Eylül Üniversitesi, İİBF, Ekonometri Bölümü, ozlem.kiren@deu.edu.tr

$$y_i^* = x_i\beta + \varepsilon_i \quad (5.1)$$

ε hata terimi ve $i = 1, \dots, N$ dir.

Tek bağımsız değişken durumunda eşitlik (5.2)'deki gibi ifade edilebilir:

$$y_i^* = \alpha + \beta X_i + \varepsilon_i \quad (5.2)$$

İki durumlu değişkenler için gerçekleştirilen modeller, y^* için j sıralı kategorilere bölünerek;

$$y_i^* = m \quad \tau_{m-1} \leq y_i^* \leq \tau_m \quad m = 1, \dots, j \text{ için} \quad (5.3)$$

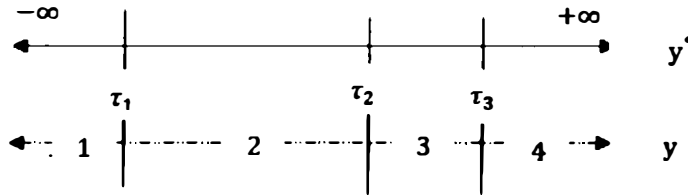
elde edilir.

Eşitlik (5.3)'teki τ 'lar “eşik değerleri” veya “kesim noktaları” olarak adlandırılır. Bir başka ifadeyle, eşik parametreleri değişik kategorilerin sınırlarını işaret eder. Eğer bir tahvil değerlendirme derece değerleri AAA, AA, A, B şeklinde büyükten küçüğe doğru sıralanıyor ise; bir tahvil B ile değerlendirildiğinde A ile derecelendirilen bir tahvilden daha düşük bir kategoride yer alıyor anlamına gelmektedir. Benzer şekilde A kategorisi de tahvilde AA ile derecelendirilen kategorinin altında yer alacak ve bu şekilde devam edecektir (Hill vd. 2017: 709; Gujarati, 2011: 170). Uç kategoriler (1 ve J) için eşik değerleri açık uçlu olarak tanımlanır. $\tau_0 = -\infty$ ve $\tau_J = +\infty$ 'dur. Sıralı değişkene örnek olarak “bireylere hangi sosyal gruba ait oldukları” sorulduğunda elde edilen cevap değişkenine ait sonuçlar sürekli gizli değişken için

$$y_i = \begin{cases} 1 \Rightarrow \text{En alt grup} & \tau_0 = -\infty \leq y_i^* < \tau_1 \\ 2 \Rightarrow \text{Çalışan Grup} & \tau_1 \leq y_i^* < \tau_2 \\ 3 \Rightarrow \text{Orta grup} & \tau_2 \leq y_i^* < \tau_3 \\ 4 \Rightarrow \text{En Üst Grup} & \tau_3 \leq y_i^* < \tau_4 = \infty \end{cases} \quad (5.4)$$

ifade edilir. Sonuç olarak gizli değişken y_i^* bir kesme noktasına geçtiğinde gözlemlenen kategoride değişmektedir (Greene, 2007: 832; Long ve Freese, 2014: 310-311). Gizli değişken ile gözlemlenen kategoriler arasındaki ilişki Şekil 5.1'de gösterilmektedir.

Şekil 5.1. Gizli Değişkenden Gözlemlenmiş Kategorilerin Elde Edilişi

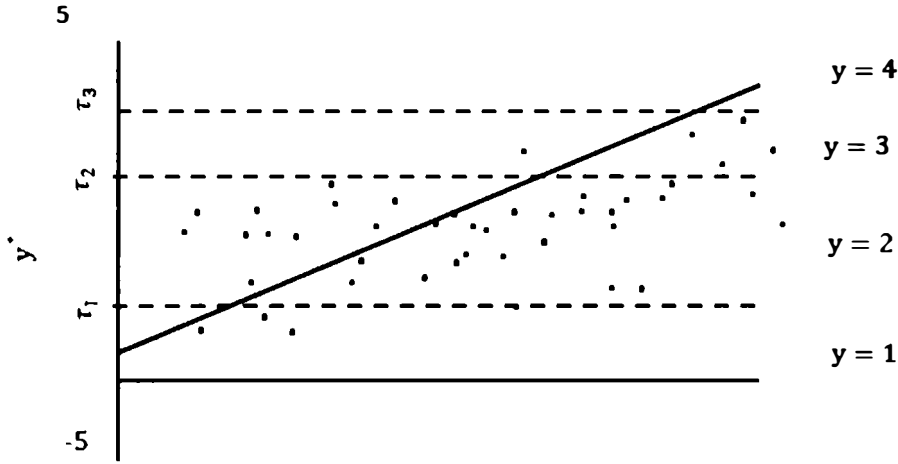


Kaynak: Long, 1997: 117

Şekil 5.1'de düz çizgi gizli değişken y^* 'ı temsil etmektedir. Kesim noktaları (τ_1, τ_2, τ_3) noktali yatay çizgilerle belirtilir. Gözlemlenen değişken y 'nin y^* aralığındaki değerleri noktali çizgiler ile belirtilmiştir (Long, 1997: 116-117).

Tek bağımsız değişkenli yapısal model $y_i^* = \alpha + \beta X_i + \varepsilon_i$ Şekil 5.2’de kesim noktaları ile birlikte gösterilmiştir. Şekil 5.2 iki durumlu regresyona benzemektedir, tek farkı τ_1, τ_2 ve τ_3 kesme noktalarını temsil eden üç yatay çizginin olmasıdır. Eşik değerleri bağımlı değişken kategorilerinin olasılık değerlerinin belirlenmesi için kullanılmaktadır ve gözlemlenen y değerlerinin dörde bölünmesi ile y^* elde edilmektedir.

Şekil 5.2. Gizli değişken y^* ’in Regresyonu



Kaynak: (Long, 1997: 118)

Örneğin x ’in verilen değerleri için $y = m$ gözlemlene olasılığı, y^* ’in τ_{m-1} ve τ_m arasına düştüğü dağılım bölgesine karşılık gelmektedir (Long ve Freese, 2014:312).

$$Pr(y = m | x) = Pr(\tau_{m-1} \leq y_i^* \leq \tau_m | x)$$

ifadesinde y^* yerine $(x\beta + \varepsilon)$ yazıldığında SRM’de tahmin edilen olasılık için standart formül eşitlik (5.5)’deki gibi elde edilebilir. Burada F, ε ’nın birikimli dağılım fonksiyonudur;

$$Pr(y = m | x) = F(\tau_m - x\beta) - F(\tau_{m-1} - x\beta) \quad (5.5)$$

Sıralı probit modelde $F; Var(\varepsilon) = 1$ ile normal dağılım, sıralı lojistik regresyonda ise $Var(\varepsilon) = \frac{\pi^2}{3}$ ile lojistik dağılım göstermektedir. $y = 1$ için $F(-\infty - x\beta) = 0$ olduğundan sağ taraftaki ikinci terim düşmekte ve $y = J$ için ilk terim $F(\infty - x\beta) = 1$ olmaktadır (Fullerton ve Xu, 2016: 22; Long ve Freese, 2014: 312).

y^* ’ın x üzerindeki regresyonunu tahmin etmek için en yüksek olabilirlik yöntemi kullanılmaktadır. En yüksek olabilirlik yönteminin uygulanabilmesi için, hata terimlerinin dağılımının belli olması gerekmektedir. Sıralı regresyon modellerinde çoğunlukla hata terimlerinin normal veya lojistik dağılımlı olduğu varsayılmaktadır. Hata teriminin dağılımına ait diğer varsayımlar cauchy, weibull, poisson ve binom dağılımlı olabilmeleridir, fakat çoğunlukla bu dağılımlar kullanılmamaktadır (McCullagh, 1980: 115; Long, 1997: 119).

Sıralı probit model için hata terimi ε , 0 ortalama ve 1 varyans ile normal dağılımlıdır. Olasılık yoğunluk fonksiyonu ise eşitlik (5.6)’daki gibidir.

$$\phi(\epsilon) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\epsilon^2}{2}\right) \quad (5.6)$$

Birikimli dağılım fonksiyonu eşitlik (5.7)'deki gibidir.

$$\Phi(\epsilon) = \int_{-\infty}^{\epsilon} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) dt \quad (5.7)$$

Sıralı logit model için hata terimi ϵ , 0 ortalama ve $\pi^2/3$ varyans ile lojistik dağılım göstermektedir. Olasılık yoğunluk fonksiyonu,

$$\lambda(\epsilon) = \frac{\exp(\epsilon)}{[1 + \exp(\epsilon)]^2} \quad (5.8)$$

Birikimli dağılım fonksiyonu,

$$\Lambda(\epsilon) = \frac{\exp(\epsilon)}{1 + \exp(\epsilon)} \quad (5.9)$$

Eşitlik (5.9)'daki gibidir (Long, 1997: 119; Borooah, 2001: 9).

Hata terimlerinin dağılımı belirlendikten sonra verilen x değişkenine bağlı olarak y değişkeninin gözlenen değerlerinin olasılıkları hesaplanabilir. Hata terimleri $E(y^* | x) = \alpha + \beta x$ regresyon doğrusu etrafında lojistik veya normal dağılım göstermektedir. m sonuç kategorisinin olasılığı; hata teriminin dağılımının τ_{m-1} ve τ_m eşik noktaları arasındaki alanına karşılık gelmektedir ve bu alan aşağıdaki gibi hesaplanmaktadır:

Kategorik bağımlı değişkenin birinci kategorisi $y = 1$ için olasılık hesaplanmak istendiğinde; y^* gizli değişkeni $\tau_0 = -\infty$ ve τ_1 arasında olacaktır. $y = 1$ 'in olasılığı için eşitlik (5.10)'daki gibidir.

$$Pr(y_i = 1 | x_i) = Pr(\tau_0 \leq y_i^* < \tau_1 | x_i) \quad (5.10)$$

Eşitlik (5.10)'da $y^* = x\beta + \epsilon_i$ y_i^* yerine koyulursa,

$$Pr(y_i = 1 | x_i) = Pr(\tau_0 \leq x_i\beta + \epsilon_i < \tau_1 | x_i) \quad (5.11)$$

ve (5.11) nolu eşitsizlikten $x\beta$ çıkarılırsa,

$$Pr(y_i = 1 | x_i) = Pr(\tau_0 - x_i\beta \leq \epsilon_i < \tau_1 - x_i\beta | x_i) \quad (5.12)$$

eşitlik (5.12) elde edilir. Rastgele bir değişkenin seçilen iki değer arasında olma olasılığı, bu iki değere karşılık gelen birikimli dağılım fonksiyonundan elde edilen sonuçlar arasındaki farktır (Greene ve Hensher, 2009: 93; Long, 1997: 121).

$y = 1$ için eşitlik (5.13)'te verilmiştir.

$$\begin{aligned} Pr(y_i = 1 | x_i) &= Pr(\epsilon_i < \tau_1 - x_i\beta | x_i) - Pr(\epsilon_i \leq \tau_0 - x_i\beta | x_i) \\ &= F(\tau_1 - x_i\beta) - F(\tau_0 - x_i\beta) \end{aligned} \quad (5.13)$$

Eşitlik (5.13)'teki işlemler, gözlenen her bir kategorinin verilen x değeri için $y = m$ olasılığını hesaplamak için genelleştirilebilir:

$$Pr(y_i = m | x_i) = F(\tau_m - x_i\beta) - F(\tau_{m-1} - x_i\beta) \quad (5.14)$$

$Pr(y_i = 1 | x_i)$ hesaplanırken ikinci terim olan $F(\tau_0 - x_i\beta) = F(-\infty - x_i\beta) = 0$ ve $Pr(y_i = j | x_i)$ hesaplanırken ilk terim $F(\tau_j - x_i\beta) = F(\infty - x_i\beta) = 1$ olacaktır.

Bu nedenle dört kategoriye sahip bir değişken için olasılık hesaplamaları aşağıda belirtilen biçimde gösterilebilir (Long, 1997: 120-121).

$$Pr(y_i = 1 | x_i) = \Phi(\tau_1 - x_i\beta) \quad (5.15)$$

$$Pr(y_i = 2 | x_i) = \Phi(\tau_2 - x_i\beta) - \Phi(\tau_1 - x_i\beta) \quad (5.16)$$

$$Pr(y_i = 3 | x_i) = \Phi(\tau_3 - x_i\beta) - \Phi(\tau_2 - x_i\beta) \quad (5.17)$$

$$Pr(y_i = 4 | x_i) = 1 - \Phi(\tau_3 - x_i\beta) \quad (5.18)$$

Eşitlik (5.15)-(5.18) ile belirtilen olasılıklı model “oranların fark oranı modeli” olarak da bilinmektedir. Çünkü $y = m$ kategorisi için fark oranı

$$OR(m) = \frac{Pr(Y_i \leq m)}{Pr(Y_i > m)} \quad (5.19)$$

olmaktadır. Fark oranının tüm kategoriler için sabit olduğu varsayılmaktadır (Long ve Freese: 2014: 335; Borooah, 2001: 10).

5.2. Sıralı Modelin Tahmini

Sıralı modellerin tahmininde en yüksek olabilirlik yöntemi kullanılmaktadır. Yapısal modelin parametre vektörü β ; sabit terimi β_0 ilk satırında yer alan ve eşik parametre vektörü τ değerlerini içeren bir vektör olsun. Model tanımlaması için β_0 ve τ_1 sıfır alınarak (5.14) nolu eşitlikten

$$Pr(y_i = m | x_i, \beta, \tau) = F(\tau_m - x_i\beta) - F(\tau_{m-1} - x_i\beta) \quad (5.20)$$

olasılık değeri (5.20) gibi ifade edilebilir. Modeli tanımlayabilmek için, β_0 veya τ_1 sıfıra eşit olarak alınmaktadır. y 'nin değeri ne olursa olsun i 'inci gözlem için gözlemleme olasılığı

$$p_i = \begin{cases} Pr(y_i = 1 | x_i, \beta, \tau) & \text{Eğer } y = 1 \\ \vdots & \vdots \\ Pr(y_i = m | x_i, \beta, \tau) & \text{Eğer } y = m \end{cases}$$

bulunur. Eğer gözlemler bağımsız ise olabilirlik fonksiyonu

$$L(\beta, \tau | y, X) = \prod_{j=1}^N p_i \quad (5.21)$$

(5.21)'deki gibidir. Eşitlik (5.20) ve eşitlik (5.21) birleştirildiğinde eşitlik (5.22) elde edilir:

$$\begin{aligned} L(\beta, \tau / Y, X) &= \prod_{j=1}^J \prod_{Y_i=j} P(Y_i = j / X_i, \beta, \tau) \\ &= \prod_{j=1}^J \prod_{Y_i=j} [F(\tau_j - X_i\beta) - F(\tau_{j-1} - X_i\beta)] \end{aligned} \quad (5.22)$$

Logaritmik olabilirlik fonksiyonu ise (5.23)'teki gibidir:

$$\ln L(\beta, \tau / Y, X) = \sum_{j=1}^J \sum_{Y_i=j} \ln [F(\tau_j - X_i\beta) - F(\tau_{j-1} - X_i\beta)] \quad (5.23)$$

Eşitlik (5.23)'te verilen fonksiyonun maksimize edilmesi ile β ve τ parametreleri tahmin edilir. Elde edilen tahminciler tutarlı, asimptotik normal ve asimptotik etkin tahmin edicilerdir (Power ve Xie, 1999: 216; Long, 1997: 124).

5.2.1. Sıralı Logit Model Tahmini

İki durumlu standart regresyon modellerinde Y ve X arasındaki ilişki incelendiğinde eğer Y bağımlı değişkeni sürekli bir değişken değil, iki durumlu (yani 0 ve 1 değerlerini alabilen) bir değişken ise basit en küçük kareler yöntemi geçersiz olacaktır (Bender ve Grouven; 1997- 546). X ve Y arasındaki ilişkiyi tanımlamak için olayların olasılık değerlerini $p = P(Y = 1)$ dikkate almak gerekmektedir. Y, 0 ve 1 iki değeri alırken, olasılık değerleri de 0 ve 1 değerleri arasında olacaktır. Fark oranı değeri, $(p/1 - p)$ sadece pozitif değerler alıp bu değerlerin logaritması $\ln(p/1 - p)$ "logit" olarak isimlendirilmektedir eşitlik (5.24)'te verilmiştir.

$$\text{logit} = \ln\left(\frac{p}{1-p}\right) = \alpha + \beta X \quad (5.24)$$

Logit değerleri $+\infty$ ve $-\infty$ sonsuz arasında yer alacaktır. Eşitlik (5.24)'te verilen ifade matematiksel olarak eşitlik (5.25)'e eşittir.

$$p = \frac{\exp(\alpha + \beta X)}{1 + \exp(\alpha + \beta X)} \quad (5.25)$$

Eşitlik (5.25)'nin sağında yer alan ifade logit model olarak isimlendirilmektedir. Y ile ilgili X değişkeni arasındaki fark oranı 3. Bölümde ayrıntılı olarak açıklandığı gibi (OR) logit modelde $OR = \exp(\beta_i)$ kullanılarak hesaplanabilmektedir.

Y değişkeni j kategoriye sahip yanıt değişkeni iken J sıralı kategoriler dikkate alındığında birikimli olasılıklar (eşitlik (5.26)), birikimli fark oranı (eşitlik (5.27)) ve birikimli logitler (eşitlik (5.28)) sırası ile aşağıdaki gibidir:

$$P(Y \leq i) = p_1 + \dots + p_i \quad (5.26)$$

$$\text{odd}(Y \leq i) = \frac{p(Y \leq i)}{1 - p(Y \leq i)} = \frac{p_1 + \dots + p_i}{p_{i+1} + \dots + p_J} \quad (5.27)$$

$$\text{logit}(Y \leq i) = \ln\left(\frac{p(Y \leq i)}{1 - p(Y \leq i)}\right) \quad (5.28)$$

Buna göre sıralı değişken için birikimli lojistik model

$$\text{logit}(Y \leq i) = \alpha_i + \beta_{i1}X_1 + \dots + \beta_{im}X_m \quad (5.29)$$

şeklinde olacaktır. Birikimli fark oranları aşağıdaki şekilde verilebilir:

$$\text{odd}(Y \leq i) = \exp(\alpha_i) \exp(\beta_{i1}X_1 + \dots + \beta_{im}X_m) \quad (5.30)$$

Eşitlik (5.30)'da j-1 fark oranının sadece sabit terim α_i 'e bağlı olarak farklılık göstereceği anlamına gelmektedir. Bağımlı değişkenin kategorilerine göre β değişmemektedir. Böylece model x_i ve Y arasındaki ilişkinin j'dan bağımsız olduğunu varsayar. McCullagh bu varsayıma orantısal oran varsayımı (veya paralel regresyon varsayımı), modele de orantısal oran modeli adını vermiştir. McCullagh sıralı logit model için orantılı olasılık modeli terimini kullanmıştır (Bender ve

Grouven, 1997: 547). Birikimli logit modeller, paralel eğriler varsayımının sağlanıp sağlanmadığı durumlara göre orantısal oran modeli, orantısal olmayan oran modeli ve kısmi orantısal oran modeli olarak 3'e ayrılmaktadır. Bu bölümde sadece orantısal oran modeli ve kısmi orantısal oran modeli (genelleştirilmiş logit model) anlatılacaktır.

Lojistik dağılım altında tesadüfi değişken X için birikimli dağılım fonksiyonu (5.31)'deki gibidir:

$$Pr = (X \leq x) = \Lambda(\varepsilon) = \frac{\exp(x)}{1 + \exp(x)} = \frac{1}{(1 + \exp(-x))} \quad (5.31)$$

Buna göre hata terimi lojistik dağılımı gösterdiği varsayımı altında üç kategorili sıralı değişken için birikimli dağılım fonksiyonu (5.32)- (5.34) eşitliklerindeki gibi olacaktır (Borooah, 2001:11):

$$Pr(y_i = 1 | x_i) = \Lambda(\tau_1 - X_i\beta) = \frac{1}{[1 + \exp(x_i\beta - \tau_1)]} \quad (5.32)$$

$$\begin{aligned} Pr(y_i = 2 | x_i) &= \Lambda(\tau_2 - x_i\beta) - \Lambda(\tau_1 - x_i\beta) \\ &= \frac{1}{[1 + \exp(x_i\beta - \tau_2)]} - \frac{1}{[1 + \exp(x_i\beta - \tau_1)]} \end{aligned} \quad (5.33)$$

$$\begin{aligned} Pr(y_i = 3 | x_i) &= 1 - \Lambda(\tau_2 - x_i\beta) \\ &= 1 - \frac{1}{[1 + \exp(x_i\beta - \tau_2)]} \end{aligned} \quad (5.34)$$

5.2.2. Sıralı Probit Model

X değişkeni ile 0 ortalama ve 1 varyansla standart normal birikimli dağılım eşitlik (5.35)'teki gibidir.

$$Pr(X < x) = \Phi(x) = \int_0^x \left(\frac{1}{2\pi} \right) \exp\left(-\frac{x^2}{2}\right) dx \quad (5.35)$$

Hata terimi birikimli standart normal dağılım gösterdiğinde

$$Pr(Y_i \leq j | x_i) = \Phi(\alpha + \beta x) \quad (5.36)$$

olacaktır. Sıralı probit model sıralı logit modele oldukça benzerdir. İki model arasındaki kullanım farklılığı birbirinden bağımsız olarak geliştirilen iki ayrı disiplin olduğu söylenebilir. McKelvey ve Zavonia'nın (1975) sıralı probit model çalışması sosyal bilimlerde, McCullagh'ın (1980) sıralı logit modeli çalışması bioistatistik alanında kullanmıştır. Sıralı logit ve probit modellerde ortak etkiler benzer yorumlara sahiptir. Bununla birlikte sıralı probit modelinde kategorik bağımsız değişkenlerin etkileri için fark oranı yorumu yoktur (Hill vd, 2017: 711; Power ve Xie, 1999:215).

Sıralı probit modele ait olasılık eşitlikleri 3 kategorili bağımlı değişken için aşağıda verilmiştir.

$$Pr(y_i = 1 | x_i) = \Phi(\tau_1 - x_i\beta)$$

$$Pr(y_i = 2 | x_i) = \Phi(\tau_2 - x_i\beta) - \Phi(\tau_1 - x_i\beta) \quad (5.37)$$

$$Pr(y_i = 3 | x_i) = 1 - \Phi(\tau_2 - x_i\beta)$$

Sıralı probit modelinde de parametre tahminleri en yüksek olabilirlik yöntemi ile gerçekleştirilmektedir (Borooah, 2001: 12).

5.3. Model Tahmin Sonuçlarının Yorumlanması

Sıralı logit ve sıralı probit model tahminleri için farklı yorumlamalar yapılabilmektedir. Sıralı logit ve sıralı probit model katsayılarının doğrudan yorumlanmasından ziyade y^* değişkeninin yeniden ölçeklendirilmesi ile marjinal katsayılar elde edilip yorumlanabilir ya da fark oranı değerleri elde edilerek yorum yapılabilir. Bu yorumlar bağımsız değişkenin polinomial ya da etkileşim terimi şeklinde olmadığı durumlar için geçerlidir (Long ve Freese, 2014: 332).

5.3.1. Fark Oranlarının Elde Edilmesi

Sıralı logit modellerde fark oranları tahmin edilen regresyon katsayılarının üstel değeri alınarak yorumlanması yapılır (e^{β_k}) (Long ve Freese, 2014: 335; Bender ve Grouven, 1997:546; Fullerton ve Xu: 2016: 25). Ancak bu değerler sıralı probit modeller için geçerli değildir. Fark oranının 1'den büyük olması bağımlı değişkenin daha üst bir düzeyinde olma şansının yüksek, 1'den küçük olması ise bağımlı değişkenin daha üst bir düzeyinde olma şansının düşük olduğu şeklinde yorumlanmaktadır (Long, 1997: 80). Elde edilen fark oranları diğer değişkenlerin etkileri sabit tutulduğunda, bağımsız x değişkenindeki bir birimlik artış karşısında bağımlı değişkenin daha yüksek kategorisi ile daha düşük kategorisi karşılaştırıldığındaki değişim (e^{β_k}) kadardır şeklinde yorumlanır. Ayrıca fark oranında yüzdesel değişim yorumu yapılması için $(e^{\beta_k} - 1) * 100$ değeri hesaplanır. (e^{β_k}) değeri çarpım katsayısıdır. Fark oranı için model katsayı değerleri "pozitif" etkilerin 1'den büyük olduğu ve negatif etkilerin 0 ile 1 arasında olduğu anlamına gelmektedir. Pozitif ve negatif etkinin büyüklükleri karşılaştırılırken, negatif etkinin tersi alınarak karşılaştırma yapılmalıdır (Borooah, 2001:10; Long, 1997: 82; Agresti, 2010: 89).

5.3.2. Marjinal Etkilerin Yorumu

Sıralı modellere ait bir başka yorum ise marjinal değişimler için yapılmaktadır. Marjinal etkilerin hesaplanması ve yorumlanması sürekli ve kategorik değişkenler için aşağıda ayrı olarak anlatılmıştır.

m. sonucun olasılığındaki marjinal değişiklik sürekli değişkenler için aşağıdaki şekilde hesaplanmaktadır:

$$\frac{\partial Pr(y=m|x)}{\partial x_k} = \frac{\partial F(\tau_m - x\beta)}{\partial x_k} - \frac{\partial F(\tau_{m-1} - x\beta)}{\partial x_k} \quad (5.38)$$

Kısmi değişim veya marjinal etkiler tüm değişkenlerin düzeylerine bağlıdır. Hangi değişkene bağlı olarak hesaplanacağına karar verilmesi gerekmektedir. Marjinal etki çoğunlukla tüm değişkenlerin ortalama değerlerinde hesaplanmaktadır (Long, 1997: 135; Borooah, 2001: 12-13):

$$\frac{\partial Pr(y=m|\bar{x})}{\partial x_k} = \beta_k [f(\tau_{m-1} - \bar{x}) - f(\tau_m - \bar{x})] \quad (5.39)$$

Marjinal etki; bağımsız değişkendeki bir birimlik değişimin bağımlı değişkenin ilgili kategorisinde gerçekleşme olasılığına katkısı şeklinde yorumlanmaktadır.

Kukla değişkenler için kesikli değişim eşitlik (5.40) şeklinde hesaplanmakta ve tahmin edilen olasılıktaki değişim, x_k kukla değişkeninin x_s başlangıç değerinden x_E son değerine kadar olan değişimi göstermektedir (Long, 1997: 135-136; Borooah, 2001:14-15).

Kategorik değişkenler için marjinal etkiler (örneğin 0 dan 1 olan değişim)

$$\frac{\Delta Pr(y=m|x)}{\Delta x_k} = Pr(y = m|x, x_k = x_E) - Pr(y = m|x, x_k = x_s) \quad (5.40)$$

eşitlik (5.40) şeklinde hesaplanır. Elde edilen bu değer, x_k değişkeni x_s den x_E değerine değiştiğinde sonuç m için tahmini olasılık $\Delta Pr(y = m|x)/\Delta x_k$ kadar değişmektedir şeklinde yorumlanmaktadır.

5.4. Paralel Eğriler Varsayımı

Sıralı logit ve sıralı probit modellerde en önemli varsayım modelde eğim parametrelerin bağımlı değişkenin kategorilerine göre değişmediğidir. Sıralı logit modeldeki orantılı olasılıklar varsayımı, daha genel olarak hem sıralı logit hem de probit modellerinde paralel regresyon varsayımına karşılık gelmektedir (Long, 1997:140; Borooah, 2001: 14). Paralel regresyonlar varsayımı sıralı modelin y 'nin m değerine eşit ya da küçük olduğunu birikimli olasılık cinsinden yazılımı ile açıklanabilir:

$$Pr(y_i \leq m | x) = F(\tau_m - x\beta) \quad (5.41)$$

Birikimli olasılık; birikimli dağılım fonksiyonu F 'in $\tau_m - x\beta$ daki değeridir. β tüm m 'ler için aynı olduğundan, denklem (5.41) ikili cevap değişkenli modeller grubunun farklı eşik değerleri ile açıklanabilir (Liao, 1994: 41; DeMaris, 2004: 305; Long, 1997: 140):

$$\tau_m - x\beta = (\tau_m - x\beta) - \sum_{k=1}^K \beta_k X_k$$

$y \leq 1$ için model $\tau_1 - \beta_0$ sabit terimi ile

$$Pr(y \leq 1|x) = F((\tau_1 - \beta_0) - \sum_{k=1}^K \beta_k X_k)$$

$y \leq 2$ için model $\tau_2 - \beta_0$ sabit terimi ile

$$Pr(y \leq 2|x) = F((\tau_2 - \beta_0) - \sum_{k=1}^K \beta_k X_k)$$

x_k 'lara ait katsayı tahminleri değişmezken sabit terimler değişmektedir.

$$\frac{Pr(y \leq 1|x)}{\partial x} = \frac{Pr(y \leq 2|x)}{\partial x} = \dots = \frac{Pr(y \leq j|x)}{\partial x}$$

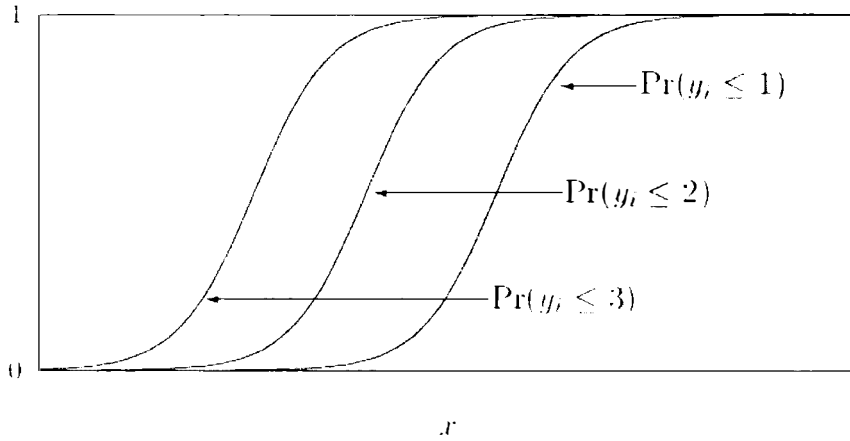
Bu durumda regresyon eğrileri paraleldir (Long, 1997: 140-141). Paralel regresyon varsayımı $\beta_1 = \beta_1 = \dots \beta_{j-1}$ olduğunu ifade etmektedir. Paralellik testine göre hipotez

H_0 : İlgili regresyon katsayıları bağımlı değişkenin tüm kategorilerinde aynıdır.

H_1 : İlgili regresyon katsayıları bağımlı değişkenin tüm kategorilerinde farklıdır.

Paralel regresyon varsayımının geçerliliği çok durumlu logit model tahmin edilerek test edilir. Çok durumlu logit model J tane kategorinin β_k eğim katsayılarının farklı olmasına olanak sağlamaktadır. Sıralı modeller K tane parametre tahminlerken çok durumlu logit model $K(J-1)$ parametre tahminlemektedir. L_1 değeri sıralı modelden elde edilen olabilirlik değerini; L_2 ise çok durumlu modelden elde edilen olabilirlik değerini gösterebilir. H_0 hipotezini test etmek için $2(L_2 - L_1)$ değeri ki-kare tablosundan elde edilen $\chi^2_{K(J-2)}$ değeri ile karşılaştırılır. Burada dikkat edilmesi gereken nokta; sıralı modeller çok durumlu modeller içinde yuvalanmış olmadığından ve iki ayrı modelden elde edilen olabilirlik oranı karşılaştırıldığından yapılan testin olabilirlik oran testi olmadığıdır. Eğer araştırmacının paralel eğim varsayımının geçersiz olduğuna dair nedenleri varsa, bağımlı değişkenin sıralı olduğuna bakmadan çok durumlu model kullanmalıdır (Borooah, 2001: 15). Şekil 5.3'te sıralı bir değişken için paralel eğim varsayımı gösterilmektedir. İlerleyen kısımlarda Stata programı ile gerçekleştirilebilen paralel regresyon varsayımı testleri daha ayrıntılı anlatılacaktır.

Şekil 5.3. Paralel Eğriler Varsayımının Gösterimi



Kaynak: (Powers ve Xie, 1999:213).

5. 5. Sıralı Model Tahminleri İçin Stata Uygulaması

Bu kısımda Türkiye İstatistik Kurumu tarafından 2018 yılında gerçekleştirilen “Yaşam Memnuniyeti Anketi” verileri kullanılarak Türkiye’de yaşayan bireylerin yaşam memnuniyeti düzeylerinin tespitine yönelik analizler yapılmıştır.

Çalışmada veriler hanehalklarındaki 18 yaş ve üzerindeki bireylerle yüz yüze görüşme yöntemiyle toplanmıştır. 2018 yılında araştırma kapsamında 4733 hane ve 9719 kişi ile görüşme yapılmıştır (TÜİK, 2018: 6). Çalışmada sıralı model yaklaşımını kullanmak için Türkiye’de yaşam memnuniyeti düzeyi olarak “yaşam memnuniyeti” bağımlı değişken olarak kullanılmıştır. 11’li likert ölçeği ile (“0: Hiç memnun değil”, “10: Çok memnun” olarak) ölçülmüş ve 11 kategoriden 3 kategoriye indirgenmiştir. Verilerin frekans dağılımına uygun olarak kategoriler “0-4” “5-6” “7-10” aralıklarında birleştirilmiştir. “0-4” arasındaki değerler “Memnun değil”, “5-6” değerleri “Orta derecede memnun”, “7-10” arasındaki değerler “Memnun” olarak tanımlanmıştır.

lanmıştır. Bağımlı değişken sıralı model analizine uygun olacak şekilde 1= Memnun Değil, 2= Orta derecede memnun ve 3=Memnun değerleri ile kodlanmıştır.

Tablo 5.1. Değişkenlerin Tanımları

Değişkenler	Açıklamaları
Bağımlı Değişken	
memnuniyet	Bir bütün olarak düşündüğünüzde, son zamanlardaki yaşam memnuniyetinizi değerlendiriniz. 1. Memnun Değil 2. Orta düzeyde 3. Memnun
Bağımsız Değişkenler	
cinsiyet	Birey erkek ise 1, değilse 0 değerini alan kukla değişken
yas	Ferdin bitirdiği yaş (18 yaş ve üzeri)
evli	Birey evli ise 1, değilse 0 değerini alan kukla değişken
egitimsiz	Birey Bir okul bitirmedi ise 1 değilse 0 değerini alan kukla değişken
ilkokul	Birey İlkokul mezunu ise 1 değilse 0 değerini alan kukla değişken
ortaokul	Birey İlköğretim-ortaokul mezunu ise 1 değilse 0 değerini alan kukla değişken
lise	Birey Lise ve dengi okul mezunu ise 1 değilse 0 değerini alan kukla değişken
universite	Birey yükseköğretim mezunu ise 1 değilse 0 değerini alan kukla değişken
gel1	Birey düşük gelir grubunda ise 1 değilse 0 değerini alan kukla değişken
gel2	Birey orta gelir grubunda ise 1 değilse 0 değerini alan kukla değişken
gel3	Birey yüksek gelir grubunda ise 1 değilse 0 değerini alan kukla değişken
gelkarsilama	Bu gelirle hanenizin temel ihtiyaçlarını hangi düzeyde karşılayabiliyorsunuz? 1. Çok zor 2. Zor 3. Orta 4. Kolay 5. Çok kolay
sos_hayat	Sosyal hayatınızdan (eğlence, kültürel ve sportif faaliyetler gibi) memnunuz? 1. Hiç memnun değil 2. Memnun değil 3. Orta 4. Memnun 5. Çok memnun

Gelir grubu değişkeni araştırmada 5 şıklı (“1:0-1741 TL”, “2:1742-2499 TL”, “3:2500-3499 TL”, “4:3500-5126 TL”, “5:5127 TL ve üzeri”) olarak cevaplandırılmıştır. Bu değişken modele eklenirken 1. 2. ve 3. kategoriler birleştirilip “Düşük gelirli”, 4. kategori “Orta gelirli”, 5. kategori “Yüksek gelirli” şeklinde tanımlanarak 3 kategoride toplanmıştır. Çalışmada kullanılan verilere ait tanımlayıcı istatistik tabloları Stata programında **describe**, **summarize** ve **tabulate** komutları ile elde edilebilmektedir.

Tablo 5.2. bolum5.dta Veri Seti İçeriği**describe**

	storage	display	value	
variable name	type	format	label	variable label
yas	int	%8.0g		Yaş
cinsiyet	byte	%8.0g		Cinsiyet
evli	byte	%8.0g		Evli
memnuniyet	byte	%11.0g		Yaşamdan memnuniyet düzeyi
egitimsiz	byte	%8.0g		Okul Bitirmedi
ilkokul	byte	%8.0g		İlkokul Mezunu
ortaokul	byte	%8.0g		Ortaokul veya ilköğretim mezunu
lise	byte	%8.0g		Lise mezunu
universite	byte	%8.0g		Üniversite ve Üstü Mezunu
gelkarsilama	byte	%9.0g		Gelir karşılama düzeyi
sos_hayat	byte	%9.0g		Sosyal Hayattan memnuniyet
gel1	byte	%8.0g		Düşük Gelir
gel2	byte	%8.0g		Orta Gelir
gel3	byte	%8.0g		Yüksek Gelir

Stata’da tanımlayıcı istatistikleri elde etmek için “**sum**” komutu kullanılır. Tablo5.3’te tanımlayıcı istatistikler verilmiştir.

Tablo 5.3. Değişkenlere Ait Tanımlayıcı İstatistikler

sum memnuniyet cinsiyet yas evli egitimsiz ilkokul ortaokul lise universite gel1 gel2 gelkarsilama sos_hayat

Variable	Obs	Mean	Std. Dev.	Min	Max
-----+-----					
memnuniyet	9566	2.027493	.7715382	1	3
cinsiyet	9566	.4650847	.4988055	0	1
yas	9566	44.43153	16.1751	18	96
evli	9566	.742212	.4374395	0	1
egitimsiz	9566	.1291031	.3353315	0	1
-----+-----					

ilkokul	9566	.3421493	.4744541	0	1
ortaokul	9566	.1452018	.3523226	0	1
lise	9566	.2021744	.4016426	0	1
universite	9566	.1813715	.3853458	0	1
gell	9566	.5958603	.4907504	0	1
-----+-----					
gel2	9566	.2433619	.4291342	0	1
gel3	9566	.1607778	.367345	0	1
gelkarsilama	9566	2.572758	.9034325	1	5
sos_hayat	9566	3.144679	.9728013	1	5

Yukarıdaki Tablo 5.3'e göre toplam gözlem sayısı 9566'dır ve tüm değişkenlere ait ortalama ve standart sapma değerleri ayrıca gösterilmektedir. Tablo 5.3 incelendiğinde bağımlı değişkenin ortalamasının yaklaşık 2 olduğu görülmektedir. Bu değer bize çalışmaya katılan bireylerin yaşam memnuniyet düzeylerinin "orta derecede" olduğunu göstermektedir. Bireylerin %46'ı erkek, %74'ü evli, %34'ünün ilkokul mezunu olduğu görülmektedir. Geliri 3499 TL'nin altında olan bireylerin oranı %59'dur.

Tüm değişkenler için frekans, yüzde ve birikimli yüzdeler Stata'da "**tab**" komutu ile elde edilebilir. Bağımlı değişken "memnuniyet" için tab komut sonucu Tablo 5.4'te verilmiştir.

Tablo 5.4. Bağımlı Değişkene Ait İstatistikler

tab memnuniyet

memnuniyet	Freq.	Percent	Cum.
-----+-----			
Memnun Değil	2,719	28.42	28.42
Orta	3,865	40.40	68.83
Memnun	2,982	31.17	100.00
-----+-----			
Total	9,566	100.00	

Tablo 5.4'e göre bireylerin %31'i yaşamlarından memnun iken, %40'ı yaşamdan orta derecede memnun olduklarını ve %29'u da yaşamdan memnun olmadıklarını belirtmişlerdir.

5.5.1. Sıralı Logit Model İçin Stata Tahminleri

Bu bölümde sıralı model tahminlerine yer verilmiştir. İlk olarak sıralı logit model tahmini için Stata uygulaması anlatılacaktır. Sıralı logit model tahmini için "**ologit**" komutu kullanılarak model tahmini yapılmaktadır. Komutlar hakkında daha ayrıntılı açıklamalara **help** komutu yardımı ile ulaşılabilir.

ologit depvar [indepvars] [if] [in] [weight] [, options]

depvar: Kategorik bağımlı değişken; 1, 2 ve 3 şeklinde

indepvars: Bağımsız değişkenler listesi

if: Örnekleme sınırlandırmak için if komutu kullanılabilir.

Örneğin sadece erkek bireyler için hesaplanmak isteniyorsa

ologit memnuniyet cinsiyet yas evli ilkokul ortaokul lise universite gel2 gel3 gelkarsilama sos_hayat if cinsiyet==1 yazılmalıdır.

Tablo 5.5. Sıralı Logit Model Tahmin Sonuçları

ologit memnuniyet cinsiyet yas evli ilkokul ortaokul lise universite gel2 gel3 gelkarsilama sos_hayat, nolog

Ordered logistic regression	Number of obs	=	9566
	LR chi2(11)	=	1501.60
	Prob > chi2	=	0.0000
Log likelihood = -9648.1202	Pseudo R2	=	0.0722

memnuniyet	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
cinsiyet	-.1374002	.0405661	-3.39	0.001	-.2169083	-.057892
yas	-.0056668	.0014535	-3.90	0.000	-.0085157	-.0028179
evli	.3540799	.0467186	7.58	0.000	.2625132	.4456466
ilkokul	.3600899	.0666121	5.41	0.000	.2295327	.4906472
ortaokul	.4821915	.0845628	5.70	0.000	.3164514	.6479315
lise	.7424256	.0803945	9.23	0.000	.5848552	.8999959
universite	.8347755	.0859506	9.71	0.000	.6663154	1.003236
gel2	.2541004	.0487291	5.21	0.000	.1585932	.3496076
gel3	.4095542	.0639512	6.40	0.000	.2842122	.5348963
gelkarsilama	.3120314	.0242793	12.85	0.000	.2644449	.3596179
sos_hayat	.482712	.0213367	22.62	0.000	.4408929	.5245312
/cut1	1.845245	.1223036			1.605535	2.084956
/cut2	3.788041	.1274557			3.538232	4.037849

Tablo 5.5 incelendiğinde katsayı tahminlerinin işaretleri bağımsız değişken ile bağımlı değişken arasındaki ilişkinin yönü hakkında bilgi vermektedir. Katsayı tahminlerinin daha anlaşılır olması için fark oranı yada marjinal etkilerin yorumlanması tercih edilmektedir. Fark oranı ve marjinal etki yorumlarından önce eşik değerlerin yorumu yapılacaktır. Tablo 5.5'te elde edilen eşik noktalarına bakıldığında $\_cut1=1.845$ 'dir. Bağımsız değişkenler sıfır iken hiç memnun değil kategorisini, orta derecede memnun ve memnun kategorilerinden ayırt etmek için kullanılan gizli değişken üzerindeki kesme noktasıdır. 1.845 memnun değil kategorisinin üst sınırıdır. $\_cut2=3.788$ bağımsız değişkenler sıfır iken, bireyin hiç memnun değil ve orta derecede memnun kategorilerini, memnun kategorisinden ayırt etmek için kullanılan gizli değişken üzerindeki kesme noktasıdır. 3.788, orta derece memnun olanların üst sınırıdır. Buna göre 1.845'den daha küçük gizli değişken değerine sahip olan bireyler hiç memnun değil kategorisine, 1.845 ile 3.788 arasında gizli değişken değerine sahip bireyler orta derecede memnun kategorisine, 3.788 ve daha yüksek gizli değişken değerine sahip olan bireyler de memnun kategorisine girecektir. Tablo 5.5'te eşik değerleri %95 güvenle anlamlıdır. Yukarıdaki açıklamalardan anlaşılacağı üzere, bireyin yaşamdan hiç memnun olmama olasılığı $Pr(\hat{\alpha} + \hat{\beta}x_i + \varepsilon_i \leq 1.845)$, yaşamdan orta düzeyde memnun olma olasılığı $Pr(1.845 \leq \hat{\alpha} + \hat{\beta}x_i + \varepsilon_i \leq 3.788)$ ve yaşamdan memnun olma olasılığı $Pr(\hat{\alpha} + \hat{\beta}x_i + \varepsilon_i \geq 3.788)$ şeklinde de ifade edilir.

Model tahminleri yorumlanmadan modelin geçerliliğinin araştırılması gerekmektedir. Model geçerliliği istatistiksel modelin, gözlem setine ne kadar iyi uyduğunu (uyum iyiliğini) açıklamaktadır. Sıralı regresyon modellerinde uyum iyiliği ölçüsü olarak olabilirlik oran (Likelihood Ratio- LR) istatistiği kullanılmaktadır (Liu, 2016: 200). LR istatistiği tüm değişkenlerin olduğu kısıtsız modelin olabilirliğinin logaritması (L_2) ve sadece sabit terimden oluşan kısıtlı modelin olabilirliğinin logaritması (L_1) ile hesaplanmaktadır. Test istatistiği $LR = -2(\ln L_2 - \ln L_1)$ dir. Bu teste ait sıfır hipotezi H_0 : Tahmin edilen model geçersizdir şeklindedir (Powers ve Xie, 1999: 66-67).

Tablo 5.5'te sıralı logit model LR değeri $LR = 1501.60$ olasılık değeri $p = 0.00$ dır. Bu sonuca göre H_0 hipotezi 0.01 önem düzeyinde reddedilmektedir ve tahmin edilen Tablo 5.5'teki modelin geçerli olduğunu göstermektedir. Tahmin edilen sıralı logit model varsayımları sağlandığında sıralı logit modelinin en yüksek olabilirlik tahminleri asimptotik olarak normal dağılmaktadır. Bu sebeple katsayı tahminlerinin anlamlılık testleri $H_0: \beta_k = 0$ hipotezi $z = (\hat{\beta}_k - \beta^*) / \hat{\sigma}_{\beta_k}$ ile test edilebilir (Long ve Freese, 2014: 321). Tablo 5.5'te verilen sıralı logit katsayılarının z değerleri incelendiğinde tüm değişkenlerin katsayılarının yaşamdan memnun olma düzeyi üzerinde anlamlı bir etkiye sahip olduğu söylenebilir ($p < 0.01$). Tablo 5.5'te elde edilen sıralı logit model tahmin sonuçlarının Stata programında fark oranlarının hesaplanması için model tahmin komut satırının sonuna “, or” eklenmektedir. Elde edilen fark oranı sonuçları Tablo 5.6'da verilmiştir.

dy/dx w.r.t. : cinsiyet yas evli ilkokul ortaokul lise universite gel2 gel3
gelkarsilama sos_hayat

```
at      : cinsiyet      =   .4650847 (mean)
        yas            =   44.43153 (mean)
        evli           =   .742212 (mean)
        ilkokul        =   .3421493 (mean)
        ortaokul       =   .1452018 (mean)
        lise           =   .2021744 (mean)
        universite     =   .1813715 (mean)
        gel2           =   .2433619 (mean)
        gel3           =   .1607778 (mean)
        gelkarsilama   =   2.572758 (mean)
        sos_hayat      =   3.144679 (mean)
```

	Delta-method					
	dy/dx	Std. Err.	z	P> z	[95% Conf. Interval]	
cinsiyet	.0264476	.0078093	3.39	0.001	.0111417	.0417535
yas	.0010908	.0002797	3.90	0.000	.0005425	.0016391
evli	-.0681555	.0090029	-7.57	0.000	-.0858008	-.0505102
ilkokul	-.0693124	.0128315	-5.40	0.000	-.0944616	-.0441631
ortaokul	-.0928152	.0162884	-5.70	0.000	-.1247399	-.0608905
lise	-.1429067	.0154939	-9.22	0.000	-.1732741	-.1125393
universite	-.1606828	.0165547	-9.71	0.000	-.1931295	-.1282361
gel2	-.0489108	.0093832	-5.21	0.000	-.0673016	-.0305201
gel3	-.0788336	.0123115	-6.40	0.000	-.1029636	-.0547035
gelkarsilama	-.0600617	.0046805	-12.83	0.000	-.0692353	-.0508882
sos_hayat	-.0929154	.0041359	-22.47	0.000	-.1010216	-.0848092

margins, dydx(*) predict(outcome(#2)) atmeans

Conditional marginal effects Number of obs = 9566

Model VCE : OIM

Expression : Pr(memnuniyet==2), predict(outcome(#2))

dy/dx w.r.t. : cinsiyet yas evli ilkokul ortaokul lise universite gel2 gel3
gelkarsilama sos_hayat

```
at      : cinsiyet      =   .4650847 (mean)
        yas            =   44.43153 (mean)
        evli           =   .742212 (mean)
        ilkokul        =   .3421493 (mean)
        ortaokul       =   .1452018 (mean)
        lise           =   .2021744 (mean)
        universite     =   .1813715 (mean)
        gel2           =   .2433619 (mean)
        gel3           =   .1607778 (mean)
        gelkarsilama   =   2.572758 (mean)
        sos_hayat      =   3.144679 (mean)
```

	Delta-method					
	dy/dx	Std. Err.	z	P> z	[95% Conf. Interval]	
cinsiyet	.0018147	.0007186	2.53	0.012	.0004063	.003223
yas	.0000748	.0000276	2.71	0.007	.0000207	.000129
evli	-.0046764	.0013757	-3.40	0.001	-.0073727	-.0019801
ilkokul	-.0047558	.001525	-3.12	0.002	-.0077447	-.0017669
ortaokul	-.0063684	.00201	-3.17	0.002	-.0103079	-.0024288
lise	-.0098053	.0027906	-3.51	0.000	-.0152747	-.004336
universite	-.011025	.003126	-3.53	0.000	-.0171518	-.0048982
gel2	-.003356	.0010925	-3.07	0.002	-.0054972	-.0012147
gel3	-.0054091	.0016622	-3.25	0.001	-.008667	-.0021512
gelkarsilama	-.0041211	.0011339	-3.63	0.000	-.0063434	-.0018987
sos_hayat	-.0063753	.0017016	-3.75	0.000	-.0097104	-.0030401

margins, dydx(*) predict(outcome(#3)) atmeans

Conditional marginal effects Number of obs = 9566

Model VCE : OIM

Expression : Pr(memnuniyet==3), predict(outcome(#3))

dy/dx w.r.t. : cinsiyet yas evli ilkokul ortaokul lise universite gel2 gel3
GELKARSİLAMA sos_hayat

at	:	cinsiyet	=	.4650847 (mean)
		yas	=	44.43153 (mean)
		evli	=	.742212 (mean)
		ilkokul	=	.3421493 (mean)
		ortaokul	=	.1452018 (mean)
		lise	=	.2021744 (mean)
		universite	=	.1813715 (mean)
		gel2	=	.2433619 (mean)
		gel3	=	.1607778 (mean)
		gelkarsilama	=	2.572758 (mean)
		sos_hayat	=	3.144679 (mean)

Delta-method						
	dy/dx	Std. Err.	z	P> z	[95% Conf. Interval]	
cinsiyet	-.0282624	.0083448	-3.39	0.001	-.0446179	-.011907
yas	-.0011656	.0002991	-3.90	0.000	-.0017518	-.0005794
evli	.0728322	.0096093	7.58	0.000	.0539983	.0916662
ilkokul	.0740685	.0136968	5.41	0.000	.0472233	.1009136
ortaokul	.099184	.0173895	5.70	0.000	.0651012	.1332669
lise	.1527127	.0165348	9.24	0.000	.1203051	.1851204
universite	.1717086	.0176932	9.70	0.000	.1370305	.2063867
gel2	.052267	.0100223	5.22	0.000	.0326237	.0719103
gel3	.084243	.0131619	6.40	0.000	.0584462	.1100398
gelkarsilama	.0641831	.0049991	12.84	0.000	.0543851	.0739811
sos_hayat	.0992911	.0043862	22.64	0.000	.0906944	.1078879

Tablo 5.7 incelendiğinde diğer değişkenler sabit tutulduğunda bireylerin yaşı arttıkça yaşamdan memnun olma olasılığı azalmaktadır. Bir başka ifadeyle, birey bir yaş daha yaşlandığında yaşamdan hiç memnun olmama kategorisinde olma olasılığı %0.109 ve orta derecede yaşamdan memnun olma kategorisindeki olasılığı %0.0074 artarken; yaşamdan memnun olma kategorisinde olma olasılığını %0.116 azaltmaktadır.

Bireylerin geliriyle hanesinin temel ihtiyaçlarını karşılama düzeyi arttıkça daha fazla yaşamdan memnun olmaktadır. Daha açık anlatımla, bireylerin geliri hanenin ihtiyaçlarını karşılaması

kolaylaştıkça yaşamdan memnun olmama kategorisindeki olasılığı %6 ve orta derecede yaşamdan memnun olma kategorisindeki olasılığı %0.41 azalırken, yaşamdan memnun olma kategorisindeki olasılığı %6.4 artmaktadır.

Sosyal hayatınızdan memnun musunuz sorusuna verilen yanıtlar incelendiğinde; bireylerin sosyal hayatlarından memnuniyet düzeyleri arttıkça yaşamdan memnun olmama olasılığını %9 ve yaşamdan orta düzeyde memnun olma olasılığı %0.6 azalırken, yaşamdan memnun olma olasılığı ise %9.9 artmaktadır.

Tablo 5.7 incelendiğinde erkeklerin kadınlara göre yaşamdan memnun olmama olasılıkları 0.029 ve orta derecede memnun olma olasılığı 0.0018 daha fazla iken erkeklerin yaşamdan memnun olma kategorisindeki olasılığı 0.028 daha azdır. Bulunan bu sonuç model tahminindeki hem cinsiyet değişkeni katsayısının işaretiyle hem de fark oranı yorumu ile benzerdir. Ayrıca, orta gelirli bireylerin düşük gelirli bireylere göre yaşamdan memnun olmama kategorisindeki olasılığı %4.8 ve orta derecede memnun olma kategorisindeki olasılığı %0.33 azalırken, bununla birlikte orta gelirli bireylerin düşük gelirli bireylere göre yaşamdan memnun olma kategorisinde olma olasılığı %5.2 daha fazladır. İlkokul, ortaokul, lise ve üniversite eğitim seviyelerine sahip bireylerin eğitimsiz bireylere göre yaşamdan memnun olmama olasılıkları sırasıyla %6.9 %9.2, %14 ve %15 daha azdır. Bununla birlikte ilkök, ortaokul, lise ve üniversite eğitim seviyelerine sahip bireylerin eğitimsiz bireylere göre yaşamdan memnun olma olasılıkları sırasıyla %7.4, %9.9, %15.2 ve %17.1 daha fazladır. Diğer kategorilerle ilgili benzer yorumlar yapılabilir.

5.5.2. Sıralı Probit Model için Stata Tahminleri

Sıralı probit model tahmini için “**oprobit**” komutu kullanılarak model tahminleri gerçekleştirilmektedir. Komutlar hakkında daha ayrıntılı açıklamalara **help** komutu yardımı ile ulaşılabilir.

```
oprobit depvar [indepvars] [if] [in] [weight] [, options]
```

Tablo 5.8. Sıralı Probit Model Tahmin Sonuçları

```
oprobit memnuniyet cinsiyet yas evli ilkokul ortaokul lise universite gel2 gel3
gelkarsilama sos_hayat, nolog
```

```
Ordered probit regression          Number of obs   =      9566
                                LR chi2(11)          =     1492.37
                                Prob > chi2           =      0.0000
Log likelihood = -9652.7343        Pseudo R2       =      0.0718
```

memnuniyet	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
-----+-----						
cinsiyet	-.0823485	.024273	-3.39	0.001	-.1299228	-.0347743
yas	-.0032628	.0008697	-3.75	0.000	-.0049673	-.0015583
evli	.2117211	.0278529	7.60	0.000	.1571303	.2663118

ilkokul	.2197309	.0398987	5.51	0.000	.1415309	.297931
ortaokul	.2975794	.0505313	5.89	0.000	.1985399	.396619
lise	.4474032	.0480278	9.32	0.000	.3532704	.541536
universite	.4973345	.0512602	9.70	0.000	.3968664	.5978027
gel2	.1532241	.0291796	5.25	0.000	.0960333	.210415
gel3	.247732	.038339	6.46	0.000	.1725889	.3228751
gelkarsilama	.1838214	.0144054	12.76	0.000	.1555874	.2120555
sos_hayat	.2865053	.0125657	22.80	0.000	.261877	.3111337

/cut1	1.094986	.0727388			.9524203	1.237551
/cut2	2.266029	.0746687			2.119681	2.412377

Benzer şekilde Tablo 5.8 sıralı probit model LR değeri $LR = 1497.32$ olasılık değeri $p=0.00$ 'dır. Bu sonuca göre "tahmin edilen model geçersizdir" H_0 hipotezi 0.01 önem düzeyinde reddedilebilmektedir ve tahmin edilen sıralı probit modelin geçerli olduğu ifade edilmektedir. Tablo 5.8'de verilen eşik değerleri incelendiğinde $\_cut1=1.094$ 'dir. Bağımsız değişkenler sıfır iken hiç memnun değil kategorisini, orta derecede memnun ve memnun kategorilerinden ayırt etmek için kullanılan gizli değişkendir. İkinci eşik değeri $\_cut2=2.266$ dir. Bağımsız değişkenler sıfır iken bireyin hiç memnun değil ve orta derecede memnun kategorilerini, memnun kategorisinden ayırt etmek için kullanılan gizi değişken üzerindeki eşik değeridir. Bireyin yaşamdan hiç memnun olmama olasılığı $Pr(\hat{\alpha} + \hat{\beta}x_i + \varepsilon_i \leq 1.094)$, yaşamdan orta düzeyden memnun olma olasılığı $Pr(1.094 \leq \hat{\alpha} + \hat{\beta}x_i + \varepsilon_i \leq 2.266)$ ve yaşamdan memnun olma olasılığı $Pr(\hat{\alpha} + \hat{\beta}x_i + \varepsilon_i \geq 2.266)$ değerleri arasındadır. Tablo 5.8'deki eşik değerleri %95 güvenle anlamlıdır.

Sıralı probit modellerinin en yüksek olabilirlik tahminleri asimptotik olarak normal dağılımlıdır. Bu sebeple katsayı tahminlerinin anlamlılık testleri $H_0: \beta_k = 0$ hipotezi $z = (\hat{\beta}_k - \beta^*) / \hat{\sigma}_{\beta_k}$ ile test edilebilir (Long ve Freese, 2014: 321). Tablo 5.8 de verilen sıralı probit model tahmininde katsayılara ait z değerleri incelendiğinde tüm değişkenlerin katsayılarının yaşamdan memnun olma düzeyi üzerinde anlamlı bir etkiye sahip olduğu söylenebilir ($p < 0.01$).

Sıralı probit model varsayımlarından biri olan normal varsayım test sonucu Tablo 5.9'da verilmiştir. Buna göre hataların normal dağılıma sahip olmadığı görülmektedir ($p < 0.00$). Model tahmin sonuçlarının anlatılabilmesi için normallik varsayımı göz ardı edilip yorumlamaya devam edilecektir.

Tablo 5.9. Normal Dağılım Testi

```
. quietly oprobit memnuniyet cinsiyet yas evli ilkokul ortaokul lise universite gel2
gel3 gelkarsilama sos_hayat
```

```
. predict residual
```

(option pr assumed; predicted probabilities)

```
. sktest residual
```

Skewness/Kurtosis tests for Normality

----- joint -----

Variable	Obs	Pr(Skewness)	Pr(Kurtosis)	adj chi2(2)	Prob>chi2
residual	9.6e+03	0.0000	0.0013		0.0000

Tablo 5.8’de verilen katsayı tahmin değerlerinin işareti ve önem düzeyi bağımlı değişken ve bağımsız değişkenler arasındaki yön ve istatistiki önemi belirtirken katsayı büyüklükleri yorumlanmamaktadır (Agresti, 2010:123). Sıralı probit model tahminlerine ait katsayı yorumları ortalama değerlerde hesaplanan marjinal etkiler üzerinden yapılabilmektedir. Marjinal etki sonuçları Tablo 5.10’da verilmiştir.

Tablo 5.10. Marjinal Etkiler

quietly oprobit memnuniyet cinsiyet yas evli ilkokul ortaokul lise universite gel2
gel3 gelkarsilama sos_hayat

```
margins, dydx(*) predict(outcome(#1)) atmeans
```

Conditional marginal effects	Number of obs	=	9566

Model VCE : OIM

Expression : Pr(memnuniyet==1), predict(outcome(#1))

dy/dx w.r.t. : cinsiyet yas evli ilkokul ortaokul lise universite gel2 gel3
gelkarsilama sos.hayat

```
at      : cinsiyet      = .4650847 (mean)
        yas            = 44.43153 (mean)
        evli           = .742212 (mean)
        ilkokul        = .3421493 (mean)
```

```

    ortaokul      =    .1452018 (mean)
    lise          =    .2021744 (mean)
    universite    =    .1813715 (mean)
    gel2          =    .2433619 (mean)
    gel3          =    .1607778 (mean)
    gelkarsilama  =    2.572758 (mean)
    sos_hayat     =    3.144679 (mean)

```

	Delta-method					
	dy/dx	Std. Err.	z	P> z	[95% Conf. Interval]	
cinsiyet	.0269568	.0079472	3.39	0.001	.0113805	.042533
yas	.0010681	.0002847	3.75	0.000	.00051	.0016261
evli	-.0693069	.0091246	-7.60	0.000	-.0871907	-.051423
ilkokul	-.0719289	.0130689	-5.50	0.000	-.0975435	-.0463143
ortaokul	-.0974126	.0165519	-5.89	0.000	-.1298536	-.0649715
lise	-.1464574	.0157423	-9.30	0.000	-.1773117	-.1156031
universite	-.1628024	.0167973	-9.69	0.000	-.1957245	-.1298803
gel2	-.0501579	.0095556	-5.25	0.000	-.0688865	-.0314293
gel3	-.081095	.012552	-6.46	0.000	-.1056965	-.0564936
gelkarsilama	-.0601739	.0047249	-12.74	0.000	-.0694346	-.0509132
sos_hayat	-.0937875	.0041446	-22.63	0.000	-.1019108	-.0856641

margins, dydx(*) predict(outcome(#2)) atmeans

Conditional marginal effects Number of obs = 9566

Model VCE : OIM

Expression : Pr(memnuniyet==2), predict(outcome(#2))

dy/dx w.r.t. : cinsiyet yas evli ilkokul ortaokul lise universite gel2 gel3
gelkarsilama sos_hayat

```

at      : cinsiyet      =    .4650847 (mean)
        : yas          =    44.43153 (mean)
        : evli         =    .742212 (mean)
        : ilkokul      =    .3421493 (mean)

```



```

ortaokul      = .1452018 (mean)
lise          = .2021744 (mean)
universite    = .1813715 (mean)
gel2          = .2433619 (mean)
gel3          = .1607778 (mean)
gelkarsilama  = 2.572758 (mean)
sos_hayat     = 3.144679 (mean)

```

	Delta-method					
	dy/dx	Std. Err.	z	P> z	[95% Conf. Interval]	
cinsiyet	.0014062	.0005594	2.51	0.012	.0003097	.0025027
yas	.0000557	.0000211	2.64	0.008	.0000144	.000097
evli	-.0036153	.0010781	-3.35	0.001	-.0057284	-.0015023
ilkokul	-.0037521	.0012093	-3.10	0.002	-.0061222	-.001382
ortaokul	-.0050815	.001606	-3.16	0.002	-.0082292	-.0019337
lise	-.0076399	.0021999	-3.47	0.001	-.0119515	-.0033282
universite	-.0084925	.0024375	-3.48	0.000	-.0132699	-.0037151
gel2	-.0026165	.0008588	-3.05	0.002	-.0042997	-.0009332
gel3	-.0042303	.0013125	-3.22	0.001	-.0068027	-.0016578
gelkarsilama	-.0031389	.0008758	-3.58	0.000	-.0048555	-.0014224
sos hayat	-.0048924	.0013255	-3.69	0.000	-.0074902	-.0022945

```
margins, dydx(*) predict(outcome(#3)) atmeans
```

Conditional marginal effects Number of obs = 9566

Model VCE : OIM

```
Expression      : Pr(memnuniyet==3), predict(outcome(#3))
```

dy/dx w.r.t. : cinsiyet yas evli ilkokul ortaokul lise universite gel2 gel3

gelkarsilama sos hayat

```
at          : cinsiyet          =    .4650847 (mean)
              yas               =    44.43153 (mean)
              evli              =    .742212 (mean)
              ilkokul           =    .3421493 (mean)
```

ortaokul	=	.1452018 (mean)
lise	=	.2021744 (mean)
universite	=	.1813715 (mean)
gel2	=	.2433619 (mean)
gel3	=	.1607778 (mean)
gelkarsilama	=	2.572758 (mean)
sos_hayat	=	3.144679 (mean)

	Delta-method					
	dy/dx	Std. Err.	z	P> z	[95% Conf. Interval]	
cinsiyet	-.0283631	.0083607	-3.39	0.001	-.0447497	-.0119764
yas	-.0011238	.0002996	-3.75	0.000	-.001711	-.0005366
evli	.0729225	.0095973	7.60	0.000	.0541122	.0917328
ilkokul	.0756813	.0137405	5.51	0.000	.0487503	.1026123
ortaokul	.1024944	.0174031	5.89	0.000	.0683851	.1366038
lise	.1540978	.0165466	9.31	0.000	.1216671	.1865285
universite	.1712955	.0176694	9.69	0.000	.1366642	.2059269
gel2	.0527746	.0100514	5.25	0.000	.0330741	.072475
gel3	.0853256	.0132134	6.46	0.000	.0594279	.1112233
gelkarsilama	.0633131	.0049684	12.74	0.000	.0535751	.0730511
sos_hayat	.0986802	.0043395	22.74	0.000	.090175	.1071855

Tablo 5.10 incelendiğinde, diğer değişkenler sabit tutulduğunda bireylerin yaşı arttıkça yaşamdan memnuniyetin azaldığı görülmektedir. Bir başka ifadeyle, birey bir yaş daha yaşlandığında, yaşamdan hiç memnun olmama kategorisindeki olasılığını %0.10 ve orta derecede yaşamdan memnun olma kategorisinde olma olasılığını %0.0055 arttırırken; yaşamdan memnun olma kategorisinde olma olasılığını %0.11 azaltmaktadır.

Bireylerin geliriyle hanenin temel ihtiyaçlarını karşılama düzeyi arttıkça daha fazla yaşamdan memnun olmaktadır. Daha açık anlatımla, bireylerin gelirleriyle hanenin ihtiyaçlarını karşılaması kolaylaştıkça yaşamdan memnun olmama kategorisindeki olasılığı %6 ve orta derecede yaşamdan memnun olma kategorisindeki olasılığı %0.31 azalırken, yaşamdan memnun olma kategorisindeki olasılığı %6.33 artmaktadır.

Sosyal hayatınızdan memnun musunuz sorusuna verilen yanıtlar incelendiğinde; bireylerin sosyal hayatlarından memnuniyet düzeyleri arttıkça yaşamdan memnun olmama olasılığını %9.3 ve yaşamdan orta düzeyde memnun olma olasılığı %0.48 azalırken, yaşamdan memnun olma olasılığı ise %9.8 artmaktadır.

Erkeklerin kadınlara göre yaşamdan memnun olmama olasılıkları 0.026 ve orta derecede memnun olma olasılığı 0.0014 daha fazla iken erkeklerin yaşamdan memnun olma kategorisindeki olasılığı kadınlara göre 0.028 daha azdır. Ayrıca, orta gelirli bireylerin düşük gelirli bireylere göre yaşamdan memnun olmama olasılığı %5 ve orta derecede memnun olma olasılığı %0.26 daha az iken, bununla birlikte orta gelirli bireylerin düşük gelirli bireylere göre yaşamdan memnun olma olasılığı %5.2 daha fazladır. İlkokul mezunu bireylerin eğitimsiz bireylere göre yaşamdan memnun olmama olasılığı %7.1 daha az iken, ortaokul, lise ve üniversite eğitim seviyelerine sahip bireylerin yaşamdan memnun olmama olasılıkları sırasıyla %9.7, %14.6 ve %16.2 daha azdır. Bununla birlikte, ilkokul mezunu bireylerin eğitimsiz bireylere göre yaşamdan memnun olma olasılığı %7.5 daha fazla iken, ortaokul, lise ve üniversite eğitim seviyelerine sahip bireylerin yaşamdan memnun olma olasılıkları sırasıyla %10.2, %15.4 ve %17.1 daha fazladır.

5.5.3. Paralel Regresyon Varsayımı Testi

Sıralı modeller için paralel regresyon varsayımı testi yapılması gerekmektedir. Paralel regresyon varsayımı $\beta_1 = \beta_2 = \dots \beta_{j-1}$ olduğunu ifade etmektedir. Bir başka ifadeyle; ilgili regresyon katsayılarının bağımlı değişkenin tüm kategorilerinde aynı olduğu hipotezi test edilmektedir. Paralel regresyon varsayımını test etmek için Stata programında “**oparallel**”, “**brant**” ve “**omodel**” komutları kullanılmaktadır. Bu komutlar “**net search komut adı**” yazılarak Stata programına indirilebilir. Testlerin gerçekleştirilebilmesi için kısıtsız ve kısıtlı modeller hesaplanmalıdır (Long ve Freese, 2014: 328). Aşağıda verilen hipoteze ait test sonuçları sırasıyla “**oparallel**”, “**omodel**” ve “**brant, detail**” komutları kullanılarak analiz edilecektir.

H_0 : İlgili regresyon katsayıları bağımlı değişkenin tüm kategorilerinde aynıdır.

H_1 : İlgili regresyon katsayıları bağımlı değişkenin tüm kategorilerinde farklıdır.

Stata programında “**oparallel**” komutu ile Wolfe Gould, Brant, Score, Likelihood ratio ve Wald test sonuçları elde edilebilmektedir.

“**oparallel**” komutu ile elde edilen her bir test sıralı logit modeller için hesaplanırken oprobit model için hesaplanamamaktadır.

“**oparallel komutu**” test sonucu Tablo 5.11’de verilmiştir. Test komutunun yanında yer alan **ic** kısaltması bilgi kriterleri “**information criteria**” sonuçlarını göstermektedir.

Tablo 5.11. oparallel test sonucu

ologit memnuniyet cinsiyet yas evli ilkokul ortaokul lise universite gel2 gel3
gelkarsilama sos_hayat,nolog

oparallel, ic

Tests of the parallel regression assumption

	Chi2	df	P>Chi2
-----+-----			

Wolfe Gould	37.95	11	0.000
Brant	41.63	11	0.000
score	38.52	11	0.000
likelihood ratio	37.92	11	0.000
Wald	38.54	11	0.000

Information criteria

	ologit	gologit	difference
-----+-----			
AIC	19322.24	19306.32	15.92
BIC	19415.40	19478.30	-62.91

Verilen sonuçlarda “likelihood ratio test” ve “Wald” testi genelleştirilmiş logit modele dayanmaktadır. “Wolfe Gould” testi yaklaşık LR testi, “Brant” Brant test sonucunu ve “score” ise Skor test sonucunu göstermektedir. Tablo 5.11’deki tüm test sonuçlarına göre H_0 hipotezi reddedilmektedir ($p < 0.01$). Paralel regresyon varsayımının sağlanmadığı ifade edilebilir. Ancak paralellik varsayımı sağlanmış kabul edilerek sıralı logit model yorumları yapılmıştır. Ayrıca AIC ve BIC istatistikleri, modellerin uyumunu değerlendirirken genelleştirilmiş modelle arasındaki değişimi değerlendirmek için kullanılabilir. Tablo 5.11’de verilen AIC değerleri karşılaştırıldığında; her iki modelin AIC değerleri arasındaki fark fazla olmasa da daha küçük AIC değerine sahip olan paralel regresyon varsayımı gevşetilebilen genelleştirilmiş logit (gologit) modelinin daha uygun olduğunu göstermektedir. Konunun ilerleyen kısmında genelleştirilmiş sıralı logit model uygulanacaktır.

Sıralı probit modeller için “**oparallel**” komutu kullanılmamaktadır, fakat Wolfe Gould testi olan “yaklaşık LR” test sonucu elde edilebilmektedir. Yaklaşık LR testi oprobit ya da ologitten gelen log olabilirlik, $j-1$ ikili logit yada probit modellerin havuzlanması ile elde edilen olabilirlik değerlerini karşılaştırarak ve $y \leq m$ ile tanımlanan ikili sonuçlar arasındaki korelasyon için bir düzenleme yaparak hesaplanmaktadır. “**omodel**” komutu (Wolfe ve Gould, 1980) ile gerçekleştirilebilir. Tablo 5.12’de sırası ile ologit ve oprobit modelleri tekrar tahmin edildikten sonra bulunan yaklaşık LR testi sonucu verilmiştir.

Tablo 5.12. Yaklaşık LR Test Sonucu

```
omodel logit memnuniyet cinsiyet yas evli ilkokul ortaokul lise universite gel2 gel3
gelkarsilama sos_hayat
```

(Model çıktısı için bkz. Tablo 5.5)

Approximate likelihood-ratio test of proportionality of odds
across response categories:

```
chi2(11) =      37.95
Prob > chi2 =    0.0001
```

```
omodel probit memnuniyet cinsiyet yas evli ilkokul ortaokul lise universite gel2 gel3
gelkarsilama sos_hayat
```

Model çıktısı için bkz. Tablo 5.9’da verilmişti.

Approximate likelihood-ratio test of equality of coefficients
across response categories:

```
chi2(11) =      38.49
Prob > chi2 =    0.0001
```

Tablo 5.12’de verilen ki-kare değerleri Tablo 5.11’deki Wolfe- Gould test sonuç ile aynıdır. Paralel regresyon varsayımı 0.01 önem düzeyinde yine reddedilmektedir ($p < 0.01$). Elde edilen test sonucuna göre bağımlı değişkenin her bir kategorisi için tahminlenen regresyon parametrelerinin aynı olmadığı sonucuna ulaşılmaktadır.

“**oparallel**” komutu ile her bir değişken için ayrı ayrı paralel regresyon testi sonuçları elde edilememektedir. Ancak “**brant, detail**” komutu ile bireysel olarak her değişkenin katsayısı için H_0 hipotezi test sonucu verilmektedir. Brant (1990), her bir değişkenin katsayısı için ayrı ayrı paralellik varsayımı testinin gerçekleştirilebileceğini belirtmiştir. Tablo 5.13’te Brant testine ait sonuçlar verilmiştir.

Tablo 5.13. Brant Test Sonucu

ologit memnuniyet cinsiyet yas evli ilkokul ortaokul lise universite gel2 gel3
gelkarsilama sos_hayat,nolog

Model çıktısı Tablo 5.5’de verilmişti.

brant, detail

Estimated coefficients from j-1 binary regressions

	y>1	y>2
cinsiyet	-.17524443	-.10534147
yas	-.00239003	-.00838188
evli	.37697365	.3471988
ilkokul	.37110085	.44691269
ortaokul	.49801653	.58951262
lise	.71255256	.86223553
universite	.65753316	1.0256004
gel2	.28432223	.23231131
gel3	.4759262	.3652938
gelkarsilama	.31111462	.30738177
sos_hayat	.48549945	.47613478
_cons	-1.9908052	-3.7457083

Brant Test of Parallel Regression Assumption

Variable	chi2	p>chi2	df
-----+-----			
All	41.63	0.000	11
-----+-----			
cinsiyet	1.58	0.209	1
yas	8.61	0.003	1
evli	0.22	0.642	1
ilkokul	0.59	0.443	1
ortaokul	0.57	0.450	1
lise	1.68	0.195	1
universite	9.18	0.002	1
gel2	0.60	0.438	1

gel3	1.51	0.219	1
gelkarsilama	0.01	0.911	1
sos_hayat	0.10	0.750	1

A significant test statistic provides evidence that the parallel regression assumption has been violated.

Tablo 5.13 Brant test sonucu incelendiğine; yaş ve eğitim değişkeninin ve üniversite kategorisine ait katsayıların test sonuçlarına bakıldığında H_0 hipotezi 0.01 önem düzeyinde reddedilmektedir ($p < 0.01$). Bu sonuçlar varsayımın sağlanmadığını ologit modelinin geçerli olmadığını göstermektedir. Ayrıca tüm parametreler için $\beta_1 = \beta_2 = \dots \beta_{j-1}$ olduğunu belirten test sonucu ki-kare değeri 41.63 değerine göre de H_0 hipotezi reddedilmektedir. Paralel eğriler varsayımı sağlanmamaktadır. Paralel regresyon varsayımı sağlanmadığında paralel eğimler varsayımı gerektirmeyen alternatif modeller ele alınmalıdır. Bu modeller paralel regresyon varsayımının aranmadığı genelleştirilmiş sıralı logit ve sıralı probit modelleridir (Greene ve Hensher, 2009: 151).

5.6. Genelleştirilmiş Sıralı Logit Model

Genelleştirilmiş sıralı logit model, paralel eğriler varsayımı sağlanmadığında sıralı logit modele alternatif olarak geliştirilmiştir. Paralellik varsayımının sağlanmadığı durumlarda sıralı logit modelin sonuçlarına güvenilmemektedir. Bu durumda tahmin için çok durumlu logit model de uygulanabilir. Fakat bu uygulamada bağımlı değişkenin sıralı yapısı göz ardı edilip sadece kategorik olduğu düşünülmekte ve hatalı sonuçlar alınabilmektedir. Böylece alternatif model olarak, varsayım şartının aranmadığı genelleştirilmiş logit model kullanılmaktadır (Greene ve Hensher, 2009: 153-154).

Genelleştirilmiş sıralı logit model yorumu yapılırken bağımlı değişkenin kategorileri iki boyutlu indirgenerek, ikili logit model katsayıları olarak yapılmaktadır. Yaşam memnuniyeti bağımlı değişkeni 1- memnun değilim, 2- orta derecede memnun ve 3- memnun şeklinde olduğunda ilk değerlendirme, bağımlı değişkenin ilk kategorisi memnun değilim kategorisine karşılık orta derecede memnun ve memnun olma kategorileri olarak kodlanan bir ikili logit modele göre, ikinci değerlendirme ise bağımlı değişkenin memnun değilim ve orta derecede memnun olma kategorilerine karşılık gelen memnun olma kategorisi olarak kodlanan diğer bir ikili logit modele göre yapılır (Williams, 2006: 59).

M kategorili sıralı yapıdaki bağımlı değişken olduğu varsayıldığında bu durumda, genelleştirilmiş sıralı logit model için oluşturulan m-1 adet birikimli logit eşitlik (5.41)'deki gibidir (Fullerton ve Xu, 2012: 183).

$$\ln \left(\frac{P(Y \leq m|x)}{P(Y > m|x)} \right) = \tau_m - x\beta_m \quad 1 \leq m < M \quad (5.41)$$

Bağılı değişken 3 kategorili olduğu durumda oluşturulan her bir birikimli logit için 2 ayrı kesim noktası (eşik değer) bulunmaktadır. Açıklayıcı değişkenin bağımlı değişkene etkisi her

logitte farklıdır. Bağımlı değişkenin 1. kategorisini 2, 3. kategorileri ile karşılaştıran logitte açıklayıcı değişkenin etkisi β_1 vektörü ile 1 ve 2. kategorisini 3. kategori ile karşılaştıran logitte açıklayıcı değişkenin etkisi β_2 vektörü ile ifade edilmektedir. Dolayısıyla oluşturulan birikimli logitlerin bağımlı değişkene etkisi bağımlı değişkenin her kategorisinde birbirinde farklıdır. Bu durum paralel eğriler varsayımının karşılanmadığının ifadesidir.

Genelleştirilmiş sıralı logit model tahminin komutu Stata'da “**gologit2**” şeklindedir ve çalıştırmak için “**findit gologit2**” komutunun yüklenmesi gerekir.

Tablo 5.14. Genelleştirilmiş Sıralı Logit Model Tahmini

gologit2 memnuniyet cinsiyet yas evli ilkokul ortaokul lise universite gel2 gel3 gelkarsilama sos_hayat

Generalized Ordered Logit Estimates Number of obs = 9,566
 LR chi2(22) = 1539.52
 Prob > chi2 = 0.0000
 Pseudo R2 = 0.0740
 Log likelihood = -9629.1604

 Coef. Std. Err. z P>|z| [95% Conf. Interval]
 -----+-----

1 Yaşamdan ‘‘Memnun Olmama’’ya Karşılık ‘‘Orta Derecede Memnun ve Memnun Olma’’

cinsiyet	-.1685969	.0501207	-3.36	0.001	-.2668317	-.0703621
yas	-.0026555	.0017706	-1.50	0.134	-.0061259	.0008149
evli	.3831703	.0560759	6.83	0.000	.2732636	.493077
ilkokul	.364885	.0761968	4.79	0.000	.2155421	.5142279
ortaokul	.4881926	.0994909	4.91	0.000	.2931941	.6831912
lise	.7120951	.0956673	7.44	0.000	.5245906	.8995996
universite	.6637285	.1033004	6.43	0.000	.4612634	.8661936
gel2	.2786597	.0611976	4.55	0.000	.1587146	.3986048
gel3	.4680861	.0858624	5.45	0.000	.2997989	.6363733
gelkarsilama	.3142622	.0300804	10.45	0.000	.2553058	.3732187
sos_hayat	.4865506	.0251597	19.34	0.000	.4372384	.5358627
_cons	-1.993014	.1450684	-13.74	0.000	-2.277343	-1.708685

2 Yaşamdan ‘‘Memnun Olmama ve Orta Derecede Memnun Olma’’ya Karşılık ‘‘Memnun Olma’’

cinsiyet	-.1106372	.048156	-2.30	0.022	-.2050211	-.0162532
yas	-.0084748	.0017851	-4.75	0.000	-.0119735	-.0049761
evli	.3373452	.0569627	5.92	0.000	.2257003	.4489901
ilkokul	.4327672	.0925556	4.68	0.000	.2513615	.6141729
ortaokul	.5643932	.1108977	5.09	0.000	.3470376	.7817487
lise	.8464696	.1046941	8.09	0.000	.641273	1.051666
universite	1.007106	.108866	9.25	0.000	.7937329	1.22048
gel2	.2325743	.0570951	4.07	0.000	.12067	.3444786
gel3	.3644533	.071839	5.07	0.000	.2236514	.5052552
gelkarsilama	.3118836	.028992	10.76	0.000	.2550604	.3687069
sos_hayat	.473789	.0266434	17.78	0.000	.4215689	.5260092
_cons	-3.72084	.1602366	-23.22	0.000	-4.034898	-3.406782

Tablo 5.14'te elde edilen Modelde 1 ile gösterilen “memnun değil (1) kategorisine karşı orta derece memnun (2) ve memnun olanları (3)” göstermektedir. Burada pozitif ve anlamlı katsayı tahminleri bağımsız değişkendeki bir birim artışın memnun değil kategorisinden orta derecede memnun ve memnun olmaya geçiş olasılığının arttığını, negatif ve anlamlı katsayılar da memnun değil kategorisinde kalma olasılığının arttığını ifade eder. Memnun değil ve orta derecede memnun olanlara karşı memnun olanlar incelendiğinde, katsayı işaret ve anlamlılıklarıyla ilgili yorumlar benzer şekilde yapılır. Aşağıda genelleştirilmiş sıralı logit model kurulacak ancak bu kez hangi değişkenlerin paralellik varsayımını ihlal ettiği görülecektir. Modelde 1 ve 2 ile gösterilen kısımda değişkenlere ait aynı katsayılar paralellik varsayımını sağlayanları gösterirken, farklı olan katsayılar paralellik varsayımını bozanları ifade etmektedir. Bu sonuçları görebilmek için modele “**autofit lforce**” ifadesi eklenmelidir.

```
gologit2 memnuniyet cinsiyet yas evli ilkokul ortaokul lise universite sos_hayat  
gelkarsilama, autofit lforce
```

Bu komutu çalıştırmak araştırmacıya bırakılmıştır. Model aşağıda sadece fark oranlarıyla tahmin edilecektir.

Tablo 5.15. Genelleştirilmiş Sıralı Logit Model –Paralellik Varsayımı ve Wald Testi-

```
gologit2 memnuniyet cinsiyet yas evli ilkokul ortaokul lise universite gel2 gel3  
gelkarsilama sos_hayat ,or autofit lforce
```

Testing parallel lines assumption using the .05 level of significance...

```
Step 1: Constraints for parallel lines imposed for gelkarsilama (P Value = 0.9434)
Step 2: Constraints for parallel lines imposed for sos_hayat (P Value = 0.6483)
Step 3: Constraints for parallel lines imposed for ortaokul (P Value = 0.5367)
Step 4: Constraints for parallel lines imposed for ilkokul (P Value = 0.7188)
Step 5: Constraints for parallel lines imposed for evli (P Value = 0.5439)
Step 6: Constraints for parallel lines imposed for gel2 (P Value = 0.5079)
Step 7: Constraints for parallel lines imposed for lise (P Value = 0.3280)
Step 8: Constraints for parallel lines imposed for gel3 (P Value = 0.3306)
Step 9: Constraints for parallel lines imposed for cinsiyet (P Value = 0.1893)
Step 10: Constraints for parallel lines are not imposed for
        yas (P Value = 0.00004)
        universite (P Value = 0.00094)
```

Wald test of parallel lines assumption for the final model:

```
( 1) [1]gelkarsilama - [2]gelkarsilama = 0
( 2) [1]sos_hayat - [2]sos_hayat = 0
( 3) [1]ortaokul - [2]ortaokul = 0
( 4) [1]ilkokul - [2]ilkokul = 0
( 5) [1]evli - [2]evli = 0
( 6) [1]gel2 - [2]gel2 = 0
( 7) [1]lise - [2]lise = 0
( 8) [1]gel3 - [2]gel3 = 0
( 9) [1]cinsiyet - [2]cinsiyet = 0
      chi2( 9) = 5.15
      Prob > chi2 = 0.8210
```

An insignificant test statistic indicates that the final model does not violate the proportional odds/ parallel lines assumption

If you re-estimate this exact same model with gologit2, instead of autofit you can save time by using the parameter

```
pl(gelkarsilama sos_hayat ortaokul ilkokul evli gel2 lise gel3 erkek)
```

```
-----
Generalized Ordered Logit Estimates      Number of obs   =      9,566
                                         LR chi2(13)     =     1534.36
                                         Prob > chi2     =      0.0000
Log likelihood = -9631.7416              Pseudo R2       =      0.0738
-----
```

```
( 1) [1]gelkarsilama - [2]gelkarsilama = 0
( 2) [1]sos_hayat - [2]sos_hayat = 0
( 3) [1]ortaokul - [2]ortaokul = 0
( 4) [1]ilkokul - [2]ilkokul = 0
( 5) [1]evli - [2]evli = 0
( 6) [1]gel2 - [2]gel2 = 0
( 7) [1]lise - [2]lise = 0
( 8) [1]gel3 - [2]gel3 = 0
( 9) [1]cinsiyet - [2]cinsiyet = 0
-----
```

	memnuniyet	Odds Ratio	Std. Err.	z	P> z	[95% Conf. Interval]
1 Yaşamdan ‘‘Memnun Olmama’’ya Karşılık ‘‘Orta Derecede Memnun ve Memnun Olma’’						
cinsiyet		.8715701	.0353802	-3.39	0.001	.804913 .9437473
yas		.9978921	.001687	-1.25	0.212	.9945912 1.001204
evli		1.434317	.0670917	7.71	0.000	1.308667 1.572031
ilkokul		1.468049	.0988996	5.70	0.000	1.286461 1.675268
ortaokul		1.665401	.1417573	5.99	0.000	1.409501 1.967761
lise		2.157487	.1747401	9.49	0.000	1.840803 2.528652
universite		2.028052	.1942442	7.38	0.000	1.68094 2.446843
gel2		1.291875	.063021	5.25	0.000	1.174077 1.421491
gel3		1.497773	.0956976	6.32	0.000	1.321478 1.697587
gelkarsilama		1.36788	.0332628	12.88	0.000	1.304216 1.434653
sos_hayat		1.618085	.0345186	22.56	0.000	1.551824 1.687174
_cons		.133538	.017426	-15.43	0.000	.1034016 .1724576

2 Yaşamdan ‘‘Memnun Olmama ve Orta Derecede Memnun Olma’’ya Karşılık ‘‘Memnun Olma’’						
cinsiyet		.8715701	.0353802	-3.39	0.001	.804913 .9437473
yas		.9909264	.0016731	-5.40	0.000	.9876526 .994211
evli		1.434317	.0670917	7.71	0.000	1.308667 1.572031
ilkokul		1.468049	.0988996	5.70	0.000	1.286461 1.675268
ortaokul		1.665401	.1417573	5.99	0.000	1.409501 1.967761
lise		2.157487	.1747401	9.49	0.000	1.840803 2.528652
universite		2.553029	.2303451	10.39	0.000	2.139226 3.046876
gel2		1.291875	.063021	5.25	0.000	1.174077 1.421491
gel3		1.497773	.0956976	6.32	0.000	1.321478 1.697587
gelkarsilama		1.36788	.0332628	12.88	0.000	1.304216 1.434653
sos_hayat		1.618085	.0345186	22.56	0.000	1.551824 1.687174
_cons		.0250825	.003306	-27.96	0.000	.0193722 .032476

Tablo 5.15 incelendiğinde yaş değişkeninin katsayısı ($p=0.00004<0.01$) ve eğitim değişkenlerinden üniversite kategorisinin katsayısı ($p=0.00094<0.01$) incelendiğinde, modelin iki kısmında da farklı olduğu görülür. Diğer bir deyişle söz konusu iki katsayı, paralellik varsayımını bozmaktadır. Çıktının 1-8. adımındaki değişkenlerin katsayıları incelendiğinde modelin iki kısmında da parametrelerin aynı olduğu yani paralellik varsayımını bozmadıkları görülmektedir. Zaten bu sonuçlar Wald testiyle de onaylanmaktadır. Test sonucu 5.15 olup ($p = 0.8210 > 0.01$) sıfır hipotezi reddedilmektedir. Modelde 1 ve 2 nolu kısımda fark oranları aynı olan katsayılar sıralı logit modelindeki gibi yorumlanmaktadır. İlk olarak varsayımın bozulmasına neden olan modelin birinci kısmındaki katsayı tahmini yorumlanacak olursa üniversite mezunu olanların yaşamlarından memnun olmama olasılığı, (orta derece memnun ve memnun olanların olasılığına göre) okul bitirmeyen bireylere göre 2 kat daha fazladır şeklinde yorumlanır. Bu kez modelin ikinci kısmındaki katsayı yorumlanacak olursa, yine üniversite mezunu olanların yaşamından memnun olma ve kendini orta derece memnun hissetmesi olasılığı (memnun olması olasılığına göre) okul bitirmeyen bireylere göre 2.5 kat daha fazladır şeklinde yorum yapılır. Yani, eğitim seviyesi arttıkça yaşamdan duyulan memnuniyet düzeyindeki değişme de görülmektedir.

Ayrıca hem birinci hem de ikinci kısımda aynı olan, diğer bir deyişle paralellik varsayımını sağlayan katsayının yorumu; bireylerin kazandığı gelirle temel ihtiyaçlarını karşılaması kolaylaştıkça yaşamlarından memnun olma olasılığı, kendini orta seviyede memnun hisseden ve hiç memnun olmayanların olasılıklarına göre 1.3662 kat daha fazladır” şeklindedir. Dikkat edilecek olursa bulunan sonuç Tablo 5.6’ daki katsayı ile aynıdır.

Genelleştirilmiş logit modelin marjinal etkilerin tahmini, sıralı logit ve probit modelleri gibi yapılmaktadır. Model tahmin edildikten sonra marjinal etkiler komutu çalıştırılmalıdır.

Komutlar

describe

sum memnuniyet cinsiyet yas evli egitimsiz ilkokul ortaokul lise universite gel1 gel2
gelkarsilama sos_hayat

tab memnuniyet

ologit memnuniyet cinsiyet yas evli ilkokul ortaokul lise universite gel2 gel3 gelkarsilama
sos_hayat, nolog

ologit memnuniyet cinsiyet yas evli ilkokul ortaokul lise universite gel2 gel3 gelkarsilama
sos_hayat, or

quietly ologit memnuniyet cinsiyet yas evli ilkokul ortaokul lise universite gel2 gel3
gelkarsilama sos_hayat

margins, dydx(*) predict(outcome(#1)) atmeans

margins, dydx(*) predict(outcome(#2)) atmeans

margins, dydx(*) predict(outcome(#3)) atmeans

oprobit memnuniyet cinsiyet yas evli ilkokul ortaokul lise universite gel2 gel3 gelkarsilama
sos_hayat, nolog

quietly oprobit memnuniyet cinsiyet yas evli ilkokul ortaokul lise universite gel2 gel3
gelkarsilama sos_hayat

predict residual

sktest residual

quietly oprobit memnuniyet cinsiyet yas evli ilkokul ortaokul lise universite gel2 gel3
gelkarsilama sos_hayat

margins, dydx(*) predict(outcome(#1)) atmeans

margins, dydx(*) predict(outcome(#2)) atmeans

margins, dydx(*) predict(outcome(#3)) atmeans

ologit memnuniyet cinsiyet yas evli ilkokul ortaokul lise universite gel2 gel3 gelkarsilama
sos_hayat,nolog

oparallel, ic

omodel logit memnuniyet cinsiyet yas evli ilkokul ortaokul lise universite gel2 gel3 gelkarsilama
sos_hayat

omodel probit memnuniyet cinsiyet yas evli ilkokul ortaokul lise universite gel2 gel3
gelkarsilama sos_hayat

ologit memnuniyet cinsiyet yas evli ilkokul ortaokul lise universite gel2 gel3 gelkarsilama
sos_hayat,nolog

brant, detail

gologit2 memnuniyet cinsiyet yas evli ilkokul ortaokul lise universite gel2 gel3 gelkarsilama
sos_hayat

gologit2 memnuniyet cinsiyet yas evli ilkokul ortaokul lise universite gel2 gel3 gelkarsilama
sos_hayat ,or autofit lforce

5.7. KAYNAKLAR

- Agresti, A. (2010). Analysis of Ordinal Categorical Data, 2nd Edition, Wiley and Sons, Canada
- Bender, R. ve Grouven, U. (1997). Ordinal logistic regression in medical research, *Journal of The Royal College of Physicians of London*, 31(5), 546-551.
- Borooah, V.K.(2001). Logit and Probit Ordered and Multinomial Models, Sage Publications. London.
- Brant, R. (1990). Assessing Proportionality in the Proportional Odds Model for Ordinal Logistic Regression. *Biometrics*, 46, 1171-1178
- Demaris, A.(2004). Regression with Social Data, Modeling Continous and Limited Response Variable, Wiley and Sons, Canada.
- Greene, W. ve Hensher, D.A. (2009). Modeling Ordered Choices: A Primer, Cambridge University Press https://www.researchgate.net/publication/227388830_Modeling_Ordered_Choices_A_Prime .

- Greene, W. (2007) *Econometrics Analysis*. 6th. Edition. The McGraw-Hill Companies, NewYork.
- Gujarati, D. (2011). *Econometrics by Example*, Palgrave Macmillian,
- Fullerton, A. S., ve Xu, J. (2012). The proportional odds with partial proportionality constraints model for ordinal response variables. *Social science research*, 41(1), 182-198.
- Fullerton, A.S ve Xu, J. (2016). *Ordered Regression Models, Parallel, Partial and Non-Parallel Alternatives*, CRC Press.
- Katchova, A. *Ordered Probit and Logit Model*,
<https://sites.google.com/site/econometricsacademy/econometrics-models/ordered-probit-and-logit-models> (Erişim Tarihi :17.09.2020).
- Liu, X. (2016). *Applied Ordinal Logistic Regression Using Stata*. U.S.A.: Sage Publications.
- Liao, T.F. 81994). *Interpreting Probability Models Logit, Probit and Generalized Linear Models*, Sage publications, London.
- Long, J. S. (1997). *Regression Models for Categorical and Limited Dependent Variable*, Sage Publications. California.
- Long, J.S. ve Freese, J. (2014). *Regression Models for Categorical Dependent Variables Using Stata*, 3th.Edition, Stata Press, US.
- McCullagh, P. (1980). Regression Models for Ordinal Data. *Journal of the Royal Statistical Society*, 42(2), 109-142.
- McKelvey, R.D. ve Zavoina, W. (1975). A Statistical Model for the Analysis of Ordinal Level Dependent Variable, *Journal of Mathematical Sociology*, 4(1), 103-120.
- O'Connell, A.A, (2006). *Logistic Regression Models for Ordinal Response Variables*, Sage Publication, London.
- Powers, D.A. ve Xie, Y. (1999). *Statistical Methods for Categorical Data Analysis*, Academic Press
- Simithson, M. Ve Merkle, E.C. (2014). *Generalized Linear Models for Categorical and Continous Limited Dependent Variable*, CRC Press.
- Türkiye İstatistik Kurumu (TÜİK). (2018). *Yaşam Memnuniyeti Araştırması*.
- Williams, R. (2005). *Gologit2: A Program for Generalized Logistic Regression/Partial Proportional Odds Models for Ordinal Dependent Variables*. <https://www.stata.com/meeting/4nasug/gologit2.pdf>. (20/02/2020).
- Williams, R. (2006). Generalized Ordered Logit/Partial Proportional Odds Models for Ordinal Dependent Variables. *The Stata Journal*. 6(1): 58-82.
- Williams, R. (2016). Understanding and Interpreting Generalized Ordered Logit Models. *The Journal Of Mathematical Sociology*. 40(1): 7-20.
- Winship, C., ve Mare, R. D. (1984). Regression models with ordinal variables. *American sociological review*, 49(4), 512-525.

BÖLÜM 6

ARDIŞIK LOGİT MODEL VE STATA UYGULAMALARI¹

Özlem Kiren Gürler²

Hatice Hicret Özkoç³

Bağımlı değişken bazı durumlarda kategorik, bazen de sıralı bir yapı göstermektedir. Kimi zaman bağımlı değişkenin sıra bilgisi kategorilerin ardışık yapıda olduğu durumda ortaya çıkmaktadır. Örneğin, (1) kısa, (2) orta ve (3) uzun vadeli işsizlik kategorileriyle bir sınıflandırma yapıp, bu kategorilerden birinde olmanın belirleyicileri araştırılır. Eğer birey orta vadeli işsiz kategorisindeyse, daha önce kısa vadeli işsiz kategorisinde bulunmuştur; uzun vadeli işsizlik kategorisine de kısa ve orta vadeli işsizlik sınıflandırmasından sonra ancak ulaşılmıştır. Bir başka örnek, birey üniversitede okuyorsa, daha önce sırayla ilköğretim ve liseyi bitirmiş demektir. Genelleme yapıldığında bireyin 3. kategoriye gelmesi için 1. ve 2. kategoriye tamamlamış olması gerekmektedir. Dolayısıyla bağımlı değişken aslında ardışık bir karar alma sürecini içermektedir. Bu bölümde, yanıt değişkenindeki sıralı ve ardışık bilgileri açıkça hesaba katan bu tür sıralı sonuç modelleri tartışılacaktır.

Ardışık (sequential) logit model ardışık-cevap(yanıt) modelleri, hiyerarşik cevap modelleri olarak da ifade edilmektedir. İki durumlu logit modellerinin uygulandığı durumlarda ardışık modellerde uygulanabilmektedir. Bazı durumlarda yanıt değişkenlerinin sonuçları düzeyli olmaktadır ve ayrık, ilgisiz kategorilerden oluşmamaktadır. Bir başka ifadeyle yanıt kategorileri aşamalı bir dizi olarak ortaya çıkmaktadır.

Ardışık modellerde yanıt değişkeni düzeyleri aşamalı bir sıralamaya sahiptir ve bir sonraki aşamadaki yanıt düzeyi bir önceki aşamadaki yanıt düzeyinde yuvalanmıştır (Liao, 1994:25). Çok durumlu yanıt modellerinin tahmini daha az sayıda yanıt içeren ardışık model tahminine indirgelebilmektedir (Amemiya, 1981:1524).

Maddala (1983)'de sosyal bilimler literatüründe çok sık karşılaşılan ardışık yanıt modelleri için aşağıdaki örnekten bahsetmektedir. Eğitim düzeyi ile ilgili olan örnekte yanıt değişkeni $y_i = 1, 2, 3, 4$ değerlerini almaktadır. Bu değerler

$y = 1$, Eğer birey lise eğitimini bitirmemiş ise

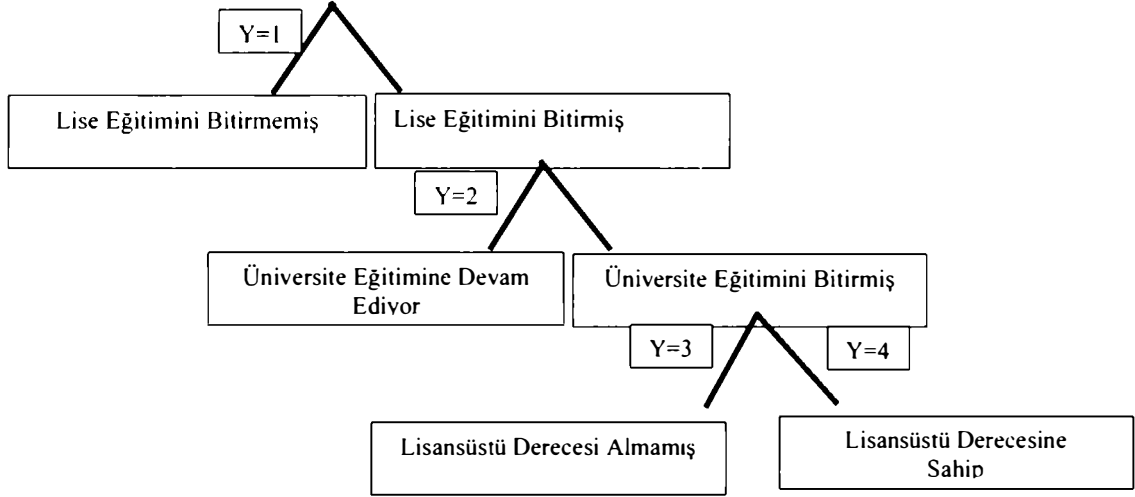
¹ Bu bölümün teorik kısımları ve verileri Özlem Kiren Gürler tarafından hazırlanmış olan "Türkiye'de Çocuk Yoksulluğu: Birleştirilmiş Yatay Kesit Verileriyle Tercih Modellerinin Uygulanması" başlıklı doktora tezin-den yararlanılarak oluşturulmuştur.

² Dr. Öğr. Üyesi, Dokuz Eylül Üniversitesi, İİBF, Ekonometri Bölümü, ozlem.kiren@deu.edu.tr

³ Doç. Dr., Muğla Sıtkı Kocaman Üniversitesi, İİBF, İşletme Bölümü, hatice.ozkoc@mu.edu.tr

$y = 2$, Eğer birey lise eğitimini bitirmiş, fakat üniversite eğitimine devam ediyor ise
 $y = 3$, Eğer birey üniversite eğitimini bitirmiş, fakat lisansüstü derecesi almamış ise
 $y = 4$, Eğer birey lisansüstü dereceye sahip ise
 şeklinde tanımlanmaktadır (Liao, 1994:25-26). Bu durum Şekil 6.1’de gösterilmiştir.

Şekil 6.1. Eğitim Düzeyleri Ardışık Yapı Örneği



Kaynak: (Maddala, 1983: 49)

Ardışık modellerin en önemli özelliği aşamalı yanıt (item) çözüm yöntemi kullanmasıdır. Bu modellerde adım adım çözme yönteminden yararlanılmaktadır. Modellerin çözümlerinde adımlar ardışıktır. Problem çözümünde gelecek adımın başarılı bir şekilde gerçekleşmesi, bir önceki adımın başarılı sonuçlanmasına bağlıdır. Bir düzeyden diğer bir düzeye geçiş modellenmektedir ve yanıt tahmincileri geçiş zorluğu açısından doğrudan yorumlanmaktadır. (Tutz, 1997:139; Winkelmann ve Boes, 2006: 195). Ardışık model yaklaşımında her bir karar süreci birbirini izleyen sırada gerçekleştiğinden bağımsız değişkenler istenirse süreçlerde değişebilir ve alternatifler de tek tek analiz edilebilir (Sakellario ve Lall, 1997; Cartwright ve Patrinos, 1997, Grootaert, 1997, 1998). Ardışık analizde modelden elde edilen olasılıklar bir önceki seçime bağlıdır. Başka bir ifadeyle, elde edilen sonuçlar seçeneklerin sırasına bağlıdır. İlgisiz alternatiflerin bağımsızlığı varsayımı gerekmemektedir.

6.1. Koşullu Geçişlerin Modellenmesi

Ardışık modellerde temel fikir, yanıt mekanizmasının ilk kategoride başlamasıdır. Birey ilk kategoride kalıp kalmamaya ya da bir üst düzeye geçmeye karar vermektedir. Eğitim örneğinde olduğu gibi birey lise eğitimini bitirdikten sonra üniversite eğitimine devam edebilir ya da üniversiteyi bitirdiyse diğer karar adımına geçebilir. İki durumlu karar, 0/1 şeklinde kodlanabilmektedir. $y_i^{(1)} = 1$ sürecin ilk kategoride durup durmadığını ifade etmektedir ve bu durumda $y_i = 1$ olarak gözlenmektedir. Eğer süreç devam ediyorsa $y_i^{(1)} = 0$ ’dır. Sürecin devamındaki koşul,

ikinci kategoriden üçüncü kategoriye geçiş $y_i = 2$ 'ye karşı $y > 2$ ile belirlenmektedir ve süreç $y_i^{(2)} = 1$ ise tamamlanmaktadır. Genel olarak j . kategoriden $j + 1$. kategoriye geçiş aşağıdaki gibi gösterilmektedir (Winkelmann ve Boes, 2006: 195):

$$y_i^{(j)} \begin{cases} 1 & \text{eğer süreç tamamlandıysa, } j+1.\text{kategoriye geçiş yok} \\ 0 & \text{eğer süreç devam ediyorsa, } j+1.\text{kategoriye geçiş var} \end{cases} \quad (6.1)$$

j . kategori, $j-1$. kategorinin reddedilme olasılığına göre seçilmektedir. Bu yüzden son geçişin modellenmesi gerekmektedir. Kısaca özetlenecek olursa geçişlerin koşullu modellendiği varsayılmaktadır, bir başka ifadeyle j . kategoriye ulaşıldığında sürecin devam edip etmeyeceğine ya da sonlandırılacak olup olmadığına kararı verilmektedir. Eğitim örneğinde olduğu gibi liseyi bitirmiş bir bireyin üniversite eğitimine devam edip etme ya da üniversite eğitimini bitirip sonraki karar adımına geçişi bir önceki adımlara bağlıdır.

Sıralı yanıtların koşullu olasılık modelinin formüle edilebilmesi için koşullu geçiş olasılıklarının belirlenmesi gerekmektedir. $y_i = j$ olayı olasılığı $y_i \geq j$ koşulluna bağlı olarak, aşağıdaki gibi gösterilebilmektedir:

$$P(y_i = j | y_i \geq j, x_i) = F(\alpha_j + x_i'\beta) \quad (6.2)$$

(6.2) denkleminde α_j her bir kategorideki sabit terim, $x_i'\beta$ doğrusal açıklayıcı değişken indeksi (sabit terim hariç) ve $F(\cdot)$, birim aralığa $\alpha_j + x_i'\beta$ 'yi eşleyen monoton örten bir dönüşümdür. Ardışık logit model için (6.2) eşitliğinin gösterimi (6.3) eşitliği ile verilmektedir:

$$F(\alpha_j + x_i'\beta) = \Lambda(\alpha_j + x_i'\beta) \quad (6.3)$$

Λ ; kümülatif lojistik yoğunluk fonksiyonudur. Koşullu geçiş olasılıkları belirlendikten sonra, koşulsuz olasılıklar $P(y_i = j | x_i)$, ($y_i \geq j$ üzerinde koşulsuz fakat x_i üzerinde hâlâ koşullu olasılıklar) tekrarlanan ilişkiler ile aşağıdaki gibi elde edilmektedir:

$$P_{ij} = P(y_i = j | x_i) = P(y_i = j | y_i \geq j, x_i) P(y_i \geq j | x_i) \quad (6.4)$$

Aslında, koşullu geçiş olasılıkları y_i olasılık fonksiyonunu tamamen tanımlamaktadır. Örneğin üç durumlu yanıt değişkeni için;

$$P(y_i = 1) = P(y_i = 1 | y_i \geq 1) P(y_i \geq 1) = P(y_i = 1 | y_i \geq 1)$$

$$\begin{aligned} P(y_i = 2) &= P(y_i = 2 | y_i \geq 2) P(y_i \geq 2) \\ &= P(y_i = 2 | y_i \geq 2) [1 - P(y_i = 1 | y_i \geq 1)] \end{aligned}$$

$$\begin{aligned} P(y_i = 3) &= P(y_i = 3 | y_i \geq 3) P(y_i \geq 3) \\ &= P(y_i = 3 | y_i \geq 3) [1 - P(y_i = 1) - P(y_i = 2)] \\ &= P(y_i = 3 | y_i \geq 3) [1 - P(y_i = 1 | y_i \geq 1)] [1 - P(y_i = 2 | y_i \geq 2)] \end{aligned}$$

bulunur. Olasılık gösterimlerinde x_i koşulu, basit gösterim sağlayabilmek için belirtilmemiştir. Genel olarak koşullu olasılıklar aşağıdaki gibi ifade edilmektedir:

$$P_{ij} = P(y_i = j | y_i \geq j, x_i) \prod_{r=0}^{j-1} [1 - P(y_i = r | y_i \geq r, x_i)]$$

$$= F(\alpha_j + x_i' \beta) \prod_{r=0}^{j-1} [1 - F(\alpha_r + x_i' \beta)] \quad (6.5)$$

(6.5) eşitliği her kategori β_j parametre vektörüne göre genelleştirilirse (6.6) nolu eşitlik elde edilir (Winkelmann ve Boes, 2006: 196):

$$P_{ij} = P(y_i = j | x_i) = F(\alpha_j + x_i' \beta_j) \prod_{r=0}^{j-1} [1 - F(\alpha_r + x_i' \beta_r)] \quad (6.6)$$

Burada $\alpha_0 = -\infty$ iken $F(-\infty) = 0$ dir. Ardışık modelin her bir kategorisinde iki durumlu logit model tahmini yapılmaktadır. Ardışık modelin en önemli özelliklerinden biri de, uygun bir olasılık fonksiyonunun varlığını sağlamak için parametre uzayı üzerinde herhangi bir kısıtlamanın olmamasıdır. Ardışık modellerde olasılıklar basit bir denklem şeklinde aşağıda ifade edilmektedir:

$$P_{ij} = P_i \cdot P_{j|i} \quad (6.7)$$

P_i , y_i sonucunun olasılığı, $P_{j|i}$ y_i sonucunun koşullu olasılığıdır. Ardışık modellerde önemli noktalardan biri, her bir aşamadaki tercih olasılığının, diğer aşamalardaki tercih olasılıklarından bağımsız olmasıdır. Başka bir ifadeyle, $y_1, y_2, \dots$ değişkenleri kavramsal olarak farklı olmalı ve istatistiksel olarak da birbirinden bağımsız olmalıdır (Liao, 1994:29).

6.2. Ardışık Logit Modellerinin Yorumlanması

Ardışık logit modeller, iki durumlu logit model serilerinden meydana geldiği için yorumlanması iki durumlu logit modellerin yorumlanmasına benzerdir. Önemli farklılıklar aşağıda açıklanmaktadır:

i. η 'deki Marjinal Etkiler veya Dönüştürülmüş η ⁴

Ardışık modelin her basamağı dönüştürülmüş ya da dönüştürülmemiş kendi η değerine (logite) sahiptir. Her bir β_k ve x_k setinde benzersiz (tek) η veya dönüştürülmüş η 'e karşılık gelen; $\beta_{k1}, x_{k1}; \beta_{k2}, x_{k2}; \beta_{k3}, x_{k3}, \dots$ sayısı ile bulunmaktadır. Bundan dolayı ardışık modelde dönüştürülmüş ya da dönüştürülmemiş η üzerindeki marjinal etki, her basamakta logit modelde yorumlandığı gibi yorumlanmaktadır.

⁴ $\eta, x_1, x_2, \dots, x_k$ değişkenleri tarafından elde edilen doğrusal tahmincidir. Lojistik model için bağlantı fonksiyonu (logit) $\eta = \log \left[\frac{P(y=1)}{1-P(y=1)} \right] = \sum_{k=1}^K \beta_k x_k$ şeklindedir.

Marjinal etki, açıklayıcı değişkendeki değişimin bağımlı değişkenin kategorilerden birisini seçme olasılığını nasıl değiştirdiğini ölçmektedir. Fark oranında x_k 'nın marjinal etkisi $\exp(\beta_k)$ olarak gösterilmektedir (Liao, 1994:29-30).

ii. Açıklayıcı Değişkenlerin Verilen Değerlerine Bağlı Olarak Tahmin Edilmiş Olasılıklar

Tahmin edilmiş olasılıklar iki durumlu logit modellerde hesaplandığı gibidir. Tek farkı; iki durumlu modellerde olasılıklar tek aşamada elde edilirken, ardışık modellerde tahmin edilmiş olasılıklar, ilgili aşamalardan elde edilmiş olasılıkların çarpımını içermektedir. Örneğin,

$$P(y_2 = 1) = P(y_1 = 1) \cdot P(y_2 = 1 | y_1 = 1) = \frac{e^{\sum_{k=1}^{K_1} \beta_{k1} x_{k1}}}{1 + e^{\sum_{k=1}^{K_1} \beta_{k1} x_{k1}}} \cdot \frac{e^{\sum_{k=2}^{K_2} \beta_{k2} x_{k2}}}{1 + e^{\sum_{k=2}^{K_2} \beta_{k2} x_{k2}}} \quad (6.8)$$

(6.8) eşitliğinde k_1 ve k_2 alt indisleri 1 ve 2 nolu adımlardaki x , bağımsız değişkenleri göstermektedir (Liao: 1994: 31).

iii. Bir Olayın Olasılığı Üzerindeki Marjinal Etki

Ardışık modelin olay olasılığındaki bağımsız değişkenin marjinal etkisini yorumlamak, iki durumlu sonuç modellerin sonuçlarının yorumlanması gibidir. Çünkü inceleme altındaki ardışık model genellikle bir dizi iki durumlu sonuç modellerinden oluşmaktadır. Bundan dolayı x_k 'ya göre ardışıkta genel olasılığın kısmi türevleri ele alınan her basamaktaki iki durumlu yanıt modelinin kısmi türevinden elde edilebilir.

Ardışık modelde marjinal olasılık etki (MPE) ilk kategoriden başlayarak hesaplanmaktadır.

$$MPE_{i1\ell} = f(\alpha_1 + x_i' \beta_1) \beta_{1\ell} \quad (6.9)$$

Burada $f(z) = dF(z)/dz$. Denklem (6.9) x 'in ℓ -elemanındaki bir birimlik artış karşısında olasılıktaki yaklaşık değişimi ifade etmektedir. Geri kalan kategori için $j = 2, \dots, J$ MPE (6.10) nolu denklemden gibi ifade edilmektedir:

$$MPE_{i1\ell} = f(\alpha_j + x_i' \beta_j) \beta_{j\ell} \prod_{r=1}^{j-1} [1 - F(\alpha_r + x_i' \beta_r)] - F(\alpha_j + x_i' \beta_j) \sum_{r=1}^{j-1} MPE_{ir\ell} \quad (6.10)$$

6.3. Ardışık Logit Model İçin Stata Uygulaması

Bu çalışmada kullanılan veri kaynağı bireylerin ve bunların oluşturduğu hanehalkının tüketimlerini, gelir düzeylerini, sosyo-ekonomik kesimler ve nüfus tabakalarına ve bölgelere göre ortaya çıkaran Türkiye İstatistik Kurumu (TÜİK) tarafından gerçekleştirilen "Hanehalkı Bütçe Anketleri"dir. 2002 yılından itibaren 2008 yılına kadar her yıl düzenli olarak uygulanan hanehalkı bütçe anketleri birlikte kullanılmıştır.

Tablo 6.1. Değişkenlerin Tanımları

Değişken	Değişken tanımları
degree	Çocukların kararları
p_harc_medyn	Yoksulluk çizgisi: hesaplanan harcamaya bağlı olarak hane yoksul ise 1, değilse 0
cinsiyet	Çocuk erkek ise 1, değilse 0 değerini alan kukla değişken
anne_egts	Bitirilen eğitim yılı olarak hanedeki annenin eğitimi düzeyi
baba_egts	Bitirilen eğitim yılı olarak hanedeki babanın eğitimi düzeyi
kent	Hane kentte yaşıyorsa 1, kırdı yaşıyorsa 0 değerini alan kukla değişken
coc_sayisi	Hanedeki çocuk sayısı

Stata uygulamasında, çocuk eğitimi ve yoksulluk arasındaki ilişki incelenmiştir. Her haneden, zorunlu eğitimin bittiği yaş temel alınarak 15-19 yaş arasında çocuklarla örneklem oluşturulmuştur. ILO, 18 yaş altındaki bireyleri çocuk saymaktadır. Türkiye’de 2006 yılından itibaren TÜİK tarafından yaş değişkeni sınıflandırılmış olarak elde edildiğinden çocuğun üst yaş sınırı 19 alınmıştır. Ayrıca çocukların bulunduğu hanelerde anne ve babaya ait sosyo-ekonomik ve demografik değişkenler kullanılmıştır. Yapılan analizlerde 2002-2008 verileri birleştirilmiştir. Birleştirilmiş veri setinde 15-19 yaş arasındaki 28717 çocuk analizde kullanılmıştır. Bu çalışmada çocuk iş gücü, 15-19 yaşlarında eğitim çağında olup çocuk nüfusunda ücret geliri elde eden, hanehalkının sahip olduğu toprak veya aile işletmelerinde çalışan çocukları ifade etmektedir. Kullanılan değişkenlere ait tanımlamalar Tablo 6.2’de verilmiştir.

Tablo 6.2. describe Komutu

describe degree p_harc_medyn erkek_coc anne_egts baba_egts kent coc_sayisi				
	storage	display	value	
variable name	type	format	label	variable label
degree	byte	%8.0g	degree	degree
p_harc_medyn	byte	%8.0g		
erkek_coc	byte	%8.0g		cins== 1.0000
anne_egts	byte	%8.0g		Anne_Egts
baba_egts	byte	%8.0g		Baba_Egts
kent	byte	%8.0g		Kent
coc_sayisi	byte	%8.0g		Coc_Sayisi

Tablo 6.3'te bağımlı değişkenden elde edilen yanıtlar, kendi içinde okuma ve çalışma durumuna göre şekillenen ardışık bir yapıyı göstermektedir.

Tablo 6.3. Çocukların Eğitime Devam ve İşgücüne Katılım Kararları (Bağımlı Değişken)

tab degree

degree	Freq.	Percent	Cum.
1-Sadece_okuyanlar	13,922	48.48	48.48
2- Hem_okuyan_hem_calisanlar	890	3.10	51.58
3- Sadece_calisan	5,647	19.66	71.24
4- Hem_calismayan_hem_okumayan	8,258	28.76	100.00
Total	28,717	100.00	

Eğitime devam eden herhangi bir işte çalışmayan çocukların oranı %48.48, hem okuyup hem çalışan %3.1, okula devam etmeyip sadece çalışanlar %19.66 ve herhangi bir işte çalışmayan ve okula devam etmeyen çocukların oranı ise %28.76'dır.

Çalışmadaki hanelere ait demografik özellikler ile çocukların bulunduğu hanedeki anne ve babalarının sosyo-ekonomik özellikleri Tablo 6.4'te gösterilmiştir.

Tablo 6.4. Tanımlayıcı İstatistikler

sum p_harc_medyn erkek_coc anne_egts baba_egts kent coc_sayisi

Variable	Obs	Mean	Std. Dev.	Min	Max
p_harc_medyn	28717	.2079953	.4058805	0	1
erkek_coc	28717	.5005397	.5000084	0	1
anne_egts	28341	4.093504	3.408291	0	17
baba_egts	26333	6.385752	3.584601	0	17
kent	28717	.7031723	.4568679	0	1
coc_sayisi	28717	3.374691	1.953067	0	17

Hanehalkının refah ve yoksulluk seviyelerinin belirlenmesinde farklı ölçüm yöntemleri kullanılmaktadır. Literatürde hanehalkının yoksulluk düzeyinin belirlenmesinde gelir yada tüketim bilgileri kullanılarak yapılan araştırmalar mevcuttur. Çalışmada çocuk yoksulluk düzeyinin belirlenmesi için yapılan ölçüm ve analizlerde hanehalkı tüketim harcamasına göre yoksulluk sınırı kullanılmıştır. Kullanılan nihai harcama ve gelir değişkenleri Eurostat ölçeğine göre yeniden düzenlenmiştir. Hanenin yıllık tüketim harcamasından çocuk geliri çıkarıldıktan sonra elde edilen harcamanın medyan değerinin yarısına göre yoksulluk çizgisi değişkeni oluşturulmuştur. Elde edilen yoksulluk çizgisi (p_harc_medyn) medyan değerinin altında harcamaya sahip olan haneler yoksul sayılmıştır. Bu amaçla iki durumlu oluşturulan bağımsız değişkende, yoksul hanelere 1, diğer hanelere 0 değeri verilmiştir. Buna göre yoksulluk sınırının altında olan hanelerin oranı %20'dir. Çalışmadaki kız ve erkek çocuk oranı eşittir (%50). Annenin eğitim yılı ortalaması 4'dür. Bu değer annelerin ilkökul mezuniyet düzeyine bile sahip olmadığını göstermektedir⁵. Babanın eğitim durumu incelendiğinde annelerde olduğu gibi sahip olunan eğitim seviyesi en çok ilkökul mezuniyetidir. Ayrıca hanelerde bulunan çocuk sayısı ise ortalama olarak yaklaşık 3'dür. Hanelerin %70'i kentte oturmaktadır.

6.4. Ardışık Model Sonuçları

Hanehalkı karar sürecinde çocuk iş gücü arzını ekonometrik olarak modellemek için birçok yöntem bulunmaktadır. Yapılan çalışmada, hanehalkı karar sürecinde çocukların eğitimlerine devam etmesi en iyi seçenek olarak ele alınırken, ikinci en iyi seçenek çocukların hem çalışıp hem de eğitimlerine devam etmesi durumudur. Bu seçenekleri çocukların sadece çalışma durumu ve eğitimine devam etmeyip aynı zamanda çalışmayan çocuklar takip etmektedir. Ardışık karar sürecinde ilk seçenekte ebeveynler çocuklarını okula gönderip göndermemeye karar verirken, ikinci seçenekte çocuklarını hem okutup hem de çalıştırıp çalıştırmamaya karar vermektedir. Üçüncü ve dördüncü seçeneklerde ise sırasıyla çocuklarını çalıştırıp çalıştırmamaya ve okutmayıp ev ile ilgili işlerle ilgilenip ilgilendirmemeye karar vermektedir. Sıralamadaki temel kriter şu şekilde ifade edilebilir: ilk aşamada çocuk refahı önemli iken, ikinci aşamada çocuğun geliri hanehalkı refahına katkıda bulunduğundan çocuğun okula gitme oranı azalabilmektedir. Eğer hanehalkı refahına çocuğun hem çalışıp hem okula gitmesi yerine sadece çalışmasının daha fazla katkı yaptığı düşünülürse, çocuğun hem çalışıp hem okuması tercihinden vazgeçilecektir. Bununla birlikte çocuğun zaman tahsisi, esas faaliyetinin ne olduğuna da bağlıdır. Çocuğun esas faaliyeti eğitime devam etmek, çalışmak ve okumak, sadece çalışmak ya da ne okumak ne de çalışmak olabilmektedir.

Çocuk iş gücü ve eğitim kararında nitel tercih modellerinden hangisinin kullanılacağı hanehalkı karar sürecinde okula devam etme ya da iş gücüne katılma kararlarından hangisinin daha uygun olduğuna bağlıdır. Wahba (2008), ardışık analizde seçilen ilk kategorinin çocuğun refahını etkileyen en önemli tercih olması gerektiğini belirtmektedir. Gelişmemiş yoksul ülkelerde yapılacak ilk tercihin çocuğun çalıştığı durumun olacağını ve bu sıranın farklı çalışmalarda değişebileceğini ifade etmiştir. Bununla birlikte Türkiye gibi gelişmekte olan bir ülkede eğitim çağındaki çocukların önceliğinin eğitim olduğu ve bu amaçla seçilecek ilk kategorinin çocuğun okula devam

⁵ Annelerin yaşları dikkate alındığında 5 yıllık ilkökul mezuniyeti açısından bu yorum yapılmıştır.

etme tercihi olması gerektiğidir. Ardışık analizde kullanılan bağımlı değişkenler aşağıdaki gibi gösterilmektedir:

P_1 = Çocuğun sadece eğitime devam etmesi olasılığı

P_2 = Çocuğun okula devam ederken çalışması olasılığı

P_3 = Çocuğun sadece çalışması olasılığı

P_4 = Çocuğun eğitime devam etmeyip çalışmaması olasılığı

Ardışık logit modelinde dört seçenek için olasılıklar aşağıdaki gibi belirlenmektedir:

$$P_1 = F(\beta_1'X)$$

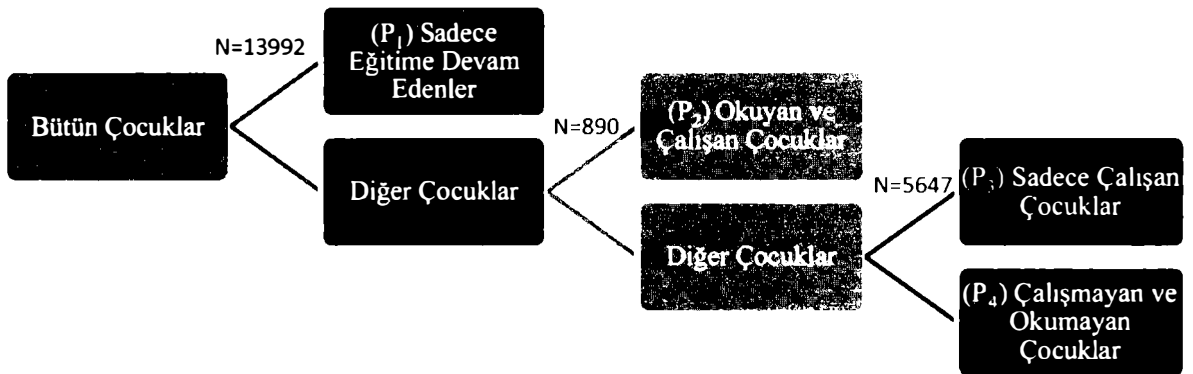
$$P_2 = [1 - F(\beta_1'X)] F(\beta_2'X)$$

$$P_3 = [1 - F(\beta_1'X)] [1 - F(\beta_2'X)] F(\beta_3'X)$$

$$P_4 = [1 - F(\beta_1'X)] [1 - F(\beta_2'X)] [1 - F(\beta_3'X)]$$

P_1 , P_2 , P_3 ve P_4 iki durumlu değişkenlerdir. F lojistik dağılım fonksiyonunu ifade ederken, β_1 , β_2 ve β_3 modeldeki parametre vektörlerini göstermektedir. X açıklayıcı değişken vektörüdür. β_1 parametresi bütün örneklem üzerinden hesaplanırken, β_2 parametresi sadece eğitime devam eden çocuklara ait gözlemler çıkartıldıktan sonra kalan gözlemler üzerinden tahminlenmektedir. β_3 parametresi ise sadece eğitime devam eden, hem eğitime devam edip hem de çalışan çocuklar çıkarıldıktan sonra kalan gözlemlerden tahminlenmektedir. Karar ağacı Şekil 6.2'de gösterilmiştir.

Şekil 6.2. 15-19 Yaş Arası Çocukların Eğitimlerine Devam Etme ve İş Gücüne Katılma Durumu



Stata programında ardışık logit model analizi “**seqlogit**” komutu ile gerçekleştirilmektedir. Bu komut aşağıdaki gibidir.

```
seqlogit depvar [indepvars] [if] [in] [weight] , tree(tree) [ ofinterest(varname)
over(varlist) x#(varlist) levels(#=#, #=#, etc.) sd(numlist) deltasd(varname numlist) rho(#) {
pr(numlist) | mn(# # , # # [, # # etc.) ] | uniform } draws(#) drawstart(#) or constra-
ints(numlist) robust cluster(clustervar) nolog level(#) maximize_options ]
```

Ardışık logit modelde **seqlogit** komutu kullanırken **tree** (ağaç) yapılandırması önemlidir. Ağacın yapılandırmasında verinin dal yapısına bir başka ifade ile karar sürecine dikkat edilmelidir. 4 kategorili bir bağımlı değişkende ağaç yapısı “**tree** (1:2 3 4, 2: 3 4, 3:4)” şeklindedir. Böylece 3 model tahmini gerçekleştirilmektedir. İlk modelde 1. kategoriye karşı 2, 3 ve 4. kategorileri; ikinci model 2. kategoriye karşı 3 ve 4. kategorileri ve sonuncu modelde ise 3. kategoriye karşı 4. kategori modellenmektedir.

Bağımlı değişkenin yapısı ve karar ağacı dikkate alındığında her 3 modelde de iki durumlu logit model hesaplanmaktadır. Burada bahsedilen örnekte karar ağacı örneğinin yapısı gereği ilk model olan “sadece okuyan çocuklarda” sadece okuyan çocuklara 1, diğer durumlara 0 değeri alan bağımlı değişkenle bir logit model tahminlenir.

Ardışık analiz sonucunda elde edilen model sonuçları bağımlı değişken “degree” ve bağımsız değişkenler “p_harc_medyn erkek_coc anne_egts baba_egts kent coc_sayisi” olmak üzere her bir kategori için sırasıyla Tablo 6.5’de gösterilmiştir.

Tablo 6.5. Ardışık Model Tahmini

```
seqlogit degree p_harc_medyn erkek_coc anne_egts baba_egts kent coc_sayisi,
tree(2 3 4:1, 3 4: 2, 4:3)
```

Transition tree:

Transition 1: 2 3 4 : 1

Transition 2: 3 4 : 2

Transition 3: 4 : 3

Computing starting values for:

Transition 1

Transition 2

Transition 3

Iteration 0: log likelihood = -26727.259

Iteration 1: log likelihood = -26727.259

Number of obs = 25957
 LR chi2(18) = 5374.90
 Log likelihood = -26727.259
 Prob > chi2 = 0.0000

degree	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
-----+-----						
_1v2_3_4	Sadece Eğitime Devam Eden Çocuklar					
p_harc_medynd	-.3262879	.0379797	-8.59	0.000	-.4007267	-.251849
erkek_coc	.328124	.0263734	12.44	0.000	.276433	.379815
anne_egts	.0824324	.0052605	15.67	0.000	.0721219	.0927429
baba_egts	.087321	.0047174	18.51	0.000	.0780751	.0965669
kent	.5408638	.0296228	18.26	0.000	.4828042	.5989235
coc_sayisi	-.0310101	.0081427	-3.81	0.000	-.0469695	-.0150508
_cons	-1.325706	.0515277	-25.73	0.000	-1.426699	-1.224714
-----+-----						
_2v3_4	Okula Devam Ederken Çalışan Çocuklar					
p_harc_medynd	-.1150688	.1017975	-1.13	0.258	-.3145882	.0844507
erkek_coc	1.0382	.0803792	12.92	0.000	.8806595	1.19574
anne_egts	.0750047	.0154236	4.86	0.000	.044775	.1052344
baba_egts	-.0399283	.0146011	-2.73	0.006	-.068546	-.0113105
kent	-.0284649	.0791506	-0.36	0.719	-.1835973	.1266675
coc_sayisi	-.1431242	.0245248	-5.84	0.000	-.191192	-.0950564
_cons	-2.885151	.1427915	-20.21	0.000	-3.165017	-2.605285
-----+-----						
_3v4	Sadece Çalışan Çocuklar					
p_harc_medynd	.1328967	.0491759	2.70	0.007	.0365137	.2292796
erkek_coc	1.20439	.0404706	29.76	0.000	1.125069	1.283711
anne_egts	-.0058777	.0086101	-0.68	0.495	-.0227532	.0109979
baba_egts	-.1231141	.0081224	-15.16	0.000	-.1390337	-.1071945
kent	-1.062458	.041473	-25.62	0.000	-1.143744	-.9811726
coc_sayisi	-.0801226	.0108252	-7.40	0.000	-.1013396	-.0589056
_cons	.6027412	.0722674	8.34	0.000	.4610997	.7443827
-----+-----						

Tablo 6.5'in yorumunu kolaylaştırabilmek ve daha anlaşılır hâle getirmek için aşağıdaki Tablo 6.5 tekrar düzenlenmiştir.

Tablo 6.5. Ardışık Logit Model Karar Aşamaları Sonuçları

	Model		
	Birinci Aşama: P₁ Sadece Eğitime Devam Etme Ola- sılığı	İkinci Aşama: P₂ Okula Devam Ederken Çalışma Olasılığı	Üçüncü Aşama: P₃ Sadece Çalışma Olasılığı
Değişkenler	Katsayı (Std Hata)	Katsayı (Std Hata)	Katsayı (Std Hata)
Sabit	-1.325 (0.051)*	-2.885 (0.142)*	0.602 (0.072)*
Yoksulluk Çizgisi (p_harc_medyn)	-0.326 (0.037)*	-0.115 (0.101)	0.132 (0.049)*
Erkek Çocuk (erkek_coc)	0.328 (0.026)*	1.038 (0.080)*	1.204 (0.040)*
Annenin Eğitim Durumu (anne_egts)	0.082 (0.005)*	0.075 (0.015)*	-0.005 (0.008)
Babanın Eğitim Durumu (baba_egts)	0.087 (0.004)*	-0.039 (0.014)*	-0.123 (0.008)*
Kent	0.540 (0.029)*	-0.028 (0.079)	-1.062 (0.041)*
Çocuk Sayısı (coc_sayisi)	-0.031 (0.008)*	-0.143 (0.024)*	-0.080 (0.010)*
N	25957		

Referans Sınıflar: Her üç aşamada da temel sınıflar kırdıya yaşayan, kız çocukları ve çocuk geliri olmadan eşdeğer fert başına hane harcaması medyan harcamanın üstünde yer alan hanelerdir. * %1 önem seviyesinde anlamlıdır. **%5 önem seviyesinde anlamlıdır.

Birinci Karar Aşaması: Sadece Eğitime Devam Eden Çocukların Ardışık Logit Model Sonuçları

Tablo 6.5’de ikinci sütunda sadece eğitime devam eden çocukların 1 ve diğer durumların 0 olduğu bağımlı değişkenin ilk kategorisi için elde edilen model sonuçları verilmiştir. Bütün katsayılar istatistiki olarak anlamlı bulunmuştur ($p < 0.01$). Ardışık modeller iki durumlu logit modellerden oluştuğu için yorumlar iki durumlu logit modellerle aynıdır. Tablo 6.5’de verilen model sonuçlarında katsayı büyüklükleri yerine katsayı işaretleri incelenerek yorum yapılabilir. Elde edilen sonuçlar incelendiğinde yoksul hanelerin yoksul olmayan hanelere göre çocuklarını okula gönderme olasılıklarının daha az olduğu görülmektedir. Eğitimden uzaklaşan çocuk, hane gelirine katkıda bulunmak için geleceğindeki refahından vazgeçmektedir. Yapılan analiz sonucunda erkek çocuklarının kız çocuklarına göre eğitimlerine devam etme olasılığının daha fazla olduğu görülmüştür. Bu durum zorunlu eğitimi tamamladıktan sonra kız çocuklarının ya eğitimlerine ara verdiği ya da tamamen eğitimden uzaklaştırılması ile açıklanabilmektedir. Ebeveynlerin eğitim düzeyleri incelendiğinde annenin ve babanın eğitim seviyesi arttıkça çocukların eği-

timlerine devam etme olasılıklarının arttığı görülmektedir. Kentte yaşayan çocukların kırdaki yaşayan çocuklara göre eğitimlerine devam etme olasılığı daha fazladır. Hanedeki çocuk sayısı arttıkça çocukların eğitime devam olasılıkları azalmaktadır.

İkinci Karar Aşaması: Okula Devam Ederken Çalışan Çocukların Ardışık Logit Model Sonuçları

İkinci karar aşama tahminleri “sadece eğitime devam eden çocuklara” ait gözlemler çıkarıldıktan sonra tahminlenmiştir. Hem çalışan hem de okuyan çocuk sayısı birleştirilmiş veri seti içerisinde sadece %3’dür (Tablo 6.3). Bu karar aşamasında, ilk karar aşamasından farklı olarak, çocuk geliri olmadan eşdeğer fert başına hane harcamasının medyan değerine göre elde edilen yoksulluk çizgisi değişkeni ve kent değişkeni katsayıları haricinde tüm katsayılar istatistiki olarak anlamlı bulunmuştur ($p < 0.01$). Tablo 6.5 incelendiğinde erkek çocuklarının kız çocuklarına göre hem okuyup hem çalışma olasılığının daha fazla olduğu ifade edilmektedir. Babanın eğitimindeki her bir yıllık artış çocukların hem okuyup hem de çalışma olasılığını azaltırken, annenin eğitimindeki artış bu olasılığı arttırmaktadır. Çocuk sayısı arttıkça çocukların hem çalışıp ve hem de okuma olasılıkları azalmaktadır.

Üçüncü Karar Aşaması: Sadece Çalışan Çocukların Ardışık Logit Model Sonuçları

Üçüncü aşama “eğitimine devam etmeyen, sadece iş gücü piyasasında yer alan çocuklardan” oluşmaktadır. Bu aşamada “sadece eğitime devam eden çocuklarla” birlikte “hem okuyan hem de çalışan çocuklara” ait gözlemler veri setinden çıkarılmıştır. Anne eğitimi dışındaki katsayılar istatistiki olarak anlamlı bulunmuştur ($p < 0.01$). Harcama esaslı hanenin yoksulluk durumunu gösteren yoksulluk çizgisi değişkeni incelendiğinde, yoksul hanelerde yaşayan çocukların iş gücünde yer alma olasılığının daha fazla olduğu görülmektedir (%13). Bu durum yoksul hanelerde çocukların sadece eğitimleri için değil, yiyecek, ısınma gibi gereksinimler için de çocuğun geliri ve iş gücüne ihtiyaç duyulduğunu göstermektedir. Tablo 6.5’de çocuk cinsiyeti incelendiğinde erkek çocuklarının kız çocuklarına göre iş gücünde yer alma olasılığının daha fazla olduğu görülmektedir. Erkek çocukları kız çocuklarına göre daha fazla iş gücü piyasasında yer alırken eğitimlerinden de uzaklaşmaktadır. Bununla birlikte anne ve babanın eğitim yıllarındaki her bir yıllık artış çocuğun iş gücünde yer alma olasılığını azaltmaktadır. Kentte yaşayan hanelerin çocukları kırdaki yaşayan hanelerin çocuklarına göre daha az iş gücüne katılmaktadır.

Fark (Odd) Oranlarının ve Marjinal Etkilerin Hesaplanıp Yorumlanması

Ardışık logit model tahminlerinde ayrıca fark oranları da tahminlenip yorumlanabilir. Fark oranı $\exp(\beta)$ olarak hesaplanmaktadır ve yorumları iki durumlu logit modellerle aynı şekilde yapılmaktadır. Fark oranının 1’den büyük olması bağımlı değişkenin daha üst bir düzeyinde olma şansının yüksek, 1’den küçük olması ise, bağımlı değişkenin daha üst bir düzeyinde olma şansının düşük olduğu şeklinde yorumlanmaktadır. Ayrıca "pozitif" etkilerin 1’den büyük olduğu ve negatif etkilerin 0 ile 1 arasında olduğu anlamına gelmektedir. Bu nedenle pozitif ve negatif etkinin büyüklükleri karşılaştırılırken, negatif etkinin tersi alınarak karşılaştırılması yapılmalıdır (Long, 1997: 80-82).

Ardışık logit modelde fark oranlarının hesaplanması için Stata programında “seqlogit” komutunun sonunda “*or*” komutu kullanılmaktadır. Her üç aşama için elde edilen fark oranları Tablo 6.6’de verilmiştir.

Tablo 6.6. Fark Oranı Sonuçları

```
seqlogit degree p_harc_medynd erkek_coc anne_egts baba_egts kent coc_sayisi
, tree(2 3 4:1, 3 4: 2, 4:3) or
```

Transition tree:

Transition 1: 2 3 4 : 1

Transition 2: 3 4 : 2

Transition 3: 4 : 3

Computing starting values for:

Transition 1

Transition 2

Transition 3

Iteration 0: log likelihood = -26727.259

Iteration 1: log likelihood = -26727.259

Number of obs = 25957

LR chi2(18) = 5374.90

Log likelihood = -26727.259

Prob > chi2 = 0.0000

degree	Odds Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
-----+-----						
_1v2_3_4	Sadece Eğitimine Devam Eden Çocuklar					
p_harc_medynd	.7215974	.0274061	-8.59	0.000	.6698331	.7773621
erkek_coc	1.388361	.0366159	12.44	0.000	1.318419	1.462014
anne_egts	1.085925	.0057126	15.67	0.000	1.074786	1.09718
baba_egts	1.091247	.0051478	18.51	0.000	1.081204	1.101383
kent	1.71749	.0508769	18.26	0.000	1.620613	1.820158
coc_sayisi	.9694658	.007894	-3.81	0.000	.9541165	.9850619

_cons		.2656153	.0136865	-25.73	0.000	.2401003	.2938418
-----+-----							
_2v3_4		Okula Devam Ederken Çalışan Çocuklar					
p_harc_medyn		.8913048	.0907326	-1.13	0.258	.7300895	1.088119
erkek_coc		2.824128	.2270011	12.92	0.000	2.41249	3.306004
anne_egts		1.077889	.0166249	4.86	0.000	1.045793	1.110971
baba_egts		.9608584	.0140296	-2.73	0.006	.9337505	.9887532
kent		.9719364	.0769294	-0.36	0.719	.8322709	1.13504
coc_sayisi		.8666464	.0212544	-5.84	0.000	.825974	.9093217
_cons		.0558464	.0079744	-20.21	0.000	.0422134	.0738821
-----+-----							
_3v4		Sadece Çalışan Çocuklar					
p_harc_medyn		1.142132	.0561654	2.70	0.007	1.037189	1.257694
erkek_coc		3.334726	.1349585	29.76	0.000	3.080431	3.610013
anne_egts		.9941396	.0085597	-0.68	0.495	.9775037	1.011059
baba_egts		.8841627	.0071815	-15.16	0.000	.8701987	.8983509
kent		.3456052	.0143333	-25.62	0.000	.318624	.3748713
coc_sayisi		.9230032	.0099917	-7.40	0.000	.9036261	.9427958
_cons		1.82712	.1320412	8.34	0.000	1.585817	2.105141
-----+-----							

Birinci karar aşamasında (sadece okuyan çocuklar) yoksulluk çizgisi için (p_harc_medyn)fark oranı 1'den küçük olduğu için tersi alınarak elde edilen değer $1/0.721 = 1.386$ 'dır. Buna göre yoksul olmayan hanelerin yoksul olan hanelere göre çocuklarını okula gönderme olasılığının 1.38 kat daha fazla olduğunu göstermektedir.

Erkek çocukların kız çocuklarına göre okula devam etme olasılığı 1.38 kat daha fazladır. Anne eğitim seviyesi ve baba eğitim seviyeleri arttıkça çocukların okula devam etme olasılıkları sırasıyla 1.08 ve 1.09 kat daha artmaktadır. Kentte yaşayan çocuklar kırdaki yaşayan çocuklara göre 1.71 kat daha fazla eğitimlerine devam edebilmektedir. Hanelerde yaşayan çocuk sayısı arttıkça çocukların eğitimlerine devam etme olasılıkları yüzde 3 ($0.969-1*100$) azalmaktadır.

İkinci karar aşamasına (hem okuyan hem çalışanlar çocuklar) ait fark oranları incelendiğinde yoksul olmayan hanelerin yoksul hanelere göre çocuklarını hem okutup hem de çalıştırma olasılığı $1/0.891 = 1.122$ kat daha fazladır. Erkek çocuklarının kız çocuklarına göre hem bir işte çalışıp hem de okumaya devam etme olasılığı ise 2.82 kat daha fazladır. Annenin eğitim düzeyi arttıkça çocukların bir işte çalışıp hem de okuma olasılığı 1.07 kat atarken, babanın eğitim düzeyindeki artış karşısında çocukların bu olasılığı 0.960 kat arttırmaktadır. Diğer bir deyişle yaklaşık yüzde 3 azalmaktadır ($0.964-1*100$). Kente yaşayan çocuklar kırdaki yaşayan çocuklara göre $1/0.971 = 1.029$ kat daha az hem çalışıp hem de okumaktadır. Bir başka ifadeyle yüzde 2 daha

Komutlar

```
describe
```

```
tab degree
```

```
sum p_harc_medyn erkek_coc anne_egts baba_egts kent coc_sayisi
```

```
seqlogit degree p_harc_medyn erkek_coc anne_egts baba_egts kent coc_sayisi,  
tree(2 3 4:1, 3 4: 2, 4:3)
```

```
seqlogit degree p_harc_medyn erkek_coc anne_egts baba_egts kent coc_sayisi ,  
tree(2 3 4:1, 3 4: 2, 4:3) or
```

```
margins, dydx(*) atmeans predict (trpr transition(1) choice(1))
```

```
margins, dydx(*) atmeans predict (trpr transition(2) choice(1))
```

```
margins, dydx(*) atmeans predict (trpr transition(3) choice(1))
```

6.5. KAYNAKLAR

- Amemiya, T. (1981). Qualitative Response Models: A Survey. *Journal of Economic Literature*, 19 (4), 1483-1536.
- Cartwright, K., ve Patrinos, H. A. (1997). Children's work activities in urban Bolivia. *World Bank Working Paper*, 1-30.
- Grootaert, C. (1998). Child labor in Cote d'Ivoire: Incidence and Determinants. *The World Bank Policy Research Working Paper* (1905).
- Grootaert, C. (1997). The Determinants of Poverty in Cote d'Ivoire in the 1980s. *Journal of African Economies*, 6(2), 169-196.
- Liao, T. F. (1994). *Interpreting Probability Models. Logit, Probit and Other Generalized Linear Models*. Sage Publications
- Maddala, G. S. (1983). *Limited dependent and qualitative variables in econometrics* (No. 3). Cambridge university press.
- Sakellariou, C. ve Lall A.(2000). Child Labour in the Philipiness: Determinants and Effects, *Asian Economic Journal*, 14(3), s.233-253
- Tutz, G. (1997). Sequential Models for Ordered Responses. W. v. Linden içinde, *Hanbook of Modern Item Response Theory* (s. 139-152). Springer.
- Wahba, J. (2008). Child Labor and Poverty Transmission: No Room for Dreams . *ERF Working Paper 0108* .
- Winkelmann, R. ve Boes, S. (2006). *Analysis of Microdata*. Newyork: Springer.

BÖLÜM 7

YUVALANMIŞ (NESTED) LOGİT MODELİ VE STATA UYGULAMALARI

Öznur Özdamar Giovanis¹

Eleftherios Giovanis²

7.1. YUVALANMIŞ LOGİT MODELİNİN TEORİK ÇERÇEVESİ

Enflasyon oranı, tüketim miktarı, işsizlik oranı gibi değişkenler nicel değişkenler olup, bir durumun sürekli değerler alan göstergeleridir. Oysa ekonometride modellenmek istenen her ekonomik değişken nicel özellik göstermeyebilmektedir. Bu tür değişkenler nitel değişkenler olarak adlandırılmaktadır. Bir ankette “*belediye hizmetlerinden memnun musunuz?*” sorusuna verilen “evet” ve “hayır” cevapları iki seçenekli olduğu için iki durumlu nitel değişken olarak isimlendirilmektedir. Benzer şekilde “*hâlihazırda bir işte çalışıyor musunuz?*” sorusu bireylerin istihdam statülerini “evet” ve “hayır” cevapları ile görmemizi sağlayan iki durumlu nitel bir değişkendir. Nitel değişkenler ikiden fazla değer de alabilir. Cevapları “ücretli”, “kendi hesabına”, “işveren”, “işsiz” gibi ikiden fazla değer alan çalışma durumu değişkeni çok durumlu nitel değişkene örnek olarak gösterilebilir. Bağımlı değişkeni bunlar gibi iki veya çok durumlu nitel değişkenlerden oluşan modeller ise *Nitel Tercih Modelleri* (Qualitative Choice Models) olarak adlandırılır. Bu modellerin bir diğer adı *Kesikli Tercih Modelleridir* (Discrete Choice Models). Nitel Tercih Modelleri literatürde pek çok konuda kendine uygulama alanı bulmuştur. Otobüs, otomobil, uçak gibi ulaşım aracı talebi ve seyahat edilecek yer seçimlerinde (Bhat vd., 1998; Train, 1998), otomobil markası ve model seçimlerinde (Bhat ve Pulugurta, 1998), konut tercihi ve yaşanılacak yer tercihlerinde (Waddell, 1993; Sermons ve Koppelman, 1998), finansal şirketlerin yatırım kararlarında (Corres vd., 1993) ilgili bireylerin veya karar vericilerin davranışlarını tahmin etmek ve birden fazla alternatif arasında verilecek kararı etkileyen faktörleri analiz etmek için *nitel tercih modelleri* kullanılmıştır.

Kitabın daha önceki bölümlerinde ayrıntılarıyla anlatıldığı gibi *nitel tercih modelleri* bağımlı değişkenin iki veya ikiden fazla değer almasına göre *ikili veya çoklu tercih modelleri* olarak adlandırılmakta olup, *çoklu tercih modelleri* bağımlı değişkenin sıralı değerler alıp almamasına göre farklılaşmaktadır. *Çok durumlu (multinomial) modeller*, bağımlı değişkeni sıralı olmayan tercihlere sahip olan modeller arasında en çok kullanılan modellerdendir.

¹ Doç. Dr., İzmir Bakırçay Üniversitesi, İktisadi ve İdari Bilimler Fakültesi, İktisat Bölümü.
E-mail: oznurozdamar@gmail.com

² Dr., Manchester Metropolitan University, Department of Economics, Policy and International Business.
Adnan Menderes Üniversitesi, Nazilli İİBF, Maliye Bölümü. E-mail: giovanis95@gmail.com

Çok durumlu modellerin en önemli kısıtı, **İlişkisiz Alternatiflerin Bağımsızlığı**³ (Independence of Irrelevant Alternatives-IIA) varsayımı altında çalışmasıdır. Varsayım, modelden elde edilecek parametre tahminlerinin değişkendeki tercihlerin (durumların) sayısından bağımsız olarak aynı kalacağını söylemektedir (Maddala, 1983). Hata terimlerinin bağımsız ve sabit varyanslı olduğu varsayımını takip eden IIA'da, var olan tercihlerden birinin eksilmesi veya yeni bir tercih eklenmesi durumunda diğer tercihlerin göreceli olasılık değerlerinin değişmeyeceği ifade edilmektedir. Bu durum ancak ve ancak tercihlerin birbirinden bağımsız olduğu durumda geçerli olabilmektedir (McFadden, 1974). Tercihlerin birbirinden bağımsız olmadığı durumda IIA varsayımının geçerli olmayacağını McFadden (1974), kırmızı otobüs ve otomobil olmak üzere iki tercihli bir ulaşım seçiminden yola çıkarak açıklamıştır. Kırmızı otobüs ve otomobilin her birinin ulaşımında kullanılma olasılığının $\frac{1}{2}$ olduğu varsayıldığında, kırmızı otobüs yerine otomobil kullanmanın göreceli olasılık değeri $\frac{\frac{1}{2}}{\frac{1}{2}} = 1$ 'e eşit olacaktır. Ulaşım alternatiflerinin mavi otobüs tercihi de ek-

lenerek üçe çıkarıldığı düşünüldüğünde, IIA varsayımının geçerli olabilmesi için kırmızı otobüs yerine otomobil kullanmanın göreceli olasılık değerinin yine 1'e eşit olması yani değişmemesi gerekmektedir. Bunun sağlanması her bir ulaşım tercihinin seçilme olasılıklarının $\frac{1}{3}$ 'e eşit olmasına bağlıdır. Ancak gerçek hayatta, otobüse binmeyi tercih eden bireylerin kırmızı veya mavi otobüse binme konusunda bir fark gözetmeyeceği düşünüldüğünde; otomobil, kırmızı otobüs ve mavi otobüsün seçilme olasılıklarının sırasıyla $\frac{1}{2}$, $\frac{1}{4}$ ve $\frac{1}{4}$ olması beklenir. Daha gerçekçi senaryoda ise IIA sağlanmamakta, kırmızı otobüs yerine otomobil kullanmanın göreceli olasılık değeri 1'e eşit olmamaktadır ($\frac{\frac{1}{2}}{\frac{1}{4}} \neq 1$) Cheng ve Long, 2007). Görüldüğü gibi IIA varsayımının

en büyük kısıtı, her bir tercihe rekabette aynı şansı tanıyor olmasıdır. Gerçek hayatta seçimler arasına yeni tercihlerin dâhil edilmesi ya da bir tercihin eksilmesi özellikle birbirleriyle ilişkili alternatiflerin seçilme olasılıklarını etkileyebildiğinden IIA geçerli olmayabilmektedir. Bu nedenle IIA altında çalışan çok durumlu modellere alternatif yöntemler geliştirilmiş olup bu yöntemlerden biri *Yuvalanmış Logit Modeli*'dir (Williams, 1977; McFadden, 1978; Daly ve Zachary, 1978).

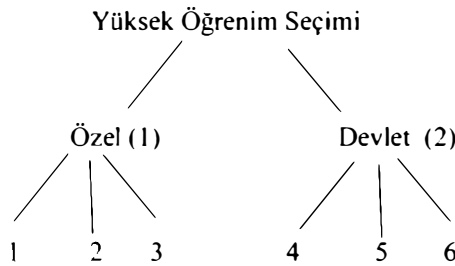
Yuvalanmış logit modelinde birbirine çok benzeyen, biri diğerinin yerine ikame edilebilecek tercihleri gruplama yoluyla analiz gerçekleştirilmektedir. Ayrıca tercihler arasında bir hiyerarşi söz konusu olup bazı ikincil (alt-aşama tercihler), birincil (üst-aşama) tercihlerin içine yuvalanmaktadır. Örnek olarak bireylerin yükseköğretimde okul tercihlerinin *yuvalanmış logit modeli* ile analiz edileceği düşünülsün. Karar alıcı olan öğrencilerin, sırasıyla devlet veya özel bir üniversitede eğitim alma kararları üst-aşama tercihleri olsun. Şekil 7.1'deki hiyerarşik karar alma yapısında görüldüğü gibi, öğrenciler üst aşamada devlete bağlı ya da özel bir kurumda yükseköğrenim almaya karar verilerse alt aşamada hangi özel üniversitede (1,2,3) veya hangi devlet üniversitesinde (4,5,6) eğitim almak istediklerine karar vermek durumundadır. İki aşamalı bu hiye-

³ Bazı kaynaklarda "İlgisiz Alternatiflerin Bağımsızlığı" olarak da adlandırılmaktadır.

rarşik yapıda üst-aşama tercihleri *yuva* olarak adlandırılmakta, alt-aşama tercihleri ise yuvaların içindeki *alternatifler* arasından yapılmaktadır. Bu örnekte basitliği sağlamak amacıyla her bir yuvada üçer alternatifin olduğu varsayılmıştır. Özetle, yükseköğretim tercihlerinde öğrenciler için 3 tane üst-aşama seçim yuvası, 6 tane alt-aşama seçim alternatifi bulunmaktadır.

Bu hiyerarşik yapı bir karar ağacı gibi görülürse yükseköğretim seçimi bu ağacın kökünü (root), özel üniversitede veya devlet üniversitesinde eğitim alma seçimi ana dallarını (limbs), bu ana dallar altında faaliyet gösteren üniversiteler de yan dallarını (twig) oluşturmaktadır.

Şekil 7.1. Yüksek Öğrenim Tercihlerinin Hiyerarşik Yapıda Gösterimi



$B = \{1,2\}$ üst-aşama seçimlerini ($b = 1,2 \in B$), $A_1 = \{1,2,3\}$ ve $A_2 = \{4,5,6\}$ ise üst-aşama seçimlerin içinde yer alan alt-aşama tercihlerini ($j = 1,2,3,4,5,6 \in A_b$) göstermek üzere, iki aşamalı seçimle karşılaşan öğrencinin üst-aşama seçimi $S_1 = b \in B$ ve alt-aşama seçimi $S_2 = j \in A_b$ olarak ifade edilebilir.

Yuvalanmış logit modelinin temelinde *tesadüfi fayda teorisi* (random utility theory) yatmaktadır. Bu teoriye göre bireyler, kendilerine en yüksek faydayı sağlayacak seçimi yapma eğilimindedir. Örneğin 2 ile ifade edilen *devlet* yuvası içerisinde 6 nolu okula gitmeyi tercih eden bir öğrencinin elde edeceği fayda, eşitlik (7.1)'de gösterildiği gibidir.

$$U_{26} = \mathbf{Z}'_{26} \boldsymbol{\beta}_2 + \varepsilon_{26} \quad (7.1)$$

Z_{26} fayda fonksiyonunun deterministik parçasını göstermekte olup, 6 nolu üniversitenin gözlemlenebilir karakteristiklerini temsil eden değişkenleri içeren vektördür (öğrenci sayısı, başarı sırası, sosyal imkânları, yurt kapasitesi gibi). Alt-aşama alternatiflerini etkileyebilecek bu değişkenler *alternatif-spesifik* (alternative specific) değişkenler olarak adlandırılmaktadır. ε_{26} ise fayda fonksiyonunun stokastik parçasını temsil eden özdeş ve bağımsız hata terimidir. *Gumbel'in Uç Değerler* (Gumbel's Extreme Values-GEV) dağılımına sahip olan hata terimlerinin, üst-aşama seçimlerini oluşturan yuvaların içindeki alternatifler arasında korelasyonlu olmalarına izin verilirken, yuvalar arasında korelasyonsuz oldukları varsayılmaktadır. Örneğin *özel* ve *devlet* seçiminden elde edilen faydayı gösterecek fonksiyonlardaki hata terimlerinin (ε_1 ve ε_2) birbirleri ile korelasyonlu olmadığı varsayılırken, *özel* içine yuvalanmış bir üniversite için yazılan fayda fonksiyonundaki hata terimi (ε_{11} , ε_{12} veya ε_{13}) ile *devlet* içine yuvalanmış bir üniversite için yazılan fayda fonksiyonundaki hata teriminin (ε_{24} , ε_{25} veya ε_{26}) ilişkili olabilmesine izin verilmektedir. Ayrıca eğer öğrenci, 6 nolu üniversitede okumayı tercih etmiş ise, bu alternatifi diğer alternatiflere yeğlemiş olduğu varsayılmaktadır. Diğer bir ifade ile 6 nolu alternatiftan elde edeceği

faýdanın, diğerk alternatiflerden elde edeceđi faýdadan yüksek olduđu kabul edilmektedir ($U_{26} > U_{25}$ gibi).

(7.1) nolu eřitlik herhangi bir yuva (b) ierisinden seilebilecek j alternatifi iin genelleřtirilirse, (7.2)'de gsterildiđi řekliyle yazılabilir.

$$U_{bj} = \mathbf{Z}'_{bj} \boldsymbol{\beta}_b + \varepsilon_{bj} \quad (7.2)$$

(7.2) nolu eřitlik ile gsterilen faýda fonksiyonunun hata terimi olan ε_{bj} 'nin GEV kmlatif dađılım fonksiyonu, incelenen rneđe uyarlanarak (7.3) nolu eřitlikteki gibi ifade edilebilir.

$$F(\varepsilon) = \exp[-G(e^{-\varepsilon_{11}}, e^{-\varepsilon_{12}}, e^{-\varepsilon_{13}}, e^{-\varepsilon_{24}}, e^{-\varepsilon_{25}}, e^{-\varepsilon_{26}})] \quad (7.3)$$

$G(\cdot)$ fonksiyonunun zel bir gsterimi, Cameron ve Trivedi (2005:508)'nin yaklařımı takip edilerek eřitlik (7.4)'teki gibi yazılabilir.

$$G(\mathbf{Y}) = G(Y_{11}, Y_{12}, Y_{13}, Y_{24}, Y_{25}, Y_{26}) = \sum_{b=1}^2 (\sum_{j=1}^6 Y_{bj}^{1/\rho_b})^{\rho_b} \quad (7.4)$$

ρ_b aynı yuva iindeki farklı iki alternatife ait hata terimleri (ε_{11} ve ε_{12} veya ε_{bj} ve ε_{bl} gibi) arasındaki korelasyonun bir fonksiyonudur ve $\sqrt{1 - \text{corr}[\varepsilon_{bj}, \varepsilon_{bl}]}$ 'ye eřittir, $\rho_b \in [0,1]$.

$\text{Corr}[\varepsilon_{bj}, \varepsilon_{bl}] = 0$ olduđuunda yani hata terimleri korelasyonsuz olduđuunda ρ_b 1'e eřit olmaktadır. Bu nedenle ρ_b benzersizlik parametresi (dissimilarity parameter) olarak da anılmaktadır. Tm yuvalar iin ayrı ayrı tahmin edilen benzersizlik parametrelerinin istatistiksel olarak anlamlı bir řekilde 1'e eřit olması alternatifler arasında benzerliđin olmadıđını, yani modelin yuvalanmıř logit modeli yerine ok durumlu logit modeli ile analiz edilmesinin yeterli olacađını gstermektedir (Huybers, 2003).

Cameron ve Trivedi (2005: 526-527)'nin yaptđı detaylı ispat, incelenilen rneđe uyarlandđında b yuvası iindeki j niversitesinde eđitim grme olasılıđı (7.5) nolu eřitlikteki gibi gsterilebilir.

$$\text{Prob}(S_2 = j | S_1 = b) = p_{j|b} = \frac{\exp(\mathbf{Z}'_{bj} \boldsymbol{\beta}_b / \rho_b)}{\sum_{l \in A_b} \exp(\mathbf{Z}'_{bl} \boldsymbol{\beta}_b / \rho_b)} \quad (7.5)$$

Ancak yksekđretim seiminin tm tercih ařamaları dikkate alınarak modellenebilmesi iin, ncelikle st-ařama seimlerinin modellenmesi gerekmektedir. Bu bađlamda đrencilerin ncelikle devlet veya zel niversiteye gitmesi durumunda elde edeceđi faýda dzeyi (7.6) nolu eřitlikteki gibi ifade edilebilir.

$$U_b = \mathbf{X}'_b \boldsymbol{\alpha} + \varepsilon_b \quad (7.6)$$

Devlet veya zel niversitede đrenim seimini etkileyecek aıklayıcı deđiřkenler $\mathbf{X}$ vektr ile gsterilmiřtir. Genellikle, bu tercihi yapacak bireylerin kiřisel karakteristikleri modelde aıklayıcı deđiřken olarak yer almaktadır. Diğerk bir ifade ile $\mathbf{X}$ vektr st-ařama seimlerine etki edeceđi dřnlen deđiřkenlerden oluřmaktadır. Bu deđiřkenler *bireye zg* ya da *birey-spesifik* (individual-specific veya case-specific) deđiřkenler olarak adlandırılmaktadır.

Cameron ve Trivedi (2005: 526-527)'nin yaptđı detaylı ispat bu konuya da uyarlandđında đrencilerin devlet niversitesine ya da zel niversiteye gitme olasılıđı eřitlik (7.7)'deki gibi yazılabilir.

$$\text{Prob}(S_1 = b) = p_b = \frac{\exp(\mathbf{X}'_b \boldsymbol{\alpha} / \rho_b l_b)}{\sum_{m \in B} \exp(\mathbf{X}'_m \boldsymbol{\alpha} / \rho_m l_m)} \quad (7.7)$$

Yuvalanmış logit modeli, (7.5) ve (7.7) eşitlikleriyle gösterilen, bireylerin öğrenim görecekları kurum türü seçimi ile gidecekleri üniversite tercihlerini bir arada analiz etme şansı tanımaktadır. Bu bağlamda öğrencinin b yuvası içindeki j üniversitesinde eğitim görmesinin koşullu olasılığı eşitlik (7.8)'de olduğu gibi ifade edilebilir.

$$p_{bj} = \text{Prob}(S_1 = b) * \text{Prob}(S_2 = j | S_1 = b) = p_b * p_{j|b}$$

$$p_{bj} = \left[\frac{\exp(X'_{b\alpha} / \rho_b I_b)}{\sum_{m \in B} \exp(X'_{m\alpha} / \rho_m I_m)} \right] * \left[\frac{\exp(Z'_{bj} \beta_b / \rho_b)}{\sum_{l \in A_b} \exp(Z'_{bl} \beta_b / \rho_b)} \right] \quad (7.8)$$

I_b ve I_m değişkenleri *kapsayıcı değer/fayda* (inclusive value/utility) olarak adlandırılmaktadır. Bunlar, bireylerin b yuvasının ve m ile gösterilen diğer yuvaların içinde yer alan alt-aşama alternatif seçimlerinden elde etmeyi bekledikleri faydayı göstermektedir (Montgomery, 2002). Bu nedenle I_b (7.5) nolu eşitlikteki kesirin bölen kısmının logaritması alınmış hâlden başka bir şey değildir.

$$I_b = \log(\sum_{l \in A_b} \exp(Z'_{bl} \beta_b / \rho_b)) \quad (7.9)$$

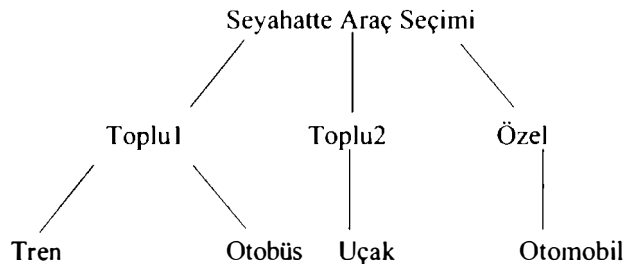
Yuvalanmış logit analizi, *tam bilgiye dayalı en çok olabilirlik tahmincisinin* (full information maximum likelihood estimator) eşitlik (7.10)'daki denklemi α , β_b ve ρ_b 'ye göre maksimize etmesi ile gerçekleştirilir (Cameron ve Trivedi; 2005: 510). y_{ibj} ; i . bireyin ($i=1, \dots, N$) b_j alternatifini seçmesi durumunda 1, aksi hâlde 0 değerini alan *seçim* (choice) değişkenidir.

$$\ln L = \sum_{i=1}^N \sum_{b=1}^B y_{ib} \ln p_{ib} + \sum_i^N \sum_b^B \sum_j^A y_{ibj} \ln p_{ij|b} \quad (7.10)$$

7.2. YUVALANMIŞ LOGİT MODELİ: STATA UYGULAMASI

Stata uygulaması için ele alınan örnek, bireylerin seyahat etmede kullandıkları araçların seçimine ilişkindir. Analizde kullanılacak **seyahat.dta**⁴ veri seti 1986 yılında Avustralya'nın Sidney, Canberra ve Melbourne şehirleri arasında seyahat gerçekleştiren 210 hanehalkını içermektedir (Hensher ve Greene, 2002)⁵. İlgili ankette bireylere uçak, tren, otobüs ve otomobil araçlarından seyahat etmede hangisini tercih ettikleri sorulmuştur. Seçimlerin temsil edildiği iki aşamalı bir hiyerarşik yapı Şekil 7.2'de gösterilmiştir.

Şekil 7.2. Seyahatte Araç Seçiminin Hiyerarşik Yapıda Gösterimi



⁴ seyahat.dta veri setine karekod ile ulaşabilirsiniz.

⁵ Orijinal veri setindeki değişkenlerin isimleri İngilizce olup, anlatımı kolaylaştırmak adına değişken adları Türkçeleştirilmiştir.

Bu yapı karar ağacı olarak düşünülürse, seyahatte araç seçimi karar ağacının köküdür. *Toplu1* ve *Toplu2* iki farklı toplu ulaşım türünü gösteren ana dallardır. *Özel* ise şahsi araçla seyahati temsil eden üçüncü ana dal olup kapsadığı tek alternatif otomobildir. *Toplu1*'e tren ve otobüs gibi iki alternatif kara ulaşım aracı yuvalanmışken, *Toplu2* içerisine tek alternatif olan hava ulaşım araçlarından *uçak* yuvalanmıştır. **Toplu1**, **Toplu2** ve **Özel** olarak adlandırılan bu 3 yuva (ana dal) üst-aşama seçim grubunu oluşturmaktadır. Alt-aşama seçim grubu ise *tren*, *otobüs*, *uçak* ve *otomobil* olmak üzere 4 tanedir. Hiyerarşik yapının seçimi, tamamen araştırmacıya ve analiz edilmek istenen konuya bağlıdır. Eğer ilgili ankette havayolu ulaşımının özel şirketler tarafından sağlandığı belirtilseydi üst-aşama seçim grubu **Kamu** ve **Özel** olarak da ayrılabilirdi ve **Özel** yuvasının içine *otomobil* ve *uçak* dâhil edilebilirdi. Bir diğer hiyerarşik yapının üst-aşama seçim grubu **Hava** ve **Kara** ulaşımı olarak belirlenebilir, **Kara** yuvasının altında *otomobil*, *tren*, *otobüs*, **Hava** yuvasının altında *uçak* yer alabilirdi. Özetle bu hiyerarşik yapının nasıl oluşturulacağı bütünüyle araştırmacının analiz etmek istediği konuya ve veri setinin özelliğine bağlı bir durumdur. Şekil 7.2'de gösterildiği şekliyle bireylerin seyahat seçimlerini modellemek için Stata 15 programında ilk olarak yuvaların (üst-aşama seçimlerin) ve onların içindeki alternatiflerin (alt-aşama seçimlerin) [7.1]⁶ nolu komut ile tanımlanması yapılmalıdır.

nlogitgen yuva = arac(toplu1: otobus | tren, toplu2: ucak özel: otomobil) [7.1]

Bu komut ile orijinal veri setinde daha önce olmayan *yuva* değişkeni yeni bir değişken olarak oluşturulmuştur. Yuva değişkeni *toplu1*, *toplu2* ve *ozel* olmak üzere üç kategoriden oluşan nitel bir değişkendir. *yuva* değişkeninin oluşturulmasında bir diğer nitel değişken olan *arac* değişkeni kullanılmış olup; uçak, tren, otobüs ve otomobil olarak sırasıyla 1, 2, 3 ve 4 değerlerini alan kategorilerden oluşmaktadır. [7.1] nolu komutta *yuva* değişkeninin oluşturulması için eşitliğin sağ tarafına *arac* yazılarak parantez içinde her bir üst-aşama seçimi isimlendirilmiş ve içerdiği alt-aşama alternatifi tanımlanmıştır. Örneğin üst-aşama seçimlerinin ilki *toplu1* olarak isimlendirildikten sonra (:) işareti koyulmuş ve içerdiği alternatifler olan otobüs ve tren tanımlanmıştır. *toplu1*'in içinde otobüs ve tren olmak üzere iki alternatif olduğundan bunlar (|) ayırıcısı ile ayrılmıştır. Ancak *toplu2* veya *ozel* yuvaları içerisinde tek alternatif olduğundan (|) ayırıcısı kullanılmamıştır. Dikkat edilmesi gereken nokta, bir yuvanın alternatifleri ile birlikte tanımlanması yapıldıktan sonra virgül konularak diğer yuvanın ve alternatiflerinin tanımlanmasının yapılmasıdır.

Bu veri setinde olmasa dahi bazı veri setlerinde kategorik değişkenlerin kategorileri, isimleri yerine sayılarla ifade edilebilir. Örneğin *arac* değişkeni içerisinde ulaşım araçları uçak, tren, otobüs veya otomobil olarak değil de sayılarla ifade edilmiş olabilirdi. Öyle bir durumda yuvaların oluşturulması için [7.2] ile gösterilen komut çalıştırılmalıdır.

nlogitgen yuva =arac (toplu1: 3 | 2, toplu2: 1 , özel: 4) [7.2]

Yuvalanmış logit modelinin analizi için yapılması gerekli olmasa da, kullanılan karar ağacının temsili gösterimi [7.3] nolu komut ile elde edilmiş ve Şekil 7.3'te gösterilmiştir.

nlogittree arac yuva [7.3]

⁶ Kitap bölümünde tüm eşitlik/denklem ve kullanılan Stata 15 komutları başta bölüm numarası gelecek şekilde numaralandırılmış olup, karışıklığa neden olmamak için eşitliklerin veya denklemlerin numaraları yuvarlak parantez (), komutların numaraları ise köşeli parantez [] içine alınmıştır.

Şekil 7.3. nlogittree Komutu ile Hiyerarşik Karar Ağacının Elde Edilmesi

tree structure specified for the nested logit model

```

yuva      N      arac      N
-----
toplul 420 --- tren      210
          +- otobus      210
toplul2 210 --- ucak      210
ozel    210 --- otomobil 210
-----
                    total 840

```

N = number of observations at each level

Şekil 7.3'te belirtildiği gibi N gözlem sayısıdır. Toplam gözlem sayısının 840 olmasının nedeni, veri setindeki 210 bireyin her birinin uçak, otobüs, tren ve otomobil olmak üzere 4'er tane seçim alternatifinin olmasıdır. Diğer bir ifade ile *toplul* yuvasında iki, diğer yuvalarda birer alternatif olduğundan 210 kişinin her birinin, alt seçim grubundaki toplam 4 alternatiften seçim yapma hakkı vardır.

Yuvalanmış logit modelinin analizine geçilmeden önce “**describe**” komutu ile elde edilen değişken tanımları Tablo 7.1'de ve “**sum**” komutu ile elde edilen tanımlayıcı istatistikler Tablo 7.2'de gösterilmiştir.

Tablo 7.1. Değişkenlerin Tanımları

describe

Contains data from C:\Users\Asus\Desktop\seyahat.dta

```

obs:      840
vars:      9                      7 Sep 2020 18:15
size:     10,920

```

variable name	storage type	display format	value label	variable label
secim	byte	%8.0g		Seçilen araç türü
termzaman	byte	%8.0g		Evden terminale kadar geçen zaman
seyahatzaman	int	%8.0g		Seyahatte geçen toplam süre
maliyet	int	%8.0g		Seyahatin maliyeti
gelir	byte	%8.0g		Hanehalkı geliri (bin dolar)
kisi	byte	%8.0g		Seyahat eden toplam kişi sayısı
fkimlik	int	%9.0g		Fert kimlik numarası
arac	byte	%8.0g	travel	Seyahat araçları
yuva	int	%8.0g	lb_yuva	

Tablo 7.2. Değişkenlerin Tanımlayıcı İstatistikleri

Variable	Obs	Mean	Std. Dev.	Min	Max
secim	840	.25	.4332707	0	1
termzaman	840	34.58929	24.94861	0	99
seyahatzaman	840	486.1655	301.4391	63	1440
maliyet	840	110.8798	47.97835	30	269
gelir	840	34.54762	19.67604	2	72
kisi	840	1.742857	1.01035	1	6
fkimlik	840	105.5	60.65721	1	210
arac	840	2.5	1.1187	1	4
yuva	840	1.75	.8296502	1	3

Veri setinde *fkimlik* her bir bireye özgü olan fert kimlik numaralarını göstermektedir. Bu değişken veri setindeki birinci kişi için 1, ikinci kişi için 2, üçüncü kişi için 3 ve sonuncu kişi için 210 değerini almaktadır. *gelir* hanehalkı gelirini ifade etmektedir. Bireylerin gelir durumları toplu taşıma mı yoksa özel araçla mı seyahat edileceğini belirleyen önemli bir faktördür. *gelir* değişkeni veri setinde bin dolar cinsinden değer almaktadır. *kisi*, veri setindeki bireyin ailesinden ya da arkadaşlarından kaç kişi ile aynı araçta seyahat edeceğini gösteren değişkendir. Toplu taşıma araçları daha fazla bireyle birlikte seyahat etme imkânı tanıdığından, *kisi* bireylerin toplu taşıma aracıyla mı yoksa özel araçlarıyla mı seyahat edeceklerini belirleyen bir diğer önemli değişkendir. Özetle *kisi* ve *gelir* değişkenleri *bireye özgü* ya da *birey-spesifik* (individual-specific veya case-specific) değişkenler olarak adlandırılmakta olup bireylerin üst-aşama seçimlerini etkilemektedir. *maliyet*, dolar cinsinden seyahat maliyetini gösteren değişkendir. Bu değişken toplu taşıma biletlerini veya özel araçla seyahat söz konusu ise otomobilin yakıt masrafını içermektedir. *termzaman*, evden toplu taşıma aracının kalktığı yere ulaşınca ya kadarki süreyi göstermektedir. Uçak için terminale, otobüs için garaja, tren için tren garına gidinceye kadar geçen zamanı temsil etmektedir. Otomobil ile seyahat edenler için otomobilin zaten evde olduğu varsayıldığından bu değişken sıfır değerini almaktadır. *seyahatzaman*, araçla varış yerine ulaşınca ya kadar geçen süreyi göstermekte olup, bu değişkene *termzaman* değişkeni dâhil değildir. *maliyet*, *termzaman* ve *seyahatzaman* hiyerarşik yapının (karar ağacının) alt-aşama seçimlerini etkileyen *alternatif-spesifik* değişkenlerdir. Yani bu değişkenlerin, bireylerin alt-aşama seçim grubundaki alternatif araçlar arasından hangi aracı seçtiğini gösteren *secim* değişkenini etkilediği varsayılır. *secim*, bireyin seçtiği araç için 1, diğer alternatifler için 0 değerini almaktadır.

Tablo 7.1'de değişkenler hakkında daha ayrıntılı bilgi sahibi olmak için analizde kullanılmış olunan veri setinden ufak bir kesit sunulmuştur. Veri setindeki ilk üç birey seyahat edeceği araç olarak otomobili seçmiş olduğundan *secim* değişkeni otomobil için 1, diğer araçlar için 0 değerini almıştır. Veri setindeki 7. bireyin (*fkimlik*=7) ise uçakla seyahat etmeyi tercih ettiği görülmektedir. Hiyerarşik yapının alt-aşama seçimlerini etkileyen alternatif-spesifik değişkenler (*maliyet*, *termzaman*, *seyahatzaman*) görüldüğü gibi seçilen seyahat aracına göre değişkenlik göstermektedir. Örneğin birinci kişi (*fkimlik*=1) için uçak, tren ve otobüs terminallerine ya da aracın kalkacağı yere evden ulaşım süresi sırasıyla 69, 34, 35 dakika olarak rapor edilmiştir. Daha önce açıklanan nedenlerle otomobil için bu süre 0'dır. *gelir* ve *kisi* değişkenleri seyahat aracına göre değişkenlik göstermemektedir. Diğer bir ifade ile seyahat aracının türüne göre bireylerin gelir du-

rumları ya da birlikte seyahat edecekleri kişi sayısı değişmemektedir. Zaten seyahat öncesi bu değişkenlerin sayısal değerleri bellidir. Örneğin birinci kişinin (*fkimlik*=1) yıllık hanehalkı geliri 35 bin dolar olup, tek başına seyahat edecektir (*kisi* =1). İkinci kişinin ise (*fkimlik*=2) yıllık hanehalkı geliri 30 bin dolar olup seyahat edecek kişi sayısı 2'dir (*kisi*=2) ve seçilen seyahat aracına göre değişkenlik göstermemektedir. Bu nedenle *gelir* ve *kisi*, alt-aşama seçimlerini değil üst-aşama seçimlerini (yuvaları) etkileyen değişkenlerdir.

Tablo 7.3. Temsili Örneklemin Stata Görünümü

secim	termzaman (dk)	seyahatzaman (dk)	maliyet (\$)	gelir (yıllık bin \$)	kisi	fkimlik	arac	yuva
0	69	100	70	35	1	1	uçak	Toplu2
0	34	372	71	35	1	1	tren	Toplu1
0	35	417	70	35	1	1	otobüs	Toplu1
1	0	180	30	35	1	1	otomobil	Özel
0	64	68	68	30	2	2	uçak	Toplu2
0	44	354	84	30	2	2	tren	Toplu1
0	53	399	85	30	2	2	otobüs	Toplu1
1	0	255	50	30	2	2	otomobil	Özel
0	69	125	129	40	1	3	uçak	Toplu2
0	34	892	195	40	1	3	tren	Toplu1
0	35	882	149	40	1	3	otobüs	Toplu1
1	0	720	101	40	1	3	otomobil	Özel
1	45	115	160	45	1	7	uçak	Toplu2
0	34	945	213	45	1	7	tren	Toplu1
0	35	935	167	45	1	7	otobüs	Toplu1
0	0	821	125	45	1	7	otomobil	Özel

[7.4] nolu komut ile yuvalanmış logit modelinin analizi gerçekleştirilir.

**nlogit secim maliyet termzaman seyahatzaman || yuva: gelir kisi , base(toplu1) || arac:,
noconstant case(fkimlik)** [7.4]

Bu komut ile eşanlı olarak birden fazla model analiz edilmektedir. İlk analiz edilen model bağımlı değişkenin "*secim*" olduğu modeldir. Bu ilk modelle bireylerin hiyerarşik yapıdaki alt-aşama seçimlerini belirlemede rol oynayan *maliyet*, *termzaman*, *seyahatzaman* değişkenlerinin etkisi analiz edilmekte ve her biri için ayrı bir parametre tahmin edilmektedir.

(||) ayırıcısı eşanlı olarak analiz edilen modelleri birbirinden ayırmak için kullanılmaktadır. [7.4] nolu komutta **secim maliyet termzaman seyahatzaman** olarak ilk model yazıldıktan sonra (||) ayırıcısı kullanılmış ve bağımlı değişkenin "*yuva*" olduğu ikinci model yazılmıştır. Bu modelde bireylerin üst-aşama seçimlerini belirleyen *gelir* ve *kisi* değişkenlerinin etkisi analiz edilmiştir. [7.4] nolu komutta "**yuva:**" yazıldıktan sonra bu değişkenler komuta eklenmekte ve her biri için bir parametre tahmin edilmektedir. Üst-aşama seçim yuvalarına etki eden faktörlerin analizi gerçekleştirilirken, çok durumlu logit modelinde olduğu gibi kategorilerden biri referans kategori olarak seçilmektedir. [7.4] nolu komutta **base (toplu1)** yazarak Stata 15 programına referans kategorinin *toplul* olduğu belirtilmiş olup, araştırmacılar diğer kategorilerden de herhangi birini referans kategori olarak belirlemede

serbesttir. Spesifik olarak herhangi bir kategorinin referans kategori olduğu belirtilmezse, yani komutta **base()** opsiyonu kullanılmazsa Stata 15 en çok gözleme sahip kategoriye referans kategori olarak kendiliğinden belirlemektedir. Şekil 7.4'te görüldüğü gibi bu kategoriler arasında yine en çok gözleme sahip olan *toplul* olduğundan, **base()** opsiyonu yazılmamış olsaydı da Stata 15'in bu kategoriye referans kategori olarak seçeceği açıktır. [7.4] nolu komutta, **yuva: gelir kisi, base(toplu1)** yazılarak ikinci model analiz edildikten sonra, bağımlı değişkenin *arac* olduğu üçüncü model analiz edilebilir. Komutta görülebileceği gibi **arac:** yazıldıktan sonra her hangi bir bağımsız değişken eklemekten virgül (,) koyularak model sonlandırılmış ve dolayısıyla modelin analiz edilmesi tercih edilmemiştir. Spesifik olarak otomobil, otobüs, uçak ve tren gibi her bir araç türü seçimini etkileyen faktörler ayrı ayrı tahminlenmek istenseydi bağımlı değişkeni *arac* olan model, tahmin edilebilirdi. Bu tamamen araştırmacının hangi etkiyi görmek istediği ile alakalı bir durum olup, Stata 15'te *nlogit* komutu, tüm modellerin analiz edilmemesi esnekliğini tanımaktadır. Dikkat edilmesi gereken nokta, aynı bağımsız değişkenin birden fazla farklı model içinde kullanılamayacağıdır. Örneğin *gelir* değişkeni hem *arac* hem de *secim* değişkeninin bağımlı değişken olduğu modellerde bağımsız değişken olarak yer alamamaktadır. Son olarak [7.4] nolu komutun sonuna *case (fkimlik)*'in yazılması yuvalanmış logit modeli tahminlerini elde edebilmek için önemlidir. Çünkü bu ifade ile seyahat tercihinin karar verenlerin farklı fert kimlik numaralarına (*fkimlik*) sahip olan bireyler olduğu belirtilmektedir. Tablo 7.4'te [7.4] nolu komut ile gerçekleştirilen analizin sonuçları gösterilmektedir.

[7.4] nolu komut çalıştırıldıktan sonra, Stata 15 öncelikle Tablo 7.4'teki karar ağacını göstermekte ve daha sonra katsayı tahminlerini sunmaktadır. **k** ile ifade edilen değişkende görüldüğü gibi 210 kişinin 63 tanesi tren, 30 tanesi otobüsü tercih ederek *toplul* üst-aşama yuvası ile seyahat etmeyi seçmiştir. *toplul* 93, *toplul2* 58, özel araçla seyahat ise 59 kişi tarafından tercih edilmiştir.

secim değişkeninin bağımlı değişken olduğu modelin tahmininde *termzaman* ve *seyahatzaman* değişkenlerine ait katsayıların negatif işaretli ve istatistiksel olarak anlamlı olması, bireylerin aracın kalkacağı terminale gidiş zamanını ve seyahatin toplam süresini artıran araçları daha az olasılıkla tercih ettiklerini göstermektedir. Örneğin otomobil ile seyahat etmek bir terminale gitmeyi gerektirmediğinden evden aracın çıktığı terminale varış süresi arttıkça bireylerin otomobil gibi seyahat araçlarını daha yüksek olasılıkla tercih ettikleri söylenebilir. İşareti negatif olmasına rağmen beklenenin aksine *maliyet* değişkeni istatistiksel olarak anlamsız çıktığından, bu değişkenin etkisine yönelik yorum yapmak yanlış olacaktır.

yuva değişkenine ait kategorilerin bağımlı değişken olduğu ikinci modelde *gelir* ve *kisi* değişkenlerinin etkisi analiz edilmiştir. Bu değişkenlere ait tahminler çok durumlu logit analizinde olduğu gibi temel kategori olan *toplul*'e göre yorumlanmaktadır. Örneğin *toplul2* kategorisinin bağımlı değişken olduğu modelde hanehalkı gelirine ait parametrenin anlamlı ve pozitif çıkmış olması, bireylerin gelirleri arttıkça *toplul*'deki araçlardan ziyade *toplul2*'deki araçlarla seyahati tercih ettiklerini göstermektedir. *toplul2* içine uçak, *toplul* içine ise otobüs ve tren yuvalandığından, gelir durumu iyi olan bireylerin otobüs ve trene kıyasla daha yüksek olasılıkla uçakla seyahat etmeyi seçtikleri sonucuna ulaşılmıştır. Diğer yandan, seyahat eden kişi sayısı arttıkça bireylerin otobüs ve trene kıyasla daha az olasılıkla uçak ile seyahati tercih ettikleri görülmektedir. *ozel* adlı yuvanın bağımlı değişken olduğu modelde hanehalkı geliri negatif katsayılı bulursa da istatistiksel olarak anlamsız çıkmıştır. Birlikte seyahat edecek kişi sayısını temsil eden değişkenin parametresi negatif ve anlamlı bulunduğundan, kişi sayısı arttıkça bireylerin otobüs ve trene kıyasla daha az olasılıkla otomobil ile seyahat ettikleri söylenebilmektedir.

Tablo 7.4. Yuvalanmış Logit Modelinin Tahmini ([7.4] nolu komut ile)

```
nlogit secim maliyet termzaman seyahatzaman || yuva: gelir kisi , base(toplu1)
|| arac:, noconstant case(fkimlik)
```

```
tree structure specified for the nested logit model
```

```
yuva      N      arac      N      k
-----
toplu1 420 --- tren      210  63
      +- otobus      210  30
toplu2 210 --- ucak      210  58
ozel   210 --- otomobil 210  59
-----
total  840 210
```

k = number of times alternative is chosen

N = number of observations at each level

```
RUM-consistent nested logit regression      Number of obs      =      840
Case variable: fkimlik                      Number of cases     =      210
Alternative variable: arac                  Alts per case: min  =       4
                                           avg   =      4.0
                                           max   =       4
                                           Wald chi2(7)       =      54.01
                                           Prob > chi2        =      0.0000
```

```
Log likelihood = -221.22808
```

secim	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
arac					
maliyet	-.0049253	.0044435	-1.11	0.268	-.0136343 .0037837
termzaman	-.0439965	.0089336	-4.92	0.000	-.0615059 -.026487
seyahatzaman	-.0035435	.0007565	-4.68	0.000	-.0050261 -.0020608

```
yuva equations
```

toplu1						
gelir	0	(base)				
kisi	0	(base)				
toplu2						
gelir	.0244671	.0095136	2.57	0.010	.0058208 .0431135	
kisi	-.9553967	.2020117	-4.73	0.000	-1.351332 -.5594611	
ozel						
gelir	-.0089212	.0083229	-1.07	0.284	-.0252338 .0073913	
kisi	-.6257574	.1636603	-3.82	0.000	-.9465257 -.3049891	

```
disimilarity parameters
```

yuva				
/toplu1_tau	.4681289	.1499751	.1741832	.7620746
/toplu2_tau	1	231608.4	-453943.1	453945.1
/ozel_tau	1	211173.1	-413890.7	413892.7

```
LR test for IIA (tau=1): chi2(3) = 4.08
```

```
Prob > chi2 = 0.2525
```

Wald χ^2 (Wald chi2) istatistiği ve istatistiğe ait prob değeri (0.000) kurulan yuvalanmış logit modelinin istatistiksel olarak anlamlı olduğunu göstermektedir.

/toplu1_tau olarak gösterilen benzersizlik parametresi 1'den farklı çıkmış olup benzersizlik parametrelerinin (dissimilarity parameters) hepsinin aynı anda istatistiksel olarak anlamlı bir şekilde 1'e eşit olma durumu sağlanamamıştır. Bu durum çok durumlu logit yerine yuvalanmış logit modeli kullanılmasının doğru bir tercih olduğunu desteklemektedir. Ancak IIA'nın geçerli olup olmadığının testi ise tam tersine varsayımının geçerli olduğunu, analizin çok durumlu logit ile yapılabileceğini göstermektedir. Çünkü IIA'nın ihlal edilmediğini söyleyen sıfır hipotezi, 4.08 olarak hesaplanan LR test istatistiğine ait 0.2525 prob değeri ile reddedilememiştir. Hatırlanacağı üzere ancak ve ancak IIA'nın geçerli olmadığı durumda yani benzer alternatiflerin mevcudiyetinde yuvalanmış logit modeli uygulanmaktadır. Benzersizlik parametreleri ve IIA testi arasındaki bu çelişki analiz edilen modelin de doğruluğu noktasında şüphe doğurmaktadır. Bu durum, bazı yuvalarda sadece tek bir alternatifin olmasından kaynaklanmaktadır. *toplu2* ve *ozel* yuvaları içinde sadece bir alternatif vardır ve bu yuvalar için benzersizlik parametreleri 1 çıkmıştır. Düşünüldüğünde aslında bu durum şaşırtıcı değildir. Aynı yuva içindeki farklı alternatiflerin seçimindeki hata terimlerinin korelasyonsuz olması benzersizlik parametrelerinin 1 olmasına neden olmaktadır. Zaten tek alternatifin olduğu durumda hata terimlerinin korelasyonundan söz etmek mümkün olmadığından, bu yuvalara ait benzersizlik parametrelerinin 1 çıkması beklenen bir durumdur. Bu gibi durumda Stata 15'te constraint opsiyonu kullanılarak, tek alternatifli yuvalarda "benzersizlik parametresinin 1'e eşit olduğu" kısıtı komutta belirtilmelidir (DeMeritt, 2019).

constraint 1 [toplu2_tau]_cons=1

constraint 2 [ozel_tau]_cons=1

İncelediğimiz örnekteki iki yuvada tek alternatif olduğundan bu kısıtlar yukarıdaki iki komut ile belirtilmiş ve nihai analiz [7.5] nolu komutla gerçekleştirilmiştir.

nlogit secim maliyet termzaman seyahatzaman || yuva: gelir kisi , base(toplu1) ||

arac:, noconstant case(fkimlik) constraints(1,2)

[7.5]

Analiz gereğine uygun bir şekilde ilgili kısıtlar belirtilerek gerçekleştirildiğinde, IIA testi için LR χ^2 (chi2) istatistiğinin yine 4.08 olarak bulunduğu, ancak bu kez IIA varsayımının geçerli olduğunu söyleyen sıfır hipotezinin %5 anlamlılık düzeyinde reddedildiği görülmektedir (prob=0.0433). Bu bulgu benzersizlik parametrelerinin gösterdiği sonuçla çelişmemekte, analizde yuvalanmış logit yönteminin uygulanmasının doğru olduğuna işaret etmektedir. Ayrıca açıklayıcı değişkenlere ait parametre tahminlerinin de değişmeden aynı kaldığı görülmektedir.

Tablo 7.5. Yuvalanmış Logit Modelinin Tahmini ([7.5] nolu komut ile)

```
nlogit secim maliyet termzaman seyahatzaman || yuva: gelir kisi , base(toplu1) ||
arac:, noconstant case(fkimlik) constraints(1,2)
```

```
RUM-consistent nested logit regression      Number of obs      =      840
Case variable: fkimlik                       Number of cases     =      210
```

```
Alternative variable: arac                   Alts per case: min =      4
                                              avg =      4.0
                                              max =      4
```

```
Log likelihood = -221.22808                  Wald chi2(7)        =      54.01
                                              Prob > chi2         =      0.0000
```

```
( 1) [toplu2_tau]_cons = 1
```

```
( 2) [ozel_tau]_cons = 1
```

secim	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
arac						
maliyet	-.0049253	.0044435	-1.11	0.268	-.0136343	.0037837
termzaman	-.0439965	.0089336	-4.92	0.000	-.0615059	-.026487
seyahatzaman	-.0035435	.0007565	-4.68	0.000	-.0050261	-.0020608

```
yuva equations
```

toplu1						
gelir		0 (base)				
kisi		0 (base)				
toplu2						
gelir	.0244671	.0095136	2.57	0.010	.0058208	.0431135
kisi	-.9553967	.2020117	-4.73	0.000	-1.351332	-.5594611
ozel						
gelir	-.0089212	.0083229	-1.07	0.284	-.0252338	.0073913
kisi	-.6257574	.1636603	-3.82	0.000	-.9465257	-.3049891

```
disimilarity parameters
```

yuva					
/toplu1_tau	.4681289	.1499751		.1741832	.7620746
/toplu2_tau		1 (constrained)			
/ozel_tau		1 (constrained)			

```
LR test for IIA (tau=1): chi2(1) = 4.08      Prob > chi2 = 0.0433
```

Yuvalanmış logit modeli analiz edildikten sonra bazı tahmin sonrası komutlarla (post-estimation tools) faydalı istatistiklerin elde edilmesi de mümkündür. Örneğin [7.5] nolu komut çalıştırıldık-

tan sonra [7.6] nolu “**estat alternatives**” komutu ile değişkenlere ait önemli özet istatistikler elde edilebilir.

estat alternatives

[7.6]

Tablo 7.6. Hiyerarşik Yapıda Üst ve Alt Aşama Seçimleri İçin Özet İstatistikler

estat alternatives

Alternatives summary for yuva

index	Alternative value	label	Cases present	Frequency selected	Percent selected
1	1	toplu1	420	93	44.29
2	2	toplu2	210	58	27.62
3	3	ozel	210	59	28.10

Alternatives summary for arac

index	Alternative value	label	Cases present	Frequency selected	Percent selected
1	1	ucak	210	58	27.62
2	2	tren	210	63	30.00
3	3	otobus	210	30	14.29
4	4	otomobil	210	59	28.10

Yorumlanması gayet kolay olan Tablo 7.6’da, ilk olarak karar ağacındaki üst-aşama seçimleri gösteren *toplu1*, *toplu2* ve *ozel* yuvaları için gözlem sayısı (cases present), her bir yuvanın kaç birey tarafından seçildiği (seçilme sıklığı) ve seçilme yüzdeleri (percent selected) verilmiştir. Örneğin *toplu1* yuvasının 210 bireyden 93 kişi tarafından seçildiği, *toplu2* ve *ozel* yuvalarını ise sırasıyla 58 ve 59 kişinin tercih ettiği görülmektedir. *toplu1* yuvasını tercih edenlerin yüzdesi %44.29, *toplu2* ve *ozel* yuvalarını tercih edenlerin yüzdesi ise sırasıyla %27.62 ve %28.10’dur⁷. Benzer şekilde alt-aşama seçim grupları için de ilgili istatistikler Tablo 7.6’da paylaşılmıştır. Alt-aşama alternatiflerden en çok tercih edilen aracın tren olduğu görülmektedir. Otobüs ile seyahati seçen 30 birey ile treni seçen 63 kişinin toplamı, bu iki alternatif *toplu1* içinde yuvalandığından, *toplu1*’in gözlem sayısı olan 93’e eşittir. İlave olarak, veri setindeki bireylerin %28.10’unun otomobil, %27.62’sinin ise uçak ile seyahat ettikleri diğer elde edilen istatistikler arasındadır.

Analiz sonrası bazı olasılık değerleri tahmin edilerek de önemli bilgilere ulaşmak mümkündür. Veri setindeki her bir bireyin üst-aşama seçim grubundan bir alternatifini seçme olasılığı (p_1), alt-aşama seçim grubundan bir alternatifini seçme olasılığı (p_2) ve veri bir üst-aşama seçimi altında,

⁷ Stata, Türkçe dilinde çalışan bir program olmadığından ondalık ayracını nokta (.), binlik ayracını ise virgül (,) olarak kullanmaktadır. Metinde Stata sonuçlarından bahsedilirken orijinaline sadık kalınarak sayısal ifadeler türkçeleştirilmemiştir.

alt-aşama seçim grubundan bir alternatifin seçilme olasılığı (**condp**⁸), [7.7] ve [7.8] komutları ile elde edilebilir. Yuvalanmış logit model analizi sonrası tahmin edilebilecek bir diğer değişken kapsayıcı değer/fayda (inclusive value) olarak adlandırılan **iv**'dir. Bu değeri tahmin etmede [7.9] nolu komut kullanılabilir.

predict p* [7.7]

predict condp, condp hlevel(2) [7.8]

predict iv, iv [7.9]

Tablo 7.7. Hiyerarşik Yapıda Üst ve Alt Aşama Seçimlerine ait Olasılıklar ve IV Değerleri

fkimlik	arac	yuva	Secim	p1	p2	condp	iv
1	ucak	toplu2	0	0.085288	0.085288	1	-3.73488
1	tren	toplu1	0	0.211002	0.127542	0.604461	-6.25487
1	otobus	toplu1	0	0.211002	0.083459	0.395539	-6.25487
1	otomobil	ozel	1	0.70371	0.70371	1	-0.78559
2	ucak	toplu2	0	0.093892	0.093892	1	-3.39165
2	tren	toplu1	0	0.278715	0.214051	0.767991	-7.43468
2	otobus	toplu1	0	0.278715	0.064665	0.23201	-7.43468
2	otomobil	ozel	1	0.627393	0.627393	1	-1.14985
6	ucak	toplu2	0	0.075194	0.075194	1	-3.73488
6	tren	toplu1	1	0.260543	0.135844	0.521388	-6.31926
6	otobus	toplu1	0	0.260543	0.124699	0.478612	-6.31926
6	otomobil	ozel	0	0.664263	0.664263	1	-1.21814
7	ucak	toplu2	1	0.764917	0.764917	1	-3.17539
7	tren	toplu1	0	0.068172	0.02629	0.385637	-11.6368
7	otobus	toplu1	0	0.068172	0.041882	0.614363	-11.6368
7	otomobil	ozel	0	0.166911	0.166911	1	-3.52486

İlgili olasılık değerleri ve kapsayıcı değerler, her bir birey için tahmin edildikten sonra bu değişkenlerin aldığı değerleri yorumlamak adına Tablo 7.7'de veri setinden ufak bir kesit sunulmuştur. Birinci birey (*fkimlik*=1) için inceleme yapıldığında, bu bireyin üst-aşama seçim grubundan *toplu2*'yi seçme olasılığının sadece %8.5 olduğu, ancak %21.1 olasılıkla *toplu1* ve %70.4 olasılık ile *ozel* yuvasını seçtiği görülmektedir. Birinci bireyin özel otomobiliyle seyahat etme olasılığının diğer alternatifleri seçme olasılığından yüksek olması, kapsayıcı değerlere bakılarak da anlaşılabilir. Kapsayıcı değer, daha önce ifade edildiği gibi yuvaların içinde yer alan alt-aşama alternatif seçimlerinden elde edilmesi beklenen fayda düzeyidir. 1. birey için *iv* değişkeninin değerlerine bakıldığında, en yüksek değer (-0.78) ile *ozel* yuvasına ait olduğu görülmektedir. *toplu1* ve *toplu2* yuvaları sırasıyla -6.25 ve -3.73 olmak üzere daha düşük *iv* değerlerine sahiptir. Bu durum bireyin özel otomobille seyahat etmekten elde edeceği faydanın diğerlerinden elde edeceği faydadan yüksek olduğunu göstermektedir. 1. birey için *p2* değişkeninin aldığı değerlere bakıldığında ise en yüksek olasılıkla (%70.4) otomobili seyahat etmede tercih ettiği, daha sonra %12.8, %8.5 ve %8.3 olasılık ile sırasıyla tren, uçak ve otobüsü ulaşımda kullandığı görülmek-

⁸ **condp**; conditional probability'nin kısaltması olup, koşullu olasılık anlamına gelmektedir.

tedir. **condp** değişkeni ise *toplu2* yuvası içerisinde sadece uçak, *ozel* yuvası içerisinde sadece otomobil alternatifi olduğundan 1 değerini almıştır. Çünkü birinci birey tarafından *toplu2* yuvasının seçilmesi durumunda uçağın araç olarak tercih edilme olasılığı başka alternatif olmadığından 1'dir. Aynı durum otomobil için de geçerlidir. Ancak *toplu1* yuvasının altında iki alternatif olduğundan, birinci bireyin *toplu1* ulaşım yolunu seçtiği durumda, treni otobüse yeğlediği görülmektedir. Çünkü üst-aşama seçimi *toplu1* olduğunda birey, %60.4 olasılıkla trenle, %39.6 olasılıkla otobüsle seyahat etmektedir. Birinci birey için yapılan bu yorumlar veri setindeki tüm bireylerin ilgili değerleri incelenerek onlar için de yapılabilir.

7.3. YUVALANMIŞ LOGİT MODELİNDE KULLANILAN STATA KOMUTLARI

7. bölümde kullanılan Stata 15 komutları toplu olarak aşağıda verilmiştir.

```
use "DOSYAKONUMU\seyahat.dta" , clear
nlogitgen yuva = arac(toplu1: otobus | tren , toplu2: ucak , ozel: otomobil)
nlogittree arac yuva
describe
sum
nlogit secim maliyet termzaman seyahatzaman || yuva: gelir kisi , base(toplu1)
|| arac:, noconstant case(fkimlik)
constraint 1 [toplu2_tau]_cons=1
constraint 2 [ozel_tau]_cons=1
nlogit secim maliyet termzaman seyahatzaman || yuva: gelir kisi , base(toplu1) ||
arac:, noconstant case(fkimlik) constraints(1,2)
nlogit secim maliyet termzaman seyahatzaman || yuva: gelir kisi , base(toplu1)
|| arac:, noconstant case(fkimlik)
estat alternatives
predict p*
predict condp, condp hlevel(2)
predict iv, iv
```

7.4. KAYNAKLAR

- Bhat, C.R., Govindarajan, A. ve Pulugurta, V. (1998). Disaggregate attraction-end choice modeling: Formulation and empirical analysis. *Transportation Research Record*, 1645, 60-68
- Bhat, C.R. ve Pulugurta, V. (1998). A comparison of two alternative behavioral mechanisms for household auto ownership decisions. *Transportation Research*, 32B, 10-24.
- Cameron, A. C., ve Trivedi, P. K. (2005). *Microeconometrics: Methods and Applications*. Cambridge University Press.
- Cameron, A. C., ve Trivedi, P. K. (2009). *Microeconometrics using Stata*, vol. 5. College Station, TX: Stata Press, USA.
- Cheng, S., ve Long, J. S. (2007). Testing for IIA in the Multinomial Logit Model. *Sociological Methods & Research*, 35(4), 583-600.
- Corres, S., Hajivassiliou, V.A. ve Ioannides, Y.M. (1993). An Empirical Analysis of the Dynamics of Qualitative Decisions of Firms, *Economic Working Papers Series*, Yale University.
- Daly, A. ve Zachary, S. (1978). Improved Multiple Choice Models. Hensher, D.A. ve Dalvi, M.Q. (Ed.). *Determinants of Travel Choice* içinde (335-357). National Research Council (U.S.).

- DeMeritt J.H.R (2009), Discrete Choice Models II, Lecture Notes, Erişim Tarihi: 09.07.2020. https://jdemeritt.weebly.com/uploads/2/2/7/7/22771764/unordered_multiresponse2.pdf
- Hensher, D. A. and Greene, W. H. (2002). Specification and Estimation of the Nested Logit Model: Alternative Normalisations. *Transportation Research Part B: Methodological*, 36(1), 1-17.
- Huybers, T. (2003). Domestic tourism destination choices: A choice modelling analysis. *International Journal of Tourism Research*, 5(6), 445-459.
- Maddala, G.S. (1983). *Limited Dependent and Qualitative Variables in Econometrics*. Cambridge University Press, Cambridge.
- McFadden, D. (1974). 'Conditional Logit Analysis of Qualitative Choice Behavior. P. Zarembka (Ed.), *Frontiers of Econometrics* içinde (105-142). New York: Academic Press.
- McFadden, D. (1978). *Modelling the Choice of Residential Location*. Institute of Transportation Studies, University of California.
- Montgomery, M. (2002). A Nested Logit Model of the Choice of a Graduate Business School. *Economics of Education Review*, 21(5), 471-480.
- Sermons, M.W. ve Koppelman, F.S. (1998) Factor Analytic Approach to Incorporating Systematic Taste Variation into Models of Residential Mode Choice. *Transportation Research Record*. 1617, 194-202.
- Train, K. (1998). Recreation Demand Models with Taste Differences over People. *Land Economics*, 74, 230-240.
- Waddell, P. (1993). Exogenous Workplace Choice in Residential Location Models; Is the Assumption Valid? *Geographical Analysis*, 25, 65-82.
- Williams, H. C. W. L. (1977). On the Formation of Travel Demand Models and Economic Evaluation Measures of User Benefit. *Environment and Planning, Part A*, V.9, 285-344.

BÖLÜM 8

SAYMALI MODELLER VE STATA UYGULAMALARI

Şenay Üçdoğruk Birecikli* Arzu Kökcen Eryavuz[†]
Mehmet Ozan Özdemir[‡]

Sayma değişkenleri negatif olmayan tam sayı değerleri alan değişkenlerdir. Ekonomi, demografi, sağlık, ulaşım, ekoloji gibi alanlarda yapılan çalışmalarda açıklayıcı değişkenler ile analiz etmek istediğimiz bağımlı değişkenin negatif olmayan tam sayı değerleri aldığı durumlarla sıkça karşılaşmaktadır. Günlük içilen sigara sayısı, bir firmanın bir yıl içinde yaptığı patent başvuru sayısı, belirli bir yılda birinin kaç kez tutuklandığı gibi örnekler verilebilir. Bu örneklerin iki ortak özelliği bulunmaktadır. Bunlardan ilki doğal üst sınırlarının olmamasıdır. Diğeri ise cevap değişkeninin anakütlenin bazı üyeleri için sıfır değerini alacak olmasıdır. Ancak bazı sayma değişkenleri için üst sınır bulunmaktadır. Örneğin, bir ailedeki üniversite mezunu çocuk sayısı ise ilgilenildiğinde üst sınır ailedeki çocuk sayısıdır.

Sayma değişkenleri ile ilgili yapılan genel hata sürekli değişkenlermiş gibi ele alınarak doğrusal regresyon modeli uygulanmasıdır. Bu durum sapmalı, tutarsız ve etkin olmayan tahminciler elde edilmesine neden olmaktadır (Long, 1997: 217). EKK yönteminin uygulanması sonucunda sayma değişkeninin doğasına aykırı bir şekilde bağımlı değişkenin tahmin değerleri negatif değerli elde edilebilir (Wooldridge, 2010: 723)

Sayma verilerinin özelliği sayımların sabit bir alanda ya da sabit bir gözlem periyodunda meydana gelmesi ve sayımların nadiren gerçekleşmesidir. Sayma verileri yalnızca negatif olmayan tam sayı değerler alması ve değerlerinin çoğunun sınırlı aralığı nedeniyle potansiyel olarak yüksek derecede çarpık bir dağılım sergilemektedir. Bu nedenle, cevap değişkeninin normal dağılıma sahip olduğunu varsaymak mantıksız olmakla birlikte analiz için doğrusal regresyon modelinin kullanılması bazı problemler ortaya çıkarmaktadır (Zainordin, 2009:2-3). Literatürde sayma verilerinin özelliklerini dikkate alan birçok model bulunmaktadır. Bu modeller içerisinde yaygın

* Prof. Dr., Dokuz Eylül Üniversitesi, İktisadi ve İdari Bilimler Fakültesi, Ekonometri Bölümü.
E-mail: s.ucdogruk@deu.edu.tr

[†] Dr., Araştırmacı. E-mail: arzukocen@gmail.com

[‡] Dr., Dokuz Eylül Üniversitesi, İktisadi ve İdari Bilimler Fakültesi, Ekonometri Bölümü.
E-mail: ozan.ozdemir@deu.edu.tr

olarak kullanılanlar Poisson regresyon modeli (PRM) ve negatif binom regresyon modelidir (NBRM).

Bu bölümde Stata uygulamaları TÜİK tarafından gerçekleştirilen 2012 Sağlık Araştırmaları Anketi verileriyle yapılmıştır. Bağımlı değişken olan bireyin hastanede geçirdiği yataklı tedavi gün sayısı ile bireyin kronik hastalığının olup olmaması, bireyin beden kitle indeksi ve bireyin uzman doktora gitme sayısı (bağımsız değişkenler) arasındaki ilişki için araştırma yapılacaktır. Vücut ağırlığı ve boy uzunluğu kullanılarak hesaplanan beden kitle indeksi (BKİ) zayıf (bki_Z), normal kilolu (bki_N) ve fazla kilolu (bki_F) olmak üzere üç kategorilidir. Öncelikle çalışmada kullanılacak değişkenlere ait bilgiler ve tanımlayıcı istatistikler verilmektedir:

Tablo 8.1. yataklitedavisayisi.dta Veri Seti İçeriği

```
use "DOSYAKONUMU\yataklitedavisayisi.dta"
```

```
describe
```

storage	display	value		
variable	name	type	format	label
				variablelabel
doktor		byte	%8.0g	Uzman Doktora Gitme Sayısı
tedavisayisi		byte	%8.0g	Yataklı Tedavi Gün Sayısı
kronikhastalik		float	%9.0g	Yok ise 0, Var ise 1
bki		float	%9.0g	Beden Kitle İndeksi
bki_Z		byte	%8.0g	Zayıf
bki_N		byte	%8.0g	Normal Kilolu
bki_F		byte	%8.0g	Fazla Kilolu

Tablo 8.2. Değişkenlere Ait Tanımlayıcı İstatistikler

sum	tedavisayisi	doktor	bki_Z	bki_N	bki_F	kronikhastalik	
Variable		Obs	Mean	Std. Dev.	Min	Max	
tedavisayisi		12817	.1880315	1.59389	0	90	
doktor		12817	.3516424	.973115	0	20	
bki_Z		12817	.0358118	.1858279	0	1	
bki_N		12817	.4094562	.4917527	0	1	
bki_F		12817	.554732	.4970148	0	1	
kronikhast~k		12817	.4904424	.4999281	0	1	

Tablo 8.2 incelendiğinde; hastanede geçirilen yataklı tedavi gün sayısının ortalaması 0.19, bireylerin uzman doktora gitme sayısının ortalaması 0.35'tir. Beden kitle endeksi incelendiğinde bireylerin %55'i fazla kilolu iken sadece %4'ü zayıftır. Bireylerin %49'unun ise en az bir kronik hastalığı bulunmaktadır.

Saymalı modellerin kullanılabilmesi için ilk koşul bağımlı değişkenin saymalı yapıda olmasıdır. Çalışmada bağımlı değişken olarak alınan yataklı tedavi gün sayısı negatif olmayan tamsayılar-

dan oluşmaktadır. Bağımlı değişkenin değerlerinin yoğunluk ve yüzdesel dağılımı (Tablo 8.3) ile ya da histogram (Şekil 8.1) ile görülebilir.

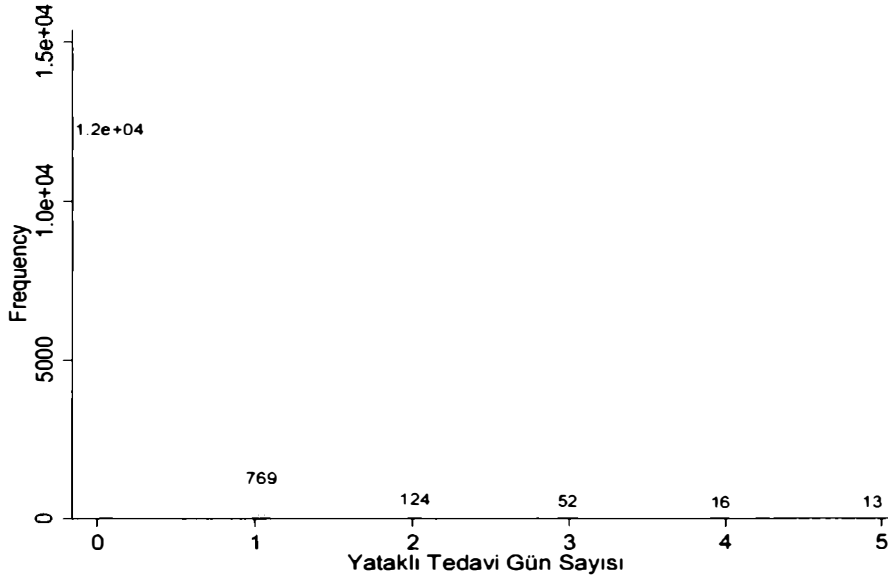
Tablo 8.3. Yataklı Tedavi Gün Sayısı Dağılımı

tab tedavisayisi

Yataklı Tedavi Gün Sayısı	Freq.	Percent	Cum.
0	11,775	91.87	91.87
1	769	6.00	97.87
2	124	0.97	98.84
3	52	0.41	99.24
4	16	0.12	99.37
5	13	0.10	99.47
6	5	0.04	99.51
7	13	0.10	99.61
9	1	0.01	99.62
10	13	0.10	99.72
11	1	0.01	99.73
12	4	0.03	99.76
13	1	0.01	99.77
14	1	0.01	99.77
15	4	0.03	99.80
16	1	0.01	99.81
17	3	0.02	99.84
18	3	0.02	99.86
19	3	0.02	99.88
20	3	0.02	99.91
22	1	0.01	99.91
24	1	0.01	99.92
25	1	0.01	99.93
27	1	0.01	99.94
30	3	0.02	99.96
36	1	0.01	99.97
45	2	0.02	99.98
60	1	0.01	99.99
90	1	0.01	100.00
Total	12,817	100.00	

Şekil 8.1. Yataklı Tedavi Gün Sayısına Ait Histogram

histogram tedavisayisi if tedavisayisi<6, freq addlabel



Tablo 8.3'e bakıldığında bağımlı değişken son 12 ay içinde hastanede geçirilen gün sayısıdır ve minimum 0 iken maksimum 90 gündür. Hastanede geçirilen gün sayısının %91.87'si sıfırlardan %6'sı birlerden oluşmaktadır. 11775 kişi hastaneye yatmadan ayakta tedavi edilmişken, geriye kalan 1052 kişi bir ve daha fazla gün hastanede yatarak tedavi görmüştür. Şekil 8.1'de yer alan histogram incelendiğinde ise gözlemlerin sağa çarpık dağılıma sahip olduğu kolaylıkla görülmektedir.

8.1. POISSON REGRESYON MODELİ

Sayma verilerinin analizi için genellikle yeterli olmasa da en temel ve başlangıç noktası olan model PRM'dir. PRM, verilen $\mathbf{x} = (x_1, x_2, \dots, x_k)$ için her bir y_i 'nin Poisson dağılımına sahip bir anakütleden geldiğini varsaymaktadır. Olasılık kütle fonksiyonu (probability mass function)

$$\Pr(Y = y_i | \mathbf{x}_i) = \frac{\exp(-\mu) \mu^{y_i}}{y_i!}, \quad y_i = 0, 1, 2, \dots \quad (8.1)$$

şeklinde. Olasılık kütle fonksiyonu tamamen koşullu ortalamayla belirlenmektedir. Koşullu ortalama;

$$E(y_i | \mathbf{x}_i) = \mu_i = \exp(\mathbf{x}_i' \boldsymbol{\beta}) \quad (8.2)$$

olarak tanımlanmaktadır. Eşitlik (8.2)'de yer alan eşitlik üstel ortalama fonksiyonu olarak adlandırılmaktadır. Eşitlik (8.1) ve (8.2)'yi içeren model PRM olarak bilinmektedir. Literatürde log-doğrusal model olarak da adlandırılmaktadır. Bunun nedeni koşullu ortalamasının logaritmasının $\ln E(y_i | \mathbf{x}_i) = \mathbf{x}_i' \boldsymbol{\beta}$ şeklinde parametrelere göre doğrusal olmasıdır (Cameron ve Trivedi, 2013: 71).

Poisson dağılımında ilk iki moment birbirine eşittir. Bu durum Poisson dağılımının eş yayılım (equi-dispersion) özelliğini göstermektedir. İlk iki momentin birbirine eşit olmasından dolayı varyans;

$$V(y_i | x_i) = \exp(x_i' \beta) \quad (8.3)$$

şeklinde dir. Eşitlik (8.3) PRM'nin doğası itibarıyla değişen varyanslı olduğunu ifade etmektedir (Cameron ve Trivedi, 2005: 668). Saymalı verilerde varyans ve ortalama arasındaki varsayım yayılım parametresi " ϕ " üzerinden $V(Y) = \phi E(Y)$ şeklinde tanımlanmaktadır. Yayılım parametresinin '1' değerinden küçük çıkması eksik yayılım, '1' değerinden büyük çıkması aşırı yayılım ve '1' olarak elde edilmesi eş yayılım olduğunu ifade etmektedir[§]. Hastanede geçirilen yataklı tedavi gün sayısının varyansı ile ortalama değerlerini karşılaştırabilmek için ilgili değışkene ait detaylı özet istatistikler aşağıdaki gibi elde edilir.

Tablo 8.4. Yataklı Tedavi Gün Sayısı Özet İstatistikleri

sum tedavisayisi, detail

Yataklı Tedavi Gün Sayısı				

	Percentiles	Smallest		
1%	0	0		
5%	0	0		
10%	0	0	Obs	12817
25%	0	0	Sum of Wgt.	12817
50%	0		Mean	.1880315
		Largest	Std. Dev.	1.59389
75%	0	45		
90%	0	45	Variance	2.540484
95%	1	60	Skewness	27.53558
99%	3	90	Kurtosis	1128.657

Bağımlı değışken olan hastanede geçirilen yataklı tedavi gün sayısının ortalaması 0.188 ve varyansı 2.540 olarak elde edilmiştir. ϕ yayılım parametresinin 13.511 ($\approx 2.540/0.188$) olması nedeniyle aşırı yayılım olduğu görölmektedir. Ayrıca varyans değeri nin, ortalama değeri nden büyük olması da veri kümesinde aşırı yayılımın varlığını desteklemektedir.

PRM'de parametrelerin tahminlenmesi için farklı yöntemler bulunmaktadır. Bu tahminleme yöntemlerinden biri En Çok Olabilirlik Yöntemidir. $(y_i | x_i)$ gözlemlerinin birbirinden bağımsız olduğu varsayımı altında log-olabilirlik fonksiyonu;

$$\ln L(\beta) = \sum_{i=1}^n (y_i x_i' \beta - \exp(x_i' \beta) - \ln y_i!) \quad (8.4)$$

şeklinde elde edilir. $\hat{\beta}$ parametrelerinin tahmini;

[§] <http://tarimdergisi.yyu.edu.tr/say16%282%29pdfler/87-92.pdf>

$$\sum_{i=1}^n (y_i - \exp(\mathbf{x}_i' \boldsymbol{\beta})) \mathbf{x}_i = 0 \quad (8.5)$$

ifadesinin çözümünden elde edilmektedir. Eğer $\mathbf{x}_i$ sabit terimi de içeriyorsa eşitlik (8.5)'den artıkların $y_i - \exp(\mathbf{x}_i' \hat{\boldsymbol{\beta}})$ toplamının sıfıra eşit olduğu sonucu elde edilmektedir (Cameron ve Trivedi, 2005: 668). $\hat{\boldsymbol{\beta}}$ 'ların hesaplanması için kullanılan standart yöntem Newton-Raphson iterasyon yöntemidir. Log-olabilirlik fonksiyonu global olarak konkav (içbükey) olduğundan yakınsama kesinlikle sağlanmaktadır. Uygulamada 10 iterasyondan daha azı yeterli olmaktadır. Eğer y_i eşitlik (8.2)'de yer alan ortalama ile Poisson dağılımına sahipse $\hat{\boldsymbol{\beta}}$ için;

$$\hat{\boldsymbol{\beta}} \approx N[\boldsymbol{\beta}, V(\hat{\boldsymbol{\beta}})] \quad (8.6)$$

sonucuna ulaşılmaktadır. Eşitlik (8.6)'da yer alan $V(\hat{\boldsymbol{\beta}})$ ise;

$$V(\hat{\boldsymbol{\beta}}) = \left(\sum_{i=1}^n \mu_i \mathbf{x}_i \mathbf{x}_i' \right)^{-1} \quad (8.7)$$

şeklinde elde edilmektedir (Cameron ve Trivedi, 2013: 71-72).

Stata Tahmini (poisson)

PRM'nin tahminlenmesi için Stata komutu olarak **poisson** kullanılmaktadır. Komutun temel yapısı aşağıda belirtilmektedir.

poisson depvar [indepvars] [if] [in] [weight] [, options]

depvar negatif olmayan sayma değişkenini, indepvars ise modelde kullanılan değişkenleri tanımlamaktadır. Komut yapısı hakkında detaylı bilgi için **help poisson** komutu yardımıyla açılan bilgilendirme ekranından detaylara ulaşabilirsiniz.

Saymalı modellerde regresyon katsayılarının doğru bir şekilde yorumlanması oldukça önemli bir konudur. $E(y_i | \mathbf{x}_i) = \mathbf{x}_i' \boldsymbol{\beta}$ şeklindeki doğrusal regresyon modelleri için $\boldsymbol{\beta}$ katsayıları doğrudan açıklayıcı değişkendeki bir birimlik değişimin koşullu ortalama üzerine etkisi şeklinde yorumlanmaktadır. Ancak doğrusal olmayan modeller için bu durum geçerli değildir ve katsayı yorumlamaları için düzenlemeler yapılmalıdır. Eşitlik (8.2)'de yer alan üstel koşullu ortalama dikkate alınarak marjinal etkiler zincir kuralı yardımıyla aşağıdaki gibi hesaplanabilir.

$$\frac{\partial E(y | \mathbf{x})}{\partial x_j} = \frac{\partial \mu}{\partial x_j} \quad j = 1, 2, \dots, k \quad (8.8)$$

$$\frac{\partial E(y | \mathbf{x})}{\partial x_j} = \frac{\partial \exp(\mathbf{x}' \boldsymbol{\beta})}{\partial x_j} \quad j = 1, 2, \dots, k \quad (8.9)$$

$$\frac{\partial E(y | \mathbf{x})}{\partial x_j} = \frac{\partial \exp(\mathbf{x}' \boldsymbol{\beta})}{\partial \mathbf{x}' \boldsymbol{\beta}} \frac{\partial \mathbf{x}' \boldsymbol{\beta}}{\partial x_j} \quad j = 1, 2, \dots, k \quad (8.10)$$

$$\frac{\partial E(y|\mathbf{x})}{\partial x_j} = e^{x'\beta} \beta_j = \exp(\mathbf{x}'\beta) \beta_j \quad j = 1, 2, \dots, k \quad (8.11)$$

Eşitlik (8.11) j . açıklayıcı değişken x_j 'deki değişimin koşullu ortalama üzerindeki etkisini göstermektedir. Diğer bir deyişle x_j 'deki birim değişimin marjinal etkisini ifade etmektedir (Hilbe, 2011: 126).

Tablo 8.5. Yataklı Tedavi Gün Sayısı Poisson Regresyon Modeli

poisson tedavisayisi doktor kronikhastalik bki_N bki_F, nolog

Poisson regression	Number of obs	=	12817
	LR chi2(4)	=	1529.67
	Prob > chi2	=	0.0000
Log likelihood = -8370.446	Pseudo R2	=	0.0837

tedavisayisi	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
+						
doktor	.2128719	.0067501	31.54	0.000	.1996419	.2261019
kronikhastalik	1.347365	.0519662	25.93	0.000	1.245513	1.449217
bki_N	.5863476	.1635483	3.59	0.000	.2657989	.9068964
bki_F	.4725242	.162674	2.90	0.004	.1536891	.7913592
_cons	-3.197394	.1636832	-19.53	0.000	-3.518207	-2.876581

Elde edilen tahmin sonuçlarına göre katsayılar istatistiki olarak anlamlıdır. Katsayıların işaretleri incelendiğinde, uzman doktora gitme sayısı, kronik hastalığın varlığı, zayıf bireylere göre normal ve fazla kilolu olmanın yataklı tedavi gün sayısı üzerinde pozitif etkisi bulunmaktadır. Marjinal etkiler yorumlanmak istenirse kullanılacak komut **margins, dydx(*)**'tir.

Tablo 8.6. Yataklı Tedavi Gün Sayısı Poisson Regresyon Modeline Ait Marjinal Etkiler

margins, dydx(*)

Average marginal effects	Number of obs	=	12817
Model VCE : OIM			
Expression : Predicted number of events, predict()			
dy/dx w.r.t. : doktor kronikhastalik bki_N bki_F			

	Delta-method				[95% Conf. Interval]	
	dy/dx	Std. Err.	z	P> z		
+						
doktor	.0400266	.0015086	26.53	0.000	.0370699	.0429833
kronikhastalik	.2533471	.0110504	22.93	0.000	.2316887	.2750054
bki_N	.1102518	.0308341	3.58	0.000	.0498181	.1706856
bki_F	.0888494	.0306413	2.90	0.004	.0287935	.1489053

Tablo 8.6'ya bakıldığında açıklayıcı değişkenlere ait marjinal etkiler sırasıyla yaklaşık; 0.040, 0.253, 0.110 ve 0.089 olarak elde edilmiştir. Doktora gitme sayısındaki artış yataklı tedavi gün sayısını 0.040 artırmaktadır. En az bir kronik rahatsızlığı bulunanların kronik hastalığı bulunmayanlara göre 0.253, normal kilolu olanların zayıf kilolu olanlara göre 0.110 ve fazla kiloluların zayıf kilolu olanlara göre 0.089 gün daha fazla hastanede yatarak tedavi edilmektedir.

Saymalı modellerde katsayıların yorumlanması incidence rate ratio (IRR)'ler hesaplanarak da gerçekleştirilebilmektedir. IRR bağımlı kukla değişkenli modellerdeki risk oranı ile aynı şekilde elde edilir. Ancak saymalı modellerde IRR olarak adlandırılmaktadır.

Elde edilen tahminler doğrudan yorumlanamayacağı için Stata programında 2 farklı komut ile incidence-rate ratio (irr) değerleri elde edilebilmektedir. Aşağıda önce PRM ve daha sonra da NBR modelleri uygulanacaktır.

Tablo 8.7. Yataklı Tedavi Gün Sayısı Poisson Regresyon Modeline Ait IRR Değerleri

poisson tedavisayisi doktor kronikhastalik bki_N bki_F , nolog irr

Poisson regression				Number of obs	=	12817
				LR chi2(4)	=	1529.67
				Prob > chi2	=	0.0000
				Pseudo R2	=	0.0837
Log likelihood =	-8370.446					
tedavisayisi	IRR	Std. Err.	z	P> z	[95% Conf. Interval]	
doctor	1.237226	.0083514	31.54	0.000	1.220965	1.253703
kronikhastalik	3.847274	.199928	25.93	0.000	3.474717	4.259776
bki_N	1.797412	.2939636	3.59	0.000	1.304473	2.476624
bki_F	1.604038	.2609352	2.90	0.004	1.166128	2.206393
_cons	.0408686	.0066895	-19.53	0.000	.0296526	.056327

Bununla birlikte, model tahminlendikten sonra **listcoef** komutu çalıştırılarak da IRR değerleri elde edilebilir.

Tablo 8.8. Yataklı Tedavi Gün Sayısı Poisson Regresyon Modeline Ait IRR Değerleri

listcoef

poisson (N=12817): Factor Change in Expected Count
Observed SD: 1.5938896

tedavisayisi	b	z	P> z	e^b	e^bStdX	SDofX
doctor	0.21287	31.536	0.000	1.2372	1.2302	0.9731
kronikhast~k	1.34736	25.928	0.000	3.8473	1.9613	0.4999
bki_N	0.58635	3.585	0.000	1.7974	1.3342	0.4918
bki_F	0.47252	2.905	0.004	1.6040	1.2647	0.4970

Tablo 8.8’de yer alan model sonuçları incelendiğinde kullanılan tüm bağımsız değişkenlerin katsayılarının istatistiki olarak anlamlı olduğu görülmektedir. Diğer değişkenler sabit tutulduğunda uzman doktora gitme sayısındaki artış hastanedeki beklenen tedavi gün sayısını 1.2372 faktör artırdığı görülmektedir. Diğer bir deyişle, $[100*(1.2372-1)]$ değeri hesaplanırsa uzman doktora gitme sayısındaki artışın, yataklı tedavi gün sayısını %23.72 artırdığı söylenir. Araştırmadaki diğer değişkenler sabit tutulduğunda kronik rahatsızlığı olanların yataklı tedavi gün sayısı 3.8473 faktör daha artmaktadır. Yani, kronik rahatsızlığı olanlar, olmayanlardan $[100*(3.8473-1)]$ %285 (2.85 kat) daha fazla hastanede yatmaktadır. Yine, diğer değişkenler sabit tutulduğunda beden kitle indeksi (BKİ)’si normal olanların yataklı tedavi gün sayısı zayıf olanlara göre 1.7974 faktör daha fazladır. BKİ’si normal olanlar, zayıf olanlardan $[100*(1.7974-1)]$ %79.74 daha fazla gün hastanede yatarak tedavi almaktadır. BKİ’si fazla olanların yataklı tedavi gün sayısı zayıf olanlara göre 1.6040 faktör daha fazladır. Diğer bir deyişle, BKİ’si fazla olanlar, zayıf olanlardan $[100*(1.6040-1)]$ %60.4 daha fazla gün hastanede yataklı tedavi görmektedir.

Saymalı modellerde modelin genel olarak performansını değerlendirmek için uyum iyiliği istatistikleri kullanılmaktadır. Genelleştirilmiş doğrusal modeller için ortak uyum iyiliği ölçüleri artıkların ağırlıklandırılmış toplamı olan Pearson ve sapma (deviance) istatistikleridir. Ortalaması μ_i ve varyansı ω_i olan y_i ’nin herhangi bir model tahmini için standart uyum iyiliği ölçüsü olan Pearson istatistiği;

$$P = \sum_{i=1}^n \frac{(y_i - \hat{\mu}_i)^2}{\hat{\omega}_i} \quad (8.12)$$

şeklinde dir. Burada $\hat{\mu}_i$ ve $\hat{\omega}_i$ sırasıyla μ_i ve ω_i ’nin tahmincileridir. Eğer ortalama ve varyans doğru bir şekilde tanımlandıysa

$$E\left(\sum_{i=1}^n \frac{(y_i - \hat{\mu}_i)^2}{\hat{\omega}_i}\right) = n \quad (8.13)$$

sonucu elde edilmektedir. Bunun nedeni;

$$E\left(\frac{(y_i - \hat{\mu}_i)^2}{\hat{\omega}_i}\right) = 1 \quad (8.14)$$

olmasıdır. Pearson istatistiği μ_i ’nin tahmini nedeniyle serbestlik derecesindeki kaybı uyarlamak için uygulamada $(n-k)$ serbestlik derecesi ile karşılaştırılmaktadır. PRM’de ortalama ve varyans birbirine eşit olduğundan Pearson istatistiği PRM için;

$$P_{PRM} = \sum_{i=1}^n \frac{(y_i - \hat{\mu}_i)^2}{\hat{\omega}_i} \quad (8.15)$$

şeklinde yazılabilir (Winkelmann, 2008: 119). Literatürde $P_{PRM} > n - k$ olması gerçek varyansın ortalamayı aştığını ifaden eden aşırı yayılımı kanıt olarak gösterilmektedir. $P_{PRM} > n - k$ sonucu

$$E\left(\frac{(y_i - \hat{\mu}_i)^2}{\hat{\mu}_i}\right) > 1 \quad (8.16)$$

olduğunu belirtmektedir. $P_{PRM} < n - k$ olması ise eksik yayılımı (underdispersion) göstermektedir. Bu yorumlamaların geçerli olabilmesi için μ_i 'nin doğru tanımlanması gerekmektedir. Bununla birlikte, $P_{PRM} \neq n - k$ olması koşullu ortalamanın veya dağılım varsayımının yanlış tanımlandığını ifade etmektedir (Cameron ve Trivedi, 2013: 189).

Diğer bir uyum iyiliği ölçüsü sapma istatistiği;

$$D = \sum_{i=1}^n \left(y_i \ln \left(\frac{y_i}{\hat{\mu}_i} \right) - (y_i - \hat{\mu}_i) \right) \quad (8.17)$$

şeklinde. Sapma istatistiği ulaşılabilir maksimum log-olabilirlik ile modelin tahmininden elde edilen log-olabilirlik arasındaki farkı ölçmektedir (Winkelmann, 2008: 119).

Sabit terim içeren üstel PRM'de eşitliğin sağ tarafındaki ikinci terimin toplamı sıfırdır. Böylece, PRM için sapma istatistiği;

$$D_{PRM} = \sum_{i=1}^n y_i \ln \left(\frac{y_i}{\hat{\mu}_i} \right) \quad (8.18)$$

olarak hesaplanmaktadır. Sapma istatistiği yaklaşık olarak $(n-k)$ serbestlik derecesi ile ki-kare dağılımına sahiptir. NBRM için ise sapma istatistiği;

$$D_{NBRM} = \sum_{i=1}^n \left(y_i \ln \left(\frac{y_i}{\hat{\mu}_i} \right) - (y_i - \alpha^{-1}) \ln \left(\frac{y_i + \alpha^{-1}}{\hat{\mu}_i + \alpha^{-1}} \right) \right) \quad (8.19)$$

şeklinde elde edilmektedir. (Cameron ve Trivedi, 2013: 190-191).

Saymalı modellerde kullanılan uyum iyiliği ölçülerinden bir diğeri Pseudo R^2 'dir. Doğrusal olmayan modellerde R^2 için tek bir tanım olmadığından "sözde" anlamına gelen pseudo ifadesi ile birlikte kullanılmaktadır. Cameron ve Windmeijer saymalı modellerde Pseudo R^2 'nin arzu edilen özelliklerini aşağıdaki gibi sıralamışlardır (Cameron ve Windmeijer, 1996: 209):

1. $0 \leq R^2 \leq 1$
2. Serbestlik derecesi düzeltilmesi olmadan modele ilave bir açıklayıcı değişken eklenmesi R^2 değerini azaltmaz.
3. Artıkların kareler toplamına dayanan R^2 ile açıklanan kareler toplamı ile uyumaktadır.
4. R^2 ile eğim parametrelerinin anlamlılık testleri arasında bir benzerlik vardır.
5. R^2 verinin bilgi içeriği açısından yorumlamaya sahiptir.

Cameron ve Windmeijer (1996) tarafından PRM ve NBRM için sapma artıklarına dayanan R^2 ölçüsü geliştirilmiştir. PRM için sapma R^2 ;

$$R_{D,PRM}^2 = \frac{\sum_{i=1}^n \left(y_i \ln \left(\frac{\hat{\mu}_i}{\bar{y}} \right) - (y_i - \hat{\mu}_i) \right)}{\sum_{i=1}^n y_i \ln \left(\frac{y_i}{\bar{y}} \right)} \quad (8.20)$$

olarak elde edilmektedir (Cameron ve Windmeijer, 1996: 211). NBRM için sapma R^2 ;

$$R_{D,NBRM}^2 = 1 - \frac{\sum_{i=1}^n \left(y_i \ln \left(\frac{y_i}{\hat{\mu}_i} \right) - (y_i + \hat{\alpha}^{-1}) \ln \left(\frac{y_i + \hat{\alpha}^{-1}}{\hat{\mu}_i + \hat{\alpha}^{-1}} \right) \right)}{\sum_{i=1}^n \left(y_i \ln \left(\frac{y_i}{\bar{y}} \right) - (y_i + \hat{\alpha}^{-1}) \ln \left(\frac{y_i + \hat{\alpha}^{-1}}{\bar{y} + \hat{\alpha}^{-1}} \right) \right)} \quad (8.21)$$

formülü ile hesaplanmaktadır (Cameron ve Windmeijer, 1996: 214).

Poisson regresyon modelinin geçerliliği, poisson uyum iyiliğini test eden *estatgof* ya da *poisgof* komutu ile incelenebilir.

Tablo 8.9. Yataklı Tedavi Gün Sayısı Poisson Regresyon Modeline Ait Uyum İyiliği Test Sonuçları

estat gof

```
Deviance goodness-of-fit = 14324.44
Prob > chi2(12812)       = 0.0000

Pearson goodness-of-fit  = 108624.3
Prob > chi2(12812)      = 0.0000
```

Test istatistiği değerleri sırasıyla 14324.44 ve 108624.3'tür ve her iki istatistiğin de anlamlı ($p < 0.05$) oldukları görülmektedir. Sonuç olarak PRM'nin verilere uyum sağlamadığı söylenebilir. Negatif Binom Regresyon Modelini kullanarak analizler bir kez daha aşağıda gerçekleştirilecektir. PRM'de aşırı yayılım varsa tahminlerin etkin olmadığını ve standart hataların aşağıya doğru sapmalı olduğu hatırlanmalıdır.

En çok olabilirlik yöntemi ile tahminlenen modellerin performanslarının değerlendirilmesi ve karşılaştırılmasında log-olabilirlik değeri, Akaike (1973) tarafından ortaya koyulan Akaike Bilgi Kriteri (AIC) ve Schwarz (1978) tarafından önerilen Bayes Bilgi Kriteri (BIC) yaygın olarak kullanılmaktadır. Modele parametre ilave edildikçe log-olabilirlik değeri arttığından AIC ve BIC değerleri parametrelere göre maksimize edilmiş log-olabilirlik değeri ve parametre sayısına bağlı ceza terimi olmak üzere iki bileşenden meydana gelmektedir. Ceza terimi bileşeni, modelde yer alan parametre sayısı ile uyum iyiliği ölçüsü arasındaki dengeyi sağlamaktadır.

$$AIC = -2 \ln L + 2k \quad (8.22)$$

şeklinde elde edilirken BIC değeri ise;

$$BIC = -2 \ln L + k \ln(n) \quad (8.23)$$

olarak hesaplanmaktadır. Burada L olabilirlik değerini, k parametre sayısını ve n gözlem sayısını göstermektedir. Modeller karşılaştırılırken AIC ve BIC değerlerinin en küçük olduğu model tercih edilmektedir (Yip ve Yau, 2005: 158).

STATA'nın **fitstat** komutunu kullanarak Log-olabilirlik, McFadden's R^2 (Pseudo R^2), LR test istatistiği değeri, Akaike Information Criterion (AIC) ve Bayesian Information Criterion (BIC)

gibi çeşitli iyi uyum istatistikleri hesaplanmaktadır. Bu komutu kullanarak elde edilen modele ilişkin uyum ölçüleri aşağıda yer almaktadır. Uyum istatistiklerinden model karşılaştırması yapılırken en uygun modelin seçiminde yararlanılmaktadır. Elde edilen uyum istatistikleri 8.4 no'lu kısımda yer alan Tablo 8.30'dan sonra yorumlanacaktır.

Tablo 8.10. Yataklı Tedavi Gün Sayısı Poisson Regresyon Modeline Ait Uyum İyiliği İstatistikleri

fitstat

Measures of Fit for poisson of tedavisayisi

Log-Lik Intercept Only:	-9135.283	Log-Lik Full Model:	-8370.446
D(12812):	16740.892	LR(4):	1529.675
		Prob > LR:	0.000
McFadden's R2:	0.084	McFadden's Adj R2:	0.083
ML (Cox-Snell) R2:	0.113	Cragg-Uhler(Nagelkerke) R2:	0.148
AIC:	1.307	AIC*n:	16750.892
BIC:	-104441.765	BIC':	-1491.840
BIC used by Stata:	16788.185	AIC used by Stata:	16750.892

Sayma verilerinin analizinde kullanılan PRM oldukça sınırlayıcıdır. PRM'deki temel sorunlardan biri poisson dağılımı nedeniyle y 'nin tüm momentlerinin μ 'nün bir fonksiyonu olmasıdır. Poisson olasılık kütle fonksiyonu sıfır sayma olasılığının örnekleme gerçekte gözlenenden daha az tahminlenmesine yol açmaktadır. Bu durum aşırı sıfır (excess zeros) problemi olarak adlandırılmaktadır.

PRM'deki ikinci ve en belirgin sorun, ortalama ve varyansın eşit olması varsayımının sağlanamamasıdır. Sayma verilerinde genellikle varyans ortalamadan büyük çıkmaktadır. Bu sorun aşırı yayılım (overdispersion) olarak bilinmektedir. Aşırı yayılım, cevaplar arasındaki pozitif korelasyondan ya da cevap olasılıkları ve sayılarındaki aşırı değişkenlikten kaynaklanmaktadır. Aşırı yayılım verinin dağılım varsayımlarının sağlanamadığı durumlarda da gözlemlenmektedir. Aşırı yayılım standart hataların olması gerekenden küçük tahminlenmesine sebep olmaktadır. Böylece, gerçekte istatistiksel olarak anlamlı olmayan bir değişkenin katsayısının istatistiksel olarak anlamlı olduğu sonucuna ulaşılabilir (Hilbe, 2011: 141-142).

PRM'nin tahminlenmesinden sonra aşırı yayılım olup olmadığının test edilmesi gerekmektedir. Aşırı yayımlı saymalı modellerin çoğu;

$$V(y_i | x_i) = \mu_i + \alpha g(\mu_i) \quad (8.24)$$

formunda tanımlanmaktadır. Burada α bilinmeyen bir parametre ve $g(\cdot)$ bilinen bir fonksiyondur. $g(\cdot)$ fonksiyonu genellikle $g(\mu) = \mu^2$ ya da $g(\mu) = \mu$ şeklinde tanımlanmaktadır. Aşırı yayılımın test edilmesinde hipotezler;

$$H_0 : \alpha = 0 \quad (\text{Poisson regresyon modeli})$$

$$H_1 : \alpha \neq 0 \quad \text{veya} \quad H_1 : \alpha > 0 \quad (\text{Negatif binom regresyon modeli})$$

şeklindedir. Sıfır hipotezi aşırı yayılım olmadığını ifade etmektedir. $\alpha = 0$ için $V(y_i | x_i) = \mu_i$ olmaktadır. Test istatistiği PRM'nin tahminlenmesi ve $\hat{\mu}_i$ 'nin elde edilmesinin ardından eşitlik (8.23)'teki sabit terim içermeyen yardımcı regresyonun tahmin edilmesi ile hesaplanmaktadır.

$$\frac{(y_i - \hat{\mu}_i)^2 - y_i}{\hat{\mu}_i} = \alpha \frac{g(\hat{\mu}_i)}{\hat{\mu}_i} + u_i \quad (8.25)$$

Burada u_i hata terimidir. α katsayısının istatistiksel olarak anlamlı olup olmadığının test edilmesi gerekmektedir. Sıfır hipotezi altında α için t istatistiği asimptotik olarak normal dağılmaktadır (Cameron ve Trivedi, 2005: 670-671).

Aşırı yayılımın test edilmesinde Olabilirlik Oranı (LR) ve Wald testleri de kullanılmaktadır. Eşitlik (8.24)'ün geçerli olduğu durumda LR testi için hipotezler;

$$H_0 : \alpha = 0 \quad (\text{Poisson regresyon modeli})$$

$$H_1 : \alpha > 0 \quad (\text{Negatif binom regresyon modeli})$$

şeklindedir. LR test istatistiği;

$$LR = 2(\ln L_1 - \ln L_0) \quad (8.26)$$

olarak hesaplanmaktadır. Burada, L_0 sıfır hipotezinin doğru olduğu varsayımı altında hesaplanan olabilirlik değeri iken L_1 alternatif hipotezin doğru olduğu varsayımı altında hesaplanan olabilirlik değeridir. LR test istatistiği serbestlik derecesi 1 olan ki-kare dağılımına sahiptir (Long ve Freese, 2014: 511). Wald testi ise;

$$W = \frac{\hat{\alpha}^2}{V(\hat{\alpha})} \quad (8.27)$$

şeklinde hesaplanmaktadır. Wald test istatistiği de 1 serbestlik dereceli ki-kare dağılımına sahiptir (İsmail ve Zamani, 2013: 15).

Aşırı yayılımı test etmek için model tahminlendikten sonra aşağıdaki komutlar kullanılabilir:

Tablo 8.11. Aşırı Yayılım Test Sonuçları

predict muhat, n

quietly generate ystar = ((tedavisayisi-muhat)^2 - tedavisayisi)/muhat

regress ystar muhat, noconstant noheader

ystar	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
muhat	23.38864	6.363516	3.68	0.000	10.9152	35.86208

Tablo 8.11’de yer alan muhat eşitlik 8.25’deki $\hat{\mu}$ ’ye eşittir. Elde edilen test istatistiği değeri 3.68’dir ve katsayı istatistiksel olarak anlamlıdır ($p < 0.05$). Bu sonuç aşırı yayılım olduğunu göstermektedir.

Poisson regresyon modelinde parametre tahminleri eş yayılım varsayımı altında tutarlıdır. Yapılan uygulama sonucunda Poisson regresyon modelinde ortalama ve varyansın eşit olması varsayımı sağlanamadığından dirençli standart hatalar hesaplanarak aşırı yayılım nedeniyle ortaya çıkan tutarsızlık ortadan kaldırılabilir. Dirençli PRM için aşağıdaki komut kullanılarak sonuçlar elde edilir:

Tablo 8.12. Yataklı Tedavi Gün Sayısı Dirençli Poisson Regresyon Modeli

poisson tedavisayisi doktor kronikhastalik bki_N bki_F, robust nolog

Poisson regression				Number of obs	=	12817
				Wald chi2(4)	=	314.56
				Prob > chi2	=	0.0000
Log pseudolikelihood = -8370.446				Pseudo R2	=	0.0837
tedavisayisi		Coef.	Robust Std. Err.	z	P> z	[95% Conf. Interval]
+						
doktor		.2128719	.0818381	11.58	0.000	.1768457 .248898
kronikhastalik		1.347365	.1331972	10.12	0.000	1.086303 1.608426
bki_N		.5863476	.2853246	2.06	0.040	.0271216 1.145574
bki_F		.4725242	.2665888	1.77	0.076	-.0499803 .9950286
_cons		-3.197394	.2634162	-12.14	0.000	-3.71368 -2.681108

Tablo 8.12’de yer alan model sonucu incelendiğinde, Wald-ki kare test istatistiğinin (314.56) prob değeri 0.05’ten küçük olduğu için istatistiksel olarak anlamlı bir model elde edildiği söylenir. Katsayılar a ait dirençli standart hataların elde edilmesiyle aşırı yayılım etkisi ortadan kaldırılmıştır. Poisson regresyon modelinde aşırı yayılım olması durumunda alternatif olarak aşırı yayılıma izin veren Negatif Binom regresyon modeli kullanılabilir.

8.2. NEGATİF BİNOM REGRESYON MODELİ

Koşullu ortalama ve koşullu varyansın birbirine eşit olduğu varsayımı PRM’nin en önemli dezavantajı olarak kabul edilmektedir. Uygulamada genellikle koşullu varyans koşullu ortalamadan büyük çıkmaktadır. PRM’ye alternatif olarak eş yayılım sınırlamasını esneten ve koşullu varyansın koşullu ortalamayı aşmasına izin veren Negatif Binom Regresyon Modeli (NBRM) ortaya koyulmuştur. NBRM çeşitli şekillerde elde edilebilmektedir. Bunlar içerisinde yaygın olarak kullanılan yöntem PRM’nin koşullu ortalamasındaki gizli heterojenliğe izin verecek şekilde model yapısının oluşturulmasıdır (Long, 1997: 230). PRM’deki koşullu ortalama gözlenemeyen heterojenliği dikkate alarak;

$$E(y_i | x_i, \varepsilon_i) = \exp(x_i' \beta + \varepsilon_i) = h_i \mu_i \quad (8.28)$$

şeklinde yazılır. Burada $h_i = \exp(\varepsilon_i)$ 'dir. h_i ortalaması $E(h_i) = 1$, varyansı $V(h_i) = v^{-1} = \alpha$ olan gama dağılımına sahiptir ve olasılık yoğunluk fonksiyonu (8.29)'daki gibidir.

$$f(h_i) = \frac{v^v h_i^{v-1} \exp(-vh_i)}{\Gamma(v)}, \quad h_i \geq 0, v > 0 \quad (8.29)$$

Eşitlik (8.28) PRM'de yeniden yazılarak

$$\Pr(Y = y_i | \mathbf{x}_i, h_i) = \frac{\exp(-\mu_i h_i) (\mu_i h_i)^{y_i}}{y_i!}, \quad y_i = 0, 1, 2, \dots \quad (8.30)$$

elde edilmektedir. Poisson ve gama karma modeli olarak da bilinen NBRM poisson ve gama fonksiyonları kullanılarak

$$\begin{aligned} \Pr(Y = y_i | \mu, v) &= \int_0^\infty \Pr(Y = y_i | \mathbf{x}_i, h_i) f(h_i) dh_i \\ \Pr(Y = y_i | \mu, v) &= \int_0^\infty \frac{e^{-\mu_i h_i} (\mu_i h_i)^{y_i}}{y_i!} \frac{v^v h_i^{v-1} e^{-vh_i}}{\Gamma(v)} dh_i \\ \Pr(Y = y_i | \mu, v) &= \int_0^\infty \frac{e^{-h_i(\mu_i + v)} h_i^{y_i + v - 1} \mu_i^{y_i} v^v}{y_i! \Gamma(v)} dh_i \\ \Pr(Y = y_i | \mu, v) &= \frac{\mu_i^{y_i} v^v}{y_i! \Gamma(v)} \int_0^\infty e^{-h_i(\mu_i + v)} h_i^{y_i + v - 1} dh_i \\ \Pr(Y = y_i | \mu, v) &= \frac{\mu_i^{y_i} v^v}{y_i! \Gamma(v)} \int_0^\infty e^{-h_i(\mu_i + v)} h_i^{y_i + v - 1} dh_i \\ \Pr(Y = y_i | \mu, v) &= \frac{\mu_i^{y_i} v^v}{y_i! \Gamma(v)} \frac{\Gamma(y_i + v)}{(\mu_i + v)^{y_i + v}} \\ \Pr(Y = y_i | \mu, v) &= \frac{\Gamma(y_i + v)}{\Gamma(y_i + 1) \Gamma(v)} \left(\frac{\mu_i}{\mu_i + v} \right)^{y_i} \left(\frac{v}{\mu_i + v} \right)^v \end{aligned} \quad (8.31)$$

şeklinde elde edilmektedir (Cameron ve Trivedi, 1986: 32-33; Greene, 2008: 586). NBRM için koşullu ortalama ve koşullu varyans aşağıdaki gibidir.

$$E(y_i | \mathbf{x}_i) = \mu_i = \exp(\mathbf{x}_i' \boldsymbol{\beta}) \quad (8.32)$$

$$V(y_i | \mathbf{x}_i) = \mu_i + \frac{\mu_i^2}{v} \quad (8.33)$$

Eşitlik (8.31)'de yer alan $v^{-1} = \alpha$ yayılım parametresini göstermektedir. Eşitlik (8.33)'te yer alan varyans incelendiğinde $\mu_i > 0$ ve $v > 0$ iken varyansın ortalamayı aştığı ve NBRM'nin aşırı yayılımı izin verdiği görülmektedir. Eşitlik (8.31)'de yer alan NBRM literatürde Negbin-2 olarak da bilinmektedir. NBRM için varyans ve ortalama arasındaki ilişki genel olarak (8.31)'de $v = \alpha^{-1} \mu_i^{2-P}$ şeklinde yazılarak elde edilir. Burada P varyans fonksiyonunun derecesini belirtmektedir ve Negbin-P olarak adlandırılır. Negbin-P modelinde $E(y_i | \mathbf{x}_i) = \mu_i$ ve varyans:

$$V(y_i | \mathbf{x}_i) = \mu_i + \alpha \mu_i^P \quad (8.34)$$

$P = 1$ için varyans fonksiyonu doğrusaldır ve Negbin-1 modeli elde edilmektedir. $P = 2$ olduğunda ise varyans fonksiyonu karesel formdadır ve Negbin-2 modeli olarak adlandırılmaktadır (Cameron ve Trivedi, 1986: 33). NBRM, içerisinde PRM'yi barındırmaktadır. Eğer $\alpha = 0$ ise eşitlik (8.31) PRM'ye indirgenmektedir (Lawless, 1987: 210). Literatürde yaygın olarak ortalaması μ ve varyansı $\mu + \alpha\mu^2$ olan Negbin-2 modeli kullanılmaktadır. Eşitlik (8.31)'de yer alan olasılık kütle fonksiyonu $v^{-1} = \alpha$ dönüşümü uygulanarak eşitlik (8.35)'teki gibi yazılır.

$$\Pr(Y = y_i | \mu, \alpha) = \frac{\Gamma(y_i + \alpha^{-1})}{\Gamma(y_i + 1)\Gamma(\alpha^{-1})} \left(\frac{\mu_i}{\mu_i + \alpha^{-1}} \right)^{y_i} \left(\frac{\alpha^{-1}}{\mu_i + \alpha^{-1}} \right)^{\alpha^{-1}} \quad (8.35)$$

Burada $\alpha \geq 0$ ve $y = 0, 1, 2, \dots$ şeklindedir. Eşitlik (8.35)'te yer alan $\Gamma(\cdot)$ gama fonksiyonudur. y 'nin tam sayı olması durumunda

$$\frac{\Gamma(y + \alpha)}{\Gamma(\alpha)} = \prod_{j=0}^{y-1} (j + \alpha) \quad (8.36)$$

eşitliği geçerlidir. Eşitlik (8.36) dikkate alınarak eşitlik (8.35)'te yer alan ifadenin logaritması

$$\frac{\Gamma(y_i + \alpha^{-1})}{\Gamma(\alpha^{-1})} = \sum_{j=0}^{y_i-1} (j + \alpha^{-1}) \quad (8.37)$$

şeklinde yazılır. Eşitlik (8.37) kullanılarak üstel ortalama için log-olabilirlik fonksiyonu aşağıdaki gibi elde edilmektedir:

$$\ln L(\alpha, \beta) = \sum_{i=1}^n \left(\left(\sum_{j=0}^{y_i-1} \ln(j + \alpha^{-1}) \right) - \ln y_i! - (y_i + \alpha^{-1}) \ln(1 + \alpha \exp(\mathbf{x}_i \beta)) + y_i \ln \alpha + y_i \mathbf{x}_i \beta \right) \quad (8.38)$$

$\hat{\beta}$ ve α parametrelerinin tahmini En Çok Olabilirlik yöntemi kullanılarak eşitlik (8.39) ve eşitlik (8.40)'ta yer alan eşitliklerin çözümüyle elde edilmektedir (Winkelmann, 2008: 135).

$$\sum_{i=1}^n \frac{y_i - \mu_i}{1 + \alpha \mu_i} \mathbf{x}_i = 0 \quad (8.39)$$

$$\sum_{i=1}^n \left(\frac{1}{\alpha^2} \left(\ln(1 + \alpha \mu_i) - \sum_{j=0}^{y_i-1} \frac{1}{(j + \alpha^{-1})} \right) + \frac{y_i - \mu_i}{\alpha(1 + \alpha \mu_i)} \right) = 0 \quad (8.40)$$

Cameron ve Trivedi (1986) tarafından önerilen Negbin-P modelinin $P = 1$ için özel bir hâli olan Negbin-1 modelinin log-olabilirlik fonksiyonu;

$$\ln L(\alpha, \beta) = \sum_{i=1}^n \left(\left(\sum_{j=0}^{y_i-1} \ln(j + \alpha^{-1} \exp(\mathbf{x}_i \beta)) \right) - \ln y_i! - (y_i + \alpha^{-1} \exp(\mathbf{x}_i \beta)) \ln(1 + \alpha) + y_i \ln \alpha \right) \quad (8.41)$$

şeklinde elde edilmektedir. $\hat{\beta}$ ve α parametrelerinin tahmini En Çok Olabilirlik yöntemi kullanılarak eşitlik (8.42) ve eşitlik (8.43)'ün çözümüyle elde edilmektedir (Winkelmann, 2008: 136).

$$\sum_{i=1}^n \left(\left(\sum_{j=0}^{y_i-1} \frac{\alpha^{-1} \mu_i}{(j + \alpha^{-1} \mu_i)} \right) \mathbf{x}_i + \alpha^{-1} \mu_i \mathbf{x}_i \right) = 0 \quad (8.42)$$

$$\sum_{i=1}^n \frac{1}{\alpha^2} \left(- \left(\sum_{j=0}^{y_i-1} \frac{\mu_i}{(j + \alpha^{-1})} \right) - \alpha^{-2} \mu_i \ln(1 + \alpha) - \frac{\alpha}{1 + \alpha} + y_i \alpha \right) = 0 \quad (8.43)$$

Şimdi aşağıda NBR modeli uygulamasına geçilecektir.

Stata Tahmini (nbreg)

NBRM'nin tahminlenmesi için Stata komutu olarak **nbreg** kullanılmaktadır. Komutun temel yapısı aşağıda belirtilmektedir.

nbreg depvar [indepvars] [if] [in] [weight] [,nbreg_options]

depvar negatif olmayan sayma değişkenini, **indepvars** ise modelde kullanılan bağımsız değişkenleri tanımlamaktadır. Komut yapısı hakkında detaylı bilgi için **help nbreg** komutu yardımıyla açılan bilgilendirme ekranından detaylara ulaşabilirsiniz.

Aşırı yayılımı test etmek önemlidir. Bunu yapmanın çeşitli yolları bulunmaktadır. Aşağıda ele alınan yaklaşımlar $\alpha = 0$ olduğunda NBRM'nin PRM'ye indirgenmesinden yararlanmaktadır:

1. $H_0: \alpha = 0$ hipotezi test edilir. Stata'nın **nbreg** komutu bunu otomatik olarak raporlar.

Tablo 8.13. Yataklı Tedavi Gün Sayısı Negatif Binom Regresyon Modeli

nbreg tedavisayisi doktor kronikhastalik bki_N bki_F, nolog

Negative binomial regression				Number of obs	=	12817
				LR chi2(4)	=	469.75
Dispersion	=	mean		Prob > chi2	=	0.0000
Log likelihood	=	-4788.8963		Pseudo R2	=	0.0468

tedavisayisi		Coef.	Std. Err.	z	P> z	[95% Conf. Interval]

doktor		.5645216	.0496915	11.36	0.000	.4671281 .661915
kronikhastalik		1.271045	.0835792	15.21	0.000	1.107233 1.434857
bki_N		.4357068	.2568064	1.70	0.090	-.0676245 .9390381
bki_F		.4926729	.2543888	1.94	0.053	-.0059199 .9912658
_cons		-3.323095	.2522385	-13.17	0.000	-3.817473 -2.828717

/lnalpha		2.504895	.0482216			2.410382 2.599407

alpha		12.24227	.5903421			11.13822 13.45576

Likelihood-ratio test of alpha=0: chibar2(01) = 7163.10 Prob>=chibar2 = 0.000

Çıktının son satırından görüldüğü gibi LR test istatistiği 7163.10 olarak elde edilmiştir. Test sonucuna göre ($\text{prob} < 0.05$) $\alpha = 0$ olduğunu söyleyen sıfır hipotezi reddedilmektedir ve aşırı yayılım olduğu sonucuna ulaşılmaktadır.

2. $\ln(\alpha) = 1$ için Wald testi yapılabilir ($\alpha = 0$ 'ın testine karşılık gelir.).

test [lnalpha]_cons = 1

```
( 1)  [lnalpha]_cons = 1
      chi2( 1) = 973.93
      Prob > chi2 = 0.0000
```

Bunu doğrulamak için Tablo 8.13'te yer alan lnalpha (2.504895) değeri 1'den çıkarılır ve standart hatasına bölünerek test istatistiği elde edilebilir:

$$\frac{2.504895 - 1}{0.0482216} = \frac{1.504895}{0.0482216} = 31.20790$$

Karesi alınırsa, 973.933 elde edilir. Elde edilen sonucun Stata'dan elde edilen sonuçla aynı olduğu görülmektedir.

3. Hem PRM hem de NBRM tahmin edilerek aşırı yayılımı test etmek amacıyla LR ki-kare testi yapılabilir.

Tablo 8.14. LR-Ki-Kare Test Sonucu

```
quietly poisson tedavisayisi doktor kronikhastalik bki_N bki_F , nolog
est store poisson
quietly nbreg tedavisayisi doktor kronikhastalik bki_N bki_F , nolog
est store nbreg
lrtest poisson nbreg, stats force
```

```
Likelihood-ratio test                                LR chi2(1) = 7163.10
(Assumption: poisson nested in nbreg)                 Prob > chi2 = 0.0000
Akaike's information criterion and Bayesian information criterion
```

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
poisson	12817	-9135.283	-8370.446	5	16750.89	16788.18
nbreg	12817	-5023.769	-4788.896	6	9589.793	9634.544

Note: N=Obs used in calculating BIC; see [R] BIC note

Tablo 8.14'te yer alan LR test istatistiği 7163.10 olarak elde edilmiş ve test sonucunda ($p < 0.000$) sıfır hipotezi reddedilerek aşırı yayılım olduğu bulgusuna ulaşılmıştır. Aşırı yayılım var olduğunda PRM kullanılamamaktadır. Bu nedenle, NBRM tercih edilmelidir. Elde edilen tahminler doğrudan

yorumlanamayacağı için incidence-rate ratio (IRR) değerinin elde edilmesi gereklidir. PRM'de olduğu gibi 2 farklı komut ile bu değerler elde edilebilmektedir (**listcoef** komutu sonuçları verilmemiştir).

Tablo 8.15. Yataklı Tedavi Gün Sayısı Negatif Binom Regresyon Modeline Ait IRR Değerleri

nbreg tedavisayisi doktor kronikhastalik bki_N bki_F, nolog irr

Negative binomial regression				Number of obs	=	12817
				LR chi2(4)	=	469.75
Dispersion = mean				Prob > chi2	=	0.0000
Log likelihood = -4788.8963				Pseudo R2	=	0.0468

tedavisayisi		IRR	Std. Err.	z	P> z	[95% Conf. Interval]

kronikhastalik	doktor	1.758606	.0873877	11.36	0.000	1.595406 1.938501
	bki_N	1.546055	.3970369	1.70	0.090	.9346113 2.55752
	bki_F	1.636685	.4163543	1.94	0.053	.9940976 2.694643
	_cons	.0360411	.009091	-13.17	0.000	.0219833 .0590886
	/lnalpha	2.504895	.0482216			2.410382 2.599407

	alpha	12.24227	.5903421			11.13822 13.45576

Likelihood-ratio test of alpha=0: chibar2(01) = 7163.10 Prob>=chibar2 = 0.000

Model sonuçları incelendiğinde uzman doktora gitme sayısı ve kronik hastalık durumu katsayıları %5 düzeyinde istatistiki olarak anlamlı iken BKİ'si normal ve fazla kiloda olanlar kategorilerinin katsayılarının %10 düzeyinde anlamlı olduğu görülmektedir. Marjinal etkilere ait sonuçlar ise yine **margins**, **dydx(*)** komutu ile elde edilmektedir.

Tablo 8.16. Yataklı Tedavi Gün Sayısı Negatif Binom Regresyon Modeline Ait Marjinal Etkiler

margins, dydx(*)

Average marginal effects

Number of obs = 12817

Model VCE : OIM

Expression : Predicted number of events, predict()

dy/dx w.r.t. : doktor kronikhastalik bki_N bki_F

		Delta-method				
	dy/dx	Std. Err.	z	P> z	[95% Conf. Interval]	
doctor	1.378834	1.266166	1.09	0.276	-1.102806	3.860474
kronikhastalik	3.104505	2.571653	1.21	0.227	-1.935843	8.144853
bki_N	1.064206	1.060282	1.00	0.316	-1.013909	3.142321
bki_F	1.203345	1.166618	1.03	0.302	-1.083185	3.489875

Tablo 8.16'da yer alan açıklayıcı değişkenlere ait marjinal etkilerin prob. değerleri 0.05'ten büyük olduğu için katsayılar istatistiksel olarak anlamlı değildir. Eğer bu durumu göz ardı ederek yorumlama yapılacak olursa; doktora gitme sayısındaki artış yataklı tedavi gün sayısını 1.379 artırmaktadır. En az bir kronik rahatsızlığı bulunanların kronik rahatsızlığı bulunmayanlara göre 3.105, normal kilolu olanların zayıf olanlara göre 1.064 ve fazla kiloluların zayıf olanlara göre 1.203 gün daha fazla hastanede yatarak tedavi edilmektedir.

PRM ve NBRM'nin yan yana karşılaştırması yapılmak istenirse est table komutu kullanılmalıdır:

Tablo 8.17. Poisson ve Negatif Binom Regresyon Modellerinin Karşılaştırılması

est table poisson nbreg, t varlabel varwidth(32) stats(alpha N) b(%9.3f)

Variable		poisson	nbreg
tedavisayisi			
Uzman Doktora Gitme Sayısı		0.213	0.565
		31.54	11.36
Yok ise 0, Var ise 1		1.347	1.271
		25.93	15.21
Normal Kilolu		0.586	0.436
		3.59	1.70
Fazla Kilolu		0.473	0.493
		2.90	1.94
Constant		-3.197	-3.323
		-19.53	-13.17
lnalpha			
Constant			2.505
			51.95
Statistics			
alpha			12.242
N		12817	12817

legend: b/t

Tablo 8.17'deki Stata komutunda yer alan t varlabel, değişkeni açıkça ifade etmek için kullanılmaktadır. varwidth(32) sütunlar arasındaki boşlukları ayarlamaktadır. stats(alpha N), alpha istatistiğini ve gözlem sayısını çıktıya eklemektedir. b(%9.3f) ise ondalık basamak sayısını ayarlamak için kullanılmaktadır. Tablo 8.17 incelendiğinde tahminlenen katsayıların benzer oldukları görülmektedir. PRM'den elde edilen t istatistikleri NBRM'den elde edilen t istatistiklerine göre daha yüksek değerlere sahiptir.

Poisson ve Negatif Binom regresyon modellerinden elde edilen tahmin edilmiş ve gerçek olasılıklar karşılaştırılmak istendiğinde **countfit** komutu kullanılır.

Tablo 8.18. PRM İçin Gerçek ve Tahminlenen Olasılıkların Karşılaştırılması

countfit tedavisayisi doktor kronikhastalik bki_N bki_F, maxcount (5) prm nograph no-estimates nofit

Comparison of Mean Observed and Predicted Count

Model	Maximum Difference	At Value	Mean Diff
PRM	-0.080	1	0.028

PRM: Predicted and actual probabilities

Count	Actual	Predicted	Diff	Pearson
0	0.919	0.840	0.079	95.710
1	0.060	0.140	0.080	581.438
2	0.010	0.018	0.008	47.262
3	0.004	0.002	0.002	21.681
4	0.001	0.000	0.001	24.031
5	0.001	0.000	0.001	54.505
Sum	0.995	1.000	0.170	824.627

Tablo 8.18'den PRM için gerçek ve tahminlenen olasılıklar arasındaki en büyük farkın 0.080 olduğu ve mutlak farkların ortalamasının 0.028 olduğu görülmektedir.

Tablo 8.19. NBRM için Gerçek ve Tahminlenen Olasılıkların Karşılaştırılması

countfit tedavisayisi doktor kronikhastalik bki_N bki_F, maxcount (5) nbreg nograph no-estimates nofit

Comparison of Mean Observed and Predicted Count

Model	Maximum Difference	At Value	Mean Diff
NBRM	0.016	1	0.005

NBRM: Predicted and actual probabilities

Count	Actual	Predicted	Diff	Pearson
0	0.919	0.921	0.002	0.046
1	0.060	0.044	0.016	77.069
2	0.010	0.015	0.005	24.515
3	0.004	0.007	0.003	16.884
4	0.001	0.004	0.003	23.678
5	0.001	0.002	0.001	10.402
Sum	0.995	0.993	0.031	152.594

Tablo 8.19 incelendiğinde NBRM için gerçek ve tahminlenen olasılıklar arasındaki en büyük farkın 0.016 olduğu ve mutlak farkların ortalamasının 0.005 olduğu görülmektedir.

Tablo 8.18 ve Tablo 8.19'dan görüleceği üzere, özellikle 0 ve 1 sayımlarında tahmin edilen olasılıklar ile gerçek olasılıklar arasındaki fark NBRM'de PRM'ye göre daha azdır. Gerçek ve tahminlenen olasılıklar arasındaki mutlak farkların toplamı karşılaştırıldığında, NBRM'den tahminlenen olasılıklarının PRM'den tahminlenenlere göre daha iyi uyduğu sonucuna ulaşılmaktadır. Her ne kadar bu değerler uyum iyiliğinin kaba bir göstergesi olsa da bu sayıların Pearson ki-kare uygunluk testinin temeli olarak kullanılmasında dikkatli olunmalıdır (Cameron ve Trivedi, 2009: 566)

Son olarak fitstat'' komutu kullanılarak Log-olabilirlik, McFadden's R^2 (Pseudo R^2), Akaike Information Criterion (AIC) ve Bayesian Information Criterion (BIC) gibi çeşitli iyi uyum istatistikleri aşağıda yer almaktadır.

Tablo 8.20. Yataklı Tedavi Gün Sayısı Negatif Binom Regresyon Modeline Ait Uyum İyiliği İstatistikleri

fitstat

Measures of Fit for nbreg of tedavisayisi

Log-Lik Intercept Only:	-5023.769	Log-Lik Full Model:	-4788.896
D(12811):	9577.793	LR(4):	469.746
		Prob > LR:	0.000
McFadden's R^2 :	0.047	McFadden's Adj R^2 :	0.046
ML (Cox-Snell) R^2 :	0.036	Cragg-Uhler(Nagelkerke) R^2 :	0.066
AIC:	0.748	AIC*n:	9589.793
BIC:	-111595.406	BIC':	-431.912
BIC used by Stata:	9634.544	AIC used by Stata:	9589.793

Tablo 8.20'de yer alan NBRM'ye ait uyum istatistikleri modellerin karşılaştırılmasında ve en uygun modelin seçilmesinde kullanılmaktadır. Elde edilen uyum istatistikleri 8.4 no'lu kısımda yer alan Tablo 8.30'dan sonra yorumlanacaktır.

Öncelikle PRM tahminlenerek aşırı yayılımın varlığı test edilmiştir. Aşırı yayılım söz konusu olduğundan dirençli Poisson regresyon modeli ve Negatif binom regresyon modelleri tahminlenmiştir. Şekil 8.1'de yer alan bağımlı sayma değişkenine ait histogramda, hastanede geçirilen gün sayısının %91.87'si sıfır değerinden oluştuğu için bu aşamadan sonra sıfır yığılmalı modeller anlatılarak Stata uygulaması yapılacaktır.

8.3. SIFIR YIĞILMALI SAYMA MODELLERİ

Bağımlı değişkene ait sayma verilerinin beklenenden fazla sıfır değerine sahip olması durumunda, sıfır değerlerini dikkate alan sıfır yığılmalı (zero inflated) modellerin kullanılması gerekmektedir. Sıfır yığılmalı sayma verilerinin analizi için literatürde birçok yöntem mevcuttur. Sıfır Yığılmalı Poisson regresyon (Zero Inflated Poisson = ZIP) modeli ve Sıfır Yığılmalı Negatif Bi-

nom regresyon (Zero Inflated Negative Binomial = ZINB) modeli yaygın olarak kullanılan temel modellerdendir.

8.3.1. Sıfır Yığılmalı Poisson Regresyon Modeli

Lambert (1992) tarafından sıfır değerlerinin yoğun olduğu sayma verilerinin analiz edilmesi amacıyla ZIP modeli ortaya konulmuştur. Örneğin; motorlu araç kasko verilerinde sürücülerin hasarsızlık indiriminden yararlanmak amacıyla küçük tazminat taleplerini bildirmemesi nedeniyle sıfır verileri yoğunluk gösterebilir (Yip ve Yau, 2005). ZIP modeli hem sıfır sürecinde hem de Poisson dağılımında açıklayıcı değişkenlerin dahil edilmesine izin verilerek, sıfırda dejenere olmuş bir dağılımın Poisson dağılımı ile birleştirilmesiyle elde edilmektedir (Ismail ve Zamani, 2013: 9). ZIP modeli olasılık kütle fonksiyonu aşağıdaki gibi tanımlanmaktadır (Lambert, 1992: 3).

$$\Pr(Y_i = y_i | \mathbf{x}_i) = \begin{cases} p_i + (1 - p_i)e^{-\mu_i}, & y_i = 0 \\ (1 - p_i)\frac{\mu_i^{y_i}}{y_i!}e^{-\mu_i}, & y_i > 0 \end{cases} \quad (8.44)$$

Burada $0 \leq p_i \leq 1$ ve $\mu_i > 0$ şeklindedir. ZIP modeli için koşullu ortalama ve koşullu varyans sırasıyla Eşitlik (8.45)'te ve eşitlik (8.46)'da yer almaktadır.

$$E(y_i | \mathbf{x}_i) = (1 - p_i)\mu_i \quad (8.45)$$

$$V(y_i | \mathbf{x}_i) = (1 - p_i)\mu_i(1 + p_i\mu_i) \quad (8.46)$$

ZIP modeli $p_i = 0$ olduğundan PRM'ye indirgenmekte ve $p_i > 0$ olduğunda ise aşırı yayılım gözlemlenmektedir (Ismail ve Zamani, 2013: 9). Parametreler $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)$ ve $\mathbf{p} = (p_1, \dots, p_n)$, sırasıyla aşağıdaki bağlantı fonksiyonlarına sahiptirler.

$$\ln(\boldsymbol{\mu}) = \mathbf{B}\boldsymbol{\beta} \quad (8.47)$$

$$\text{logit}(\mathbf{p}) = \ln\left(\frac{\mathbf{p}}{1 - \mathbf{p}}\right) = \mathbf{G}_\gamma \quad (8.48)$$

$\mathbf{B}$ ve $\mathbf{G}$ ortak değişken matrislerini, $\boldsymbol{\beta}$ ve $\boldsymbol{\gamma}$ bilinmeyen parametre vektörlerini ifade etmektedir. $\mathbf{B}$ matrisi sayma değişkenindeki Poisson çıktısından sorumluyken $\mathbf{G}$ matrisi ise aşırı sıfırlardan sorumludur. ZIP modeli için log-olabilirlik fonksiyonu;

$$\ln L(\boldsymbol{\gamma}, \boldsymbol{\beta}) = \sum_{y_i=0} \ln(e^{G_\gamma} + e^{-\mu_i^{\mathbf{B}, \boldsymbol{\beta}}}) + \sum_{y_i>0} (y_i \mathbf{B}_i \boldsymbol{\beta} - e^{\mathbf{B}_i \boldsymbol{\beta}}) - \sum_{i=1}^n \ln(1 + e^{G_\gamma}) - \sum_{y_i>0} \ln(y_i!) \quad (8.49)$$

şeklindedir. Eşitlik (8.49)'da yer alan $\mathbf{B}_i$ ve $\mathbf{G}_i$, $\mathbf{B}$ ve $\mathbf{G}_\gamma$ matrislerinin i . satırını ifade etmektedir (Lambert, 1992: 4). Lambert (1992) parametrelerin tahminlenmesinde EM (Expectation-Maximization) algoritmasını önermiştir. Greene (1994) gradyan vektörünü elde ederek BHHH algoritması ile En Çok Olabilirlik tahminlerinin elde edilmesini sağlamıştır (Long, 1997: 244)

ZIP modeli ile PRM'yi karşılaştırmak amacıyla Vuong (1989) tarafından önerilen test istatistiği kullanılmaktadır. Vuong test istatistiği:

$$v = \frac{\sqrt{n} \left((1/n) \sum_{i=1}^n m_i \right)}{\sqrt{(1/n) \sum_{i=1}^n (m_i - \bar{m})^2}} = \frac{\sqrt{n\bar{m}}}{S_m} \quad (8.50)$$

ile bulunur. Eşitlik (8.50)'de yer alan m_i eşitlik (8.51)'deki gibi hesaplanmaktadır.

$$m_i = \ln \left(\frac{f_1(y_i | \mathbf{x}_i)}{f_2(y_i | \mathbf{x}_i)} \right) \quad (8.51)$$

$f_1(y_i | \mathbf{x}_i)$ ve $f_2(y_i | \mathbf{x}_i)$ karşılaştırılmak istenen modellere ait olasılık kütle fonksiyonlarını göstermektedir. Örneğin; $f_1(y_i | \mathbf{x}_i)$ ZIP dağılımının kütle fonksiyonunu gösterirken, $f_2(y_i | \mathbf{x}_i)$ Poisson dağılımına ait kütle fonksiyonunu belirtmektedir. Eşitlik (8.50)'de yer alan $\bar{m}$ ortalamayı, S_m standart sapmayı ve n örneklem büyüklüğünü göstermektedir. Vuong istatistiği $E(m_i)$ 'nin sıfıra eşit olup olmadığını test etmektedir. Vuong (1989) v test istatistiğinin asimptotik olarak standart normal dağılıma sahip olduğunu göstermiştir. Test istatistiği iki yönlüdür. Elde edilen test istatistiği değeri mutlak değerce t tablo değerinden daha küçükse test sonucu herhangi bir modelin lehine karar vermemektedir. t tablo değerinden daha büyük pozitif test istatistiği değeri $f_1(y_i | \mathbf{x}_i)$ ile gösterilen modelin, t tablo değerinden daha küçük negatif test istatistiği değeri $f_2(y_i | \mathbf{x}_i)$ ile gösterilen modelin seçilmesi gerektiğini ifade etmektedir (Greene, 2018: 906).

Vuong testi bir örnek üzerinden ele alınsın.

H_0 : PRM tercih edilmelidir.

H_1 : ZIP modeli tercih edilmelidir.

Elde edilen v test istatistiği değeri %95 güven düzeyinde t tablo değeri 1.96 ile karşılaştırılır. Eğer v değeri 1.96'dan daha büyükse sıfır hipotezi reddedildiğinden ZIP modeli tercih edilmelidir. Eğer v değeri -1.96'dan daha küçükse sıfır hipotezi reddedilemediğinden PRM tercih edilmelidir. Elde edilen test istatistiği değeri -1.96 ile 1.96 arasında ise herhangi bir model lehine karar verilememektedir (Shankar ve diğerleri, 1997: 831).

Stata Tahmini (zip)

İki kısımdan oluşan sıfır değer ağırlıklı poisson tahminleri için Stata komutu olarak **zip** kullanılmaktadır. Komutun temel yapısı aşağıda belirtilmektedir.

zip depvar [indepvars] [if] [in] [weight], inflate(varlist) [options]

inflate(varlist), saymanın logit (varsayılan) olma olasılığını belirleyen değişkenleri (varsa) ifade ederken, diğer seçenekler poisson ile aynıdır. Zip komutunu kullanmak için Statada inflate kısmına gözlemin sıfır değerini almasına sebep olabilecek değişkenler yazılır. inflate () seçeneği yalnızca sabit terim içerebileceği gibi ZIP modelinde yer alan değişkenleri de içerebilir. Fakat

sıfır saymaya neden olduğunu düşündüğünüz herhangi bir değişkeniniz yoksa sadece *inflate* (*_cons*) diye belirtebilirsiniz ve Stata bu olasılığı tüm gözlemler için aynı olduğu varsayımına göre hesaplar. Komut yapısı hakkında detaylı bilgi için **help zip** komutu yardımıyla açılan bilgilendirme ekranından detaylara ulaşabilirsiniz.

Sayma sürecini dikkate alan ilk kısımda tüm değişkenler kullanılmışken, sayımın sıfır olup olmadığını etkilemesi beklenen *inflate* kısmında sadece son 12 ay içerisinde bireylerin uzman doktora gitme sayısı değişkeni kullanılmıştır. ZIP modeli ile PRM'yi karşılaştırmak amacıyla kullanılan Vuong istatistiği içinde **vuong** komutu Tablo 8.21'de yer alan komutun sonuna eklenmiştir.

Tablo 8.21. Yataklı Tedavi Gün Sayısı ZIP Modeli

zip tedavisayisi doktor kronikhastalik bki_N bki_F, inflate (doktor) vuong nolog

Zero-inflated Poisson regression				Number of obs	=	12817
				Nonzero obs	=	1042
				Zero obs	=	11775
Inflation model = logit				LR chi2(4)	=	409.41
Log likelihood = -6237.629				Prob > chi2	=	0.0000

tedavisayisi		Coef.	Std. Err.	z	P> z	[95% Conf. Interval]

tedavisayisi						
doktor		.0749784	.0093482	8.02	0.000	.0566562 .0933006
kronikhastalik		1.206088	.0668003	18.06	0.000	1.075162 1.337014
bki_N		.6797953	.1914331	3.55	0.000	.3045934 1.054997
bki_F		.5332921	.1902785	2.80	0.005	.160353 .9062311
_cons		-.9729096	.1956486	-4.97	0.000	-1.356374 -.5894454

inflate						
doktor		-.3292178	.0283291	-11.62	0.000	-.3847419 -.2736938
_cons		2.199283	.0430812	51.05	0.000	2.114845 2.28372

Vuong test of zip vs. standard Poisson:				z =	6.48	Pr>z = 0.0000

Tablo 8.21'de yer alan model çıktısına bakıldığında 12817 gözlemin 11775'i için bağımlı değişkenin sıfır değerini aldığı, 1042'sinin ise 1 değerini aldığı görülmektedir. LR-ki kare değeri 409.41 ve prob<0.000 olduğundan modelin anlamlı olduğu söylenir. Elde edilen tahmin sonuçlarına göre; modelin birinci kısmı Poisson katsayıları gibi yorumlanır. Katsayılar istatistiksel olarak anlamlıdır. Uzman doktora gitme sayısı, kronik hastalığın varlığı, zayıf bireylere göre normal ve fazla kilolu olma bireyin hastanede geçirdiği yataklı tedavi gün sayısı üzerinde pozitif etkilidir. İkinci kısım (logit), sıfırları tahmin etmektedir. Birey ya doktora gitmiyordur veya yatarak tedavi olan hasta değildir ve katsayının işareti negatif çıkmıştır. Modelde birinci ve ikinci kısımda aynı değişkenler yer alıyorsa genellikle katsayıların işaretleri ters yönlüdür.

Tablo 8.21'in son satırından görüleceği üzere Vuong testi sıfır değer ağırlıklı poisson modeli ile sıradan bir poisson regresyon modelini karşılaştırır.

$H_0 : V(\text{ZIP} / \text{PRM}) < \text{kritikdeğer}$ (PRM modeli tercih edilmelidir.)

$H_1 : V(\text{ZIP} / \text{PRM}) > \text{kritikdeğer}$ (ZIP modeli tercih edilmelidir.)

ZIP modelini ve PRM'yi karşılaştırmak için Tablo 8.21'de yer alan Vuong istatistiğine bakıldığında, $6.48 > 1.96$ veya $\text{prob.} < 0.000$ olduğundan H_0 hipotezi reddedilmektedir. ZIP modeli PRM'ye tercih edilir. Artık Tablo 8.21'deki model ile çalışılacaktır; ancak Tablo 8.21'den elde edilen tahminler doğrudan yorumlanamayacağı için incidence rate ratio (IRR) değerlerinin hesaplanması gerekir ve bunun için model tahminine `irr` komutu eklenir.

Tablo 8.22. Yataklı Tedavi Gün Sayısı ZIP Modeline Ait IRR Değerleri

zip tedavisayisi doktor kronikhastalik bki_N bki_F, irr inflate (doktor) vuong nolog

Zero-inflated Poisson regression				Number of obs	=	12817
				Nonzero obs	=	1042
				Zero obs	=	11775
Inflation model = logit				LR chi2(4)	=	409.41
Log likelihood = -6237.629				Prob > chi2	=	0.0000

tedavisayisi		IRR	Std. Err.	z	P> z	[95% Conf. Interval]

tedavisayisi						
doktor		1.077861	.0100761	8.02	0.000	1.058292 1.097792
kronikhastalik		3.340392	.2231392	18.06	0.000	2.930468 3.807658
bki_N		1.973474	.3777881	3.55	0.000	1.356073 2.871967
bki_F		1.704535	.3243363	2.80	0.005	1.173925 2.474977
_cons		.3779817	.0739516	-4.97	0.000	.2575932 .5546348

inflate						
doktor		-.3292178	.0283291	-11.62	0.000	-.3847419 -.2736938
_cons		2.199283	.0430812	51.05	0.000	2.114845 2.28372

Vuong test of zip vs. standard Poisson:				z =	6.48	Pr>z = 0.0000

Tablo 8.22'de yer alan ilk kısım için yorumlar aynı PRM'de olduğu gibi yapılmaktadır. İlk kısım için kullanılan tüm açıklayıcı değişkenlerin katsayılarının istatistiksel olarak anlamlı olduğu görülmektedir. Diğer değişkenler sabit tutulduğunda uzman doktora gitme sayısındaki artışın, yataklı tedavi gün sayısını %7.79 artırdığı söylenir. Kronik rahatsızlığı olanlar, olmayanlardan %234.04 (2.34 kat) daha fazla hastanede yatmaktadır. BKİ'si normal olanlar, zayıf olanlardan %97.35 daha fazla gün hastanede yatarak tedavi alırken, BKİ'si fazla olanlar, zayıf olanlardan %70.45 daha fazla gün hastanede yataklı tedavi görmektedir. İkinci kısım (logit) için: Son 12 ay içerisinde bireyin fazladan bir kez daha uzman doktora gitmesi, hastanede yataklı tedavi hizmeti almama olasılığını yaklaşık $[(e^{-0.329} - 1) * 100]$ %28 azaltmaktadır yorumu yapılır.

Marjinal etkiler ise yine **margins dydx(*)** komutu ile elde edilir.

Tablo 8.23. Yataklı Tedavi Gün Sayısı ZIP Regresyon Modeline Ait Marjinal Etkiler**margins, dydx(*)**

Average marginal effects				Number of obs	=	12817	
Model VCE : OIM							
Expression : Predicted number of events, predict()							
dy/dx w.r.t. : doktor kronikhastalik bki_N bki_F							

		Delta-method					
		dy/dx	Std. Err.	z	P> z	[95% Conf. Interval]	

doktor		.0646639	.0046092	14.03	0.000	.0556301	.0736978
kronikhastalik		.2240553	.015428	14.52	0.000	.1938171	.2542936
bki_N		.1262858	.0359335	3.51	0.000	.0558575	.1967141
bki_F		.0990698	.035589	2.78	0.005	.0293166	.168823

Tablo 8.23'te açıklayıcı değişkenlere ait marjinal etkiler sırasıyla yaklaşık; 0.065, 0.224, 0.126 ve 0.099 olarak elde edilmiştir. Doktora gitme sayısındaki artış yataklı tedavi gün sayısını 0.065 arttırmaktadır. En az bir kronik rahatsızlığı bulunanlar kronik rahatsızlığı bulunmayanlara göre 0.224, normal kilolu olanlar zayıf olanlara göre 0.126 ve fazla kilolular zayıf olanlara göre 0.099 gün daha fazla hastanede yatarak tedavi görmektedir.

Elde edilen ZIP modeline ilişkin uyum ölçüleri için **fitstat** komutu kullanılır.

Tablo 8.24. Yataklı Tedavi Gün Sayısı ZIP Modeline Ait Uyum İyiliği İstatistikleri**fitstat**

Measures of Fit for zip of tedavisayisi			
Log-Lik Intercept Only:	-6573.298	Log-Lik Full Model:	-6237.629
D(12810):	12475.258	LR(5):	671.338
		Prob > LR:	0.000
McFadden's R2:	0.051	McFadden's Adj R2:	0.050
ML (Cox-Snell) R2:	0.051	Cragg-Uhler(Nagelkerke) R2:	0.080
AIC:	0.974	AIC*n:	12489.258
BIC:	-108688.481	BIC':	-624.046
BIC used by Stata:	12541.468	AIC used by Stata:	12489.258

Tablo 8.24'te yer alan istatistikler modellerin karşılaştırılması ve en uygun modelin belirlenmesi amacıyla kullanılmaktadır. Elde edilen uyum istatistikleri 8.4 no'lu kısımda yer alan Tablo 8.30'dan sonra yorumlanacaktır.

8.3.2. Sıfır Yığılmalı Negatif Binom Regresyon Modeli

Sayma değişkenindeki sıfırların yoğunluğu ile birlikte aşırı yayılımı dikkate alan ZINB modeli Greene (1994) tarafından ZIP modeline alternatif olarak geliştirilmiştir. ZINB modeli hem sıfır sürecinde hem de negatif binom dağılımında açıklayıcı değişkenlerin dahil edilmesine izin verilerek, sıfırda dejenere olmuş bir dağılımın negatif binom dağılımı ile birleştirilmesiyle elde edilmektedir (Ismail ve Zamani, 2013: 10). ZINB modeli için olasılık kütle fonksiyonu aşağıdaki gibidir (Long, 1997: 245).

$$\Pr(Y_i = y_i | \mathbf{x}_i) = \begin{cases} p_i + (1 - p_i) \left(\frac{\alpha^{-1}}{\mu_i + \alpha^{-1}} \right)^{\alpha^{-1}}, & y_i = 0 \\ (1 - p_i) \frac{\Gamma(y_i + \alpha^{-1})}{\Gamma(y_i + 1) \Gamma(\alpha^{-1})} \left(\frac{\mu_i}{\mu_i + \alpha^{-1}} \right)^{y_i} \left(\frac{\alpha^{-1}}{\mu_i + \alpha^{-1}} \right)^{\alpha^{-1}}, & y_i > 0 \end{cases} \quad (8.52)$$

ZINB modeli için koşullu ortalama ve koşullu varyans sırasıyla Eşitlik (8.53) ve eşitlik (8.54)'te yer almaktadır.

$$E(y_i | \mathbf{x}_i) = (1 - p_i) \mu_i \quad (8.53)$$

$$V(y_i | \mathbf{x}_i) = (1 - p_i) \mu_i (1 + \mu_i (p_i + \alpha)) \quad (8.54)$$

ZINB modeli $p_i = 0$ olduğundan NBRM'ye indirgenmektedir. Eğer $\alpha = 0$ ise ZINB modeli ZIP modeli hâline gelmektedir (Ismail ve Zamani, 2013: 10). Parametrelerin tahminlenmesi Newton-Raphson yöntemi gibi nümerik optimizasyon yöntemleri ile gerçekleştirilmektedir.

Stata Tahmini (zinb)

İki kısımdan oluşan sıfır değer ağırlıklı negatif binom tahminleri için Stata komutu olarak **zinb** kullanılmaktadır. Komutun temel yapısı aşağıda belirtilmektedir.

zinb depvar [indepvars] [if] [in] [weight], inflate(varlist) [options]

ZINB modelinde negatif binom sayma modeli ve aşırı sıfırları tahmin etmek için logit modeli olmak üzere iki bölüm vardır. ZINB hem bir sayım modeline hem de bir logit modeline sahip olduğundan her iki modelin de iyi tahminleri olmalıdır ve iki modelin de mutlaka aynı değişkenleri kullanması gerekmez. Sayımın sıfır olup olmadığını etkilemesi beklenen değişken veya değişkenler **inflate(varlist)** kısmında belirtilir. Komut yapısı hakkında detaylı bilgi için **help zinb** komutu yardımıyla açılan bilgilendirme ekranından detaylara ulaşabilirsiniz.

Sayma sürecini dikkate alan ilk kısımda tüm değişkenler kullanılmışken, sayımın sıfır olup olmadığını etkilemesi beklenen inflate kısmında sadece son 12 ay içerisinde bireylerin uzman doktora gitme sayısı değişkeni kullanılmıştır. ZINB modelini ve NBRM'yi karşılaştırmayı sağlayan ve aynı zamanda ZIP modeli için de $\alpha = 0$ olasılık oranı testini veren Vuong istatistiği için “**vuong**” seçeneği Stata'da en sona eklenmiştir.

Tablo 8.25. Yataklı Tedavi Gün Sayısı ZINB Modeli

zinb tedavisayisi doktor kronikhastalik bki_N bki_F, inflate (doktor) vuong zip nolog

Zero-inflated negative binomial regression				Number of obs	=	12817
				Nonzero obs	=	1042
				Zero obs	=	11775
Inflation model = logit				LR chi2(4)	=	252.39
Log likelihood = -4757.171				Prob > chi2	=	0.0000

tedavisayisi		Coef.	Std. Err.	z	P> z	[95% Conf. Interval]

tedavisayisi						
doktor		.2844715	.0470475	6.05	0.000	.1922601 .3766829
kronikhastalik		1.208825	.0857611	14.10	0.000	1.040736 1.376914
bki_N		.5432054	.2596866	2.09	0.036	.0342291 1.052182
bki_F		.5480653	.256926	2.13	0.033	.0444997 1.051631
_cons		-2.792548	.2621829	-10.65	0.000	-3.306417 -2.278679

inflate						
doktor		-14.61646	847.6642	-0.02	0.986	-1676.008 1646.775
_cons		-.1655304	.1589209	-1.04	0.298	-.4770096 .1459489

/lnalpha		2.140738	.0632508	33.85	0.000	2.016769 2.264707

alpha		8.505714	.5379931			7.514007 9.628307

Likelihood-ratio test of alpha=0: chibar2(01) = 2960.92 Pr>=chibar2 = 0.0000						
Vuong test of zinb vs. standard negative binomial: z = 2.96 Pr>z = 0.0015						

Öncelikle ZINB ve ZIP modellerini karşılaştırmak ve aşırı yayılımın varlığını araştırmak gerekmektedir.

$H_0 : \alpha = 0$ (ZIP modeli tercih edilmelidir.)

$H_1 : \alpha \neq 0$ (ZIP yerine ZINB tercih edilmelidir.)

hipotezleri dikkate alınırsa,

2960.92 test istatistiğine ait prob. değeri $0.000 < p$ (likelihood ratio test for $\alpha=0$ için) olduğundan H_0 hipotezi reddedilmektedir. ZIP modeli yerine ZINB modeli tercih edilmelidir. İstatistiksel olarak aşırı yayılım çıkmıştır.

ZINB modeli ve NBRM'yi karşılaştıran Vuong istatistiğine bakıldığında;

$H_0 : V(\text{ZINB} / \text{NBRM}) < \text{kritikdeğer}$ (NBRM tercih edilmelidir.)

$H_1 : V(\text{ZINB} / \text{NBRM}) > \text{kritikdeğer}$ (ZINB modeli tercih edilmelidir.).

Vuong testinde $2.96 > 1.96$ veya prob. değeri $0.0015 < p$ olduğundan H_0 hipotezi reddedilmektedir. ZINB modeli NBRM'ye tercih edilmektedir.

Tablo 8.25'te elde edilen çıktı yorumlandığında Negatif binom regresyon modelinin hastanede geçirilen yataklı tedavi gün sayısını (sayımını) tahmin eden tüm açıklayıcı değişkenlerinin katsayıları istatistiksel olarak anlamlıdır. Logit model kısmında yer alan uzman doktora gitme sayısının, hastanede yataklı tedavi hizmeti alma olasılığı üzerinde anlamlı bir etkisi yoktur, katsayı istatistiksel olarak anlamsızdır. Bu sonuç, Vuong test sonucu ile çelişmektedir ve modelin dirençli tahmincilerinin elde edilmesi gerekir.

Dirençli ZINB modeli aşağıdaki gibi tahminlenmiştir.

Tablo 8.26. Yataklı Tedavi Gün Sayısı Dirençli ZINB Regresyon Modeli

zinb tedavisayisi doktor kronikhastalik bki_N bki_F, inflate (doktor) robust nolog

Zero-inflated negative binomial regression				Number of obs	=	12817
				Nonzero obs	=	1042
				Zero obs	=	11775
Inflation model = logit				Wald chi2(4)	=	119.59
Log pseudolikelihood = -4757.171				Prob > chi2	=	0.0000

tedavisayisi	Coef.	Robust Std. Err.	z	P> z	[95% Conf. Interval]	

tedavisayisi						
doktor	.2844715	.0736258	3.86	0.000	.1401676	.4287754
kronikhastalik	1.208825	.1248151	9.68	0.000	.9641918	1.453458
bki_N	.5432054	.2701354	2.01	0.044	.0137498	1.072661
bki_F	.5480653	.2632778	2.08	0.037	.0320503	1.06408
_cons	-2.792548	.27152	-10.28	0.000	-3.324717	-2.260379

inflate						
doktor	-14.61646	3.269011	-4.47	0.000	-21.0236	-8.209313
_cons	-.1655304	.1971877	-0.84	0.401	-.5520112	.2209504

/lnalpha	2.140738	.0940368	22.76	0.000	1.956429	2.325047

alpha	8.505714	.7998499			7.074024	10.22716

Tablo 8.26'da yer alan dirençli ZINB modelinin logit kısmında doktora gitme sayısı istatistiksel olarak anlamlı hâle gelmiştir. Katsayıların işaretleriyle ilgili yorumlar Tablo 8.21'deki ZIP modeli gibi yapılır. Tablo 8.26'dan elde edilen tahminler doğrudan yorumlanamayacağı için IRR değerlerinin hesaplanması gerekir ve bunun için model tahminine **irr** komutu eklenir.

Tablo 8.27. Yataklı Tedavi Gün Sayısı Dirençli ZINB Modeline Ait IRR Değerleri

zinb tedavisayisi doktor kronikhastalik bki_N bki_F, inflate (doktor) irr robust nolog

Zero-inflated negative binomial regression Number of obs = 12817
 Nonzero obs = 1042
 Zero obs = 11775

Inflation model = logit Wald chi2(4) = 119.59
 Log pseudolikelihood = -4757.171 Prob > chi2 = 0.0000

tedavisayisi	IRR	Robust Std. Err.	z	P> z	[95% Conf. Interval]	
tedavisayisi						
doktor	1.329059	.0978531	3.86	0.000	1.150467	1.535376
kronikhastalik	3.349546	.418074	9.68	0.000	2.622667	4.277882
bki_N	1.721516	.4650425	2.01	0.044	1.013845	2.923148
bki_F	1.729903	.4554451	2.08	0.037	1.032569	2.898172
_cons	.0612649	.0166346	-10.28	0.000	.0359827	.104311
inflate						
doktor	-14.61646	3.269011	-4.47	0.000	-21.0236	-8.209313
_cons	-.1655304	.1971877	-0.84	0.401	-.5520112	.2209504
/lnalpha	2.140738	.0940368	22.76	0.000	1.956429	2.325047
alpha	8.505714	.7998499			7.074024	10.22716

Tablo 8.27’de yer alan ilk kısım için kullanılan tüm açıklayıcı değişkenlerin katsayılarının istatistiksel olarak anlamlı olduğu görülmektedir. Diğer değişkenler sabit tutulduğunda uzman doktora gitme sayısındaki artışın, yataklı tedavi gün sayısını %32.91 arttırdığı söylenir. Kronik rahatsızlığı olanlar, olmayanlardan %234.95 (2.35 kat) daha fazla hastanede yatmaktadır. BKİ’si normal olanlar, zayıf olanlara göre %72.15 daha fazla gün hastanede yatarak tedavi alırken BKİ’si fazla olanlar, zayıf olanlara göre %72.99 daha fazla gün hastanede yataklı tedavi görmektedir. İkinci kısma (logit) bakıldığında son 12 ay içerisinde bireyin fazladan bir kez daha uzman doktora gitmesi, hastanede yataklı tedavi hizmeti almama olasılığını yaklaşık $[(e^{-14.61}-1)*100]$ %99.99 azaltmaktadır yorumu yapılır.

Marjinal etkiler ise yine **margins dydx(*)** komutu ile elde edilir.

Tablo 8.28. Yataklı Tedavi Gün Sayısı Dirençli ZINB Regresyon Modeline Ait Marjinal Etkiler**margins, dydx(*)**

Average marginal effects

Number of obs = 12817

Model VCE : Robust

Expression : Predicted number of events, predict()

dy/dx w.r.t. : doktor kronikhastalık bki_N bki_F

	Delta-method					
	dy/dx	Std. Err.	z	P> z	[95% Conf. Interval]	
-----+-----						
doktor	.6689327	.1418361	4.72	0.000	.3909391	.9469264
kronikhastalık	.2616532	.0617338	4.24	0.000	.1406572	.3826492
bki_N	.1175782	.063917	1.84	0.066	-.0076968	.2428531
bki_F	.1186301	.0607458	1.95	0.051	-.0004295	.2376897
-----+-----						

Tablo 8.28'de bağımsız değişkenlere ait marjinal etkiler sırasıyla yaklaşık; 0.669, 0.262, 0.118 ve 0.119 olarak elde edilmiştir. Beden kitle indeksi normal ve fazla olanlara ait marjinal etkilerin istatistiki olarak %10'da anlamlı olduğuna dikkat edilmelidir. Doktora gitme sayısındaki artış yataklı tedavi gün sayısını 0.669 arttırmaktadır. En az bir kronik rahatsızlığı bulunanlar kronik rahatsızlığı bulunmayanlara göre 0.262, normal kilolu olanlar zayıf olanlara göre 0.118 ve fazla kilolular zayıf olanlara göre 0.119 gün daha fazla hastanede yatarak tedavi görmektedir.

Elde edilen modele ilişkin uyum ölçülerini görmek için **fitstat** komutu kullanılır. Sonuçlar Tablo 8.29'daki gibidir:

Tablo 8.29. Yataklı Tedavi Gün Sayısı ZINB Modeline Ait Uyum İyiliği İstatistikleri**fitstat**

Measures of Fit for zinb of tedavisayisi

Log-Lik Intercept Only:	-5023.769	Log-Lik Full Model:	-4757.171
D(12809):	9514.341	LR(5):	533.198
		Prob > LR:	0.000
McFadden's R2:	0.053	McFadden's Adj R2:	0.051
ML (Cox-Snell) R2:	0.041	Cragg-Uhler(Nagelkerke) R2:	0.075
AIC:	0.744	AIC*n:	9530.341
BIC:	-111639.940	BIC':	-485.905
BIC used by Stata:	9590.009	AIC used by Stata:	9530.341

Tablo 8.29'da yer alan ZINB modeline ait uyum istatistiklerinden modellerin karşılaştırılması ve en uygun modelin seçilmesi amacıyla yararlanılmaktadır. Elde edilen uyum istatistikleri 8.4 no'lu kısımda yer alan Tablo 8.30'dan sonra yorumlanacaktır.

8.4. SAYMALI MODELLERİN KARŞILAŞTIRILMASI

Uygulamada tahminlenen Poisson, negatif binom ve sıfır yığılmalı modelleri birlikte karşılaştırmak amacıyla Statada aşağıdaki **countfit** komutu kullanılmaktadır.

Tablo 8.30. Saymalı Modellerin Karşılaştırılması

countfit tedavisayisi doktor kronikhastalik bki_N bki_F, inflate (doktor) replace

Birinci kısım

Variable	PRM	NBRM	ZIP	ZINB
tedavisayisi				
Uzman Doktora Gitme Sayısı	1.237	1.759	1.078	1.329
	31.54	11.36	8.02	6.05
Yok ise 0, Var ise 1	3.847	3.565	3.340	3.350
	25.93	15.21	18.06	14.10
Normal Kilolu	1.797	1.546	1.973	1.722
	3.59	1.70	3.55	2.09
Fazla Kilolu	1.604	1.637	1.705	1.730
	2.90	1.94	2.80	2.13
Constant	0.041	0.036	0.378	0.061
	-19.53	-13.17	-4.97	-10.65
lnalpha				
Constant		12.242		8.506
		51.95		33.85
inflate				
Uzman Doktora Gitme Sayısı			0.719	0.000
			-11.62	-0.02
Constant			9.019	0.847
			51.05	-1.04
Statistics				
alpha		12.242		
N	12817	12817	12817	12817
ll	-8370.446	-4788.896	-6237.629	-4757.171
bic	16788.185	9634.544	12541.468	9590.009
aic	16750.892	9589.793	12489.258	9530.341

legend: b/t

İkinci Kısım

Comparison of Mean Observed and Predicted Count

Model	Maximum Difference	At Value	Mean Diff
PRM	-0.080	1	0.017
NBRM	0.016	1	0.003
ZIP	0.028	1	0.006
ZINB	0.018	1	0.004

Üçüncü kısım

PRM: Predicted and actual probabilities

Count	Actual	Predicted	Diff	Pearson
0	0.919	0.840	0.079	95.710
1	0.060	0.140	0.080	581.438
2	0.010	0.018	0.008	47.262
3	0.004	0.002	0.002	21.681
4	0.001	0.000	0.001	24.031
5	0.001	0.000	0.001	54.505
6	0.000	0.000	0.000	8.532
7	0.001	0.000	0.001	131.276
8	0.000	0.000	0.000	0.774
9	0.000	0.000	0.000	0.446
Sum	0.996	1.000	0.172	965.655

NBRM: Predicted and actual probabilities

Count	Actual	Predicted	Diff	Pearson
0	0.919	0.921	0.002	0.046
1	0.060	0.044	0.016	77.069
2	0.010	0.015	0.005	24.515
3	0.004	0.007	0.003	16.884
4	0.001	0.004	0.003	23.678
5	0.001	0.002	0.001	10.402
6	0.000	0.002	0.001	11.431
7	0.001	0.001	0.000	0.049
8	0.000	0.001	0.001	9.823
9	0.000	0.001	0.000	5.336
Sum	0.996	0.997	0.033	179.233

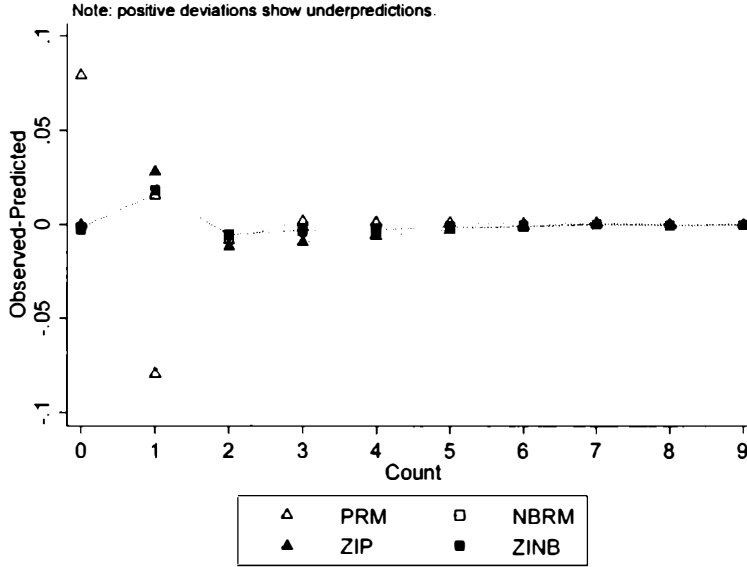
ZIP: Predicted and actual probabilities

Count	Actual	Predicted	Diff	Pearson
0	0.919	0.919	0.000	0.000
1	0.060	0.032	0.028	324.049
2	0.010	0.022	0.012	85.262
3	0.004	0.014	0.010	85.627

ZIP	BIC=-1.087e+05	AIC=	0.974	Prefer	Over	Evidence

vs ZINB	BIC=-1.116e+05	dif=	2951.459	ZINB	ZIP	Very strong
	AIC= 0.744	dif=	0.231	ZINB	ZIP	
	LRX2= 2960.917	prob=	0.000	ZINB	ZIP	p=0.000

Şekil 8.2. Saymalı Modellere Ait Artıkların Karşılaştırılması



Tablo 8.30'da yer alan çıktının birinci kısmı, modellerin her birine ait parametre tahminlerini ve uyum iyiliği istatistiklerini özetler. Tüm modellerden elde edilen parametre tahminleri birbirlerine yakın olarak elde edilmiştir. En küçük AIC ve BIC ise Tablo 8.25'teki ZINB modelinden elde edilmiştir.

Tablo 8.30'da yer alan çıktının ikinci kısmı, her bir model için gözlenen ve tahmin edilen sayımlardaki maksimum ve ortalama farkları göstermektedir. NBRM'nin bu tahminde daha iyi performans gösterdiği söylenebilir. NBRM için gözlenen ve tahmin edilen değerler arasında ortalama mutlak farkın (0.003) diğer modellere göre (PRM için 0.017, ZIP için 0.006, ZINB için 0.004) daha düşük olduğu görülmektedir.

Tablo 8.30'da yer alan çıktının üçüncü kısmında her bir modele ait sayma değerleri için tahminlenen ve gerçek olasılıklar yer almaktadır. Mutlak fark ile birlikte verilerin gerçek dağılımını ve model tarafından önerilen dağılımı karşılaştıran Pearson Chi-Square istatistiği de karar vermede önemlidir. Pearson sütununun toplamına bakmak, tahmin edilen olasılıkların gerçek olasılıklara ne kadar yakın olduğuna dair bir fikir vermektedir. NBRM için elde edilen 179.233 değeri diğer modellerden elde edilenlere (PRM için 965.655, ZIP için 608.310, ZINB için 230.580) göre daha düşüktür. Bu nedenle, negatif binom regresyon modelinin diğer modellere göre daha iyi olduğu söylenebilir.

Tablo 8.30'da yer alan çıktının dördüncü kısmında, modelleri birbiri ile karşılaştırmaya uygun testler yer almaktadır. Modeller AIC ve BIC kullanarak karşılaştırıldığında negatif binom, ZIP ve ZINB regresyon modelleri poisson regresyon modeline göre tercih edilirken; negatif binom ve ZINB regresyon modelleri ZIP regresyon modeline göre tercih edilir. ZINB regresyon modeli ise negatif binom regresyon modeline göre daha iyidir. Vuong testi dikkate alındığında da ZINB modeli tercih edilir ve test sonucuna bakıldığında istatistiksel olarak anlamlıdır. Bu nedenle, çıktının dördüncü kısmında yer alan modellere ait uyum istatistikleri Şekil 8.2'de yer alan model artıkları da gördüklerimizi destekler. Artıkların küçük olması iyi bir modelin göstergesidir ve bu nedenle sıfıra en yakın çizgileri olan modeller verilerimiz için dikkate alınmalıdır. Fakat sadece grafiğe dayanarak karar verirken dikkatli olunmalıdır. Şekil 8.2 incelendiğinde NBRM ile ZINB modelinin çok benzer bir performans sergilediği ve bu iki model arasından seçim yapılabileceği söylenebilir. Bu durumda NBRM ile ZINB modeli arasında tercih yapmak için Vuong istatistiğine bakılır. Vuong istatistiğine göre ZINB modeli NBRM'ye tercih edilmektedir.

PRM'nin en büyük başarısızlığı sıfırların sayısını 0.1 değeri kadar eksik tahminlemesidir. ZIP modeli sıfırları tahminlemede daha iyidir fakat birden üçe kadar olan sayma değerleri için zayıf tahminlere sahiptir. NBRM sıfırları çok iyi tahminler ve aynı zamanda birden üçe kadar olan sayma değerleri için çok daha iyi tahminler sağlamaktadır. ZINB modeli sıfırları olduğundan daha az, birleri daha fazla, diğer saymaları ise NBRM'ye benzer şekilde tahminlemektedir (Long, 2006: 247).

Bağımlı değişken son 12 ay içinde hastanede geçirilen gün sayısı olarak alınmıştır. Bağımlı değişkene ait varyans değerinin ortalama değerinden büyük olması PRM'nin sağlaması gereken eş yayılım varsayımının ihlal edildiğini ve aşırı yayılımın varlığını ortaya koymaktadır. Eş yayılım varsayımı sağlanamadığında PRM'ye ait parametre tahminleri etkin değildir ve standart hatalar sapmalıdır. Bu nedenle, aşırı yayılım altında tutarlı parametre tahminleri veren NBRM'ye geçilmiştir. Tahmin edilen olasılıklar ile gerçek olasılıklar arasındaki fark NBRM'de PRM'ye göre daha azdır. Gerçek ve tahminlenen olasılıklar arasındaki mutlak farkların toplamı karşılaştırıldığında NBRM'den tahminlenen olasılıklarının PRM'den tahminlenenlere göre daha iyi olduğu görülmüştür. Bağımlı değişkenin %91.87'si sıfırlardan oluşmaktadır ve gözlemler sağa çarpık dağılıma sahiptir. Bu nedenle, sıfır değerlerini dikkate alan sıfır yığılmalı sayma modelleri tahminlenmiştir. İlk olarak ZIP modeli tahminlenmiş ve aşırı yayılımın varlığı test edilmiştir. Test sonucuna göre aşırı yayılım olduğu görülmüş ve aşırı yayılıma izin veren ZINB modeli tahminlenmiştir. ZINB modeli ile NBRM arasında tercih yapmak için Vuong istatistiği kullanılmış ve test sonucunda ZINB modeli NBRM'ye tercih edilmiştir. Modellere ait uyum istatistikleri de ZINB modelinin daha başarılı olduğunu göstermiştir. ZINB modelinin Logit model kısmında yer alan değişkenin bağımlı değişken üzerinde istatistiksel olarak anlamlı bir etkisinin olmadığı görülmüş ve model dirençli olarak tekrar tahmin edilmiştir. Tahminlenen dirençli ZINB modeline bakıldığında Logit kısmında yer alan değişkenin istatistiksel olarak anlamlı olduğu sonucu elde edilmiştir. Ele alman veri setine uygun olan nihai modelin dirençli ZINB modeli olduğu sonucuna ulaşılmıştır.

STATA KOMUTLARI

```
use "C:\Users\Hp\Desktop\yataklitedavisayisi.dta", clear
describe
sum tedavisayisi doktor bki_Z bki_N bki_F kronikhastalik
tab tedavisayisi
histogram tedavisayisi if tedavisayisi<6, freq addlabel
sum tedavisayisi, detail
poisson tedavisayisi doktor kronikhastalik bki_N bki_F, nolog
margins, dydx(*)
poisson tedavisayisi doktor kronikhastalik bki_N bki_F , nolog irr
listcoef
estat gof
fitstat
predict muhat, n
quietly generate ystar = ((tedavisayisi-muhat)^2 - tedavisayisi)/muhat
regress ystar muhat, noconstant noheader
poisson tedavisayisi doktor kronikhastalik bki_N bki_F, robust nolog
nbreg tedavisayisi doktor kronikhastalik bki_N bki_F , nolog
test [lnalpha]_cons = 1
quietly poisson tedavisayisi doktor kronikhastalik bki_N bki_F , nolog
est store poisson
quietly nbreg tedavisayisi doktor kronikhastalik bki_N bki_F , nolog
est store nbreg
lrtest poisson nbreg, stats force
nbreg tedavisayisi doktor kronikhastalik bki_N bki_F , nolog irr
margins, dydx(*)
est table poisson nbreg, t varlabel varwidth(32) stats(alpha N) b(%9.3f)
countfit tedavisayisi doktor kronikhastalik bki_N bki_F , maxcount (5) prn nograph
noestimates nofit
countfit tedavisayisi doktor kronikhastalik bki_N bki_F, maxcount (5) nbreg nograph
noestimates nofit
fitstat
zip tedavisayisi doktor kronikhastalik bki_N bki_F, inflate (doktor) vuong nolog
zip tedavisayisi doktor kronikhastalik bki_N bki_F, irr inflate (doktor) vuong nolog
margins, dydx(*)
fitstat
zinb tedavisayisi doktor kronikhastalik bki_N bki_F, inflate (doktor) vuong zip nolog
zinb tedavisayisi doktor kronikhastalik bki_N bki_F, inflate (doktor) robust nolog
zinb tedavisayisi doktor kronikhastalik bki_N bki_F, inflate (doktor) irr robust
nolog
margins, dydx(*)
fitstat
countfit tedavisayisi doktor kronikhastalik bki_N bki_F, inflate (doktor) replace
```

8.5. KAYNAKLAR

- Akaike, H. (1973). Maximum Likelihood Identification of Gaussian Autoregressive Moving Average Models. *Biometrika*. 60(2): 255-265.
- Cameron, A. C. ve Windmeijer, F. A. (1996). R-Squared Measures for Count Data Regression Models with Applications to Health-Care Utilization. *Journal of Business & Economic Statistics*. 14(2): 209-220.
- Cameron, A. C. ve Trivedi, P. K. (2005). *Microeconometrics: Methods and Applications*. New York: Cambridge University Press.
- Cameron, A. C. ve Trivedi, P. K. (2009). *Microeconometrics Using Stata*. Vol. 5. College Station, TX: Stata Press
- Cameron, A. C. ve Trivedi, P. K. (2013). *Regression Analysis of Count Data* (Vol. 53). New York: Cambridge University Press. Second Edition.
- Greene, William H. (1994). Accounting for Excess Zeros and Sample Selection in Poisson and Negative Binomial Regression Models. NYU Working Paper No. EC-94-10. Available at SSRN: <https://ssrn.com/abstract=1293115>
- Greene, W. H. (2018). *Econometric Analysis*. New York: Pearson Education. Eighth Edition.
- Hilbe, J. M. (2011). *Negative Binomial Regression*. United Kingdom: Cambridge University Press.
- Ismail, N. ve Zamani, H. (2013). Estimation of Claim Count Data Using Negative Binomial, Generalized Poisson, Zero-Inflated Negative Binomial And Zero-Inflated Generalized Poisson Regression Models. *In Casualty Actuarial Society E-Forum*. 41(20): 1-28.
- Lambert, D. (1992). Zero-inflated Poisson Regression, with an Application to Defects in Manufacturing. *Technometrics*. 34(1): 1-14.
- Lawless, J. F. (1987). Negative Binomial and Mixed Poisson Regression. *The Canadian Journal of Statistics/La Revue Canadienne de Statistique*. 15(3): 209-225.
- Long, J. S. ve Freese, J. (2006). *Regression Models for Categorical Dependent Variables Using Stata*. USA: Stata press. Second Edition.
- Long, J. S. ve Freese, J. (2014). *Regression Models for Categorical Dependent Variables Using Stata*. USA: Stata press. Third Edition.
- Long, J. S. (1997). *Regression Models for Categorical and Limited Dependent Variables*. USA: SAGE Publications.
- Schwarz, G. (1978). Estimating The Dimension of A model. *The Annals of Statistics*. 6(2): 461-464.
- Shankar, V., Milton, J., ve Mannering, F. (1997). Modeling Accident Frequencies as Zero-Altered Probability Processes: An Empirical Inquiry. *Accident Analysis & Prevention*. 29(6): 829-837.
- Vuong, Q. H. (1989). Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses. *Econometrica*. 57(2): 307-333.
- Winkelmann, R. (2008). *Econometric Analysis of Count Data*. Verlag Berlin Heidelberg: Springer. Fifth Edition.
- Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data*. London: MIT Press. Second Edition.
- Yip, K. C. ve Yau, K. K. (2005). On Modeling Claim Frequency Data in General Insurance with Extra Zeros. *Insurance: Mathematics and Economics*. 36(2): 153-163.
- Zainordin, R. (2009). *Count Data Analysis Using Poisson Regression and Handling of Overdispersion*. PhD Thesis. Universiti Teknologi Malaysia.

BÖLÜM 9

SINIRLI BAĞIMLI DEĞİŞKENLİ MODELLER VE STATA UYGULAMALARI

Seda Şengül ¹

Doğrusal regresyon modelinde, örneklemdaki bütün değişkenlerin (bağımlı, bağımsız) tüm gözlemleri bilinmektedir. Bu bölümde bağımlı değişkene ait verilerin tümünün gözlenmediği (sansürlenmiş-censoring) ve bağımlı değişkenin tüm verilerinin gözlendiği, ancak anakütlenin tamamını temsil etmeyen, anakütlenin belli bir alt kümesinden sistematik olarak seçilmiş örneklemelerden (budanmış-kırpılmış-truncated) tahmin edilen regresyon modelleri üzerinde durulacaktır. Sansürlenmiş ve budanmış olan örneklemelerin ortak özelliği, sıradan en küçük kareler tahmincisiyle tahmin edilmeleri durumunda parametrelerin sapmalı (yanlı - biased) ve tutarsız olmalarıdır. Sansürlü ve budanmış verilerle tahmin edilen modeller literatürde sınırlı bağımlı değişkenli modeller veya gizli değişkenli modeller olarak adlandırılmaktadır. Bu bölümde sansürlü ve budanmış veriler tanıtıldıktan sonra sınırlı bağımlı değişkenli modellerden Tobit modeli, Heckman Seçim modeli ve Çift Eşik (Double Hurdle) modeli Stata uygulamalarıyla açıklanacaktır.

9.1. Budanmış (Truncated) ve Sansürlü (Censored) Veriler

Budama (kırpma-kesme) etkisi, örnekteki veriler sadece bir anakütlenin alt kümesinden çekildiğinde ortaya çıkmaktadır. Alt kümenin örnekleme, bağımlı değişkenin değerine bağlıdır. Örneğin yoksulların harcamalarının belirleyicileri üzerine bir çalışma yapılmak istendiği varsayalım. Bağımlı değişken yoksulluk sınırı altında kalan hanehalklarının harcaması olsun. Geliri yoksulluk sınırının altında kalan hanehalkları alt kümeye, bu alt kümeden çekilen örneklem ise budanmış veriye örnek olarak gösterilebilir.

Sansürleme, bağımlı değişkene ait verilerin bir kısmının kayıp veya sansürlendiği, fakat bağımsız değişkenlere ait verilerin tam olarak gözlendiği durumlarda ortaya çıkmaktadır. Bağımlı değişkenin gözlem değerleri soldan (alttan), sağdan (üstten) veya hem alttan hem de üstten sansürlü olabilir. Sansürlü veriye örnek olarak Tobin'in (1958) orijinal makalesindeki dayanıklı tüketim malları harcaması verilebilir. Dayanıklı mallara ya harcama yapılmıştır ya da yapılmamıştır. Bu nedenle harcama verileri ya pozitif ya da sıfır olarak kaydedilmiştir. Sıfır değerler dayanıklı mallara hiç harcama yapılmadığını göstermekte olup soldan sansürlüdür.

¹ Prof. Dr., Çukurova Üniversitesi İktisadi ve İdari Bilimler Fakültesi Ekonometri Bölümü, 01330 ADANA

Aynı şekilde, gelirle ilgili bir çalışmada yüksek gelirli bireyler örnekleme yer almasına rağmen tüm yüksek gelirli bireylerin gelirlerinin belli bir eşik değere (örneğin 10000 TL) eşit kabul edilmesi durumunda bağımlı değişken sağdan sansürlenmiş kabul edilmektedir. Budanmış verideki bilgi kaybı, sansürlenmiş verideki bilgi kaybından fazladır. Budanmış veride anakütlenin belli bir alt kümesinden örnek seçildiği için örnekleme değişmekte, sansürlü veride ise bağımsız değişkene/değişkenlere ait verilerin eksiksiz olarak örnekleme yer alması nedeniyle örnekleme değişmemektedir.

9.1.1 Budanmış ve Sansürlenmiş Dağılımlar Üzerine Bazı Temel Bilgiler

y , μ ortalama ve σ^2 varyansa sahip normal dağılım bir değişken ise y 'nin olasılık yoğunluk fonksiyonu eşitlik (9.1)'deki gibi yazılır.

$$f(y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-\mu)^2}{2\sigma^2}} \quad (9.1)$$

Burada $y \sim N(\mu, \sigma^2 I)$ olarak gösterilmektedir. Bir normal dağılım standart normal dağılımın $N(0,1)$ bir fonksiyonu olarak yazılabildiği gibi normal dağılıma sahip bir değişken, normal dağılımın doğrusal dönüşümlü bir fonksiyonu olarak da yazılabilir. Yani $y \sim N(\mu, \sigma^2 I)$ ise $(a + by) \sim N(a + b\mu, b^2 \sigma^2 I)$ 'dir. Daha yaygın bir dönüşüm $a = \frac{-\mu}{\sigma}$ ve $b = \frac{1}{\sigma}$ şeklindedir. $z = \frac{y-\mu}{\sigma}$ değişkeni standart normal değişkene dönüşmektedir ve $N(0,1)$ ile eşitlik (9.2)'de verilen olasılık yoğunluk fonksiyonuna sahiptir.

$$f(z) = \phi(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) \quad (9.2)$$

$\phi(z)$ notasyonu genellikle olasılık dağılım fonksiyonu için kullanılırken, $\Phi(z)$, birikimli dağılım fonksiyonu (cdf) için kullanılmaktadır. $y \sim N(\mu, \sigma^2 I)$ ise standart normal dağılım fonksiyonu eşitlik (9.3)'te, cdf ise (9.4)'te verilen eşitlikler şeklinde yazılabilir.

$$f(y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-\mu)^2}{2\sigma^2}} = \frac{1}{\sigma} \phi\left(\frac{y-\mu}{\sigma}\right) \quad (9.3)$$

$$\begin{aligned} P(Y \leq y) &= \Phi\left(\frac{y-\mu}{\sigma}\right) \\ P(Y > y) &= 1 - \Phi\left(\frac{y-\mu}{\sigma}\right) \end{aligned} \quad (9.4)$$

9.1.1.1 Budanmış Normal Dağılım

y gözlemlenmiş bağımlı değişken olsun. Normal regresyondan farklı olarak y , y^* gizli bağımlı değişkenin tam olarak gözlenen değeridir. Budanmış verilerin örnekleminin, büyük bir anakütlenin alt kümesinden çekilmiş gözlemlerden oluştuğundan bahsedilmiştir. Aşağıda gösterilen budama eşliğine (τ) göre $y^* \geq \tau$ için $y = y^*$ olarak gözlemlendiği varsayalım. Yani aslında y^* 'm eşik değerinden küçük değerleri için gözlemler kaybedilir. Böyle bir durumda $y|y > \tau$ koşuluyla y değişkeni budanmış normal dağılıma uymaktadır. Burada dikkat edilmesi gereken önemli nokta normal dağılım değişkeninin bir kısmının budanmış olmasıdır. Bu durumda, budanmış değişkenin dağılım alanı artık bire eşit değildir. Dağılımın kalan kısmının tekrar bire eşit olmasını sağlamak için yeniden ölçeklendirmek gerekmektedir (Greene, 2003). Böylelikle y değişkeni

$f(y)$ olasılık yoğunluk fonksiyonuna sahip sürekli bir değişken ve τ bir sabit olmak koşulu ile eşitlik 9.5'teki gibi yazılır.

$$f(y | y > \tau) = \frac{f(y)}{P(y > \tau)} \quad (9.5)$$

Ayrıca, eşitlik (9.5)'ten eşitlik (9.6)'da elde edilir.

$$P(y > \tau) = 1 - \Phi\left(\frac{\tau - \mu}{\sigma}\right) = 1 - \Phi(\alpha) \quad (9.6)$$

Burada $\alpha = \frac{\tau - \mu}{\sigma}$ 'dir ve $\Phi(\cdot)$ standart birikimli normal dağılım fonksiyonudur. Bu koşullar altında budanmış normal dağılımın yoğunluk fonksiyonu eşitlik 9.7'de verilmiştir.

$$f(y | y > \tau) = \frac{f(y)}{1 - \Phi(\alpha)} = \frac{\frac{1}{\sigma} \phi\left(\frac{y - \mu}{\sigma}\right)}{1 - \Phi\left(\frac{\tau - \mu}{\sigma}\right)} = \frac{\frac{1}{\sigma} \phi\left(\frac{y - \mu}{\sigma}\right)}{1 - \Phi(\alpha)} \quad (9.7)$$

Burada $\phi(\cdot)$ standart normal dağılım fonksiyonudur. Budanmış normal dağılım fonksiyonunun en çok olabilirlik fonksiyonu ise eşitlik 9.8'de verilmiştir.

$$L = \prod_{i=1}^N \frac{f(y)}{1 - \Phi(\alpha)} \quad (9.8)$$

Logaritması alındığında ise en çok olabilirlik fonksiyonu

$$\ln L = \sum_{i=1}^N (\ln[f(y)] - \ln[1 - \Phi(\alpha)]) \quad (9.9)$$

eşitlik (9.9)'daki gibi yazılabilir. Budanmış dağılımda *budama*, dağılımın sol kısmından yapılmış ise budanmış verinin ortalaması orijinal verinin ortalamasından büyük, sağ kısmından yapılmışsa, budanmış verinin ortalaması orijinal verinin ortalamasından küçük olacaktır. Ayrıca, budama işlemi varyansı da küçültmekte, budanmış verinin varyansı orijinal verinin varyansından daha küçük olacaktır. Budanmış normal dağılımın momentleri $y \sim N(\mu, \sigma^2 I)$ ise τ budama noktası olmak üzere (9.10) ve (9.11) eşitliklerinde verilmiştir.

$$E[y | y > \tau] = \mu + \sigma \lambda(\alpha) \quad (9.10)$$

ve

$$\text{Var}[y | y > \tau] = \sigma^2 [1 - \delta(\alpha)] \quad (9.11)$$

dir. Burada $\alpha = \frac{\tau - \mu}{\sigma}$ 'dir ve $\phi(\alpha)$ standart normal dağılımdır.

$$\delta(\alpha) = \lambda(\alpha)[\lambda(\alpha) - \alpha]$$

ve

$$\lambda(\alpha) = \frac{\phi(\alpha)}{[1 - \Phi(\alpha)]} \quad (9.12)$$

olmak üzere $\lambda(\alpha)$ ters Mills oranıdır (Inverse Mills Ratio). Ters Mills oranının fazla budamanın (λ daha büyük) bir ölçüsü olduğu da düşünülebilir.

9.1.1.2 Budanmış Regresyon Modeli

$y \sim N(x\beta, \sigma^2)$ olmak üzere budama noktasından (τ) daha büyük y değerleri için y değişkeninin dağılımı ile ilgilenilmesi durumunda eşitlik (9.13) elde edilir.

$$E[y | y > \tau] = X\beta + \sigma \left[\frac{\phi\left(\frac{\tau - x\beta}{\sigma}\right)}{1 - \Phi\left(\frac{\tau - x\beta}{\sigma}\right)} \right] \quad (9.13)$$

Sonuç olarak, koşullu ortalama β, x, σ ve τ 'nin doğrusal olmayan bir fonksiyonudur. Artık, en küçük kareler tahmincisinin böyle bir veri için neden doğru bir tahminci olmadığı daha kolay görülmektedir. Burada budanmış veriden kaynaklı sapma vardır. Ayrıca bu durum, hata terimlerinin değişen varyanslı olacağı anlamına da gelmektedir (Cameron ve Trivedi, 2005).

9.1.1.3 Sansürlü Normal Dağılım

Bir dağılım soldan sansürlendiğinde, τ 'nin altındaki tüm gözlemlerin τ_y 'de sansürleneceği anlamına gelmektedir.

$$y = \begin{cases} y^* & \text{if } y^* > \tau \\ \tau_y & \text{if } y^* \leq \tau \end{cases} \quad (9.14)$$

τ ve τ_y 'nin kullanılması sadece τ ve τ_y 'nin sıfırda belirlenmesinin genelleştirilmesidir. Eğer y değişkeni $f(y)$ sürekli olasılık yoğunluk fonksiyonuna sahip ve τ bir sabit ise $f(y)$ fonksiyonu eşitlik (9.15)'e eşittir.

$$f(y) = [f(y^*)]^{y_i} [F(\tau)]^{1-y_i} \quad (9.15)$$

Başka bir ifade ile y 'nin yoğunluk fonksiyonu ($y > \tau$ koşuluyla) y^* 'ın yoğunluk fonksiyonuyla aynıdır ve $y^* < \tau$ olasılığına ($y = \tau$ koşuluyla) eşittir. Ayrıca y_i değişkeni $y > \tau$ iken 1 değerini alan kukla değişkendir, yani gözlem sansürlenmemiştir ve $y = \tau$ iken, y_i değişkeni 0 değerini alır, yani gözlem sansürlenmiştir. Aşağıdaki eşitlikler eşitlik 9.4'te de verilmiştir.

$$P(\text{Sansürlü}) = P(y^* \leq \tau) = \Phi\left(\frac{\tau - \mu}{\sigma}\right) = 1 - \Phi\left(\frac{\mu - \tau}{\sigma}\right) \quad (9.16)$$

ve

$$P(\text{Sansürlsüz}) = 1 - \Phi\left(\frac{\tau - \mu}{\sigma}\right) = \Phi\left(\frac{\mu - \tau}{\sigma}\right) \quad (9.17)$$

Böylece, olabilirlik fonksiyonu eşitlik (9.18)'de verildiği şekilde yazılabilir:

$$L = \prod_i^N \left[\frac{1}{\sigma} \phi\left(\frac{y - \mu}{\sigma}\right) \right]^{y_i} \left[1 - \Phi\left(\frac{\mu - \tau}{\sigma}\right) \right]^{1-y_i} \quad (9.18)$$

Sansürlenmiş verinin beklenen değeri ise eşitlik (9.19)'daki gibi yazılabilir:

$$\begin{aligned} E[y] &= (P(\text{sansürlsüz}) \times E[y | y > \tau]) + (P(\text{sansürlü}) \times E[y | y = \tau_y]) \\ &= \left\{ \Phi\left(\frac{\mu - \tau}{\sigma}\right) [\mu + \sigma \lambda(\alpha)] \right\} + \Phi\left(\frac{\tau - \mu}{\sigma}\right) \tau_y \end{aligned} \quad (9.19)$$

$\tau = 0$ özel durumu için,

$$E[y] = \Phi\left(\frac{\mu}{\sigma}\right)[\mu + \sigma\lambda] \quad (9.20)$$

Burada $\lambda = \frac{\phi(\frac{\mu}{\sigma})}{\Phi(\frac{\mu}{\sigma})}$ 'dir.

Ters Mills Oranı her biri $(\frac{\mu}{\sigma})$ 'da değerlendirilen standart normal olasılık dağılım fonksiyonu ile standart normal birikimli dağılım fonksiyonu arasındaki orandır.

9.2 Tobit Modeli

Sansürlü verilerle yapılan tahminde en yaygın kullanılan model Tobit (ya da sansürlü regresyon modeli de denir) modelidir. Tobit modelinde, y^* olarak adlandırılan, nicel anlamı olan gizli bir değişken mevcuttur. Eğer y^* anakütledeki her birim için gözlenebilseydi, en küçük kareler tahmincisi kullanılabilirdi. Fakat anakütlenin bir kısmı için (alttan, üstten, hem alttan hem de üstten sansürlü olabilir) y^* gözlenemediğinden bir takım veri problemleri ortaya çıkmaktadır. Sansürlü verilere, alkol tüketimi, sigara tüketimi, evli kadınların çalışma saatleri (her kadın çalışmaz), sosyal yardım kurumlarına yapılan bağışlar (çoğu insan bağış yapmaz) vb. çok sayıda örnek verilebilir. Bu örneklerin ortak özelliği, anakütleyi oluşturan her birim için bu örneklerle ait verilerin mevcut olmamasıdır. Herkes lüks mal, alkol veya sigara tüketmez, yardım kuruluşlarına bağış yapmaz.

Tobit modelinde tüm sansürlü gözlemler köşe çözümleri (corner solution) olarak karakterize edilmektedir. Çünkü y gözlemlenebilir bir seçim veya sonuçtur. Bağımlı değişkenin aldığı sıfır değeri rastgele bir değer değildir. Gerçek sıfır değeridir ve bağımlı değişkenin ekonomik nedenlerden dolayı aldığı değerdir, bağımlı değişkenin değeri olarak her zaman sıfır değeri gözlenebilir. Örneğin, ekonomik nedenlerle bir birey kültür, eğlence harcaması yapmayabilir, konut satın almayabilir ve bu durumda sıfır gözlem kaydedilir. Genellikle y , ekonomik nedenlerden dolayı alttan sansürlenmiş bir köşe çözüm değişkeni olarak düşünülmekte ve bu değişkenin aldığı değerler bakımından, sıfır değerli gözlemlerin toplanmasından oluşan kesikli kısım ile pozitif gözlemlerden oluşan sürekli kısım nedeniyle y 'nin hibrit olduğu varsayılmaktadır.

Tobit modeli eşitlik (9.21) şeklinde ifade edilir.

$$y^* = x_i' \beta + \varepsilon_i \quad (9.21)$$

Burada $\varepsilon_i \sim N(0, \sigma^2)$ dir. y^* , gözlem değerlerinin τ eşik değerinden büyük olduğu durumlar için gözlemlendiği, diğer durumlar için sansürlendiği bir değişkendir. Gözlemlenmiş y_i değişkeni aşağıdaki şekilde tanımlanır;

$$y_i = \begin{cases} y^* & \text{eğer } y^* > \tau \\ \tau_y & \text{eğer } y^* \leq \tau \end{cases} \quad (9.22)$$

Tobit modelde, $\tau = 0$ olduğu varsayımı ile (yani gözlemler 0'da sansürlenmiş olduğu varsayımıyla) bu değişken eşitlik (9.23)'e eşit olacaktır.

$$y_i = \begin{cases} y^* & \text{eğer } y^* > 0 \\ 0 & \text{eğer } y^* \leq 0 \end{cases} \quad (9.23)$$

y_i değişkeni (9.24)'te sunulan eşitlik şeklinde de yazılabilir. Bu eşitliklerde y_i soldan(alttan) sansürlenmiştir.

$$y_i = \max(y^*, 0) \quad (9.24)$$

Eşitlik (9.25)'te y_i sağdan (üstten) sansürlenmiştir.

$$y_i = \begin{cases} y^* & \text{if } y^* < \tau \\ \tau_y & \text{if } y^* \geq \tau \end{cases} \quad (9.25)$$

Daha önce eşitlik 9.18'de verilen sansürlü normal dağılımın olabilirlik fonksiyonu eşitlik (9.26)'da tekrar (kolay takibi sağlamak için) verilmiştir.

$$L = \prod_i^N \left[\frac{1}{\sigma} \phi\left(\frac{y_i - \mu}{\sigma}\right) \right]^{y_i} [1 - \Phi\left(\frac{\mu - \tau}{\sigma}\right)]^{1-y_i} \quad (9.26)$$

Burada τ sansür noktasıdır. Geleneksel Tobit modelinde, sansür 0 noktasındadır ve ortalama $x_i\beta$ olarak parametrize edilmiştir. Bu durumda, Tobit modelinin olabilirlik fonksiyonu eşitlik (9.27)'deki, olabilirlik fonksiyonunun logaritması ise eşitlik (9.28)'deki gibi yazılır.

$$L = \prod_i^N \left[\frac{1}{\sigma} \phi\left(\frac{y_i - x_i\beta}{\sigma}\right) \right]^{y_i} [1 - \Phi\left(\frac{x_i\beta}{\sigma}\right)]^{1-y_i} \quad (9.27)$$

$$\ln L = \sum_{i=1}^N \{y_i(-\ln \sigma + \ln \phi\left(\frac{y_i - x_i\beta}{\sigma}\right)) + (1 - y_i)\ln(1 - \Phi\left(\frac{x_i\beta}{\sigma}\right))\} \quad (9.28)$$

Bir bütün olarak yukarıdaki fonksiyonun iki bölümden oluştuğu görülebilir. İlk kısım sansürlenmemiş klasik regresyon modeline karşılık gelmekte, ikinci kısım gözlemlerin sansürlendiği ilgili olasılıklara karşılık gelmektedir.

Yukarıdaki eşitliklerde (9.27 ve 9.28) yer alan x_i : Tobit modelindeki açıklayıcı değişkenleri, β : parametreleri, σ : Tobit modelinin standart sapmasını, ϕ olasılık dağılım fonksiyonunu, Φ birikimli dağılım fonksiyonunu göstermektedir.

9.2.1 Tobit Modelinde Beklenen Değerler

Sigelman ve Zeng (2000) çalışmalarında üç farklı beklenen değerden söz edilmiştir. Bunları açıklamak için sansürün 0 noktasında yapıldığı varsayılmıştır.

1) y^* 'nin beklenen değeri:

$$E[y^*] = x_i\beta \quad (9.29)$$

2) $y|y > 0$ için beklenen değer:

$$E[y | y > 0] = x_i\beta + \sigma\lambda(\alpha) \quad (9.30)$$

Burada $\lambda(\alpha) = \frac{\phi(\frac{x_i\beta}{\sigma})}{\Phi(\frac{x_i\beta}{\sigma})}$ ters Mills oranıdır.

3) y 'nin beklenen değeri:

$$E[y] = \Phi\left(\frac{x_i\beta}{\sigma}\right)[x_i\beta + \sigma\lambda(\alpha)] \quad (9.31)$$

Burada $\lambda(\alpha)$ ters Mills oranıdır.

9.2.2 Tobit Modelinde Marjinal Etkiler

Tobit modeli için üç farklı beklenen değer olduğu gibi üç farklı marjinal etkide hesaplanabilmektedir. Bu marjinal etkiler sırasıyla aşağıda verilmiştir.

1) Gizli bağımlı değişken y^* 'ın marjinal etkisi eşitlik (9.32)'den hesaplanmaktadır.

$$\frac{\partial E[y^*]}{\partial x_k} = \beta_k \quad (9.32)$$

x_k bağımsız değişkenindeki bir birimlik değişimin gizli bağımlı değişken y^* 'da meydana getirdiği değişikliği göstermektedir.

2) Sansürlsüz gözlemler için y 'nin beklenen değerinin marjinal etkisi eşitlik (9.33)'te verilmiştir.

$$\frac{\partial E[y|y>0]}{\partial x_k} = \beta_k \{1 - \lambda(\alpha) \left[\frac{x_i\beta}{\sigma} + \lambda(\alpha) \right]\} \quad (9.33)$$

Burada $\lambda(\alpha) = \frac{\phi(\frac{x_i\beta}{\sigma})}{\Phi(\frac{x_i\beta}{\sigma})}$ 'dir. Bu marjinal etki ise x_k değişkenindeki bir birimlik değişimin sansürlenmemiş bağımlı değişkende kaç birimlik bir değişim yaratacağını göstermektedir.

3) Sansürlü ve sansürlsüz bütün gözlemlerin beklenen değerinin marjinal etkisi (y^* gözlemlenmediği için y^* 'ın yerine y koyulduğunda y 'nin beklenen değerinin marjinal etkisi (sansürlü ve sansürlsüz verilerin tümü için)) eşitlik (9.34)'teki gibi olacaktır.

$$\frac{\partial E[y]}{\partial x_k} = \Phi\left(\frac{x_i\beta}{\sigma}\right)\beta_k \quad (9.34)$$

x_k bağımsız değişkenindeki bir birimlik değişimin sansürlenmiş ve sansürlenmemiş bağımlı değişken üzerindeki etkisini gösterir.

9.2.3 Sansürlü Veriler Neden En Küçük Kareler ile Tahmin Edilemez?

Genel olarak, tüm örnekleme ya da sadece sansürlenmemiş gözlemlere en küçük kareler tahmincisinin uygulanması tutarlı olmayan tahminler üretmektedir. Sansürlenmemiş verilere en küçük kareler tahmincisinin uygulanması durumunda eşitlik (9.30)'dan eşitlik (9.35) elde edilir.

$$\begin{aligned} y_i &= x_i\beta + \sigma\lambda\left(\frac{x_i\beta}{\sigma}\right) + \varepsilon_i \\ E[\varepsilon_i | x_i, y_i > 0] &= 0 \end{aligned} \quad (9.35)$$

Sansürlenmemiş verilerin kullanılması durumunda $\sigma \lambda \left(\frac{x_i \beta}{\sigma} \right)$ ifadesi ihmal edilir. Bu da açıklayıcı değişkenlerle hata teriminin ilişkili olmasına, dolayısıyla parametre tahminlerinin tutarsız olmasına neden olur.

Tüm örnekleme en küçük kareler tahmincisinin uygulanması durumunda da en küçük kareler tahmincileri tutarsız tahmin ediciler olacaklardır. Çünkü $E[y]$ 'i x, β ve σ nin doğrusal olmayan bir fonksiyonudur (bakınız eşitlik 9.34) (Cameron ve Trivedi, 2005; Greene, 2003; Maddala, 1987).

y ve x normal dağılmış ve sansür alttan (soldan) yapılmışsa en küçük kareler tahmincisinden elde edilen eğim parametresinin, sansürlü örneklemin oranı kadar doğru tahmin edilmiş olduğu söylenebilir. Uygulamada bu durum, en küçük kareler tahmincisindeki tutarsızlığın yaklaşık olarak tahmin edilmesini sağlamaktadır.

9.2.4 Tobit ve Probit Modelleri Arasındaki İlişki

Tobit ve Probit modelleri birçok bakımdan birbirine benzemektedir. Tobin'in Probit'i olarak adlandırılan (Goldberger, 1964) Tobit modeli ölçüm modeli dışında Probit modeli ile yapısal olarak aynı modeli kullanır. Yani y den türetilen y^* farklıdır. Tobit modelde, y^* 'ın değeri, $y^* > 0$ olduğunda $y^*=y$ 'dir. Probit modelde ise $y^* > 0$ olduğunda $y^*=1$ 'dir.

Tobit modeli ile Probit modeli arasında, Tobit modelinin doğru kurulması durumunda eşitlik (9.36)'da sunulan ilişki mevcuttur.

$$\beta_{\text{probit}} = \left(\frac{1}{\sigma_{\text{tobit}}} \right) \beta_{\text{tobit}} \quad (9.36)$$

$$\beta_{\text{tobit}} = \beta_{\text{probit}} \sigma_{\text{tobit}}$$

9.2.5 Tobit Modelinde Değişen Varyans

Tobit modelin ε_i hata terimi normal dağılım göstermiyor veya değişen varyans sorunu varsa, en çok olabilirlik tahmincileri tutarlı değildir. Fakat değişen varyans sorunu olsa bile uygun dönüşümlerle tutarlı tahminciler elde etmek mümkündür. Değişen varyans sorununu çözmek için, (9.37)'de sunulan eşitlikten hesaplanan standart sapmalar, (9.28)'deki eşitlikteki standart sapmalar yerine kullanılmakta ya da Stata paket programında yer alan dirençli standart hatalar hesaplanmaktadır (Tablo 9.6'da sunulan Stata uygulamasına bakınız.).

$$\sigma_i = \exp(w_i \gamma) \quad (9.37)$$

$$\ln L = \sum_{i=1}^N \{ y_i (-\ln(\exp(w_i \gamma) + \ln \phi(\frac{y_i - x_i \beta}{\exp(w_i \gamma)}))) + (1 - y_i) \ln(1 - \Phi(\frac{x_i \beta}{\exp(w_i \gamma)})) \} \quad (9.38)$$

Burada; w : x açıklayıcı değişkenleri arasından değişen varyans sorununa neden olabilecek açıklayıcı değişkenler vektörünü, γ_i : parametre vektörünü göstermektedir (Greene, 2003).

9.2.6 Tobit Modelinin Stata Uygulaması

Türkiye İstatistik Kurumu tarafından yapılan 2018 Türkiye Hanehalkı Bütçe Anketinden sağlanan veriler kullanılarak Türkiye’de sağlık harcamalarının belirleyicileri Tobit modeliyle tahmin edilmiştir.

Bu veri setinde yer alan 21019 gözlemden 8166’sı sansürlü olduğundan Tobit modeli için uygun bir veri setidir (saglik.dta). Tablo 9.1’de Tobit modelinde kullanılan değişkenler (saglik.dta) tanımlanmış, Tablo 9.2’de ise tüm gözlemler ve sansürlenmemiş (pozitif gözlem, bağımlı değişkene ait verilerin olduğu örneklem) gözlemlerin yer aldığı örneklem için değişkenlerin tanımlayıcı istatistikleri verilmiştir.

Tablo 9.1. Tobit Modelinde Kullanılan Değişkenlerin Tanımları

Variable	Değişken tanımları
logsaghar	Sağlık harcamasının logaritması
loghar	Toplam harcamanın logaritması
yas	Hanehalkı reisinin yaşı
egit1	Hanehalkı reisi hiç okul bitirmediyse 1, diğer durumda 0 değerini alan kukla değişken (referans değişkeni)
egit2	Hanehalkı reisi ilkokul mezunu ise 1, değilse 0 değerini alan kukla değişken
egit3	Hanehalkı reisi ortaokul-lise mezunu ise 1, değilse 0 değerini alan kukla değişken
egit4	Hanehalkı reisi üniversite mezunu ise 1, değilse 0 değerini alan kukla değişken
Oecd	OECD eşdeğerlik ölçęi ile düzeltilmiş hane genişliği
evli	Hanehalkı reisi evli ise 1, değilse 0 değerini alan kukla değişken
cinsiyet	Hanehalkı reisi erkek ise 1, değilse 0 değerini alan kukla değişken
kirl	Hanehalkı reisi kırdaki yaşıyorsa 1, değilse 0 değerini alan kukla değişken

Stata paket programında değişkenlerin tanımlarını listelemek için “describe” komutu kullanılır.

describe

```
\saglik.dta
  obs:      21,019
  vars:       12
  size:     945,855
21 Aug 2020 02:55
```

```
--
-- variable name  storage  display  value
-- variable name  type     format   label    variable label
```

```
--
yas             byte      %8.0g
oecd            float     %8.0g
logsaghar       float     %9.0g
kirl            float     %9.0g
loghar          float     %9.0g
```

evli	float	%9.0g
egit1	float	%9.0g
egit2	float	%9.0g
egit3	float	%9.0g
cinsiyet	float	%9.0g
egit4	float	%9.0g
saghar	float	%9.0g

Stata’da tanımlayıcı istatistikler Tablo 9.2’de ve 9.3’te verilen “sum” komutu kullanılarak elde edilir. Stata’da bütün gözlemlerin ve sansürlenmemiş gözlemlerin hesaplanan tanımlayıcı istatistikleri Tablo 9.2’de ve Tablo 9.3’te verilmiştir. Bütün gözlemlerin %38.85’i (8166 gözlem) sıfırda sansürlenmiştir. Diğer bir ifadeyle, hanelerin %38.85’i sağlık harcaması yapmamıştır. Hanehalkı reisinin ortalama yaşı bütün gözlemlerde 32.44, pozitif gözlemlerde (sansürlenmemiş) 33.13’tür. Eğitim değişkeni 4 kategoride incelenmiş ve hanehalkı reisinin hiç okul bitirmediği haneler kukla değişken tuzağına düşmemek için referans (karşılaştırma) değişken olarak seçilmiştir.

Tablo 9.2. Bütün Gözlemler İçin Değişkenlerin Tanımlayıcı İstatistikleri

sum logsaghar loghar yas oecd cinsiyet kir1 evli egit2 egit3 egit4

Variable	Obs	Mean	Std. Dev.	Min	Max
logsaghar	21,019	2.217789	2.161323	0	9.055156
loghar	21,019	7.41332	.8813916	3.178054	11.01572
yas	21,019	32.44498	4.779042	20	42
oecd	21,019	2.22374	.8521548	1	10.3
cinsiyet	21,019	.0729816	.2601125	0	1
kir1	21,019	.7933298	.4049264	0	1
evli	21,019	.9177887	.274693	0	1
egit2	21,019	.1122794	.3157174	0	1
egit3	21,019	.4805176	.4996322	0	1
egit4	21,019	.3728531	.483575	0	1

Tablo 9.3. Pozitif Gözlemler İçin Değişkenlerin Tanımlayıcı İstatistikleri

sum logsaghar loghar yas oecd cinsiyet kir1 evli egit2 egit3 egit4 if logsaghar>0

Variable	Obs	Mean	Std. Dev.	Min	Max
logsaghar	12,853	3.626835	1.59017	0.0198026	9.055156
loghar	12,853	7.528331	.8301314	3.178054	11.01572
yas	12,853	33.1381	4.592819	20	42
oecd	12,853	2.292492	.9294707	1	10.3
cinsiyet	12,853	.0667548	.2496067	0	1
kir1	12,853	.7690811	.421437	0	1
evli	12,853	.9316891	.2522885	0	1
egit2	12,853	.1066677	.3087023	0	1
egit3	12,853	.4731191	.4992963	0	1
egit4	12,853	.3852019	.4866619	0	1

Tablo 9.4. Tobit Modelinin En Çok Olabilirlik Tahmincisiyle Tahmini**tobit logsaghar loghar yas oecd cinsiyet kir1 evli egit2 egit3 egit4, ll(0)**

Tobit regression	Number of obs	=	21,019
	LR chi2(9)	=	2159.89
	Prob > chi2	=	0.0000
Log likelihood = -39406.829	Pseudo R2	=	0.0267

logsaghar	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
loghar	.9030775	.027916	32.35	0.000	.8483599	.9577951
yas	.0944108	.0055961	16.87	0.000	.083442	.1053797
oecd	.3039678	.0264645	11.49	0.000	.2520954	.3558402
cinsiyet	.1589426	.0996686	1.59	0.111	-.0364155	.3543007
kir1	-.035587	.0647008	-0.55	0.582	-.1624055	.0912315
evli	.7867269	.0958602	8.21	0.000	.5988334	.9746204
egit2	-.3648563	.1390488	-2.62	0.009	-.6374028	-.0923099
egit3	.3498609	.1282023	2.73	0.006	.0985746	.6011473
egit4	.6485804	.129373	5.01	0.000	.3949994	.9021614
_cons	-10.11973	.3193135	-31.69	0.000	-10.74561	-9.49385
/sigma	3.065585	.0211499			3.024129	3.10704
8,166 left-censored observations at logsaghar <= 0						
12,853 uncensored observations						
0 right-censored observations						

Tablo 9.4'te Tobit modelinin parametre tahminleri verilmiştir. Olabilirlik oran testi (LR=2159.89) modelin bir bütün olarak anlamlı olduğunu göstermektedir (prob>chi2=0.0000). Açıklayıcı değişkenlerin katsayılarının her birinin sıfırdan farklı olup olmadığına bakıldığında, hanehalkı reisinin kırsal alanda yaşıyor olması ve cinsiyet değişkenlerinin katsayılarının p değerlerinin sırasıyla 0.582 ve 0.111 olduğu görülmektedir. Bu p değerleri de $\alpha=0.05$ değerinden (α : anlamlılık düzeyi - 1. tip hata yapma düzeyi) büyük olduğu için H_0 hipotezi reddedilmemektedir. Bu katsayıların %5 anlamlılık düzeyinde istatistiki olarak anlamsız oldukları, sıfırdan farklı olmadıkları söylenebilir. Modeldeki diğer değişkenlerin katsayılarının (p değeri< $\alpha=0.05$ olduğu için her tekil katsayı için H_0 hipotezi reddedilmektedir.) %5 anlamlılık düzeyinde istatistiki olarak anlamlı oldukları, yani sıfırdan farklı oldukları görülmektedir (Tablo 9.4).

Tobit modelinin katsayılar da doğrusal olmamasının yanı sıra modeldeki açıklayıcı değişkenlerin katsayıları, açıklayıcı değişkenlerin gizli bağımlı değişken y^* ($E[y^*|x]$) üzerindeki etkilerini ölçmektedir. Yani en küçük kareler tahmincisi de olduğu gibi bir bağımsız değişkendeki bir birimlik değişme, gözlenen bağımlı değişkenin ortalama değeri üzerindeki marjinal etkisini vermektedir. Açıklanmak istenen değişken y 'dir. Bu nedenle Tobit modelinde marjinal etkiler hesaplanır ve yorumlanır.

Tablo 9.5. Marjinal Etkiler

```

margins, predict(e(0,.)) dydx(_all)
Average marginal effects
Model VCE      : OIM
Expression     : E(logsaghar|logsaghar>0), predict(e(0,.))
dy/dx w.r.t.  : loghar yas oecd cinsiyet kirl evli egit2 egit3 egit4

```

	Delta-method					
	dy/dx	Std. Err.	z	P> z	[95% Conf. Interval]	
loghar	.432785	.0135447	31.95	0.000	.4062378	.4593321
yas	.0452448	.0026828	16.86	0.000	.0399867	.050503
oecd	.1456716	.0126907	11.48	0.000	.1207983	.1705448
cinsiyet	.0761706	.0477657	1.59	0.111	-.0174485	.1697898
kirl	-.0170545	.0310069	-0.55	0.582	-.077827	.043718
evli	.3770259	.0459564	8.20	0.000	.2869529	.4670989
egit2	-.1748514	.0666411	-2.62	0.009	-.3054656	-.0442372
egit3	.1676651	.0614388	2.73	0.006	.0472473	.2880829
egit4	.3108215	.0620061	5.01	0.000	.1892918	.4323511

En çok olabilirlik tahmincisiyle tahmin edilen Tobit modelinde marjinal etkiler açıklayıcı değişkenlerin ortalama değeri üzerinden hesaplanır. Tablo 9.5'te Tobit modelinin marjinal etkileri verilmiştir. Diğer değişkenler sabitken, Türkiye'de hanehalkının toplam harcamasının %1 oranında artması toplam sağlık harcamasını %0.43 (sağlık harcaması ve toplam harcama logaritmik olarak alındığından) oranında artıracaktır. Hanehalkı reisinin yaşının artması hanenin sağlık harcamasını %4.52, OECD eşdeğerlik ölçeği ile düzeltilmiş hanehalkı büyüklüğündeki artış ise hanenin sağlık harcamasını %14.56 oranında arttıracaktır. Hanehalkı reisinin evli olduğu hanelerin sağlık harcaması hanehalkı reisinin evli olmadığı hanelere göre (0.377 kadar), hanehalkı reisinin erkek olduğu hanelerin sağlık harcaması hanehalkı reisinin kadın olduğu hanelere göre (0.076 kadar) daha fazladır. Hanehalkı reisinin eğitimi arttıkça hanelerin sağlık harcamasının da arttığı belirlenmiştir. Hanehalkı reisinin ilkökul mezunu olduğu hanelerin sağlık harcaması hanehalkı reisinin hiç okul okumadığı hanelerin sağlık harcamasından (-0.1748) daha az iken hanehalkı reisinin üniversite öğrenim düzeyine sahip olduğu hanelerin sağlık harcaması hanehalkı reisinin hiç okul okumadığı hanelerin sağlık harcamasından (0.3108 kadar) daha fazladır.²

Tobit modelinde hata teriminin değişken varyanslı olması durumunda en çok olabilirlik tahmincisi tutarsız olacağından Stata'da dirençli standart hatalar hesaplanmış ve Tablo 9.6'da verilmiştir. Tablo 9.4 ve Tablo 9.6'daki Stata çıktıları karşılaştırıldığında standart hatalarda önemli farklılıkların olmadığı görülmektedir. Ancak bağımsız değişkenlerle hata teriminin ilişkili olması durumunda standart hataların önemli ölçüde değişebileceği gözardı edilmemelidir.

² Chay vd. (2001).

Tablo 9.6. Tobit Modelinin Dirençli (Robust) Tahmini

tobit logsaghar loghar yas oecd cinsiyet kir1 evli egit2 egit3 egit4, ll(0) vce(robust)

Tobit regression	Number of obs	=	21,019
	F(9, 21010)	=	229.23
	Prob > F	=	0.0000
Log pseudolikelihood = -39406.829	Pseudo R2	=	0.0267

		Robust					
logsaghar	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]		
loghar	.9030775	.033432	27.01	0.000	.8375482	.9686068	
yas	.0944108	.005702	16.56	0.000	.0832345	.1055872	
oecd	.3039678	.0241022	12.61	0.000	.2567257	.3512099	
cinsiyet	.1589426	.1012667	1.57	0.117	-.0395479	.3574331	
kirl	-.035587	.0609566	-0.58	0.559	-.1550667	.0838927	
evli	.7867269	.099463	7.91	0.000	.5917717	.9816821	
egit2	-.3648563	.1416242	-2.58	0.010	-.6424506	-.0872621	
egit3	.3498609	.1301513	2.69	0.007	.0947543	.6049676	
egit4	.6485804	.1311774	4.94	0.000	.3914626	.9056982	
_cons	-10.11973	.331887	-30.49	0.000	-10.77025	-9.469205	
/sigma	3.065585	.01801			3.030284	3.100886	
8,166 left-censored observations at logsaghar <= 0							
12,853 uncensored observations							
0 right-censored observations							

Tablo 9.7’de Tobit modelinin EKK ile Tablo 9.8’de ise budanmış regresyon modelinin en çok olabilirlik tahmincisiyle tahminleri verilmiştir.

Tablo 9.7. Sansürlü Regresyon Modelinin EKK ile Tahmini

regress logsaghar loghar yas oecd cinsiyet kir1 evli egit2 egit3 egit4

Source	SS	df	MS	Number of obs	=	21,019
				F(9, 21009)	=	277.35
Model	10426.417	9	1158.49078	Prob > F	=	0.0000
Residual	87755.3692	21,009	4.17703694	R-squared	=	0.1062
				Adj R-squared	=	0.1058
Total	98181.7861	21,018	4.67131916	Root MSE	=	2.0438

logsaghar	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
loghar	.6241167	.016686	37.40	0.000	.5914109 .6568225
yas	.0497512	.0034938	14.24	0.000	.042903 .0565994
oecd	.1838428	.0169307	10.86	0.000	.1506574 .2170282
cinsiyet	.1139434	.0613165	1.86	0.063	-.0062418 .2341285
kirl	-.054696	.0412587	-1.33	0.185	-.1355662 .0261742
evli	.4375873	.0580769	7.53	0.000	.3237522 .5514224
egit2	-.2605327	.0871992	-2.99	0.003	-.4314497 -.0896156
egit3	.1585891	.0807103	1.96	0.049	.0003907 .3167875
egit4	.3770429	.0815251	4.62	0.000	.2172473 .5368384
cons	-4.986053	.1933016	-25.79	0.000	-5.364939 -4.607167

Tablo 9.8. Budanmış Regresyon Modelinin En Çok Olabilirlik Tahmincisiyle Tahmini

```
truncreg logsaghar loghar yas oecd cinsiyet kir1 evli egit2 egit3
egit4, ll(0)
```

(note: 8,166 obs. truncated)

Fitting full model:

```
Iteration 0: log likelihood = -23046.972
Iteration 1: log likelihood = -23001.577
Iteration 2: log likelihood = -23001.444
Iteration 3: log likelihood = -23001.444
```

Truncated regression

Limit: lower =	0	Number of obs =	12,853
upper =	+inf	Wald chi2(9) =	1934.70
Log likelihood =	-23001.444	Prob > chi2 =	0.0000

logsaghar	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
loghar	.8502742	.0204413	41.60	0.000	.8102099 .8903384
yas	-.0265993	.0035471	-7.50	0.000	-.0335514 -.0196472
oecd	.0454381	.0155662	2.92	0.004	.014929 .0759472
cinsiyet	.1404919	.0652439	2.15	0.031	.0126161 .2683676
kir1	-.0433436	.0387132	-1.12	0.263	-.11922 .0325328
evli	.1710668	.0648407	2.64	0.008	.0439814 .2981523
egit2	-.1500765	.0870211	-1.72	0.085	-.3206348 .0204817
egit3	-.0253736	.0792924	-0.32	0.749	-.1807838 .1300367
egit4	.1068332	.079638	1.34	0.180	-.0492544 .2629207
_cons	-2.213425	.216807	-10.21	0.000	-2.638358 -1.788491
/sigma	1.534467	.0109094	140.66	0.000	1.513085 1.555849

Tobit modelinin en çok olabilirlik ile tahmininden (Tablo 9.4) elde edilen katsayı değerlerinin EKK ile tahminden elde edilen katsayı değerlerinden daha büyük olduğu görülmektedir. Ayrıca, Tobit modelinin ve budanmış regresyon modelinin katsayı büyüklüklerinin, işaretlerinin ve standart hatalarının farklı olduğu görülmektedir. Budanmış regresyondaki katsayı değerleri daha küçüktür (Tablo 9.4, Tablo 9.8 ve Tablo 9.9).

Bu bölümde kullanılan Stata komutları aşağıda toplu olarak verilmiştir.

```
use "DOSYAKONUMU\saglik.dta", clear
```

```
describe
```

```
sum logsaghar loghar yas oecd cinsiyet kir1 evli egit2 egit3 egit4
```

```
sum logsaghar loghar yas oecd cinsiyet kir1 evli egit2 egit3 egit4 if logsaghar>0
```

```
tobit logsaghar loghar yas oecd cinsiyet kir1 evli egit2 egit3 egit4, ll(0)
```

margins, predict(e(0,.)) dydx(_all)

tobit logsaghar loghar yas oecd cinsiyet kir1 evli egit2 egit3 egit4, ll(0) vce(robust)

regress logsaghar loghar yas oecd cinsiyet kir1 evli egit2 egit3 egit4

truncreg logsaghar loghar yas oecd cinsiyet kir1 evli egit2 egit3 egit4, ll(0)

9.3. Heckman Seçim Modeli

Örneklem seçim problemi, örneklemin rassal seçilmediği (nonrandom sample), ancak gözlemlerin anakütleden sistematik olarak seçildiği durumlarda ortaya çıkar. Budama (kırpma-kesme) ve sansürleme gibi özel durumlar rassal olmayan örneklem seçiminin (nonrandom sample selection) veya rastlantısal (incidental truncation) budamanın özel durumlarıdır.

Rassal olmayan örneklem seçimine anket tasarımı örnek olarak verilebilir. Anketteki sorulara cevap verenlerin belli soruları cevaplayamaması/cevaplamaması, bağımlı ve bağımsız değişkenlerde eksik veriye neden olabilir. Diğer bir örnekte anketin, anakütlenin seçilmiş belli bir kesimine uygulanmasıdır. Yoksullarla ilgili bir araştırma için anketle veri toplandığı varsayılın. Anakütleden geliri yoksulluk sınırının üstünde olanlar çıkarılabilir, örnekleme alınmayabilir, örnekleme yoksulluk sınırının altında olanlar oluşturabilir (Wooldridge, 2010; Greene, 2003) .

Rastlantısal örnekleme açıklayıcı değişkenin verilerinin gözlenmesi bakımından bir sorun yokken, bağımlı değişkenin verileri gözlenmeyebilir. Rastlantısal örnekleme, örnekleme tasarlanan anket belirlememekte, insanların kararı, seçimi belirlemektedir. Eğer bir birey işgücü piyasasına katılmak (çalışmak) istiyorsa örnekleme yer alır, işgücü piyasasına katılmak istemiyorsa (çalışmamak) örnekleme yer almaz. Burada birey kendi kişisel tercihindan dolayı örnekleme yer almaz. Bu tür budamanın yol açtığı sapma örneklem seçim yanlılığı (sample selection bias) olarak adlandırılır³.

Rastlantısal örnekleme örneklem seçim yanlılığını düzeltmek için geliştirilen ve uluslararası literatürde yaygın olarak kullanılan model Heckman seçim modelidir (Heckman, 1976; 1979; Greene, 2003; Maddala, 1987; Wooldridge, 2003)⁴. Heckman (1976, 1979) bu tür durumlar için seçim problemini dışlanmış bir değişken problemi olarak ele alan basit bir pratik çözüm önermiştir.

Heckman seçim modelinde y_i^* ve z_i^* olarak gösterilen iki gizli değişken vardır (iki karar vardır: i) katılım kararı (to participate) ii) miktar kararı (how much-quantity)).

$$z_i^* = w_i\gamma + v_i \quad (9.39)$$

$$y_i^* = x_i\beta + \varepsilon_i \quad (9.40)$$

³ İşgücü piyasasına katılma yaşına gelmiş bir bireye teklif edilen ücret, rezervasyon ücretinin altındaysa birey ücreti kabul etmeyebilecek ve işgücü piyasasına katılmayı seçebilecektir. Burada veri toplama mekanizmasından dolayı bağımlı değişken budanmamış, bireysel bir seçim veya bireysel bir karar sonucu veri budanmıştır. Rassal olmayan örnekleme rastlantısal örnekleme arasındaki en büyük fark; ilkinde örnekleme, veri toplama mekanizmasına dayalı bir budama sonucu belirlenirken, ikincisinde örnekleme bir seçim, tercih ve karara bağlı olarak belirlenir (rezervasyon ücreti, bir işçinin belirli bir işi kabul etmek için isteyeceği en düşük ücrettir).

⁴ Heckman seçim modeli, literatürde II. Tip Tobit modeli (Amemiya, 1985), probit seçim modeli (Wooldridge, 2003) olarak da bilinir. Heckman bu çalışmasıyla Nobel ödülü almıştır.

$$z_i = \begin{cases} 1 & z_i^* > 0 \text{ ise} \\ 0 & \text{diğer durumda} \end{cases} \quad (9.41)$$

$$y_i = \begin{cases} y_i^* & z_i^* = 1 \text{ ise} \\ n.a & \text{diğer durumda} \end{cases} \quad (9.42)^5$$

Bir gösterge olan z_i 'yi yalnızca gizli değişken z_i^* pozitif olduğunda gözlemlenir. $y_i = y_i^*$ değişkeninin değeri ise sadece $z_i^* = 1$ olduğunda gözlenir (Eşitlik (9.41), (9.42)) .

İki eşitlikte aynı açıklayıcı değişkenler olabileceği gibi farklı açıklayıcı değişkenlerde (x ve w) olabilir. İki eşitlikteki hata terimlerinin iki değişkenli (bivariate) normal dağılım gösterdiği varsayılır.

$$\begin{pmatrix} v_i \\ \varepsilon_i \end{pmatrix} \sim N \left(0, \begin{bmatrix} 1 & \rho\sigma_\varepsilon \\ \rho\sigma_\varepsilon & \sigma_\varepsilon^2 \end{bmatrix} \right) \quad (9.43)$$

İlk eşitlik (karar eşitliği z_i^*) bir gözlemin örnekleme olup olmadığını, ikinci eşitlik ise y_i 'nin değerini (çıktı eşitliği $y_i^* = y_i$) belirler⁶.

$$E(y_i | x_i, z_i) = E(y_i^* | z_i = 1, x_i, z_i) = x_i\beta + \rho\sigma_\varepsilon \frac{\phi(w_i\gamma)}{\Phi(w_i\gamma)} \quad (9.44)$$

$\lambda(a) \equiv \phi(a) / \Phi(a)$ ters Mills oranıdır.

$$E(y_i | x_i, z_i) = E(y_i^* | z_i = 1, x_i, z_i) = x_i\beta + \rho\sigma_\varepsilon = x_i\beta + \rho\sigma_\varepsilon \lambda(w_i\gamma) \quad (9.45)$$

y değişkeninin beklenen değeri $z_i=1$ olduğunda y_i^* 'ın koşullu beklenen değerine eşittir (eşitlik (9.44) ve eşitlik (9.45)). Burada λ : ters Mills oranını, ρ : eşitlik (9.39) ve eşitlik (9.40)'taki hata terimleri arasındaki korelasyon katsayısını göstermektedir.

9.3.1 Örneklem Seçim Modellerinin Tahmini

Gözlemlenen y_i değişkeninin bağımlı değişken olduğu eşitlik 9.46, en küçük kareler tahmincisi ile tahmin edilirse β tahmincileri sapmalı olur. $\rho\sigma_\varepsilon \frac{\phi(w_i\gamma)}{\Phi(w_i\gamma)}$ terimi dışlandığında, bu terim hata teriminin içinde yer alacağından β 'nın sapmalı olacağı ve gerekli bir değişkenin dışlanması kaynaklı model tanımlama hatasının olacağı yukarıda açıklanmıştır.

$$y_i = x_i\beta + u_i \quad (9.46)$$

Eğer $\rho \neq 0$ ve z_i, x_i ile ilişkili ise u_i hata terimi de x_i ile ilişkilidir. Sonuçta ortaya çıkan sapmaya seçim yanlılığı (selection bias) veya örneklem seçim yanlılığı (sample selection bias) denir.

⁵ n.a: not available (mevcut değil)

⁶ Bu yapının standart Tobit modelinin ($w_i = x_i, \gamma = \beta, \sigma_v = \sigma_\varepsilon$ ve $\rho = 1$) genel bir durumu olduğu unutulmamalıdır.

9.3.1.1. Heckman Seçim Modelinin En Çok Olabilirlik Tahmini

9.39 ve 9.40'da sunulan eşitlikler (katılım (seçim) kararı ve miktar kararı (çıktı eşitliği)) normal dağılım varsayımı altında en çok olabilirlik tahmincisiyle eşanlı olarak tahmin edilir (Eşitlik 9.47). $\rho = 0$ olduğunda, log olabilirlik fonksiyonu, ayrı ayrı tahmin edilen probit ve doğrusal regresyon modeline indirgenir.

$$\begin{aligned} \text{LogL}(\beta, \gamma, \rho, \sigma | x, w) = \sum_{z_i=0} [\log 1 - \Phi(w_i \gamma)] + \sum_{z_i=1} [-\log(\sigma) + \log \left[\Phi \left(\frac{y_i - x_i \beta}{\sigma} \right) \right] + \\ \log \left[1 - \Phi \left(\frac{-w_i \gamma + \rho \left(\frac{y_i - x_i \beta}{\sigma} \right)}{\sqrt{1 - \rho^2}} \right) \right]] \end{aligned} \quad (9.47)$$

Seçim modelinin en çok olabilirlik tahmincisi, standart en çok olabilirlik tahmincisinin özellikleri olan; tutarlılık, etkinlik, asimptotik normallik özelliklerine sahiptir. Heckman seçim modelinin en çok olabilirlik tahmincisi, hata terimlerinin birlikte normal dağıldığı varsayımına dayanmaktadır. Bu çok güçlü ve çoğu zaman gerçekçi olmayan bir varsayımdır. Değişen varyans sorununun var olması durumunda, değişen varyans sorunu dirençli standart hatalar hesaplanarak ya da eşitlik (9.37)'de Tobit modeli için sunulan eşitlikten yararlanarak hesaplanan standart sapmalar eşitlik 9.47'de kullanılarak değişen varyans sorunu düzeltilmiş olur.

9.3.1.2. Heckman'ın İki Aşamalı Tahmin Süreci

Heckman (1976, 1979), örneklem seçim yanlılığını düzeltmek ve tutarlı tahminçiler elde etmek için birinci aşamasının standart bir probit modeline, ikinci aşamasının doğrusal regresyon modeline dayandığı iki aşamalı bir tahmin önermiştir. İki aşamalı bu süreç, tamamen gözlenen y'lerin koşullu ortalamalarını temel almaktadır (eşitlik 9.45) ve Heckit modeli olarak da adlandırılmaktadır. Heckman'ın iki aşamalı tahmin sürecinde;

1. Aşamada, tüm gözlem seti kullanılarak eşitlik (9.39)'da sunulan probit modeli (seçim eşitliği-karar eşitliği) en çok olabilirlik tahmincisiyle tahmin edilir. Bu aşamada seçim yanlılığını düzeltmek için λ ile gösterilen ters Mills oranı hesaplanır.

$$\begin{aligned} z_i^* &= w_i \gamma + v_i \\ z_i^* > 0 \text{ ise } z_i &= 1, \text{ aksi takdirde } 0 \end{aligned} \quad (9.48)$$

2. Aşamada, ters Mills oranının araç değişken olarak eklendiği doğrusal regresyon modeli (sonuç eşitliği) tüm gözlem setinin alt örneğinin ($z_i=1$ olan gözlemler) oluşturduğu veri setini kullanarak EKK ile tahmin edilir (Eşitlik 9.49).

$$y_i = x_i \beta + \beta_\lambda \hat{\lambda}_i + u_i \quad (9.49)$$

EKK tahmini ile $\hat{\beta}, \hat{\beta}_\lambda, \hat{\sigma}_\varepsilon$ ve dolayısıyla $\hat{\rho} = \hat{\beta}_\lambda / \hat{\sigma}_\varepsilon$ elde edilir.

$\hat{\beta}_\lambda$ istatistiki olarak sıfırdan farklı olduğunda, β tahminçileri tutarlı olacaktır. Sıfırdan farklı olmaması durumunda ters Mills oranının değişken olarak eşitlik 9.49'a eklenmemesi ise model tanımlama hatasına yol açacaktır. Heckman'ın iki aşamalı tahmincisi tutarlıdır ancak etkin değil-

dir. Bu nedenle dirençli standart hatalar hesaplanmalıdır⁷. w_i 'deki değişkenlerin hiç biri x_i 'nin dışında bırakılmadığında, yani dışlama kısıtlaması (exclusion restriction) olmadığında, güçlü bir çoklu doğrusallık sorununa ve daha büyük standart hatalara neden olur. Bu durum, denklem (9.49)'da ters Mills oranının doğrusal olmamasıyla tanımlanır. Ancak w_i ve x_i 'deki değişkenlerin aynı olmaması durumunda, dışlama kısıtı (w_i 'de x_i 'de olmayan değişkenlerin yer alması) ile çok doğrusallık daha küçük bir sorun olur. (Wooldridge 2003; 2010, Long, 1997; Maddala, 1987, Greene, 2003).

9.3.2. Heckman Seçim Modellerinin Stata Uygulamaları

TÜİK 2019 İşgücü Araştırması veri setinden seçilen 10144 bu örneklemdeki değişkenlerin tanımları, tanımlayıcı istatistikleri Tablo 9.9 ve Tablo 9.10'da verilmiştir. Türkiye'de bireylerin ücretini etkileyen değişkenleri belirlemeye yönelik bir uygulama yapılmak istenmektedir. Örneklem 15-65 (aktif nüfus) yaş aralığındaki bireylerden oluşmaktadır. Bu bireylerden bazıları kendi bireysel tercihleri nedeniyle; örneğin, çalışma saatinin veya piyasadaki ücretin uygun olmaması vb. nedenlerden dolayı çalışmak istemeyebilirler ve sıfır gözlemler kaydedilebilir. Bu sıfır gözlemlerin örneklemden çıkarılması, yukarıda anlatıldığı gibi, seçim yanlılığına neden olacağı için Heckman seçim modelleri (iki aşamalı tahmin süreci ve en çok olabilirlik tahmincisi) kullanılarak Stata'da tahmin yapılacak ve yorumlanacaktır.

Tablo 9.9. Heckman Seçim Modelinde Kullanılan Değişkenlerin Tanımları

Variable	Değişken tanımları
Ücret	Ücret (TL/ay)
Cinsiyet	Birey erkek ise 1, değilse 0 değerini alan kukla değişken
Yas	Yaş (yıl)
Yaskare	Yaşın karesi
Egitim	Eğitim değişkeni 0, 1, 2 ve 3 olmak üzere 4 değer almaktadır. Eğitim=0 birey hiç okul okumadı, 1= birey ilköğretim düzeyine sahip, 2= birey ilköğretim, ortaokul-lise öğrenim düzeyine sahip, 3=birey üniversite öğrenim düzeyine sahip ("eğitim"= 0; birey hiç okul okumadı değişkeni referans değişken olarak alınmıştır)
Medeni	Bireyin medeni durumu 1, 2 ve 3 değerlerini almaktadır. medeni=1 ise birey bekâr, medeni=2 ise evli, medeni=3 boşanmış veya dula göstermektedir. (Referans değişken olarak bekârlar (medeni =1) alınmıştır.)
Deneyim	Bireyin çalıştığı süre
logücret	Ücretin logaritması
Katılım	Birey çalışıyorsa 1, diğer durumda 0 değerini alan katılım eşitliğinin bağımlı değişkeni

Stata'da değişken tanımları describe komutu ile belirlenir.

describe

obs: 10,144

⁷ Ayrıntılı bilgi için Greene, 2003, Maddala, 1987'e bakınız.

```
vars:          13
size:        385,472
4 Oct 2020 16:19
```

variable name	storage type	display format	value label	variable label
referans_yil	int	%8.0g		
yas	byte	%8.0g		
deneyim	byte	%8.0g		
ucret	long	%8.0g		
medeni	byte	%8.0g		
egitim	byte	%8.0g		
cinsiyet	float	%9.0g		
yaskare	float	%9.0g		
logucret	float	%9.0g		
katilim	float	%9.0g		
med1	float	%9.0g		
med2	float	%9.0g		
med3	float	%9.0g		

Tablo 9.10. Heckman Seçim Modelinde Kullanılan Değişkenlerin Tanımlayıcı İstatistikleri⁸

```
sum katilim ucret logucret yas yaskare cinsiyet deneyim med1 med2 med3
```

Variable	Obs	Mean	Std. Dev.	Min	Max
katilim	10,144	.8984621	.3020545	0	1
ucret	10,144	2476.28	2077.558	1	60000
logucret	10,144	6.959242	2.411669	0	11.0021
yas	10,144	37.23393	11.18577	15	65
yaskare	10,144	1511.475	854.385	225	4225
cinsiyet	10,144	.6735016	.4689551	0	1
deneyim	10,144	5.804121	7.749732	0	44
med1	10,144	.2584779	.4378196	0	1
med2	10,144	.6934148	.4610983	0	1
med3	10,144	.0481073	.2140034	0	1

```
tab eğitim
```

egitim	Freq.	Percent	Cum.
0	407	4.01	4.01
1	2,412	23.78	27.79
2	1,928	19.01	46.80
3	5,397	53.20	100.00
Total	10,144	100.00	

```
tab medeni
```

medeni	Freq.	Percent	Cum.

⁸ Medeni durum değişkeninin bekâr, evli, boşanmış-dul kategorileri için 1,2 ve 3 değerlerini aldığı Tablo 9.9'da anlatılmıştı. Modele i.medeni şeklinde eklenmesi durumunda, ilk kategori referans alınmakta bu tür değişkenleri oluşturan değerlerin dağılımı için Tab komutu kullanmak uygundur. Bazı durumlarda araştırmacı her kategori için kendisi kukla değişken oluşturmak ve referans değişkeni kendi belirlemek isteyebilir. Bu durumda da gen med1=medeni==1; gen med2= medeni==2 şeklinde Stata'da kukla oluşturulur.

1	2,622	25.85	25.85
2	7,034	69.34	95.19
3	488	4.81	100.00
Total	10,144	100.00	

Heckman seçim modelinde eğitim ve medeni durum değişkenleri de kategorik olduğundan, bu değişkenlerin frekansları, yüzdelik ve birikimli yüzdelik değerleri tab komutuyla belirlenmiştir. İşgücü piyasasına katılım, cinsiyet, ücret, deneyim, yaş gibi sürekli değişkenlerin ortalama değerleri, standart sapmaları ve en küçük-en büyük değerleri sum komutuyla elde edilmiştir.

Tablo 9.10 incelendiğinde; 2019 yılında bireylerin %89.84'ünün çalıştığı (işgücü piyasasına katıldığı), ortalama ücretin 2476.28 TL/ay, ortalama yaşın 37.23, ortalama çalışma süresinin 5.80 yıl olduğu, bireylerin %67.35'inin erkek, %69.34'ünün evli, %53.20'sinin üniversite mezunu olduğu belirlenmiştir.

Tablo 9.11'de Heckman Seçim Modelinin İki Aşamalı Tahmini verilmiştir. Tablo 9.11'de görüldüğü gibi Stata çıktısının ilk bölümünde sonuç modelinin, ikinci bölümünde de karar (seçim modeli) modelinin parametre tahminleri yer almaktadır. Her iki eşitlikte aynı açıklayıcı değişkenler kullanılmıştır. Çocuk sayısı, çocukların yaşı vb. değişkenler işgücü piyasasına katılım kararını etkileyen, ancak ücreti etkilemeyen değişkenlerdir. Bu değişkenler veri setinde yer alsaydı seçim eşitliğine açıklayıcı değişken olarak eklenebilirdi. Sonuç eşitliğine de ücreti etkilemeyecekleri için eklenmezlerdi (Patrinos ve ark., 2020). Böylece, seçim eşitliğinde ve sonuç eşitliğinde farklı değişkenler kullanılmış ve dışlama kısıtı (exclusion restriction) sağlanmış olurdu. Sonuç eşitliğinde, seçim sapmasını düzeltmek amacıyla araç değişken olarak eklenen ters Mills oranının katsayısı lambda istatistiki olarak %5 anlamlılık düzeyinde sıfırdan farklı bulunmuştur. Bu da sıfır gözlemlerin örneklemden budanması (kırılması) durumunda örnekleme seçim yanlılığı probleminin yani, gerekli bir değişkenin model dışında bırakılmasından kaynaklı model tanımlama hatasının olacağı anlamına gelmektedir. Lambda katsayısı için sıfır hipotezinin reddedilmemesi durumu ise örnekleme seçim yanlılığı probleminin olmadığı, iki eşitliğin hata terimleri arasında ilişki olmadığı anlamına gelmektedir. Bu durumda, budanmış verilerle regresyon denklemi EKK ile tahmin edilebilir. Ancak, EKK tahmincileri etkinlik özelliklerini sağlayamayacaktır. Tablo 9.11'deki Rho katsayısı seçim ve sonuç eşitliklerindeki hata terimleri arasındaki korelasyon katsayısını göstermektedir.

Seçim eşitliğinde cinsiyet ve medeni durumla ilgili değişkenlerin katsayıları istatistiki olarak sıfırdan farksız, (H_0 hipotezi reddedilememektedir), yas, deneyim, eğitim değişkenlerinin katsayıları ise sıfırdan farklı bulunmuştur. Diğer değişkenler sabit tutulduğunda yaşın ve yaşın karesi değişkenlerinin bireyin işgücü piyasasına katılım kararı üzerindeki etkisine bakıldığında; bireyin yaşı ortalama bir yıl arttıkça işgücü piyasasına katılım kararının belli bir yaşa kadar artış göstereceği, belli bir yaştan sonra ise azalış göstereceği belirlenmiştir. Yani yaş (yaşın katsayısı pozitif, yaşın karesi negatif işaretli olduğundan) ile çalışma kararı arasında içbükey (konkav) bir ilişki vardır. Bireylerin deneyimlerindeki ortalama birer yıllık artış bireyin işgücü piyasasına katılım olasılığında artışa yol açacaktır.

Eğitim değişkeninin işgücü piyasasına katılım kararı üzerindeki etkisinin pozitif olduğu, eğitim düzeyi arttıkça işgücü piyasasına katılım olasılığının arttığı söylenebilir. Örneğin ilköğretim düzeyindeki bireylerin, ortaokul-lise öğrenim düzeyindeki bireylerin ve üniversite öğrenim düzeyindeki bireylerin hiç okula gitmeyen bireylere göre çalışma olasılıkları daha yüksektir.

Sonuç (miktar eşitliği) eşitliğine bakıldığında ise; bu eşitlikteki değişkenlerin katsayılarının hepsinin tekil olarak sıfırdan farklı olduğu görülmektedir. Bağımlı değişken ücretin logaritmasıdır. Bu durumda, diğer değişkenler sabit tutulduğunda yaş bir yıl arttığında bireylerin ortalama ücreti %8.71 oranında artmakta, ancak yaştaki artışın ortalama ücretler üzerindeki etkisi belli bir yaştan sonra azalış göstermektedir (yaşın karesinin katsayı işareti negatiftir.). Deneyimin artması (çalışılan sürenin bir yıl artması durumunda) ücretleri ortalama %2.22 oranında arttıracaktır. Eğitim arttıkça ücretlerin arttığı görülmektedir. Ortaokul- lise öğrenim düzeyindeki bireylerin ortalama ücretleri hiç okula gitmeyen bireylerin ücretlerinden daha yüksektir. Evli bireylerin ortalama ücretleri de bekâr bireylerin ücretlerinden daha yüksektir.

Tablo 9.11. Heckman Seçim Modelinin İki Aşamalı Tahmini

```
heckman logucret cinsiyet yas yaskare deneyim i.egitim i.medeni, twostep select(katılım = cinsiyet yas yaskare deneyim i.egitim i.medeni) rhosigma
```

```
Heckman selection model -- two-step estimates      Number of obs      =      10,144
(regression model with sample selection)           Censored obs       =       1,030
                                                    Uncensored obs     =       9,114
                                                    Wald chi2(9)       =      4448.18
                                                    Prob > chi2        =       0.0000
```

	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	

logucret						
cinsiyet	.170464	.0107407	15.87	0.000	.1494125	.1915154
yas	.0871662	.0033666	25.89	0.000	.0805677	.0937647
yaskare	-.0010716	.0000414	-25.88	0.000	-.0011527	-.0009904
deneyim	.0222293	.0008314	26.74	0.000	.0205997	.0238589
egitim						
1	.1145727	.0306002	3.74	0.000	.0545974	.1745479
2	.2409108	.0317947	7.58	0.000	.1785943	.3032273
3	.5787683	.0305716	18.93	0.000	.518849	.6386876
medeni						
2	.0908031	.0158348	5.73	0.000	.0597675	.1218387
3	.0556396	.0277808	2.00	0.045	.0011903	.1100889
_cons	5.40983	.0675789	80.05	0.000	5.277378	5.542283

katilim						
cinsiyet	-.030751	.0479901	-0.64	0.522	-.1248098	.0633078
yas	.0361593	.0130749	2.77	0.006	.0105331	.0617856
yaskare	-.0004985	.000161	-3.10	0.002	-.0008142	-.0001829
deneyim	.7557123	.0265116	28.50	0.000	.7037506	.807674
egitim						
1	.4842326	.0878613	5.51	0.000	.3120276	.6564376
2	.7084253	.0932319	7.60	0.000	.5256941	.8911564

3		.9320422	.0909979	10.24	0.000	.7536896	1.110395
medeni							
2		-.0482752	.0741524	-0.65	0.515	-.1936113	.0970609
3		-.1949092	.1165849	-1.67	0.095	-.4234113	.0335929
_cons		-.6462661	.2284086	-2.83	0.005	-1.093939	-.1985935
mills lambda		-.1428381	.031433	-4.54	0.000	-.2044455	-.0812306
rho		-0.30552					
sigma		.46752101					

Heckman seçim modelinin en çok olabilirlik tahmincisiyle elde edilen parametre tahminleri Tablo 9.12'de verilmiştir. Sonuçlar Heckman'ın İki Aşamalı tahmin sonuçlarıyla benzerlik göstermektedir. Lambda katsayısı ve Rho katsayısı istatistiki olarak sıfırdan farklı elde edilmiştir. Yukarıda yapılan yorumlara ek olarak seçim ve miktar kararları arasında bir ilişki olduğu da söylenebilir. Parametre yorumları benzer olduğu için yorumlar öğrengiye bırakılmıştır.

Tablo 9.12. Heckman Seçim Modelinin En Çok Olabilirlik Tahmincisiyle Tahmini

```
heckman logucret cinsiyet yas yaskare deneyim i.egitim i.medeni, select(katılım = cinsiyet yas yaskare deneyim i.egitim i.medeni)
```

```
Iteration 0: log likelihood = -8209.3129
Iteration 1: log likelihood = -8172.0709
Iteration 2: log likelihood = -8169.6708
Iteration 3: log likelihood = -8169.6565
Iteration 4: log likelihood = -8169.6565
```

Heckman selection model	Number of obs	=	10,144
(regression model with sample selection)	Censored obs	=	1,030
	Uncensored obs	=	9,114
Log likelihood = -8169.657	Wald chi2(9)	=	5359.18
	Prob > chi2	=	0.0000

	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
logucret					
cinsiyet	.1693877	.010706	15.82	0.000	.1484043 .190371
yas	.0889598	.0033306	26.71	0.000	.0824319 .0954876
yaskare	-.0010958	.0000409	-26.80	0.000	-.001176 -.0010157
deneyim	.0232476	.0007911	29.39	0.000	.0216972 .0247981
egitim					
1	.1259086	.0304927	4.13	0.000	.066144 .1856732
2	.2546115	.0316307	8.05	0.000	.1926164 .3166066
3	.5964164	.0302855	19.69	0.000	.5370578 .655775
medeni					
2	.0906224	.0157859	5.74	0.000	.0596826 .1215621
3	.0518546	.027694	1.87	0.061	-.0024247 .1061339

	_cons	5.347954	.0657728	81.31	0.000	5.219042	5.476866

katılım							
cinsiyet		-.0126995	.0467689	-0.27	0.786	-.1043649	.0789658
yas		.0415223	.0127766	3.25	0.001	.0164806	.0665641
yaskare		-.0005631	.000157	-3.59	0.000	-.0008709	-.0002553
deneyim		.565093	.0231403	24.42	0.000	.5197389	.6104471

egitim							
1		.4900365	.0856973	5.72	0.000	.3220729	.658
2		.7095543	.0909147	7.80	0.000	.5313647	.8877439
3		.9152882	.0890094	10.28	0.000	.740833	1.089743

medeni							
2		-.0629694	.0720157	-0.87	0.382	-.2041177	.0781788
3		-.2038194	.1134923	-1.80	0.073	-.4262602	.0186213
_cons		-.6848964	.2237913	-3.06	0.002	-1.123519	-.2462736

/athrho		-.1173198	.0497621	-2.36	0.018	-.2148516	-.0197879
/lnsigma		-.7653131	.0074662	-102.50	0.000	-.7799466	-.7506797

rho		-.1167845	.0490834			-.2116056	-.0197853
sigma		.4651882	.0034732			.4584305	.4720456
lambda		-.0543268	.0228881			-.0991865	-.009467

LR test of indep. eqns. (rho = 0): chi2(1) = -302.58 Prob > chi2 = 1.0000							

Tablo 9.13. Heckman Seçim Modelinin EÇO ile Dirençli Tahmini

```
heckman logucret cinsiyet yas yaskare deneyim i.egitim i.medeni, se-
lect(katılım = cinsiyet yas yaskare deneyim i.egitim i.medeni) vce(robust)
```

```
Iteration 0: log pseudolikelihood = -8218.0248
Iteration 1: log pseudolikelihood = -8171.6385
Iteration 2: log pseudolikelihood = -8169.6663
Iteration 3: log pseudolikelihood = -8169.6565
Iteration 4: log pseudolikelihood = -8169.6565
```

Heckman selection model	Number of obs	=	10,144
(regression model with sample selection)	Censored obs	=	1,030
	Uncensored obs	=	9,114
	Wald chi2(9)	=	6262.76
Log pseudolikelihood = -8169.657	Prob > chi2	=	0.0000

	Coef.	Robust Std. Err.	z	P> z	[95% Conf. Interval]	
logucret						
cinsiyet	.1693877	.0112742	15.02	0.000	.1472906	.1914847
yas	.0889598	.0038974	22.83	0.000	.081321	.0965986
yaskare	-.0010958	.000048	-22.82	0.000	-.0011899	-.0010017
deneyim	.0232476	.0007257	32.03	0.000	.0218252	.02467
eqitim						

1		.1259086	.0358551	3.51	0.000	.0556339	.1961833
2		.2546115	.0370086	6.88	0.000	.182076	.327147
3		.5964164	.0358493	16.64	0.000	.5261532	.6666797
medeni							
2		.0906224	.0147196	6.16	0.000	.0617724	.1194723
3		.0518546	.0284831	1.82	0.069	-.0039713	.1076804
_cons		5.347954	.0761103	70.27	0.000	5.198781	5.497127

katılım							
cinsiyet		-.0126995	.0462591	-0.27	0.784	-.1033656	.0779666
yas		.0415223	.0133803	3.10	0.002	.0152974	.0677473
yaskare		-.0005631	.0001609	-3.50	0.000	-.0008786	-.0002477
deneyim		.565093	.1650895	3.42	0.001	.2415236	.8886624
egitim							
1		.4900365	.0790873	6.20	0.000	.3350282	.6450447
2		.7095543	.0855734	8.29	0.000	.5418334	.8772751
3		.9152882	.0830806	11.02	0.000	.7524532	1.078123
medeni							
2		-.0629694	.0771607	-0.82	0.414	-.2142016	.0882627
3		-.2038194	.1114886	-1.83	0.068	-.422333	.0146942
_cons		-.6848964	.2268028	-3.02	0.003	-1.129422	-.240371

/athrho		-.1173198	.0331323	-3.54	0.000	-.1822579	-.0523817
/lnsigma		-.7653131	.0144379	-53.01	0.000	-.793611	-.7370153

rho		-.1167845	.0326804			-.1802662	-.0523339
sigma		.4651882	.0067164			.4522089	.4785401
lambda		-.0543268	.0153826			-.084476	-.0241775

Wald test of indep. eqns. (rho = 0): chi2(1) = 12.54 Prob > chi2 = 0.0004							

Tablo 9.12 ve Tablo 9.13'te verilen Stata çıktıları karşılaştırıldığında standart hatalarda önemli farklılıkların olmadığı gözlenmiştir.

Bu bölümde kullanılan Stata komutları aşağıda toplu olarak verilmiştir.

```
use "DOSYAKONUMU\ucret.dta", clear
describe
```

```
gen yaskare=yas^2
```

```
replace ucet=1 if ucet==0
```

```
gen logucet=log(ucet)
```

```
gen med1=medeni==1
```

```
gen med2=medeni==2
```

```
gen med3=medeni==3
gen katilim= ucret>0
```

```
sum katilim ucret logucret yas yaskare cinsiyet deneyim med1 med2 med3
```

```
tab egitim
```

```
tab medeni
```

```
heckman logucret cinsiyet yas yaskare deneyim i.egitim i.medeni, twostep select(katilim = cinsiyet yas yaskare deneyim i.egitim i.medeni) rhosigma
```

```
heckman logucret cinsiyet yas yaskare deneyim i.egitim i.medeni, select(katilim = cinsiyet yas yaskare deneyim i.egitim i.medeni)
```

```
heckman logucret cinsiyet yas yaskare deneyim i.egitim i.medeni, select(katilim = cinsiyet yas yaskare deneyim i.egitim i.medeni) vce(robust)
```

9.4 Çift Eşik (Double Hurdle) Modeli

Cragg (1971), Çift Eşik modelini Tobit modeline alternatif olarak önermiştir. Çift Eşik modeli; katılım (decision to participate- whether or not participate) ve ne kadar katılım (how much to participate) kararlarının ayrı ayrı belirlendiği iki aşamalı stokastik süreçten oluşmaktadır. Çift Eşik modelinde bağımlı değişkenin sansürlü olması, diğer bir ifadeyle bağımlı değişkenin belli değerleri için sıfır gözlemlerinin toplanma nedeni, birimlerin/bireylerin bireysel kararının, tercihlerinin bir sonucu olduğuna dayandırılmaktadır.

Örneklem neden sansürlenir? Sıfır gözlemler neden kaydedilir?

i) Verilerin toplandığı dönemde tüketiciler o malı satın almamış olabilirler.

Tüketicilerin satın alma dönemiyle verilerin toplanma dönemi denk gelmeyebilir. Tüketiciler, incelenen malı veri toplama dönemi boyunca stoklarından tüketiyor olabilir (infrequency of purchase). Örneğin beyaz eşya, otomobil, konut, cep telefonu vb. sık sık satın alınmaz.

ii) Araştırılmak istenen malın piyasadaki fiyatının yüksek olması veya gelirin düşük olması (köşe çözüm- corner solution) gibi ekonomik nedenlerden dolayı sıfır gözlemler toplanmış, kaydedilmiş olabilir. Yani sıfır gözlemlerin nedeni gerçekten ekonomik nedenler (köşe çözüm) olabilir.

iii) Sıfır gözlemler, gerçekte sıfır oldukları için isteğe bağlı olarak kaydedilmiş olabilir. O mal tüketicilerin tüketim malları sepetinde yer almayabilir (vejetaryenler için et tüketimi, bazı insanlar için de sigara, alkol tüketimi örnek olarak verilebilir-arbitrary). Örneğin insanların yaptıkları bağış miktarı ile ilgili bir çalışma yapılmak istendiği varsayalım. Çoğu insan gerçekte, bağış yapmaz. Bu durumda verilerin önemli bir bölümü sıfıra eşit olur.

Çift Eşik modelinde bireylerin tüketim kararlarını iki adımda verdikleri varsayılır⁹. İlk olarak, birey pazara katılmak isteyip istemediğini belirler. Buna katılım kararı (participation decision) denir. Daha sonra birey, kendi koşullarını göz önüne alarak en uygun tüketim miktarını (0 olabilir) belirler. Buna miktar kararı (quantity decision) denir.

Çift Eşik modelinin Tobit modeline göre iki önemli üstünlüğü vardır. Bunlardan ilki; Tobit modelinde iki kararın ayrı ayrı alındığı varsayılmakla birlikte, modele dâhil edilen değişkenlerin bu kararlar üzerindeki etkisi aynı büyüklükte ve yöndedir (tek eşitlik). Çift Eşik modelinde ise her iki karar eşitliğinde aynı değişkenler kullanılabileceği gibi farklı değişkenler de kullanılabilir. Aynı değişkenlerin kullanılması durumunda da bu değişkenlerin her iki karar üzerindeki etkisi farklı yönde ve farklı büyüklükte elde edilebilir (hem katılım eşitliği hem de miktar eşitliği için tahmin edilen parametreler vardır). Çift Eşik modelinin diğer üstünlüğü ise sıfır gözlemlerin nedeninin sadece ekonomik nedenlere (köşe çözüm) bağlı olmamasıdır.

Çift Eşik modelinde pazara katılım eşitliği Probit modeline, tüketim miktarı eşitliği Tobit modeline dayanmaktadır. Pozitif tüketim (y_i), gizli pazara katılım (latent participation = y_1^*) ve gizli tüketim miktarı (latent consumption = y_2^*) kararları pozitif olduğunda gözlenmektedir.

Çift Eşik modelinde iki karar sürecini gösteren eşitlikler aşağıda verilmiştir.

$$y_i = y_2^* \quad \text{Eğer } y_1^* > 0 \text{ ve } y_2^* > 0 \quad (9.50)$$

$$y_i = 0 \quad \text{Diğer durumlarda}$$

$$y_1^* = z_i\alpha + u_1, \quad u_1 \sim N(0,1) \quad (9.51)$$

$$y_2^* = x_i\beta + u_2, \quad u_2 \sim N(0,\sigma^2) \quad (9.52)$$

Standart normal dağılım gösteren u_1 ve normal dağılım gösteren u_2 hata terimleri Çift Eşik modelinde iki değişkenli normal dağılım göstermektedir (9.53. eşitlik). ρ : iki hata terimi arasındaki korelasyon katsayısını z_i , katılım eşitliğindeki açıklayıcı değişkenleri, x_i miktar eşitliğindeki açıklayıcı değişkenleri, α ve β katsayıları göstermektedir.

$$(u_1, u_2) \approx BVN(0, \Gamma); \Gamma = \begin{bmatrix} 1 & \rho\sigma_{12} \\ \rho\sigma_{12} & \sigma^2 \end{bmatrix} \quad (9.53)$$

Çift Eşik modelinde pazara katılım kararı ile miktar kararı birbirinden bağımsız alınabileceği gibi (tüketici, pazara katılım kararı alabilir, ancak satın alacağı, tüketileceği miktarla ilgili karar almayabilir), iki kararı birlikte, eşanlı olarak da alabilir, dolayısıyla iki karar birbirine bağlı ola-

⁹ Çift eşik modeliyle yapılan ilk çalışmalar tüketim çalışmaları olduğundan (Cragg, 1971; Jones, 1992; Yen & Jones, 1997) açıklamalar tüketim üzerinden yapılacaktır.

bilir. İlki **Bağımsız Çift Eşik** modeli, ikincisi **Bağımlı Çift Eşik** modeli olarak literatürde adlandırılmaktadır. (Blaylock ve Blisard, 1992; Gao vd, 1995; Yen, 1994; Yen ve Jones 1997).

Bağımlı Çift Eşik modelinin en çok olabilirlik fonksiyonu eşitlik (9.54)'te, olabilirlik fonksiyonunun logaritması ise eşitlik (9.55)'te verilmiştir. Bu eşitliklerde yer alan ρ katsayısı 0 olduğunda Bağımlı Çift Eşik modeli, Bağımsız Çift Eşik modeline indirgenir. Bağımsız Çift Eşik modelinde $\Phi(z_i\alpha) = 1$ olduğunda ise Bağımsız Çift Eşik modeli Tobit modeline indirgenir (Atkinson vd. 1984; Blaylock ve Blisard, 1992; Blisard ve Blaylock, 1993; Yen, 1994).

$$L = \prod_{y_i=0} \left[1 - \Phi \left(z_i\alpha, \frac{x_i\beta}{\sigma}, \rho \right) \right] \prod_{y_i>0} \left[\Phi \left(\frac{z_i\alpha + \frac{\rho}{\sigma}(y_i - x_i\beta)}{\sqrt{1-\rho^2}} \right) \right] \frac{1}{\sigma} \phi \left[\frac{(y_i - x_i\beta)}{\sigma} \right] \quad (9.54)$$

$$\log L = \sum_{y_i=0} \log \left[1 - \Phi \left(z_i\alpha, \frac{x_i\beta}{\sigma}, \rho \right) \right] + \sum_{y_i>0} \left(\log \left[\Phi \left\{ \frac{z_i\alpha + \frac{\rho}{\sigma}(y_i - x_i\beta)}{\sqrt{1-\rho^2}} \right\} \right] - \log(\sigma) + \log \left\{ \phi \left(\frac{y_i - x_i\beta}{\sigma} \right) \right\} \right) \quad (9.55)$$

9.4.1 Çift Eşik (Double Hurdle) Modelinde Değişen Varyans

Çift Eşik modelinde de Tobit ve Heckman modellerinde olduğu gibi değişen varyans sorunu varsa, en çok olabilirlik tahminçileri tutarlı olmayacaktır. Fakat uygun dönüştürme yapıldığında değişen varyans olsa bile tutarlı tahminçiler elde etmek mümkündür. Değişen varyans sorununu çözmek için, (9.37)'de sunulan eşitlikten hesaplanan standart sapmalar, eşitlik (9.55)'teki standart sapmalar yerine kullanılmaktadır (Arabmazar ve Schmitd, 1981; 1982). Ayrıca, karar eşitliklerindeki hata terimlerinin normal dağılım göstermemesi durumunda da en çok olabilirlik tahminçisi sapmalı ve tutarsızdır (Arabmazar ve Schmitd, 1981; 1982). Çift Eşik modelinde bu sorunu çözmek için Box-Cox dönüşümü ya da ters hiperbolik sinüs (inverse hyperbolic sine) dönüşümünün uygulandığı çok sayıda çalışma literatürde mevcuttur. Bu dönüşümleri kullanan çalışmalara örnek olarak Yen (1993), Jones ve Yen (1994; 1997), vb. verilebilir.

9.4.2 Çift Eşik (Double Hurdle) Modelinde Marjinal Etkiler

Çift Eşik modelinin iki aşamalı karar yapısından dolayı üç tip marjinal etki hesaplanmaktadır (Yen ve Su, 1995; Su ve Yen, 1996).

Bu marjinal etkilerden ilki, açıklayıcı değişkenin pazara katılım olasılığı $P(Y_i^* > 0)$ üzerindeki etkisini ölçmektedir. Pazara katılım eşitliğinde marjinal etki, z_j 'ye göre pazara katılım olasılığının türevinin alınmasıyla hesaplanmaktadır (9.56 eşitlik).

$$\frac{\partial P(y_i^* > 0)}{\partial z_j} = \frac{\partial \Phi(z_i\alpha)}{\partial z_j} = \phi(z_i\alpha)\alpha_j \quad (9.56)$$

Çift Eşik modelinde x_j değişkeninin tüketim olasılığı üzerindeki marjinal etkisi eşitlik (9.57)'de verilmiştir.

$$\frac{\partial P(y_i > 0)}{\partial x_j} = \Phi\left(\frac{x_i\beta}{\sigma}\right)\phi(z_i\alpha)\alpha_j + \Phi(z_i\alpha)\phi\left(\frac{x_i\beta}{\sigma}\right)\sigma^{-1} \times \left[\beta_j - \left(\frac{x_i\beta}{\sigma}\right)\frac{\partial\sigma}{\partial x_j}\right] \quad (9.57)$$

Çift Eşik modeli için tüketimin koşullu düzeydeki marjinal etkisi eşitlik (9.58)'de verilmiştir.

$$\frac{\partial E(y_i|y_i > 0)}{\partial x_j} = \beta_j + \left[\frac{\phi\left(\frac{x_i\beta}{\sigma}\right)}{\Phi\left(\frac{x_i\beta}{\sigma}\right)}\right]\frac{\partial\sigma}{\partial x_j} - \left[\frac{\phi\left(\frac{x_i\beta}{\sigma}\right)}{\Phi\left(\frac{x_i\beta}{\sigma}\right)}\right] \left[\beta_j - \left(\frac{x_i\beta}{\sigma}\right)\frac{\partial\sigma}{\partial x_j}\right] \times \left[\left(\frac{x_i\beta}{\sigma}\right) + \frac{\phi\left(\frac{x_i\beta}{\sigma}\right)}{\Phi\left(\frac{x_i\beta}{\sigma}\right)}\right] \quad (9.58)$$

Çift Eşik modelinde, tüketimin koşulsuz düzeyi için de (toplam) marjinal etki hesaplanmaktadır. Bu marjinal etki, açıklayıcı değişkenin tüketim üzerindeki toplam etkisini ölçmektedir. Tüketimin koşulsuz düzeyi için marjinal etki, tüketim olasılığı ve tüketimin koşullu marjinal etkilerinin toplamıdır (Yen ve Su, 1995).

9.5. Çift Eşik Modelinin Stata Uygulaması

Bu bölümde kullanılan veriler daha önce, Tobit Modelinin Stata Uygulamasında kullanılan verilerin aynısı olduğundan; verilerin özelliği, değişkenlerin tanımları, tanımlayıcı istatistikleri tekrar verilmeyecektir (bakınız Tablo 9.1, Tablo 9.2 ve Tablo 9.3).

Çift Eşik modelinin en çok olabilirlik tahmincisiyle tahmin edilen parametre tahminleri Tablo 9.14'te verilmiştir. Çift Eşik modelinin Stata çıktısındaki ilk bölümde miktar eşitliğinin parametre tahminleri, ikinci bölümde katılım kararı eşitliğinin parametre tahminleri yer almaktadır. Üçüncü bölümde yer alan sigma tüketim kararı eşitliğinin standart sapmasını göstermektedir (katılım eşitliğinde $var(u_1)=1$). Ortak varyans ise iki eşitlikteki hata terimleri arasındaki ortak varyanstır. Korelasyon katsayısı yerine kullanılabilir. Tablo 9.14'te ortak varyansın istatistiki olarak anlamsız olduğu belirlenmiştir. Yani, bu örnekte haneler sağlık harcaması için pazara katılım kararı (sağlık harcaması yapmaya karar verme) ve ne kadar sağlık harcaması yapacakları kararını bağımsız olarak almaktadır. Çift Eşik modelinde parametreler doğrusal olmadığı için doğrudan yorumlanamamaktadır.

Tablo 9.14. Çift Eşik Modelinin En Çok Olabilirlik ile Tahmini

dblhurdle logsaghar loghar yas oecd cinsiyet evli egit2 egit3 egit4 kir1, peq(loghar yas oecd cinsiyet evli egit2 egit3 egit4 kir1) ll(0)

Iteration 0:	log likelihood = -42195.028	(not concave)
Iteration 1:	log likelihood = -42099.63	(not concave)
Iteration 2:	log likelihood = -39565.813	(not concave)
Iteration 3:	log likelihood = -39348.602	(not concave)
Iteration 4:	log likelihood = -38777.326	(not concave)


```

Iteration 5:  log likelihood = -37919.202   (not concave)
Iteration 6:  log likelihood = -37598.877
Iteration 7:  log likelihood = -36776.225
Iteration 8:  log likelihood = -36406.71
Iteration 9:  log likelihood = -36394.42
Iteration 10: log likelihood = -36394.305
Iteration 11: log likelihood = -36394.305

```

Double-Hurdle regression				Number of obs	=	21,019
logsaghar	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
logsaghar						
loghar	.7813299	.0195582	39.95	0.000	.7429966	.8196632
yas	-.0230254	.0045358	-5.08	0.000	-.0319155	-.0141354
oecd	.0415569	.017004	2.44	0.015	.0082298	.0748841
cinsiyet	.1294713	.0649247	1.99	0.046	.0022212	.2567214
evli	.1629544	.0663272	2.46	0.014	.0329554	.2929534
egit2	-.1575865	.0871193	-1.81	0.070	-.3283371	.0131641
egit3	-.039679	.0799696	-0.50	0.620	-.1964165	.1170585
egit4	.1111643	.0812142	1.37	0.171	-.0480126	.2703412
kir1	-.0612699	.0384001	-1.60	0.111	-.1365327	.013993
_cons	-1.791924	.3294662	-5.44	0.000	-2.437666	-1.146182
peq						
loghar	.1357776	.0117791	11.53	0.000	.1126909	.1588642
yas	.0493983	.0023265	21.23	0.000	.0448384	.0539581
oecd	.1315643	.0117665	11.18	0.000	.1085024	.1546262
cinsiyet	.0296585	.0400255	0.74	0.459	-.0487901	.1081071
evli	.2548558	.0376699	6.77	0.000	.1810242	.3286874
egit2	-.1022124	.0567842	-1.80	0.072	-.2135075	.0090826
egit3	.1945416	.0527149	3.69	0.000	.0912222	.2978609
egit4	.2820678	.0532355	5.30	0.000	.1777282	.3864074
kir1	-.0061916	.0278282	-0.22	0.824	-.0607338	.0483506
_cons	-2.991965	.1319639	-22.67	0.000	-3.250609	-2.73332
/sigma	1.534507	.0110353			1.512878	1.556135
/covariance	.0438057	.1068519	0.41	0.682	-.1656203	.2532316

Pazara katılım için hesaplanan marjinal etki, açıklayıcı değişkenin pazara katılım olasılığı üzerindeki etkisini ölçerken, tüketim olasılığı için hesaplanan marjinal etki, açıklayıcı değişkenin hâlihazırda (pozitif tüketim) yapılan tüketim üzerindeki etkisini ölçmektedir. Tüketicinin koşullu düzeydeki marjinal etkisi ise bireyin sağlık harcaması yapıp yapmamasına bakılmaksızın tüm bireylerin sağlık harcaması yapma kararı vermeleri durumunda açıklayıcı değişkenin sağlık harcaması üzerindeki etkisinin ölçmektedir (Yen ve Su, 1995; Şengül, 2002).

Tablo 9.15. Pazara Katılım Olasılığı İçin Hesaplanan Marjinal Etki

```
margins, dydx( loghar yas oecd cinsiyet evli egit2 egit3 egit4 kir1 ) predict(ppar)
```

```

Average marginal effects      Number of obs      =      21,019
Model VCE      : OIM

```

```

Expression      : predict(ppar)
dy/dx w.r.t.    : loghar yas oecd cinsiyet evli egit2 egit3 egit4 kir1

```


Tablo 9.16 ve Tablo 9.17'deki marjinal etkilerin toplamının koşulsuz marjinal etkiyi verdiğinden yukarıda bahsedilmişti. Bu durumda, loghar değişkeninin katsayı toplamı olan 1.358 ($0.6463+0.7117=1.358$) değeri (Toplama harcama ve sağlık harcaması değişkenleri logaritmik olarak modellerde yer aldıklarından katsayı değerleri esnekliğe karşılık gelmektedir) sağlığın lüks mal grubunda yer aldığını göstermektedir.

Bu bölümde kullanılan Stata komutları aşağıda toplu olarak yer almaktadır.

use "DOSYAKONUMU\saglik.dta", clear
describe

```
dblhurdle logsaghar loghar yas oecd cinsiyet evli egit2 egit3 egit4 kir1, ll(0) peq(loghar yas oecd
cinsiyet evli egit2 egit3 egit4 kir1 )
predict p, ppar
```

```
predict d, ycond
```

```
predict s, yexpected
```

```
margins, dydx( loghar yas oecd cinsiyet evli egit2 egit3 egit4 kir1 ) predict(ppar)
```

```
margins, dydx( loghar yas oecd cinsiyet evli egit2 egit3 egit4 kir1 ) pre-
dict(yexpected)
```

```
margins, dydx( loghar yas oecd cinsiyet evli egit2 egit3 egit4 kir1 ) predict(ycond)
```

Kaynaklar

- Amemiya, T. (1985). *Advanced Econometrics*. Harvard University Press. chapter 10.
- Atkinson A., J. Gomulka, and N. Stern, (1984), Household expenditure on tobacco 1970-1980: evidence from the family expenditure survey. ESRC programme on taxation, incentives, and the distribution of income. London School of Economics. Discussion paper No.60.
- Arabmazar, A., & Schmidt, P. (1981). Further evidence on the robustness of the Tobit estimator to heteroskedasticity. *Journal of Econometrics*, 17(2), 253-258.
- Arabmazar, Abbas ve Peter Schmidt 1982 "An investigation of the robustness of the Tobit estimator to non-normality." *Econometrica* 50:1055-68.
- Blaylock, J. R., & Blisard, W. N. (1992). US cigarette consumption: The case of low-income women. *American Journal of Agricultural Economics*, 74(3), 698-705.
- Blisard, N., & Blaylock, J. (1993). Distinguishing between market participation and infrequency of purchase models of butter demand. *American Journal of Agricultural Economics*, 75(2), 314-320.
- Cameron, A. C., & Trivedi, P. K. (2005). *Microeconometrics: methods and applications*. Cambridge university press.
- Chay, Kenneth Y. & James L. Powell (2001). Semiparametric Censored Regression Models. *Journal of Economic Perspectives*, 15(4), 29-42.

- Cragg, J. G. (1971). Some statistical models for limited dependent variables with application to the demand for durable goods. *Econometrica: Journal of the Econometric Society*, 829-844.
- Gao, X., Wailes, E. J., & Cramer, G. L. (1995). Double-hurdle model with bivariate normal errors: an application to US rice demand. *Journal of Agricultural and applied economics*, 27(1379-2016-113242), 363-376.
- Goldberger, A. S. (1964). *Econometric theory*. New York: Wiley.
- Greene, W. H. (2003). *Econometric Analysis*, Prentice Hall. New Jersey, 16.
- Heckman, James J. (1976). A Life-Cycle Model of Earnings, Learning, and Consumption."J.P.E. 84, no. 4, pt. 2 :S11-S44
- Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica: Journal of the econometric society*, 153-161.
- Jones, A. M. (1992). A note on computation of the double-hurdle model with dependence with an application to tobacco expenditure. *Bulletin of economic Research*, 44(1), 67-74.
- Long, J. S. (1997). *Regression Models for Categorical and Limited Dependent Variables*. London: Sage Publications.
- Maddala, G. S. (1987). Limited Dependent and Qualitative Variables in Econometrics. *Econometric Society Monographs* No.3.
- Sigelman, L., & Zeng, L. (2000). Analyzing censored and sample-selected data with Tobit and Heckit models. *Political analysis*, 167-182.
- Su, S. J. B., & Yen, S. T. (1996). Microeconomic models of infrequently purchased goods: an application to household pork consumption. *Empirical economics*, 21(4), 513-533.
- Şengül, S. (2002). Zeytinyağı Talebinin Çift Eşik Modeli İle Tahmini. *Hacettepe Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 20(2), 93-114.
- Tobin, J. (1958). Estimation of relationships for limited dependent variables. *Econometrica: journal of the Econometric Society*, 24-36.
- Wooldridge, J. M. (2003). *Introductory Econometrics – A Modern Approach*, 2nd ed., 2003 (South-Western: Mason, OH).
- Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. Cambridge: MIT press.
- Yen, S. T. (1994). Cross-section estimation of US demand for alcoholic beverage. *Applied Economics*, 26(4), 381-392.
- Jones, A. M. & Yen, S. T. (1994) A Box-Cox double-hurdle model. Institute for Fiscal Studies, Working Paper, W94/6, London
- Yen, S. & Jones, A. (1997) Household consumption of cheese: an inverse hyperbolic sine double-hurdle model with dependent errors, *American Journal of Agricultural Economics*, 79, 246–51.
- Yen, S. T., & Su, S. J. (1995). Modeling US butter consumption with zero observations. *Agricultural and Resource Economics Review*, 24(1203-2016-94988), 47-55.

BÖLÜM 10

SWITCHING REGRESYON MODELİ VE STATA UYGULAMASI

Seda Şengül¹

Salih Çam²

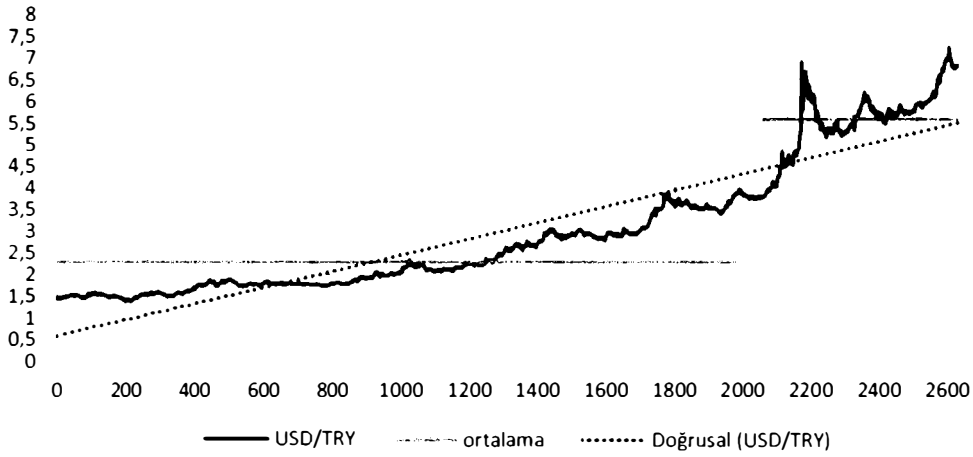
10.1. Rejim Kavramı

Ekonometrik modellerde genel olarak model katsayılarının zaman veya birime göre değişmediği, sabit kaldığı varsayılmaktadır. Fakat gerçek hayatta model katsayıları değişkenlik gösterebilmektedir. Dolayısıyla uygulamada katsayıların değişmesine izin veren modeller kullanmak daha uygun olmaktadır (Deutsch vd., 1994). Uygulanan makroekonomik politikalardaki değişiklikler, savaş ve doğal afetler, ekonomideki kriz dönemleri, kur şokları gibi birçok olgu değişkenler arasındaki ilişkilerin zamanla farklılık göstermesine ve model katsayılarının değişmesine neden olmaktadır. Bir modeldeki katsayılar üç şekilde değişebilir; Modelin sabit katsayısı değişebilir, eğim katsayıları değişebilir veya hem sabit hem de eğim katsayıları değişebilir. Switching regresyon modellerinde yukarıda bahsi geçen her durumun (savaş, politika değişikliği vs.) öncesi ve sonrası bir rejim olarak isimlendirilmektedir (Hamilton, 1994). Dolayısıyla switching regresyon modellerinin arkasında yatan mantık her bir rejimde model katsayılarının farklı olmasıdır. Peki, rejimden ne anlaşılmalıdır ve bir modelde rejimler nasıl belirlenmelidir? Zaman serisi verilerinde rejim genellikle belli bir nokta öncesi ve sonrası (savaş, kriz, politika değişikliği, salgın hastalık, yeni bir kanun, bir zirai ilacın kullanımı, ekonomik daralma ve genişleme dönemi vs.) olarak tanımlanmaktadır. Örneğin, aşağıda 1 Ocak 2010 ile 16 Haziran 2020 tarihleri arasında günlük Dolar döviz kuru (USD/TRY) kapanış fiyatları grafiği verilmiştir.

¹Prof.Dr. Çukurova Üniversitesi İktisadi ve İdari Bilimler Fakültesi Ekonometri Bölümü, ssengul@cu.edu.tr

² Arş.Gör. Çukurova Üniversitesi İktisadi ve İdari Bilimler Fakültesi Ekonometri Bölümü, scam@cu.edu.tr

Şekil 10.1 Günlük Dolar kuru kapanış fiyatları



Şekilde döviz kurunun 2018 döviz şoku öncesi ve sonrası dönemleri için düzey ortalamalarının oldukça farklı olduğu görülmektedir. Ocak 2010 ile Ocak 2018 tarihleri arasında USD/TRY döviz kurunun ortalaması 2.30 TL civarındayken, Ocak 2018 ile Haziran 2020 arasındaki ortalama yaklaşık olarak 5.58 TL civarındadır. Kırmızı yatay çizgiler ilgili dönemlerin ortalamasını göstermektedir. Dönem ortalamaları arasındaki fark şekilden rahatlıkla görülmektedir. Ocak 2018 dönemini t_0 olarak kabul edersek, tüm örneklem döneminin bir model olarak tahmin edilmesinden ziyade, t_0 öncesi dönemin bir rejim, t_0 sonrası dönemin ise bir rejim olarak modellenmesi daha doğru olacaktır.

Zaman serilerinde rejim ve rejim değişikliği yukarıdaki gibi tanımlanabilir. Yatay kesit verilerde, dolayısıyla mikroekonometrik çalışmalarda ise rejimler genel olarak bir duruma katılım, bir gruba üyelik vs. olarak tanımlanmaktadır. Örneğin, yeni geliştirilmiş bir zirai ilacın ürün verimliliğine etkisi araştırılırken ilacı kullananlar bir rejimi, kullanmayanlar diğer rejimi oluşturmaktadır (Alene ve Manyong, 2007). Burada rejim belirlendikten sonra ilacı kullananlarda ve ilacı kullanmayanlarda ürün verimliliğinin ne olduğu karşılaştırılmaktadır (yani, rejimler ayrı ayrı karşılaştırılmaktadır). Obezite ile ilgili bir çalışmada sigara içmenin obezite üzerindeki etkisini araştırırken sigara kullanılması bir rejim, kullanılmaması diğer rejimi oluşturmaktadır. Diğer aşama ise sigara içenlerde ve sigara içmeyenlerde obeziteyi hangi değişkenler nasıl etkilemektedir sorusunun cevabı araştırılmaktadır. Benzer şekilde, bilinçli tüketimin obezite üzerindeki etkisini incelerken ürünün ambalajının arkasındaki besin değerlerine göre tüketim yapan bireyler bir rejimde, ambalajın arkasındaki besin değerlerine dikkat etmeden tüketim yapanlar diğer bir rejimde yer almaktadır. Böyle bir modelde karar denkleminin bağımlı değişkeni, bireyler gıdaları besin değerlerine bakarak tüketiyorsa 1, besin değerlerine bakmadan tüketiyorsa 0 değerlerini alan kukla değişkenidir. Rejim1'in bağımlı değişkeni besin değerlerini dikkate alan bireylere ait gözlemleri içerirken, rejim2'nin bağımlı değişkeni besin değerlerine bakmadan tüketim yapan bireylere ait gözlemleri içermektedir (Loureiro vd. 2012). Diğer bir ifadeyle, rejim1 ambalajın arkasındaki besin değerleri bilgisini dikkate alanların gıda tüketimi, rejim 2'de ambalajın arkasındaki besin değerlerini dikkate almayanların gıda tüketimini etkileyen değişkenleri belirleyerek karşılaştırma olanağı verdiği gibi besin değerini bilerek tüketim yapanların özelliklerinin de belirlen-

mesine olanak tanır. Yani zaman serisi modellerinde rejim belirlenirken genelde verinin yapısı göz önüne alınırken, yatay kesit verilerde ise rejimler araştırmacının incelemek istediği olguya bağlı olarak seçilmektedir.

Rejimler belirlendikten sonra modelin tahmini yapılmalıdır. Switching regresyon modellendikten sonra model tahmini yapılırken hata terimlerinin birbiri ile korelasyonuna bağlı olarak içsel (endogenous) switching regresyon modeli veya dışsal (exogenous) switching regresyon modeli tahmin edilmektedir.

10.2. Switching Regresyon Tahmin Problemi

Switching regresyon modelinde rejimlere ait regresyonların hata terimlerinin birbiri ile ilişkisine bağlı olarak model içsel veya dışsal olarak tahmin edilmektedir. İki rejimli bir switching regresyon modelinde σ_{u1u} ve σ_{u2u} sırasıyla rejim bir ile karar denkleminin hata terimleri arasındaki kovaryans ve rejim iki ile karar denkleminin hata terimleri arasındaki kovaryans olduğu varsayıl-sın. Eğer $\sigma_{u1u} = \sigma_{u2u} = 0$ ise model dışsal (exogenous) switching regresyon, σ_{u1u} ve σ_{u2u} sıfırdan farklı ise model içsel (endogenous) switching regresyon modeli olarak adlandırılmaktadır (Maddala ve Nelson, 1975). Burada içsellik kavramının açıklanması gerekmektedir. Örneğin, yeni üretilmiş bir tarımsal teknolojinin ürün verimliliğine etkisi incelenmek istensin. Örneklem-deki bir grup çiftçi yeni üretilen teknolojiyi kullanırken, bir grup çiftçi bu teknolojiyi kullanma-maktadır. Çiftçiler bu teknolojiyi kendi istekleriyle kullanmış olabilir veya tarım politikası gereği belli bir kesim çiftçiye bu teknoloji kullanılmak istenmiş olabilir. Burada yeni teknolojiyi kul-lananlar birinci rejimi oluştururken, bu teknolojiyi kullanmayanlar ikinci rejimi oluşturmakta-dır. Model iki rejimli tahmin edilip sonuçlar değerlendirilirse önemli bir problem gözden kaçmış olacaktır. Çünkü yeni teknolojiyi kullananlar arasında teknolojiye uyum açısından bir içsellik sorunu olabilir. Zira teknolojiye uyum aynı zamanda çiftçilerin eğitimleriyle de yakından ilgili-dir. Yani teknolojinin kullanımı ve adaptasyonu eğitimin bir fonksiyonudur. Dolayısıyla yeni teknolojiyi kullanmaktan bağımsız olarak, teknolojiye uyum sağlama yeteneği daha yüksek olan (daha eğitilmiş) çiftçilerin diğerlerine göre daha üretken olacağı aşikârdır ve eğitim, içsel olarak yeni teknolojinin etkisini olduğundan çok daha yüksek gösterecektir (Alene ve Manyong, 2007). Başka bir ifade ile teknoloji kullanımının verimlilik üzerindeki etkisi tekil olmayacak aynı za-manda eğitilmiş çiftçilerin ürün verimliliğine etkisini de içerecektir. Bu ise model katsayılarının sapmalı olmasına neden olacaktır. Dolayısıyla switching regresyon modeli tahmin edilirken iç-sellik problemine dikkat edilmelidir (Hausman, 1983).

10.2.1. Dışsal (Exogenous) Switching Regresyon Modeli

Dışsal switching regresyon modellerinde karar regresyonunun hata terimleri ile rejim regresyonunun hata terimleri arasındaki kovaryans sıfırdır. Başka bir ifade ile bu modellerde hem karar regresyo-nundaki (0,1) ikili değişkeni hem de rejim regresyonundaki bağımlı değişkenleri aynı anda etkileyen herhangi bir değişken yoktur. Böyle bir durumda rejim olarak seçilen olgunun incelenen konu üze-rinde hiçbir etkisi yok demektir. Yani rejim anlamsızdır ve rejimler arasında bir fark yoktur. Fakat gerçek hayatta bu duruma çok az rastlanmaktadır. Modellerin çoğunda rejimler anlamlıdır ve EKK ile elde edilecek parametreler sapmalı ve tutarsız sonuçlar vermektedir. Bu noktada içsel switching regresyon modellerini kullanmak tahmincilerin tutarlı ve sapmasız olması sağlayacaktır.

10.2.2. İçsel (Endogenous) Switching Regresyon Modeli

Switching regresyon modelleri tahmin edilirken modelin içsel bir model veya dışsal bir model olması tahmin problemi açısından önemlidir. Yatay kesit verilerinde genellikle içsel switching regresyon modeli tahmin edilmektedir. Çünkü karar regresyonundaki bağımlı değişkeni etkileyen bağımsız değişkenler aynı zamanda rejim regresyonlarındaki bağımlı değişkenleri de etkileyebilir. Yani karar regresyonunda ve her iki rejimde aynı açıklayıcı değişkenler yer alabilir. Başka bir ifade ile içsel switching regresyon modellerindeki karar regresyonundaki hata terimi ile rejim regresyonlarındaki hata terimleri arasındaki kovaryans sıfırdan farklı olabilir (Maddala ve Nelson, 1975). İçsel bir switching regresyon modeli ve bu modelin tahmini aşağıdaki şekilde ifade edilebilir:

$$Y^* = Z\beta + u \quad (\text{Karar denklemi}) \quad (10.1)$$

$$y_1 = X\beta_1 + u_1 \quad (1. \text{ Rejim}) \quad (10.2)$$

$$y_2 = X\beta_2 + u_2 \quad (2. \text{ Rejim}) \quad (10.3)$$

Burada $I_i = \begin{cases} 1 & Y \geq \gamma \\ 0 & Y < \gamma \end{cases}$ olarak tanımlanmıştır. $Y = [y_1 \ y_2]'$ burada rejim1 ve rejim2'deki tüm gözlemleri içeren sürekli değişkendir. Y^* gözlenmeyen bir değişken olmak üzere, I_i aşağıdaki koşullara bağlı olarak rejim1 veya rejim2'yi temsil etmekte ve sistemin hangi durumda hangi rejimde olacağını göstermektedir. Y değişkeni uygulamada bireylerin BKİ değerlerini ifade etmektedir. Ayrıca γ burada Y değerlerinin hangi rejimlerden geldiğini temsil eden eşik veya sınır değeridir. Y^* ise Y değişkeninin aldığı değerlere (γ 'dan büyük veya küçük olmasına bağlı olarak) bağlı olarak 1 veya 0 değeri alan kukla değişkendir. Eşitlik (10.1) karar denklemidir ve bağımlı değişkeni (0,1) iki değerli olduğu için logit model veya probit model kullanılarak tahmin edilebilir. (10.2) ve (10.3) denklemleri sırasıyla rejim 1 ve rejim 2'yi ifade etmektedir. Denklem (10.2) ve (10.3)'teki X matrisi aynı veya farklı dışsal değişkenlerden oluşan açıklayıcı değişkenler matrisidir. $Cov(u, u_1) = Cov(u, u_2) = 0$ olduğu (dışsal switching regresyon modeli) durumda u, u_1 ve u_2 hata terimlerinin beklenen değeri sıfır olacaktır. Bu durumda EKK ile tahmin edilen regresyon katsayıları sapmasız ve tutarlı olacaktır. Fakat hata terimleri arasındaki kovaryanslar sıfırdan farklı ise EKK tahminçileri sapmalı ve tutarsız olacağından içsel switching regresyon modeli kullanılmalıdır.

$I_i = 1$ eğer $Y \geq \gamma$ (rejim1) ve $I_i = 0$ eğer $Y < \gamma$ (rejim2) iki rejimi ifade eden bir değişken olsun. O hâlde

$I_i = 1$ iken $y_1 = X\beta_1 + u_1$ regresyonu **rejim 1**

$I_i = 0$ iken $y_2 = X\beta_2 + u_2$ regresyonu **rejim 2**'yi ifade edecektir.

İçsel switching regresyon modellerini tahmin etmek için kullanılacak en etkin yöntem En Çok Olabilirlik yöntemidir (Adela ve Aurbacher, 2018).

Bu bölümde yapılacak uygulama temelinde switching regresyon denkleminin yapısı aşağıdaki şekilde açıklanabilir. Y^* değişkeni karar denkleminde bireyler spor yapıyorsa 1, yapmıyorsa 0 değerini alan bağımlı değişkendir. Rejim 1'deki bağımlı değişken spor yapan bireylerin BKİ (Beden Kitle Endeksi) iken rejim 2'deki bağımlı değişken spor yapmayan bireylerin BKİ değerleridir. Modelin mantığı; sporun BKİ değerleri üzerindeki etkisini ortaya koyabilmektir. Bireyler

spor yapmaya ya da spor yapmamaya karar verebilir. Bireylerin spor yapma kararı almasında etkili olan değişkenlerin (yaş, cinsiyet, sağlık durumu, eğitim, gelir vs.) birçoğu aynı zamanda bireylerin BKİ'lerini de etkilemektedir. Yani bazı değişkenler hem karar denklemindeki bağımlı değişkeni hem de rejimlerdeki bağımlı değişkeni aynı anda etkileyecektir. Dolayısıyla bir içsellik problemi ortaya çıkacaktır. Tahmin yapılırken içsellik problemi göz ardı edilir, örneğin model tahmini için EKK kullanılırsa artık karar denklemi hata terimi u veriyken, $E(u_1)$ ve $E(u_2)$ 'nin beklenen değerleri 0 olmayacaktır ve $Cov(u, u_1) \neq Cov(u, u_2) \neq 0$ olacaktır (Abdulai ve Huffman, 2014). Dolayısıyla u_1 ve u_2 hata terimlerinin koşullu beklenen değerleri,

$$E(u_1|I = 1) = \sigma_{uu_1} \frac{\varphi(\frac{Z\beta}{\sigma})}{\vartheta(\frac{Z\beta}{\sigma})} \equiv \sigma_{uu_1} \lambda_1 \quad (10.4)$$

$$E(u_2|I = 0) = \sigma_{uu_2} \frac{-\varphi(\frac{Z\beta}{\sigma})}{1 - \vartheta(\frac{Z\beta}{\sigma})} \equiv \sigma_{uu_2} \lambda_2 \quad (10.5)$$

olacaktır. Burada φ ve ϑ sırasıyla standart olasılık yoğunluk fonksiyonu ve standart birikimli dağılım fonksiyonudur. $Z\beta$ 'dan elde edilen φ ve ϑ oranlarına ters Mill oranı denilmektedir ve λ_1 ve λ_2 ile gösterilmektedir. Tahmin edilen σ_{uu_1} ve σ_{uu_2} kovaryans katsayıları sıfırdan istatistiksel olarak farklıysa bu içsellik probleminin varlığına ve örneklem seçim yanlılığına (sapmasına) işaret edecektir (Maddala, 1986; Maddala ve Nelson, 1975). Örneklem seçim yanlılığından dolayı sistem rejim birdeyken rejim ikideki bireylerin gözlem değerleri yer almamakta, yani kırılmakta, rejim ikideyken rejim birdeki bireylerin gözlemleri kırılmaktadır. Yani sistem rejim birdeyken rejim ikideki bağımlı değişkenin gözlem değerleri gözlenememektedir. Kırpmadan dolayı her rejim için bağımlı değişkenin beklenen değerinde de bir sapma meydana gelmektedir (Long vd., 2017). Sonuç olarak, ρ_{uu_1} ve ρ_{uu_2} karar denklemi hata terimi ile rejim denklemi hata terimi arasındaki korelasyon katsayılarını göstermekte olup, bağımlı değişkenin koşullu beklenen değeri üzerinden hem içinde bulunulan rejim için hem de diğer rejim için işlenmiş ve işlenmemiş ortalama müdahale³ etkisi hesaplanmaktadır. Burada işlenmiş ve işlenmemiş ortalama iyileştirme etkileri, hesaplanan ve örneklem seçim probleminden dolayı hesaplanmayan ortalama etkileri ifade etmektedir.

$$E(y_{1i}|I = 1, X_{1i}) = \beta_1 X_{1i} + \sigma_{uu_1} \rho_1 \frac{\varphi(Z\beta)}{\vartheta(Z\beta)} \quad (10.6)$$

$$E(y_{2i}|I = 0, X_{2i}) = \beta_1 X_{2i} - \sigma_{uu_2} \rho_1 \frac{\varphi(Z\beta)}{(1 - \vartheta(Z\beta))} \quad (10.7)$$

$$E(y_{2i}|I = 1, X_{1i}) = \beta_2 X_{1i} + \sigma_{uu_2} \rho_2 \frac{\varphi(Z\beta)}{\vartheta(Z\beta)} \quad (10.8)$$

$$E(y_{1i}|I = 0, X_{2i}) = \beta_2 X_{2i} - \sigma_{uu_1} \rho_2 \frac{\varphi(Z\beta)}{(1 - \vartheta(Z\beta))} \quad (10.9)$$

İşlenmiş ortalama iyileştirme etkisi için denklem (10.6) ile (10.8) arasındaki fark işlenmemiş ortalama iyileştirme etkisi için de denklem (10.7) ile (10.9) arasındaki fark alınır.

Yukarıda ifade edildiği gibi içsel (endogenous) switching regresyon modelleri EKK ile tahmin edilirse tahmin edilen katsayılar sapmalı ve tutarsız olacaktır. Sapmalı olmasının nedeni hem

³ “Average treatment effect on the treated and un treated” den çevrilmiştir.

karar denklemindeki bağımlı değişkeni hem de rejimlerdeki bağımlı değişkenleri etkileyen aynı değişkenlerin varlığıdır. Uygulamadaki örnekte bireylerin spor yapmasını ve aynı zamanda obez olmasını etkileyen ortak değişkenler vardır. EKK tahmincilerinin tutarlılık ve sapmasız olmamasının bir diğer nedeni, örneklem seçim yanlılığıdır. Sonuçta rejim değişikliği olmuş bir ilişkide sadece bir rejim varmış gibi tahmin yapmak hem katsayılarda bir sapmaya neden olacak, hem de katsayılara ilişkin varyanslar olduğundan büyük hesaplanacaktır ve etkinlik kaybına sebep olacaktır. Dolayısıyla switching regresyon modelleri tam bilgi altında en çok olabilirlik metodu kullanılarak tahmin edilmelidir (Lokshin ve Sajaia, 2004).

10.3. Switching Regresyon Modelinin Stata Uygulaması

Bu bölümde önce kullanılan veri seti ve model tanıtılacak, sonra Stata paket programı yardımı ile elde edilen uygulama sonuçları yorumlanacaktır.

Uygulamada Türkiye’de Sporun BKİ üzerindeki etkisi araştırılmak istenmektedir ve bu bağlamda Türkiye İstatistik Kurumu (TÜİK) 2016 yılı Hanehalkı Sağlık Araştırması Anketi verileri kullanılmıştır. Anket verilerinden öncelikle bireylerin BKİ (Beden Kitle Endeksi) değerleri $\frac{Kilo}{(Boy)^2}$ formülü ile hesaplanmıştır. Bu hesaplama ve uygulamada kullanılan BKİ aralıkları Dünya Sağlık Örgütü’nün obezite sınıflaması için kullandığı ölçülerdir. Buna göre BKİ değerleri 0-18 arasında olan bireyler aşırı zayıf, 18-24.99 arasında olan bireyler normal kilolu, 25-29.99 arasında olan bireyler kilolu ve 30 veya daha büyük BKİ değerine sahip bireyler obez grubunda yer almaktadır. Spor/egzersiz yapmanın BKİ değeri üzerindeki etkisinin ölçülmeye çalışıldığı uygulamada spor/egzersiz yapanlar bir rejimi oluştururken, spor/egzersiz yapmayanlar diğer rejimi oluşturmaktadır. Yukarıda modeller tanımlanırken de ifade edildiği gibi burada tahmin edilen model içsel (endogenous) bir modeldir. Model tahmini aşamasında öncelikle karar denkleminin çıktıları değerlendirilmektedir. Bu çalışmada karar denklemini bireylerin spor yapmasını etkileyen faktörlerin belirlenmesi amaçlandığından, karar denkleminin bağımlı değişkeni birey spor yapıyorsa 1, yapmıyorsa 0 değerini alan kukla değişkendir. Karar denkleminin bağımlı değişken 0 ve 1 değerlerini aldığı için denklem logit model (probit modeliyle de tahmin edilebilir) ile tahmin edilmektedir. Karar denkleminde bireylerin spor yapma olasılıklarını etkileyen faktörler belirlendikten sonra spor yapmanın bireylerin BKİ değerleri üzerindeki etkileri için rejim1 ve rejim2 regresyon modelleri tahmin edilmektedir. Rejim1 ve rejim2 regresyonlarında bağımlı değişkenler sırasıyla spor yapan bireylerin BKİ değerleri ve spor yapmayan bireylerin BKİ değerleridir. Karar denkleminin ve rejim denklemlerinin bağımlı değişkenini etkileyebilecek bazı değişkenlerin aynı olmasından dolayı içsellik sorunuyla karşılaşılabilir. Kullanılan modeller aşağıdaki şekilde gösterilebilir.

$$Y^* = Z\beta + u \quad (\text{Karar denklemini}) \quad (10.10)$$

$$y_1 = X\beta_1 + u_1 \quad (1. \text{ Rejim}) \quad (10.11)$$

$$y_2 = X\beta_2 + u_2 \quad (2. \text{ Rejim}) \quad (10.12)$$

Burada Y^* birey spor yapıyorsa 1, birey spor yapmıyorsa 0 değerini alan ve logit model ile tahmin edilen karar denkleminin bağımlı değişkenidir. Z spor yapmayı etkilediği düşünülen bağımsız değişkenler matrisi, X rejim1 ve rejim2 regresyonlarında kullanılan açıklayıcı değişkenler matrisi, β_1 ve β_2 ise katsayılar matrisidir.

Uygulamadaki araştırma sorusu “Spor yapmak bireylerin BKİ değerleri üzerinde, dolayısıyla obez olma olasılıkları üzerinde bir etkiye sahip midir?” şeklindedir. Dolayısıyla karar denkleminde öncelikle bireylerin spor yapma olasılığını etkileyen faktörler belirlenmelidir. Bireylerin spor yapma olasılıklarını etkileyen faktörler belirlendikten sonra, spor yapmanın BKİ değeri üzerindeki etkisi rejim1 ve rejim2 denklemleri yardımı ile belirlenmektedir. Karar denkleminde örneklemdaki tüm bireylere ait gözlemler yer almaktadır. Bu uygulama için karar denklemi 4677 birey için tahmin edilmiştir.

Rejim1 spor yapan bireyleri temsil eden denklemdir ve y_1 spor yapan bireylerin BKİ değerleridir. Örneklemdaki bireylerin 1710 tanesi spor yapmaktadır ve rejim1 denklemi için gözlem sayısı 1710’dur. Rejim2 spor yapmayan bireyleri temsil eden regresyon denklemidir ve y_2 spor yapmayan bireylerin BKİ değerleridir. Burada spor yapmayan 2967 bireyin gözlem değerleri mevcuttur ve bağımlı değişken rejim2 için bu bireylerin BKİ değerleridir. Rejim1 ve rejim 2’de BKİ değerleri doğrudan alındığı için bireylerin bir kısmı obez iken bir kısmı normal kilolu veya zayıftır.

Uygulamada, düzenli egzersiz/spor yapan bireyler bir rejimde, düzenli egzersiz/spor yapmayan bireyler ise diğer rejimde yer almaktadır. Model tahmini yapılırken eğitim, hanehalkı geliri vb. değişkenler alt gruplara ayırarak kukla değişkenler oluşturuldu ve her grubun etkisi ayrıntılı olarak belirlendi. Uygulamaya başlamadan önce değişkenlerin özet istatistikleri ve tanımları Tablo 10.1’de özetlenmiştir.

Tablo 10 .1. Değişkenler ve tanımları

<i>Değişken</i>	<i>Tanım</i>
BKİ	Bireylerin vücut kitle endeksini gösteren değişkendir.
Cinsiyet	Birey erkek ise 1, değilse 0 kukla değişkenidir.
Yaş	Bireyin yaşını göstermektedir.
Eğit1	Birey herhangi bir okul mezunu değil ise 1, aksi hâlde 0 kukla değişkenidir.
Eğit2	Birey ilkökul mezunu ise 1, değilse 0 kukla değişkenidir.
Eğit3	Birey lise mezunu ise 1, değilse 0 kukla değişkenidir.
Eğit4	Birey herhangi bir Yükseköğretim kurumundan mezun ise 1, değilse 0 değişkenidir.
Gelir1	H.H geliri 0-1814 TL arasında ise 1, değilse 0 değişkenidir.
Gelir2	H.H geliri 1815-3720 TL arasında ise 1, değilse 0 değişkenidir.
Gelir3	H.H geliri 3721+ TL arasında ise 1, değilse 0 değişkenidir.
Egzersiz	Birey günde en az bir saat spor yapıyorsa 1, değilse 0 kukla değişkenidir.
Diyabet	Birey kronik diyabet hastasıysa 1, değilse 0 kukla değişkenidir.
Depresyon	Birey depresyondaysa 1, değilse 0 kukla değişkenidir.
Sigara	Birey sigara içiyorsa 1, değilse 0 kukla değişkenidir.
Alkol	Birey alkol kullanıyorsa 1, değilse 0 kukla değişkenidir.
Erkekyaş	Erkek bireylerin yaşını gösteren değişkendir.
Sağlık1	Sağlık durumu iyi veya çok iyi ise 1, değilse 0
Sağlık2	Sağlık orta ise 1, değilse 0
Sağlık3	Sağlık durumu kötü veya çok kötü ise 1, değilse 0

Bu değişkenler TÜİK'in 2016 yılı Türkiye Sağlık Araştırması Anketi'ndeki sorulardan yola çıkarak oluşturulmuştur. Örnek olarak ankette eğitim durumu ile ilgili soruya verilecek on bir farklı cevap mevcuttur. Eğitim değişkeninin detayları "tab" komutu yardımı ile gösterilebilir. Tab komutu ile Stata'da herhangi bir değişkenin frekansını, yüzdelik değerleri ve birikimli yüzdelik değerleri elde edilir.

tab EGITIM

Tablo 10.2 Eğitim Değişkenine Verilen Cevapların Dağılımı

EGITIM	Freq.	Percent	Cum.
0	170	3.63	3.63
1	1,581	33.80	37.44
3	978	20.91	58.35
4	251	5.37	63.72
5	473	10.11	73.83
6	76	1.62	75.45
7	7	0.15	75.60
8	285	6.09	81.70
21	510	10.90	92.60
22	8	0.17	92.77
23	338	7.23	100.00
Total	4,677	100.00	

Bu tabloda EGITIM sorusuna verilen cevapların dağılımı yer almaktadır. Buna göre katılımcıların %3.63'ü "Bir okul bitirmedim" yanıtını verirken, %0.15'i "Doktora mezunuyum" yanıtını vermiştir. Bu uygulamada eğitim sorusu dört farklı alt gruba ayrılmıştır⁴. Değişken tablosundaki "Erkekyaş" değişkeni etkileşim değişkenidir.

⁴ Eğitim değişkenleri oluşturulurken aşağıdaki komutlar kullanılmıştır:

gen eđit1= EGITIM==0

gen eđit2 (EGITIM==1)|(EGITIM>20 & EGITIM<24)

gen eđit3=EGITIM==3

gen eđit4=(EGITIM==4)|(EGITIM>4 & EGITIM<9)

Burada (...)|(...) komutu "veya" anlamındadır. Örneğin "**gen eđit4=(EGITIM==4)|(EGITIM>4 & EGITIM<9)**" komutu eğitim ile ilgili soruya 4 veya 5, 6, 7, 8 cevaplarından biri olması durumunda eđit4=1 aksi durumda eđit4=0 değişkenini oluştur demektir.

sum BKI yas yaskare cinsiyet egit1 egit2 egit3 egit4 gelir1 gelir2 gelir3 egzersiz sigara depresyon
erkekyas saglik1 saglik2 saglik3

Tablo 10.3 Değişkenlerin Özet İstatistikleri

Variable	Obs	Mean	Std. Dev.	Min	Max
BKI	4,677	26.3838	5.104695	13.98401	51.90311
cinsiyet	4,677	.4455848	.4970833	0	1
Yas	4,677	42.74278	16.91577	15	97
Yaşkare	4,677	2113.028	1585.045	225	9409
Eğit1	4,677	.0972846	.2963766	0	1
Eğit2	4,677	.5210605	.4996097	0	1
Eğit3	4,677	.2091084	.4067154	0	1
Eğit4	4,677	.1725465	.3778951	0	1
Gelir1	4,677	.1436818	.3508043	0	1
Gelir2	4,677	.4770152	.4995248	0	1
Gelir3	4,677	.379303	.4852655	0	1
Sigara	4,677	.3121659	.4634267	0	1
Alkol	4,677	.1421852	.3492773	0	1
Erkekyas	4,677	18.99829	24.02167	0	93
Egzersiz	4,677	.365619	.4816548	0	1
Sağlık1	4,677	.0912978	.288063	0	1
Sağlık2	4,677	.2747488	.4464353	0	1
Sağlık3	4,677	.6339534	.4817739	0	1

Yukarıdaki Stata çıktısında uygulamada kullanılan tüm değişkenlerin özet istatistikleri yer almaktadır. Tüm değişkenler için gözlem sayısı 4677'dir. Bu gözlemler 2016 Türkiye Sağlık Araştırması Anketinden rastgele seçilmiştir. Tablodaki "mean" sütunu özellikle (0,1) ikili (kukla) değişkenler için önemlidir. Cinsiyet değişkeninin ortalaması yaklaşık olarak 0.4455 olarak hesaplanmıştır. Yani örneklemdaki bireylerin yaklaşık olarak %44.55'i erkektir. Ayrıca örneklemdaki bireylerin ortalama yaşı 42.74'tür. Bireylerin %14.36'sı düşük gelir grubunda yer alırken %47.70'si orta gelir grubunda, %37.93'ü ise yüksek gelir grubunda yer almaktadır. Rejimleri oluşturan egzersiz değişkenine göre ankete katılan bireylerin %36.6'sı günlük en az 30 dakika spor yapmaktadır. Bu egzersiz değişkeni ile ilgili detaylar "tab" komutu ile incelenmiştir.

tab Egzersiz

Tablo 10.4 Egzersiz Değişkeninin Özet İstatistikleri

Egzersiz	Freq.	Percent	Cum.
0	2,967	63.44	63.44
1	1,710	36.56	100.00
Total	4,677	100.00	

Görüldüğü gibi gözlemler ilk alt örnekleme ayrılmakta; 1710 birey rejim1'in (spor yapanların) gözlemlerini, 2967 birey rejim2'nin (spor yapmayanların) gözlemlerini oluşturmaktadır. Değişkenlere ait özellikler incelendikten sonra model tahmini yapılabilir.

“10.2 Switching Regresyon Tahmin Problemi” alt bölümünde bu modellerin EKK ile tahmin edilmesinin sapmalı tahminlere neden olacağı belirtilmişti. Dolayısıyla bu modellerin endogenous (içsel) olarak tahmin edilmesi gerekmektedir. Stata paket programda bu tür modeller tahmin edilirken “movestay” komut kullanılmaktadır⁵.

⁵ Movestay komutu En çok olabilirlik (EÇOF) fonksiyonunun birinci ve ikinci dereceden türevlerini alarak optimizasyon yapmaktadır. Komut dirençli tahminci seçeneği de sunmaktadır. Movestay komutunun genel yazılışı şu şekildedir:

movestay (depvar1 [=] varlist1) [(depvar2 [=] varlist2)] [if exp] [in range] [weight], select(depvar, = varlist,) [robust cluster(varname) maximize_option]

Regresyon modellerindeki bağımsız değişkenlerin aynı olması ve sadece bir bağımlı değişkenin olması durumunda iki farklı regresyon denklemi tanımlamaya gerek yoktur. Bir denklemin tanımlanması yeterli olmaktadır. Alternatif olarak iki farklı bağımlı değişkenin mevcut olması durumunda ise parantez içindeki depvar1 ve depvar2 regresyonların ikisi de tanımlanmalıdır. Movestay komutu çalıştırıldıktan sonra “**mspredict**” komutu yardımı ile (10.6) ile (10.9) arasındaki koşullu beklenen değerler de hesaplanabilir (Lokshin ve Sajaia, 2004).

movestay BKI yas yaskare cinsiyet egit2 egit3 egit4 gelir2 gelir3 sigara erkekyas depresyon, select(egzersiz = saglik1 saglik2 egit2 egit3 gelir1 gelir2)

Tablo 10.5 Switching Regresyon Model Çıktıları

Fitting initial values

Iteration 0: log likelihood = -16534.124

Iteration 1: log likelihood = -16478.313

Iteration 2: log likelihood = -16469.247

Iteration 3: log likelihood = -16468.835

Iteration 4: log likelihood = -16468.833

Endogenous switching regression model Number of obs = 4677

Wald chi2(11) = 318.36

Log likelihood = -16468.833 Prob > chi2 = 0.0000

	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
-----+-----						
BKI_1						
yas	.5307409	.0401772	13.21	0.000	.4519951	.6094868
yaskare	-.0052649	.0004582	-11.49	0.000	-.006163	-.0043668
cinsiyet	2.71168	.7018126	3.86	0.000	1.336153	4.087208
egit2	.154637	.6173213	0.25	0.802	-1.055291	1.364565
egit3	-.3492328	.669524	-0.52	0.602	-1.661476	.9630101
egit4	-1.435694	.6922671	-2.07	0.038	-2.792513	-.0788758
gelir2	.3961033	.4084437	0.97	0.332	-.4044317	1.196638
gelir3	.4721545	.4327224	1.09	0.275	-.3759659	1.320275
sigara	-1.011951	.2863901	-3.53	0.000	-1.573265	-.4506363
erkekyas	-.0085349	.0161334	-0.53	0.597	-.0401557	.0230859
depresyon	-.5373986	.4666433	-1.15	0.249	-1.452003	.3772055
_cons	7.410002	1.107328	6.69	0.000	5.239678	9.580325
-----+-----						
BKI_0						
yas	.5119177	.0251205	20.38	0.000	.4626824	.5611531
yaskare	-.0041627	.0002632	-15.81	0.000	-.0046786	-.0036468
cinsiyet	3.214875	.494057	6.51	0.000	2.246542	4.183209
egit2	-.0393718	.3120893	-0.13	0.900	-.6510557	.572312
egit3	-.9598293	.3761313	-2.55	0.011	-1.697033	-.2226256
egit4	-1.442324	.3940024	-3.66	0.000	-2.214554	-.6700933
gelir2	.4301861	.2560108	1.68	0.093	-.0715859	.9319582
gelir3	.6581819	.2809301	2.34	0.019	.107569	1.208795
sigara	-.9727619	.2169758	-4.48	0.000	-1.398027	-.5474971
erkekyas	-.0709855	.0103823	-6.84	0.000	-.0913344	-.0506366
depresyon	-.0690656	.2929997	-0.24	0.814	-.6433344	.5052032
_cons	14.51007	.6542569	22.18	0.000	13.22775	15.79239
-----+-----						
egzersiz						
yas	.0127981	.0058235	2.20	0.028	.0013844	.0242119
yaskare	-.0001878	.0000646	-2.91	0.004	-.0003145	-.0000612
cinsiyet	.1950578	.1047052	1.86	0.062	-.0101606	.4002763
egit2	.3877904	.0833491	4.65	0.000	.2244292	.5511516
egit3	.4218925	.0922646	4.57	0.000	.2410572	.6027278
egit4	.372803	.09617	3.88	0.000	.1843131	.5612928
gelir2	-.065855	.0424824	-1.55	0.121	-.149119	.0174089
sigara	.053404	.0439391	1.22	0.224	-.0327151	.1395231
erkekyas	.003004	.0023601	1.27	0.203	-.0016218	.0076298
depresyon	.0320808	.0688266	0.47	0.641	-.1028169	.1669784

saglik1		-.4442324	.0640086	-6.94	0.000	-.569687	-.3187779
saglik2		-.1900737	.0357203	-5.32	0.000	-.2600843	-.1200631
gelir1		-.1138873	.0635669	-1.79	0.073	-.2384761	.0107016
_cons		-.8979698	.1527268	-5.88	0.000	-1.197309	-.5986308

/lns1		1.856095	.0301484	61.57	0.000	1.797006	1.915185
/lns2		1.559669	.0195454	79.80	0.000	1.521361	1.597978
/r1		1.641779	.0820245	20.02	0.000	1.481014	1.802544
/r2		.3130476	.0931104	3.36	0.001	.1305545	.4955407

sigma_1		6.398704	.1929109			6.03156	6.788197
sigma_2		4.757248	.0929822			4.578453	4.943025
rho_1		.9277209	.0114288			.9016578	.9470689
rho_2		.303207	.0845504			.1298177	.4586029

LR test of indep. eqns. :				chi2(1) =	120.97	Prob > chi2 =	0.0000

Model bölümünde açıklandığı gibi switching regresyon modellerinde (iki rejimli modeller için) üç regresyon denklemi tahmin edilmektedir. İlk denklem karar denklemi olarak isimlendirilmektedir. Bu uygulamada spor (egzersiz yapmanın) yapmanın bireylerin BKİ değerleri üzerindeki etkisi incelenmiştir. Dolayısıyla karar denkleminde öncelikle bireylerin spor yapmasını etkileyen değişkenler belirlenmektedir. Bu modelde bağımlı değişken 0 veya 1 değerlerini almaktadır ve model logit model ile tahmin edilmektedir. Karar denkleminde bireylerin spor yapmasını etkilediği düşünülen yaş, cinsiyet ve eğitim gibi değişkenler modele bağımsız değişken olarak dâhil edilmiştir. Stata çıktısında karar denklemine ait sonuçlar çıktının en altında (burada üçüncü kısımda) yer almaktadır.

Çıktının üçüncü kısmında yer alan karar (Egzersiz) denklemindeki değişkenlerin katsayılarına baktığımızda yaş ve yaşkare değişkenlerin katsayılarının istatistiki olarak anlamlı oldukları, bireylerin yaşı arttıkça bireylerin spor yapma olasılıklarının arttığı, ancak yaşın etkisinin giderek azaldığı görülmektedir. Çünkü yaşkare değişkeninin katsayı işareti negatiftir. Erkeklerin spor yapma olasılıkları kadınlara göre ortalama 0.19 kadar daha fazladır. Bireylerin eğitim düzeyleri arttıkça spor yapma olasılıkları da artmaktadır. Tüm eğitim düzeyleri için en düşük eğitim düzeyine (referans grup) kıyasla bireylerin spor yapma olasılıkları daha yüksektir ve katsayıları istatistiksel olarak anlamlıdır. Burada gelir ve sağlık durumu ile ilgili katsayılar incelenirken dikkatli olmak gerekmektedir. Çünkü her iki değişken grubu için de en yüksek gruptaki değişkenler referans gruplar olarak seçilmiştir. Örneğin Gelir3 değişkeni en yüksek hanehalkı gelir düzeyine sahip bireyleri temsil etmektedir. Dolayısıyla Gelir1 ve Gelir2 değişkenlerinin katsayılarının negatif çıkması, yüksek gelir grubundaki bireylerin spor yapma olasılıklarının düşük gelir gruplarına göre daha yüksek olduğunu ya da aynı anlama gelecek şekilde düşük ve orta hanehalkı gelirine sahip bireylerin spor yapma olasılıklarının yüksek gelir grubundaki bireylere göre daha düşük olduğunu göstermektedir. Benzer bir yorum sağlık durumu değişkenleri için de yapılabilir. Sağlık3 değişkeni kendini çok iyi hisseden bireyleri temsil ettiğine göre, kendini kötü hisseden (sağlık durumunu kötü olarak tanımlayan) veya sağlık durumunu orta düzeyde iyi olduğunu hisseden bireylerin spor yapma olasılıkları sağlıklı hisseden bireylere kıyasla daha düşüktür. Son olarak karar denklemi için depresyondaki bireylerin spor yapma olasılıkları depresyonda olmayan bireylere göre daha yüksektir.

Birinci rejime ait sonuçlar BKI_1 çıktısında verilmektedir. Gelir düzeyi arttıkça spor yapan bireylerin vücut kitle endeksleri de artmaktadır. Örneğin üst gelir (Gelir3) grubundaki spor yapan bireylerin vücut kitle endeksleri en düşük gelire sahip bireylere göre ortalama 0.48 daha fazladır. Spor yapan bireyler için bireylerin eğitim düzeyi arttıkça ortalama BKI değerleri düşmektedir. Düşük eğitim düzeyleri için katsayı pozitif ama istatistiksel olarak anlamsızken, eğitim düzeyi arttıkça katsayının işareti negatife dönmekte ve katsayı istatistiksel olarak anlamlı olmaktadır. Rejim birdeki bireyler için yaş arttıkça BKI'nin de ortalama olarak arttığı görülmektedir. Fakat yaşın etkisi belli bir noktadan sonra azalarak devam etmektedir yani Yaşkare değişkeninin katsayısı işareti negatiftir. Spor yapan bireylerde sigara içenlerin BKI değerleri sigara içmeyenlere göre ortalama 0.88 kadar daha azdır.

Stata çıktısının ikinci kısmında rejim2 regresyon denkleminin katsayıları yer almaktadır. Bu sonuçlar BKI_0 kısmında yer almaktadır. Buna göre spor yapmayan bireyler için yaş arttıkça ortalama BKI'de artmaktadır. Fakat tıpkı rejim birde olduğu gibi yaşın etkisi belli bir noktadan sonra azalarak devam etmektedir. Yani yaş ile BKI arasındaki ilişki konkavdır. Spor yapmayan bireyler arasında erkek bireylerin ortalama BKI değerleri kadın bireylerin BKI değerlerinden yaklaşık olarak 3.18 daha fazladır. İkinci rejimdeki bireyler için bireylerin eğitim düzeyleri arttıkça ortalama BKI değerlerinin azaldığı görülmektedir. Sonuçlara göre en yüksek eğitim grubundaki bireyler en düşük eğitim grubundaki bireylere kıyasla ortalama BKI değerleri 1.41 daha azdır. Spor yapamayan bireylerde gelir düzeyi arttıkça ortalama BKI değerlerinin de arttığı görülmektedir. Gelir değişkenlerinin katsayıları hem pozitif işaretli hem de istatistiksel olarak anlamlı bulunmuştur.

Sigara içmenin spor yapmayan bireylerin BKI değerleri üzerindeki etkileri incelendiğinde görülmüyor ki, sigara içen bireylerin BKI değerleri sigara içmeyenlere göre ortalama 0.92 daha düşüktür. Ayrıca rejim ikideki bireylerin depresyonda olması BKI değerlerini negatif yönde etkilemektedir. Fakat katsayı istatistiksel olarak anlamlı değildir. Son olarak rejim iki için depresyondaki bireylerin BKI değeri depresyonda olmayan bireylere göre ortalama 0.05 daha azdır.

Stata çıktısındaki karar denklemi, rejim1 denklemi ve rejim2 denkleminin katsayıları yorumlanmıştır. Fakat katsayıların altında birtakım istatistikler yer almaktadır. Örneğin $\ln s_1$, $\ln s_2$, r_1 , r_2 , ρ_1 ve ρ_2 istatistikleri verilmiştir. Burada ρ_1 σ_{uu1} katsayısını, ρ_2 σ_{uu2} katsayısını temsil etmektedir. Görüldüğü gibi iki katsayı da sıfırdan farklıdır ve istatistiksel olarak anlamlıdır. Yani incelenen ilişkinin içsel switching regresyon modeli olarak tahmin edilmesi uygundur. Bu ilişki EKK ile tahmin edilirse sapmalı ve tutarsız katsayılar elde edilecektir. Bu katsayılar ek olarak r_1 ve r_2 sırasıyla denklem (10.6) ve (10.9)'daki ρ_1 ve ρ_2 katsayılarıdır. $\ln s_1$ ve $\ln s_2$ ise En Çok olabilirlik tahmininde yer alan sigmaları temsil etmektedir.

Movestay komutu kullanılarak switching regresyon modeli bir kere tahmin edildikten sonra "mspredict" komutu yardımı ile denklem (10.6) ile (10.9) arasında verilen bağımlı değişkenin koşullu beklenen değerleri kolaylıkla hesaplanabilir. Bunun için aşağıdaki komutları girmek yeterlidir.

mspredict A, yc1_1
mspredict B, yc1_2
mspredict C, yc2_2
mspredict D, yc2_1

Burada A, B, C ve D keyfi verilmiş isimlerdir ve her biri bağımlı değişkenin bir koşullu beklenen değerini temsil etmektedir⁶. Bu komutlar girildikten sonra Stata paket programı A, B, C ve D parametrelerini depolamaktadır (saklamaktadır). Sonrasında “sum” komutu ile bu parametrelerin detayları görülebilir.

sum A B C D

Tablo 10.6 Koşullu Beklenen Değerler

Variable	Obs	Mean	Std. Dev.	Min	Max
A	1,710	25.81898	2.199155	20.07644	31.08979
B	2,967	16.37955	2.087698	9.014247	21.35516
C	2,967	26.72569	2.471291	19.81148	30.61992
D	1,710	28.56041	2.400735	22.65534	33.2798

Yukarıdaki çıktının “mean” sütunu ilgili koşullu beklenen değerleri vermektedir. “İçsel (Endogenous) Switching Regresyon Modeli” alt başlığı altında verilen iyileştirilmiş ortalama etkiler bu ortalamalardan hareketle hesaplanabilir. Son olarak yukarıdaki tablonun son iki sütunundaki değerler koşullu beklenen değerlere ait güven aralıklarıdır. Bu değerler her bir rejimin gözlenen ve gözlenmeyen değerlerine ilişkin beklenen değerlerdir ve örneklem dağılımı ile ilgili bilgi vermektedir.

Bu bölümde kullanılan Stata komutları aşağıda listelenmiştir:

```
import excel "C:\Users\Switching\Desktop\Kitapbölüm\obez.xlsx", sheet("Sayfa2") firstrow
gen cinsiyet= CINSIYET==1
tab EGITIM
gen eđit1= EGITIM==0
gen eđit2 (EGITIM==1)|(EGITIM>20 & EGITIM<24)
gen eđit3=EGITIM==3
gen eđit4=(EGITIM==4)|(EGITIM>4 & EGITIM<9)
sum eđit1 eđit2 eđit3 eđit4
```

⁶ **mspredict A, yc1_1** komutu rejim1 deki gözlenen değerler için (denklem (10.6)),

mspredict B, yc1_2 komutu rejim1 deki gözlenmeyen değerler için (denklem (10.7)),

mspredict C, yc2_2 komutu rejim2 deki gözlenen değerler için (denklem (10.8)),

mspredict D, yc2_1 komutu rejim1 deki gözlenmeyen değerler için kullanılmaktadır (denklem (10.9)),

```

tab HANE_GELIR
gen gelir1= HANE_GELIR==1
gen gelir2= HANE_GELIR==2
gen gelir3= HANE_GELIR==3
gen gelir4= HANE_GELIR==4
gen sađlık1= GENELSAGLIKDURUM==4| GENELSAGLIKDURUM==5
gen sađlık2= GENELSAGLIKDURUM==3
gen sađlık3= GENELSAGLIKDURUM==1| GENELSAGLIKDURUM==2
tab YURUME_ZAMAN
tab BISIKLET_ZAMAN
gen Egzersiz= YURUME_ZAMAN>1| BISIKLET_ZAMAN>1
tab Egzersiz

sum BKI yas yaskare alchol gender educ1 educ2 educ3 educ4 income1 income2 income3 exercisel dia-
bet smoker depression yaserk

movestay BKI yas yassquare Cinsiyet Eđit2 Eđit3 Eđit4 Gelir2 Gelir3 Sigara Erkekyas Depresyon, se-
lect( Egzersiz = sađlık1 sađlık2 Eđit2 Eđit3 Gelir1 Gelir2)

mspredict A, ycl_1
mspredict B, ycl_2
mspredict C, yc2_2
mspredict D, yc2_1

sum A B C D

```

10.4. KAYNAKLAR

- Abdulai, A., ve Huffman, W. (2014). The adoption and impact of soil and water conservation technology: An endogenous switching regression application. *Land economics*, 90(1), 26-43.
- Adela, F. A., ve Aurbacher, J. (2018). Access to irrigation water-poverty nexus: Application of an Endogenous Switching Regression in Ethiopia.
- Alene, A. D., ve Manyong, V. M. (2007). The effects of education on agricultural productivity under traditional and improved technology in northern Nigeria: an endogenous switching regression analysis. *Empirical economics*, 32(1), 141-159
- Deutsch, M., Granger, C. W., ve Teräsvirta, T. (1994). The combination of forecasts using changing weights. *International Journal of Forecasting*, 10(1), 47-57
- Hamilton, J. D. (1994). *Time series analysis* (Vol. 2, pp. 690-696). New Jersey: Princeton
- Hausman, J. A. (1983). Specification and estimation of simultaneous equation models. *Handbook of econometrics*, 1, 391-448."

- Lokshin, M., ve Sajaia, Z. (2004). Maximum likelihood estimation of endogenous switching regression models. *The Stata Journal*, 4(3), 282-289.
- Loureiro, M. L., Yen, S. T., ve Nayga, Jr, R. M. (2012). The effects of nutritional labels on obesity. *Agricultural economics*, 43(3), 333-342.
- Maddala, G. S. (1986). Disequilibrium, self-selection, and switching models. *Handbook of econometrics*, 3, 1633-1688.
- Maddala, G. S., ve Nelson, F. D. (1975). Switching regression models with exogenous and endogenous switching. In *Proceedings of the American Statistical Association* (Vol. 5, pp. 423-426). Business and Economics Section.
- Ypma, T. J. (1995). Historical development of the Newton–Raphson method. *SIAM review*, 37(4), 531-551.

BÖLÜM 11

DOĞRUSAL ARAÇ DEĞİŞKENLİ REGRESYON MODELİ (IV, 2SLS, GMM) VE STATA UYGULAMASI

Merve ERTOK ONURLU¹

Ekonometride sıkça karşılaşılan sorunlardan bir tanesi modelde yer alan açıklayıcı değişken veya değişkenler ile hata terimi arasında ortaya çıkan korelasyon sorunudur. İlerleyen bölümlerde “içsellik problemi” olarak adlandırılacak olan bu tarz bir durumda, klasik doğrusal regresyon modeli varsayımlarından hata terimi u ve açıklayıcı değişken X ’in korelasyonsuz olması varsayımı ihlal edilmiş olacak ve bu varsayımın ihlali sonucunda ise açıklayıcı değişken X ’in açıklanan değişken Y üzerindeki etkisi sapmalı ve tutarsız olarak tahmin edilecektir. Araştırmacının sunacağı ilk çözüm, açıklayıcı değişken X ile korelasyonlu fakat gözlemlenemeyen stokastik değişken u hata terimi ile korelasyonsuz olan uygun bir araç değişken seçmek olacaktır. İşte ekonometride bu yönteme “Araç Değişkenler Yöntemi” adı verilir. Bu bölümde klasik doğrusal regresyon modeli çerçevesinde Araç Değişkenler Yöntemi (IV Tahmin Yöntemi) ele alınacak ve diğer tahmin yöntemlerinden İki Aşamalı En Küçük Kareler Yöntemi (2AEKK) ve Genelleştirilmiş Momentler Metodu (GMM) da incelenecektir.

11.1. İçsellik Kavramı

Klasik Doğrusal Regresyon Modelinin en önemli varsayımlarından biri, modeldeki bağımsız değişken/değişkenler ile hata teriminin (u) birbirlerinden bağımsız olmalarıdır. Kısaca (11.1) eşitliğinde gösterilen, X açıklayıcı değişkeninin değeri veri iken, hata teriminin koşullu beklenen değerinin sıfır olmasıdır.

$$E(u_i|X_i) = 0 ; \quad i = 1, \dots, N \quad (11.1)$$

N gözlem sayısını göstermek üzere, (11.1) varsayımının sağlandığı durumlar için açıklayıcı değişkenlerin dışsal (egzojen) olduğu varsayılır. Bu varsayım, Trivedi ve Cameron (2009) tarafından kullanılan örnek incelenerek ele alınsın. Aşağıdaki (11.2) regresyon modelinde açıklanan değişken Y ’nin “cepten yapılan sağlık harcaması”, açıklayıcı değişken X ’in ise bireyin sahip

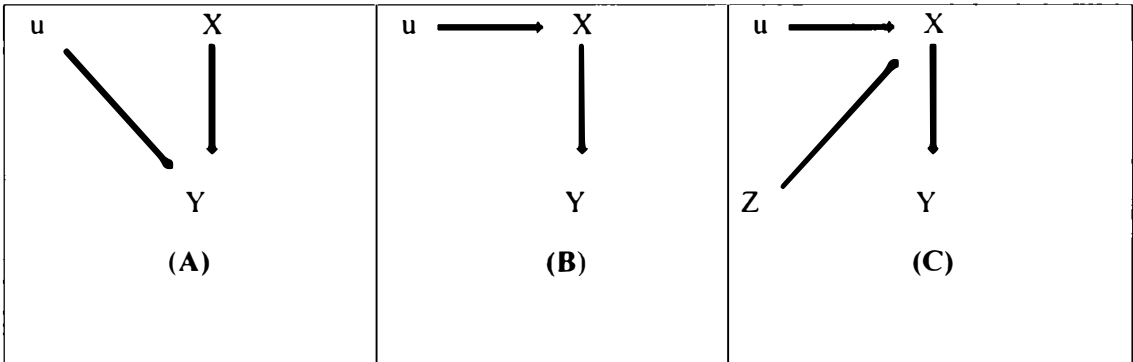
¹ Dr. Öğr. Üyesi, Çanakkale Onsekiz Mart Üniversitesi, Biga İktisadi ve İdari Bilimler Fakültesi, Ekonometri Bölümü. m.ertok.onurlu@comu.edu.tr . <https://orcid.org/0000-0001-5187-1568>.

olduğu “sağlık sigortası türü” olduğu varsayalım. Bireylerin gelirleri sahip olunan sigorta türünü ve kapsamını etkileyebilir. Bu durumda bireysel gelir değişkeninin model dışında kalması hata terimi u ile sigorta türü değişkeninin ilişkili olması, dolayısıyla (11.1)’de bahsi geçen varsayımın ihlal edilmesi sonucunu ortaya çıkarır ve sağlık sigortası türü değişkeninin içsel (endojen) açıklayıcı değişken olmasına sebebiyet verir. Bu durumda (11.2) modelinin En Küçük Kareler (EKK) Yöntemi ile tahmin edilen $\hat{\beta}$ tahmincisi sadece sağlık sigortası türünün sağlık harcaması üzerindeki direk etkisini değil, gelirin de etkisini taşıyacağından, olması gerektiğinden daha büyük tahmin edilir. (11.1) varsayımı ihlal edildiğinden EKK tahmincisi $\hat{\beta}$ sapmalı ve tutarsız olur.

$$Y = X_i'\beta + u \quad (11.2)$$

EKK yöntemi kapsamında $\hat{\beta}$ tahmincisinin doğrusal en iyi sapmasız (DESTE)² tahminci özelliğini sağladığı varsayılır. Bu varsayıma ek olarak hata teriminin normal dağılıma sahip sıfır ortalamalı ve sabit varyanslı olduğu kabul edilir. Teorik olarak tahminciler ölçme hataları, dışlanan açıklayıcı değişken, eşanlı denklem yanlılığı gibi çeşitli sebepler ile DESTE özelliklerini yitirebilirler (Ersöz ve Toktamış, 2017: 90). Bu durumda (11.1) varsayımı geçerli olmaz ve modelde yer alan açıklayıcı değişkenler ile hata terimi belirli bir yönde aynı anda hareket eder. Aşağıdaki Şekil 11.1’i incelensin. İçsel açıklayıcı değişkenlerin yer almadığı bir regresyon modeli için Şekil 11.1’de yer alan Diyagram A’yı çizmek mümkün olacaktır. Diyagram A’da, hata terimi u ile açıklayıcı değişken X arasında ilişki olmaması durumu gösterilmiştir. Diyagram B’de ise hata terimi u ile açıklayıcı değişken X arasında belirli bir ilişkinin olduğu gösterilmiş olup, bu durumda varsayım (11.1) ihlal edilmiş olur. Diyagram C’de ise bir sonraki bölümde detaylandırılacak olan uygun araç değişken Z belirlenerek içsel açıklayıcı değişken sorununa çözüm getirilmiş ve (11.1) varsayımı sağlanmıştır.

Şekil 11.1. Değişkenler ve hata terimi arasındaki ilişki diyagramları



11.2. Araç Değişken (IV-Instrumental Variable)

Araç değişken kavramı X açıklayıcı değişkeni ve u hata terimi arasında ilişki olduğu zaman (11.1) varsayımı geçersiz iken ve $\hat{\beta}$ tahmincisinin X ’in Y üzerindeki etkisini olduğundan daha büyük tahmin ettiği ve tutarlılık özelliğini sağlamadığı durumlarda ekonometrik çalışma yapan-

² DESTE özelliği İngilizce ifade ile best linear unbiased (BLUE) estimator özelliği olarak literatürde sıkça yer almaktadır.

ların faydalandığı bir değişken türüdür. Z araç değişkeni (Şekil 11.1, Diyagram C); X açıklayıcı değişkeni ile ilişkili olmalı, Y açıklanan değişkenini X açıklayıcı değişken üzerinden dolaylı olarak etkilemeli ve aynı zamanda u hata terimi ile ilişkili olmamalıdır. Bu durumda EKK tahmini sonucunda (11.1) varsayımı geçerli olacak olup ve $\hat{\beta}$ tahmindisi ise DESTTE özelliğini sağlayan tutarlı tahminci olacaktır.

Araç değişken yönteminin detaylarına geçmeden önce Z araç değişkeni belirlenirken dikkat edilmesi gereken bazı önemli özellikler aşağıdaki gibi sıralanabilir (Wooldridge, 2013: 508; Verbeek, 2012):

- a) *Seçilen araç değişkenin (Z) ilgili olma düzeyi:* Seçilen araç değişken Z'nin, X açıklayıcı değişkeni ile pozitif veya negatif ilişkili olması gerekir. Bir diğer ifade ile regresyon modelinde yer alan X açıklayıcı değişkeni ile Z araç değişkeni arasındaki kovaryans 0'dan farklı olmalıdır ($Cov(X, Z) \neq 0$). Buna literatürde araç değişken uygunluğu denir. Değişkenler arasındaki bu ilişkilerin belirlenmesi sırasında iktisat teorisi göz önüne alınarak, seçilen değişkenlerin iktisadi teori kapsamında birbirleri ile ilişkili olmaları istenir. Örneğin, tüketim ve gelir modeli ele alınsın. Mikro iktisadi teoriye göre bireylerin gelirlerindeki artış tüketimlerini pozitif yönde etkilemektedir. Fakat gelirin dışında tüketimi etkileyen eğitim, yaş, cinsiyet gibi faktörlerin model dışında kalması gelir değişkeni ile hata teriminin korelasyonlu olmasına neden olur. Bu durumda gelir değişkeni yerine gelir ile korelasyonlu bir başka değişkenin seçilmesi ve bu seçimin mikro iktisat teorisi kapsamında yapılması uygun olacaktır.
- b) *Dışsal araç değişken:* Regresyon modelinde kullanılacak olan araç değişken Z'nin hata terimi u ile korelasyonlu olmaması yani dışsal olması gerekir ($Cov(Z, u) = 0$). Bu özelliğin tam olarak sağlanması araç değişkenin bağımlı değişken üzerinde hiçbir dolaylı etkisinin olmaması anlamına gelmektedir. Bu yüzden ampirik çalışmalarda maalesef mevcut kriteri tam anlamı ile sağlayacak araç değişken bulmak oldukça zordur.
- c) *Seçilecek olan araç değişkenin modelde önceden yer almaması:* Z araç değişkeni modelde önceden yer alan açıklayıcı değişkenlerden bir tanesi olmamalıdır.

11.3. Doğrusal Araç Değişkenli Regresyon Modeli Tahmini

11.3.1. Basit Doğrusal Regresyon Modeli Kapsamında Araç Değişkenler Yöntemi ve Stata Uygulaması

Bu bölümde araç değişkenler yönteminin işleyişi ve nasıl tahmin edileceği bir örnek üzerinden ayrıntılı olarak incelenecektir:

$$\text{ölümloran}_i = \beta_0 + \beta_1(\text{haftaiçidoğanbebeksayısı}) + u_i; \quad (11.3)$$

$$i = 1, \dots, n \text{ (i. Hastane)}$$

Yukarıdaki regresyon denklemi kullanılarak i. hastanede doğan bebeklerin ölüm oranları ile doğdukları gün arasında nedensellik ilişkisi kurmak istenildiği varsayılınsın. EKK yöntemi varsayımlarının sağlanması durumunda ortalamadan farkları kullanarak β_1 parametresi EKK yöntemi

ile tahmin edilsin. Y_i ölüm oranı ve X_i hafta içi doğan bebek sayısının hafta sonu doğan bebek sayısına oranı şeklinde olmak üzere aşağıdaki model tanımlansın:

$$Y_i = \beta_0 + \beta_1 X_i + u_i \quad (11.4)$$

$$\hat{\beta}_{1(SEKK)} = \frac{\sum_{i=1}^n x_i y_i}{\sum x_i^2} \quad ; \quad x_i = X_i - \bar{X} ; y_i = Y_i - \bar{Y} \quad (11.5)$$

Teorik çerçevede düşünüldüğünde bebek ölüm oranları bebeğe veya anneye ait demografik faktörler (annenin yaşı, ırkı, önceki gebelik sayısı, genel sağlık durumu, eğitimi ve bebeğin cinsiyeti gibi) tarafından etkilenmektedir. Bu değişkenlerden herhangi birinin model dışında bırakılması hata terimi u_i ile hafta içi doğan bebek sayısı değişkenini korelasyonlu hâle getirir. Böylelikle içsel açıklayıcı değişken sorunu ortaya çıkar. Bu durumda 11.2 başlığı altında incelenen araç değişken özelliklerini sağlayan ve X_i =hafta içi doğan bebek sayısının hafta sonu doğan bebek sayısına oranı değişkenini etkileyen fakat hata terimi u_i ile korelasyonlu olmayan bir araç değişken seçmek gerekir. Sağlık ekonomisi literatüründe hafta içi doğan bebeklerin, hafta sonu doğan bebeklere göre ölüm risklerinin daha az olduğu birçok ampirik çalışmada görülmektedir (Macfarlane vd. (2019); Gould vd. (2003)). Bu durumun temelde hafta içleri ve hafta sonları arasında var olan anne-çocuk sağlığı hizmeti kalite göstergeleri arasındaki farktan ileri geldiği düşünülebilir. Hastanelerde çalışan ebe, hemşire veya uzman hekim sayıları kalite göstergeleri olarak ele alındığında; hafta içi çalışan sayısı hafta sonu çalışanların sayılarına kıyasla daha yüksek olacağından genellikle riskli gebelik veya hamilelik durumları için hastane çalışanlarının müdahalesi daha erken olur ve böylelikle bebeklerin yaşam şansları da yükselir. Önceden planlanabilen doğum türleri (sezeryan doğumlar) için ise hastaneler personel sayılarının daha fazla olduğu hafta içlerini tercih edebilirler (Gould vd., 2003: 2959). Bu durumda hastanelerde hafta içi ve hafta sonu çalışan personel sayıları oranı 11.2 başlığı altında incelenen araç değişken özelliklerini sağlayan ve hastane hizmet kalitesini etkileyen, fakat hata terimi u_i ile ilişkili olmayan bir Z araç değişkeni olarak seçilebilir. EKK ve araç değişken tahmincileri aşağıdaki gibi yazılabilir (Gujarati, 2015: 452):

$$\hat{\beta}_{1(IV)} = \frac{\sum_{i=1}^n z_i y_i}{\sum z_i^2} \quad Z_i = Z_i - \bar{Z} ; y_i = Y_i - \bar{Y} \quad (11.6)$$

Açıklanan değişken Y_i ise:

$$y_i = \hat{\beta}_1 x_i + (u_i - \bar{u}); \hat{\beta}_{1(IV)} = \frac{\sum_{i=1}^n z_i [\beta_1 x_i + (u_i - \bar{u})]}{\sum_{i=1}^n z_i x_i} = \beta_1 + \frac{\sum z_i (u_i - \bar{u})}{\sum z_i x_i} \quad (11.7)$$

şeklinde yazmak mümkün olacaktır. Sabit terim tahmincisi $\hat{\beta}_0$ ise:

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_{1(IV)}\bar{X} \quad (11.8)$$

formülü ile hesaplanır. (11.8)'de araç değişken tahmincisi ile EKK tahmincileri arasındaki ilişki görülmektedir. Büyük örneklem için (11.9) eşitliğini yazmak mümkündür. *plim* olasılık limiti olmak üzere (Wooldridge, 2013: 514; Cameron ve Trivedi, 2005: 101):

$$plim(\hat{\beta}_{1(IV)}) = plim\left(\frac{\sum_{i=1}^n z_i[\beta_1 x_i + (u_i - \bar{u})]}{\sum_{i=1}^n z_i x_i}\right) = \beta_1 + plim\left(\frac{Cov(z_i u_i)}{Cov(z_i x_i)}\right) \quad (11.9)$$

(11.8)'e göre $Cov(Z_i, u_i) = 0$ varsayıldığında, büyük örneklem için

$$plim(\hat{\beta}_{1(IV)}) = \beta_1 \quad (11.10)$$

yazılabilir. Bu durumda (11.10) eşitliğine göre IV tahmincisi $\hat{\beta}_{1(IV)}$ büyük örneklem için tutarlılık özelliğini taşır.

IV tahmincisi büyük örneklem için aynı zamanda normal dağılıma sahip olacağından $\hat{\beta}_{1(IV)}$ tahmincisi (11.11)'deki gibi ifade edilebilir (Gujarati, 2015:453):

$$\hat{\beta}_{1(IV)} \sim N\left(\beta_1, \frac{\sigma^2}{\sum x_i^2} \frac{1}{\rho_{XZ}^2}\right) \quad (11.11)$$

σ^2 hata terimi u_i 'ye ait varyans değeri ve ρ_{XZ}^2 ise X_i ve Z_i değişkenlerine ait korelasyon katsayısının karesi olmak üzere (11.11)'e göre IV tahmincisinin büyük örneklem için ortalaması parametre değeri olan β_1 'e eşit, varyansı ise $\frac{\sigma^2}{\sum x_i^2} \frac{1}{\rho_{XZ}^2}$ değerine eşit olur. Bunun yanısıra ρ_{XZ}^2 değeri 1'e ne kadar yakın ise Z_i araç değişkeni o kadar güçlü, ne kadar küçükse ρ_{XZ}^2 değeri o kadar zayıf bir araç değişken olmaktadır. EKK tahmincisi $\hat{\beta}_{1(EKK)}$ için ise (11.12)'deki gibi normal dağılıma sahip olduğu hatırlansın (Güriş ve Çağlayan, 2013: 106-107):

$$Var(\hat{\beta}_{1(EKK)}) = \sigma^2 / \sum x_i^2 \text{ olmak üzere } \hat{\beta}_1 \sim N(\beta_1, \sigma^2 / \sum x_i^2) \quad (11.12)$$

Örnek

Literatürde eğitim yılı ile ücret değişkenleri arasındaki nedensellik ilişkisini inceleyen çok sayıda regresyon modelleri vardır (Gujarati 2015; Cameran ve Trivedi, 2005). Literatürü takip ederek aşağıdaki örnek ele alınmıştır.

Eğitim yılının ücretler üzerindeki etkisini incelerken aşağıdaki log-doğrusal model (11.13) kullanılsın:

$$\ln(\text{ücret}_i) = \beta_0 + \beta_1 \text{eğitim}_i + u_i, \quad i = 1, \dots, n \quad (11.13)$$

i indisi bireyleri temsil etmek üzere, bireyin aldığı ücret ile eğitim yılı arasındaki nedensellik ilişkisi EKK yöntemi ile tahmin edilmek istendiğinde yetenek değişkeninin modelde yer almadığı ve çoğu zaman toplanan veriler arasında yer almaması sebebiyle değişkenin etkisi hata terimi u_i 'de toplanır ve bu durumda (11.1) varsayımı ihlal edilir. Buna göre yetenek değişkenini temsilen bir Z araç değişkeni bulmak gerekir. Bu Z araç değişkeni, eğitim yılıyla ilişkili, fakat hata terimi u_i ile korelasyonsuz olmalıdır. Yapılan çalışmalarda (Card (1995); Cameran ve Trivedi (2005); Gujarati (2015)), eğitim-ücret modeline uygun araç değişken olarak bireyin bulunduğu yer ile üniversiteye veya koleje olan uzaklık önerilmiştir. Üniversiteye daha uzak olan bölgelerde bulunan bireylerin uzaklığa bağlı olarak daha az ihtimalle üniversite veya kolej eğitimi almaya eğilimli oldukları varsayılmıştır. Bu durumda açıklayıcı değişken eğitim yılı ile araç değişken üniversiteye/koleje olan uzaklık korelasyonludur. Bir diğer ifade ile alt bölüm 11.2'deki (a) özelliği sağlanmış olur. Angrist ve Krueger (1991) ikinci bir alternatif araç değişken olarak "bireyin doğduğu çeyrek yıl" değişkenini önermişlerdir. Hatta bu değişkenin 11.2 alt başlığında belirtilen tüm varsayımları çoğunlukla sağladığı da vurgulanmıştır.

Bir başka açıklayıcı örnek olarak bireyin işteki performans notu ile işe geç kaldığı günler arasındaki nedensellik ilişkisi incelensin. Aşağıdaki regresyon modeli ele alınsın:

$$\text{performans}_i = \partial_0 + \partial_1 \text{geçkalmangün}_i + u_i, \quad i = 1, \dots, n \quad (11.14)$$

Bireyin işteki performansı ile işe geç geldiği gün sayısı arasındaki nedensellik ilişkisi incelenirken bireyin işteki performansını etkileyen, fakat modelden dışlanmış olan değişkenlerin varlığı ve bu değişkenler ile işe geç kalması arasında bir ilişki olup olmadığı ele alınsın. İş motivasyonu yüksek bireylerin işe daha erken saatlerde veya tam zamanında gelecekleri düşünülürse, motivasyon değişkeninin modelde yer almaması (etkilerinin hata terimi u_i 'de yer alması) durumunda açıklayıcı değişken ile hata terimi arasında korelasyon ortaya çıkarır (Wooldridge, 2013: 509). Bu durumda hata terimi ile ilişkili olmayan, fakat işe geç kalınan gün sayısı ile ilişkili olan uygun bir Z araç değişkeni bulunması gerekir. Örneğin, yaşanan yerdeki hava koşulları bireyin işe geç kalma durumunu etkiler. Olumsuz hava şartları veya trafik yoğunluğu gibi faktörler bireyin işe tam zamanında ulaşmasını etkileyen unsurlar olabilir. 11.2'deki (a) varsayımını test etmek için aşağıdaki regresyon modeli (11.15) oluşturulup EKK yöntemi ile tahmin edilmesi gerekir.

$$\text{geçkalmangün}_i = \delta_0 + \delta_1 \text{trafikyoğunluğu}_i + u_i \quad (11.15)$$

Eğer $\hat{\delta}_1$ tahmini büyük güven düzeylerinde (%99, %95 gibi) istatistiksel açıdan anlamlı ise (a) varsayımı geçerli olur. Alt başlık 11.2'deki (b) varsayımı için ise trafik yoğunluğunun hata terimi u_i ile ilişkili olup olmadığını tartışmak gerekir. Hata terimi ve trafik yoğunluğu değişkeni arasında ilişkili olma durumu söz konusu olabilir. Bireylerin iş yerlerine yakın yerlerde yaşamaları onların trafikten daha az etkilenmelerine sebebiyet verir. Öte yandan iş yerine yakın veya uzak mesafelerde oturmaları bireylerin kişisel tercihleri ile alakalı olabileceği gibi gelir seviyeleri ile

de ilişkili olabilir. Bu durumda, bu faktörlerin modele dâhil edilmesi şartı ile trafik yoğunluğu araç değişken olarak kullanılabilir (Wooldridge, 2013: 509).

Şimdi ise TÜİK 2013 Hanehalkı Bütçe Anketi³ verileri kullanılarak bireyin eğitim yılı ve yıllık ücret bilgisi verileri ile kurulan regresyon modeli EKK yöntemi ile tahmin edilecektir. Hanehalkı Bütçe Anketi Türkiye'deki yaklaşık her ili kapsayan, bireylere ait ücret, tüketim ve gelir bilgilerinin yanı sıra birçok demografik faktörlere ilişkin bilgi veren ve düzenli aralıklar ile yapılan bir araştırmadır (Türkiye İstatistik Kurumu web sitesi, 2020). Veriler Türkiye İstatistik Kurumu'ndan elde edilmiştir. Veri seti oldukça geniş olduğundan analize veri setinin tamamı değil, sadece evli ve erkek bireylerden oluşan bir örneklem dâhil edilmiştir. Örneklem hacmi toplamda 4236 bireyden oluşmaktadır. Uygulamalı örnekler ve model tahminleri için Stata Paket Programı kullanılacaktır. Kullanılacak olan veri UCRET.dta verisi olarak isimlendirilmiştir.

İlerleyen analizlerde bireylerin yıllık ücretleri (doğal logaritması) ve eğitim yılları değişkenleri kullanılarak (11.16) modeli EKK yöntemi ile tahmin edilecektir. Dışsal olarak adlandırılacak olan değişkenler, veri setine ilişkin bilgiler ve değişkenlere ilişkin tanımlayıcı istatistikler Tablo 11.1'de yer almaktadır. Kullanılan örneklem toplamda 4236 evli erkek bireyden oluşmaktadır.

Tablo 11.1. “UCRET.dta” Veri Seti İçeriği ve Tanımlayıcı İstatistikler (N = 4236)

Panel (A): “UCRET.dta” Veri Seti İçeriği

```
use "DOSYAKONUMU\UCRET.dta", clear
global x2list bireyno yas kent esegitim is_tecrube Cocsay0_5 Coc_Sayisi yil
describe lnucrer egitim $x2list
```

variable name	storage type	display format	value label	variable label
lnucrer	float	%9.0g		Yıllık ücret logaritması
egitim	float	%9.0g		Bireyin eğitim yılı
bireyno	str4	%9s		Örneklemdeki bireye ait sıra
yas	byte	%8.0g		Bireyin yaşı
kent	byte	%8.0g		Kentsel/kırsal bölge (kent = 1, kırsal = 0)
esegitim	float	%9.0g		Eşin eğitim yılı
is_tecrube	byte	%8.0g		Toplam iş tecrübesi yıl
Cocsay0_5	byte	%8.0g		Çocuk sayısı 0-5 yaş arası
Coc_Sayisi	byte	%8.0g		Toplam çocuk sayısı
yil	float	%9.0g		2013

³ Daha detaylı bilgi için <http://www.tuik.gov.tr/MicroVeri/HBA2013/turkce/metaveri/kapsam/index.html> linkinden erişim sağlayınız.

Panel (B): Tanımlayıcı İstatistikler

summarize lnucret egitim \$x2list

Variable	Obs	Mean	Std. Dev.	Min	Max
lnucret	4,236	9.483245	.9046225	4.317488	12.38839
egitim	4,236	8.745042	4.097186	0	17
bireyno	0				
yas	4,236	40.87724	9.144571	20	80
kent	4,236	.7832861	.4120547	0	1
esegitim	4,236	7.052408	4.281708	0	17
is_tecrube	4,236	10.01959	9.319026	0	50
Cocsay0_5	4,236	.5023607	.7070194	0	5
Coc_Sayisi	4,236	1.643532	1.321847	0	11
yil	4,236	2013	0	2013	2013

$$\ln(ucret_i) = \beta_0 + \beta_1 egitim_i + u_i, \quad i = 1, \dots, n \quad (11.16)$$

“i” indisi hanehalkında yer alan ve hanehalkından sorumlu evli erkek bireyleri temsil etmektedir. $\ln(ucret_i)$ i bireye ait yıllık ücret değişkeninin doğal logaritmasını ifade ederken $egitim_i$ değişkeni ise i bireyine ait eğitim yılı değişkenini temsil etmektedir.

Aşağıdaki Stata komutu kullanılarak eşitlik (11.16)’daki log-doğrusal model tahmin edilmiştir. Tablo 11.2 elde edilen tahmin sonuçlarını göstermektedir.

Tablo 11.2. Eğitimin Erkeğin Ücreti Üzerindeki Etkisi

regress lnucret egitim

Source	SS	df	MS	Number of obs	=	4,236
Model	991.005894	1	991.005894	F(1, 4234)	=	1695.55
Residual	2474.6716	4,234	.584476051	Prob > F	=	0.0000
Total	3465.67749	4,235	.818341793	R-squared	=	0.2859
				Adj R-squared	=	0.2858
				Root MSE	=	.76451

lnucret	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
egitim	.1180663	.0028673	41.18	0.000	.1124449 .1236876
_cons	8.45075	.0276895	305.20	0.000	8.396464 8.505036

Elde edilen sonuçlara göre (11.16) modelini tekrar yazarsak aşağıdaki model elde edilir:

$$\ln(\widehat{ucret}_i) = 8.450 + 0.118 \text{ egitim}_i \quad (11.17)$$

Gözlem sayısı (n) 4236 olan eşitlik (11.17)'deki regresyon modelinde bireyin eğitim gördüğü süre 1 yıl artınca, yıllık ücretinin %11.8 oranında arttığı ve bu artışın %1 anlamlılık düzeyinde istatistiksel olarak anlamlı olduğu görülmektedir.

Stata programında “**robust**” opsiyonu sayesinde robust veya dirençli tahmin ediciler elde edilir. Bir diğer ifade ile Huber-White sandwich tahmincisi elde edilir. Bu sayede örnekleme olası aykırı değerlerden veya varsayımlardan sapmalardan etkilenmeyecek tahminciler elde edilmiş olur (Altın ve Şenoğlu, 2008: 292). Şimdi ise model (11.16) “**robust**” opsiyonu kullanılarak tahmin edilsin:

Tablo 11.3. Robust Standart Hatalar ile Eğitimin Ücret Üzerindeki Etkisi

regress lnucret egitim, robust

Linear regression	Number of obs	=	4,236
	F(1, 4234)	=	1696.66
	Prob > F	=	0.0000
	R-squared	=	0.2859
	Root MSE	=	.76451

lnucret	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]	
egitim	.1180663	.0028663	41.19	0.000	.1124467	.1236858
_cons	8.45075	.0313017	269.98	0.000	8.389383	8.512118

Tablo 11.2 ve Tablo 11.3 karşılaştırıldığında, parametrelere ait tahmin değerleri arasında hiçbir farklılıklar bulunmamışken, standart hatalar ve t hesap değerleri arasında küçük farklılıklar gözlemlenmektedir. Fakat “**robust**” opsiyonu aykırı değerler, varsayımlardan sapmalar, normallik varsayımı ihlali, değişen varyans sorunu vb. durumlardan daha az etkilenecek olan tahmincilerin elde edilmesini sağlayacağından geri kalan analizler de bu opsiyon kullanılacaktır (Tiku vd., 1986).

Eğitim yılı değişkeni yerine eşin eğitimi araç değişken olarak modele eklenilmek istenildiğinde, öncelikle aşağıdaki model tahmin edilir:

$$\text{egitim}_i = \partial_0 + \partial_1 \text{esegitim}_i + u_i \quad (11.18)$$

Tablo 11.4. Eşin Eğitiminin Erkeğin Eğitimi Üzerindeki Etkisi**regress egitim esegitim, robust**

Linear regression	Number of obs	=	4,236
	F(1, 4234)	=	4355.45
	Prob > F	=	0.0000
	R-squared	=	0.4470
	Root MSE	=	3.0473

	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]	
egitim						
esegitim	.6397394	.0096936	66.00	0.000	.6207348	.658744
_cons	4.233339	.0878558	48.19	0.000	4.061096	4.405583

Tablo 11.4'teki sonuçlara göre erkeğin eşinin eğitim yılında meydana gelecek olan 1 yıllık artış erkeğin eğitim yılını yaklaşık olarak 0.64 yıl arttırmaktadır ve bu etki %1 anlamlılık düzeyinde istatistiksel olarak anlamlıdır. Araç değişken olarak kullanılan eşin eğitim yılı değişkeni, içsel açıklayıcı değişken olarak kabul edilen erkeğin eğitim yılı değişkeninde meydana gelen değişimlerin yaklaşık olarak %45'ini açıklamaktadır ($R^2 \cong 0.45$). Bu durumda erkeğin eğitim yılı değişkeni yerine eşinin eğitimi araç değişken olarak kullanarak aşağıdaki model tahmin edilir:

$$\ln(ucret_i) = \beta_0 + \beta_1 esegitim_i + u_i \quad (11.19)$$

Tablo 11.5. Eşin Eğitiminin Erkeğin Ücreti Üzerindeki Etkisi**regress lnucret esegitim, robust**

Linear regression	Number of obs	=	4,236
	F(1, 4234)	=	1127.87
	Prob > F	=	0.0000
	R-squared	=	0.1917
	Root MSE	=	.81341

	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]	
lnucret						
esegitim	.0925004	.0027543	33.58	0.000	.0871005	.0979003
_cons	8.830894	.026315	335.58	0.000	8.779303	8.88248

Tablo 11.5'teki sonuçlara göre eşin eğitim yılında meydana gelecek olan 1 yıllık artış, erkeğin ücreti üzerinde %9.2'lik artışa sebebiyet vermektedir. Bu etki eşitlik (11.17)'de tahmin edilen %11.8'lik etkiden daha küçüktür. (11.17)'deki ücret-eğitim modelinde erkeğin eğitim yılının içsel değişken olabileceğine dair bir ipucu elde edilir. Bir diğer ifade ile eğitim değişkeninde 11.1 alt bölümünde bahsedilen ölçme hataları veya dışlanan değişkenler gibi sebeplerden dolayı böyle bir sapma elde edilmiş olur (Wooldridge, 2013: 512). Fakat erkeğin eğitim yılı değişkeninin gerçekten içsel değişken olup olmadığı bir sonraki bölümlerde daha detaylı test edilecektir. Wooldridge (2013) benzer bir şekilde çocuğun eğitim yılı ile ücret ilişkisini tahmin ettiği modelinde çocuğun eğitim yılı yerine babanın eğitim yılını kullanmış ve benzer sonuçlar elde etmiştir. Card (1994) yaptığı bir çalışmada eğitim değişkeninin içsel olmasından dolayı özellikle anket verileri ile yapılan mikro ekonometrik çalışmalarda eşitlik (11.16)'daki EKK tahmincisi $\hat{\beta}_1$ 'de yetenek sapması (ability bias) olarak isimlendirilen yaklaşık olarak %10'luk bir sapmanın oluşacağını belirtmiştir (Dickson, 2009: 3).

11.3.2. Zayıf Araç Değişken Problemi ve Stata Uygulaması

Zayıf araç değişken kavramı özellikle içsel açıklayıcı değişken ile kullanılan araç değişken arasında negatif veya pozitif güçlü bir ilişki yoksa söz konusu olur. İlerleyen bölümlerde zayıf araç değişken kavramına çözüm olarak İki Aşamalı En Küçük Kareler Yöntemi ele alınacaktır. Literatürde aynı zamanda Kısıtlı En Küçük Kareler Yöntemi de çözüm olarak önerilmiştir (Wooldridge, 2007: 1). Eşitlik (11.19)'daki IV tahmincisinin varyansı araç değişken Z_i ile açıklayıcı değişken X_i arasında güçlü bir ilişki olmadığı zaman artmaya devam eder. Buna bağlı olarak da IV tahmincisine ait standart hatalar da devamlı artar. Araç değişken Z_i ile hata terimi u_i arasında hiç ilişki olmaması durumunda $Cov(z_i u_i)$ ifadesi sifıra eşittir ve IV tahmincisi ise tutarlıdır. Önceden bahsedildiği üzere bu özelliği sağlayan araç değişken bulmak oldukça zordur. Bu yüzden ikinci durumda eşitlik (11.20)'deki $Cov(z_i u_i)$ ifadesi sifıra çok yakın dahi olsa paydadaki ifade $Cov(z_i x_i)$ çok küçük iken IV tahmincisi $\hat{\beta}_{1(IV)}$ sapmalıdır ve tutarlılık özelliğini sağlamaz. IV ve EKK tahmincileri için sırasıyla aşağıdaki eşitlikler yazılabilir (Wooldridge, 2013: 514-515):

$$plim(\hat{\beta}_{1(IV)}) = plim\left(\frac{\sum_{i=1}^n z_i [\beta_1 x_i + (u_i - \bar{u})]}{\sum_{i=1}^n z_i x_i}\right) = \beta_1 + plim\left(\frac{Cov(z_i u_i)}{Cov(z_i x_i)}\right) \quad (11.20)$$

$$plim(\hat{\beta}_{1(SEKK)}) = plim\left(\frac{\sum_{i=1}^n x_i [\beta_1 x_i + (u_i - \bar{u})]}{\sum_{i=1}^n x_i^2}\right) = \beta_1 + plim\left(\frac{Cov(x_i u_i)}{Var(x_i)}\right) \quad (11.21)$$

Zayıf araç değişken olması durumunda yukarıdaki eşitlikler (11.20) ve (11.21) kapsamında, $\hat{\beta}_{1(IV)}$ ve $\hat{\beta}_{1(EKK)}$ 'deki sapmaların yönü hakkında birkaç farklı durum ile karşı karşıya kalınabilir. Örneğin eğer zayıf araç değişkenler problemi kapsamında $Cov(z_i x_i)$ pozitif fakat sifıra oldukça yakın; $Cov(z_i u_i)$ ve $Cov(x_i u_i)$ işaretleri pozitif ise bu durumda her iki tahminciye ait sapma yukarı doğru olur. Hangi sapmanın daha büyük olacağı ise $plim\left(\frac{Cov(z_i u_i)}{Cov(z_i x_i)}\right)$ ve $plim\left(\frac{Cov(x_i u_i)}{Var(x_i)}\right)$ değerlerine bağlıdır.

Evli erkeklerin ücreti ve eğitim yılı örneğine geri dönülürse, zayıf araç değişken olabileceği düşünülen bir araç değişken bulunarak analiz tekrarlansın:

$$\ln(ucret_i) = \beta_0 + \beta_1 egitim_i + u_i \quad (11.22)$$

Erkeğin eğitim yılı yerine veri setinde sahip olunan 0-5 yaş arası çocukların sayısı araç değişken olarak kullanılsın. Literatüre göre, eğitim seviyesi arttıkça bireyin sahip olduğu çocuk sayısının azalması beklenir. Erkeğin eğitimi üzerinde 0-5 yaş arası çocuk sayısının etkisine ait regresyon tahmini Tablo 11.6'da verilmiştir.

Tablo 11.6. 0-5 Yaş Arası Çocuk Sayısının Erkeğin Eğitimi Üzerindeki Etkisi: Zayıf Araç Değişken Problemi

regress egitim Cocsay0_5, robust

Linear regression	Number of obs	=	4,236
	F(1, 4234)	=	0.37
	Prob > F	=	0.5411
	R-squared	=	0.0001
	Root MSE	=	4.0975

	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]
egitim					
Cocsay0_5	-.0554803	.0907757	-0.61	0.541	-.2334482 .1224877
_cons	8.772914	.0777053	112.90	0.000	8.62057 8.925257

İktisadi teoriye uygun olarak 0-5 yaş arası çocuk sayısında meydana gelecek olan 1 artış erkeğin eğitim yılını 0.05 yıl azaltır. Tahmine ait p olasılık değeri 0.541 olarak elde edildiğinden, bu etkinin istatistiksel olarak anlamsız olduğu görülmektedir. Aynı zamanda modele ait R^2 değeri de sıfıra çok yakın elde edilmiştir. Bu iki değer ele alındığında 0-5 yaş arası çocuk sayısı değişkeninin eğitim değişkeninde meydana gelen değişimleri açıklamadığı ve zayıf bir araç değişken olduğu sonucuna varılır. Bu değişken, araç değişken olarak kullanılıp, eşitlik (11.22)'deki model yeniden tahmin edildiğinde:

$$\ln(ucret_i) = \beta_0 + \beta_1 (Cocsay0\_5_i) + u_i \quad (11.23)$$

Tablo 11.7. 0-5 Yaş Arası Çocuk Sayısının Erkeğin Ücreti Üzerindeki Etkisi: IV Tahmini**regress lnucrer Cocsay0_5, robust**

Linear regression	Number of obs	=	4,236
	F(1, 4234)	=	0.00
	Prob > F	=	0.9981
	R-squared	=	0.0000
	Root MSE	=	.90473

lnucrer	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]
Cocsay0_5	.0000474	.0194229	0.00	0.998	-.0380316 .0381264
_cons	9.483221	.0179875	527.21	0.000	9.447956 9.518486

Model tahmin sonucuna göre 0-5 yaş arası çocuk sayısının erkeğin aldığı ücret üzerinde anlamlı bir etkisi olmadığı görülmektedir ($p = 0.988$). Ayrıca modelde bütünüyle anlamlı değildir ($p(F_{hesap}) = 0.987$).⁴

11.4. Çoklu Regresyon Kapsamında Araç Değişkenler Yöntemi

Şimdiye kadar araç değişken kullanımı basit regresyon kapsamında açıklayıcı ve uygulamalı örnekler ile açıklanmıştır. Eşitlik (11.2)'de yer alan model genişletildiğinde ve çoklu regresyon modeli uygun bir şekilde yeniden ifade edildiğinde, iki açıklayıcı değişkenli durum için aşağıdaki model yazılır:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i \quad (11.24)$$

Bir önceki bölümde olduğu gibi X_{1i} açıklayıcı değişkeninin içsel, X_{2i} açıklayıcı değişkeninin ise dışsal açıklayıcı değişken olduğu varsayalım. Başka bir ifade ile aşağıdaki varsayımlar kabul edilsin:

$$\text{Cov}(X_{1i}u_i) \neq 0 \quad \text{ve} \quad \text{Cov}(X_{2i}u_i) = 0; \quad u_i \sim N(0, \sigma^2) \quad (11.25)$$

Buna göre X_{1i} açıklayıcı değişkeninin çeşitli sebepler ile (ölçme hatası, dışlanan değişkenler vb.) hata terimi ile ilişkili olduğu düşünülün. Bu durumda (11.24) modelinde sadece bir tane içsel açıklayıcı değişken olur ve bu değişkenin yerine kullanılabilir Z_i araç değişkenini seçmek gerekir. Ekonometrik teoride (11.24) modeli *yapısal denklem* olarak ifade edilir. Denklemdaki katsayılar ise *yapı katsayıları* olarak isimlendirilir (Gujarati ve Porter, 2009: 690). Bu durumda X_{1i} içsel açıklayıcı değişkeni yerine alt başlık (11.2)'de belirtilen varsayımları sağlayan bir araç değişken seçmek gerekir. Yeri gelmiş iken, X_{2i} 'nin modelde hâlihazırda yer alan açıklayıcı değişken olması sebebi ile içsel açıklayıcı değişken yerine araç değişken olarak seçilemeyeceğinin unutulmaması gerekmektedir (Wooldridge, 2013: 517). Bu durumda çoklu doğrusal reg-

⁴ Modelin basit doğrusal regresyon modeli kapsamında tahmin edildiği hatırlansın. Çoklu doğrusal regresyon kapsamında yaş ve benzeri kontroller modele eklendiğinde daha robust (dirençli) tahminler elde edilecektir.

resyon kapsamında, bir ikinci denklem kurarak içsel açıklayıcı değişken X_{1i} 'yi hem eşitlik (11.24)'teki dışsal açıklayıcı değişken X_{2i} hem de $Cov(Z_{1i}, u_i) = 0$ koşulunu sağlayan Z_{1i} araç değişkeni cinsinden aşağıdaki gibi yazmak mümkün olacaktır:

$$X_{1i} = \emptyset_0 + \emptyset_1 Z_{1i} + \emptyset_2 X_{2i} + v_i \quad (11.26)$$

$Cov(Z_{1i}, v_i) = 0$; $Cov(X_{2i}, v_i) = 0$; $v_i \sim N(0, \sigma_v^2)$ ve $\emptyset_1 \neq 0$ olduğu unutulmasın (Wooldridge, 2013: 518). (11.26) denklemi *indirgenmiş yapı denklemi* ve denklemdeki katsayılar ise *indirgenmiş yapısal katsayılar* olarak isimlendirilir. Asıl soru bu denklemlerin nasıl tahmin edileceğidir. Buna göre (11.26) *indirgenmiş yapısal denklemi* EKK yöntemi ile tahmin edilir. (11.26)'daki indirgenmiş yapısal katsayılarını tahmin ederek (11.24)'teki bütün katsayılar tahmin edebiliyorsa bu durumda (11.24) denkleminin ***tam belirlenmiş*** olduğu söylenir. Eğer yapısal katsayılar indirgenmiş yapısal katsayı tahminlerinden elde edilemiyorsa ***belirlenmemiş*** veya ***eksik belirlenmiş denklem***, yapısal katsayılar için indirgenmiş katsayı tahminleri aracılığı ile birden fazla tahmin veriyorsa ***aşırı belirlenmiş denklem*** tanımlamaları kullanılır. Hem (11.24) hem de (11.26) denklemlerine iki kez EKK yöntemi uygulanmış olacağından bu yönteme İki Aşamalı En Küçük Kareler (2AEKK)⁵ yöntemi adı verilmektedir. 11.4.1. Bölümde 2AEKK yöntemi daha detaylı sayısal bir örnek üzerinden incelenecektir.

11.4.1. İki Aşamalı En Küçük Kareler Yöntemi (2AEKK)

Yöntemin aşamaları tekrar ele alınsın:

1. Aşama: Birinci aşamada içsel açıklayıcı değişken olduğu düşünülen X_1 açıklayıcı değişkeni açıklanan değişken olarak kullanarak, Z_1 araç değişkeni ve yapısal modelde yer alan diğer dışsal açıklayıcı değişken/değişkenleri ise açıklayıcı değişkenler olarak ele alınıp, indirgenmiş yapısal regresyon modeli EKK yöntemiyle tahmin edilir.
2. Aşama: 1. aşamadan elde edilen $\hat{X}_1$ kullanılarak (11.21) yapısal denkleminde X_1 yerine yazılır.
3. Aşama: Bu aşamada ise yapısal denklem (11.21) EKK yöntemi kullanılarak tahmin edilir ve böylelikle yapısal denklem katsayıları da tahmin edilmiş olur.

Kısaca bu araç değişken yöntemine 2AEKK yöntemi adı verilir. Tam belirlenme durumu için sayısal bir örnek üzerinden 2AEKK yöntemi incelensin.

11.4.2. İki Aşamalı En Küçük Kareler Yöntemi (2AEKK) Stata Uygulaması

Eğitim-ücret örneği tekrar ele alınsın. TÜİK 2013 Hanehalkı Bütçe Anketi'ni kullanarak bireyin eğitim yılı ile yıllık ücret bilgisi verileri 2AEKK yöntemi ile incelensin.

Bireylerin yıllık ücreti ve eğitim yılları değişkenleri ile bireye ilişkin veri setinde mevcut demografik değişkenler kullanılarak aşağıdaki çoklu regresyon modeli ele alınsın:

$$\ln(ucret)_i = \tau_0 + \tau_1 eğitim_i + \tau_2 yas_i + \tau_3 kent_i + u_i, \quad i = 1, \dots, n \quad (11.27)$$

⁵ 2EKK Yöntemi İngilizce karşılığı Two Stage Least Squares (2SLS)'dir.

$egitim_i$ eğitim yılını, yas_i bireyin yaşını, $kent_i$ kukla değişkeni bireyin kentsel veya kırsal bölgede yaşayıp yaşamadığını göstermektedir⁶. Bir önceki bölümde evli çalışan erkeklerde eğitim yılı değişkeninin içsel değişken olması ihtimali tartışılmıştı. Yetenek değişkeni gibi model dışında kalan değişkenler eğitim yılı değişkenini etkiler, fakat yetenek değişkeninin doğrudan ölçülemeyen bir değişken olması sebebi ile temsilen farklı bir araç değişken bulunmalı ve yukarıdaki (11.27) yapısal denklemi 2AEKK yöntemi ile tahmin edilmelidir. Bu esnada modelde yer alan yaş ve bireyin sahip olduğu çocuk sayısı değişkenlerinin ise hata terimi u_i ile ilişkili olmadığı düşünülün. Eğitim yılı değişkeninin ise gerçekte içsel olup olmadığının istatistiksel açıdan nasıl test edileceği bir sonraki bölümde detayları ile incelenecektir. (11.27) modeli için eğitim yılı değişkeni yerine UCRET.dta veri setinde yer alan eşin eğitim yılı değişkeni araç değişken olarak kullanılsın. Bu durumda indirgenmiş yapısal denklemi aşağıdaki gibi yazılır:

$$egitim_i = \varphi_0 + \varphi_1 eegitim_i + \varphi_2 yas_i + \varphi_3 kent_i + v_i \quad (11.28)$$

Öncelikli olarak kıyaslama yapabilmek adına (11.27) denklemini EKK yöntemi ile UCRET.dta veri seti kullanılarak tahmin edildin⁷:

Tablo 11.8. Eğitimin Erkeğin Ücreti Üzerindeki Etkisi

regress lnucret egitim kent yas, robust

Linear regression				Number of obs	=	4,236
				F(3, 4232)	=	600.98
				Prob > F	=	0.0000
				R-squared	=	0.3238
				Root MSE	=	.74414

lnucret		Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]
-----+-----						
egitim		.1054315	.0027431	38.43	0.000	.1000535 .1108095
kent		.3972291	.0343386	11.57	0.000	.3299074 .4645508
yas		-.0074899	.0015684	-4.78	0.000	-.0105647 -.004415
_cons		8.556262	.0738809	115.81	0.000	8.411417 8.701108

Tablo 11.8’de, (11.27) modelinin EKK tahmin sonuçları yer almaktadır. Buna göre eğitim yılında meydana gelecek olan 1 yıllık artış (yaş ve kent değişkenleri sabitken) bireyin yıllık ücretini yaklaşık olarak %10.5 oranında arttırmaktadır ve bu artış %1 anlam düzeyinde istatistiksel olarak anlamlıdır ($p_{\hat{\tau}_1} < 0.01$). Diğer değişkenlerin etkileri sabit iken, buna karşılık erkeğin yaşında meydana gelen 1 yıllık artış aynı anlam seviyesi için bireyin ücretini %0.7 oranında azaltmaktadır. Eğitim yılı ve yaş değişkenleri sabit iken, kentsel alanda yaşayan

⁶ $kent_i = 1$ eğer birey kentsel alanda yaşıyorsa ve $kent_i = 0$ kırsal alanda yaşıyorsa.

bireylerin kırsal alanda yaşayan bireylere kıyasla daha yüksek ücret aldığı sonucuna varılır. Kentsel yerlerde yaşayanların ücretleri üzerindeki etkisi $\hat{\tau}_3 = 0.397$ olarak bulunmuştur. Bu değişken kukla olduğundan, değişkenin ücretler üzerindeki etkisi log-doğrusal model için yarı esneklik üzerinden yorumlanır. 0.397 katsayısının anti-logaritması hesaplanır ($\text{antilog}(0.397) = 1.49$) ve hesaplanan değerden 1 değeri çıkarılır. $1.49 - 1 = 0.49$ (Gujarati, 2015: 459). 0.49 değeri de 100 ile çarpıldığında %49 değeri elde edilir. Kentsel bölgelerde yaşayan evli erkekler, kırsal bölgelerde yaşayan evli erkeklere kıyasla yaklaşık olarak %49 oranında daha yüksek ücret almaktadır.

Daha sonra (11.28)'deki indirgenmiş yapısal model tahmin edilir ve eğitim-hat (tahmin edilmiş eğitim değişkeni) değerleri hesaplanır.

Tablo 11.9. İndirgenmiş Yapısal Model Tahmini: 2AEKK 1. Aşaması

```
regress egitim esegitim yas kent, robust
quietly predict egitimhat
```

Linear regression	Number of obs	=	4,236
	F(3, 4232)	=	1474.53
	Prob > F	=	0.0000
	R-squared	=	0.4502
	Root MSE	=	3.0391

	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]	
egitim						
esegitim	.6246936	.0107048	58.36	0.000	.6037065	.6456807
yas	.0014895	.0056039	0.27	0.790	-.009497	.0124761
kent	.5885716	.1159876	5.07	0.000	.361175	.8159681
_cons	3.817541	.2762023	13.82	0.000	3.276039	4.359042

Tablo 11.9'daki Stata çıktısı indirgenmiş yapısal modele ilişkin tahmin sonuçlarını göstermektedir. Sonrasında ise yine (11.28) modeli kullanılarak “**predict**” komutu ile $\widehat{egitim}_i$ ($egitim_i - hat$) değişkeni tahmin edilmiştir.

Tablo 11.10. Eğitimin Erkeğin Ücreti Üzerindeki Etkisi: 2AEKK 2. Aşaması

```
regress lnucret egitimhat yas kent, robust
```

Linear regression	Number of obs	=	4,236
	F(3, 4232)	=	395.76
	Prob > F	=	0.0000
	R-squared	=	0.2307
	Root MSE	=	.79373

		Robust				
lnucret	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
egitimhat	.1240555	.0045262	27.41	0.000	.1151818	.1329292
yas	-.0061955	.0017237	-3.59	0.000	-.0095749	-.0028161
kent	.3551986	.0379005	9.37	0.000	.2808937	.4295035
_cons	8.373408	.0875013	95.69	0.000	8.201859	8.544956

Tablo 11.10'da yer alan tahmin sonuçları Stata programında doğrudan “**ivregress**” komutu kullanılarak da tahmin edilebilir. Aşağıda yer alan Stata komutunda içsel olduğu test sonucunda karar verilen açıklayıcı değişken yerine “**varlist2**” yazılacaktır. Bu durumda evli erkekler için eğitimin ücretlere göre getirisi modelinde “**varlist2**” yerine “**egitim**”, “**varlist_iv**” yerine ise dışsal olduğu düşünülen ve eğitim yılı ile korelasyonlu olduğu varsayılan eşin eğitim değişkeni “**esegitim**” kullanılır.⁸ 2AEKK yöntemi uygulandığında “**estimator**” yerine de “**2sls**” yazılır.

ivregress estimator depvar [varlist1] (varlist2 = varlist_iv) [if] [in] [weight] [, options]

Tablo 11.11. Eğitimin Erkeğin Ücreti Üzerindeki Etkisi: 2AEKK (ivregress)

ivregress 2sls lnucret yas kent (egitim = esegitim), robust

Instrumental variables (2SLS) regression	Number of obs	=	4,236
	Wald chi2(3)	=	1339.44
	Prob > chi2	=	0.0000
	R-squared	=	0.3173
	Root MSE	=	.74737

		Robust				
lnucret	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
egitim	.1240555	.0042714	29.04	0.000	.1156837	.1324273
yas	-.0061955	.0016122	-3.84	0.000	-.0093553	-.0030357
kent	.3551986	.0355147	10.00	0.000	.2855911	.4248061
_cons	8.373408	.0834909	100.29	0.000	8.209768	8.537047

Instrumented: **egitim**

Instruments: **yas kent esegitim**

Tablo 11.11'e göre erkeğin eğitim yılında meydana gelecek olan 1 yıllık artış (yaş ve kent değişkenleri sabitken) erkeğin yıllık ücretini yaklaşık olarak %12 oranında arttırmaktadır ve bu artış %1 anlam düzeyinde istatistiksel olarak anlamlıdır ($p < 0.01$). Diğer değişkenlerin etkileri sabitken, bu-

⁸ Stata uygulamalarında kullanılan “**egitim**” değişkeninin içsel olup olmadığının istatistiksel olarak nasıl test edileceği 11.5 başlığı altında açıklanmıştır. Şimdilik bu değişkenin uygun bir araç değişken olduğu düşünülsün.

na karşılık erkeğin yaşında meydana gelen 1 yıllık artış aynı anlam seviyesi için erkeğin ücretini %6 oranında azaltmaktadır. Eğitim yılı ve yaş değişkenleri sabit iken kentsel alanda yaşayan bireyler, kırsal alanda yaşayan bireylere kıyasla daha yüksek ücret aldığı belirlenmiştir. Tablo 11.10 ve Tablo 11.11'deki sonuçlar incelendiğinde her iki tabloda da elde edilen katsayı tahminlerinin aynı olduğu görülmektedir. Fakat tahminlere ait standart hatalar farklılık göstermektedir. Tablo 11.11'de elde edilen tahminlere ait standart hataların, Tablo 11.10'da elde edilen standart hatalardan küçük olduğu görülmektedir. “**regress**” komutu kullanarak 2 aşamada gerçekleştirilen IV tahmininde 2. aşamada “**egitim**” yerine tahmin değerleri “**egitimhat**” kullanıldığı için iki tabloda yer alan tahmin edilmiş katsayılar ait standart hatalar farklı elde edilmiştir. Tablo 11.10'da 2. aşama tahmini sırasında elde edilen standart hataların yanlış olduklarını da belirtmek gerekecektir (Gujarati, 2015: 460; Yamano, 2009: 8). Bunun altında yatan sebep, içsel olduğu düşünülen “**egitim**” değişkeninin aslında “**egitimhat**” ve (11.28)'de verilen indirgenmiş yapısal denklemde yer alan “ $\hat{v}_i$ ” kalıntılarından oluşuyor olmasıdır (Maddala, 2001: 374).

“**ivregress**” komutunu kullanarak ilk aşamada gerçekleşen tahmin sonuçlarını da görebilmek amacı ile “**robust**” opsiyonundan hemen önce “**first**” yazılarak aşağıdaki Stata komutu kullanılabilir (Tablo 11.12):

Tablo 11.12. Eğitimin Erkeğin Ücreti Üzerindeki Etkisi: 2AEKK (ivregress)
(Sırasıyla 1. ve 2. Aşamalar)

ivregress 2sls lnucret yas kent (egitim = esegitim), first robust

First-stage regressions

				Number of obs	=	4,236
				F(3, 4232)	=	1474.53
				Prob > F	=	0.0000
				R-squared	=	0.4502
				Adj R-squared	=	0.4498
				Root MSE	=	3.0391
		Robust				
egitim		Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
yas		.0014895	.0056039	0.27	0.790	-.009497 .0124761
kent		.5885716	.1159876	5.07	0.000	.361175 .8159681
esegitim		.6246936	.0107048	58.36	0.000	.6037065 .6456807
_cons		3.817541	.2762023	13.82	0.000	3.276039 4.359042

Instrumental variables (2SLS) regression

Number of obs	=	4,236
Wald chi2(3)	=	1339.44
Prob > chi2	=	0.0000
R-squared	=	0.3173
Root MSE	=	.74737

		Coef.	Robust Std. Err.	z	P> z	[95% Conf. Interval]	
lnucret							
egitim		.1240555	.0042714	29.04	0.000	.1156837	.1324273
yas		-.0061955	.0016122	-3.84	0.000	-.0093553	-.0030357
kent		.3551986	.0355147	10.00	0.000	.2855911	.4248061
_cons		8.373408	.0834909	100.29	0.000	8.209768	8.537047

Instrumented: egitim

Instruments: yas kent esegitim

“ivregress” komutunun dışında Baum, Schaefer ve Stillman (2002) tarafından yazılan “ivreg2” Stata komutu da kullanılabilir⁹.

ivreg2 depvar [varlist1] (varlist2=instlist) [if] [in] [, options]

Öncelikle “ivreg2” komutunun Stata programına indirmek gerekmektedir. “depvar” modeldeki bağımlı değişken, “varlist1” modelde yer alan dışsal olduğu düşünülen açıklayıcı değişkenler, “varlist2” ise modeldeki içsel açıklayıcı değişken ve “instlist” ise içsel açıklayıcı değişken yerine kullanılabilecek araç değişkenleri göstermektedir.¹⁰ “ivreg2” ve “ivreg” komutları arasında çok büyük farklılık bulunmamakla birlikte “ivreg2” komutu sonrasında Stata tarafından otomatik olarak bazı istatistikler raporlanmaktadır¹¹.

Tablo 11.13. Eğitimin Erkeğin Ücreti Üzerindeki Etkisi: 2AEKK (ivreg2)

ssc install ivreg2

ivreg2 lnucret yas kent (egitim = esegitim), robust

IV (2SLS) regression with robust standard errors

		Number of obs =	4236
		F(3, 4232) =	446.06
		Prob > F =	0.0000
Total (centered) SS	=	Centered R2	= 0.3173
Total (uncentered) SS	=	Uncentered R2	= 0.9938
Residual SS	=	Root MSE	= .7474

		Coef.	Robust Std. Err.	z	P> z	[95% Conf. Interval]	
lnucret							
egitim		.1240555	.0042714	29.04	0.000	.1156837	.1324273
yas		-.0061955	.0016122	-3.84	0.000	-.0093553	-.0030357

⁹ 11. Bölüm’de araç değişkenler konusu “ivregress” komutu kullanılarak işlenmektedir. Sebebi ise komutun Stata tarafından tamamen tüm alt komutları ile birlikte destekleniyor oluşudur (Cameron ve Trivedi, 2009).

¹⁰ Stata 9.2 ve daha üzeri sürümlerinde çalışan bir komuttur (Baum, 2008).

¹¹ “ivreg2” komutuna dair daha fazla bilgi için Stata Help menüsünü inceleyebilirsiniz.

kent	.3551986	.0355147	10.00	0.000	.2855911	.4248061
_cons	8.373408	.0834909	100.29	0.000	8.209768	8.537047

Anderson canon. corr. LR statistic (identification/IV relevance test):2169.432
Chi-sq(1) P-val = 0.0000

Hansen J statistic (overidentification test of all instruments): 0.000
(equation exactly identified)

Instrumented: egitim
Included instruments: yas kent
Excluded instruments: esegitim

Tablo 11.13'teki sonuçlar ile Tablo 11.12'deki sonuçlar arasında bir farklılık olmadığı görünmektedir. Tablo 11.12'de, Tablo 11.11'den farklı olarak bazı istatistiksel testler de yer almaktadır (Anderson CC LR ve Hansen J testleri). Anderson CC LR testi kullanılan *esegitim_i* araç değişkenin içsel değişken olan *egitim_i* değişkenlerine ait X ve Z matrisleri için geçerli olacak kanonik korelasyonun 0 olup olmadığı hususunda bilgi verir (Baum, 2008: 37). Teste ilişkin hesaplanan χ^2 değerine ait olasılık 0 olarak bulunduğu için minimum kanonik korelasyonun sıfıra eşit olduğunu savunan sıfır hipotezi reddedilir. Aksi durumda H_0 hipotezinin kabul edilmesi modelin eksik belirlendiği anlamına gelir. Hansen J istatistiği ise tahmin edilen regresyon modelinin aşırı belirlenip belirlenmediğine ilişkin bilgi veren bir testtir. Modelde kullanılan araç değişken sayısı, içsel açıklayıcı değişken sayısından daha fazla ise modelin aşırı belirlendiği söylenir. Bu durumda yukarıdaki Hansen J istatistiğinin sıfır olarak bulunması modelin tam belirlendiği anlamına gelir. Aşırı belirlenme durumu için ise J istatistiğinin pozitif olması beklenir (Baum, 2008: 27). Aşırı belirlenme sorunu 11.8 başlığı altında ele alınacaktır.

11.5. Açıklayıcı Değişkenler için İçsellik Sınaması ve Stata Uygulaması

Bir önceki bölümde gerçekleştirilen model tahminlerinde literatürde sıkça yer bulan eğitimin gelire göre getirisi teorisi kullanılarak eğitim değişkeninin içsel olduğu varsayılmıştı. Fakat gerçekte eğitim değişkeni içsel midir? Bu bölümde içsel olmasından şüphelenilen açıklayıcı değişken veya değişkenlerin içsellik sınamaları ele alınacaktır. Hausman (1978) testi kullanılarak eğitim değişkeninin içselliği istatistiksel olarak sınanabilir. Hausman (1978) testini uygulayabilmek için öncelikle içsel olduğundan şüphelenilen açıklayıcı değişkenin regresyon modelinde yer alan dışsal açıklayıcı değişkenler ile yerine kullanılabileceği düşünülen araç değişken veya değişkenlere göre regresyonu tahmin edilip, bu tahmine ilişkin olarak kalıntılar hesaplanır (shat veya $\hat{u}$, Tablo 11.14). Bir sonraki aşamada ise hesaplanan kalıntıların da açıklayıcı değişken olarak eklendiği ana regresyon modeli tahmin edilir (örnekte evli erkeklerin yıllık ücretlerinin logaritmasının bağımlı değişken olduğu denklem (11.27)). Tahmin sonrasında kalıntılara ait (shat) katsayı tahmini istatistiksel olarak anlamlı ise bu durumda içsel olmasından şüphelenilen açıklayıcı değişkenin gerçekten içsel olduğuna karar verilir (Tablo 11.15).

Tablo 11.14. Eğitimin Erkeğin Ücreti Üzerindeki Etkisi: Eğitim Değişkeni için Hausman Testi (1. Aşama)

regress egitim yas kent esegitim, robust

Linear regression	Number of obs	=	4,236
	F(3, 4232)	=	1474.53
	Prob > F	=	0.0000
	R-squared	=	0.4502
	Root MSE	=	3.0391

		Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]	
egitim							
yas		.0014895	.0056039	0.27	0.790	-.009497	.0124761
kent		.5885716	.1159876	5.07	0.000	.361175	.8159681
esegitim		.6246936	.0107048	58.36	0.000	.6037065	.6456807
_cons		3.817541	.2762023	13.82	0.000	3.276039	4.359042

Elde edilen regresyon sonucundan sonra Tablo 11.14'teki model için aşağıdaki Stata komutu kullanılarak kalıntılar hesaplanır:

quietly predict shat

Tablo 11.15. Eğitimin Erkeğin Ücreti Üzerindeki Etkisi: Eğitim Değişkeni için Hausman Testi (2. Aşama)

regress lnucret egitim yas kent shat, robust

Linear regression	Number of obs	=	4,236
	F(4, 4231)	=	473.93
	Prob > F	=	0.0000
	R-squared	=	0.3282
	Root MSE	=	.74182

		Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]	
lnucret							
egitim		.0929746	.0035839	25.94	0.000	.0859483	.1000008
yas		-.0061955	.0016138	-3.84	0.000	-.0093593	-.0030317
kent		.3551986	.0355686	9.99	0.000	.2854656	.4249316
shat		.0310809	.0055706	5.58	0.000	.0201595	.0420023
_cons		8.373408	.0828987	101.01	0.000	8.210883	8.535933

Tablo 11.15'e göre kalıntılara ($\hat{u}_i$) ait katsayının %1 anlamlılık düzeyinde istatistiksel olarak anlamlı olduğu görünmektedir ($p_{shat} < 0.01$). Buna göre eğitim yılı değişkeninin içsel olduğu sonucuna varılır.

İlgili açıklayıcı değişkeninin içsellik testi için bir diğer opsiyon ise “**ivregress**” komutundan hemen sonra kullanılan “**estat endogenous**” komutudur. Bu komut Robustified Durbin Wu-Hausman testi sonucunu verir (Cameron ve Trivedi, 2009:184). (11.27) modeli üzerinden incelenen bu modelin test sonuçları Tablo 11.16'da verilmiştir.

Tablo 11.16. Eğitimin Gelire Göre Getirisi: Eğitim Değişkeni için Robustified Durbin Wu-Hausman Testi

```
quietly ivregress 2sls lnucrct yas kent (egitim = esegitim), robust
estat endogenous
```

Tests of endogeneity

Ho: variables are exogenous

Robust score chi2(1) = 30.2774 (p = 0.0000)

Robust regression F(1,4231) = 31.1298 (p = 0.0000)

Tablo 11.16'daki test istatistiğine göre (Robust regression F(1,4231)) içsel olduğu düşünülen eğitim değişkeni için değişkenin dışsal olduğunu savunan H_0 hipotezi %99 güven düzeyinde reddedilmiştir. Bu sonuçlara göre hem Hausman (1978) hem de Durbin Wu-Hausman testleri tutarlı sonuçlar vermiştir.

11.6. Zayıf Araç Değişkenler için İstatistiksel Testler ve Stata Uygulaması

Önceki bölümlerde bir araç değişkenin hangi şartlarda zayıf veya güçlü araç değişken olarak isimlendirildiklerinden bahsedilmişti. “**ivregress**” komutu sonrası kullanılabilen “**estat firststage**” komutu ile araç değişken veya değişkenlerin “zayıf araç değişken” olup olmadıkları test edilebilir.

“**estat firststage**” komutundan sonra “**forcenonrobust**” opsiyonu dirençli hataların (**vce(robust)**) kullanıldığı 2AEKK yöntemi için de kullanılabilir (Stata Help Menüsü, 2020). “**forceonrobust**” komutu 2AEKK tahmininde elde edilen regresyon hatalarının Gaussian dağılımına sahip, bağımsız ve özdeş (i.i.d.) dağılmış olduklarını kabul etmekte olup buna rağmen her koşulda (değişen varyans söz konusu olsa bile) bu zayıf araç değişkenlere ilişkin testin yapılmasının istendiği manasına gelmektedir (Cameron ve Trivedi, 2009:191). “all” opsiyonu ise 2AEKK yönteminin uygulandığı modelde birden fazla içsel değişken olması durumunda tüm aşamaların her bir içsel değişken için ayrı ayrı yapılmasını sağlar (Tablo 11.17).

Tablo 11.17. Eğitimin Erkeğin Ücreti Üzerindeki Etkisi: Eğitim Değişkeni için Zayıf Araç Değişken Testi

```
quietly ivregress 2sls lnucrct yas kent (egitim = esegitim), robust
```

```
estat firststage, forcenonrobust all
```

First-stage regression summary statistics

Variable	R-sq.	Adjusted R-sq.	Partial R-sq.	Robust F(1,4232)	Prob > F
egitim	0.4502	0.4498	0.4008	3405.44	0.0000

Shea's partial R-squared

Variable	Shea's Partial R-sq.	Shea's Adj. Partial R-sq.
egitim	0.4008	0.4005

Minimum eigenvalue statistic = 2830.62

Critical Values	# of endogenous regressors:	1		
Ho: Instruments are weak	# of excluded instruments:	1		

	5%	10%	20%	30%
2SLS relative bias	(not available)			

	10%	15%	20%	25%
2SLS Size of nominal 5% Wald test	16.38	8.96	6.66	5.53
LIML Size of nominal 5% Wald test	16.38	8.96	6.66	5.53

Tablo 11.17'nin ilk bölümü 2AEKK yönteminin ilk aşamasına ait R^2 , $\bar{R}^2$ (düzeltilmiş R^2) ve kısmi (partial) R^2 'ye ilişkin bilgi vermektedir¹². Bu üç değer de çalışılan araç değişkenlerin içsel açıklayıcı değişken veya değişkenler ile yeterli düzeyde korelasyonlu olup olmadıklarına dair bilgi veren unsurlardır. Modelde tek bir içsel açıklayıcı değişken olduğunda R^2 ve kısmi R^2 birbirlerine yakın değerler alır (Drichoutis vd., 2009: 7). Bu açıklamaya uygun olarak Tablo 11.17'de R^2 (%45) ile kısmi R^2 (%40) değerlerinin birbirlerine yakın sayılacak değerler aldıkları görülmektedir. İncelenen veri setinin yatay kesit verisi olması dolayısıyla elde edilen kısmi R^2 değerinin %40 olarak bulunması *esegitim_i* değişkeni ile *egitim_i* değişkeninin yeterli düzeyde korelasyonlu olduğunu göstermektedir.

¹² Kısmi R^2 değeri 2AEKK'nın ilk aşaması için modelde yer alan diğer değişkenlerin (araç değişken hariç) etkileri sabitken, içsel açıklayıcı değişken ile söz konusu Z araç değişkenine ait kısmi korelasyonun karesi olarak ifade edilebilir (Bound vd., 1995: 444). Dolayısıyla R^2 değerine göre daha objektif bir ölçüt olacaktır.

Eğer incelenen modelde birden fazla içsel açıklayıcı değişken var ise birinci aşamada elde edilen kısmi R^2 değerine bakarak Z araç değişkeninin içsel açıklayıcı değişkenin ne kadarını açıkladığını söylemek zor olacağından, araçlar arasındaki karşılıklı korelasyonları dikkate alan Shea'nin kısmi R^2 (Shea's partial R-squared) değerinin incelenmesi gerekir (Shea, 1997; Baum vd., 2003: 15). Tablo 11.17'de bu değer yaklaşık olarak %40 bulunmuştur. Bununla birlikte birinci aşamada elde edilen kısmi R^2 değeri ile eşit bir değer elde edilmiştir. (11.27) modelinde sadece bir tane içsel açıklayıcı değişken ($egitim_i$) ve bir tane de araç değişken ($esegitim_i$) oluşu bu durumu açıklamaktadır. Bunların dışında Tablo 11.17'nin ikinci kısmında Robust F değeri de yer almaktadır. Bu değer modelde kullanılan araç değişkenlerin aynı anda sıfıra eşit olup olmadıklarına ilişkin bilgi veren bir istatistiktir. Tablo 11.17'de incelenen modelde bilindiği üzere tek bir araç değişken yer aldığından bu değer $esegitim_i$ araç değişkenine ait bir değer olmakla birlikte araç değişkenin zayıf bir araç değişken olup olmadığına dair bilgi verir. Pratik bir yöntem olarak eğer birinci aşamadan elde edilen bu değer yaklaşık olarak 10'dan daha büyük ise bu durumda araç değişkenin zayıf bir araç değişken olduğunu savunan sıfır hipotezi reddedilir (Staiger ve Stock, 1997; Andrews vd., 2019). Örnekte F değeri (Robust, F) oldukça yüksek elde edilmiş olduğundan $esegitim_i$ değişkeninin zayıf araç değişken olmadığı açıkça görülmektedir.

Üçüncü kısım ise Stock ve Yogo (2005) testleri olarak isimlendirilen testler ile ilgili bilgi vermektedir. Stock ve Yogo testlerinin ilkinde ait sonuçlar Tablo 11.17'de mevcut değildir. Bunun sebebi modelin tam belirlenmiş olmasıdır. Tablo 11.17'de sadece Stock ve Yogo testlerinin ikincisine ait sonuçlar görülmektedir. Minimum özdeğer istatistiği 2830.62 olarak bulunmuştur. Bunun anlamı varsayılan standart hatalar kullanıldığı zaman ("**robust**" opsiyonu olmadan), 1. aşama sonrasında $esegitim_i = 0$ testi ile elde edilecek olan F istatistik değeri olmasıdır (Cameron ve Trivedi, 2009: 193). Tablo 11.17'deki örnekte "**robust**" opsiyonunu kullanılarak dirençli standart hatalar elde edilmiş ve birinci aşamada sonrasında robust F değeri ise 3405.44 olarak bulunmuştur. Bu değer ise Tablo 11.17'de üçüncü kısımda elde edilen 16.38 kritik değerinden oldukça büyüktür. Sonuç olarak H_0 : $esegitim_i$ değişkeni zayıf araç değişkendir ifadesini savunan sıfır hipotezinin reddedilmesi uygun olacaktır yorumu yapılabilir.

11.7. Genelleştirilmiş Momentler Yöntemi (GMM) ve Stata Uygulaması

IV tahmini için 2AEKK dışında kullanılan yöntemlerden bir diğeri ise Genelleştirilmiş Momentler Yöntemi (GMM)'dir. Bir diğer ifade ile IV tahmini GMM'nin özel bir durumu olarak da kabul edilmektedir (Kaplan ve Bulut, 2011: 48). Matris notasyonu kullanarak IV Tahmini kapsamında Cameron ve Trivedi (2005: 172) GMM metodunu aşağıdaki gibi ele almışlardır:

$$y_i = X_i\beta + u_i \quad (11.29)$$

Yukarıdaki modeli GMM yöntemi ile tahmin etmek için öncelikle seçilecek olan araç değişkenlerin (Z_i), $Cov(Z_i, u_i) = 0$ ve $Cov(Z_i, X_i) = 0$ varsayımlarını sağlamaları gerekir. y_i bağımlı değişken vektörü, n adet gözlem için X_i açıklayıcı değişken matrisi (nxm) boyutunda ve Z_i araç değişken matrisi ise (nxs) boyutunda olmak kaydı ile örnek moment koşulu (diklik koşulu) aşağıdaki gibi yazılabilir:

$$\frac{1}{n} (\sum_{i=1}^n Z_i(Y_i - X_i\hat{\beta})) = 0 \quad (11.30)$$

Tam belirlenme durumu söz konusu ise ($m=s$) bu durumda elde edilecek olan tahminci doğrusal IV tahmincisi olacaktır.

$$\hat{\beta}_{GMM} = (\sum_{i=1}^n Z_i X_i)^{-1} (\sum_{i=1}^n Z_i Y_i) = \hat{\beta}_{IV} \text{ olacaktır.} \quad (11.31)$$

Eğer aşırı belirlenme durumu söz konusu ise de ($s>m$) bu durumda $\hat{\beta}$ 'da s adet bilinmeyenden daha fazla denklem veya moment koşulu oluşacaktır. s adet araç değişken kullanımı çözüm olarak görülse de tahmincilerde etkinlik kaybı oluşacaktır. Cameron ve Trivedi (2005) moment koşullarında karesel (kuadratik) bir yapı oluşturacak şekilde $s \times s$ boyutunda W_i ağırlık matrisini kullanarak aşağıdaki formülü minimize edecek olan $\hat{\beta}_{GMM}$ tahmincileri bulunabileceğini açıklamışlardır:

$$q_n(\hat{\beta}) = \frac{1}{n} (\sum_{i=1}^n Z_i (Y_i - X_i' \hat{\beta}))' W_i \frac{1}{n} (\sum_{i=1}^n Z_i (Y_i - X_i' \hat{\beta})) \quad (11.32)$$

Yukarıda bahsi geçen ağırlık matrisinin boyut tercihi araştırmacıya bağlıdır. Bu yüzden 2AEKK yöntemine göre daha etkin tahmincilerin elde edildiği bir yöntem olduğundan söz etmek mümkündür.

Stata programı için 2AEKK yönteminde olduğu gibi aşağıdaki genel komut kullanılır:

ivregress estimator depvar [varlist1] (varlist2 = varlist_iv) [if] [in] [weight] [, options]

“**estimator**” opsiyonu için “**gmm**” yazılarak IV tahmini yapılır. GMM tahmini için aşırı belirlenmiş olan modellerde “**robust**” opsiyonu yerine ağırlık matrisi olan W_i matrisinin türünü belirlemeye yarayan **wmatrix(wmtype)** ve etkin GMM tahmincileri elde edilmesini belirleyecek olan “**robust**” opsiyonu kullanılır (Cameron ve Trivedi, 2009:181). Eğitimin ücretlere göre getirisini örneğinde tam belirlenme durumu için $eşegitim_i$ değişkeninin araç değişken olduğu aşağıdaki model GMM yöntemi ile tahmin edilsin:

$$\ln(ucret)_i = \tau_0 + \tau_1 egitim_i + \tau_2 yas_i + \tau_3 kent_i + u_i \quad (11.33)$$

Tablo 11.18. Eğitimin Erkeğin Ücreti Üzerindeki Etkisi: 2AEKK ve GMM Yöntemleri

Panel A: 2AEKK Tahmini

ivregress 2sls lnucret yas kent (egitim = esegitim), robust

Instrumental variables (2SLS) regression	Number of obs	=	4,236
	Wald chi2(3)	=	1339.44
	Prob > chi2	=	0.0000
	R-squared	=	0.3173
	Root MSE	=	.74737

		Robust				
lnucret	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
egitim	.1240555	.0042714	29.04	0.000	.1156837	.1324273
yas	-.0061955	.0016122	-3.84	0.000	-.0093553	-.0030357

kent	.3551986	.0355147	10.00	0.000	.2855911	.4248061
_cons	8.373408	.0834909	100.29	0.000	8.209768	8.537047

Instrumented: egitim

Instruments: yas kent esegitim

Panel B: GMM Tahmini

`ivregress gmm lnucret yas kent (egitim = esegitim), wmatrix(robust)`

Instrumental variables (GMM) regression	Number of obs	=	4,236
	Wald chi2(3)	=	1339.44
	Prob > chi2	=	0.0000
	R-squared	=	0.3173
GMM weight matrix: Robust	Root MSE	=	.74737

lnucret	Coef.	Robust Std. Err.	z	P> z	[95% Conf. Interval]
egitim	.1240555	.0042714	29.04	0.000	.1156837 .1324273
yas	-.0061955	.0016122	-3.84	0.000	-.0093553 -.0030357
kent	.3551986	.0355147	10.00	0.000	.2855911 .4248061
_cons	8.373408	.0834909	100.29	0.000	8.209768 8.537047

Instrumented: egitim

Instruments: yas kent esegitim

Tablo 11.18 hem GMM hem de 2AEKK yöntemi kullanılarak (11.33) modeli için tahmin sonuçlarını göstermektedir. Panel A de 2AEKK tahminine göre bireyin eğitime ait katsayı tahmini 0.124 olarak bulunmuştur. Bunun anlamı eşin eğitimi değişkeni araç değişken olarak kullanıldığında bireyin eğitim yılında meydana gelen 1 yıllık artışın (diğer değişkenler sabitken) ücretleri %12 oranında arttırmasıdır. Bu tahmine ait standart hata değeri ise yaklaşık olarak 0.004 olarak hesaplanmıştır. GMM yönteminde ise bu katsayı tahmini 0.124 olarak hesaplanmıştır (Tablo 11.18, Panel B). Sonuçlardan görüldüğü üzere iki yöntemden elde edilen tahmin sonuçları ve standart hatalar birbirine çok yakındır.

11.8. Belirlenme Kısıtlarına İlişkin Testler

Bölüm 11.3'te bir regresyon modelinin belirlenme kısıtları ele alınmıştır. Bu kısıtları özetlemek gerekirse (Gujarati, 2015:467):

- Eğer regresyon modelinde yer alan içsel açıklayıcı değişken sayısı (m) araç değişken sayısına (s) eşit ise regresyon modeli tam belirlenmiştir denir ($m = s$). Bir regresyon denkleminin belirlenmesi için gerekli koşula literatürde rank koşulu adı da verilir.
- Eğer içsel açıklayıcı değişken sayısı (m) araç değişken sayısından (s) daha büyük ise model eksik belirlenmiş veya belirlenememiştir denir ($m > s$).
- Son olarak eğer içsel açıklayıcı değişken sayısı (m) araç değişken sayısından (s) daha küçük ise model aşırı belirlenmiştir denir ($m < s$).

Yukarıda sıralanan kısıtlamalar ile ilgili olarak bir araştırmacı bu üç durumla da karşılaşabilir. Bu üç durum için de araç değişkenlerin geçerliliğini test etmek gereklidir. Seçilecek olan araç değişkenlerin $Cov(Z_i, u_i) = 0$ varsayımını sağlamalarının yanında içsel açıklayıcı değişken ile oldukça ilişkili olmaları ($Cov(Z_i, X_i) \neq 0$) gerekliliği önceki bölümlerde tartışılmıştı. Fakat tam belirlenme kısıtı için $Cov(Z_i, u_i) = 0$ varsayımını test etmek maalesef ki çok mümkün değildir ve genellikle araç değişkenlerin seçimi ekonomik teori ve önceki yapılan çalışmalardan yararlanılarak seçilmektedir (Cameron ve Trivedi, 2009:175). Aşırı belirlenme durumunda araçların geçerliliği, eğitimin ücretlere göre getirisi modeli üzerinden test edilecektir.

11.8.1. Aşırı Belirlenme Durumunda Araçların Geçerliliğinin Test Edilmesi

Aşırı belirlenme durumunda içsel olduğundan şüphelenilen açıklayıcı değişken sayısı muhtelif araç sayısından daha küçük ise aşırı belirlenme durumunun olacağından bahsedilmişti. Bu bölümde eğitimin ücretlere getirisi örneği üzerinden eğitim değişkeni için birden fazla araç değişken olması durumunda seçilecek araç değişkenlerin uygun olup olmadığı test edilecektir. Eğitimin ücretlere göre getirisi modeli tekrar ele alınsın:

$$\ln(\text{ücret})_i = \tau_0 + \tau_1 \text{egitim}_i + \tau_2 \text{yaşı}_i + \tau_3 \text{kent}_i + u_i \quad (11.34)$$

Bir önceki bölümde eğitim değişkeninin içsel açıklayıcı değişken olduğu Hausman (1978) testi ile test edilmişti. Buna göre eğitim değişkeni için potansiyel olarak eğitim değişkeni ile ilişkili (s adet araç değişken için $Cov(Z_{s,i} X_i) \neq 0$) ve hata terimi u_i ile ilişkisiz ($Cov(Z_{s,i} u_i) = 0$) araç değişkenlerini seçmek gerekeceğinden önceki bölümlerde seçilen eşin eğitimi değişkeni ile birlikte erkeğin işteki deneyimi (yıl) ve çocuk sayısı değişkenleri araç değişkenler olarak seçilir.

11.8.2. Aşırı Belirlenme Durumunda Araçların Geçerliliğinin Test Edilmesi İçin Stata Uygulaması

İncelenen örnek için muhtelif araç değişken sayısı ($s = \text{esegitim}_i; \text{Coc\_sayisi}_i; \text{is\_tecrübe}_i$) içsel açıklayıcı değişken sayısından ($m = \text{egitim}_i$) daha büyük olacaktır ($s = 3 > m = 1$). Seçilen araç değişkenlerin geçerliliğini test edebilmek amacıyla Stata programında “**estat overid**” komutu kullanılarak araç değişkenlerin geçerli olduğunu savunan sıfır hipotezi test edilir.

Bu tahmin için GMM yöntemi kullanılır. Araç değişken olarak eşin eğitimi, bireyin sahip olduğu çocuk sayısı ve bireyin işteki tecrübesi (çalışılan yıl) kullanılır. Bu durumda 1 içsel değişken ve 3 araç değişken olduğundan ($s > m$) aşırı belirlenme durumu test edilebilecektir.

Tablo 11.19. Eğitimin Erkeğin Ücreti Üzerindeki Etkisi: GMM Yöntemi

```
ivregress gmm lnucret yas kent (egitim = esegitim is_tecrube Coc_Sayisi), wmat-
rix(robust)
```

Instrumental variables (GMM) regression	Number of obs	=	4,236
	Wald chi2(3)	=	1590.36
	Prob > chi2	=	0.0000
	R-squared	=	0.3105
GMM weight matrix: Robust	Root MSE	=	.75107

		Coef.	Robust Std. Err.	z	P> z	[95% Conf. Interval]	
lnucrer							
egitim		.1306519	.004125	31.67	0.000	.122567	.1387368
yas		-.0026601	.0015554	-1.71	0.087	-.0057087	.0003884
kent		.3834727	.0350927	10.93	0.000	.3146922	.4522532
_cons		8.158013	.0780063	104.58	0.000	8.005123	8.310902

Instrumented: egitim

Instruments: yas kent esegitim is_tecrube Coc_Sayisi

Tablo 11.20. Seçilen Araç Değişkenlerin Geçerliliği Testi (Hansen's J testi)

estat overid

Test of overidentifying restriction:

Hansen's J $\chi^2(2) = 60.3043$ ($p = 0.0000$)

“**estat overid**” komutu ile elde edilen sonuçlar Tablo 11.20’de yer almaktadır. Aynı komut 2AEKK yöntemi için de kullanılabilir. Tahmin edilen test istatistiği Hansen’s J istatistiği olup, χ^2 dağılımı göstermektedir. GMM tahmini sırasında 1 ($m=1$) adet içsel, 3 ($s=3$) adet ise dışsal olduğu düşünülen araç değişken kullanılmıştır. Buna göre elde edilen p değeri(0.000), 0.01 anlamlılık düzeyinden küçük olduğundan seçilen araç değişkenlerin dışsal olduğunu ifade eden H_0 hipotezi reddedilir. Yani seçilen araç değişkenler geçerli değildir. Buna göre, Tablo 11. 18'deki model farklı araç değişkenler ile yeniden tahmin edilmelidir.

11.9. Özet ve Sonuç

Klasik Doğrusal Regresyon Modellerinde sıkça karşılaşılan problemlerden bir tanesi olan açıklayıcı değişkenler ile hata terimi arasında ortaya çıkan korelasyon sorunudur. “içsellik problemi” olarak adlandırılan bu tarz bir durumda EKK yöntemi, açıklayıcı değişken X ’in açıklanan değişken Y üzerindeki etkisini sapmalı ve tutarsız olarak tahmin edilir. Bu durumda ilk çözüm, açıklayıcı değişken X ile korelasyonlu, fakat gözlemlenemeyen stokastik değişken u hata terimi ile korelasyonsuz olan uygun bir araç değişken seçmektir. Bu bölümde Klasik Doğrusal Regresyon Modeli çerçevesinde Araç Değişkenler Yöntemi (IV Tahmin Yöntemi) ele alınmış; İki Aşamalı En Küçük Kareler Yöntemi (2AEKK) ve Genelleştirilmiş Momentler Metodu (GMM) da detaylı olarak incelenmiştir.

Araç değişkenler yönteminde dikkat edilecek en önemli unsur kullanılacak olan araç değişkenlerin geçerliliğinin test edilmesidir. Araç değişken ile içsel açıklayıcı değişken arasındaki korelasyon ne kadar yüksek ise araç değişken o kadar güçlü, aralarındaki korelasyon ne kadar düşük ise

araç değişkenin o kadar zayıf araç değişken olduğu ifade edilir. IV tahmincileri büyük örneklemelerde tutarlı ve etkin iken küçük örneklemelerde sapmalı ve EKK tahmincilerinden daha az etkin olacaklardır (Gujarati, 2015: 453).

Model tahmininde kullanılacak olan içsel açıklayıcı değişken ve araç değişken sayıları ise dikkat edilmesi gereken bir diğer önemli unsurdur. Modeldeki içsel açıklayıcı değişken sayısı ile kullanılan araç değişken sayısı birbirine eşit ise model tam belirlenmiş, kullanılan araç değişken sayısı içsel açıklayıcı değişken sayısından fazla ise model aşırı belirlenmiş ve son olarak içsel açıklayıcı değişken sayısı kullanılan araç değişken sayısından daha büyük ise model eksik belirlenmiştir denilir. Eğer model aşırı belirlenmiş ise aşırı belirlenme kısıtının ve kullanılan araç değişkenlerin geçerliliğinin test edilmesi gerekmektedir.

Uygun araç değişken seçimi için iktisadi teorinin ve önceden yapılmış olan çalışmaların detaylı incelenmesi gerekir. Modelde eğer birden fazla içsel açıklayıcı değişken var ise bu durumda her bir içsel açıklayıcı değişken için uygun araç değişken bulmak gerekmektedir. Günümüzde de ekonometrisyenler tarafından oldukça rağbet gören ve üzerinde detaylı çalışılan bir konu olan araç değişkenler konusu birçok paket programda da geliştirilen komutlar ve kodlar sayesinde rahatlıkla analiz edilebilmektedir (Gujarati, 2015: 472). Bu sebeple araç değişkenler konusunun ilerleyen günlerde daha da fazla gelişme göstereceği şüphesizdir.

STATA KOMUTLARI

```
use "DOSYAKONUMU\UCRET.dta", clear
global x2list bireyno yas kent esegitim is_tecrube Cocsay0_5 Coc_Sayisi yil
describe lnucet egitim $x2list
summarize lnucet egitim $x2list
regress lnucet egitim
regress lnucet egitim, robust
regress egitim esegitim, robust
regress lnucet esegitim, robust
regress egitim Cocsay0_5, robust
regress lnucet Cocsay0_5, robust
regress lnucet egitim kent yas, robust
regress egitim esegitim yas kent, robust
quietly predict egitimhat
regress lnucet egitimhat yas kent, robust
ivregress 2sls lnucet yas kent (egitim = esegitim), robust
ivregress 2sls lnucet yas kent (egitim = esegitim), first robust
ssc install ivreg2
ivreg2 lnucet yas kent (egitim = esegitim), robust
regress egitim yas kent esegitim, robust
quietly predict shat
regress lnucet egitim yas kent shat, robust
quietly ivregress 2sls lnucet yas kent (egitim = esegitim), robust
estat endogenous
quietly ivregress 2sls lnucet yas kent (egitim = esegitim), robust
estat firststage, forcenonrobust all
ivregress 2sls lnucet yas kent (egitim = esegitim), robust
ivregress gmm lnucet yas kent (egitim = esegitim), wmatrix(robust)
```

```
ivregress gmm lnucrer yas kent (egitim = esegitim is_tecrube Coc_Sayisi), wmat-
rix(robust)
estat overid
```

11.10. KAYNAKLAR

- Andrews, J., J. Stock, ve Sun, L. (2019): Weak Instruments in Instrumental Variable Regression: Theory and Practice, *Annual Review of Economics*, 11(1), 727–753. doi:10.1146/annurev-economics-080218-025643.
- Angrist, J. ve Krueger, A. (1991): "Does Compulsory School Attendance Affect Schooling and Earnings," *Quarterly Journal of Economics*, 106, 979-1014.
- Altın, A., Şenoğlu, B. (2008). Konum Parametrisinin Bazı Sağlam Tahmin Edicilerinin Örnekleme Alanında Kullanılması ve Bir Tarım Uygulaması. *Afyon Kocatepe Üniversitesi Fen Ve Mühendislik Bilimleri Dergisi*, 8 (1) , 291-306 .
- 10 Haziran 2020'de <https://dergipark.org.tr/tr/pub/akufemubid/issue/1607/20063> adresinden erişildi.
- Baum, C.F., Schaffer, M.E. ve Stillman, S., 2002. IVREG2: Stata module for extended instrumental variables/2SLS and GMM estimation. Boston College Department of Economics.
- Baum, C.F., Schaffer, M. E. ve Stillman, S., (2003), Instrumental variables and GMM: Estimation and testing. *The Stata Journal*. 3, Number 1, pp. 1–31.
- Baum, F.C. (2008). Using instrumental variables techniques in economics and finance. German Stata Users Group Meeting, Berlin, June 2008.
- 20 Haziran 2020 tarihinde <https://www.stata.com/meeting/germany08/Baum.DESUG8621.beamer.pdf> adresinden erişildi.
- Bound, J., Jaeger, D.A., Baker, R.M. (1995), "Problems with instrumental variables estimation when the correlation between the instruments and the endogeneous explanatory variable is weak", *Journal of the American Statistical Association* 90: 443-450.
- Cameron, A. C. ve Trivedi, P.K. (2005), *Microeconometrics: Methods and Applications*, Cambridge University Press, 6th Edition. New York.
- Cameron, A. C. ve Trivedi, P.K. (2009), *Microeconometrics Using Stata*, Stata Press. Texas.
- Card, D. (1994). Earnings, schooling, and ability revisited. *NBER Working Paper Series* 4832.
- Card, D. (1995): "Using Geographic Variation in College Proximity to Estimate the Return to Schooling," in *Aspects of Labour Market Behaviour: Essays in Honour of John Vanderkamp*, ed, by Louis N. Christofides, E. Kenneth Grant, and Robert Swidinsky. Toronto: *University of Toronto Press*, 201-222.
- Christopher F Baum & Mark E Schaffer & Steven Stillman (2002). "IVREG2: Stata module for extended instrumental variables/2SLS and GMM estimation," Statistical Software Components S425401, Boston College Department of Economics, revised 26 Jun 2020.
- Dickson, Matt (2009). The causal effect of education on wages revisited, *IZA Discussion Papers*, No. 4419, Institute for the Study of Labor (IZA), Bonn, <http://nbn-resolving.de/urn:nbn:de:101:1-200910122609>.
- Drichoutis, A. C., P. Lazaridis ve R. M. Nayga, Jr. (2009), "Body Weight Outcomes and Food Expenditures Among Older Europeans: A Simultaneous Equation Approach", paper presented at 113th EAAE Seminar, "A Resilient European Food Industry and Food Chain in a Challenging World", 3-6 September, Chania, Crete, Greece.

- Ersöz, T., & Toktamış, Ö., (2017). Açıklayıcı Değişkenlerde Hata Olması Durumunda Regresyon Modellerinin İncelenmesi. *Erciyes Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, vol.33, 102-120.
- Gould JB, Qin C, Marks AR, Chavez G. (2003). Neonatal mortality in weekend vs weekday births. *JAMA*. 289(22):2958-2962. doi:10.1001/jama.289.22.2958.
- Gujarati, D.N. and Porter, D.C. (2009). 4th Edition. Basic Econometrics. London: The Mc Graw Hill Companies.
- Gujarati, D. (2015). *Econometrics by Example*, 2nd Edition. London: Palgrave Macmillan Education.
- Güriş, S. & Çağlayan, E., (2013). *Ekonometri Temel Kavramlar*. 4. Basım. İstanbul: DER Yayınları.
- Hausman, J. A. (1978). Specification tests in econometrics. *Econometrica* 46: 1251–1271.
- Kaplan, F., Bulut, Y. (2011). Türkiye’de Harcamalar Yöntemiyle Milli Gelir Tahmini: Eşanlı Denklem Modelleri ve Genelleştirilmiş Momentler Yöntemi ile Tahmin. *Bingöl Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*. 1 (1), 43-56. Temmuz 2020 tarihinde <http://busbed.bingol.edu.tr/pub/issue/29515/399847> adresinden erişildi.
- Macfarlane, A. J., Dattani, N., Gibson, R., Harper, G., Martin, P., Scanlon, M., Newburn, M. and Cortina-Borja, M. (2019). Births and their outcomes by time, day and year: a retrospective birth cohort data linkage study. *Health Services and Delivery Research*, 7(18), doi: 10.3310/hsdr07180.
- Maddala, G. S. (2001), *Introduction to Econometrics*, 3rd edition, John Wiley and Sons, Chichester, UK.
- Shea, J. S. (1997). Instrument relevance in multivariate linear models: A simple measure. *Review of Economics and Statistics* 79: 348–352.
- Staiger, D. ve J.H. Stock (1997): “Instrumental Variables Regression with Weak Instruments,” *Econometrica*, 65, 557 – 586.
- Stata Help Menüsü, Stata Resmi Web Sitesi (2020). 10 Haziran 2020 tarihinde <https://www.stata.com/> adresinden erişildi.
- Stata Help Menüsü, ivreg2 (2020). 21 Haziran 2020 tarihinde <http://www.repec.org/bocode/i/ivreg2.html> adresinden erişildi.
- Stock, J. H., ve M. Yogo. 2005. Testing for weak instruments in linear IV regression. In Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg, ed. D. W. K. Andrews and J. H. Stock, 80–108. New York: *Cambridge University Press*.
- Tiku, M.L., Tan, W.Y. and Balakrishnan, N., (1986). Robust inference, Marcel Dekker Inc., New York.
- Türkiye İstatistik Kurumu Resmi Web sitesi (2020). Hane Halkı Bütçe Anketi Mikro Veri Seti, 2 Temmuz 2020 tarihinde <http://www.tuik.gov.tr/MicroVeri/HBA2013/turkce/metaveri/kapsam/index.html> adresinden erişildi.
- Verbeek, M. (2012). *A guide to modern econometrics*. Chichester, England: John Wiley & Sons.
- Wooldridge, J. M. (2007). Estimation of Average Treatment Effects Under Unconfoundedness
- What’s New in Econometrics NBER, Summer 2007. Lecture 1, Monday, July 30th, 9.00-10.30am. Estimation of Average Treatment Effects Under Unconfoundedness. 9 Temmuz 2020 tarihinde <http://www.nber.org/WNE/WNEnotes.pdf> adresinden erişildi.
- Wooldridge, J.M. (2013). *Ekonometriye Giriş Modern Yaklaşım / Introductory Econometrics Modern Approach*. 2. Cilt. Çeviri Editörü: Doç. Dr. Ebru Çağlayan. Nobel Yayıncılık.
- Yamano, T. (20p). Lecture Notes on Advanced Econometrics. Lecture 8: Instrumental Variables Estimation. Fall Semester 2009. 18 Ağustos 2020 tarihinde <http://www3.grips.ac.jp/~yamanota/Lecture%20Note%20to%2010%20SLS%20&%20others.pdf> adresinde erişildi.

BÖLÜM 12

PANEL İKİ DURUMLU MODELLER VE STATA UYGULAMALARI

Çiler Sigeze¹

12.1. Panel İki Durumlu Modeller

Panel veri, bireyler, haneler, firmalar ve ülkeler gibi birimlerin belli bir zaman periyodu boyunca tekrarlı gözlemlerinden oluşan veri setleridir. Panel veri ile oluşturulan nitel bağımlı değişkenli modellerden, bağımlı değişkenin 1 ve 0 değerini aldığı iki uçlu modeller ise panel iki durumlu modeller olarak tanımlanmaktadır.

Panel iki durumlu modellerinin genel formu aşağıdaki şekilde gösterilmektedir.

$$Y_{it}^* = X'_{it} \beta + \mu_i + u_{it} \quad i=1,2,\dots,N, t=1,2,\dots,T, \quad u_{it} \sim iid(0, \sigma^2) \quad (12.1)$$
$$Y_{it} = 1 \text{ eğer } Y_{it}^* > 0, \text{ diğer durumlarda } 0$$

Burada Y_{it} , bir durumun varlığında 1, yokluğunda ise 0 değerini alan bağımlı değişkeni göstermektedir. Y_{it}^* ise gözlenemeyen gizli değişkeni, X_{it} açıklayıcı değişkenler matrisini ve μ_i rassal birim etkiyi, u_i hata terimini ifade etmektedir. Panel veri modellerinde birimlere özgü ve gözlenemeyen özellikler (unobserved heterogeneity) birim etki olarak tanımlanmaktadır. Birim etki, bireyler için deneyim, yetenek, alışkanlıklar, firmalar için yönetici ve yönetim özellikleri gibi gözlenemeyen değişkenler olabilmektedir (Greene, 2003, s. 284). Panel veri modellerinde gözlenemeyen birim etkiler, sabit ya da rassal olarak ortaya çıkabilmektedir. Bunun yanında, birim etki ile açıklayıcı değişkenler arasında ilişkinin yapısına göre sabit etkiler ve rassal etkiler modelleri oluşmaktadır. Sabit etkiler modelinde gözlenemeyen birim etki modelde yer alan açıklayıcı değişkenler ile ilişkili olabilmekte, rassal etkiler modelinde ise gözlenemeyen birim etkinin rassal ve açıklayıcı değişkenler ile ilişkisiz olduğu varsayılmaktadır. Birim etki ile açıklayıcı değişkenler arasındaki ilişkinin test edilmesine dayanarak rassal ve sabit etkiler arasında karar verilmesi amcacıyla Hausman (1978) testi kullanılmaktadır (Greene, 2003, s. 690; Cameron ve Trivedi, 2005, s. 781).

Nitel bağımlı değişkenli modeller daha çok pazarlama, işletme, psikoloji gibi sosyal bilim alanlarında birimlerin (firmalar, haneler, fertler) rassal olarak seçildiği anket çalışmalarında kullanılmaktadır. Bu şekilde her bir birime özgü gözlenemeyen birim etkinin rassal olduğu bilinmektedir. Bunun yanında ikili nitel bağımlı değişkenli modellerde gözlenemeyen birim etkinin sabit

¹ Çukurova Üniversitesi, İktisadi ve İdari Bilimler Fakültesi, Ekonometri Bölümü.
E-mail: csigeze@cu.edu.tr

olduğu modellere de rastlanılmaktadır. Panel iki durumlu modellerde, panel verideki birimler tüm zaman boyunca gözlemlendiğinde dengeli panel ve bazı birimlerin bazı zaman dilimlerinde gözlenemediği durumda dengesiz panel veri tipleri oluşmaktadır.

12.2. Rassal Etkiler İki Durumlu Modeller

Panel veri modellerinde birim etkilerin rassal olduğu varsayıldığında rassal etkiler modeli oluşturulmaktadır. Rassal etkiler iki durumlu modellerin genel formu aşağıdaki şekilde gösterilmektedir.

$$Y_{it}^* = x'_{it}\beta + \mu_i + v_{it} \quad i=1,2,\dots,N, t=1,2,\dots,T, \quad (12.2)$$

$$Y_{it} = 1 \text{ eğer } Y_{it}^* > 0, \text{ diğer durumlarda } 0$$

Doğrusal panel veri modellerinde yer aldığı gibi birleşik hata terimi eşitlik (12.3) ile ifade edilmektedir.

$$\varepsilon_{it} = \mu_i + v_{it} \quad (12.3)$$

μ_i rassal birim etkiyi ve v_{it} normal dağılımdan geldiği bilinen rassal hata terimini göstermektedir. μ_i ve v_{it} rassal değişkenleri, birbirlerinden bağımsız olmasının yanında hata terimlerinin hem birimler, hem de zamanlar arasında korelasyonsuz ve birim etkilerin de birimler arasında ilişkisiz ve normal dağılımlı olduğu bilinmektedir.

Birleşik hata terime ilişkin varyans, kovaryans ve korelasyon değerleri sırasıyla eşitlik(12.4)-eşitlik(12.6) ile gösterilmektedir (Baltagi, 2005, s. 15; Johnston ve Dinardo, 1997, s. 391; Greene, 2003, s. 691).

$$\text{var}(\varepsilon_{it}) = \sigma_\mu^2 + \sigma_v^2 \quad (12.4)$$

$$\text{cov}(\varepsilon_{it}, \varepsilon_{js}) = \sigma_\mu^2 + \sigma_v^2 = \sigma_\mu^2 + 1 \quad i = j, t = s \quad (12.5)$$

$$\text{corr}(\varepsilon_{it}, \varepsilon_{js}) = \sigma_\mu^2 / (\sigma_\mu^2 + \sigma_v^2) = \rho \quad i = j, t \neq s \quad (12.6)$$

İki durumlu modellerde, bağımlı değişkenin $Y_{it} = 1$ olması diğer deyişle bir olayın başarı olasılığı eşitlik (12.7) ile ifade edilmektedir.

$$P(Y_{it} = 1/\mu_i) = F(x'_{it}\beta) \quad (12.7)$$

$Y_{it} = 0$ olması diğer deyişle bir olayın başarısız olma olasılığı ise aşağıdaki eşitlikte verilmektedir.

$$P(Y_{it} = 0/\mu_i) = 1 - F(x'_{it}\beta) \quad (12.8)$$

Probit ve logit modelleri ile bağımlı değişkenin gerçekleşme olasılığı üzerinde açıklayıcı değişkenlerin etkisi tanımlanmaktadır. Panel logit modeli, hata teriminin lojistik olasılık dağılımlı olduğunu varsayar ve genel olarak aşağıdaki şekilde yazılmaktadır.

$$F(x'_{it}\beta) = \frac{e^{x'_{it}\beta}}{1 + e^{x'_{it}\beta}} \quad (12.9)$$

Birim etkinin dikkate alınması ile oluşturulan model aşağıdaki gibidir.

$$F(x'_{it}\beta) = \frac{e^{x'_{it}\beta + \mu_i}}{1 + e^{x'_{it}\beta + \mu_i}} \quad (12.10)$$

Hata teriminin normal dağılımlı olduğu panel probit modeli ise genel olarak (12.11) eşitliği ile gösterilmektedir.

$$F(x'_{it}\beta) = \Phi(x'_{it}\beta + \mu_i) = \int_{-\infty}^{x'_{it}\beta + \mu_i} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du \quad (12.11)$$

Rassal etkiler probit ve logit modellerin tahmini sıklıkla en çok olabilirlik yöntemi ile yapılmaktadır. Yatay-kesit veriler için olabilirlik fonksiyonu aşağıdaki gibidir.

$$P(y_i | x_i) = \int_{L_i} f(\varepsilon_i) d\varepsilon_i \quad (12.12)$$

$y_i = 0$ için $(L_i, U_i) = (-\infty, -x'_i\beta)$ ve $y_i = 1$ için $(L_i, U_i) = (-x'_i\beta, +\infty)$ değer aralığındadır.

Panel probit ve panel logit modellerin genişletilmesi sonucunda i birimin tüm zamanlar için bileşik olasılığı aşağıdaki eşitlikte verilmiştir.

$$L_i = P(y_{i1}, y_{i2}, \dots, y_{iT_i} | X) = \int_{L_{i1}}^{U_{i1}} \dots \int_{L_{iT_i}}^{U_{iT_i}} f(\varepsilon_{i1}, \varepsilon_{i2}, \dots, \varepsilon_{iT_i}) d\varepsilon_{i1} d\varepsilon_{i2} \dots d\varepsilon_{iT_i} \quad (12.13)$$

Bileşik yoğunluğun integralinin alınması zor olmasına rağmen rassal birim etkilerin varlığı bu işlemi kolaylaştırmaktadır. $(\varepsilon_{i1}, \dots, \varepsilon_{iT_i}, \mu_i)$ 'nin bileşik yoğunluğu eşitlik (12.14) ile gösterilmek üzere hata terimlerinin bileşik olasılığı eşitlik (12.15) gibi yazılmaktadır.

$$f(\varepsilon_{i1}, \varepsilon_{i2}, \dots, \varepsilon_{iT_i}, \mu_i) = f(\varepsilon_{i1}, \varepsilon_{i2}, \dots, \varepsilon_{iT_i} | \mu_i) f(\mu_i) \quad (12.14)$$

$$f(\varepsilon_{i1}, \varepsilon_{i2}, \dots, \varepsilon_{iT_i}) = \int_{-\infty}^{+\infty} f(\varepsilon_{i1}, \varepsilon_{i2}, \dots, \varepsilon_{iT_i} | \mu_i) f(\mu_i) d\mu_i \quad (12.15)$$

Eşitlik (12.15) ile belirtilen hata terimlerinin bileşik olasılığı, hata terimlerinin bağımsız olması nedeniyle aşağıdaki gibi gösterilebilir.

$$f(\varepsilon_{i1}, \varepsilon_{i2}, \dots, \varepsilon_{iT_i}) = \int_{-\infty}^{+\infty} \prod_{t=1}^{T_i} f(\varepsilon_{it} | \mu_i) f(\mu_i) d\mu_i \quad (12.16)$$

Hata terimlerinin bileşik olasılığını ifade eden eşitlik (12.16), eşitlik (12.13)'te yerine konulduğunda olabilirlik fonksiyonu yeniden yazılabilmektedir.

$$L_i = P(y_{i1}, y_{i2}, \dots, y_{iT_i} | X) = \int_{L_{i1}}^{U_{i1}} \dots \int_{L_{iT_i}}^{U_{iT_i}} \int_{-\infty}^{+\infty} \prod_{t=1}^{T_i} f(\varepsilon_{it} | \mu_i) f(\mu_i) d\mu_i d\varepsilon_{i1} d\varepsilon_{i2} \dots d\varepsilon_{iT_i} \quad (12.17)$$

İntegrallerin sırası bağımsız olduğu için yerlerinin değiştirilmesi ile eşitlik (12.17) yeniden düzenlenebilmektedir.

$$L_i = P(y_{i1}, y_{i2}, \dots, y_{iT_i} | X) = \int_{-\infty}^{+\infty} \left[\prod_{t=1}^{T_i} \left(\int_{L_{it}}^{U_{it}} f(\varepsilon_{it} | \mu_i) d\varepsilon_{it} \right) \right] f(\mu_i) d\mu_i \quad (12.18)$$

Burada birim etkiye koşullu olasılık dağılımı $\int_{L_{it}}^{U_{it}} f(\varepsilon_{it} | \mu_i) d\varepsilon_{it}$, eşitlik (12.12) dikkate alınarak, $x'_{it}\beta + \mu_i$ 'ye koşullu olasılıkları vermektedir. Böylece ikili nitel bağımlı değişkenli rassal etkiler modeli için olabilirlik fonksiyonu genel olarak aşağıdaki gibi yazılmaktadır (Greene, 2003, s. 691-692).

$$L_i = P(y_{i1}, y_{i2}, \dots, y_{iT_i} | X) = \int_{-\infty}^{+\infty} \left[\prod_{t=1}^{T_i} P(Y_{it} = y_{it} | x'_{it}\beta + \mu_i) \right] f(\mu_i) d\mu_i \quad (12.19)$$

Olabilirlik fonksiyonunun hesaplanabilmesi için denklemde yer alan integrallerin hesaplanması ve bunun için de birim etkinin dağılımının belirlenmesi gerekmektedir. Butler ve Moffitt (1982) çalışmalarında birim etkinin normal dağılımlı olduğunu varsayarak ve Gauss-Hermite kuadratik yaklaşımından yararlanarak integral için çözüm önermişlerdir. İntegralin çözümü için alternatif olarak, kuadratik yaklaşıma göre daha esnek bir yaklaşım olan MSL (maksimum simulated likelihood) yönteminden yararlanılmaktadır. MSL yaklaşımı, rassal etkiler modelinde birim etkinin herhangi bir dağılımı için logaritmik olabilirlik fonksiyonunun maksimizasyonunu sağlamaktadır. Bu yöntemlerin yanında rassal etkiler probit modelinin tahmini için GMM ve simüle edilmiş olabilirlik yöntemleri de kullanılmaktadır (Greene, 2003, s. 691-692; Contoyannis vd., 2004).

12.3. Sabit Etkiler İki Durumlu Modeller

Panel iki durumlu nitel bağımlı değişkenli modellerin sabit etkiler içermesi ile oluşan sabit etkiler iki durumlu modeller eşitlik (12.20) gibi gösterilmektedir.

$$Y_{it}^* = X_{it}' \beta + \mu_i d_{it} + v_{it} \quad i=1,2,\dots,N, \quad t=1,2,\dots,T, \quad v_{it} \sim iid(0, \sigma^2) \quad (12.20)$$

$$Y_{it} = 1 \text{ eğer } Y_{it}^* > 0, \text{ diğer durumlarda } 0$$

d_{it} , i-nci birey için 1, diğer durumlar 0 değerini alan bir kukla değişkeni, X_{it}' ise sabit terimi içermeyen açıklayıcı değişkenler matrisini göstermektedir. Buna göre sabit etkiler iki durumlu modelin logaritmik olabilirlik fonksiyonu eşitlik (12.21) ile gösterilmektedir.

$$\ln L = \sum_{i=1}^N \sum_{t=1}^T \ln P(y_{it} | \mu_i + X_{it}' \beta) \quad (12.21)$$

$P(\cdot)$, gözlemlenen çıktının olasılığı olmak üzere probit model için $\Phi[q_{it}(\mu_i + X_{it}' \beta)]$ ve logit model için $\Lambda[q_{it}(\mu_i + X_{it}' \beta)]$ şeklinde gösterilmektedir.

Doğrusal panel veri modellerinde, sabit etkiler modelinin tahmini için birim etkileri yok eden grup ortalamalarından sapmaların alındığı dönüştürme yöntemi kullanılmaktadır. Ancak doğrusal olmayan modellerde bu dönüştürmeyi yapmak mümkün olmamaktadır. Bu durumda sabit terimleri de dikkate alarak logaritmik olabilirlik fonksiyonunun birinci ve ikinci türevlerini almak gerekmektedir². Böylece, elde edilen olabilirlik fonksiyonlarının çözümü kolaylıkla sağlanabilmektedir. Ancak sabit etkiler modelinin en çok olabilirlik yöntemi ile çözümünde birim etkilerin varlığı nedeni ile parametre sayısının fazla olması bazı problemlere neden olmaktadır. *Rastlantısal (Incidental) parametre problemi* (bkz. Neyman ve Scott, 1948; Lancaster, 2000) olarak bilinen problemde birim etki μ_i 'nin parametre sayısının, gözlem sayısı N ile birlikte artması ancak bu artışın birim etki hakkında ek bir bilgi sağlamaması nedeniyle μ_i 'nin tüm tahmin edicileri N büyük olsa bile T sonlu ise sapmalı olacaktır. Doğrusal regresyon modellerinde T sonlu olduğu durumda, μ_i 'nin tahmin edicileri tutarsız olsa bile β ve μ_i 'nin asimptotik bağımsız olması nedeniyle β 'nin grup içi (within) tahmin edicisi tutarlı olmaktadır. Ancak doğrusal olmayan modellerde β ve μ_i 'nin birbirinden bağımsız olmaması nedeniyle örneklem sayısı sonsuza gitmesine rağmen $\hat{\mu}_i$ 'nin tutarsızlığı β 'nin tahmincisine yansımaktadır (Baltagi, 2003, s. 179). Pa-

² Türev alma işlemleri için, bkz. Greene, 2003, s. 695-696

parametrelerin tutarlı tahminlerini sağlamak için birim etki parametrelerinin minimum yeterli istatistikleri kullanılabilmektedir. Buna göre birim etkiyi içermeyen ve yerine başka bir istatistiğin kullanıldığı fonksiyonlar ile parametre tahminleri sağlanabilmektedir. Yeterli istatistikler probit model için uygun değilken logit model için uygundur (Greene, 2003, s. 698). Bu durumda sabit etkilerin varlığı altında probit model tahmini sağlanamazken logit model iyi bir alternatif sunmaktadır.

Panel logit model eşitlik (12.22) ile ifade edilmek üzere,

$$P(y_{it} = 1 | x_{it}) = \frac{e^{\mu_i + x'_{it}\beta}}{1 + e^{\mu_i + x'_{it}\beta}} \quad (12.22)$$

nT bağımsız gözlem için koşulsuz olabilirlik fonksiyonu eşitlik (12.23) ile gösterilmektedir.

$$L = \prod_i \prod_t (F_{it})^{y_{it}} (1 - F_{it})^{1-y_{it}} \quad (12.23)$$

Birim etkileri de içeren koşulsuz olabilirlik fonksiyonu yerine Chamberlain (1980) tarafından μ_i 'lerden bağımsız koşullu olabilirlik fonksiyonu eşitlik (12.24) önerilmiştir.

$$\begin{aligned} L^c &= \prod_{i=1}^n P(Y_{i1} = y_{i1}, Y_{i2} = y_{i2}, \dots, Y_{iT_i} = y_{iT_i} | \sum_{t=1}^{T_i} y_{it}) \\ &= \frac{\exp\left(\sum_{t=1}^{T_i} y_{it} x'_{it} \beta\right)}{\sum_{S_i \in \mathcal{S}_i} \exp\left(\sum_{t=1}^{T_i} d_{it} x'_{it} \beta\right)} \end{aligned} \quad (12.24)$$

Paydada yer alan fonksiyon, $S_i = \sum_{t=1}^{T_i} y_{it}$ ile aynı toplama sahip tüm farklı $\begin{pmatrix} T_i \\ S_i \end{pmatrix}$ kümeleri üzerinden toplanmaktadır (Greene, 2003, s. 698).

12.4. Panel İki Durumlu Modeller Stata Uygulaması

Panel logit model tahminleri için Stata komutu olarak **xtlogit**, Panel probit model tahminleri için **xtprobit** kullanılmaktadır. Komutun temel yapısı aşağıda belirtilmektedir.

xtlogit depvar [indepvars] [if] [in] [weight] [, options]

xtprobit depvar [indepvars] [if] [in] [weight] [, options]

Komut yapısı hakkında detaylı bilgi için Stata programında “**help**” menüsünden **xtlogit** ya da **xtprobit** komutu yardımıyla açılan bilgilendirme ekranından detaylara ulaşabilirsiniz. Birim etkilerin sabit olduğunu belirtmek amacıyla Stata komutunda “,” den sonra opsiyon için “**fe**” (fixed effects) yazılması gerekmektedir. Aksi takdirde xtlogit komutu **default** olarak rassal etkiler tahminleri üretir. Probit modeli, birim etkilerin sabit olduğu durumda Stata paket programında tahmin edilememektedir. Birim etkilerin rassal olduğunu belirtmek amacıyla Stata komutunda “,” den sonra opsiyon için “**re**” (random effects) yazılması gerekmektedir. Modellerin tahmininden sonra tahmin edilen katsayıların marjinal etkilerini elde etmek amacıyla Stata komut satırına “**mfx**” yazılmalıdır.

Örnek 12.1. Rassal etkiler logit, rassal etkiler probit ve sabit etkiler logit modellerinin tahmini ve açıklayıcı değişkenlerin marjinal etkilerini hesaplamak için **yoksulluk.dta** veri setinden yararlanıla-

caktır. Uygulama için kullanılan veri, Türkiye İstatistik Kurumu tarafından 2009-2012 yıllarında derlenen Gelir ve Yaşam Koşulları Araştırması (GYKA)'ndan elde edilmektedir. Dört yıl boyunca izlenen 2500 hane ve hanehalkı reisine ait özelliklerin yer aldığı dengeli panel veri kullanılmaktadır.

Stata paket programında değişkenlerin tanımlarını listelemek için aşağıda verilen “ describe” komutu kullanılır.

describe

Contains data from C:\dosyakonumu\yoksulluk.dta

obs: 10,000

vars: 9

size: 230,000

variable name	storage type	display format	value label	variable label
time	int	%8.0g		
id	long	%12.0g		
yoksulluk	byte	%8.0g		
yas	int	%8.0g		
yas2	float	%9.0g		
universite	byte	%8.0g		
gelir	float	%8.0g		

Çalışmada kullanılan bağımlı değişken hane yoksul ise 1, değil ise 0 değerini alan iki durumlu olmak üzere model aşağıdaki gibi kurulmaktadır.

$$yoksulluk_{it} = \beta_0 + \beta_1 yas_{it} + \beta_2 yas2_{it} + \beta_3 cinsiyet_{it} + \beta_4 universite_{it} + \beta_5 gelir_{it} + u_{it}$$

Çalışmada kullanılacak değişkenlere dair bilgiler Tablo 12.1’de yer almaktadır

Tablo 12.1. Modelde Kullanılan Değişkenler ve Tanımları

Değişkenler	Tanım	Açıklama
id	Hane numaraları	2500 hane
time	Yıllar	2009, 2010, 2011, 2012
yoksulluk ³	Bağımlı değişken	Hanehalkı yoksul ise 1, değil ise 0.
cinsiyet	Açıklayıcı değişken	Hanehalkı reisi kadın ise 1, değil ise 0.
yas	Açıklayıcı değişken	Hanehalkı reisinin yaşı
yas2	Açıklayıcı değişken	Hanehalkı reisinin yaşının karesi
universite	Açıklayıcı değişken	Hanehalkı reisi yükseköğretim, fakülte ve üzeri mezunu ise 1, değil ise 0.
gelir	Açıklayıcı değişken	Hanehalkı toplam geliri

³ OECD eşdeğerlik ölçeği kullanılarak eşdeğer fert başına düşen medyan gelirinin yüzde 60’ı yoksulluk sınırı olarak kabul edilmiş ve bu sınırın altında gelire sahip olan haneler yoksul olarak belirlenmiştir.

Veri dosyası Stata paket programında açıldıktan sonra, verinin panel olarak Stata'ya tanıtılması için `xtset` komutu ile birlikte birim (id) ve zamanın (time) belirtilmesi gerekmektedir.

xtset id time

Panel logit modelini sabit etkiler varsayımıyla tahmin etmek için aşağıdaki komut kullanılmaktadır.

xtlogit yoksulluk yas yas2 cinsiyet üniversite gelir, fe

Tablo 12.2. Sabit Etkiler Logit Model Tahmini

Conditional fixed-effects logistic regression	Number of obs	=	2,324
Group variable: id	Number of groups	=	581
	Obs per group:		
	min	=	4
	avg	=	4.0
	max	=	4
	LR chi2(5)	=	1141.22
Log likelihood = -304.56678	Prob > chi2	=	0.0000

yoksulluk	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
yas	.372298	.1037621	3.59	0.000	.1689279 .575668
yas2	-.0025203	.0009797	-2.57	0.010	-.0044405 -.0006001
cinsiyet	-.5945457	.6969605	-0.85	0.394	-1.960563 .7714718
üniversite	-9.396748	572.9001	-0.02	0.987	-1132.26 1113.467
gelir	-.0006777	.0000409	-16.58	0.000	-.0007578 -.0005976

Tablo 12.2'de yer alan sonuçlara göre sabit etkiler varsayımı altında logit modelin tahmini koşullu sabit etkiler lojistik regresyon modeli ile tahmin edilmiştir. Sabit etkiler regresyon modelinde yalnızca zaman içinde birimdeki (birey, hane, firma vs.) değişimden gelen bilgi dikkate alınmaktadır. Zaman içinde birimde herhangi bir değişim olmadığı durumda sabit etkiler regresyon modeli, değişim göstermeyen birimi dışlamaktadır. Bu nedenle modelde yoksulluk durumları 4 yıl boyunca hiç değişmeyen 1919 hane çıkarılmış ve geriye kalan 581 hane ve hanehalkı reisine ait bilgi ile sabit etkiler logit model tahmini gerçekleştirilmiştir.

Tablo 12.2'ye göre ilk olarak, olabilirlik oran (LR) testi (1141.22) modelin bir bütün olarak anlamlı olduğunu göstermektedir ($\text{prob} > \chi^2 = 0.0000$). Örnekte yer alan her bir değişkenin katsayısının sıfırdan farklı olup olmadığına bakıldığında, *yas*, *yas2* ve *gelir* değişkenlerinin katsayıları %5 anlamlılık düzeyinde istatistiki olarak anlamlı (p değerleri $\alpha = 0.05$ değerinden küçük olduğundan), üniversite ve cinsiyet değişkenlerinin katsayıları ise istatistiksel olarak anlamlı değildir (p değerleri $\alpha = 0.10$ değerinden büyük).

Logit model sabit etkiler varsayımı ile tahmin edilmesiyle birlikte, kullanılan örneklemin büyük bir anakütleden rassal olarak çekilmesi nedeniyle önsel olarak birim etkilerin rassal olması beklenmektedir (Baltagi, 2005). Hausman testinden yararlanarak sabit etkiler ve rassal etkiler arasında karar verebilmek amacıyla sabit etkiler logit model tahmininden sonra Hausman testinde

kullanmak amacıyla “**estimates store**” komutu yanına herhangi bir isim yazılarak sabit etkiler kaydedilmelidir.

estimates store FElogit

Stata paket programında, panel logit modelini rassal etkiler varsayımıyla tahmin etmek için aşağıdaki komut kullanılmaktadır.

xtlogit yoksulluk yas yas2 cinsiyet üniversite gelir, re

Tablo 12.3. Rassal Etkiler Logit Model Tahmini

Random-effects logistic regression	Number of obs	=	10,000
Group variable: id	Number of groups	=	2,500
Random effects u_i ~ Gaussian	Obs per group:		
	min	=	4
	avg	=	4.0
	max	=	4
Integration method: mvaghermite	Integration pts.	=	12
	Wald chi2(5)	=	569.84
Log likelihood = -2106.3995	Prob > chi2	=	0.0000

yoksulluk	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
yas	.2954704	.0465623	6.35	0.000	.2042098 .3867309
yas2	-.002946	.0004325	-6.81	0.000	-.0037936 -.0020983
cinsiyet	-1.478468	.2855747	-5.18	0.000	-2.038184 -.9187515
üniversite	-3.466824	.9902767	-3.50	0.000	-5.40773 -1.525917
gelir	-.0007935	.0000335	-23.70	0.000	-.0008591 -.0007279
_cons	1.185422	1.14588	1.03	0.301	-1.060462 3.431305
/lnsig2u	2.710797	.102297			2.510298 2.911295
sigma_u	3.878305	.1983694			3.508361 4.287258
rho	.820531	.0150642			.7890903 .8481865
LR test of rho=0: chibar2(01) = 1147.96 Prob >= chibar2 = 0.000					

Tablo 12.3'te, 2500 haneden oluşan 4 yıllık panel veri ile elde edilen rassal etkiler logit model tahmin sonuçları yer almaktadır. Wald testi sonucuna göre model genel olarak anlamlıdır. Parametrelerin anlamlılığını sınamak için t istatistiği yerine z istatistiği hesaplanmıştır. Tüm bağımsız değişkenlerin katsayıları istatistiksel olarak anlamlıdır. Tablo 12.3'ün en altında yer alan olabilirlik oranı (LR) test (1147.96) sonucuna göre $\rho \neq 0$ alternatif hipotezine karşı, $\rho = 0$ sıfır hipotezi reddedilmektedir (prob>chi2=0.000). Buna göre birim etkiye ait varyansın, birim etki ve hata terimi varyans toplamaları içindeki payını veren ρ katsayısı istatistiksel olarak anlamlıdır. Dolayısıyla modelde birim etki vardır. Birim etkinin rassal ya da sabit olup olmadığına karar verebil-

mek amacıyla birim etkiler ve açıklayıcı değişkenler arasındaki ilişkinin sınanmasına dayanan Hausman testinden yararlanılmaktadır. Rassal birim etkilerin Hausman testinde kullanılabilmesi için rassal etkiler logit modelin tahmininden sonra “**estimates store**” komutu yanına herhangi bir isim yazılarak kaydedilmelidir.

estimates store RElogit

Panel logit modelde anlamlı olan birim etkinin sabit ya da rassal olup olmadığının belirlenmesi amacıyla Hausman testinden yararlanılmaktadır. Test için gerekli olan komut “**hausman**” olmak üzere sabit etkiler logit model tahmininden sonra kaydedilen sonuçlar “FElogit” ve rassal etkiler logit model tahmininden sonra kaydedilen sonuçlar “RElogit” kullanılmalıdır.

hausman FElogit RElogit

Tablo 12.4. Hausman Testi Sonuçları

	---- Coefficients ----			
	(b)	(B)	(b-B)	sqrt(diag(V_b-V_B))
	FE	RE	Difference	S.E.
yas	.372298	.2954704	.0768276	.0927283
yas2	-.0025203	-.002946	.0004257	.0008791
cinsiyet	-.5945457	-1.478468	.883922	.6357681
üniversite	-9.396748	-3.466824	-5.929925	572.8993
gelir	-.0006777	-.0007935	.0001158	.0000234

b = consistent under Ho and Ha; obtained from xtlogit

B = inconsistent under Ha, efficient under Ho; obtained from xtlogit

Test: Ho: difference in coefficients not systematic

$$\begin{aligned}
 \text{chi2}(3) &= (b-B)'[(V_b-V_B)^{-1}](b-B) \\
 &= 2.74 \\
 \text{Prob}>\text{chi2} &= 0.4335
 \end{aligned}$$

Tablo 12.4’te panel logit modeldeki birim etkilerin açıklayıcı değişkenler ile ilişkisi Hausman testi ile sınanmaktadır. Hausman (1978), birim etkiler ile açıklayıcı değişkenler arasında ilişki olmadığını belirten sıfır hipotezi altında her iki tahmin edicinin tutarlı olduğunu ancak sıfır hipotezinin reddedildiği durumda rassal etkiler tahmin edicisinin tutarsız ve sapmalı olduğunu ifade etmiştir. Dahası her iki hipotez altında sabit etkiler tahmin edicisi tutarlı iken sıfır hipotezi reddedilmediği durumda rassal etkiler tahmin edicisi En iyi Doğrusal Sapmasız Tahmin Edici (BLUE) olmasının yanında asimptotik olarak etkin ve tutarlıdır (Baltagi, 2005, s. 66). Bu bilgiler doğrultusunda Tablo 12.4’ün en altında yer alan 2.74 değerindeki chi2(3) testinin prob. değeri 0.4335, $\alpha = 0.05$ değerinden büyük olduğu için %5 anlamlılık düzeyinde Ho hipotezi reddedilememektedir. Buna göre rassal etkiler tahmin edicisinin etkin ve tutarlı olduğu sonucuna ulaşılmaktadır.

Logit modelin rassal etkiler tahmin edicisi ile elde edilen sonuçları Tablo 12.3’te yer almaktadır. Modelde yer alan tüm açıklayıcı değişkenlerin katsayıları %1 düzeyinde anlamlı bulunmuştur. Yaş değişkeninin katsayısının hanelerin yoksul olma olasılığı üzerinde pozitif bir etkisi varken yaşınn ka-

resi negatif etkilidir. Buna göre hanehalkı reisinin yaşı arttıkça yoksul olma olasılığı artarken, ortalama yaştan sonra yaş artışı ile birlikte hanelerin yoksul olma olasılıkları düşecektir. Hanehalkı reisi üniversite mezunu olan hanelerin, hanehalkı reisi üniversite mezunu olmayanlara göre yoksul olma olasılıkları daha düşüktür. Benzer şekilde, hanehalkı reisi kadın olan hanelerin, hanehalkı reisi erkek olan hanelere göre yoksul olma olasılıkları daha azdır. Son olarak hanehalkı toplam gelirinin hanelerin yoksul olma olasılığı üzerinde beklenildiği gibi negatif bir etkisi bulunmaktadır.

Bilindiği üzere logit modelde tahmin edilen katsayı büyüklükleri, hanelerin yoksul olma olasılıkları üzerinde değişkenlerin doğrudan etkisini göstermemektedir. Logit model katsayı tahminleri, fark oranlarının (bahis oranı, odds ratio) logaritmasını vermektedir. Bu nedenle katsayı büyüklüklerinin yorumlanması şu şekilde olmaktadır. Diğer değişkenler sabit tutulduğunda hanehalkı reisi bir yıl daha yaş aldığı anda hanenin yoksul olma olasılığının yoksul olmama olasılığına oranını veren fark oranının logaritması yaklaşık olarak 0.29 artacaktır. Bu şekilde katsayı büyüklüklerinin yorumlanması çok zor olduğu için bir açıklayıcı değişkenin başarı olasılığı (bu soru için hanenin yoksul olma olasılığı) üzerindeki doğrudan etkisini veren marjinal etkiler hesaplanabilmektedir. Stata paket programında herhangi iki durumlu nitel bağımlı değişkenli modellerde açıklayıcı değişkenlerinin marjinal etkilerini hesaplamak amacıyla öncelikle model tahmini yapılır ve ardından "mfx" komutu kullanılır. Bu durumda rassal etkiler logit modelinin parametre tahminlerinin marjinal etkilerini elde etmek amacıyla sırasıyla aşağıdaki komutlar kullanılmaktadır.

**xtlogit yoksulluk yas yas2 cinsiyet üniversite gelir, re
mfx**

Tablo 12.5. Rassal Etkiler Logit Model- Marjinal Etkileri

Marginal effects after xtlogit

y = Linear prediction (predict)

= -10.606938

variable	dy/dx	Std. Err.	z	P> z	[	95% C.I.	]	X
-----+-----								
yas	.2954704	.04656	6.35	0.000	.20421	.386731		49.9746
yas2	-.002946	.00043	-6.81	0.000	-.003794	-.002098		2705.07
cinsiyet*	-1.478468	.28557	-5.18	0.000	-2.03818	-.918752		.1634
üniver~e*	-3.466824	.99028	-3.50	0.000	-5.40773	-1.52592		.1131
gelir	-.0007935	.00003	-23.70	0.000	-.000859	-.000728		22628.3
-----+-----								

(*) dy/dx is for discrete change of dummy variable from 0 to 1

Tablo 12.5'te yer alan sonuçlara göre, yas değişkeninin katsayısının etkisi pozitif ve yaşı karesinin etkisi ise negatif ve %1 önem düzeyinde anlamlı bulunmuştur. Bu bulgulara göre hanehalkı reisinin bir yıl daha yaş alması hanenin yoksul olma olasılığını %29 artırmaktadır. Belli bir yaştan sonra ise hanehalkı reisinin yaşındaki her artış hanenin yoksul olma olasılığını %0.2 azaltmaktadır.

Cinsiyet değişkeninin katsayısı ise pozitif ve %1 önem düzeyinde istatistiksel olarak anlamlı bulunmuştur. Bu bulguya göre hanehalkı reisi kadın olan hanelerin yoksul olma olasılığı hanehalkı reisi erkek olan hanelere göre %147 (1.47 kat) daha azdır. Benzer şekilde hanehalkı reisi-

nin üniversite mezunu olduğu hanelerin yoksul olma olasılığı, hanehalkı reisi diğer eğitim düzeylerinde olan hanelere göre %346 (3.46 kat) daha azdır⁴. Bu bulgulara göre hanehalkı reisinin kadın olması ve hanehalkı reisinin üniversite mezunu olması hanenin yoksul olma olasılığı üzerinde negatif etkilidir. Son olarak, hanehalkının toplam gelirindeki 1 birimlik artış, hanenin yoksul olma olasılığını %0.07 azaltmaktadır.

Logit modelin katsayı yorumları için marjinal etkilerin yanında fark oranı (bahis oranı, odds ratio, OR) da hesaplanabilmektedir.

Stata programında rassal etkiler logit modelinin parametre tahminlerinin fark oranlarını elde etmek amacıyla aşağıdaki komut kullanılmaktadır.

xtlogit yoksulluk gelir üniversite cinsiyet yas yas2, re or

Tablo 12.6. Rassal Etkiler Logit Model Fark Oranları

Random-effects logistic regression	Number of obs	=	10,000
Group variable: id	Number of groups	=	2,500
Random effects u_i ~ Gaussian	Obs per group:		
	min	=	4
	avg	=	4.0
	max	=	4
Integration method: mvaghermite	Integration pts.	=	12
Log likelihood = -2106.3995	Wald chi2(5)	=	569.84
	Prob > chi2	=	0.0000

yoksulluk	OR	Std. Err.	z	P> z	[95% Conf. Interval]	
gelir	.9992068	.0000335	-23.70	0.000	.9991412	.9992724
üniversite	.031216	.0309125	-3.50	0.000	.0044818	.2174216
cinsiyet	.2279868	.0651073	-5.18	0.000	.1302651	.3990169
yas	1.343758	.0625685	6.35	0.000	1.226556	1.47216
yas2	.9970584	.0004312	-6.81	0.000	.9962136	.9979039
_cons	3.272066	3.749394	1.03	0.301	.3462959	30.91696
/lnsig2u	2.710797	.102297			2.510298	2.911295
sigma_u	3.878305	.1983694			3.508361	4.287258
rho	.820531	.0150642			.7890903	.8481865

LR test of rho=0: chibar2(01) = 1147.96	Prob >= chibar2 = 0.000

⁴ Cinsiyet ve üniversite değişkenlerinin katsayıları teorik beklentiyle uyuşmayacak ölçüde büyük çıkmıştır. Ancak oluşturulan örnekte amaç konunun anlaşılır olmasını sağlamak olduğu için katsayı büyüklükleri sorgulanmamaktadır.

Tablo 12.6'da rassal etkiler logit model katsayılarının fark oranları (OR), standart hataları, z değerleri ve %95 güven aralıkları yer almaktadır. Tablo 12.6'da yer alan fark oranları, Tablo 12.3'te yer alan katsayıların üstel değeri alınarak hesaplanmaktadır. Örneğin yaş değişkeninin katsayısı 0.2954704'tür. Fark oranı $e^{0.2954704} = 1.343758$ olarak hesaplanır. Hanehalkı reisi bir yıl daha yaş alan hanelerin yoksul olma olasılığı yoksul olmama olasılığından 1.34 kat daha fazladır şeklinde yorumlanır. Cinsiyet değişkeni için fark oranı 0.2279868'dir. Hanehalkı reisi kadın olan hanelerin yoksul olma olasılığı, hanehalkı reisi erkek olan hanelerin yoksul olma olasılığından 0.22 kat daha azdır. Fark oranı 1'den küçük olduğu için yorumu kolaylaştırmak için fark oranı katsayısının tersi alınarak da yorum yapılabilir (Long ve Freese, 2014, s. 230). Burada dikkat edilmesi gereken önemli nokta referans sınıfın değiştirilmesi gerektiğidir. $1/0.2279=4.38$ değeri bulunur. Bu değer, hanehalkı reisi erkek olanların yoksul olma olasılığı hanehalkı reisi kadın olan hanelerin yoksul olma olasılığından 4.38 kat daha fazladır şeklinde yorumlanır. Son olarak, hanehalkı reisi üniversite mezunu olan hanelerin yoksul olma olasılığı, hanehalkı reisi üniversite mezunu olmayan hanelerin yoksul olma olasılığından yaklaşık 0.03 kat daha azdır. Fark oranı katsayısının tersi alındığında $1/0.0312= 32.051$ değeri bulunur. Yani hanehalkı reisi üniversite mezunu olmayanların yoksul olma olasılığı, üniversite mezunu olanların yoksul olma olasılığından 32 kat daha fazladır.

Bu bölümde kullanılan Stata komutları aşağıda toplu olarak yer almaktadır.

use "dosyakonu\yoksulluk.dta", clear

xtset id time

xtlogit yoksulluk yas yas2 cinsiyet üniversite gelir, fe

estimate store FElogit

xtlogit yoksulluk yas yas2 cinsiyet üniversite gelir, re

estimate store RElogit

hausman FElogit RElogit

xtlogit yoksulluk yas yas2 cinsiyet üniversite gelir, re

mfx

xtprobit yoksulluk yas yas2 cinsiyet üniversite gelir, re

mfx

xtlogit yoksulluk gelir üniversite cinsiyet yas yas2, re or

Alternatif olarak;

Sabit etkiler logit model marjinal etkileri için ilgili kod:

xtlogit yoksulluk yas yas2 cinsiyet üniversite gelir, fe

mfx

Rassal etkiler probit model marjinal etkileri için ilgili kod:

xtprobit yoksulluk yas yas2 cinsiyet üniversite gelir, re

mfx

12.5. KAYNAKLAR

- Baltagi, B. H. (2005). *Econometric Analysis of Panel Data*. John Wiley & Sons: England.
- Cameron, A. C., & Trivedi, P. K. (2005). *Microeconometrics: methods and applications*. Cambridge university press.
- Chamberlain, G. (1980). *Analysis of Covariance with Qualitative Data, the Review of Economic Studies*, Vol. 47, No. 1 II f. Econometrics Issue, 1.
- Contoyannis, P., Jones, A. M., & Rice, N. (2004). Simulation-based inference in dynamic panel probit models: an application to health. *Empirical Economics*, 29(1), 49-77.
- Greene, W. H. (2003). *Econometric Analysis*. Fifty Edition. Pearson Education: New Jersey.
- Hausman, J. A. (1978), Specification Tests in Econometrics. *Econometrica*, Vol. 46, No. 6 (Nov., 1978), pp. 1251-1271.
- Johnston, J. & Dinardo, J. (1997). *Econometrics Methods*. The McGraw-Hill Companies, Inc: Printed in the United States of America.
- Lancaster, T. (2000). The incidental parameter problem since 1948. *Journal of econometrics*, 95(2), 391-413.
- Long, J. S., & Freese, J. (2014). *Regression models for categorical dependent variables using Stata*. Stata press.
- Neyman, J., & Scott, E. L. (1948). Consistent estimates based on partially consistent observations. *Econometrica: Journal of the Econometric Society*, 1-32.
- TÜİK, Gelir ve Yaşam Koşulları Araştırması 2009-2012 panel mikro veri seti.

BÖLÜM 13

PANEL ÇOK DURUMLU MODELLER VE STATA UYGULAMALARI

Çiler Sigeze¹

13.1. Panel Çok Durumlu Modeller

Bağımlı değişkenin ikiden fazla değer aldığı nitel bağımlı değişkenli modeller, bağımlı değişkenin aldığı değerlerin belli bir sıralaması olması ve herhangi bir sıralamasının olmaması durumuna göre ikiye ayrılmaktadır. Bağımlı değişkenin aldığı değerlerin belli bir sıralaması varsa sıralı probit ya da sıralı logit modeller kullanılmaktadır. Bağımlı değişkenin aldığı değerlerin herhangi bir sırasının olmadığı durumda çok durumlu (multinomial) probit ve logit modeller oluşturulmaktadır. Bu modellere ek olarak çok durumlu logit modellerin tahmininde herhangi iki alternatifin birbirine oranının diğer alternatiflerden bağımsız olduğunu ifade eden ilişkisiz alternatiflerin bağımsızlığı (IIA, the independence of irrelevant alternatives) varsayımının model tahminlerinde yarattığı sıkıntıyı aşabilmek için logit modeline dayanan rassal parametreler logit (random parameters logit), karma (mixed) logit gibi modeller geliştirilmiştir. Çok durumlu logit modelde bağımlı değişkeni oluşturan alternatiflerin seçilme olasılığı yalnızca birimlerin özelliklerine bağlı iken, hem alternatiflerin hem de birimlerin özelliklerine göre bağımlı değişkenin seçilme olasılığını dikkate alan karma logit model, logit ailesinin en kapsamlı modelini oluşturmaktadır. Rassal parametreler logit modelde ise bazı parametrelerin rassal olduğu ve birimden birime değiştiğinin varsayılmasının yanında aynı birimlerin tekrarlı gözlemleri arasında korelasyona izin verilmektedir. Bu şekilde rassal etkilerin yer aldığı bağımlı değişkenli nitel modeller, genelleştirilmiş doğrusal karma modeller (Generalized Linear Mixed Models, GLMM) olarak da bilinmektedir (McCulloch ve Searle, 2001, s.2; Train, 1997, Revelt ve Train, 1998; Train, 2003, s. 141, Cameron ve Trivedi, 2005, s.491- 495, Croissant, 2012). Genel anlamda, nitel bağımlı değişkenli modellerde yer alan herhangi bir değişkenin rassal olarak kabul edildiği ve/veya panel veri ile oluşturulan nitel bağımlı değişkenli modellerde rassal etkinin varlığı bilindiğinde genelleştirilmiş doğrusal karma modellerin bir sınıfı oluşmaktadır.

Nitel bağımlı değişkenli modellerde rassal olarak seçilen bireylerin tekrarlı seçimleri arasında korelasyonun olması beklenmektedir. Bu korelasyonu dikkate almanın en kolay yolu birimlere özgü rassal etkinin dâhil edilmesidir. Birimlerin rassal ve tekrarlı örneklemelerinin yer aldığı çalışmalarda, her bir birime özgü gözlenemeyen birim etkinin rassal olduğu bilinmektedir. Bu açı-

¹ Çukurova Üniversitesi, İktisadi ve İdari Bilimler Fakültesi, Ekonometri Bölümü.
E-mail: csigeze@cu.edu.tr

dan, panel veriler ile oluşturulan çok durumlu logit ve probit modellerin tahmini rassal etkilerin varlığı altında geliştirilmiştir.

13.2. Rassal Etkiler Çok Durumlu Probit Model

Belli zaman periyodu boyunca aynı birimlerin (bireyler, haneler, firmalar, ülkeler vb.) tekrarlı ölçümlerinden oluşan panel verinin kullanıldığı ve bağımlı değişkenin ikiden fazla, nitel ve herhangi bir sıra içermediği durumda panel çok durumlu probit (multinomial multiperiod probit model, MMNP) model kullanılabilir. Bilindiği üzere çok durumlu probit model fayda maksimizasyonu teorisine dayanmaktadır. i. bireyin t. zamanda j. alternatifi seçmesinin fayda fonksiyonu aşağıdaki şekilde yazılmaktadır.

$$U_{ijt} = X'_{ijt} \beta_j + \varepsilon_{ijt} \quad (j=1,2,\dots,J \quad t=1,2,\dots,T) \quad (13.1)$$

Burada X_{ijt} açıklayıcı değişkenler matrisi, β_j katsayılar vektörü olmak üzere hata terimi $\varepsilon_{ijt} \sim IIDN(0, \Sigma)$ şeklinde gösterilmektedir. Tercih modellerinde faydanın değeri ölçülemediği için alternatiflerin sağladığı faydalar arasındaki fark hesaplanarak, diğer alternatife göreli olarak seçilen alternatifi faydası belirlenebilmektedir. Ancak panel çok durumlu probit model tahminlerinin yapılması için hata teriminin varyansının ilk köşegen elemanının ($\sigma_{11}=1$) bire eşit olması gereklidir. Bu amaçla model hata terimleri ile normalleştirilir. Son olarak model eşitlik (13.2)-eşitlik (13.4) ile gösterilmektedir.

$$U_{ijt}^* = (U_{ijt} - U_{i1t})(\sigma_{11} + \sigma_{jj} - 2\sigma_{1j})^{-1/2} \quad (13.2)$$

$$= [(X'_{ijt} \beta_j - X'_{i1t} \beta_1) + (\varepsilon_{ijt} - \varepsilon_{i1t})](\sigma_{11} + \sigma_{jj} - 2\sigma_{1j})^{-1/2} \quad (13.3)$$

$$= X_{ijt}^{**} \beta_j^* + \varepsilon_{ijt}^* \quad (i=1,2,\dots,N) \quad (t=1,2,\dots,T) \quad (13.4)$$

Hata terimlerinin $\varepsilon_i^* \sim IIDN(0, \Sigma^*)$ dağılımlı ve $U_{i1t}^* = 0$ ve $\varepsilon_{i1t}^* = 0$ olduğu varsayılmaktadır. Dönüştürülmüş modelden yola çıkılarak i. bireyin t. zamanda j. alternatifi seçtiği durumda fayda fonksiyonu eşitlik (13.5) ile gösterilmektedir.

$$P(U_{ijt}^* > U_{ikt}^*) = (\varepsilon_{ijt}^* - \varepsilon_{ikt}^* > X_{ikt}^{**} \beta_k^* - X_{ijt}^{**} \beta_j^*) \quad k \neq j (j=1,2,\dots,J-1) \quad (13.5)$$

Hata terimi ε_{ijt}^* 'nin ardışık olarak bağımsız olması nedeniyle J-1 seçimin faydasının hesaplanması için her bir T boyutlu J-1 tane integralin yani $T \times (J-1)$ tane integralin çözülmesi gerekmektedir. Panel çok durumlu probit modellerde alternatiflerin sayısı arttıkça integrallerin hesaplanmasında yaşanan problemlerden kaçınmak için literatürde farklı simülasyon yöntemleri kullanılmıştır. Panel çok durumlu probit modelin tahmini için GHK (Geweke-Hajivassiliou-Keane) simülasyon yöntemine dayanan SML (simulated maximum likelihood) ve MSM (method of simulated moment) yöntemleri ve Monte-Carlo simülasyon tekniğine dayanan bayesyen (bkz. Rossi ve Allenby, 1993; Geweke, Keane ve Runkle, 1997) yaklaşımlarından yararlanılmaktadır. Bu simülasyon tahmin yöntemleri GAUSSX ve LIMDEP gibi bazı yazılım paketlerinde uygulanabilmektedir. Panel çok durumlu probit modelin, Stata paket programında analizinin yapılamaması nedeniyle kitapta uygulamasına yer verilememiştir.

13.3. Rassal Etkiler Çok Durumlu Logit Model

McFadden(1974) tarafından önerilen çok durumlu logit modelin panel veri ile geliştirilmesi ile bağımlı değişkeni oluşturan tercihlerin (alternatif, kategori) her birinin gerçekleşme olasılıkları birimlerin tekrarlı gözlemleri dikkate alınarak hesaplanabilmektedir. Panel çok durumlu logit modeli oluşturmak amacıyla kullanılan regresyon modeli aşağıdaki gibidir.

$$Y_{ijt} = \alpha_j + x'_{ijt}\beta + \varepsilon_{ijt} \quad (13.6)$$

Y_{ijt} , birey i 'nin t zamanında alternatif j 'yi seçmesini, x_{ijt} bireylere ait bilgileri içeren açıklayıcı değişken matrisini, α_j alternatif j 'nin seçilmesine özgü sabit terimi ifade etmektedir. Hata terimi matrisi ε_{ijt} ise birbirinden bağımsız logistik dağılımlıdır. Panel çok durumlu logit modele göre i . bireyin t zamanında j alternatifini seçmesinin olasılığı eşitlik (13.7) ile gösterilmektedir (Sarma ve Simpson, 2007)

$$P_{ijt} = P(y_{it} = j) = \frac{\exp(\alpha_j + x'_{ijt}\beta)}{\sum_{j=1}^J \exp(\alpha_j + x'_{ijt}\beta)} \quad (13.7)$$

Çok durumlu logit modelde hataların bağımsızlığı varsayımı, panel veri ile çalışıldığında hataların zamanla bağımsız olması varsayımını da gerektirmektedir. Buna göre gözlenemeyen faktörlerin aynı bireylerin tekrarlı gözlemleri üzerinde bağımsız olduğu varsayımı geçerli olmaktadır (Train, 2003, s.56). Ancak bu kısıt, bireylerin şimdiki tercihlerinin geçmişteki tercihlerinden etkilenmesi gibi durumlarda çok gerçekçi olmamaktadır. Bu kısıtı genişletmek yani aynı bireylerin tekrarlı gözlemleri arasındaki korelasyonlu hatalarını dikkate almak için seçimler arasında rassal olarak değişen bazı parametrelerin varlığı gerekmektedir. Bu fikir, aynı bireylerin tekrarlı gözlemleri için tercih olasılıklarının aynı rassal etkiyi paylaştıkları rassal etkiler çok durumlu logit modelini ortaya çıkarmaktadır (Train, 2003, s.55; Sarma ve Simpson, 2007). Buna göre rassal etkiler çok durumlu logit model eşitlik (13.8) ile gösterilmektedir.

$$Y_{ijt} = \alpha_j + x'_{ijt}\beta + z'_{ijt}u_i + \varepsilon_{ijt} \quad (13.8)$$

Burada Y_{ijt} , birey i 'nin t . zamanda alternatif j 'yi seçmesini, x_{ijt} bireylere ait bilgileri içeren açıklayıcı değişken matrisini, α_j alternatif j 'nin seçilmesine özgü sabit terimi, u_i , model içerisinde bireylere özgü gözlenemeyen birim heterojenliği göstermektedir. z_{ijt} ise bilinen dizayn vektörünü göstermektedir. Böylece, aynı birimlerin tekrarlı gözlemlerinin olasılıklarının, aynı gözlenemeyen rassal etkiyi içerdiği rassal etkiler çok durumlu logit model aşağıda verilmektedir (Chen ve Kuo, 2001; Sarma ve Simpson, 2007).

$$P_{ijt}(Y_{ijt} = j | u_i) = \frac{\exp(\alpha_j + x'_{ijt}\beta + z'_{ijt}u_i)}{\sum_{j=1}^J \exp(\alpha_j + x'_{ijt}\beta + z'_{ijt}u_i)} \quad (13.9)$$

Burada, $z_{ijt} = 1$ durumunda rassal etkiler çok durumlu logit modelde, aynı birimlerin tekrarlı gözlemleri üzerinde aynı gözlenemeyen faktörlerin etkisi yalnızca sabit terim ile modellenmek-

tedir. Bunun yanında, aynı bireylerin tekrarlı gözlemleri arasındaki korelasyonlu hatalarını dikkate almak için eğim katsayısının değiştiği ($z_{ij} = x_{ij}$) ve hem sabit hem de eğim katsayısının değiştiği ($z_{ij} = (1, x'_{ij})'$) modeller de kullanılmaktadır (Chen ve Kuo, 2001). Bu modeller literatürde rassal parametreler logit modeli ya da karma logit modeli olarak bilinmektedir (Sarma ve Simpson, 2007; Train, 2003, s.56). Daha önce de bahsedildiği gibi rassal etkilerin yer aldığı bağımlı değişkeni nitel modeller genelleştirilmiş doğrusal karma modeller (GLMM) olarak da bilinmektedir (McCulloch ve Searle, 2001, s.2). GLMM'lerde en çok olabilirlik tahminlerinin sağlanması için tepkinin marjinal dağılımının hesaplanması gerekmektedir. Ancak bu dağılımın kapalı formda çözümü olmadığı için rassal etkilere göre integrallerin alınabilmesi amacıyla sayısal tümleştirme (numerical integration) yöntemlerinden yararlanılmaktadır². Bunun yanında, rassal etkiler çok durumlu logit modelin tahmininde marjinal olabilirlik hesaplamasında integrallerin çözülmesinde yaşanan sayısal problemlere çözüm olarak simülasyon tabanlı MSL yaklaşımı ve bayesyen yaklaşım önerilmiştir. MSL yaklaşım, veri setinde yer alan birimlerin sayısının çok olması durumunda oldukça fazla hesaplama süresi gerektirirken, bayesyen yaklaşım hesaplama kolaylığı ve zaman açısından oldukça avantaj sağlamaktadır (Haynes, vd., 2008).

13.3.1. Rassal Etkiler Çok Durumlu Logit Model Stata Uygulaması

Örnek 13.1. Rassal etkiler çok durumlu logit modelin Stata paket programında analizi için **dinamikyoksulluk.dta** veri setinden yararlanılacaktır. Bu veri seti, Örnek 12.1'de yer alan panel veri setini içerdiği için hanenin yoksulluk durumunu gösteren bağımlı değişkenden yararlanılarak yeni bir bağımlı değişken oluşturulmuştur. Bu bağımlı değişken, bir önceki yıla göre hanenin yoksulluk durumunu oluşturmaktadır. Öyle ki; hane bir önceki dönemde yoksul, bir sonraki dönemde yoksul değil ise "yoksulluktan çıkan(2)", bir önceki dönemde yoksul değil ancak bir sonraki dönemde yoksul ise "yoksulluğa giren(3)" olarak tanımlanmaktadır³. Hane her iki dönemde de yoksul değil ise "yoksul olmayan(1)" olarak tanımlanırken üst üste ya da daha sonraki dönemde de yoksul ise analiz edilen sürenin yarısında yoksul olduğu için "kronik yoksul(4)" olarak değerlendirilmektedir⁴. Çalışmada, hanelerin bir önceki yoksulluk durumlarına göre bağımlı değişkenin kategorileri oluşturulduğu için panel verinin ilk yılı olan 2009 yılına ait bağımlı değişken oluşturulamamaktadır. Böylece 2009 yılına ait gözlemler göz ardı edilerek geriye kalan üç yıllık dönemde haneye ve hanehalkı reisine ilişkin açıklayıcı değişkenler kullanılmıştır.

Stata paket programında veri dosyasını çağırabilmek amacıyla aşağıdaki komuttan yararlanılmaktadır.

use "dosyakonumu\dinamikyoksulluk.dta"

² Bkz. Hartzel, Agresti ve Caffo (2001); Haan ve Uhlenborff (2006)

³ Yoksulluktan çıkan ve yoksulluğa giren haneler genel olarak geçici yoksul haneler olarak da tanımlanabilmektedir (bkz. Bane ve Ellwood (1986)).

⁴ Bağımlı değişkenin oluşturulmasında Bane ve Ellwood (1986) ve Dartanto ve Nurkholis (2013)'ün çalışmalarından yararlanılmıştır.

Stata paket programında değişkenlerin tanımlarını listelemek için aşağıda verilen “describe” komutu kullanılmaktadır.

describe

Contains data from C:\dosyakonumu\dinamikyoksulluk.dta

obs: 7,500

vars: 10

size: 165,000

variable name	storage type	display format	value label	variable label
time	int	%8.0g		
id	long	%12.0g		
yas	int	%8.0g		
yas2	int	%8.0g		
cinsiyet	byte	%8.0g		
universite	byte	%8.0g		
gelir	float	%8.0g		
lngelir	float	%8.0g		
yoksulluk	byte	%8.0g		
yoksulluk2	byte	%8.0g		

Çalışmada kullanılacak değişkenlere dair bilgiler Tablo 13.1’de yer almaktadır. Panel veri, birimlerin belli bir zaman periyodu boyunca tekrarlı gözlemlerinden oluşmaktadır. Bu çalışmada birimler (id) 2500 haneyi, zaman bileşeni (time) ise 2010, 2011 ve 2012 yıllarını göstermektedir. Tüm birimler, tüm zamanlar boyunca gözlemlendiği için dengeli panel veri oluşmaktadır.

Tablo 13.1. Örnekte Kullanılan Değişkenler ve Tanımları

Değişkenler	Tanım	Açıklama
id	Hane numaraları	2500 hane
time	Yıllar	2010, 2011, 2012
yoksulluk2	Bağımlı değişken ⁵	Hane “yoksul olmayan” ise 1 (Referans)
		Hane “yoksulluktan çıkan” ise 2
		Hane “yoksulluğa giren” ise 3
		Hane “kronik yoksul” ise 4
cinsiyet	Açıklayıcı değişken	Hanehalkı reisi kadın ise 1, değil ise 0.
yas	Açıklayıcı değişken	Hanehalkı reisinin yaşı
yas2	Açıklayıcı değişken	Hanehalkı reisinin yaşının karesi
universite	Açıklayıcı değişken	Hanehalkı reisi yüksek okul, fakülte ve üzeri mezunu ise 1, değil ise 0.
lngelir	Açıklayıcı değişken	Hanenin eşdeğer hanehalkı kullanılabilir fert gelirinin logaritması

⁵ Literatürde bağımlı değişkenin yoksul, geçici yoksul ve kronik yoksul olarak ayrıştırıldığı çalışmalarda model yapısını sıralı logit model ya da çok durumlu logit model seçen farklı çalışmalar yer almaktadır. Bkz. McCulloch ve Baulch (1999); Alisjahbana ve Yusuf (2003); Dartanto ve Nurkholis (2013).

Rassal etkiler çok durumlu logit modeli de kapsayan genelleştirilmiş doğrusal karma modellerin ve genelleştirilmiş doğrusal gizli ve karma modeller (Generalized Linear Latent And Mixed Models, GLLAMM) olarak bilinen modellerin tahmini Stata paket programında Rabe-Hesketh, Skrondal ve Pickles (2004) tarafından geliştirilen “gllamm” paketi ile sağlanabilmektedir. Stata paket programında “gllamm” komutunu çalıştırmadan önce “help” menüsü “stata command” kısmına “gllamm” yazılarak paket indirilebilir. Alternatif olarak aşağıdaki komuttan yararlanılarak “gllamm” paketi indirilebilir.

ssc install gllamm

Örnekteki veri seti anakütleden rassal olarak seçilen hanelerden oluştuğu için birim etkilerin rassal olduğu varsayımı altında rassal etkiler çok durumlu logit model için Rabe-Hesketh, Skrondal ve Pickles (2004) tarafından geliştirilen, Haan ve Uhlenborff (2006) tarafından düzenlenen Stata paket programı kodları kullanılmaktadır⁶. Stata paket programına verinin panel yapısı, Haan ve Uhlenborff (2006) tarafından düzenlenen kodlar ile tanıtılmaktadır.

Stata paket programında “gllamm” komutu ile rassal etkiler çok durumlu logit model tahmini sonuçları Tablo 13.2’de yer almaktadır. Burada rassal etkilere göre integral hesaplanması amacıyla uyarlamalı kuadratik (adaptive quadrature) ve Newton-Raphson yaklaştırma yöntemleri kullanılmaktadır.

Tablo 13.2. Rassal Etkiler Çok Durumlu Logit Model Tahmini – Stata Programı Sonucu

gllamm model

log likelihood = -4186.3644

alt	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
sabit2	11.13567	1.16059	9.59	0.000	8.860959	13.41039
sabit3	41.31099	1.609276	25.67	0.000	38.15686	44.46511
sabit4	47.59726	1.416514	33.60	0.000	44.82094	50.37357
model2cinsiyet	-.2051527	.143282	-1.43	0.152	-.4859802	.0756748
model3cinsiyet	-1.138887	.1691913	-6.73	0.000	-1.470496	-.8072777
model4cinsiyet	-1.281841	.1373425	-9.33	0.000	-1.551027	-1.012655
model2lmgelir	-1.46787	.1187242	-12.36	0.000	-1.700566	-1.235175
model3lmgelir	-5.037351	.1711317	-29.44	0.000	-5.372763	-4.701939
model4lmgelir	-5.606185	.1542534	-36.34	0.000	-5.908516	-5.303854
model2yas	.0526571	.0245691	2.14	0.032	.0045026	.1008116
model3yas	.2022043	.0276759	7.31	0.000	.1479606	.256448
model4yas	.2035676	.0219067	9.29	0.000	.1606312	.246504
model2yas2	-.0006327	.0002326	-2.72	0.007	-.0010886	-.0001767
model3yas2	-.0021429	.0002595	-8.26	0.000	-.0026515	-.0016344
model4yas2	-.0022114	.0002071	-10.68	0.000	-.0026173	-.0018056
model2universite	-2.046565	.4568259	-4.48	0.000	-2.941927	-1.151203
model3universite	-.4752708	.48581	-0.98	0.328	-1.427441	.4768993
model4universite	-2.188611	.7378237	-2.97	0.003	-3.634719	-.7425035

⁶ Stata paket programı kodları bölüm sonunda yer almaktadır.

Variances and covariances of random effects

```

-----
***level 2 (time)
var(1): .00272353 (.01826208)
cov(2,1): -.01256576 (.04803532) cor(2,1): -.39440138
var(2): .37270713 (.39228868)
cov(3,1): -.01237618 (.04545166) cor(3,1): -.46105761
cov(3,2): .30410064 (.35489639) cor(3,2): .96843036
var(3): .26456365 (.32016105)
-----

```

Tablo 13.2'ye göre sabit2, sabit3 ve sabit4 terimleri, sırasıyla yoksul olmayan hanelere göre yoksulluktan çıkan, yoksulluğa giren ve kronik yoksul haneler için hesaplanan modellerin sabit terimlerini göstermektedir. model2, model3 ve model4 ile başlayan ve açıklayıcı değişken isimleri ile devam eden kavramlar ise sırasıyla yoksul olmayan hanelere göre yoksulluktan çıkan, yoksulluğa giren ve kronik yoksul haneler için hesaplanan modellerdeki değişkenlerin katsayı tahminlerini vermektedir.

Tablo 13.2'ye göre yoksulluktan çıkan, yoksulluğa giren ve kronik yoksul hanelerin rassal birim etkilerinin varyansları sırasıyla 0.002, 0.372 ve 0.264 bulunmuştur. Rassal birim etkiler, modelde açıklayıcı değişkenler ile dikkate alınmayan, gözlenemeyen etkileri içermek üzere rassal birim etkilerin varyans değerleri büyüdükçe gözlenemeyen rassal etkilerin daha değişken olduğu söylenebilir. Bu durumda, özellikle “yoksulluğa giren” ve “kronik yoksul” hanelerin, açıklayıcı değişkenler ile gözlenemeyen, kendilerine özgü bir heterojenliği bulunmaktadır. Modelde gözlenemeyen rassal birim etkilerin varlığı yani birim içi heterojenliğin derecesi en düşük 0.002 varyans değeri ile yoksulluktan çıkan haneler için hesaplanmıştır. Yani, yoksulluktan çıkan hanelerin gözlenemeyen rassal birim etkileri oldukça azdır.

“Yoksul olmayan” hanelere göre oluşturulan üç model için rassal birim etkilerin korelasyonlarına bakıldığında ise yoksulluktan çıkan ve yoksulluğa giren hanelerin gözlenemeyen özellikleri arasında 0.394 değerinde negatif korelasyon bulunmuştur. Bu bulguyla tutarlı olarak, yoksulluktan çıkan ve kronik yoksul hanelerin rassal etkileri arasında ise 0.461 değerinde negatif bir ilişki gözlenmiştir. Teorik beklentiye göre yoksulluğa giren haneler ile kronik yoksul hanelerin gözlenen karakterlerinin yanında gözlenemeyen rassal özellikleri ortak olmaktadır. Teorik beklentiyle tutarlı olarak, yoksulluğa giren ve kronik yoksul hanelerin rassal birim etkileri arasında 0.968 değerinde güçlü bir pozitif ilişki bulunmuştur. Bu bulgu, hanelerin yoksulluğa düşmelerinde ve kronik yoksul olmalarında hanelerin gözlenemeyen özelliklerinin benzer olduğunu göstermektedir.

Tablo 13.2'de yer alan Stata paket programında gerçekleştirilen rassal etkiler çok durumlu logit model tahmin sonuçlarına göre modelde yer alan açıklayıcı değişkenlerin önemli bir kısmı istatistiksel olarak anlamlı bulunmuştur. Hanehalkı reisi kadın olan hanelerin yoksul olmamalarına göre yoksulluğa girmeleri ve kronik yoksul olma olasılıkları hanehalkı reisi erkek olan hanelere göre daha düşük bulunmuştur. Bunun yanında, hanehalkı reisinin kadın olmasının hanenin yoksul olmasına göre yoksulluktan çıkması üzerinde istatistiksel olarak anlamlı bir etkisi bulunmamaktadır.

Hane gelirinin beklenildiği gibi hanenin yoksul olmasına göre hanenin yoksulluğa girmesi ve kronik yoksul olması üzerindeki etkisi negatiftir. Hanehalkı reisinin yaşının hanenin yoksul olmamasına göre yoksulluktan çıkan, yoksulluğa giren ve kronik yoksul olma olasılıkları üzerinde pozitif, hane-

halkı reisinin yaşının karesinin ise negatif etkisi bulunmaktadır. Bu bulgular, hanehalkı reisinin yaşı arttıkça hanelerin geçici yoksul (yoksulluktan çıkan ve/veya yoksulluğa giren) ve kronik yoksul olma olasılıklarının arttığını bununla birlikte belli bir yaştan sonra ise yaş artışı ile birlikte hanelerin geçici ve kronik yoksul olma olasılıklarının azaldığını göstermektedir. Bunun yanında, yoksul olmayan hanelerin hanehalkı reisinin üniversite mezunu olması ile yoksulluğa giren hanelerin hanehalkı reisinin üniversite mezunu olması arasında da anlamlı bir fark bulunmamaktadır.

Stata paket programında rassal etkiler çok durumlu logit modelin en çok olabilirlik yöntemi ile tahmininin oldukça fazla zaman gerektirmesi nedeniyle Markov Zinciri Monte Carlo (MCMC) simülasyon tekniğini kullanarak rassal birim etkilere göre integrallerin hesaplanmasını sağlayan bayesyen yöntem uygun bir alternatif sunmaktadır. Rassal etkiler çok durumlu logit modelin katsayı tahminlerinin bayesyen tahmini için R paket programında Hadfield (2010) tarafından MCMC tekniği kullanılarak GLMM'leri tahmin etmek için geliştirilen "MCMCglmm" isimli paketten yararlanılmaktadır.

Rassal etkiler çok durumlu logit modelin tahmini için R programı kodları aşağıdaki gibidir⁷.

```
library(MCMCglmm)
library(nnet)
library(parallel)
setwd("C:/Users/dosyakonumu")
dinamikyoksulluk<-read.csv("dinamikyoksulluk.csv",header=TRUE, sep=";")
yoksulluk2<-dinamikyoksulluk$yoksulluk2
lngelir<-dinamikyoksulluk$lngelir
yas<-dinamikyoksulluk$yas
yas2<-dinamikyoksulluk$yas2
universite<-dinamikyoksulluk$universite
cinsiyet<-dinamikyoksulluk$cinsiyet
k <- 4
I <- diag(k-1)
J <- matrix(rep(1,(k-1)^2),c(k-1,k-1))
priors = list(R = list(fix = 1, V = (I+J), n = 3),
              G = list(G1 = list(V = diag(3), n =3)))
analiz <-mclapply(1:2, function(i){
  MCMCglmm(yoksulluk2 ~ -1+trait+trait: (cinsi-
yet+lngelir+yas+yas2+universite),
  random = ~ us(trait):id,
  rcov = ~ idh(trait):units,
  prior = priors,verbose=TRUE,
  nitt = 250000, thin = 1000, burnin =50000,
  family = "categorical",data = dinamikyoksulluk)},mc.cores=1)
analiz <-lapply(analiz,function(m)m$Sol)
analiz <-do.call(mcmc.list,analiz)
summary(analiz)
gelman.diag(analiz)
```

⁷ R komutlarında "k" bağımlı değişkenin olduğu kategori sayısını belirtmek üzere değişkenlerin isimleri değiştirilerek, rassal etkiler çoklu logit model için R kodları kolaylıkla çalışmanız için uyarlanabilir.

R programına verinin yüklenmesi için öncelikle excel dosyasının .csv formatında çalışma klasörüne kaydedilmesi gerekmektedir. Ardından, `setwd()` fonksiyonu kullanılarak veri dosyasının konumu belirtilerek ve `read.csv()` fonksiyonu kullanılarak `dinamikyoksulluk.csv` isimli veri dosyası “dinamikyoksulluk” ismi ile R programına aktarılmaktadır.

R paket programında MCMC simülasyon yöntemi ile çoklu logit modelin tahmini için öncelikle “MCMCglmm” paketinin yüklenmesi gerekmektedir. R paket programında paketler daha önce yüklenmiş ise kütüphaneden (`library(MCMCglmm)`) gibi çağrılması yeterli olacaktır. MCMC simülasyon yönteminde, markov zincirlerinin oluşturulması amacıyla kullanılan algoritmaların iterasyon sayısının (nitt) fazla olması güvenilir sonuçların elde edilmesini sağlamaktadır. Bunun yanında parametrelerin başlangıç değerlerinin etkisinden arındırılması amacıyla yakınsamanın sağlandığı “burnin” sayısı ve algoritmada ardışık gözlemler arasındaki korelasyonun etkisinden kurtulmak amacıyla kullanılan “thin” sayısı da MCMC simülasyon yöntemlerinde oldukça önemlidir. Oluşturulan algoritmada farklı başlangıç noktaları ile birden fazla zincir oluşturularak iterasyon sayısı arttırılabilmektedir. Bu amaçla “mclapply” isimli fonksiyon oluşturan komuttan yararlanılmakta ve iki zincir için algoritma tekrarlanmaktadır. Çok durumlu logit modelin analizi için oluşturulan fonksiyonunun hesaplanması amacıyla “parallel” ve “nnet” isimli paketlerin yüklenmesi gerekmektedir.

Panel çok durumlu logit modelin R programında MCMC simülasyon yöntemi ile tahmini için Metropolis-Hastings algoritması kullanılmaktadır. R kodlarında belirtildiği gibi 250 bin iterasyonun belirlendiği algoritmada yakınsamanın gerçekleştiği iterasyon sayısı (burn-in period) 50 bin olarak belirlenmiştir. Böylelikle ilk 50 bin iterasyon sayısı, toplam iterasyon sayısından çıkarılarak parametrelerin başlangıç değerlerinin etkisi arındırılmıştır. Bunun yanında algoritmada ardışık gözlemler arasındaki olası korelasyonun etkisini yok etmek amacıyla seyretme (thinning) sayısı ise 1000 alınmıştır. Buna göre yalnızca bininci iterasyon kaydedilmekte ve böylece otokorelasyon sorunu da ortadan kaldırılmış olmaktadır. Böylece bir önceki değerin bir sonraki değeri etkilemediği daha güvenilir tahminler elde edilmiştir. MCMC yöntemindeki zincirlerin yakınsadığının test edilmesi amacıyla Gelman-Rubin (1992) yakınsama testi, Heidelberger-Welch (1992) yakınsama testi ve Geweke(1992) testi kullanılmaktadır. Bu analiz için Gelman-Rubin (1992) yakınsama testi kullanılmıştır. Gelman-Rubin istatistiğine göre yakınsamanın sağlanması için hesaplanan potansiyel ölçeği azaltma fonksiyonu (potential scale reduction function, Psrf) değerinin ve her bir katsayı için nokta tahminlerinin 1.2’den küçük olması gerekmektedir. Yapılan test sonucunda Psrf değeri ve bütün katsayılar için nokta tahmini 1.2’den küçük çıkmıştır ve yakınsama sağlanmıştır.

R programında rassal etkiler çok durumlu logit modelin MCMC simülasyon yöntemi ile tahmini veren çıktı, bayesyen yaklaşıma göre tahmin edilen parametrelerin dağılım ortalamasını, standart sapmasını ve her bir değişken için %2.5, %25, %50, %75 ve %97.5 kantil değerlerini vermektedir. Buna göre katsayıların istatistiksel olarak anlamlı olup olmadıkları %2.5 ve %97.5 kantil değerleri dikkate alınarak yalnızca %5 önem düzeyinde incelenebilmektedir. Rassal etkiler çok durumlu logit modelin MCMC tekniği ile elde edilen sonuçları Tablo 13.3’te yer almaktadır.

Tablo 13.3. Rassal Etkiler Çok Durumlu Logit Model Tahmini – R Programı Sonucu

Iterations = 50001:249001

Thinning interval = 1000

Number of chains = 2

Sample size per chain = 200

Empirical mean and standard deviation for each variable,
plus standard error of the mean:

	Mean	SD	Naive SE	Time-series SE
Traityoksulluk2.2	14.7145091	1.6667730	8.334e-02	8.340e-02
traityoksulluk2.3	49.0709441	2.2343287	1.117e-01	1.361e-01
traityoksulluk2.4	58.1000831	2.0694428	1.035e-01	1.036e-01
traityoksulluk2.2:cinsiyet	-0.2798245	0.1695844	8.479e-03	7.970e-03
traityoksulluk2.3:cinsiyet	-1.3264935	0.1934680	9.673e-03	1.043e-02
traityoksulluk2.4:cinsiyet	-1.5807307	0.1757835	8.789e-03	8.903e-03
traityoksulluk2.2:lngelir	-1.9188520	0.1352548	6.763e-03	7.200e-03
traityoksulluk2.3:lngelir	-5.9580661	0.2049717	1.025e-02	1.325e-02
traityoksulluk2.4:lngelir	-6.8059012	0.1812275	9.061e-03	9.190e-03
traityoksulluk2.2:yas	0.0648254	0.0281062	1.405e-03	1.399e-03
traityoksulluk2.3:yas	0.2311015	0.0336776	1.684e-03	1.686e-03
traityoksulluk2.4:yas	0.2366766	0.0273763	1.369e-03	1.368e-03
traityoksulluk2.2:yas2	-0.0007845	0.0002635	1.317e-05	1.313e-05
traityoksulluk2.3:yas2	-0.0024517	0.0003153	1.577e-05	1.578e-05
traityoksulluk2.4:yas2	-0.0025637	0.0002577	1.289e-05	1.287e-05
traityoksulluk2.2:universite	-2.3552421	0.5280948	2.640e-02	4.816e-02
traityoksulluk2.3:universite	-0.6784598	0.5345219	2.673e-02	3.250e-02
traityoksulluk2.4:universite	-2.9708887	1.0099954	5.050e-02	1.498e-01

2. Quantiles for each variable:

	2.5%	25%	50%	75%	97.5%
traityoksulluk2.2	11.561462	13.5781994	14.7051105	15.7596184	17.8677771
traityoksulluk2.3	44.870778	47.5131118	49.0100171	50.4204810	53.6435105
traityoksulluk2.4	54.315859	56.8972440	57.9764165	59.2934178	62.4074284
traityoksulluk2.2:cinsiyet	-0.639341	-0.3863085	-0.2875398	-0.1713239	0.0620152
traityoksulluk2.3:cinsiyet	-1.721730	-1.4554139	-1.3131725	-1.2015966	-0.9468658
traityoksulluk2.4:cinsiyet	-1.889805	-1.7106787	-1.5875078	-1.4644924	-1.2241745
traityoksulluk2.2:lngelir	-2.170769	-2.0080528	-1.9222443	-1.8216286	-1.6733413
traityoksulluk2.3:lngelir	-6.364130	-6.0968008	-5.9609311	-5.8217839	-5.5522380
traityoksulluk2.4:lngelir	-7.165415	-6.9114921	-6.7965908	-6.6899936	-6.4419086
traityoksulluk2.2:yas	0.010144	0.0458109	0.0666294	0.0838582	0.1183077
traityoksulluk2.3:yas	0.167827	0.2076323	0.2312095	0.2515408	0.2980732
traityoksulluk2.4:yas	0.182659	0.2173441	0.2369435	0.2554149	0.2895356
traityoksulluk2.2:yas2	-0.001299	-0.0009637	-0.0007941	-0.0006072	-0.0002645
traityoksulluk2.3:yas2	-0.003116	-0.0026481	-0.0024481	-0.0022219	-0.0018876
traityoksulluk2.4:yas2	-0.003063	-0.0027370	-0.0025527	-0.0023931	-0.0020643
traityoksulluk2.2:universite	-3.593076	-2.6621014	-2.3062026	-1.9862321	-1.4831308
traityoksulluk2.3:universite	-1.689752	-1.0370337	-0.6601501	-0.3140125	0.3676307
traityoksulluk2.4:universite	-5.357243	-3.4989806	-2.8044893	-2.2695121	-1.3912662

Rassal etkiler çok durumlu logit modelin MCMC simülasyon yöntemine tahmininde “yoksulluk2” olarak tanımlanan bağımlı değişkenin birinci kategorisi olan hanenin “yoksul olmayan”

olması referans kategori olarak belirlenmiştir. Buna göre Tablo 13.3'te yer alan traityoksulluk2.2, traityoksulluk2.3 ve traityoksulluk2.4 sırasıyla bağımlı değişkenin “yoksulluktan çıkan”, “yoksulluğa giren” ve “kronik yoksul” olmasını ifade etmektedir. Tablo 13.3'ün ilk üç satırında belirtilen tahminler, bağımlı değişkenin referans kategorisine göre oluşturulan modellerin sabit terimlerini göstermektedir. Benzer şekilde, hanenin “yoksul olmayan” olmasına göre “yoksulluktan çıkan”, “yoksulluğa giren” ve “kronik yoksul” olması olasılığı üzerinde açıklayıcı değişkenlerin etkisi sırasıyla Tablo 13.3'te yer almaktadır.

Tablo 13.3'te yer alan rassal etkiler çok durumlu logit modelin MCMC tekniğine dayanan bayesyen yaklaşım ile hesaplanan katsayıları, Tablo 13.2'de yer alan rassal etkiler çok durumlu logit modelin en çok olabilirlik yöntemiyle hesaplanan katsayı büyüklükleri, katsayı işaretleri ve katsayıların anlamlılıkları benzerlik göstermektedir. En çok olabilirlik ve bayesyen yöntem ile elde edilen katsayılar arasında önemli bir fark olmamasının nedeni bayesyen yöntemde parametrelere ilişkin önsel dağılımın bilgi içermeyen önsel dağılım olmasıdır. Rassal etkiler çok durumlu logit modelin tahmininde, her iki programdan elde edilen katsayı büyüklükleri, işaretleri ve anlamlılıkları benzerlik göstermekle birlikte Stata paket programı R programına göre analiz için daha fazla zaman gerektirmektedir.

Bu bölümde kullanılan rassal etkiler çok durumlu logit modelin tahmini için Stata komutları aşağıda yer almaktadır.

```
use "dosyakonumu\dinamikyoksulluk.dta", clear
sort id time
gen patt=_n
expand 4
sort patt
qui by patt: gen alt=_n
gen seçim=alt==yoksulluk2
tab alt, gen(sabit)
eq sabit2: sabit2
eq sabit3: sabit3
eq sabit4: sabit4
gen model2cinsiyet = sabit2*cinsiyet
gen model3cinsiyet = sabit3*cinsiyet
gen model4cinsiyet = sabit4*cinsiyet
gen model2lmgelir = sabit2*lmgelir
gen model3lmgelir= sabit3*lmgelir
gen model4lmgelir= sabit4*lmgelir
gen model2yas = sabit2*yas
gen model3yas= sabit3*yas
gen model4yas= sabit4*yas
```

```

gen model2yas2 = sabit2*yas2
gen model3yas2= sabit3*yas2
gen model4yas2= sabit4*yas2
gen model2üniversite = sabit2*üniversite
gen model3üniversite= sabit3*üniversite
gen model4üniversite = sabit4*üniversite

gllamm alt sabit2 sabit3 sabit4 model2cinsiyet model3cinsiyet model4cinsiyet mo-
del2lmgelir model3lmgelir model4lmgelir model2yas model3yas model4yas model2yas2 mo-
del3yas2 model4yas2 model2üniversite model3üniversite model4üniversite, nocons
expand(patt seçim o) i(time) /*

*/ lin(mlogit) family(binom) nrf(3) eqs(sabit2 sabit3 sabit4) nip(4) /*

*/ adapt

```

13.4. KAYNAKLAR

- Alisjahbana, A., & Yusuf, A. A. (2003). Poverty dynamics in Indonesia: panel data evidence. *Department of Economics, Padjadjaran University*.
- Bane, M. J. & Ellwood, D. T. (1986). Slipping Into and Out of Poverty: The Dynamics of Spells." *Journal of Human Resources*, 1986, 21(1), pp. 1-23.
- Cameron, A. C., & Trivedi, P. K. (2005). *Microeconometrics: methods and applications*. Cambridge university press.
- Chen, Z., & Kuo, L. (2001). A note on the estimation of the multinomial logit model with random effects. *The American Statistician* 55(2), 89-95.
- Croissant, Y. (2012). Estimation of multinomial logit models in R: The mlogit Packages. R package version 0.2-2. URL: <http://cran.r-project.org/web/packages/mlogit/index.html>.
- Dartanto, T., & Nurkholis (2013). The determinants of poverty dynamics in Indonesia: evidence from panel data. *Bulletin of Indonesian Economic Studies*, 49(1), 61-84.
- Gelman, A., & Rubin, D. B. (1992). A single series from the Gibbs sampler provides a false sense of security. *Bayesian statistics*, 4, 625-631.
- Geweke, J. (1992). Bayesian comparison of econometric models. *Federal Reserve bank of Minneapolis working paper*, 532.
- Geweke, J., Keane, M., & Runkle, D. (1997). Statistical Inference in the Multinomial Multiperiod Probit Model. *Research Department Staff Report 177*. Federal Reserve Bank of Minneapolis.
- Haan, P., & Uhlenborff, A. (2006). *Estimation of multinomial logit models with unobserved heterogeneity using maximum simulated likelihood* (No. 573). DIW Discussion Papers.
- Hadfield, J. D. (2010). MCMC methods for multi-response generalized linear mixed models: the MCMCglmm R package. *Journal of Statistical Software*, 33(2), 1-22.
- Hadfield, J. (2016). MCMCglmm course notes. <http://cran.us.r-project.org/web/packages/MCMCglmm/vignettes/CourseNotes.pdf>
- Hartzel, J., Agresti, A., & Caffo, B. (2001). Multinomial logit random effects models. *Statistical Modelling*, 1(2), 81-102.

- Haynes, M., Western, M., Yu, L., & Spellak, M. (2008, August). Analysing nominal data from a panel survey: Employment transitions of Australian women. In *103rd Annual Meeting of the American Sociological Association* (pp. 1-4).
- McCulloch, N., & Baulch, B. (1999). *Distinguishing the chronically from the transitorily poor: Evidence from rural Pakistan*. Institute of Development Studies at the University of Sussex.
- McCulloch, C. E., & Searle, S. R. (2001). *Generalized, Linear, and Mixed Models*, Wiley Series in Probability and Statistics.
- Rabe-Hesketh, S., Skrondal, A., & Pickles, A. (2002). Reliable estimation of generalized linear mixed models using adaptive quadrature. *The Stata Journal*, 2(1), 1-21.
- Rabe-Hesketh, S., Skrondal, A., & Pickles, A. (2004). Generalized multilevel structural equation modeling. *Psychometrika*, 69(2), 167-190.
- Rabe-Hesketh, S., Skrondal, A., & Pickles, A. (2004). GLLAMM manual. University of California, Berkeley U.C. Berkeley Division of Biostatistics Working Paper Series No: 160.
- Revelt, D., & Train, K. (1998). Mixed logit with repeated choices: households' choices of appliance efficiency level. *Review of economics and statistics*, 80(4), 647-657.
- Rossi, P. E., & Allenby, G. M. (1993). A Bayesian approach to estimating household parameters. *Journal of Marketing Research*, 171-182.
- Sarma, S., & Simpson, W. (2007). A panel multinomial logit analysis of elderly living arrangements: Evidence from aging in Manitoba longitudinal data, *Canada. Social science & medicine*, 65(12), 2539-2552.
- Train, K. (2003). *Discrete Choice Methods with Simulation*, Cambridge University Press.
- Train, K. (1997), Halton sequences for mixed logit. Department of Economics, UCB.
- TÜİK, Gelir ve Yaşam Koşulları Araştırması 2009-2012 panel mikro veri seti.
- Ziegler, A. (2002). Simulated classical tests in the multiperiod multinomial probit model. ZEW Discussion Papers, No. 02-38, Zentrum für Europäische Wirtschaftsforschung (ZEW), Mannheim.

BÖLÜM 14

PANEL SIRALI MODELLER VE STATA UYGULAMALARI

Canan GÜNEŞ*

Bazı durumlarda kategorik değişkenler kendi doğasından gelen sıralı bir yapıya sahiptir. 5. Bölümde bahsedildiği üzere bağımlı değişkenin sıralı ölçeğe sahip olduğu durumlarda sıralı modellere başvurulmaktadır. Panel veriler ile çalışıldığında ise sıralı modeller yerini panel sıralı modellere bırakmaktadır.

Sosyal ve davranış bilimlerinde sıralı ölçekli bağımlı değişkenler oldukça yaygın olarak kullanılmaktadır. Sıralı bağımlı değişkenlere bireylerin kendi sağlık durumlarını kötü, orta, iyi ve mükemmel olarak değerlendirdikleri çalışmalar örnek olarak verilebilir. Ankete katılanların belirli bir duruma karşı tutumlarını likert ölçeği ile ölçen çalışmalarla da sıklıkla karşılaşmaktadır. Gelir düzeyinin belirli aralıklarla tanımlandığı değişkenler de sıralı ölçeğe örnektir. Bu bölümde panel sıralı modellerden, panel sıralı probit modeli ve panel sıralı logit modeli tanıtılacaktır.

14.1. Panel Sıralı Modeller

Ekonometrik çalışmalarda yer alan bireyler, hanhalkları, firmalar, ülkeler, bölgeler ve şehirler gibi birimler arasında davranışsal farklılıklar vardır. Birimler arasında davranışsal farklılıkların olması, birimlerin heterojen olduğunu gösterir. Panel veriler ile çalışmak gözlenemeyen birime özgü heterojenliğin, diğer adıyla gözlenemeyen birim etkinin hesaba katılmasını sağlar. Panel veri analizinde sabit ve rassal etkiler arasındaki seçim, gözlenemeyen birim etki ve açıklayıcı değişkenler arasındaki ilişkinin varlığına göre belirlenir. Birim etkinin açıklayıcı değişkenler ile ilişkisiz olması durumunda rassal etkiler tercih edilirken; birim etki ile açıklayıcı değişkenlerin korelasyonlu olması durumunda ise sabit etkiler kullanılır.

Kesikli tercih modelleri, sınırlı bağımlı değişkenli modeller ve süre modelleri gibi parametrelerde doğrusal olmayan modellere genel olarak doğrusal olmayan modeller de denir. Doğrusal olmayan panel veri modellerinde rassal etkiler tahmincisi, sabit etkilere göre daha hassas tahminler verir. Bunun nedenlerinden biri, birime özgü özellikleri yansıtan zaman değişmezi değişkenleri-

* Çanakkale Onsekiz Mart Üniversitesi, Biga İktisadi ve İdari Bilimler Fakültesi, Ekonometri Bölümü.
E-mail: canangunes@comu.edu.tr.

nin¹ etkisini tahmin edebilmesidir. Sabit etkiler modelinde, zaman değişmezi değişkenleri değişken dönüşümü sonucunda modelden düşer ve tahmin edilemez. Sosyal bilimlerde yapılan çalışmalarda cinsiyet, medeni durum gibi zaman değişmezi değişkenlerinin modele dâhil edilememesi önemli bir sorundur.

Rassal etkiler tahmincisinin daha hassas olmasının bir diğer nedeni ise tahminlerde kullanılan bilgiler ile ilişkilidir. Sabit etkiler tahmincisi, açıklayıcı değişkenlerin bağımlı değişken üzerindeki etkilerini tahmin ederken her bir birey için yalnızca değişkenlerin zaman içindeki değişimlerinden gelen bilgileri kullanır. Fakat farklı açıklayıcı değişken değerlerine karşılık, farklı bireyler arasında bağımlı değişkenin nasıl değiştiğine ilişkin bilgiden faydalanmaz. Buna karşılık, rassal etkiler tahmincisi her iki bilgi kaynağını da kullanır (Hill ve diğerleri, 2011: 557).

Son neden ise panel verinin zaman boyutu ile ilişkilidir. Doğrusal olmayan panel veri modellerinde sabit etkiler varsayımı ile zaman boyutu T sonsuza giderken en çok olabilirlik yöntemi tutarlı tahminler verir. Ancak panel verilerde genellikle zaman boyutu küçüktür. Zaman boyutu küçük olduğunda birim etki hakkında bilgi elde etmek için sınırlı sayıda gözlem vardır. Bu durumda rastlantısal parametre problemi (Neyman ve Scott (1948), ve Lancaster (2000)) ortaya çıkar. Rastlantısal parametre problemine göre sabit etkiler varsayımı altında model tahmin edildiğinde T sonlu olduğu sürece gözlenemeyen birim etki tahminleri sapmalıdır.

Panel sıralı modellerde sabit etkiler sıralı logit modelini tahmin etmek için farklı yöntemler önerilmiştir. Ancak literatürde bu tahmin edicilerden hangisinin tercih edileceği ya da tahmin edicilerin tutarlılıkları konusunda henüz bir fikir birliği yoktur. Bu nedenle rassal etkiler sıralı modeller, sabit etkilerden daha çok kullanılır. Bu bilgiler ışığında panel sıralı modeller sadece rassal etkiler için açıklanacaktır. Ancak rassal etkiler sıralı modellerde birim etki ile açıklayıcı değişkenler arasında korelasyona izin verilmediği unutulmamalıdır. Birim etki ile açıklayıcı değişkenler arasında korelasyon olması durumunda Mundlak (1978) düzeltmesi² kullanılarak rassal etkiler varsayımı altında model tahmin edilebilir ve kısmen de olsa bu sorun aşılabılır (Longhi ve Nandi, 2015: 209).

Genel olarak bir panel sıralı model aşağıdaki şekilde gösterilebilir:

$$y_{it} = \beta'x_{it} + \alpha_i + u_{it} \quad i = 1, 2, \dots, N \quad t = 1, 2, \dots, T \quad (14.1)$$

Burada y_{it} , 1'den J 'ye kadar değerler alan sıralı ölçekli bağımlı değişkendir. x_{it} açıklayıcı değişkenler; β ise eğim parametreleri vektörüdür. α_i , birimlere özgü heterojenliği temsil eden gözlenemeyen birim etkidir. α_i ; sıfır ortalamalı, σ_α^2 sabit varyanslı ve u_{it} 'den (tüm t 'ler için) bağım-

¹ Birimlere özgü özellikleri yansıtan değişkenlere zaman değişmezi değişkeni denir. Zaman değişmezi değişkenleri, birimlere göre değişir ama zamana göre sabit kalır. Örneğin; cinsiyet, ırk gibi. Panel veri analizinde kullanılan açıklayıcı değişken türleri için bkz. Tatoğlu (2018: 39).

² Rassal etkiler sıralı modelin Mundlak düzeltmesinde, gözlenemeyen birim etki ve açıklayıcı değişkenler arasındaki olası korelasyonu kontrol etmek için modele açıklayıcı değişkenlerin zamana göre birim ortalamaları dahil edilir. Daha fazla bilgi için; bkz. Mundlak (1978), Greene ve Hensher (2010: 281), Greene (2012: 836-837).

sızdır. Birim etkinin rassal olduğu (birim etki ile açıklayıcı değişkenler arasında korelasyon olmadığı) varsayılır. u_{it} ise hata terimidir. Hata terimi u_{it} sıfır ortalamalı ve sabit varyanslıdır.

Panel sıralı modelleri açıklayabilmek için gizli değişken yaklaşımı kullanılabilir. Gizli değişken doğrudan gözlenemez. Ancak gözlenebilen değişkenlerle oluşturulan istatistiksel model ile ilişkilendirilerek gizli değişken çıkarsanabilir. Gizli değişken yaklaşımında, sıralı ölçekli bağımlı değişken y_{it} , gizli sürekli değişken y_{it}^* olarak aşağıdaki gibi gösterilir.

$$y_{it}^* = \beta' x_{it} + \alpha_i + u_{it} \quad (14.2)$$

J adet sıralı kategoriden oluşmak üzere bağımlı değişken y_{it} ile gizli değişken y_{it}^* arasındaki ilişki aşağıdaki gibidir (Rabe-Hesketh ve Skrondal, 2012: 577).

$$\begin{aligned} y_{it} &= 1, \tau_0 < y_{it}^* \leq \tau_1 \text{ için} \\ y_{it} &= 2, \tau_1 < y_{it}^* \leq \tau_2 \text{ için} \\ y_{it} &= 3, \tau_2 < y_{it}^* \leq \tau_3 \text{ için} \\ &\vdots \\ y_{it} &= J, \tau_{j-1} < y_{it}^* \leq \tau_j \text{ için} \end{aligned} \quad (14.3)$$

τ 'lar kesme noktaları ya da eşik parametreleri olarak adlandırılır. Eşik parametreleri model ile birlikte tahmin edilir ve kategorilerin sınırlarını gösterir. Eşik parametreleri $-\infty$ ve $+\infty$ arasında değerler alırlar. Burada $\tau_0 = -\infty$ ve $\tau_j = \infty$ 'dur. Sıralı bağımlı değişken J adet kategoriden oluşmak üzere, $J-1$ adet eşik değeri $(\tau_1, \tau_2, \dots, \tau_{j-1})$ tahmin edilir.

Rassal etkiler sıralı modellerde hata terimi varyansı 1'e eşit ve standart normal dağılım gösteriyorsa panel sıralı probit model söz konusudur. Hata terimi varyansı $\pi^2/3$ 'e eşit ve lojistik dağılım gösteriyorsa panel sıralı logit model geçerlidir. Koşulsuz olasılığın elde edilişi daha kolay ve anlaşılır olduğu için öncelikle rassal etkiler sıralı probit model açıklanmıştır. Daha sonra rassal etkiler sıralı logit modele yer verilmiştir.

14.2. Rassal Etkiler Sıralı Probit Model

Rassal etkiler sıralı probit modeli, en çok olabilirlik yöntemi kullanılarak tahmin edilebilir. y_{it} 'nin j 'nin üstünde bir kategoriye düşme birikimli olasılığı;

$$Pr(y_{it} > j | \tau, x_{it}, \alpha_i) = \Phi(\beta' x_{it} + \alpha_i - \tau_j) \quad (14.4)$$

ya da y_{it} 'nin j ya da altında bir kategoriye düşme birikimli olasılığı;

$$Pr(y_{it} \leq j | \tau, x_{it}, \alpha_i) = \Phi(\tau_j - \beta' x_{it} - \alpha_i) \quad (14.5)$$

şeklinde hesaplanır. $\alpha_i \sim IIDN(0, \sigma_\alpha^2)$ olmak üzere τ_j eşik değerler kümesini temsil etmektedir. Sıralı ölçekli bağımlı değişken J kategoriden oluşmak üzere; $\tau_1, \tau_2, \dots, \tau_{j-1}$ eşik değerlerdir. $\Phi(\cdot)$, standart normal birikimli dağılım fonksiyonudur.

Açıklayıcı değişken y_{it} için j 'inci kategorinin gözlemlenme olasılığı aşağıdaki gibi elde edilir.

$$\begin{aligned} Pr(y_{it} = j | \tau, x_{it}, \alpha_i) &= Pr(\tau_{j-1} < \beta'x_{it} + \alpha_i + u_{it} \leq \tau_j) \\ &= Pr(\tau_{j-1} - \beta'x_{it} - \alpha_i < u_{it} \leq \tau_j - \beta'x_{it} - \alpha_i) \\ &= \Phi(\tau_j - \beta'x_{it} - \alpha_i) - \Phi(\tau_{j-1} - \beta'x_{it} - \alpha_i) \end{aligned} \quad (14.6)$$

Burada β vektörü sabit terim içermez, çünkü sabit terimin etkisi eşik parametreleri ile tahmin edilir. Rassal etkiler sıralı probit modelde hata terimi u_{it} , sıfır ortalama ve birim varyans ile standart normal dağılır. Ayrıca u_{it} , gözlenemeyen birim etki α_i 'den bağımsızdır.

Birim etkinin bir değeri (i. birim) için koşullu olasılık aşağıda verilmiştir.

$$Pr(y_i | \alpha_i) = \prod_{t=1}^T \sum_{j=1}^J I(y_{it} = j) [\Phi(\tau_j - \beta'x_{it} - \alpha_i) - \Phi(\tau_{j-1} - \beta'x_{it} - \alpha_i)] \quad (14.7)$$

Birim etkinin normal dağılım gösterdiği varsayımı ile koşulsuz olasılık ise

$$\begin{aligned} P(y_i) &= \int_{-\infty}^{\infty} \prod_{t=1}^T \sum_{j=1}^J I(y_{it} = j) [\Phi(\tau_j - \beta'x_{it} - \alpha) - \Phi(\tau_{j-1} - \beta'x_{it} - \alpha)] \\ &\quad \times \frac{1}{\sqrt{2\pi}\sigma_\alpha} e^{-0.5\left(\frac{\alpha}{\sigma_\alpha}\right)^2} d\alpha \end{aligned} \quad (14.8)$$

şeklinde. Değişken değişimi kullanılarak,

$$v = \frac{\alpha}{\sqrt{2}\sigma_\alpha} \Rightarrow dv = \frac{d\alpha}{\sqrt{2}\sigma_\alpha} \quad (14.9)$$

elde edilir ve yerine konulduğunda koşulsuz olasılığa ulaşılır (Croissant ve Millo, 2019: 217-218).

$$\begin{aligned} P(y_i) &= \frac{1}{\sqrt{\pi}\sigma_\alpha} \int_{-\infty}^{\infty} \prod_{t=1}^T \sum_{j=1}^J I(y_{it} = j) \\ &\quad \times [\Phi(\tau_j - \beta'x_{it} - \sqrt{2}\sigma_\alpha v) - \Phi(\tau_{j-1} - \beta'x_{it} - \sqrt{2}\sigma_\alpha v)] \times e^{-v^2} dv \end{aligned} \quad (14.10)$$

Denklem 14.10'da yer alan integral için kapalı bir form yoktur. Ancak Butler ve Moffitt (1982) tarafından geliştirilen Gauss-Hermite kareleme yöntemi (Gauss-Hermite quadrature) kullanılarak etkin bir şekilde tahmin edilebilir. Bu yöntemde v 'nin farklı değerleri (v_r ile gösterilsin) için Denklem 14.10'da yer alan fonksiyon değerlendirilir ve bu değerlerin doğrusal bileşeni hesaplanır.

$$P(y_i) \approx \frac{1}{\sqrt{\pi}\sigma_\alpha} \sum_{r=1}^R \prod_{t=1}^T \sum_{j=1}^J I(y_{it} = j) \times [\Phi(\tau_j - \beta'x_{it} - \sqrt{2}\sigma_\alpha v_r) - \Phi(\alpha_{j-1} - \beta'x_{it} - \sqrt{2}\sigma_\alpha v_r)] \times e^{-v_r^2} dv \quad (14.11)$$

Gauss-Hermite kareleme yönteminde yakınsamanın doğruluğu seçilen kareleme noktasının sayısına bağlıdır. Kareleme noktası arttıkça yakınsama iyileşir. Fakat kareleme noktası arttıkça hesaplama süresi uzayacaktır, bu da araştırmacıya zaman maliyeti yaratır. Modeli tahmin edebilmek için Gauss-Hermite kareleme yöntemi yerine en çok benzetimli olabilirlik (maximum simulated likelihood) yaklaşımı³ da kullanılabilir. Söz konusu hesaplama yöntemlerinden hiçbirinin belli bir avantajı yoktur. Ancak kareleme boyutu arttıkça en çok benzetimli olabilirlik yaklaşımı daha hızlı yakınsar (Greene ve Hensher, 2010: 277).

14.2.1. Rassal Etkiler Sıralı Probit Model Stata Uygulaması

Standard&Poor's (S&P) kredi derecelendirme kuruluşu tarafından ülkelere verilen kredi notları üzerinde etkili olan faktörleri incelemek amacıyla 58 ülkenin 2000-2014 dönemi verileri toplanmıştır. Veriler için kitabın web sayfasında **kredinotu.dta** veri setine bakabilirsiniz.

Tablo 14.1. kredinotu.dta Veri Seti

describe

```
obs:      870
vars:      9
size:     26,970
```

variable name	storage type	display format	value label	variable label
ulkeler	long	%16.0g	ulke	ülke isimleri
yil	int	%8.0g		yıl, 2000-2014
sp	str4	%9s		s&p
spnot	byte	%8.0g		s&p numaralandırılmış kredi notları
kn	float	%9.0g		Ülke kredi notları; 1:çok düşük, 2:düşük, 3:yüksek, 4:çok yüksek
advanced	byte	%8.0g		Gelişmiş ülke kuklası; 1: gelişmiş, 0:diğer
lac	float	%9.0g		1: Latin Amerika ve Karayipler ülkeleri, 0:diğer
inf	float	%8.0g		Enflasyon, ortalama tüketici fiyatları (%)
unemploy	float	%8.0g		işsizlik oranı (%)
lngdppercapita	float	%9.0g		Kişi başına düşen GSMH'nın logaritması

Sorted by: ulkeler yil

³ Ayrıntılı bilgi için bknz. Greene ve Hensher (2010: 277).

S&P tarafından ülkelere verilen uzun vadeli kredi notlarına göre en yüksek not AAA iken en düşük not D'dir. Seçilen ülkeler ve dönem için veri setine bakıldığında en yüksek notun AAA ve en düşük notun da CC olduğu görülür. Bu hâliyle toplam 19 kategori vardır. Basitleştirmek amacıyla kredi notları 4 kategoride toplanmıştır. Kategoriler belirlenirken öncelikle S&P tarafından verilen kredi notlarının anlamları ve daha sonra frekansları dikkate alınmıştır⁴. CC'den B(-)'ye kadar olan notlar "1", B'den B(+)’ya kadar olan notlar "2", BB(-)'den BBB(+)'ya kadar olan notlar "3" ve A(-)'den AAA'ya kadar notlar da "4" olarak belirlenmiştir. Bu durumda bir ülkenin alabileceği kredi notları "1=çok düşük, 2=düşük, 3=yüksek, 4=çok yüksek" şeklinde sıralanmıştır. Bu hâliyle ülkelerin kredi notunu temsil eden "kn" değişkeninin ikiden fazla kategoriden oluşan sıralı ölçekli bir nitel değişken olduğu açıktır.

Modelde kullanılan açıklayıcı değişkenler; ülke gelişmiş ise "1", değilse "0" değerini alan kukla değişken (**advanced**), ülke Latin Amerika ve Karayipler ülkesi ise "1", değilse "0" değerini alan kukla değişken (**lac**), ortalama tüketici fiyatları ile enflasyon (%) (**inf**), işsizlik oranı (%) (**unemp**) ve ABD Doları cinsinden kişi başına düşen Gayri Safi Milli Hasıla (GSMH)'nın logaritması (**lngdppercapita**)'dır⁵. Veri setinde yer alan değişkenlere ilişkin tanımlayıcı istatistikler Tablo 14.2'de verilmiştir.

Tablo 14.2. Tanımlayıcı İstatistikler

sum ulkeler yil kn advanced lac inf unemploy lngdppercapita

Variable	Obs	Mean	Std. Dev.	Min	Max
-----+-----					
ulkeler	870	29.5	16.7503	1	58
yil	870	2007	4.322979	2000	2014
kn	870	3.067816	.9083419	1	4
advanced	870	.4885057	.5001554	0	1
lac	870	.137931	.3450259	0	1
-----+-----					
inf	870	4.300993	5.642625	-3.716	57.311
unemploy	870	7.737964	4.599089	.49	27.47
lngdpperca~a	870	9.505752	1.151553	6.13768	11.70233

Veri seti 58 ülkeye ait 2000-2014 dönemini kapsayan toplam 870 gözlemden oluşmaktadır. Dengeli panel söz konusudur. 58 ülkenin %48.85'i gelişmiş ülke, %13.79'u ise Latin Amerika ve Karayipler ülkesidir. Model tahminine geçmeden önce panel veri ile çalışıldığı için Stata'ya birim (ulkeler) ve zaman (yil) boyutunu temsil eden değişkenlerin tanımlanması gerekir.

⁴ Benzer bir çalışma için bkz. Akay ve Oskonbaeva (2018).

⁵ "kn" Trading Economies internet sitesinden; "lac", "advanced" ve "inf" IMF'nin; "unemploy" ve "lngdppercapita" ise Dünya Bankası'nın veri tabanından derlenmiştir.

Tablo 14.3. Panel Veri Setinin Birim ve Zaman Boyutunun Tanıtılması

```

xtset ulkeler yil
      panel variable:  ulkeler (strongly balanced)
      time variable:   yil, 2000 to 2014
      delta:           1 unit

```

Stata'da sabit etkiler ile panel sıralı modelleri tahmin etmek için hazır bir komut yoktur. Fakat Stata 13 ya da daha yeni bir sürümünü kullanarak rassal etkiler ile panel sıralı modeller kolaylıkla tahmin edilebilir. Rassal etkiler sıralı probit modelini tahmin etmek için menü çubuğundan aşağıdaki adımlar izlenir, açılan pencerede bağımlı ve bağımsız değişkenler tanıtıldıktan sonra model tahmin edilir.

Statistics > Longitudinal/panel data > Ordinal outcomes > Probit regression (RE)

Rassal etkiler sıralı probit modeli, menü kullanmadan, **xtoprobit** komutu kullanılarak da tahmin edilebilir. Model tahmini Tablo 14.4'te verilmiştir.

```

xtoprobit depvar [indepvars] [if] [in] [weight] [, options]

```

Tablo 14.4. Rassal Etkiler Sıralı Probit Model

```

xtoprobit kn advanced lac inf unemploy lngdppercapita, nolog

```

```

Random-effects ordered probit regression      Number of obs      =      870
Group variable: ulkeler                     Number of groups    =      58

Random effects u_i ~ Gaussian                Obs per group: min =      15
                                              avg =      15.0
                                              max =      15

Integration method: mvaghermite             Integration points =      12

Wald chi2(5) =      174.60
Log likelihood = -369.88229                   Prob > chi2         =      0.0000

```

	kn	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
advanced		1.724863	.4100431	4.21	0.000	.9211935	2.528533
lac		-2.885241	1.055458	-2.73	0.006	-4.953902	-.8165807
inf		-.0510309	.0213138	-2.39	0.017	-.0928051	-.0092567
unemploy		-.3134879	.0339347	-9.24	0.000	-.3799987	-.2469771
lngdppercapita		1.154284	.1562993	7.39	0.000	.8479431	1.460625
/cut1		2.623814	1.585331	1.66	0.098	-.4833782	5.731005
/cut2		4.90169	1.594289	3.07	0.002	1.77694	8.026439
/cut3		10.30989	1.622985	6.35	0.000	7.128899	13.49088
/sigma2_u		6.235684	1.810882			3.52931	11.01738

```

LR test vs. oprobit regression:  chibar2(01) = 458.77 Prob>=chibar2 = 0.0000

```

Tablo 14.4'te rassal etkiler sıralı probit model tahmini yer almaktadır. Model tahmini 12 kareleme noktası ile ortalama ve varyans uyarlamalı (adaptive) Gauss-Hermite kareleme yöntemi (mvaghermite) kullanılarak elde edilmiştir. Ayrıca `intmethod(ghermite)` seçeneği kullanılarak uyarlamasız (nonadaptive) Gauss-Hermite (ghermite) kareleme yöntemi ile de model tahmini yapılabilir. Kareleme yönteminin tahmin edilen modele uygunluğu önemlidir. Tablo 14.4'teki modelde uyarlamasız Gauss-Hermite kareleme yöntemi kullanılmak istendiğinde maksimizasyon süreci çözüme yakınsayamadığı için model tahmin edilememektedir. Ayrıca `intpoints()` seçeneği ile kareleme noktası sayısı değiştirilebilir. Bu seçenek belirtilmezse varsayılan 12 kareleme noktası ile model tahmin edilir. Kareleme noktası arttırıldıkça daha doğru tahmin elde edilir. Ancak kareleme noktasının arttırılması, modelin tahmin edilme süresinin uzamasına neden olacaktır.

Modelin bütüncül anlamlılığı Tablo 14.4'teki model çıktısında yer alan Wald testi ile araştırılmaktadır. 5 (modelde yer alan açıklayıcı değişken sayısı) serbestlik dereceli χ^2 dağılımı gösteren Wald test istatistiği 174.60 olarak hesaplanmıştır. Test sonucuna göre modelin bütüncül olarak anlamsız olduğunu söyleyen H_0 hipotezi reddedilir. Bu durumda modelin bütüncül olarak anlamlı olduğu söylenebilir.

Eğim katsayılarının her biri tekil olarak %5 anlamlılık düzeyinde istatistiksel olarak anlamlıdır. Açıklayıcı değişken katsayı tahminleri standart sıralı probit modelde olduğu gibi yorumlanabilir. Ancak bu yorumlar yeterince açıklayıcı olmadığı için katsayıların büyüklüklerini değil, işaretlerini yorumlamak tercih edilir. Diğer değişkenlerin etkisi sabit tutulduğunda; gelişmiş ülke olmak, gelişmiş olmayan ülkelere göre daha yüksek sıra değerli kredi notu kategorisinde bulunma olasılığını arttırırken, Latin Amerika ve Karayipler ülkesi olmak diğer ülkelere göre söz konusu olasılığı azaltır. Diğer değişkenlerin etkisi sabit tutulduğunda; kişi başına düşen GSMH'deki artış ülkelerin daha yüksek sıra değerli kredi notu kategorisinde bulunma olasılığını arttırmaktadır. Enflasyon ve işsizlik oranlarındaki artış ise daha düşük kredi notu kategorisinde bulunma olasılığını arttırır. Model çıktısında eşik değerleri (cut1, cut2, cut3) tahminleri de yer almaktadır. Bağımlı değişken 4 (J) kategoriden oluştuğu için 3 (J-1) adet eşik değeri tahmin edilmiştir. Eşik değerleri %10 anlamlılık düzeyinde istatistiksel olarak anlamlıdır. Eşik değerlerinin istatistiksel olarak anlamlı olması kredi notu kategorilerinin birbirinden farklı olduğunu gösterir. Tablo 14.4'te birim etkinin varyansı σ_α^2 (sigma2_u) tahmini de bulunmaktadır. Birim etkinin varyansı 1.8109 standart hata ile 6.2357 olarak tahmin edilmiştir.

Tablo 14.4'ün alt kısmında Olabilirlik Oran (Likelihood Ratio-LR) testi yer almaktadır. Bu test ile standart sıralı probit modeli, rassal etkiler sıralı probit modeline karşı test edilmektedir. 1 serbestlik dereceli χ^2 dağılımı gösteren LR test istatistiği 458.77 olarak elde edilmiştir ve test sonucuna göre H_0 hipotezi reddedilmektedir. Bu durumda birimler (ülkeler) arasında değişkenlik olduğu için standart sıralı probit model yerine, rassal etkiler sıralı probit model tercih edilmelidir.

Rassal etkiler sıralı probit modeli ile çalışırken katsayı büyüklükleri doğrudan yorumlanamadığı için marjinal etkiler kullanılabilir. Marjinal etkileri hesaplamak için `margins` komutu kullanılır. Tablo 14.5'te her bir kredi notu düzeyi için hesaplanmış açıklayıcı değişkenlerin marjinal etkileri

yer almaktadır. Komut dizisinde **outcome(#1)** ile $y_{it} = 1$ için, **outcome(#2)** ile $y_{it} = 2$ için, **outcome(#3)** ile $y_{it} = 3$ için ve **outcome(#4)** ile $y_{it} = 4$ için marjinal etkiler hesaplanmıştır.

Tablo 14.5. Rassal Etkiler Sıralı Probit Modeli için Marjinal Etkiler

```
margins, dydx(*) predict(pu0 outcome(#1))
```

Average marginal effects	Number of obs	=	870
Model VCE : OIM			

```
Expression      : Predicted mean (1.kn), assuming u_i=0, predict(pu0 outcome(#1))
dy/dx w.r.t.   : advanced lac inf unemploy lngdppercapita
```

	Delta-method					
	dy/dx	Std. Err.	z	P> z	[95% Conf. Interval]	
advanced	-.048416	.0197299	-2.45	0.014	-.0870858	-.0097461
lac	.0809871	.051501	1.57	0.116	-.019953	.1819273
inf	.0014324	.0007836	1.83	0.068	-.0001034	.0029682
unemploy	.0087994	.0029108	3.02	0.003	.0030944	.0145044
lngdppercapita	-.0324001	.011659	-2.78	0.005	-.0552513	-.0095489

```
. margins, dydx(*) predict(pu0 outcome(#2))
```

Average marginal effects	Number of obs	=	870
Model VCE : OIM			

```
Expression : Predicted mean (2.kn), assuming u_i=0, predict(pu0 outcome(#2))
dy/dx w.r.t. : advanced lac inf unemploy lngdppercapita
```

	Delta-method					
	dy/dx	Std. Err.	z	P> z	[95% Conf. Interval]	
advanced	-.0735479	.0315056	-2.33	0.020	-.1352977	-.011798
lac	.1230261	.0114697	10.73	0.000	.1005459	.1455064
inf	.0021759	.0010761	2.02	0.043	.0000668	.0042851
unemploy	.0133671	.0044904	2.98	0.003	.0045661	.0221681
lngdppercapita	-.0492185	.0188656	-2.61	0.009	-.0861943	-.0122426

```
margins, dydx(*) predict(pu0 outcome(#3))
```

Average marginal effects	Number of obs	=	870
Model VCE : OIM			

```
Expression      : Predicted mean (3.kn), assuming u_i=0, predict(pu0 outcome(#3))
dy/dx w.r.t.    : advanced lac inf unemploy lngdppercapita
```

	Delta-method					
	dy/dx	Std. Err.	z	P> z	[95% Conf. Interval]	
advanced	-.0446188	.0217022	-2.06	0.040	-.0871544	-.0020832
lac	.0746355	.0596164	1.25	0.211	-.0422105	.1914814
inf	.0013201	.0010426	1.27	0.205	-.0007234	.0033635
unemploy	.0081093	.0050163	1.62	0.106	-.0017224	.017941
lngdppercapita	-.029859	.0166098	-1.80	0.072	-.0624136	.0026955

margins, dydx(*) predict(pu0 outcome(#4))

Average marginal effects

Number of obs = 870

Model VCE : OIM

Expression : Predicted mean (4.kn), assuming $u_i=0$, predict(pu0 outcome(#4))

dy/dx w.r.t. : advanced lac inf unemploy lngdppercapita

	Delta-method					
	dy/dx	Std. Err.	z	P> z	[95% Conf. Interval]	
advanced	.1665826	.0345207	4.83	0.000	.0989232	.2342419
lac	-.2786487	.1035577	-2.69	0.007	-.4816181	-.0756793
inf	-.0049284	.0022589	-2.18	0.029	-.0093558	-.000501
unemploy	-.0302758	.0053859	-5.62	0.000	-.0408321	-.0197196
lngdppercapita	.1114776	.0206986	5.39	0.000	.0709092	.152046

Diğer açıklayıcı değişkenler sabit tutulduğunda; gelişmiş ülkelerin, gelişmiş ülke grubunda yer almayan ülkelere göre çok düşük ($y_{it} = 1$) kredi notu alma olasılıkları 0.0484 kadar, düşük ($y_{it} = 2$) kredi notu alma olasılıkları 0.0735 kadar ve yüksek ($y_{it} = 3$) kredi notu alma olasılıkları ise 0.0446 kadar (ya da %4.46) daha azdır. Buna karşın gelişmiş ülkelerin, gelişmiş ülke grubunda yer almayan ülkelere göre çok yüksek ($y_{it} = 4$) kredi notu alma olasılıkları 0.1666 kadar daha fazladır. Diğer değişkenler sabit tutulduğunda; kişi başına düşen GSMH'deki %1'lik artış; çok düşük kredi notu alma olasılığını 0.0324 kadar, düşük kredi notu alma olasılığını 0.0492 kadar, yüksek kredi notu alma olasılığını da 0.0299 kadar azaltmakta ve çok yüksek kredi notu alma olasılığını ise 0.1115 kadar arttırmaktadır. İstatistiksel olarak anlamlı olan diğer değişkenlerin marjinal etkileri de benzer şekilde yorumlanmaktadır.

Rassal etkiler sıralı probit modeli Stata'da tahmin edilirken eğim katsayılarına ilişkin standart hataların türü **vce(vcetype)** ile belirlenebilir. Tab14.4'te standart hata türü ile ilgili bir değişiklik yapılmadığı için gözlemlenen bilgi matrisine (observed information matrix) dayanan standart hatalar ile model tahmini verilmiştir. **vce(robust)** ile bazı spesifikasyon (tanımlama) hatalarına karşı dirençli olan Huber White Sandwich tahmincisine dayanan ve grup içi korelasyona karşı dirençli standart hatalar, **vce(cluster clustvar)** ile grup içi korelasyona izin veren ve gözlemlerin bağımsız olması gerektiği varsayımını gevşeten standart hatalar, **vce(bootstrap)** ile paramet-

rik olmayan bootstrap yöntemine dayanan standart hatalar ve `vce(jackknife)` ile jackknife yöntemine dayanan standart hatalar ile model tahminleri yapılabilir. Örnek olarak dirençli standart hatalar ile rassal etkiler sıralı probit model tahmini Tablo 14.6'da verilmiştir.

Tablo 14.6. Dirençli Standart Hatalar ile Rassal Etkiler Sıralı Probit Model

`xtoprobit kn advanced lac inf unemploy lngdppercapita, vce(robust) nolog`

Random-effects ordered probit regression	Number of obs	=	870
Group variable: <i>ulkeler</i>	Number of groups	=	58
Random effects $u_i \sim \text{Gaussian}$	Obs per group: min	=	15
	avg	=	15.0
	max	=	15
Integration method: <i>mvaghermite</i>	Integration points	=	12
	Wald $\chi^2(5)$	=	70.25
Log pseudolikelihood = -369.88229	Prob > χ^2	=	0.0000

(Std. Err. adjusted for 58 clusters in *ulkeler*)

	kn	Coef.	Robust Std. Err.	z	P> z	[95% Conf. Interval]	
advanced		1.724863	.544034	3.17	0.002	.6585763	2.79115
lac		-2.885241	1.112184	-2.59	0.009	-5.065082	-.7054001
inf		-.0510309	.0310149	-1.65	0.100	-.1118189	.0097571
unemploy		-.3134879	.0580173	-5.40	0.000	-.4271997	-.199776
lngdppercapita		1.154284	.3535902	3.26	0.001	.4612601	1.847308
/cut1		2.623814	3.442681	0.76	0.446	-4.123717	9.371344
/cut2		4.90169	3.442446	1.42	0.154	-1.84538	11.64876
/cut3		10.30989	3.877886	2.66	0.008	2.709375	17.91041
/sigma2_u		6.235684	2.363494			2.966576	13.10728

Tablo 14.6'da elde edilen dirençli standart hatalı tahminler ile Tablo 14.4'te verilen gözlemlenen bilgi matrisine dayalı standart hatalı tahminler karşılaştırıldığında eğitim katsayılarının aynı olduğu görülmektedir. Bu iki model arasında standart hatalar ve dolayısıyla anlamlılıklar farklıdır. Tablo 14.4'te eğitim katsayıları %5 anlamlılık düzeyinde istatistiksel olarak anlamlı iken, Tablo 14.6'daki katsayılar %10 anlamlılık düzeyinde anlamlıdır. Eşik değerlerine bakıldığında Tablo 14.6'daki dirençli standart hatalar ile tahmin edilen *cut1* ve *cut2*'nin istatistiksel olarak anlamsızlaştığı görülmektedir. Eşik değerlerinin istatistiksel olarak anlamsız olması dört kredi notu kategorisinin birbirinden farklı olmadığını düşündürmektedir.

14.3. Rassal Etkiler Sıralı Logit Model

Panel sıralı modellerde rassal etkiler varsayımı ile başvurulabilecek bir diğer model rassal etkiler sıralı logit modeldir. Rassal etkiler sıralı logit modele en çok olabilirlik uygulanırsa; y_{it} 'nin j 'nin üstünde bir kategoriye düşme birikimli olasılığı;

$$Pr(y_{it} > j | \tau, x_{it}, \alpha_i) = \Lambda(\beta'x_{it} + \alpha_i - \tau_j) \quad (14.12)$$

şeklinde hesaplanır. Rassal etkiler sıralı probit modelde olduğu gibi gözlenemeyen birim etki $\alpha_i \sim IIDN(0, \sigma_\alpha^2)$ 'dir ve τ_j eşik değerler kümesini temsil etmektedir. $\Lambda(\cdot)$, lojistik birikimli dağılım fonksiyonudur.

Açıklayıcı değişken y_{it} için j 'inci kategorinin gözlemlenme olasılığı aşağıdaki gibidir (Akay ve Van, 2018: 231).

$$\begin{aligned} Pr(y_{it} = j | \tau, x_{it}, \alpha_i) &= Pr(\tau_{j-1} < \beta'x_{it} + \alpha_i + u_{it} \leq \tau_j) \\ &= Pr(\tau_{j-1} - \beta'x_{it} - \alpha_i < u_{it} \leq \tau_j - \beta'x_{it} - \alpha_i) \\ &= \Lambda(\tau_j - \beta'x_{it} - \alpha_i) - \Lambda(\tau_{j-1} - \beta'x_{it} - \alpha_i) \\ &= \frac{1}{1 + \exp(-\tau_j + \beta'x_{it} + \alpha_i)} - \frac{1}{1 + \exp(-\tau_{j-1} + \beta'x_{it} + \alpha_i)} \end{aligned} \quad (14.13)$$

Burada $\tau_0 = -\infty$ ve $\tau_j = \infty$ 'dur. Rassal etkiler sıralı probit modelde olduğu gibi rassal etkiler sıralı logit modelde de sabit terim yoktur. Eşik parametreleri, sabit terimin yerini alır. Rassal etkiler sıralı logit modelde hata terimi u_{it} , sıfır ortalama ve $\pi^2/3$ varyans ile lojistik dağılır. Hata terimi u_{it} , gözlenemeyen birim etki α_i 'den bağımsızdır.

Rassal etkiler sıralı logit modeli en çok olabilirlik yöntemi ile tahmin edebilmek için, panel sıralı probit modelde olduğu gibi, koşulsuz olasılık ile olabilirlik fonksiyonu elde edilir. Daha sonra Gauss-Hermite kareleme yöntemi ya da en çok benzetimli olabilirlik yaklaşımı ile integral hesaplanarak parametre tahminleri elde edilebilir. Sıralı probit modeldeki $\Phi(\cdot)$, standart normal birikimli dağılım fonksiyonu yerine sıralı logit modelde $\Lambda(\cdot)$, lojistik birikimli dağılım fonksiyonu kullanıldığı unutulmamalıdır.

14.3.1. Rassal Etkiler Sıralı Logit Model Stata Uygulaması

Rassal etkiler sıralı logit modeli tahmin etmek için menü çubuğundan aşağıdaki adımlar izlenir, açılan pencerede bağımlı ve bağımsız değişkenler tanıttıldıktan sonra model tahmin edilir.

Statistics > Longitudinal/panel data > Ordinal outcomes > Logit regression (RE)

Rassal etkiler sıralı logit modeli, **xtologit** komutu kullanılarak da tahmin edilebilir. **kredino-tu.dta** veri seti ile modelin tahmini Tablo 14.7'de verilmiştir.

```
xtologit depvar [indepvars] [if] [in] [weight] [, options]
```

Tablo 14.7. Rassal Etkiler Sıralı Logit Model

xtologit kn advanced lac inf unemploy lngdppercapita, nolog

Random-effects ordered logistic regression	Number of obs	=	870
Group variable: <i>ulkeler</i>	Number of groups	=	58
Random effects $u_i \sim \text{Gaussian}$	Obs per group: min	=	15
	avg	=	15.0
	max	=	15
Integration method: <i>mvaghermite</i>	Integration points	=	12
Log likelihood = -367.095	Wald $\chi^2(5)$	=	153.53
	Prob > χ^2	=	0.0000

kn	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
advanced	3.488785	.8324144	4.19	0.000	1.857283	5.120288
lac	-5.400571	1.984809	-2.72	0.007	-9.290725	-1.510416
inf	-.0910412	.0399296	-2.28	0.023	-.1693018	-.0127806
unemploy	-.5544546	.0640539	-8.66	0.000	-.679998	-.4289112
lngdppercapita	2.204798	.2948999	7.48	0.000	1.626804	2.782791
/cut1	5.609985	2.919326	1.92	0.055	-.1117896	11.33176
/cut2	9.673581	2.952586	3.28	0.001	3.886619	15.46054
/cut3	20.33282	3.091006	6.58	0.000	14.27456	26.39108
/sigma2_u	21.87165	6.518073			12.19582	39.22402
LR test vs. ologit regression: $\text{chibar2}(01) = 451.04$ Prob>=chibar2 = 0.0000						

Tablo 14.7'de rassal etkiler sıralı logit modelinin tahmini yer almaktadır. Model tahmini 12 kareleme noktası ile ortalama ve varyans uyarlamalı (adaptive) Gauss-Hermite kareleme yöntemi (*mvaghermite*) kullanılarak elde edilmiştir. Rassal etkiler sıralı probit modelde olduğu gibi **intmethod(ghermite)** seçeneği kullanılarak uyarlamasız (nonadaptive) Gauss-Hermite (*ghermite*) kareleme yöntemi ile model tahmini yapılabilir. **intpoints()** seçeneği ile kareleme noktası sayısı araştırmacı tarafından değiştirilebilir. Rassal etkiler sıralı probit modelde olduğu gibi **vce(vcetype)** ile eğim katsayılarına ilişkin standart hataların türü belirlenebilir. Standart hata ile ilgili seçenekler rassal etkiler sıralı probit model ile aynıdır.

Wald testi ile modelin bütüncül anlamlılığı araştırılmaktadır. 5 (modelde yer alan açıklayıcı değişken sayısı) serbestlik dereceli χ^2 dağılımı gösteren Wald test istatistiği 153.53 olarak hesaplanmıştır. Test sonucuna göre modelin bütüncül olarak anlamsız olduğunu söyleyen temel hipotez reddedilir ve modelin bütüncül olarak anlamlı olduğuna karar verilir.

Eğim katsayılarının her biri tekil olarak %5 anlamlılık düzeyinde istatistiksel olarak anlamlıdır. Açıklayıcı değişken katsayı tahminleri standart sıralı logit modelde olduğu gibi yorumlanabilir.


```
margins, dydx(*) predict(pu0 outcome(#2))
```

Average marginal effects	Number of obs	=	870
Model VCE : OIM			

```
Expression      : Predicted mean (2.kn), assuming u_i=0, predict(pu0 outcome(#2))
dy/dx w.r.t.    : advanced lac inf unemploy lngdppercapita
```

	Delta-method					
	dy/dx	Std. Err.	z	P> z	[95% Conf. Interval]	
advanced	-.0785564	.0355989	-2.21	0.027	-.148329	-.0087838
lac	.1216037	.0116085	10.48	0.000	.0988516	.1443559
inf	.00205	.0010737	1.91	0.056	-.0000545	.0041544
unemploy	.0124846	.0045439	2.75	0.006	.0035786	.0213905
lngdppercapita	-.0496451	.0203501	-2.44	0.015	-.0895305	-.0097596

```
margins, dydx(*) predict(pu0 outcome(#3))
```

Average marginal effects	Number of obs	=	870
Model VCE : OIM			

```
Expression : Predicted mean (3.kn), assuming u_i=0, predict(pu0 outcome(#3))
dy/dx w.r.t. : advanced lac inf unemploy lngdppercapita
```

	Delta-method					
	dy/dx	Std. Err.	z	P> z	[95% Conf. Interval]	
advanced	-.0433257	.0265282	-1.63	0.102	-.09532	.0086686
lac	.0670674	.0601159	1.12	0.265	-.0507576	.1848923
inf	.0011306	.0010138	1.12	0.265	-.0008564	.0031176
unemploy	.0068855	.0049827	1.38	0.167	-.0028804	.0166515
lngdppercapita	-.0273804	.0183746	-1.49	0.136	-.063394	.0086331

```
margins, dydx(*) predict(pu0 outcome(#4))
```

Average marginal effects	Number of obs	=	870
Model VCE : OIM			

```
Expression      : Predicted mean (4.kn), assuming u_i=0, predict(pu0 outcome(#4))
dy/dx w.r.t.   : advanced lac inf unemploy lngdppercapita
```

	Delta-method					
	dy/dx	Std. Err.	z	P> z	[95% Conf. Interval]	
advanced	.1745473	.0392785	4.44	0.000	.0975628	.2515318

lac		-.2701958	.1001571	-2.70	0.007	-.4665001	-.0738914
inf		-.0045549	.0021976	-2.07	0.038	-.0088622	-.0002476
unemploy		-.0277399	.0054091	-5.13	0.000	-.0383415	-.0171383
lngdppercapita		.1103081	.0222027	4.97	0.000	.0667916	.1538247

Diğer açıklayıcı değişkenlerin etkisi sabit tutulduğunda; enflasyon oranındaki %1'lik artış ülkelerin çok düşük ($y_{it} = 1$) kredi notu alma olasılıklarını 0.0014 kadar (ya da %0.14) ve düşük ($y_{it} = 2$) kredi notu alma olasılıklarını 0.0021 kadar arttırırken, çok yüksek ($y_{it} = 4$) kredi notu alma olasılıklarını 0.0046 kadar azaltır. İşsizlik oranı incelendiğinde; işsizlik oranındaki %1'lik artışın çok düşük kredi notu alma olasılığını 0.0084 kadar, düşük kredi notu alma olasılığını 0.0125 kadar arttırdığı ve çok yüksek kredi notu alma olasılığını 0.0277 kadar azalttığı görülmektedir. Yüksek ($y_{it} = 3$) kredi notu alma durumu için değişkenlerin marjinal etkileri %10 anlamlılık düzeyinde istatistiksel olarak anlamsızdır, bu nedenle yorumlanmamıştır. Diğer değişkenlerin istatistiksel olarak anlamlı olan marjinal etkileri de benzer şekilde yorumlanmaktadır.

Tablo 14.9. Rassal Etkiler Sıralı Logit Modeli için Bahis Oranları

xtologit kn advanced lac inf unemploy lngdppercapita, or nolog

Random-effects ordered logistic regression	Number of obs	=	870
Group variable: ulkeler	Number of groups	=	58
Random effects u_i ~ Gaussian	Obs per group: min	=	15
	avg	=	15.0
	max	=	15
Integration method: mvaghermite	Integration points	=	12
Log likelihood = -367.095	Wald chi2(5)	=	153.53
	Prob > chi2	=	0.0000

kn	Odds Ratio	Std. Err.	z	P> z	[95% Conf. Interval]
advanced	32.74615	27.25837	4.19	0.000	6.406309 167.3835
lac	.004514	.0089594	-2.72	0.007	.0000923 .220818
inf	.9129801	.036455	-2.28	0.023	.844254 .9873007
unemploy	.5743854	.0367917	-8.66	0.000	.506618 .6512178
lngdppercapita	9.068416	2.674275	7.48	0.000	5.087591 16.16407
/cut1	5.609985	2.919326	1.92	0.055	-.1117896 11.33176
/cut2	9.673581	2.952586	3.28	0.001	3.886619 15.46054
/cut3	20.33282	3.091006	6.58	0.000	14.27456 26.39108
/sigma2_u	21.87165	6.518073			12.19582 39.22402

LR test vs. ologit regression: chibar2(01) = 451.04 Prob>=chibar2 = 0.0000

Uygulamada logit modelleri ile bahis oranlarını (odds ratio⁶) hesaplamak ve bu bahis oranları üzerinden katsayı yorumlarını yapmak genellikle tercih edilmektedir. Stata'da **logit** komutu ile kullanılan **or** seçeneği ile bahis oranları elde edilebilir. Ayrıca bahis oranları, logit modeli ile tahmin edilen β eğitim katsayılarının üsteli alınarak (e^β) da hesaplanır. Tablo 14.9'da rassal etkiler sıralı logit modeli için bahis oranları, standart hataları, z değerleri ve anlamlılıkları yer almaktadır. Diğer değişkenlerin etkisi sabit tutulduğunda; Latin Amerika ve Karayipler ülkelerinin diğer ülkelere göre daha düşük kredi notu kategorisinden daha yüksek kredi notu kategorisi lehine olan bahis oranı yaklaşık %99.5 ($[1-0.004514]*100$) kadar azdır. Enflasyon oranındaki %1'lik artış daha düşük kredi notu kategorisinden daha yüksek kredi notu kategorisi lehine olan bahis oranını yaklaşık %9 ($[1-0.9129801]*100$) azaltır (Gujarati, 2015: 276).

Bu bölümde kullanılan Stata komutları aşağıda toplu olarak verilmiştir.

describe

sum *ulkeler yil kn advanced lac inf unemploy lngdppercapita*

xtset *ulkeler yil*

xtprobit *kn advanced lac inf unemploy lngdppercapita, nolog*

margins, dydx(*) predict(pu0 outcome(#1))

margins, dydx(*) predict(pu0 outcome(#2))

margins, dydx(*) predict(pu0 outcome(#3))

margins, dydx(*) predict(pu0 outcome(#4))

xtprobit *kn advanced lac inf unemploy lngdppercapita, vce(robust) nolog*

xtlogit *kn advanced lac inf unemploy lngdppercapita, nolog*

margins, dydx(*) predict(pu0 outcome(#1))

margins, dydx(*) predict(pu0 outcome(#2))

margins, dydx(*) predict(pu0 outcome(#3))

margins, dydx(*) predict(pu0 outcome(#4))

xtlogit *kn advanced lac inf unemploy lngdppercapita, or nolog*

14.4. Kaynaklar

- Akay, E. Ç. ve Oskonbaeva, Z. (2018). Panel Sıralı Modelinin Kredi Derecelendirilmesine Uygulanması. *International Conference on Eurasian Economies*. 430-437.
- Akay, E. Ç. ve Van, M. H. (2018). Panel Sıralı Nitel Tercih Modelleri. *Uygulamalı Panel Veri Ekonometrisi*. (ss. 227-242). Editör Selahattin Güriş. İstanbul: Der Yayınları.
- Butler, J. S. ve Moffitt, R. (1982). A Computationally Efficient Quadrature Procedure for One-Factor Multinomial Probit Model. *Econometrica*. 50(3): 761-764.
- Croissant, Y. ve Millo, G. (2019). *Panel Data Econometrics with R*. UK: John Wiley & Sons Ltd.

⁶ Farklı kaynaklarda fark oranı, olasılık oranı gibi isimler de kullanılmaktadır.

- Greene, W. H. ve Hensher, D. A. (2010). *Modeling Ordered Choices: A Primer*. Cambridge: Cambridge University Press.
- Greene, W. H. (2012). *Econometric Analysis, Seventh Edition*. England: Pearson Education Limited.
- Gujarati, D. (2015). *Örneklerle Ekonometri*. Çev. Nasip Bolatoğlu. Ankara: BB101 yayınları.
- Hill, R. C., Griffiths, W. E. ve Lim, G. C. (2011). *Principles of Econometrics, Fourth Edition*. United States of America: John Wiley & Sons.
- Lancaster, T. (2000). The Incidental Parameter Problem Since 1948. *Journal of Econometrics*. 95: 391-413.
- Longhi, S. ve Nandi, A. (2015). *A Practical Guide to Using Panel Data*. Sage.
- Mundlak, Y. (1978). On the Pooling of Time Series and Cross Section Data. *Econometrica*. 46(1): 69-85.
- Neyman, J. ve Scott, E. L. (1948). Consistent Estimates Based on Partially Consistent Observations. *Econometrica*. 16(1): 1-32.
- Rabe-Hesketh, S. ve Skrondal, A. (2012). *Multilevel and Logitudinal Modeling Using Stata Volume II: Categorical Responses, Counts and Survival, Third Edition*. Texas: Stata Press.
- Tatoğlu, F. Y. (2018). *Panel Veri Ekonometrisi: Stata Uygulamalı, 4. Baskı*. İstanbul: Beta.

BÖLÜM 15

PANEL SAYMALI MODELLER VE STATA UYGULAMALARI

Egemen İpek *

Saymalı veriler genellikle belirli bir zaman aralığında meydana gelen olay sayıları olarak tanımlanır. Bu nedenle saymalı veriler negatif olmayan tamsayı değerlerinden oluşur (teknik olarak bir üst sınırı olmasına da gerek yoktur). Saymalı veri özelliğindeki bağımlı değişkenin sıralı veya sürekli değişken olarak değerlendirilmesi durumunda klasik regresyon modelleri sapmalı tahminler üretir. Bu durumun temel nedeni saymalı veri özelliğindeki bağımlı değişkenin normal dağılım göstermemesidir. Özellikle örnekleme sıfırlı gözlemlerin çok olması veya örneklemin küçük değerler üzerinde yoğunlaşması durumunda bağımlı değişkenin normal dağılım göstermesi nedeniyle tahminlemede saymalı modellerinin tercih edilmesi gerekmektedir. Verilerin sahip olduğu bu yapı nedeniyle olasılık dağılımı dikkate alınmalıdır. Bu olasılık dağılımlarına bağlı iki temel model vardır. Bunlardan birincisi Panel Poisson model, ikincisi Panel Negatif Binom modelidir. Basit saymalı modellerinden farklı olarak panel saymalı modellerde gözlemlerin birimler ve zamanlar arası bağımsızlık varsayımını sağlaması gerekli değildir. Ayrıca panel saymalı modelleri birimler arası heterojenliğe izin vermektedir. Pozitif koşullu ortalamalı panel saymalı modellerinde birim etkiler modele toplamsal yerine çarpımsal olarak dâhil edilir.

Mikro panel veriler genellikle zaman boyutu kısa gözlem boyutu uzun yapıdadır. Birim sayısı $n \rightarrow \infty$ iken, zaman boyutu T sınırlıdır. Verinin birimler arası bağımsız olduğu varsayılırken birimlerin zamanla bağlantılı olduğu kabul edilir. Hausman, Hall ve Griliches (1984) tarafından yapılan ABD'deki firmaların patent başvuru sayıları ile Ar-Ge harcamaları arasında ilişkiyi analiz eden çalışma panel saymalı modelinin kullanıldığı ilk çalışma olarak literatürde sıklıkla yer almaktadır.

Bu bölümde Stata uygulamaları Hausman, Hall ve Griliches'in (1984) çalışmasına benzer şekilde Dünya Bankası veri sisteminden çeşitli ülkeler için 2000-2017 dönemine ait verilerle patent başvuru sayısı ile Ar-Ge harcamalarının GSMH içindeki payı, bir yıl içinde yayınlanan akademik makale sayısı, büyüme oranı ve bir yıl içinde kurulan yeni start-up firma sayısı arasındaki ilişki için yapılacaktır.

* Doç.Dr.Tarsus Üniversitesi, Uygulamalı Bilimler Fakültesi, Finans & Bankacılık Bölümü.
E-mail: egemenipek@tarsus.edu.tr

Veri her bir ülkenin farklı zaman dilimlerindeki gözlemlerini içermesinden dolayı panel veri özelliğindedir. Patent sayısı ile söz konusu değişkenler arasındaki ilişki belirlenirken hâliyle ülkelerin sahip olduğu gelir ve teknoloji gibi farklılıkların patent başvuru sayısında etkili olabileceği gibi ülke bazında gözlemlenemeyen etkilerin de patent sayısı üzerinde etkisinin olması muhtemeldir.

Panel saymalı modellerde, belirli zaman dönemi (bir yıl) için verinin birimler (ülkeler) arası bağımsız olduğu varsayılırken, belirli birimin (ülkenin) ise zamanla ilişkili olmasına izin verilir. Standart panel veri analizlerinde olduğu gibi, her bir birim (ülke) için ayrı kukla değişkenin tanımlandığı sabit etkiler, birim spesifik terimin belirlenmiş bir dağılımdan elde edildiği rassal etkiler (tesadüfî etkiler) ile birim ve zaman etkilerinin sabit varsayıldığı örneklem ortalaması (population-averaged) veya havuzlanmış (pooled) etkiler olmak üzere üç temel model vardır.

Zaman periyodu bütün birimler için aynı ise dengeli panel aksi durumda dengesiz panel olarak adlandırılır. Dengesiz panele neden olan gözlemler rassal olarak eksik ve aşırı derecede dengesiz değil ise panelin dengeli veya dengesiz olması Stata tahmin algoritmasında bir soruna neden olmaz. Çalışmada kullanılan veri seti dengeli panel özelliğinde olup bütün birimler için 2000-2017 yılları arasındaki gözlemler yer almaktadır.

15.1. Panel Poisson Modelleri

Panel saymalı modellerinde en çok kullanılan model poisson modelidir. Temel doğrusal panel model denklemi aşağıda gösterilmektedir.

$$y_{it} = \alpha_{it} + x'_{it}\beta_{it} + u_{it}, \quad i = 1, \dots, N, \quad t = 1, \dots, T \quad (15.1.a)$$

$$y_{it} \sim \text{Poisson}(\mu_{it}) \quad (15.1.b)$$

Burada y_{it} saymalı veri özelliğinde negatif olmayan tam sayı değerlerine sahip bağımlı değişkeni, α_{it} sabit terimi, x_{it} $k \times 1$ bağımsız değişken vektörünü, u_{it} hata terimini göstermektedir. Alt indislerden i birimi (birey, firma ülke gibi) temsil ederken, t zamanı temsil etmektedir.

Denklem (15.1.a)'da genel bir model tanımlandığı için bazı kısıtlamaların yapılması gerekmektedir. Yapılacak olan kısıtlamalar ve varsayımlara bağlı olarak model tahminleri yapılabilmektedir.

15.1.1. Örneklem Ortalaması Panel Poisson Modeli

Birim özelinde etkilerin incelenmesine geçmeden önce örneklem ortalaması panel poisson modelinde tahminleme i . birim için $y_{i1}, \dots, y_{iT}$ 'nin ortak dağılımından ziyade y_{it} 'nin birim marjinal dağılımı üzerinden yapılır. Örneklem ortalaması panel poisson modelinde birim etkilerin ortalamasına eşit olduğu varsayımında temel doğrusal panel model Denklemi (15.1) aşağıdaki gibi olur.

$$y_{it} = \alpha + x'_{it}\beta + u_{it}, \quad i = 1, \dots, N, \quad t = 1, \dots, T \quad (15.2.a)$$

$$y_{it} \sim \text{Poisson}(\mu_{it}) \quad (15.2.b)$$

(15.1.b)'de yer alan koşul için aşağıdaki varsayımın sağlanması durumunda (koşullu ortalamanın doğru belirlenmesi durumunda) tutarlı tahmin ediciler elde edilir.

$$\mu_{it} = \alpha \lambda_{it} = \alpha \exp(x'_{it}\beta), \quad (\alpha = E_{\alpha_i}[\alpha_i]) \quad (15.3)$$

Veri X_{it} durumunda Y_{it} 'nin koşullu varyansı $Var(Y_{it}|X_{it}) = v(\mu_{it})\phi$. Burada v varyans fonksiyonu, ϕ tahmin edilmesi gereken parametredir. v ve ϕ çıktı değişkeninin dağılımına bağlıdır. Örneklem ortalaması panel poisson modeli tahminleri, genelleştirilmiş tahmin denklemleri (GEE) üzerinden yapılırken modelin poisson dağılımına ve log bağlantı fonksiyonuna sahip olduğu varsayımı yapılır (StataCorp, 2019).

y_i için varyans-kovaryans matrisinin korelasyon yapısına bağlı olarak gösterimi:

$$V_i = \phi A_i^{1/2} R_i(\alpha) A_i^{1/2} \quad (15.4)$$

$A_i = \text{Diag}\{v(\mu_{i1}), \dots, v(\mu_{it})\}$ ve $R_i(\alpha)$ grup içi korelasyon yapısını temsil etmektedir. α değerinin tahmin edilmesi için iteratif algoritma uygulanarak Pearson artıkları (e_{it}) cari β değeri yardımıyla aşağıdaki gibi hesaplanır:

$$e_{it} = (y_{it} - \mu_{it}) / \sqrt{v(\mu_{it})} \quad (15.5)$$

ϕ parametresi ise aşağıdaki gibi tahmin edilir:

$$\hat{\phi} = \frac{1}{N-p} \sum_{i=1}^K \sum_{t=1}^T e_{it}^2 \quad (15.6)$$

Burada $N = \sum_{i=1}^K n_t$ toplam gözlem sayısı ve p bağımlı değişken boyutunu temsil etmektedir.

Örneklem ortalaması panel poisson modeli, GEE yönteminin özel bir durumu olduğu için β parametreleri aşağıdaki denklem yardımıyla elde edilir:

$$U(\beta) = \sum_{i=1}^K D_i' V_i^{-1} (Y_i - \mu_i) = 0, \quad (D_i = \partial \mu_i / \partial \beta') \quad (15.7)$$

15.1.1.1. Örneklem Ortalaması Panel Poisson Model Stata Uygulaması

Panel poisson tahminleri için Stata komutu olarak **xtpoisson** kullanılmaktadır. Komutun temel yapısı aşağıda belirtilmektedir.

xtpoisson depvar [indepvars] [if] [in] [weight] , pa [PA_options]

Komut yapısı hakkında detaylı bilgi için **help xtpoisson** komutu yardımıyla açılan bilgilendirme ekranından detaylara ulaşabilirsiniz. Örneklem ortalamalı tahminlerin üretilebilmesi için Stata komutunda “,” den sonra opsiyon için “**pa**” (population average) yazılması gerekmektedir. Aksi takdirde **xtpoisson** komutu default olarak rassal etkiler tahminleri üretir.

Örneklem ortalaması panel poisson tahminleri daha genel kullanıma sahip GEE komutu olan **xtgee** yardımıyla da elde edilebilmektedir. Söz konusu komutun örneklem ortalaması panel poisson için genel yapısı aşağıdaki gibidir:

```
xtgee depvar [indepvars] [in] [weight] [, family(poisson) link(log)
corr(exchangeable) options]
```

Dünya Bankası veri sisteminden temin edilen 2000-2017 yılları arasında çeşitli ülkeler için patent başvuru sayısı ile Ar-Ge harcamalarının GSMH içindeki payı arasındaki ilişkiyi ilk olarak örneklem ortalaması panel poisson modeli ile tahmin edelim. Uygulama için patent.dta[†] veri setinden yararlanılacaktır. Öncelikle panel veri seti ile çalıştığımız için veri setinin panel şeklinde tanıtılmış olması gereklidir. Bunun için **xtsetup** komutu ile birim (id) ve zamanın (time) tanıtılması gerekir. Çalışmada kullanılacak değişkenlere dair bilgiler Tablo 15.1a ve Tablo 15.b'de yer almaktadır. Tablo 15.1. b'de çalışmada kullanılan veri setinin dengeli panel olduğu görülmektedir.

Tablo 15.1a. Patent.dta Veri Seti İçeriği

```
use "DOSYAKONUMU\patent.dta"
describe
```

variable name	storage type	display format	value label	variable label
ülkeler	str20	%20s		ülke isimleri
time	int	%8.0g		zaman
patent	long	%8.0g		patent başvuru sayısı
arge	float	%8.0g		arge/gsmh
eu	float	%9.0g		avrupa kuklası
id	long	%20.0g		ülke kodları
makale	float	%8.0g		bilimsel çalışma sayısı
startup	byte	%8.0g		kurulan startup firma sayısı
gdp	float	%8.0g		büyüme oranı

Tablo 15.1b. Patent.dta Veri Setinin Zamana Bağlı Dağılımı

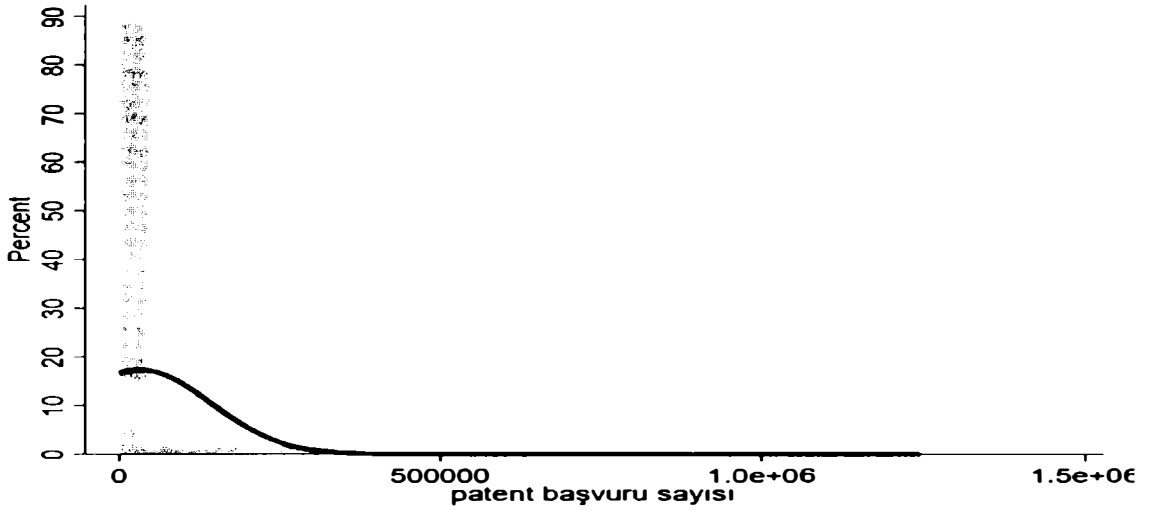
```
xtdescribe
    id: 1, 2, ..., 38                n =          38
    time: 2000, 2001, ..., 2017      T =          18
    Delta(time) = 1 unit
    Span(time) = 18 periods
    (id*time uniquely identifies each observation)
Distribution of T_i:  min      5%      25%      50%      75%      95%      max
                    18       18       18       18       18       18       18
    Freq.  Percent  Cum. | Pattern
-----+-----
    38    100.00  100.00 | 111111111111111111
-----+-----
    38    100.00      | XXXXXXXXXXXXXXXXXXXX
```

[†]patent.dta veri setine Uygulamalarla Mikro Ekonometri kitabının linkinden ulaşabilirsiniz.

Poisson modellerinin kullanılabilmesi için ilk koşul bağımlı değişkenin saymalı yapıda olmasıdır. Saymalı veriler belirli bir zaman aralığında meydana gelen olay sayıları olarak tanımlanır ve negatif olmayan tamsayı değerlerinden oluşur. Çalışmada kullanılan bağımlı değişken olan patent sayısı bir yıl içinde ülkelerin toplam başvuru yaptıkları patent sayısını temsil etmektedir. Bu nedenle bağımlı değişken negatif olmayan tamsayılardan oluşmaktadır. İkinci temel koşul ise bağımlı değişkenin sağa çarpık yapıda olmasıdır. Bunun için bağımlı değişkenin değerlerinin dağılım grafiğine bakılabilir. Gözlemlerin yoğunluk, frekans veya yüzdesel dağılımlarına göre grafikler oluşturulabilir. Grafik Şekil 15.1’de yüzdesel dağılıma ait grafik yer almakta olup ve normal dağılım mavi çizgi ile gösterilmektedir. Grafik 15.1’den gözlemlerin sağa çarpık dağılıma sahip olduğu görülmektedir. Ayrıca gözlemlerin yüzdesel dağılımını detaylı olarak görmek için `tabulate patent` komutu kullanılabilir.

Şekil 15.1. Toplam Patent Başvuru Dağılımı

histogram patent,normal percent



Üçüncü koşul olarak ise bağımlı değişkene ait koşullu ortalama ile koşullu varyansın eşit olmasıdır. Saymalı verilerde koşullu varyans ve koşullu ortalama arasındaki varsayım yayılım parametresi “ ϕ ” üzerinden tanımlanır:

$$\text{var}(Y|x) = \phi E(Y|x) \quad (15.8)$$

Eğer $\phi > 1$ ise aşırı yayılım, $\phi = 1$ ise eş yayılım ve son olarak $\phi < 1$ ise eksik yayılım durumu söz konusudur. Çalışmada kullanılan bağımlı ve bağımsız değişkenlere ait detaylı özet istatistikler aşağıdaki gibi elde edilir.

Tablo 15.2.a. Detaylı Özet İstatistikler (Patent)

summarize patent, detail

patent başvuru sayısı

Percentiles		Smallest		
1%	25	13		
5%	85	18		
10%	142	18	Obs	684
25%	444	19	Sum of Wgt.	684
50%	1315		Mean	32225.31
		Largest	Std. Dev.	110355.6
75%	4336	801135		
90%	56769	968252	Variance	1.22e+10
95%	241347	1204981	Skewness	6.344857
99%	415829	1245709	Kurtosis	55.89786

Poisson modellerinde eş yayılım varsayımının sağlanması beklenmektedir. Ancak örneğimizde de olduğu gibi ampirik olarak veri setlerinin sıklıkla aşırı yayılım sergiledikleri görülür (Tablo 15.17’de aşırı yayılım parametresinin nasıl elde edildiği anlatılmaktadır). Bu varsayımın ihlal edilmesi durumunda uygulanacak alternatif yöntemler daha sonra açıklanacak olup örneklem ortalaması panel poisson tahminlerine ait uygulama detaylarıyla aşağıda açıklanmıştır.

Tablo 15.2.b. Özet İstatistikler (Tüm Değişkenler)

summarize patent arge makale startup gdp europe

Variable	Obs	Mean	Std. Dev.	Min	Max
patent	684	32225.31	110355.6	13	1245709
arge	684	.1490047	.1012137	.012999	.455324
makale	684	14247.43	22643.64	20.7	108474
startup	680	7.348529	3.411062	2	20
gdp	684	.0316152	.0383981	-.1481	.3447
europe	684	.4210526	.4940893	0	1

Tablo 15.2.b’de modelde kullanılan bağımsız değişkenlere ait bilgilere yer verilmektedir. Ar-Ge harcamalarının GSMH içindeki payı ortalama olarak %15 iken en yüksek pay %45 ile Kore Cumhuriyetine aittir. Kukla değişkeni europe ikili değer almakta ve örneklem %42’si Avrupa Birliğine üye ülkelerden oluşmaktadır.

İlk olarak ülkelerin patent sayısı ile Ar-Ge harcamalarının GSMH içindeki payı, yayınlanan akademik makale sayısı, kurulan yeni start-up işletme sayısı arasındaki ilişkinin örneklem ortalaması panel poisson tahmini aşağıdaki gibi elde edilir.

Tablo 15.3. Örneklem Ortalaması Panel Poisson Tahmin Sonuçları

```
quietly xtset id time
```

```
xtpoisson patent arge makale startup gdp, pa nolog
```

Kullanılan opsiyonlar:

pa : örneklem ortalaması tahminler

nolog : iterasyonları göstermemesi için

```
. xtpoisson patent arge makale startup gdp ,pa nolog
```

```
GEE population-averaged model      Number of obs      =      680
Group variable:                     id      Number of groups   =      38
Link:                               log      Obs per group:
Family:                             Poisson      min =      15
Correlation:                        exchangeable      avg =      17.9
                                          max =      18
                                          Wald chi2(4)      =      6821519
Scale parameter:                    1      Prob > chi2        =      0.0000
```

	patent	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
arge		7.476831	.000291	2569.33	0.000	7.471128 7.482535
makale		-.0000116	3.43e-08	-338.62	0.000	-.0000117 -.0000115
startup		-.0020082	.0000445	-45.10	0.000	-.0020954 -.0019209
gdp		1.077045	.0031514	341.77	0.000	1.070868 1.083222
_cons		9.345597	.001187	7873.22	0.000	9.343271 9.347924

Tablo 15.3'te yer alan örneklem ortalaması panel poisson tahmin sonuçlarına göre Ar-Ge harcamalarının ve GSMH'nin patent başvuruları üzerinde pozitif ve istatistiksel bir etkiye sahip olduğu, makale sayısı ve start-up firma sayısının ise negatif, anlamlı bir etkiye sahip olduğu görülmektedir. Saymalı model tahminleri neticesinde elde edilen katsayılar Logit veya Probit modellerinde olduğu gibi doğrudan yorumlanmaya uygun değildir. Söz konusu modellerde kullanılan fark oranına (odds ratio) benzer şekilde *incidence-rate ratio* (IRR)'nin hesaplanması gerekir. Bunun için tahmin edilen katsayıların exponansiyeli ($e^{\hat{\beta}}$) manuel olarak hesaplanabileceği gibi `irr` opsiyonu kullanılarak da hesaplanabilir. Tahminleme yapıldıktan sonra `xtpoisson irr` komutu kullanılarak aynı sonuçların hesaplanması sağlanır. IRR katsayısı x_i 'deki bir birimlik değişimin değeridir. Bir diğer ifade ile açıklayıcı değişkendeki beklenen olaydaki bir birimlik artışın beklenen olay sayısına oranıdır. Ancak dikkat edilirse Tablo 15.3'te yer alan temel modeldeki katsayılara ait standart hatalar ile IRR tahminlerine ait standart hatalar farklıdır. Bunun nedeni IRR'ye ait standart hataların modele ait varyans-kovaryans matrisi üzerinden değil de *delta* yöntemi kullanılarak hesaplanmasından kaynaklanmaktadır (Hilbe, 2014). IRR standart hatalar herhangi bir değişkene ait IRR katsayısı ile Tablo 15.3'te yer alan temel modeldeki aynı değişkene ait standart hatanın çarpımı yoluyla elde edilir. Örnek olarak Tablo 15.4'teki Ar-Ge değişkeni için $2.112101 \times 0.000291 = 0.0006146$ 'dır.

Tablo 15.4. Örneklem Ortalaması Panel Poisson IRR Tahmin Sonuçları

```
quietly xtpoisson patent arge makale startup gdp, pa
xtpoisson, irr
```

veya

```
xtpoisson patent arge makale startup gdp ,pa nolog irr
```

Kullanılan opsiyonlar:

```
pa      : örneklem ortalaması tahminler
nolog  : iterasyonları göstermemesi için
irr     : incidence-rate ratio hesaplanması için
```

```
xtpoisson patent arge makale startup gdp ,pa nolog irr
```

```
GEE population-averaged model      Number of obs      =      680
Group variable:                     id      Number of groups   =      38
Link:                               log      Obs per group:
Family:                             Poisson      min =      15
Correlation:                        exchangeable      avg =      17.9
                                          max =      18
                                          Wald chi2(4)      =      6821519
Scale parameter:                    1      Prob > chi2        =      0.0000
```

patent	IRR	Std. Err.	z	P> z	[95% Conf. Interval]
arge	1766.634	5.140966	2569.33	0.000	1756.587 1776.739
makale	.9999884	3.43e-08	-338.62	0.000	.9999883 .9999885
startup	.9979938	.0000444	-45.10	0.000	.9979068 .9980809
gdp	2.935991	.0092524	341.77	0.000	2.917912 2.954181
_cons	11448.31	13.58926	7873.22	0.000	11421.7 11474.97

Note: _cons estimates baseline incidence rate (conditional on zero random effects).

Tablo 15.4'te yer alan panel poisson örneklem ortalamasına dair IRR tahmin sonuçlarına bakıldığında Ar-Ge harcamalarının GSMH oranındaki bir birimlik artışın patent sayıları üzerinde istatistiksel olarak anlamlı ve %1765'lik ($1766-1=1765$) çok ciddi bir artışa neden olduğu görülmektedir. Ülkelerin büyüme oranlarındaki bir birimlik artışın ise patent başvuru sayısında istatistiksel olarak anlamlı ve %193'lük ($2.93-1=1.93$) bir artışa neden olduğu söylenebilir. Makale sayısındaki bir birimlik artışın patent sayıları üzerinde istatistiksel olarak anlamlı ve %1'lik ($0.99-1=-0.01$) azalışa neden olduğu görülmektedir.

Modele k kategorili bir açıklayıcı değişken (kukla) eklenmesi durumunda tahmin edilen katsayının yorumlanmasında genel kural $k-1$ kategori için tahminlerin yapılp 1 kategorinin temel kategori sayılmasıdır. Tahmin edilen katsayılar bu temel kategoriye göre yorumlanır. Örnek olarak Avrupa Birliği'nde bulunan ülkeler için kukla değişken oluşturulsun (Ülke Avrupa Birliğine üye ise 1 diğer durumlarda 0 değerini almaktadır) ve model tahmini yeniden yapılsın.

Tablo 15.5. Örneklem Ortalaması Panel Poisson IRR Tahmin Sonuçları: Kukla Değişkenli

xtpoisson patent arge makale startup gdp eu, pa irr nolog

Kullanılan opsiyonlar:

pa : örneklem ortalaması tahminler
nolog : iterasyonları göstermemesi için
irr : incidence-rate ratio hesaplanması için

GEE population-averaged model		Number of obs	=	680
Group variable:	id	Number of groups	=	38
Link:	log	Obs per group:		
Family:	Poisson	min	=	15
Correlation:	exchangeable	avg	=	17.9
		max	=	18
		Wald chi2(5)	=	8716079
Scale parameter:	1	Prob > chi2	=	0.0000

patent	IRR	Std. Err.	z	P> z	[95% Conf. Interval]
arge	887.268	2.276416	2645.78	0.000	882.8175 891.7409
makale	.9999892	3.71e-08	-292.02	0.000	.9999891 .9999893
startup	.9913057	.0000383	-226.00	0.000	.9912307 .9913808
gdp	9.934944	.0388821	586.68	0.000	9.859028 10.01144
europa	.0420715	.0001269	-1050.26	0.000	.0418235 .042321
_cons	23690.95	27.55436	8660.53	0.000	23637.01 23745.02

Note: _cons estimates baseline incidence rate (conditional on zero random effects).

Tablo 15.5'te yer alan kukla değişkenli modelde temel kategori Avrupa Birliği üyesi olmayanlardır. Elde edilen katsayı 0.05 olup, diğer değişkenler sabitken Avrupa Birliği üyesi ülkelerin Avrupa Birliği üyesi olmayan ülkelere göre patent başvuru sayısı %96 (0.04-1=-0.96) daha azdır şeklinde yorum yapılabilir.

Model tahmin katsayılarını *marjinal etkiler* veya *kesikli değişim* terimleri üzerinden değerlendirmek model sonuçlarını anlamak açısından daha faydalı bir yaklaşımdır. *Marjinal etkiler* açıklayıcı değişkenin sürekli olduğu, *kesikli değişim* açıklayıcı değişkenin ikili veya faktör değişkenler (kategorik değişkendeki bir birimlik değişimi göstermek için) olduğu durumda kullanılmaktadır.

Marjinal etkiler aslında diğer tahmincilerin belirli bir seviyede (genellikle ortalama veya medyan değerleri için) sabit varsayıldığı durumda tahmin edilen tahminler ve onların tahmin edilen olasılıkları arasındaki ilişkiyi göstermek için kullanılır. Özellikle marjinal etkiler, tahmin edicinin değeri bir birim değiştiğinde çıktının olasılığının nasıl değiştiğine ilişkin ilişkiyi ölçmek için kullanılır. Aynı şekilde diğer tahminler belirli bir seviyede sabit varsayıldığında tahmincilerin belirli bir veri değişim değerine karşılık, çıktı değişkeninin anlık değişim oranı olarak da yorumlanabilir. Bu anlamıyla değişim oranı, eğim ve kısmi türev kavramlarını karşılamaktadır.

En küçük kareler yönteminde herhangi bir k'inci birey için tahmin edilen β_k regresyon katsayıları eğim olarak kullanılabilirken, poisson ve negatif binom gibi doğrusal olmayan modeller için bu şekilde kullanılamaz. Bu nedenle marjinal etkilerin hesaplanması söz konusu bilgilerin elde edilmesi açısından gereklidir.

Sürekli bağımsız değişkenler üzerinden hesaplanan marjinal etkiler için temel olarak iki farklı uygulama bulunmaktadır. Birincisi, ortalamaya göre marjinal etkilerde diğer değişkenlerin örneklem ortalamasındaki veri durumlarına göre tahminleme yapılır. İkincisi, ortalama marjinal etkilerde, modeldeki tahmin edilen saymaların (bağımlı değişkenin) ortalaması veri durum olarak dikkate alınmaktadır.

Doğrusal olmayan modellerde ortalamaya göre marjinal etkiler gözlemler boyunca değişebilen sürekli tahmincilerdir. Belirli bir sürekli tahminci için ortaya çıkan etkinin altında yatan mantık aşağıdaki gibidir (Hilbe, 2011; 126):

$$\frac{\partial E(y_i|x_i)}{\partial x_i} = \frac{\partial \mu_i}{\partial x_i} = \frac{\partial e^{x'_i\beta}}{\partial x_i} = \frac{\partial e^{x'_i\beta}}{\partial x'_i\beta} \frac{\partial x'_i\beta}{\partial x_i} = e^{x'_i\beta} \beta_k = \exp(x'_i\beta) \beta_k \quad (15.9)$$

Poisson ve negatif binom modellerinde herhangi bir sürekli tahmin edici için marjinal etkiler aşağıdaki gibi gösterilebilir:

$$\frac{\partial E(y_i|x_i)}{\partial x_{ik}} = E(y_i|x_i) \beta_k = \exp(x'_i\beta_k) \beta_k \quad (15.10)$$

Ortalama marjinal etkiler ise $\hat{\beta}_k \bar{y}$ olacak şekilde hesaplanır. Ayrıca hesaplanan marjinal etkiler (marginals effects, ME) yardımıyla esneklikler $ME \cdot y/x$ ve yarı esneklikler ise $\mu = \exp(x'_i\beta)$ şeklinde hesaplanabilmektedir. Dikkate edilirse yarı esneklikler poisson ve negatif binom modellerindeki tahmin katsayıları ile aynıdır.

Son olarak bağımsız değişkenin kesikli (ikili değişken) olması durumunda *kesikli değişim* ortalamaya göre ve ortalama olmak üzere iki farklı yaklaşım ile hesaplanmaktadır. Ortalamaya göre kesikli değişim aşağıdaki gibi formülize edilebilir:

$$\frac{\Delta \Pr(y_i|x_i=1|x_i=0)}{\Delta x_k} \quad (15.11)$$

Ortalama kesikli değişim ise kısmi etkiler olarak bilinir ve bağımsız değişkendeki değişimin bağımlı değişkende neden olduğu ortalama etkiyi ölçer. Stata programında marjinal etkiler tahminlemenin ardından kullanılan `margins` komutu yardımıyla elde edilirken, kesikli değişim için prosedür biraz daha farklıdır (Hilbe, 2011). Tablo 15.4.a ve Tablo 15.4.b'de ortalamaya göre marjinal etkiler ile ortalama marjinal etkiler gösterilmektedir.

Tablo 15.4.a. Örneklem Ortalaması Panel Poisson: Ortalamaya Göre Marjinal Etkiler

quietly xtpoisson patent arge makale startup gdp, pa
margins, dydx(*) atmean

```

Conditional marginal effects          Number of obs    =      680
Model VCE      : Conventional
Expression     : Exponentiated linear prediction, predict()
dy/dx w.r.t.  : arge makale startup gdp
at
      arge      =    1.492838 (mean)
      makale    =   14330.27 (mean)
      startup   =    7.348529 (mean)
      gdp       =    .0316441 (mean)

```

	Delta-method					
	dy/dx	Std. Err.	z	P> z	[95% Conf. Interval]	
arge	22560.88	13.26205	1701.16	0.000	22534.89	22586.88
makale	-.3502594	.000986	-355.23	0.000	-.352192	-.3483268
startup	-60.59524	1.341778	-45.16	0.000	-63.22508	-57.9654
gdp	32499.18	96.26819	337.59	0.000	32310.5	32687.86

Ortalamaya göre marjinal etkilerin hesaplandığı Tablo 15.4.a'ya göre Ar-Ge harcamalarının marjinal etkisi: Ar-Ge dışındaki diğer değişkenler örneklem ortalaması değerlerinde iken Ar-Ge harcamalarının GSMH içindeki payı bir birim arttığında patent sayısının %22560 artacağını gösterir. Makale sayısı dışındaki diğer değişkenler örneklem ortalaması değerlerinde iken makale sayısında bir adet artış ise patent sayısında %35'lik bir azalışa neden olacaktır.

Tablo 15.4.b. Örneklem Ortalaması Panel Poisson: Ortalama Marjinal Etkiler

```
quietly xtpoisson patent arge makale startup gdp, pa
margins, dydx(*)
```

```

Average marginal effects          Number of obs   =      680
Model VCE      : Conventional
Expression     : Exponentiated linear prediction, predict()
dy/dx w.r.t.  : arge makale startup gdp

```

	Delta-method					
	dy/dx	Std. Err.	z	P> z	[95% Conf. Interval]	
arge	336723.3	181.8304	1851.85	0.000	336366.9	337079.7
makale	-.5227654	.0015562	-335.92	0.000	-.5258155	-.5197153
startup	-90.43896	2.005076	-45.11	0.000	-94.36884	-86.50908
gdp	48505.33	141.35	343.16	0.000	48228.29	48782.37

Ortalama marjinal etkilerin hesaplandığı Tablo 15.4.b’de Ar-Ge’nin ortalama marjinal etkisine göre: Ar-Ge’de meydana gelen her bir birimlik artışta patent sayısı ortalama 336723.3 adet artar. İlave bir adet start-up firma kurulduğunda ise patent sayısı ortalama 90.43 adet azalır.

Tablo 15.4.c. Örneklem Ortalaması Panel Poisson: Esneklikler

quietly xtpoisson patent arge makale startup gdp, pa
margins, eyex(*)

Average marginal effects Number of obs = 680

Model VCE : Conventional

Expression : Exponentiated linear prediction, predict()

ey/ex w.r.t. : arge makale startup gdp

		Delta-method				
	ey/ex	Std. Err.	z	P> z	[95% Conf. Interval]	
arge	1.11617	.0004344	2569.33	0.000	1.115319	1.117021
makale	-.1663434	.0004912	-338.62	0.000	-.1673062	-.1653806
startup	-.0147571	.0003272	-45.10	0.000	-.0153984	-.0141158
gdp	.0340821	.0000997	341.77	0.000	.0338867	.0342776

Tablo 15.4.c'deki esneklik değerlerine göre: Ar-Ge harcamalarında %1'lik artış yaşandığında patent sayısı %1.11 artmaktadır. Aynı mantıkla makale sayısında %1'lik artış yaşandığında patent sayısı %0.16 azalmaktadır.

Tablo 15.4.d. Örneklem Ortalaması Panel Poisson: Yarı Esneklikler

quietly xtpoisson patent arge makale startup gdp, pa
margins, eydx(*)

Average marginal effects Number of obs = 680

Model VCE : Conventional

Expression : Exponentiated linear prediction, predict()

ey/dx w.r.t. : arge makale startup gdp

		Delta-method				
	ey/dx	Std. Err.	z	P> z	[95% Conf. Interval]	
arge	7.476831	.00291	2569.33	0.000	7.471128	7.482535
makale	-.0000116	3.43e-08	-338.62	0.000	-.0000117	-.0000115
startup	-.0020082	.0000445	-45.10	0.000	-.0020954	-.0019209
gdp	1.077045	.0031514	341.77	0.000	1.070868	1.083222

Dikkat edilirse Tablo 15.4.d'den elde edilen sonuçlar ile Tablo 15.3'ten elde edilen sonuçlar aynıdır. Bunun nedeni daha önceden belirtildiği üzere yarı esneklikler $\mu = \exp(x_i'\beta)$ şeklinde hesaplanmaktadır. Bu durumda yarı esneklik katsayıları, örneklem ortalaması poisson ve örneklem ortalaması negatif binom modellerindeki tahmin katsayıları ile aynı şekilde hesaplanır.

Tablo 15.5.a. Örneklem Ortalaması Panel Poisson (Kukla Değişkenli): Ortalamaya Göre Kesikli Değişim

```
qui xtpoisson patent arge makale startup gdp i.europe , pa
margins, dydx(europe) atmean
```

```
Conditional marginal effects          Number of obs    =          680
Model VCE      : Conventional
Expression     : Exponentiated linear prediction, predict()
dy/dx w.r.t.   : 1.europe
at             : arge                =    .1492838 (mean)
                makale               =   14330.27 (mean)
                startup               =    7.348529 (mean)
                gdp                  =    .0316441 (mean)
                0.europe              =    .5764706 (mean)
                1.europe              =    .4235294 (mean)
```

	Delta-method					
	dy/dx	Std. Err.	z	P> z	[95% Conf. Interval]	
1.europe	-53995.79	43.3519	-1245.52	0.000	-54080.76	-53910.82

Note: dy/dx for factor levels is the discrete change from the base level.

Avrupa Birliği üyeliğine bağlı olarak oluşturulan kukla değişken için hesaplanan ortalamaya göre kesikli değişim değeri Tablo 15.5.a'da yer almaktadır. Ortalamaya göre kesikli değişimi daha doğru tahmin etmek için kukla değişkenin önüne "i." ifadesi konulmalıdır. "i." ifadesi kategorik değişkenleri belirtmek için kullanılır. Tahminlemenin ardından **margins** komutu ile kesikli etkiler tahmin edilir ve tahmin edilen katsayının yorumu: modelde kullanılan diğer değişkenler örneklem ortalamasında iken Avrupa Birliği üyesi olan ülkelerin patent sayısı Avrupa Birliği üyesi olmayan ülkelere nazaran 53995 adet daha azdır.

Tablo 15.5.b. Örneklem Ortalaması Panel Poisson (Kukla Değişkenli): Ortalama Kesikli Değişim Etkiler

```
quietly xtpoisson patent arge makale startup gdp i.europe , pa
margins, dydx (europe)
```

```
Average marginal effects          Number of obs    =          680
Model VCE      : Conventional
Expression     : Exponentiated linear prediction, predict()
dy/dx w.r.t.   : 1.europe
```

	Delta-method					
	dy/dx	Std. Err.	z	P> z	[95% Conf. Interval]	
1.europe	-75137.79	47.85141	-1570.23	0.000	-75231.58	-75044

Note: dy/dx for factor levels is the discrete change from the base level.

Avrupa Birliği üyeliğine bağlı olarak oluşturulan kukla değişken için hesaplanan ortalama kesikli değişim değeri ise Tablo 15.5.b’de yer almaktadır. Kukla değişken için ortalama kesikli değişimi daha doğru tahmin etmek için aynı şekilde kukla değişkenin önüne “i.” ifadesi konulmalıdır. Tahminlemenin ardından margins komutu ile kesikli etkiler tahmin edilir. Ortalama kesikli değişim kısmi etki olarak bilinir ve tahmin edilen katsayı şu şekilde yorumlanır: modelde kullanılan diğer değişkenler sabit iken Avrupa Birliği üyesi olan ülkelerin patent sayısı Avrupa Birliği üyesi olmayan ülkelere nazaran 75137 adet daha azdır.

y_{it} ’nin eş yayılım (equidispersed) ve y_{it} ’nin i . birim için zamandan bağımsız olduğu varsayılmasına rağmen mevcut (default) standart hatalar yanlış olma eğilimindedir (Cameron ve Trivedi, 2015). Bu nedenle kısa panel verilerde standart olarak birim üzerinden kümeleme (cluster) yapılarak kümelenmiş dirençli (cluster-robust) standart hatalar veya panel dirençli standart hatalar kullanılır (Cameron ve Trivedi, 2015). Bu nedenle kısa panel verilerde standart olarak daha tutarlı tahminlerin elde edilmesi için birim üzerinden kümeleme yapılarak kümelenmiş dirençli standart hataların hesaplanması gerekir. Örneklem ortalaması panel poisson tahmini yapılırken iki farklı varyans tahmin edicisi opsiyonu olan **vce(robust)** veya **vce(bootstrap)** kullanılarak birim üzerinden kümelenmiş dirençli standart hatalar elde edilir.

Tablo 15.3’te yer alan temel model ile dirençli standart hataların elde edildiği modelleri tahmin sonuçlarına dair karşılaştırma yapmak için aşağıdaki komutlar yardımıyla Tablo 15.6 oluşturulmuştur.

```
qui xtpoisson patent arge makale startup gdp, pa nolog irr
estimates store temel_pa
qui xtpoisson patent arge makale startup gdp, pa nolog irr vce(robust)
estimates store robust_pa
qui xtpoisson patent arge makale startup gdp, pa nolog irr vce(boot)
estimates store bootstrap_pa
estimates table temel_pa robust_pa bootstrap_pa , b se t p
```

Tablo 15.6. Örneklem Ortalaması Panel Poisson Modeli İçin Dirençli Tahminçiler

Variable	temel_pa	robust_pa	bootstrap_pa
arge	7.4768313	7.4768313	7.4768313
	.00291003	2.0010784	2.7008161
	2569.33	3.74	2.77
	0.0000	0.0002	0.0056
makale	-.00001161	-.00001161	-.00001161
	3.428e-08	.00001909	.00004164
	-338.62	-0.61	-0.28
	0.0000	0.5432	0.7804
startup	-.00200816	-.00200816	-.00200816
	.00004453	.0081359	.01032777
	-45.10	-0.25	-0.19
	0.0000	0.8050	0.8458
gdp	1.0770449	1.0770449	1.0770449
	.00315138	.48680868	.69522591

		341.77	2.21	1.55
		0.0000	0.0269	0.1213
_cons		9.345597	9.345597	9.345597
		.00118701	.58584622	.58743156
		7873.22	15.95	15.91
		0.0000	0.0000	0.0000

legend: tahmin katsayısı/standart hata/ z değeri /p değeri

Standart hataların düzeltilmesi sonucunda modellerden elde edilen tahmin katsayıları değişmez, ancak tahmin ediciler daha tutarlı hâle gelir. Dikkat edilirse dirençli standart hataların üretilmesi sonucunda makale sayısı, start-up firma sayısı ve büyüme oranı istatistiksel olarak anlamsız olmaktadır. Tablo 15.6'da dirençli modellere ait standart hatalar ile temel modele ait standart hatalar arasında ortaya çıkan farkın nedeni (bağımlı değişken olan patent sayısının) aşırı yayılıma sahip olmasıdır. Dirençli standart hataların üretilmesi poisson modellerindeki olası aşırı yayılım probleminin çözümü için bir çözüm yoludur (Hilbe, 2014).

Liang ve Zeger (1986) GEE modellerinde korelasyon yapısının eksik belirlenmesi durumunda bile tutarlı tahminlerin elde edildiğini belirtmektedir. Ancak özellikle örneklem büyüklüğünün az olduğu durumlarda grup içi korelasyon matrisinin doğru tanımlanması durumunda tahmin edilen parametrelerin etkinliği kesin biçimde artmaktadır (Wang ve Carey 2003). Bunun için y_{it} ve y_{is} ($t \neq s$) arasındaki korelasyon yapısı ($R_{t,s}$) farklı varsayımlarla kısıtlanabilir.

Bağımsız (Independent) Korelasyon Yapısı:

$$R_{t,s} = \begin{cases} 1 & \text{eğer } t = s \text{ ise} \\ 0 & \text{diğer durumlarda} \end{cases}$$

Değişebilir (Exchangeable) Korelasyon Yapısı

$$R_{t,s} = \begin{cases} 1 & \text{eğer } t = s \text{ ise} \\ \rho & \text{diğer durumlarda} \end{cases}$$

Otoregresif (AR) Korelasyon Yapısı

$$R_{t,s} = \begin{cases} 1 & \text{eğer } t = s \text{ ise} \\ \rho^{|t-s|} & \text{diğer durumlarda} \end{cases}$$

Durağan Olmayan (g) (Nonstationary (g)) Korelasyon Yapısı

$$R_{t,s} = \begin{cases} 1 & \text{eğer } t = s \text{ ise} \\ \rho_{ts} & \text{eğer } 0 < |t - s| \leq g, \quad \rho_{ts} = \rho_{st} \\ 0 & \text{diğer durumlarda} \end{cases}$$

Korelasyon yapısının, diyagonal elemanları 1, g. yan köşegen dışındaki elemanları sıfır olacak şekilde bir kısıtlamayı ifade etmektedir.

Yapılandırılmamış (Unstructured) Korelasyon Yapısı

$$R_{t,s} = \begin{cases} 1 & \text{eğer } t = s \text{ ise} \\ \rho_{ts} & \text{diğer durumlarda,} \end{cases} \quad \rho_{ts} = \rho_{st}$$

Korelasyon yapısının, sadece diyagonal elemanlarının 1 olduğu kısıtlamayı ifade eder.

Bu varsayımlar kolay bir şekilde komutun opsiyon kısmına yazılarak modele eklenir. Korelasyon opsiyonu `corr()` olup yukarıda belirtilen korelasyon yapıları kullanılarak model tahminlerinin daha tutarlı olmaları sağlar. `xtpoisson, pa` komutunda default olarak `corr(exchangeable)` varsayımı yapmaktadır. `corr(exchangeable)` ve `corr(independence)` dışındaki korelasyon varsayımların kullanılabilmesi için (`xtset id time`) zaman boyutunun da tanımlanmış olması gerekir. `xtpoisson, pa` opsiyonunda kullanılabilecek korelasyon yapıları aşağıda belirtilmiştir.

<code>corr(independence)</code>	: gözlemler arası korelasyonun 1'e eşit olması
<code>corr(exchangeable)</code>	: gözlemlerin eş bağlantılı olması
<code>corr(unstructured)</code>	: her gözlem çiftinin farklı değere sahip olması
<code>corr(ar#)</code>	: gözlemler arası ilişkinin # kadar gecikmeli olması
<code>corr(nonstationary#)</code>	: korelasyon matrisinin #. zaman aralığında ilişkili olmaması
<code>corr(fixed matname)</code>	: kullanıcı tarafından belirlenen herhangi özel yapı

Korelasyon varsayımlarına bağlı olarak yapılan tahminlerden sonra `matrix list e(R)` yardımıyla tahminde kullanılan korelasyon matrisi görüntülenir. Farklı korelasyon opsiyonuna `corr()` bağlı olarak elde edilen sonuçlar toplu olarak Tablo 15.7 ve Tablo 15.8'de verilmiştir.

```
quietly xtpoisson patent arge makale startup gdp, pa corr(ind)
estimates store robust_ind
quietly xtpoisson patent arge makale startup gdp, pa corr(exc)
estimates store robust_exc
quietly xtpoisson patent arge makale startup gdp, pa corr(ar1)
estimates store robust_ar1
quietly xtpoisson patent arge makale startup gdp, pa corr(ar2)
estimates store robust_ar2
estimates table robust_ind robust_exc robust_ar1 robust_ar2, b se t p
```

Tablo 15.7. Düzeltilmemiş Standart Hatalar Altında Farklı Corr() Opsiyonlarına Bağlı Sonuçlar

Variable	robust_ind	robust_exc	ar1	robust_ar2
arge	7.8209655 .00189727 4122.21 0.0000	7.4768313 .00291003 2569.33 0.0000	6.0051649 .00418786 1433.94 0.0000	6.0273883 .00411616 1464.32 0.0000
makale	-.00001128 1.847e-08 -611.03 0.0000	-.00001161 3.428e-08 -338.62 0.0000	-3.765e-06 1.811e-08 -207.92 0.0000	-3.700e-06 1.687e-08 -219.34 0.0000
startup	.04043781 .00005638 717.18 0.0000	-.00200816 .00004453 -45.10 0.0000	.00238927 .00002443 97.79 0.0000	.00253212 .00002219 114.10 0.0000
gdp	2.2573034 .00485404 465.04 0.0000	1.0770449 .00315138 341.77 0.0000	.29459605 .0017862 164.93 0.0000	.25022369 .00165949 150.78 0.0000
_cons	8.5327886 .00071333 11961.85 0.0000	9.345597 .00118701 7873.22 0.0000	9.3438256 .0013043 7163.85 0.0000	9.3125859 .00124909 7455.48 0.0000
Wald	2.11e+07	6821519	2167708	2268589

legend: tahmin katsayısı/standart hata/ z değeri /p değeri

```

quietly xtpoisson patent arge makale startup gdp, pa vce(robust) corr(ind)
estimates store robust_ind
quietly xtpoisson patent arge makale startup gdp, pa vce(robust) corr(exc)
estimates store robust_exc
quietly xtpoisson patent arge makale startup gdp, pa vce(robust) corr(ar1)
estimates store robust_ar1
quietly xtpoisson patent arge makale startup gdp, pa vce(robust) corr(ar2)
estimates store robust_ar2
estimates table robust_ind robust_exc robust_ar1 robust_ar2, b se t p

```

Tablo 15.8. Dirençli Standart Hatalar Altında Farklı Corr() Opsiyonlarına Bağlı Sonuçlar

Variable	robust_ind	robust_exc	robust_ar1	robust_ar2
arge	7.8209655 2.3672218 3.30 0.0010	7.4768313 2.0010784 3.74 0.0002	6.0051649 1.1127157 5.40 0.0000	6.0273883 1.2239871 4.92 0.0000
makale	-.00001128 .00002927 -0.39 0.6998	-.00001161 .00001909 -0.61 0.5432	-3.765e-06 2.924e-06 -1.29 0.1979	-3.700e-06 2.800e-06 -1.32 0.1865

startup	.04043781	-.00200816	.00238927	.00253212
	.04394904	.0081359	.00237125	.00223313
	0.92	-0.25	1.01	1.13
	0.3575	0.8050	0.3136	0.2568
gdp	2.2573034	1.0770449	.29459605	.25022369
	3.3982371	.48680868	.14162429	.13391991
	0.66	2.21	2.08	1.87
	0.5065	0.0269	0.0375	0.0617
_cons	8.5327886	9.345597	9.3438256	9.3125859
	.90427368	.58584622	.56827528	.59270948
	9.44	15.95	16.44	15.71
	0.0000	0.0000	0.0000	0.0000
Wald	23.29	19.63	54.54	53.84

Legend: tahmin katsayısı/standart hata/ z değeri /p değeri

Tablo 15.7 ve Tablo 15.8. farklı corr() opsiyonlarına bağlı olarak model tahmin sonuçlarını göstermektedir. Bu sonuçlara göre korelasyon yapılarının tahmin sonuçlarında farklılıklara neden olduğu görülmektedir. Kurulan modeller arasında seçim yapılmasında ilk olarak korelasyon yapısının seçiminde araştırmacı tarafından yapılacak önsel varsayımlar belirleyici olmaktadır. Ardından modeller arasında istatistiksel olarak seçim yapılabilmesi için bazı yöntemler bulunmaktadır. Bu yöntemlerden olan Wald testi ve Quasi-Likelihood Bilgi Kriteri (Yarı-olabilirlik Bilgi Kriteri, QIC) aşağıda açıklanmıştır.

Modelde tahmin edilen katsayıların sıfırdan farklı olup olmadığı Wald testi yardımıyla yapılır. Wald testinde kurulan hipotezler Denklem (15.12.a) ve Denklem (15.12.b) yer almaktadır.

$$H_0: \beta_i = 0, \quad i = 1, \dots, N \quad (15.12.a)$$

$$H_1: \beta_i \neq 0 \quad (15.12.b)$$

Eğer hesaplanan değer tablo değerinden büyük ise H_0 reddedilir ve katsayıların toplu olarak sıfırdan farklı olduğu söylenir. Genellikle p değerinin 0.05'ten küçük olması durumunda H_0 reddedilir. Wald testi aynı zamanda modele açıklayıcı değişken eklenmesi veya modelden açıklayıcı değişken çıkartılması gibi durumlarda araştırmacıya ipuçları verir. Eğer Wald testi reddedilirse modelden değişken çıkartılabilir, kabul edilirse modele yeni değişkenler eklenebilir.

Tablo 15.7 ve Tablo 15.8'de görüldüğü üzere tahmin edilen modellerdeki Wald chi kare test değerleri, katsayı tahminlerindeki z değerleriyle birlikte hareket ettiği, yüksek z değerlerinin olduğu modellerde yüksek Wald test değerlerinin elde edildiği görülmektedir. Gerçekte açıklayıcı değişkenler arasında korelasyon olması durumunda, bu ilişkinin oluşmaması beklenirken, bu durum yüksek Wald değerlerinin elde edilmesindeki basit mantığı göstermektedir. Wald test değerleri modeller arası bir karşılaştırma yapmak için yeterli bilgi içermediğinden model karşılaştırması için bilgi kriterlerine bakılması gerekmektedir.

Bağımsız gözlemler için en çok olabilirlik teorisi üzerine kurulan genelleştirilmiş doğrusal modellerin (GLM) aksine GEE yarı olabilirlik (quasi-likelihood) teorisi üzerine kuruludur ve gözlemlerinin dağılımı üzerine herhangi bir varsayımda bulunmaz (Wedderburn, 1974; McCullagh

ve Nelder 1989). Bu nedenlerle GLM model seçiminde sıklıkla kullanılan Akaike Bilgi Kriteri (AIC) GEE için uygulanabilir değildir (Cui, 2007).

Örneklem ortalaması panel poisson tahminlerinde genelleştirilmiş tahmin denklemleri kullanıldığı ve benzerlik fonksiyonu tanımlanmadığı için Akaike Bilgi Kriteri (AIC) kullanılamamaktadır. Bu nedenle Quasi-likelihood Bilgi Kriteri (QIC) yardımıyla hem uygun modelin hem de uygun korelasyon yapısının seçimine karar verilebilir. QIC değerlerinin hesaplanması GEE üzerinden yapılmaktadır. Pan (2001)'nin önerdiği GLM'de model seçimi için kullanılan AIC'nin, Cui (2007) tarafından geliştirilen versiyonu olan QIC, GEE için kullanılmaktadır.

GLM varsayımı altında $g(\mu) = \beta'x$ [$g()$ herhangi bir bağlantı fonksiyonu, poisson modeli için log] AIC aşağıdaki gibi tanımlanır:

$$AIC = -2LL + 2p \quad (15.13)$$

Burada LL log olabilirlik p modeldeki parametre sayısıdır. Pan (2001) GEE için AIC'yi geliştirmiş ve ceza terimini $2p$ olarak ayarlamıştır. Bu durumda Quasi-likelihood Bilgi Kriteri:

$$QIC = -2Q(\hat{\mu}; I) + 2\text{trace}(\hat{\Omega}_I^{-1}\hat{V}_R) \quad (15.14)$$

Burada I bağımsız quasi-likelihood hesaplanmasında kullanılan kovaryans yapısını temsil ederken, $\hat{\mu} = g^{-1}(x\hat{\beta})$, $\hat{V}_R$; R korelasyon yapısını kullanan dirençli varyans tahmin edicilerini temsil etmektedir. $\hat{\Omega}_I$ bağımsız korelasyon yapısı varsayımı altında elde edilen bir diğer varsayans tahmin edicisidir. En düşük QIC değerini veren korelasyon matrisi, uygun korelasyon yapısı olarak tercih edilir.

$\text{trace}(\hat{\Omega}_I^{-1}\hat{V}_R) \approx \text{trace}(I) = p$ olduğunda QIC'nin basitleştirilmiş versiyonu olan QIC_u elde edilir.

$$QIC_u = -2Q(\hat{\mu}; I) + 2p \quad (15.15)$$

Burada $\hat{\Omega}_I$ ve $\hat{V}_R$ asimtotik olarak birbirine eşit olduğu varsayıldığı için QIC_u optimal kovaryans yapısının seçimi için kullanılamaz (Cui,2007).

Denklem (15.14) ve (15.15) yer alan quasi-likelihood genel yapısı aşağıdaki gibidir (Wedderburn, 1974):

$$Q(\mu) = \int_y^\mu \frac{y-t}{\phi V(t)} dt \quad (15.16)$$

Burada ϕ yayılım parametresi, $V()$ varyans fonksiyonunu temsil etmektedir.

QIC `gee`, `xtgee` komutları yardımıyla kolayca elde edilebilmektedir. Daha önce de belirtildiği gibi örneklem ortalaması panel poisson modelleri `xtgee` komutu yardımıyla da tahmin edilir. QIC değerlerinin hesaplanması için `qic` komutunun yüklenmesi gerekmektedir. Farklı korelasyon yapılarına (`corr(exc)`, `corr(ind)` `corr(ar1)`, vs.) göre QIC sonuçları aşağıdaki komutlarla elde edilir. Farklı korelasyon yapılarına bağlı olarak QIC sonuçları Tablo 15.5'te yer almaktadır.

```
ssc install qic
```

```
qic patent arge makale startup gdp, i(id) t(time) fam(poisson) link(log) corr(ind)
qic patent arge makale startup gdp, i(id) t(time) fam(poisson) link(log) corr(exc)
qic patent arge makale startup gdp, i(id) t(time) fam(poisson) link(log) corr(ar1)
qic patent arge makale startup gdp, i(id) t(time) fam(poisson) link(log) corr(ar2)
```

Tablo 15.9. Farklı Korelasyon Yapılarına Bağlı Olarak QIC Sonuçları

Corr(ind)	Corr(exc)	Corr(ar1)	Corr(ar2)
Corr = ind	Corr = exc	Corr = ar1	Corr = ar2
Family = poisson	Family = poisson	Family = poisson	Family = poisson
Link = log	Link = log	Link = log	Link = log
p = 5	p = 5	p = 5	p = 5
Trace = 9587985.238	Trace = 11299732.998	Trace = 6145601.221	Trace = 5840303.020
QIC = -414667232.909	QIC = -407738293.019	QIC = -419103401.545	QIC = -419788212.457
QIC_u = -433843193.385	QIC_u = -430337798.015	QIC_u = -431394593.987	QIC_u = -431468808.498

Birimler arası farklı korelasyon yapılarından en düşük QIC değeri alan modelin tercih edilmesi gerekir. Veri setinde en uygun korelasyon yapısının Tablo 15.9'da görüleceği üzere en düşük QIC değeri -419788212.378 ile **corr(ar2)** yani gözlemler arası ilişkinin 2 gecikmeli olmasına izin verilen modelin olduğu söylenebilir. Eğer gözlemler arası korelasyon yapısı bağımsız ise örneklem ortalaması panel poisson ile poisson modeli aynı sonucu verir. Bir başka ifade ile **poisson** komutu ile **xtpoisson**, **corr(ind)** eş tahminler üretir.

Bu bölümde kullanılan Stata komutları aşağıda toplu olarak verilmiştir.

```
use "DOSYAKONUMU\patent.dta", clear
xtset id time
xtsum
describe
xtdescribe
histogram patent, normal percent
summarize patent, detail
tabulate patent
summarize
quietly xtset id time
xtpoisson patent arge makale startup gdp, pa nolog
quietly xtpoisson patent arge makale startup gdp, pa
xtpoisson irr
xtpoisson patent arge makale startup gdp ,pa nolog irr
xtpoisson patent arge makale startup gdp eu, pa irr nolog
quietly xtpoisson patent arge makale startup gdp, pa
margins, dydx(*) atmean
margins, dydx(*)
margins, eyex(*)
margins, eydx(*)
quietly xtpoisson patent arge makale startup gdp i.europe , pa
margins, dydx (europe) atmean
margins, dydx (europe)
xtpoisson patent arge makale startup gdp, pa nolog irr vce(robust)
xtpoisson patent arge makale startup gdp, pa nolog irr vce(bootstrap)
quietly xtpoisson patent arge makale startup gdp, pa nolog irr
estimates store temel_pa
```



```

quietly xtpoisson patent arge makale startup gdp, pa nolog irr vce(robust)
estimates store robust_pa
quietly xtpoisson patent arge makale startup gdp, pa nolog irr vce(boot)
estimates store bootstrap_pa
estimates table temel_pa robust_pa bootstrap_pa , b se t p
quietly xtpoisson patent arge makale startup gdp, pa corr(ind)
estimates store ind
quietly xtpoisson patent arge makale startup gdp, pa corr(exc)
estimates store exc
quietly xtpoisson patent arge makale startup gdp, pa corr(ar1)
estimates store ar1
quietly xtpoisson patent arge makale startup gdp, pa corr(ar2)
estimates store ar2
estimates table ind exc ar1 ar2, b se t p
quietly xtpoisson patent arge makale startup gdp, pa vce(robust) corr(ind)
estimates store robust_ind
quietly xtpoisson patent arge makale startup gdp, pa vce(robust) corr(exc)
estimates store robust_exc
quietly xtpoisson patent arge makale startup gdp, pa vce(robust) corr(ar1)
estimates store robust_ar1
quietly xtpoisson patent arge makale startup gdp, pa vce(robust) corr(ar2)
estimates store robust_ar2
estimates table robust_ind robust_exc robust_ar1 robust_ar2, b se t p
ssc install qic
qic patent arge makale startup gdp, i(id) t(time) fam(poisson) link(log) corr(ind)
qic patent arge makale startup gdp, i(id) t(time) fam(poisson) link(log) corr(exc)
qic patent arge makale startup gdp, i(id) t(time) fam(poisson) link(log) corr(ar1)
qic patent arge makale startup gdp, i(id) t(time) fam(poisson) link(log) corr(ar2)

```

15.1.2. Sabit Etkiler Panel Poisson Modeli

Sabit etkiler panel poisson (FE) modeli temel doğrusal panel modelinden hareketle aşağıdaki gibi olur

$$y_{it} = \alpha_i + x'_{it}\beta + u_{it}, \quad i = 1, \dots, N, \quad t = 1, \dots, T \quad (15.17.a)$$

$$y_{it} \sim \text{Poisson}(\mu_{it}) \quad (15.17.b)$$

Burada sabit etkiler panel poisson modelinde birim etkileri temsil eden α_i parametresinin rassal ve X_{it} açıklayıcı değişkenleri ile potansiyel olarak ilişkili olduğu varsayılır. Denklem (15.3)'de yer alan koşul Denklem 15.18'de birim etkileri içerecek şekilde yeniden tanımlanmıştır.

$$\mu_{it} = \alpha_i \lambda_{it} = \alpha_i \exp(x'_{it}\beta) = \exp[\ln(\alpha_i + x'_{it}\beta)] \quad (15.18)$$

Kısa panel durumunda (veri asimptotik teoreminin T sabitken $n \rightarrow \infty$ olmasını ima etmesinden dolayı) β 'nin tutarlı olarak tahmin edilebilmesi zorlayıcı olabilmektedir. Bu nedenle ilk olarak sıfırsız sayıdaki $\alpha_1, \dots, \alpha_n$ sabit etkiler parametrelerini kontrol altında tutmak gerekir. Doğrusal modellerde sabit etkiler parametrelerini kontrol altında tutmak kolay iken doğrusal olmayan mo-

dellerde için pek kolay değildir. Ancak poisson modellerinde, sabit etkiler parametresi modele çarpım olarak dahil edildiği için kolayca kontrol altına alınır.

Birim etkileri içeren genel panel poisson için oluşturulan parametrik yaklaşım modeli Denklem (15.19)'da yer almaktadır.

$$f(y_i|X_i, \alpha_i, \beta) = \prod_t \exp[-\alpha_i \exp(x'_{it}\beta)][-\alpha_i \exp(x'_{it}\beta)]^{y_{it}}/y_{it}! \quad (15.19)$$

Sabit etkiler panel poisson modellerinde tutarlı β tahminlerinin elde edilebilmesi için temel olarak üç farklı öncü yaklaşım mevcuttur (Cameron ve Trivedi, 2013:351). Birincisi hem β hem de α_i 'nin müşterek olarak en çok olabilirlik yardımıyla tahmin edilmesidir. İkincisi, bireysel etkiler için yeterli istatistikler üzerinden β 'nin koşullu en çok olabilirlik yöntemi ile tahmin edilmesidir. Üçüncüsü, fark dönüşümü ile sabit etkilerin elenerek moment tabanlı olarak tahminlerin elde edilmesidir. Söz konusu üç yaklaşım hakkında daha fazla bilgi edinmek isteyen okuyucular için Winkelmann (2008: 222-227) ile Cameron ve Trivedi (2013: 352-357) çalışmaları önerilmektedir.

Birinci yaklaşımda sabit etkiler panel poisson modelinde yer alan α_i parametresi ile ve β katsayısının eşanlı olarak tahmin edilmesinde Denklem (15.19)'dan yararlanılarak elde edilen log olabilirlik denklemi aşağıda yer almaktadır:

$$\begin{aligned} \ln L(\beta, \alpha) &= \ln[\prod_i \prod_t \{\exp(-\alpha_i \lambda_{it}) (\alpha_i \lambda_{it})^{y_{it}}/y_{it}!\}] \\ &= \sum_i [-\alpha_i \sum_t \lambda_{it} + \ln \alpha_i \sum_t y_{it} + \sum_t y_{it} \ln \lambda_{it} - \sum_t \ln y_{it}!] \end{aligned} \quad (15.20)$$

Burada $\lambda_{it} = \exp(x'_{it}\beta)$ 'dir. Söz konusu denklemin α_i 'ye göre birinci türevi alınıp sıfıra eşitlenirse Denklem (15.21)'de gösterilen α_i için en çok olabilirlik tahmin edicisi elde edilir.

$$\hat{\alpha}_i = \sum_t y_{it} / \sum_t \lambda_{it}. \quad (15.21)$$

Bu ifade, Denklem (15.20)'de yerine koyulursa, yoğunlaştırılmış olabilirlik fonksiyonu (concentrated likelihood) elde edilir. Sadeleştirme sonrasında elde edilen ve Denklem (15.22)'in maksimizasyonu neticesinde asimptotik teori ile tutarlı β tahminleri elde edilir (Cameron ve Trivedi, 2013; Winkelmann, 2008)

$$\ln L_{conc}(\beta) \propto \sum_i \sum_t [y_{it} \ln \lambda_{it} - y_{it} \ln(\sum_s \lambda_{is})] \quad (15.22)$$

Denklem (15.22)'nin β 'ya göre birinci sıra koşulu Denklem (15.23.a)'da basitleştirilmiş formu ise Denklem (15.23.b)'de yer almaktadır.

$$\sum_i \sum_t [y_{it} \ln \lambda_{it} - y_{it} \ln(\sum_s \lambda_{is} X_{is}) / (\sum_s \lambda_{is})] = 0 \quad (15.23.a)$$

$$\sum_{i=1}^N \sum_{t=1}^T X_{it} \left(y_{it} - \frac{\lambda_{it}}{\bar{\lambda}_{it}} \bar{y}_i \right) = 0 \quad (15.23.b)$$

Burada $\lambda_{it} = \exp(x'_{it}\beta)$ ve $\bar{\lambda}_{it} = T^{-1} \sum_t \exp(x'_{it}\beta)$ 'dir. Panel poisson modelde α_i ve β 'nin eşanlı tahmini kısa panel durumunda tutarlı β tahminlerinin elde edilmesine olanak sağlar ve bu nedenle tesadüfi parametre problemi (incidental parameters problem) yoktur.

İkinci yaklaşım, koşullu en çok olabilirlik yönteminin α_i için yeterli istatistik üzerinden koşullandırılarak sabit etkileri elemesidir. Hausman, Hall ve Griliches (1984) tarafından önerilen bu yaklaşımda sabit etkiler panel poisson modelinin i . birimin koşullu olabilirlik katkısının birim üzerinden toplamı $\sum_t y_{it}$ olarak varsayılmakta (koşullandırılmakta) ve tutarlı β tahminlerinin α_i 'lerden bağımsız olarak elde edilmesine olanak sağlamaktadır. Bu durumda tutarlı β tahminleri Denklem (15.24)'te gösterilen koşullu log olabilirlik fonksiyonun maksimizasyonuna dayanmaktadır.

$$\ln L_{cond}(\beta) = \sum_i \left[\ln(\sum_t y_{it})! - \sum_t \ln(y_{it}!) + \sum_t y_{it} \ln \left(\frac{\exp(x'_{it}\beta)}{\sum_s \exp(x'_{is}\beta)} \right) \right] \quad (15.24)$$

Üçüncü yaklaşımda ise doğrusal sabit etkiler modellerinde moment tabanlı tutarlı β tahminlerinin elde edilebilmesi için sabit etkiler elenir. Bunun için ortalama farkın veya birinci farkın dönüşümleri yollarından herhangi biri kullanılarak sabit etkilerin elenmesi sağlanır. Denklem (15.19)'daki varsayımın i . birim için zaman boyunca ortalaması şeklinde tanımlanması ile Denklem (15.25) elde edilir.

$$E[\bar{y}_i | x_{i1}, \dots, x_{iT}, \alpha_i] = \alpha_i \bar{\lambda}_i \quad (15.25)$$

Burada $\bar{y}_i = T^{-1} \sum_t \lambda_{it}$ ve $\bar{\lambda}_i = T^{-1} \sum_t \lambda_{it}$ 'dir. Momentler tabanlı sabit etkiler poisson tahmin edicisi tutarlı β tahminlerinin elde edilebilmesi için Denklem (15.16)'da gösterilen moment koşulunu çözer.

$$\sum_i \sum_t x_{it} \left(y_{it} - \lambda_{it} \frac{\bar{y}_i}{\bar{\lambda}_i} \right) = 0 \quad (15.26)$$

15.1.2.1. Sabit Etkiler Panel Poisson Modeli Stata Uygulaması

Sabit etkiler panel poisson modelinde yukarda kısaca açıklanan üç temel yaklaşım da tutarlı tahmin ediciler üretmektedir (Cameron ve Trivedi, 2013). Stata 16 programı sabit etkiler panel poisson tahminlerini koşullu log olabilirlik fonksiyonu üzerinden yapmaktadır (StataCorp, 2019). Sabit etkiler panel poisson modeli için temel STATA komutu aşağıdaki gibidir.

```
xtpoisson depvar [indepvars] [if] [in] [weight] , fe [FE_options]
```

Patent sayısı ile Ar-Ge harcamalarının GSMH içindeki payı, yayınlanan akademik makale sayısı, kurulan yeni start-up işletme sayısı arasındaki ilişkinin sabit etkiler panel poisson tahminlerini elde etmek için aşağıdaki komut kullanılır.

Tablo 15.10. Sabit Etkiler Panel Poisson Tahmin Sonuçları**xtpoisson patent arge makale startup gdp ,fe nolog**

Kullanılan opsiyonlar:

fe : sabit etkiler

nolog : iterasyonları göstermemesi için

Conditional fixed-effects Poisson regression

Group variable: id

Number of obs = 680

Number of groups = 38

Obs per group:

min = 15

avg = 17.9

max = 18

Wald chi2(4) = 3.02e+06

Prob > chi2 = 0.0000

Log likelihood = -1960244.8

patent	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
arge	11.95565	.0070542	1694.83	0.000	11.94182	11.96947
makale	9.87e-06	8.90e-08	110.93	0.000	9.69e-06	.00001
startup	.0039339	.0001084	36.29	0.000	.0037214	.0041464
gdp	.8422943	.0103898	81.07	0.000	.8219307	.8626579

Daha önceden de belirtildiği üzere saymalı model tahminleri neticesinde elde edilen katsayılar doğrudan yorumlanmaya uygun olmadığı için *incidence-rate ratio* (irr)'nin hesaplanması gerekmektedir.

Tahmin edilen katsayıların üsteli ($e^{\hat{\beta}}$) alınarak hesaplanabileceği gibi irr opsiyonu kullanılarak da IRR elde edilebilir. Ancak önceden de belirtildiği üzere temel modeldeki katsayılar ait standart hatalar ile IRR tahminlerine ait standart hatalar kullanılan hesaplama yöntemi nedeniyle farklıdır.

Tablo 15.11. Sabit Etkiler Panel Poisson IRR Tahmin Sonuçları**xtpoisson patent arge makale startup gdp ,fe nolog irr**

Kullanılan opsiyonlar:

fe : sabit etkiler

nolog : iterasyonları göstermemesi için

irr : incidence-rate ratio hesaplanması için

xtpoisson patent arge makale startup gdp ,fe nolog irr

Conditional fixed-effects Poisson regression

Group variable: id

Number of obs = 680

Number of groups = 38

Obs per group:

min = 15

avg = 17.9

max = 18

Wald chi2(4) = 3.02e+06

Prob > chi2 = 0.0000

Log likelihood = -1960244.8

patent	IRR	Std. Err.	z	P> z	[95% Conf. Interval]	
arge	155693.8	1098.293	1694.83	0.000	153556	157861.4
makale	1.00001	8.90e-08	110.93	0.000	1.00001	1.00001
startup	1.003942	.0001088	36.29	0.000	1.003728	1.004155
gdp	2.321688	.0241219	81.07	0.000	2.274888	2.36945

Tablo 15.11'de yer alan model tahmin sonuçlarına bakıldığında Ar-Ge harcamalarının GSMH oranındaki bir birimlik artışın patent sayıları üzerinde istatistiksel olarak anlamlı ve %155692'lik ($155693-1=155692$) çok ciddi bir artışa neden olduğu görülmektedir. Ülkelerin büyüme oranlarındaki bir birimlik artışın ise patent başvuru sayısında istatistiksel olarak anlamlı ve %132'lik ($2.32-1=1.32$) bir artışa neden olduğu söylenebilir. Start-up firma sayısındaki bir birimlik artışın patent sayıları üzerinde istatistiksel olarak anlamlı ve %0.3'lük ($1.003-1=-0.003$) azalışa neden olduğu söylenebilir. Makale sayısındaki bir birimlik artışın ise patent sayıları üzerinde neredeyse etkisinin olmadığı (%0.00001) söylenebilir.

Eğer zaman boyutu çok uzun değilse, poisson ya da negatif binom gibi doğrusal olmayan modellerde sabit etkiler ve rassal etkiler tahmincileri için marjinal etkiler anlamlı bilgiler sunmaz. Dikkat edilirse Tablo 15.10.a ve Tablo 15.10.b'de yer alan marjinal etkilere dair tahmin sonuçları ile Tablo 15.10'da yer alan model tahmin sonuçları aynı değerleri vermektedir. Bu nedenle sabit etkiler panel poisson modellerinde marjinal etkiler yerine esneklikler ve yarı esnekliklerin hesaplanması önerilir. Sabit etkiler panel poisson modeline ait esneklikler ile yarı esneklikler sırasıyla Tablo 15.10.c ve Tablo 15.10.d'de gösterilmektedir.

Tablo 15.10.a. Sabit Etkiler Panel Poisson: Ortalamaya Göre Marjinal Etkiler

```
qui xtpoisson patent arge makale startup gdp ,fe nolog
margins, dydx(*) atmean noatlegend
```

```
Conditional marginal effects      Number of obs      =      680
Model VCE      : OIM
Expression      : Linear prediction, predict()
dy/dx w.r.t.    : arge makale startup gdp
```

		Delta-method				[95% Conf. Interval]	
	dy/dx	Std. Err.	z	P> z			
arge	11.95565	.0070542	1694.83	0.000	11.94182	11.96947	
makale	9.87e-06	8.90e-08	110.93	0.000	9.69e-06	.00001	
startup	.0039339	.0001084	36.29	0.000	.0037214	.0041464	
gdp	.8422943	.0103898	81.07	0.000	.8219307	.8626579	

Tablo 15.10.b. Sabit Etkiler Panel Poisson: Ortalama Marjinal Etkiler

```
qui xtpoisson patent arge makale startup gdp ,fe nolog
margins, dydx(*)
```

```
Average marginal effects      Number of obs      =      680
Model VCE      : OIM
Expression      : Linear prediction, predict()
dy/dx w.r.t.    : arge makale startup gdp
```

		Delta-method				[95% Conf. Interval]	
	dy/dx	Std. Err.	z	P> z			
arge	11.95565	.0070542	1694.83	0.000	11.94182	11.96947	
makale	9.87e-06	8.90e-08	110.93	0.000	9.69e-06	.00001	
startup	.0039339	.0001084	36.29	0.000	.0037214	.0041464	
gdp	.8422943	.0103898	81.07	0.000	.8219307	.8626579	

Tablo 15.10.c. Sabit Etkiler Panel Poisson: Esneklikler

qui xtpoisson patent arge makale startup gdp ,fe nolog
margins, eyex(*)

Average marginal effects				Number of obs	=	680	
Model VCE : OIM							
Expression : Linear prediction, predict()							
ey/ex w.r.t. : arge makale startup gdp							

		Delta-method					
		ey/ex	Std. Err.	z	P> z	[95% Conf. Interval]	

arge		.8601443	.0008517	1009.91	0.000	.858475	.8618136
makale		.100779	.0005565	181.09	0.000	.0996883	.1018698
startup		.0208584	.0005471	38.13	0.000	.0197862	.0219306
gdp		.0182182	.0002204	82.65	0.000	.0177862	.0186503

Tablo 15.10.c’de sunulan sabit etkiler panel poisson modelinden hesaplanan esneklikler: Ar-Ge harcamalarında %1’lik artış yaşandığında patent sayısı %0.86 arttığını, makale sayısında %1’lik artış yaşandığında da patent sayısı %0.10 arttığını göstermektedir.

Tablo 15.10.d. Sabit Etkiler Panel Poisson: Yarı Esneklikler

qui xtpoisson patent arge makale startup gdp ,fe nolog
margins, eydx(*)

Average marginal effects				Number of obs	=	680	
Model VCE : OIM							
Expression : Linear prediction, predict()							
ey/dx w.r.t. : arge makale startup gdp							

		Delta-method					
		ey/dx	Std. Err.	z	P> z	[95% Conf. Interval]	

arge		8.857375	.0149788	591.33	0.000	8.828017	8.886733
makale		.0000073	25.50e-08	132.83	0.000	7.20e-06	7.42e-06
startup		.0029144	.0000766	38.06	0.000	.0027643	.0030645
gdp		.6240161	.007926	78.73	0.000	.6084814	.6395509

Sabit etkiler panel poisson modeline ait yarı esnekliklerin hesaplandığı Tablo 15.10.d’ye bakıldığında: GSMH oranında her birim artış yaşandığında patent sayısının %62, start-up şirket sayısı bir adet arttığında ise patent sayısının %0.29 arttığı söylenebilir.

Poisson modellerinde eş yayılım temel varsayım olmakla birlikte örneğimizde de olduğu gibi veri setlerinin sıklıkla aşırı yayılım sergiledikleri görülür. Eş yayılım varsayımın ihlal edilmesi hâlinde, standart hataların aşağı yönlü olarak tahmin edilmesine ve değişkenlere ait katsayıların istatistiksel anlamlılıkların yukarı yönlü çıkmasına neden olabilir. Bu sorunların üstesinden gelmek adına kümelenmiş dirençli standart hataların kullanılması önerilmektedir (Hilbe, 2014; Ca-

meron ve Trivedi 2013). Sabit etkiler panel poisson tahmini yapılırken üç farklı farklı varyans tahmin edicisi opsiyonu olan `vce(robust)`, `vce(bootstrap)`, `vce(jackknife)` kullanılarak kü-melenmiş dirençli standart hatalar elde edilebilmektedir.

Sabit etkiler panel poisson modeli için standart hatada üç farklı düzeltme yapılarak elde edilen sonuçların toplu karşılaştırmasını yapabilmek için aşağıdaki komutlar yardımıyla Tablo 15.12 oluşturulmuştur. Tablo 15.10'da yer alan model, standart hatada herhangi bir düzeltmenin yapılmadığı temel (default) durumdur. Panel modellerde standart hataların düzeltilmesi için `vce(bootstrap)` veya `vce(jackknife)` opsiyonları önerilmekle birlikte, eğer `vce(jackknife)` opsiyonu kullanılan komut syntax'inde yer alıyorsa jackknife dirençli hataların oluşturulması tavsiye edilmektedir (StataCorp, 2019).

```
quietly xtpoisson patent arge makale startup gdp, fe
estimates store temel_fe
quietly xtpoisson patent arge makale startup gdp, fe vce(robust)
estimates store robust_fe
quietly xtpoisson patent arge makale startup gdp, fe vce(boot)
estimates store bootstrap_fe
quietly xtpoisson patent arge makale startup gdp, fe vce(jack)
estimates store jackknife_fe
estimates table temel_fe robust_fe bootstrap_fe jackknife_fe, b se t p
```

Tablo 15.12. Sabit Etkiler Panel Poisson Modeli İçin Dirençli Tahminciler

Variable	temel_fe	robust_fe	bootstrap_fe	jackknife_fe
arge	11.1955646	11.1955646	11.1955646	11.1955646
	.00705419	8.2735632	9.9705742	13.788449
	1694.83	1.45	1.20	0.87
	0.0000	0.1484	0.2305	0.3915
makale	9.869e-06	9.869e-06	9.869e-06	9.869e-06
	8.896e-08	4.950e-06	.00001249	.00001027
	110.93	1.99	0.79	0.96
	0.0000	0.0462	0.4296	0.3430
startup	.00393389	.00393389	.00393389	.00393389
	.0001084	.00730314	.01111055	.01641608
	36.29	0.54	0.35	0.24
	0.0000	0.5901	0.7233	0.8119
gdp	.84229486	.84229486	.84229486	.84229486
	.01038979	.59113586	.66940482	1.2647448
	81.07	1.42	1.26	0.67
	0.0000	0.1542	0.2083	0.5096
Wald	3.02e+06	847.21	3.86	2.32
p-value	0.0000	0.0000	0.4257	0.0749

gend: tahmin katsayısı/standart hata/ z değeri /p değeri

Tablo 15.12'de farklı dirençli standart hataların üretildiği sabit etkiler panel poisson modelleri sunulmuştur. Sonuçlar incelendiğinde temel (default) sabit etkiler panel poisson sonuçlarında standart hataların diğer modellere göre düşük ve aynı zamanda tahmin edilen tüm katsayıların

0.01 seviyesinde istatistiksel olarak anlamlı oldukları görülmektedir. Dirençli standart hataların üretildiği durumda ise sabit etkiler panel poisson tahminlerinde standart hatalar yükselirken aynı zamanda da tahmin edilen katsayıların istatistiksel anlamlılıkları kaybolmaktadır. Bu durumun olası nedeni bağımlı değişkenin aşırı yayılıma sahip olması gösterilebilir (Hilbe, 2014).

Tutarlı tahmin edicilerin elde edilebilmesi için dirençli modellerin tahmin edilmesi neticesinde katsayılar değişmez ancak standart hatalar ve istatistiksel anlamlılık düzeyleri değişebilmektedir. Aşırı yayılımın tespiti konusunu daha sonraya bırakmak kaydıyla eğer bağımlı değişkenin sağlaması gereken dağılım konusunda şüpheler var ise dirençli standart hataların üretildiği modellerin kullanılması önerilmektedir.

Tablo 15.12’de ilk sütunda düzeltilmemiş standart hataların olduğu temel model yer alırken diğer sütunlarda dirençli standart hataların üretildiği sabit etkiler panel poisson model sonuçları yer almaktadır. Temel modelde tahmin edilen katsayılar istatistiksel olarak anlamlı iken düzeltilmiş standart hatalı modeller istatistiksel olarak anlamlı değildir.

Bu durum poisson modellerinde aşırı yayılımın dikkate alınmaması durumunda sahte regresyonun ortaya çıkabileceğine işaret etmektedir. Eğer aşırı veya eksik yayılım söz konusu ise veya bu konuda bir şüphe var ise dirençli standart hataların üretildiği model seçilmelidir. Dirençli standart hataların üretildiği modellerden model uyum iyiliği açısından istatistiksel olarak hangisinin seçileceği konusunda Wald testi yardımcı olabilir. Bu durumda Tablo 15.12’de ikinci sütunda yer alan ve 0.05 istatistiksel anlamlılık seviyesinde katsayıların bir bütün olarak anlamlı olduğunu gösteren Wald testinden hareketle dirençli modelin seçilmesi önerilmektedir.

Poisson modellerinde aşırı yayılım aşağıdaki nedenlerden dolayı ortaya çıkabilir (Hilbe, 2014:8):

- Önemli değişkenlerin model dışında kalması,
- Verinin sapan gözlem içermesi,
- Modelin gerekli etkileşim terim veya terimlerini içermemesi,
- Değişkenlerin başka bir ölçeğe (log veya karekök gibi) dönüştürülmelerinin gerekli olması gibi nedenlerdir.

Çalışmalarda kullanılan veri setlerinin genellikle aşırı yayılıma sahip olması nedeniyle birçok araştırmacı Tablo 15.12’de yere alan sütun 1’deki temel modele ait standart hataları kullanmamaktadır (Hilbe, 2014). Eğer model gerçekten eş yayılıma sahipse veri seti poisson modele iyi şekilde uyum sağlamak ve temel standart hatalar ile model özdeş olmaktadır. Eğer model aşırı veya eksik yayılıma sahipse temel standart hataların kullanılması açıklayıcı değişkenlere ait daha tutarlı bilgilerin elde edilmesine olanak sağlarken istatistiksel anlamlılıklarının kaybolmasına neden olabilmektedir.

Bu bölümde kullanılan Stata komutları aşağıda toplu olarak yer almaktadır.

```
use "DOSYAKONUMU\patent.dta", clear
xtset id time
xtpoisson patent arge makale startup gdp ,fe nolog
xtpoisson patent arge makale startup gdp ,fe nolog irr
xtpoisson patent arge makale startup gdp ,fe irr nolog vce(robust)
xtpoisson patent arge makale startup gdp ,fe irr nolog vce(bootstrap)
```



```

xtpoisson patent arge makale startup gdp ,fe irr nolog vce(jackknife)
qui xtpoisson patent arge makale startup gdp ,fe nolog
margins, dydx(*) atmean noatlegend
margins, dydx(*)
margins, eyex(*)
margins, eydx(*)
quietly xtpoisson patent arge makale startup gdp, fe
estimates store temel_fe
quietly xtpoisson patent arge makale startup gdp, fe vce(robust)
estimates store robust_fe
quietly xtpoisson patent arge makale startup gdp, fe vce(boot)
estimates store bootstrap_fe
quietly xtpoisson patent arge makale startup gdp, fe vce(jack)
estimates store jackknife_fe
estimates table temel_fe robust_fe bootstrap_fe jackknife_fe, b se t p

```

15.1.3. Rassal Etkiler Panel Poisson Modeli

Rassal etkiler panel poisson (RE) altında temel doğrusal panel model denklemi sabit etkiler panel poisson modeli ile aynıdır. Ancak temel fark rassal etkiler panel poisson modelinde birim etkileri temsil eden α_i parametresi X_{it} açıklayıcı değişkenleri ile ilişkili olmadığı varsayılmaktadır. Bir başka ifade ile rassal etkiler modelinde α_i parametresinin bağımsız ve eş yayılıma (iid) sahip olduğu varsayılır.

$$y_{it} = \alpha_i + x'_{it}\beta + u_{it}, \quad i = 1, \dots, N, \quad t = 1, \dots, T \quad (15.17.a)$$

$$y_{it} \sim \text{Poisson}(\mu_{it}) \quad (15.17.b)$$

$$\mu_{it} = \alpha_i \lambda_{it} = \alpha_i \exp(x'_{it}\beta) = \exp[\ln(\alpha_i + x'_{it}\beta)] \quad (15.18)$$

Denklem 15.17.a Denklem 15.17.b ve Denklem 15.18 ile gösterilen sabit etkiler panel poisson modeli ile rassal etkiler panel poisson modeli aynı olmakla birlikte rassal etkiler panel poisson modelinde α_i rassal etkiler parametresinin dağılımına ilişkin olarak yapılacak olan varsayımlara bağlı olarak farklı rassal etkiler tahmincileri kullanılır. Bu varsayımlardan birincisi α_i 'nin Gamma dağılımına sahip olduğu ($\alpha_i \sim \Gamma[1, \eta]$), ikincisi $\ln \alpha_i$ 'nin normal dağılıma sahip olduğudur ($\ln \alpha_i \sim N[0, \sigma_\epsilon^2]$).

α_i 'nin gamma dağılımına sahip olduğu varsayımı altında Denklem (15.19)'da gösterilen poisson modeli rassal etkiler panel poisson modeli için aşağıdaki gibi olur:

$$f(y_i | X_i, \beta, \eta) = \left[\prod_t \frac{\lambda_{it}^{y_{it}}}{y_{it}!} \right] \times \left(\frac{\eta}{\sum_t \lambda_{it} + \eta} \right)^\eta \times (\sum_t \lambda_{it} + \eta)^{-\sum_t y_{it}} \times \frac{\Gamma(\sum_t y_{it} + \eta)}{\Gamma(\eta)} \quad (15.27)$$

Burada $\lambda_{it} = \exp(x'_{it}\beta)$ olup Denklem (15.27)'nin β 'ya göre birinci sıra koşulu sonucunda rassal etkiler panel poisson tahmin edicileri Denklem (15.28)'de yer almaktadır.

$$\sum_i \sum_t x_{it} \left(y_{it} - \lambda_{it} \frac{\bar{y}_i + \eta/T}{\bar{\lambda}_i + \eta/T} \right) = 0 \quad (15.28)$$

Burada $\bar{\lambda}_i = T^{-1} \sum_t \exp(x'_{it}\beta)$ olup en çok olabilirlik tahmin yöntemi yardımıyla β ve η doğru-
dan tahmin edilir.

Alternatif olarak α_i rassal etkiler parametresinin normal dağılıma sahip olduğu varsayılabilir. Bu durumda eksponansiyel koşullu ortalama fonksiyonu standart olarak kabul edilir. Panel seviye log olabilirlik fonksiyonu Gauss-Hermite tümlevi (quadrature) uyarlaması ile yakınsanır. Bu varsayım için rassal etkiler panel poisson modeli tahmin edilirken **normal** opsiyonu kullanılır.

15.1.3.1. Rassal Etkiler Panel Poisson Model Stata Uygulaması

Rassal etkiler panel poisson modelinde yukarıda kısaca açıklanan yöntemler temel alınarak tahminlerin elde edilmesi için kullanılan STATA komutu aşağıdaki gibidir.

```
xtpoisson depvar [indepvars] [if] [in] [weight] , re [RE_options]
```

Patent sayısı ile Ar-Ge harcamalarının GSMH içindeki payı, yayınlanan akademik makale sayısı, kurulan yeni start-up işletme sayısı arasındaki ilişkinin rassal etkiler panel poisson tahminlerini elde etmek için aşağıdaki komut kullanılır. Daha önceden de belirtildiği üzere saymalı model tahminleri neticesinde elde edilen katsayılar doğrudan yorumlanmaya uygun olmadığı için incidence-rate ratio (irr)'nin hesaplanması gerekmektedir bu nedenle kullanılan komut aşağıda yer almaktadır.

Tablo 15.13. Rassal Etkiler Panel Poisson IRR Tahmin Sonuçları

```
xtpoisson patent arge makale startup gdp ,re nolog irr
```

Kullanılan Opsiyonlar

re : rassal etkiler

nolog : iterasyonları göstermemesi için

irr : incidence-rate ratio hesaplanması için

Random-effects Poisson regression	Number of obs	=	680
Group variable: id	Number of groups	=	38
Random effects u_i ~ Gamma	Obs per group:		
	min	=	15
	avg	=	17.9
	max	=	18
	Wald chi2(4)	=	3.02e+06
Log likelihood = -1960734.4	Prob > chi2	=	0.0000

patent	IRR	Std. Err.	z	P> z	[95% Conf. Interval]	
-----+-----						
arge	155697	1098.316	1694.83	0.000	153559.2	157864.7
makale	1.00001	8.90e-08	110.92	0.000	1.00001	1.00001
startup	1.003941	.0001088	36.29	0.000	1.003728	1.004155
gdp	2.32171	.0241221	81.07	0.000	2.27491	2.369473
_cons	2231.468	638.209	26.96	0.000	1273.924	3908.749
-----+-----						
/lnalpha	1.134048	.1821647			.7770119	1.491084
-----+-----						
alpha	3.108213	.5662066			2.174964	4.441909
-----+-----						

Note: Estimates are transformed only in the first equation.

Note: _cons estimates baseline incidence rate (conditional on zero random effects).

LR test of alpha=0: chibar2(01) = 6.3e+07

Prob >= chibar2 = 0.000

Tablo 15.13'te elde edilen model çıktılarına bakıldığında “Random effects $u_i \sim \text{Gamma}$ ” ifadesi tahmin sonuçlarının α_i 'nin Gamma dağılımına sahip olduğu varsayımı altında elde edildiğini göstermektedir. Tahmin edilen katsayılar incelendiğinde Ar-Ge harcamalarının GSMH oranındaki bir birimlik artışın patent sayıları üzerinde istatistiksel olarak anlamlı ve %155696'lık (155697-1=155696) ciddi bir artışa neden olduğu görülmektedir. Ülkelerin büyüme oranlarındaki bir birimlik artışın ise patent başvuru sayısında istatistiksel olarak anlamlı ve %132'lik (2.32-1=1.32) bir artışa neden olduğu söylenebilir. Start-up sayısındaki bir birimlik artışın patent sayıları üzerinde istatistiksel olarak anlamlı ve %0.3'lük (1.003-1=-0.003) azalışa neden olduğu, makale sayısındaki artışın ise patent sayıları üzerinde nerdeyse etkisinin olmadığı (%0.00001) söylenebilir. Sabit etkiler panel poisson model sonuçlarından farklı olarak analiz sonucunda “lnalpha” ve “alpha” değerleri yer almaktadır. Analizin en sonunda yer alan “LR test of alpha=0” ifadesi ise bu değerlerin sıfıra eşit olup olmadığını test eden log olabilirlik oranı testidir. Bu test kurulan rassal etkiler panel poisson modeli ile örneklem ortalaması panel poisson modelinin karşılaştırılmasını sağlar. Örnekten de görüleceği üzere α parametresi istatistiksel olarak sıfırdan farklıdır. Bu durum rassal etkiler panel poisson modelinin havuzlanmış panel poisson modelinden istatistiksel olarak farklı olduğuna ve havuzlanmış modelin kullanılmaması gerektiğine işaret eder.

Eğer zaman boyutu çok uzun değilse, poisson ya da negatif binom gibi doğrusal olmayan modellerde sabit etkiler ve rassal etkiler tahmincileri için marjinal etkiler anlamlı bilgiler sunmaz. Bu nedenle rassal etkiler panel poisson modellerinde marjinal etkiler yerine esneklikler ve yarı esnekliklerin hesaplanması önerilir. Rassal etkiler panel poisson modeline ait esneklikler ile yarı esneklikler sırasıyla Tablo 15.13.a ve Tablo 15.13.b'de gösterilmektedir.

Tablo 15.13.a. Rassal Etkiler Panel Poisson: Esneklikler

```
qui xtpoisson patent arge makale startup gdp ,re
margins, eyex(*)
```

Average marginal effects				Number of obs	=	680	
Model VCE : OIM							
Expression : Linear prediction, predict()							
ey/ex w.r.t. : arge makale startup gdp							

		Delta-method					
		ey/ex	Std. Err.	z	P> z	[95% Conf. Interval]	

arge		.1727491	.0048522	35.60	0.000	.163239 .1822592	
makale		.0148808	.00047	31.66	0.000	.0139596 .0158021	
startup		.0030069	.0001227	24.50	0.000	.0027663 .0032475	
gdp		.0027707	.00009	30.79	0.000	.0025944 .0029471	

Rassal etkiler panel poisson modeline ait esnekliklerin hesaplandığı Tablo 15.13.a'da görüldüğü üzere, Ar-Ge harcamalarında %1'lik artış yaşandığında patent sayısı %0.17 oranında artmaktadır. Aynı mantıkla makale sayısında %1'lik artış yaşandığında patent sayısı %0.014 artmaktadır.

Tablo 15.13.b. Sabit Etkiler Panel Poisson: Yarı Esneklikler

```
qui xtpoisson patent arge makale startup gdp ,re
margins, eydx(*)
```

Average marginal effects				Number of obs	=	680	
Model VCE : OIM							
Expression : Linear prediction, predict()							
ey/dx w.r.t. : arge makale startup gdp							

		Delta-method					
		ey/dx	Std. Err.	z	P> z	[95% Conf. Interval]	

arge		1.250691	.0379067	32.99	0.000	1.176396	1.324987
makale		1.03e-06	3.26e-08	31.65	0.000	9.68e-07	1.10e-06
startup		.0004115	.0000168	24.43	0.000	.0003785	.0004445
gdp		.088114	.002883	30.56	0.000	.0824634	.0937646

Tablo 15.13.b’de rassal etkiler panel poisson modeline ait yarı esneklikler yer almaktadır. Tahmin edilen yarı esneklik katsayıları Ar-Ge harcamalarında her birim artış yaşandığında, patent sayısı %125 artmaktadır. Aynı şekilde start-up şirket sayısı bir adet arttığında patent sayısı %0.004 artarken, GSMH’de her birim artış yaşandığında patent sayısı %8 artmaktadır şeklinde yorumlanır.

Rassal etkiler panel poisson modelinde tahmin edilen katsayıların sıfırdan farklı olup olmadığı default olarak Wald testi yardımıyla yapılır. $6.3e+07$ olarak hesaplanan test istatistiğinin olasılık değeri $0.0000 < 0.05$ olduğu için modelde kullanılan katsayıların topluca anlamlı olduğu sonucuna ulaşılır. Rassal etkiler panel poisson modellerinde Wald testine alternatif olarak olabilirlik oranı (LR) testide yapılabilir. Bu durumda Tablo 15.14’de gösterilen `lrmodel` opsiyonu kullanılır. Söz konusu komut ve tahmin sonuçları aşağıda yer almaktadır. Görüleceği üzere LR test sonuçları da Wald test sonuçları ile paralellik göstermektedir.

Tablo 15.14. Panel Poisson Rassal Etkiler LR Model Tahmin Sonuçları

```
xtpoisson patent arge makale startup gdp, re nolog irr lrmodel
```

Kullanılan Opsiyonlar

re : rassal etkiler

nolog : iterasyonları göstermemesi için

irr : incidence-rate ratio hesaplanması için

lrmodel: default Wald testi yerine olabilirlik oranı (LR) model testi için

Random-effects Poisson regression		Number of obs		=	680
Group variable: id		Number of groups		=	38
Random effects u_i ~ Gamma		Obs per group:			
		min	=		15
		avg	=		17.9
		max	=		18
		LR chi2(4)	=		3.55e+06
Log likelihood = -1960734.4		Prob > chi2	=		0.0000

.....

(model tahmin sonuçları dışlandı)

Tablo 15.14’de yer alan temel model standart hatada herhangi bir düzeltmenin yapılmadığı temel durumdur. Hata terimi u_{it} ’de seri korelasyon veya panel üzerinde dağılımların eş yayılıma sahip olmaması durumunda panel değişken üzerinden kümeleme yapmak tutarlı varyans kovaryans matrisi tahmin edicilerin elde edilmesine olanak sağlar. Gelecek bölümde detaylı olarak ele alınacağı üzere veri seti aşırı yayılım gösterdiği için dirençli standart hataların üretildiği modeller daha tutarlı sonuçlar verir. Rassal etkiler panel poisson modelinde standart hataların düzeltilmesi için `vce(robust)`, `vce(bootstrap)` veya `vce(jackknife)` opsiyonları yer almaktadır (StataCorp, 2019). Üç farklı dirençli standart hataların oluşturulduğu model tahmin sonuçlarının karşılaştırılması için aşağıdaki komutlar kullanılmış olup elde edilen sonuçlar Tablo 15.15’te yer almaktadır.

```
quietly xtpoisson patent arge makale startup gdp, re
estimates store temel_re
quietly xtpoisson patent arge makale startup gdp, re vce(robust)
estimates store robust_re
quietly xtpoisson patent arge makale startup gdp, re vce(boot)
estimates store bootstrap_re
quietly xtpoisson patent arge makale startup gdp, re vce(jack)
estimates store jackknife_re
estimates table temel_re robust_re bootstrap_re jackknife_re, b se t p
```

Tablo 15.15. Rassal Etkiler Panel Poisson Modeli İçin Dirençli Tahminciler

Variable	temel_re	robust_re	bootstrap_re	jackknife_re
patent				
arge	11.955667	11.955667	11.955667	11.955667
	.00705418	8.273494	8.6881925	13.788393
	1694.83	1.45	1.26	0.87
	0.0000	0.1484	0.2060	0.3915
makale	9.868e-06	9.868e-06	9.868e-06	9.868e-06
	8.896e-08	4.949e-06	.00001104	.00001027
	110.92	1.99	0.89	0.96
	0.0000	0.0462	0.3714	0.3428
startup	.00393348	.00393348	.00393348	.00393348
	.0001084	.00730348	.00999654	.01641547
	36.29	0.54	0.39	0.24
	0.0000	0.5902	0.6940	0.8119
gdp	.84230398	.84230398	.84230398	.84230398
	.01038979	.59111968	.72835472	1.2647443
	81.07	1.42	1.16	0.67
	0.0000	0.1542	0.2475	0.5096
_cons	7.7104149	7.7104149	7.7104149	7.7104149
	.28600413	1.1635464	1.5641837	2.2806013
	26.96	6.63	4.93	3.38
	0.0000	0.0000	0.0000	0.0017
/lnalpha	1.1340481	1.1340481	1.1340481	1.1340481
	.18216466	3.3913071	.15821844	.13684549
	6.23	0.33	7.17	8.29

		0.0000	0.7381	0.0000	0.0000
Wald		3.02e+06	1345.18	2.85	2.33
p-value		0.0000	0.0000	0.5840	0.0740

 legend: tahmin katsayısı/standart hata/ z değeri /p değeri

Çalışmada kullanılan veri setinde aşırı yayılım durumu söz konusu olduğu için (aşırı yayılımın tespiti Tablo 15.17'de açıklanmıştır) dirençli standart hataların üretildiği modellerin seçilmesi gerekmektedir. Tablo 15.15'te dirençli standart hataların üretildiği rassal etkiler panel poisson model sonuçları yer almaktadır. Burada dirençli standart hataların üretildiği modellerde katsayıların istatistiksel olarak 0.05 seviyesinde anlamlı olmadıkları görülmektedir. Söz konusu durum, aşırı yayılımın dikkate alınmaması sahte regresyona neden olmaktadır (Hilbe, 2014; Cameron ve Trivedi 2013). Dirençli standart hataların üretildiği modellerden model uyum iyiliği açısından hangisinin seçileceğine ise istatistiksel olarak anlamlı bulunan Wald test istatistiği yardımcı olabilir. Bu durumda Tablo 15.15'te ikinci sütunda yer alan ve 0.05 istatistiksel anlamlılık seviyesinde katsayıların toplu anlamlı olduğu görülen dirençli model seçilebilir.

15.1.4. Model Seçimi

Bu başlık altında panel poisson modellerinde olası üç modelleme yöntemi olarak kullanılan örneklem ortalaması, sabit ve rassal etkiler modellerinin seçiminde dikkat edilmesi gereken hususlar üzerinde durulacaktır (Winkelmann, 2008:222):

1. Örneklem ortalaması panel poisson modelinde birim etkileri temsil eden α_i parametresinin birim etkilerin ortalamasına eşit olduğu ($\alpha = E_{\alpha_i}[\alpha_i]$) varsayılır.
2. Sabit etkiler panel poisson modelinde birim etkileri temsil eden α_i parametresinin X_{it} açıklayıcı değişkenleri ile potansiyel olarak ilişkili olduğu varsayılır.
3. Rassal panel poisson modelinde birim etkileri temsil eden α_i parametresinin birim etkilerin X_{it} açıklayıcı değişkenleri ile ilişkili olmadığı varsayılmaktadır.

Eğer birim etkileri temsil eden α_i parametresinin X_{it} açıklayıcı değişkenleri ile ilişkili ise havuzlanmış ve rassal etkiler modelleri tutarsız tahminler üretir. Sabit etkiler modelini sadece zaman içinde değişen etkilerin etkisi analiz edileceği zaman kullanmak gerekir. Sabit etkiler modelinde açıklanan değişken ile açıklayıcı değişkenler arasındaki ilişki gözlemler için sahip olunan belirli bir özellik üzerinden yapılacaksa tercih edilmelidir. Sabit etkiler modeli bireyler (birimler) arasındaki zamandan bağımsız etkileri kontrol altında tutar ve böylece tahmin edilen katsayılar zamandan bağımsız özelliklerin (cinsiyet, etnik kimlik, kültür gibi) dışlanması nedeniyle sapmasız olur. Teknik olarak söz konusu zamandan bağımsız özellikler birey ile tam ilişkilidir. Sabit etkiler modeli bireyde meydana gelen değişimlerin etkilerini ölçer, bu durumda zamandan bağımsız özellikler, birey için sabit olduğundan, bu tür etkilere neden olmaz. Söz konusu etkiler sabit etkiler tarafından karşılanır ve böylece açıklayıcı değişkenlerin açıklanan değişken üzerindeki net etkilerine ulaşılır.

Sabit etkiler ve rassal etkiler modellerindeki temel fark rassal etkiler panel poisson modelinde birim etkileri temsil eden α_i parametresinin sabit etkiler panel poisson modelinin aksine X_{it} açıklayıcı değişkenleri ile ilişkili olmadığının varsayılmasıdır. Hâliyle sabit etkiler ile rassal etkiler

arasındaki temel ayrım “gözlemlenemeyen bireysel etkilerin açıklayıcı değişkenler ile ilişkili olup olmadığıdır” (Green, 2008:183). Bu nedenle eğer bireysel özellikler arasındaki farkın bağımlı değişken üzerinde bir etkiye sahip olduğuna dair bir beklenti var ise rassal etkiler modelinin kullanılması gerekir. Rassal etkilerin bir avantajı da cinsiyet, etnik kimlik, kültür gibi zamandan bağımsız özelliklerin modele eklenebilmesidir.

Bu bölümde kullanılan Stata komutları aşağıda toplu olarak yer almaktadır.

```
use "DOSYAKONUMU\patent.dta", clear
xtset id time
xtpoisson patent arge makale startup gdp ,re nolog
xtpoisson patent arge makale startup gdp ,re nolog irr
xtpoisson patent arge makale startup gdp, re nolog irr lrmodel
qui xtpoisson patent arge makale startup gdp ,re
margins, eyex(*)
margins, eydx(*)
quietly xtpoisson patent arge makale startup gdp, re
estimates store temel_re
quietly xtpoisson patent arge makale startup gdp, re vce(robust)
estimates store robust_re
quietly xtpoisson patent arge makale startup gdp, re vce(boot)
estimates store bootstrap_re
quietly xtpoisson patent arge makale startup gdp, re vce(jack)
estimates store jackknife_re
estimates table temel_re robust_fe bootstrap_fe jackknife_fe, b se t p
```

15.2. Panel Negatif Binom Modelleri

Panel negatif binom (NegBin) modelleri poisson modellerinin genelleştirilmiş hâlidir. NegBin modelleri basitçe Poisson modellerine birim ve gözlemlenemeyen etkilerin koşullu ortalamaya dâhil edilmiş hâlidir (Long ve Freese, 2014). Denklem (15.1.a) ve (15.1.b)’de gösterilen doğrusal panel model denkleminde hareketle Denklem (15.29) varsayımı yapılır.

$$\mu_{it} = \phi_i \lambda_{it} = \exp(x'_{it}\beta + \phi_i) \quad (15.29)$$

Burada ϕ_i ’nin genellikle ortalaması 1, varyansı $1/\phi_i$ olan gamma dağılıma sahip bir parametreye sahip olduğu varsayılır (Greene, 2007:2). Koşullu ortalama $E(Y_i|X_i) = \exp(x'_i\beta)$ iken koşullu varyans $\text{Var}(Y_i|X_i)$ fonksiyonuna bağlı olarak farklı negatif binom modelleri kurulabilir. NegBin1 veya lineer negatif binom olarak bilenen modelde ortalama ile varyans arasındaki ilişki Denklem (15.30)’da NegBin2 veya kuadratik negatif binom olarak bilenen modelde ise ortalama ile varyans arasındaki ilişki Denklem (15.31)’de gösterilmektedir ve burada $\alpha = \frac{1}{\phi}$ ’dır.

$$\text{Var}(Y_i|X_i) = \mu_i(1 + \alpha) > E(Y_i|X_i) \quad (15.30)$$

$$\text{Var}(Y_i|X_i) = \mu_i(1 + \alpha\mu_i) > E(Y_i|X_i) \quad (15.31)$$

Görüldüğü üzere NegBin modellerinde koşullu varyansın koşullu ortalamadan büyük olduğu varsayılmaktadır. Poisson modeli $\alpha = 0$ özel durumuna karşılık gelmektedir. Bu nedenle ampi-

rik uygulamada Wald testi yardımıyla bu durum test edilebilir. Poisson modellerinde eş yayılım varsayımı yapıldığı için, aşırı yayılım durumunda poisson modellerinin kullanılması standart hataların ciddi derecede düşük tahmin edilmesine ve regresyon parametrelerinden yanıltıcı çıkarımların yapılmasına neden olabilir (Hinde ve Demetrio, 1998). Varyans fonksiyonuna bağlı olarak oluşturulan NegBin1 ve NegBin2 modellerinde katsayıların tahmini için en çok olabilirlik algoritması kullanılır.

NegBin1 modeline ait en çok olabilirlik fonksiyonu Denklem (15.32)'de yer almaktadır (Winkelmann, 2008):

$$\ell(\beta, \sigma^2) = \sum_i \left[\left(\sum_j \log(\sigma^{-2} \exp(x'_i \beta) + j - 1) - \log y_i! - (y_i + \sigma^{-2} \exp(x'_i \beta) \log(1 + \sigma^2)) + y_i \log \sigma^2 \right) \right] \quad (15.32)$$

$\hat{\beta}$ için birinci sıra koşul aşağıda gösterilmektedir.

$$\sum_i \left[\left(\sum_j \frac{\sigma^{-2} \exp(x'_i \beta)}{\sigma^{-2} \exp(x'_i \beta) + j - 1} \right) x_i + \sigma^{-2} + \exp(x'_i \beta) x_i \right] \quad (15.33)$$

NegBin2 modeline ait en çok olabilirlik fonksiyonu Denklem (15.34)'de yer almaktadır (Winkelmann, 2008):

$$\ell(\beta, \sigma^2) = \sum_i \left[\left(\sum_j \log(\sigma^{-2} + j - 1) - \log y_i! - (y_i + \sigma^{-2}) \log(1 + \sigma^2 \exp(x'_i \beta)) + y_i \log \sigma^2 \right) + y_i x'_i \beta \right] \quad (15.34)$$

NegBin2 en çok olabilirlik tahmin edicileri $\hat{\beta}$ ve $\hat{\sigma}^2$ Denklem (15.35) ve Denklem (15.36)'da yer alan birinci sıra koşulların sonuçlarından elde edilir.

$$\sum_i \frac{y_i - \exp(x'_i \beta)}{1 + \sigma^2 \exp(x'_i \beta)} x_i = 0 \quad (15.35)$$

$$\sum_i \left[\frac{1}{\sigma^4} \left(\log(1 + \sigma^2 \exp(x'_i \beta)) - \sum_j \frac{1}{\sigma^{-2} + j - 1} \right) - \frac{(y_i + \sigma^{-2}) \exp(x'_i \beta)}{1 + \sigma^{-2} \exp(x'_i \beta)} + \frac{y_i}{\sigma^2} \right] = 0 \quad (15.36)$$

15.2.1.Örneklem Ortalaması Panel Negatif Binom Modeli

Örneklem ortalaması panel NegBin tahminleri örneklem panel ortalaması poisson tahminlerinin elde edilmesi ile aynı yaklaşımı kullanır. Daha önceden belirtilgi üzere veri X_{it} durumunda Y_{it} 'nin koşullu varyansı $Var(Y_{it}|X_{it}) = v(\mu_{it})\phi$ şeklindedir ve burada v varyans fonksiyonu, ϕ tahmin edilmesi gereken parametreleri temsil etmektedir. Poisson modelinden farklı olarak v fonksiyonunun yapısı NegBin1 veya NegBin2 olabilirken ve ϕ gamma dağılıma sahiptir.

15.2.1.1. Örneklem Ortalaması Panel Negatif Binom Model Stata Uygulaması

Panel NegBin tahminleri için Stata komutu olarak `xtnbreg` kullanılmaktadır. Komutun temel yapısı aşağıda belirtilmektedir.

`xtnbreg depvar [indepvars] [if] [in] [weight] , pa [PA_options]`

Komut yapısı hakkında detaylı bilgi için **help xtnbreg** komutu yardımıyla açılan bilgilendirme ekranından detaylara ulaşabilirsiniz. Örneklem ortalamalı tahminlerin üretilebilmesi için Stata komutunda “,” den sonra opsiyon için “**pa**” (population average) yazılması gerekmektedir. Aksi takdirde **xtnbreg** komutu default olarak rassal etkiler tahminleri üretir.

Örneklem ortalaması panel NegBin tahminleri daha genel kullanıma sahip GEE komutu olan **xtgee** yardımıyla da elde edilebilmektedir. Söz konusu komutun örneklem ortalaması panel NegBin için genel yapısı aşağıdaki gibidir:

```
xtgee depvar [indepvars] [in] [weight] [, family(nbinomial) link(log)
corr(exchangeable) options]
```

Örneklem ortalaması panel NegBin tahminlerinde **xtnbreg** default olarak $\alpha = 1$ varsayımı üzerinden tahminleme yapmakta olup sınırlı opsiyona sahiptir. Ancak **xtgee** komutu daha geniş opsiyona sahip olup farklı varsayımlara bağlı örneklem ortalaması tahminlerinin elde edilmesine olanak sağlamaktadır.

İlk olarak ülkelerin patent sayısı ile Ar-Ge harcamalarının GSMH içindeki payı, yayınlanan akademik makale sayısı, kurulan yeni start-up işletme sayısı arasındaki ilişkinin örneklem ortalaması panel NegBin tahminlerini Tablo 15.12’de yer almaktadır.

Tablo 15.16. Örneklem Ortalaması Panel NegBin Model Tahmin Sonuçları

```
xtnbreg patent arge makale startup gdp ,pa nolog
```

Kullanılan opsiyonlar:

Pa : örneklem ortalaması

Nolog : iterasyonları göstermemesi için

```
GEE population-averaged model
Group variable:          id      Number of obs      =      680
Link:                   log      Number of groups   =      38
Family:                negative binomial(k=1)      min =      15
Correlation:           exchangeable      avg =      17.9
                                      max =      18
                                      Wald chi2(4)    =      5097.41
Scale parameter:        1      Prob > chi2      =      0.0000
```

patent	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
arge	27.49981	.3875239	70.96	0.000	26.74028	28.25934
makale	-4.11e-06	2.24e-06	-1.83	0.067	-8.51e-06	2.90e-07
startup	.0076824	.0049346	1.56	0.120	-.0019892	.0173541
gdp	.3739644	.2831806	1.32	0.187	-.1810593	.9289882
_cons	5.268261	.1756345	30.00	0.000	4.924024	5.612498

Tablo 15.16'da yer alan sonuçlar aynı şekilde aşağıda yer alan **xtgee** komutu yardımıyla da elde edilir.

xtgee patent arge makale startup gdp, family(nbinomial) Link(log) corr(exchangeable)

Denklem (15.29) ve Denklem (15.30)'da yer alan α parametresine bağlı olarak örneklem ortalaması panel NegBin modeli tahminleri elde edilebilir. Bunun için ilk olarak söz konusu modele dair α parametresinin tahmin edilmesi gerekmektedir. Bunun için **nberg** komutu yardımıyla söz konusu α parametresi kolayca tahmin edilebilir.

Tablo 15.17. Yayılım Parametresinin (α) Tahmin Edilmesi

nbreg patent arge makale startup gdp , nolog

Kullanılan opsiyonlar:

Nolog : iterasyonları göstermemesi için

Negative binomial regression				Number of obs	=	680
				LR chi2(4)	=	510.30
Dispersion = mean				Prob > chi2	=	0.0000
Log likelihood = -6632.4487				Pseudo R2	=	0.0370

patent		Coef.	Std. Err.	z	P> z	[95% Conf. Interval]

arge		19.63717	.8579256	22.89	0.000	17.95566 21.31867
makale		-.0000234	2.69e-06	-8.73	0.000	-.0000287 -.0000182
startup		.025467	.0228872	1.11	0.266	-.019391 .070325
gdp		-.9378977	2.305233	-0.41	0.684	-5.456071 3.580276
_cons		6.425644	.2084065	30.83	0.000	6.017175 6.834113

/lnalpha		.9178567	.043876			.8318613 1.003852

alpha		2.503918	.109862			2.297591 2.728773

LR test of alpha=0: chibar2(01) = 6.7e+07				Prob >= chibar2 = 0.000		

Hilbe (2011) pratikte α parametresinin 0 ila 4 arasında bir değer alması gerektiğini ileri sürmektedir. Söz konusu yayılım parametresi sıfıra eşitse Negbin modeli Poisson modeline dönüşür. Tablo 15.17'de yer alan alpha katsayısı (≈ 2.504) söz konusu yayılım parametresi olan α 'yı temsil etmektedir. LR testi alpha'nın istatistiksel olarak anlamlı ve sıfırdan farklı olduğunu göstermektedir. Tahmin edilen alpha katsayısına bağlı olarak örneklem ortalamalı NegBin modeli aşağıdaki komut yardımıyla elde edilir. Burada **family(nbinomial 2.504)** opsiyonu kullanılır. Tablo 15.18'de Family: negative binomial($k=.3994$) olarak yer alan ifade $\left(k = \frac{1}{\alpha} = \frac{1}{2.504} \approx 0.3994\right)$ elde edilmektedir.

Tablo 15.18. Yayılım Parametresine ($\alpha = 2.504$) Bağlı Olarak Örneklem Ortalaması Panel NegBin Model Tahmin Sonuçları

```
xtgee patent arge makale startup gdp, family(nbinomial 2.503918) link(log)
corr(exchangeable) nolog
```

```
GEE population-averaged model      Number of obs   =      680
Group variable:                    id      Number of groups =      38
Link:                              log      Obs per group:
Family:      negative binomial(k=.3994)    min =      15
Correlation:      exchangeable              avg =     17.9
                                              max =      18
                                              Wald chi2(4)    =    2036.48
Scale parameter:                    1      Prob > chi2      =     0.0000
```

patent	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
arge	27.49999	.6131082	44.85	0.000	26.29832	28.70166
makale	-4.12e-06	3.55e-06	-1.16	0.246	-.0000111	2.84e-06
startup	.0076712	.0078074	0.98	0.326	-.0076311	.0229735
gdp	.3740384	.4480268	0.83	0.404	-.5040781	1.252155
_cons	5.2681	.2778358	18.96	0.000	4.723552	5.812648

Daha önceden de belirtildiği üzere saymalı model tahminleri neticesinde elde edilen katsayılar doğrudan yorumlanmaya uygun olmadığı için incidence-rate ratio (irr)'nin hesaplanması gerekmektedir. Bunun için tahmin edilen katsayıların exponansiyeli (e^{β}) manuel olarak hesaplanabileceği gibi `xtnbreg` için `irr`, `xtgee` için `eform` opsiyonu kullanılarak söz konusu hesaplanmış katsayılar elde edilebilir. Ancak dikkat edilirse önceden de belirtildiği üzere temel modeldeki katsayılarla ait standart hatalar ile IRR tahminlerine ait standart hatalar kullanılan hesaplama yöntemi nedeniyle farklıdır.

Tablo 15.19. Örneklem Ortalaması Panel NegBin Model Tahmin IRR Sonuçları

```
xtnbreg patent arge makale startup gdp ,pa nolog irr
```

Kullanılan opsiyonlar:

Pa : örneklem ortalaması

Nolog : iterasyonları göstermemesi için

irr : incidence-rate ratio hesaplanması için

```
GEE population-averaged model      Number of obs   =      680
Group variable:                    id      Number of groups =      38
Link:                              log      Obs per group:
Family:      negative binomial(k=1)    min =      15
Correlation:      exchangeable              avg =     17.9
                                              max =      18
                                              Wald chi2(4)    =    5097.41
Scale parameter:                    1      Prob > chi2      =     0.0000
```

patent	IRR	Std. Err.	z	P> z	[95% Conf. Interval]	

arge		8.77e+11	3.40e+11	70.96	0.000	4.10e+11	1.87e+12
makale		.9999959	2.24e-06	-1.83	0.067	.9999915	1
startup		1.007712	.0049727	1.56	0.120	.9980127	1.017506
gdp		1.453485	.4115988	1.32	0.187	.8343859	2.531946
_cons		194.0782	34.08683	30.00	0.000	137.555	273.8275

Note: _cons estimates baseline incidence rate (conditional on zero random effects).

Tablo 15.20. Yayılım Parametresine ($\alpha = 2.504$) Bağlı Olarak Örneklem Ortalaması Panel NegBin Model Tahmin IRR Sonuçları

```
xtgee patent arge makale startup gdp, family(nbinomial 2.504) link(log)
corr(exchangeable) nolog eform
```

```
GEE population-averaged model
Group variable:          id      Number of obs      =      680
Link:                   log      Number of groups   =      38
Family:                 negative binomial(k=.3994)      min =      15
Correlation:            exchangeable                  avg =      17.9
                                                    max =      18
                                                    Wald chi2(4)   =      2036.41
Scale parameter:        1        Prob > chi2       =      0.0000
```

patent		IRR	Std. Err.	z	P> z	[95% Conf. Interval]
arge		8.77e+11	5.38e+11	44.85	0.000	2.64e+11 2.92e+12
makale		.9999959	3.55e-06	-1.16	0.246	.9999889 1.000003
startup		1.007701	.0078677	0.98	0.326	.9923977 1.02324
gdp		1.453593	.6512593	0.83	0.404	.6040536 3.497922
_cons		194.047	53.91407	18.96	0.000	112.5664 334.5068

Note: _cons estimates baseline incidence rate (conditional on zero random effects).

Tablo 15.19.a. Örneklem Ortalaması Panel NegBin: Ortalamaya Göre Marjinal Etkiler

```
qui xtnbreg patent arge makale startup gdp ,pa nolog
margins, dydx(*) atmean noatlegend
```

```
Conditional marginal effects      Number of obs      =      680
Model VCE      : Conventional
Expression      : Exponentiated linear prediction, predict()
dy/dx w.r.t.    : arge makale startup gdp
```

		Delta-method		z	P> z	[95% Conf. Interval]
		dy/dx	Std. Err.			
arge		326810.4	51710.19	6.32	0.000	225460.2 428160.5
makale		-.0488209	.027744	-1.76	0.078	-.1031982 .0055564
startup		91.29849	60.39337	1.51	0.131	-27.07034 209.6673
gdp		4444.229	3437.311	1.29	0.196	-2292.776 11181.23

Ortalamaya göre marjinal etkilerin sunulduğu Tablo 15.19.a'da Ar-Ge harcamalarının marjinal etkisi: Ar-Ge dışındaki diğer değişkenler örneklem ortalaması değerlerinde iken Ar-Ge harcamalarının GSMH içindeki payı bir birim attığında patent sayısı %326810 artacağını, makale dışındaki diğer değişkenler örneklem ortalaması değerlerinde iken makale sayısındaki bir adet artışın ise patent sayısında %4.8'lik bir azalışa neden olacağını göstermektedir.

Tablo 15.19.b. Örneklem Ortalaması Panel NegBin: Ortalama Marjinal Etkiler

```
qui xtnbreg patent arge makale startup gdp ,pa nolog
margins, dydx(*)
```

Average marginal effects				Number of obs	=	680	
Model VCE : Conventional							
Expression : Exponentiated linear prediction, predict()							
dy/dx w.r.t. : arge makale startup gdp							

	Delta-method						
	dy/dx	Std. Err.	z	P> z	[95% Conf. Interval]		

arge	2.62e+07	5112357	5.12	0.000	1.62e+07	3.62e+07	
makale	-3.911903	2.292233	-1.71	0.088	-8.404596	.58079	
startup	7315.536	4938.825	1.48	0.139	-2364.383	16995.46	
gdp	356105.8	272833.5	1.31	0.192	-178638	890849.5	

Ortalama marjinal etkilerin hesaplandığı Tablo 15.19.b'ye göre Ar-Ge'nin ortalama marjinal etkisi Ar-Ge'de meydana gelen her bir birimlik artışta patent sayısının ortalama 356105 adet, ilave bir adet start-up firma kurulduğunda patent sayısının ortalama 7315 adet artacağını göstermektedir.

Tablo 15.4.c. Örneklem Ortalaması Panel NegBin: Esneklikler

```
qui xtnbreg patent arge makale startup gdp ,pa nolog
margins, eyex(*)
```

Average marginal effects				Number of obs	=	680
Model VCE : Conventional						
Expression : Exponentiated linear prediction, predict()						
ey/ex w.r.t. : arge makale startup gdp						

		Delta-method				
		ey/ex	Std. Err.	z	P> z	[95% Conf. Interval]

arge		4.105277	.0578511	70.96	0.000	3.991891 4.218663
makale		-.05887	.0321563	-1.83	0.067	-.1218952 .0041553
startup		.0564544	.0362621	1.56	0.120	-.014618 .1275268
gdp		.0118338	.008961	1.32	0.187	-.0057295 .029397

Esnekliklerin hesaplandığı Tablo 15.4.c'ye göre Ar-Ge harcamalarında %1'lik artış yaşandığında patent sayısı %4.10, makale sayısında %1'lik artış yaşandığında ise patent sayısı %0.05 azalmaktadır.

Tablo 15.4.d. Örneklem Ortalaması Panel Negbin: Yarı Esneklikler

```
qui xtnbreg patent arge makale startup gdp ,pa nolog
margins, eydx(*)
```

Average marginal effects				Number of obs	=	680
Model VCE : Conventional						
Expression : Exponentiated linear prediction, predict()						
ey/dx w.r.t. : arge makale startup gdp						

		Delta-method				
		ey/dx	Std. Err.	z	P> z	[95% Conf. Interval]

arge		27.49981	.3875239	70.96	0.000	26.74028 28.25934
makale		-4.11e-06	2.24e-06	-1.83	0.067	-8.51e-06 2.90e-07
startup		.0076824	.0049346	1.56	0.120	-.0019892 .0173541
gdp		.3739644	.2831806	1.32	0.187	-.1810593 .9289882

Dikkat edilirse Tablo 15.4.d'den elde edilen sonuçlar ile Tablo 15.3'ten elde edilen sonuçlar aynıdır. Bunun nedeni daha önceden belirtildiği üzere yarı esneklikler $\mu = \exp(x_i'\beta)$ olacak şekilde hesaplanmaktadır. Bu durumda elde edilen yarı esneklik katsayıları, örneklem ortalaması poisson ve örneklem ortalaması negatif binom modellerindeki tahmin katsayıları ile aynıdır.

Tablo 15.19'da ve Tablo 15.20'de yer alan örneklem ortalaması panel NegBin IRR Tahmin sonuçlarına bakıldığında tahmin edilen katsayılar aynı iken elde edilen standart hatalar ile istatistiksel anlamlılıklar farklıdır. Ancak her iki modelde de Ar-Ge harcamalarının GSMH oranındaki bir birimlik artışın patent sayıları üzerinde istatistiksel olarak anlamlı çok ciddi bir artışa neden olduğu görülmektedir. Ülkelerin büyüme oranları, makale sayısı ve startup firma sayılarının patent sayıları üzerindeki etkisi %5 seviyesinde her iki model içinde istatistiksel olarak anlamlı değildir.

Tablo 15.20'de yer alan örneklem ortalaması panel NegBin modelinde tahmin edilen katsayıların sıfırdan farklı olup olmadığı Wald testi yardımıyla yapılır. Wald testi için hesaplanan p olasılık değeri $0.0000 < 0.05$ olduğu için modelde kullanılan katsayıların topluca anlamlı olduğu sonucuna ulaşılır. Tablo 15.19 ve Tablo 15.20'de yer alan Wald test istatistiğine bakıldığında modelde tahmin edilen katsayıların sıfırdan olduğu istatistiksel olarak desteklenmektedir.

Kısa panel verilerde standart olarak birim üzerinden kümeleme (cluster) yapılarak kümelenmiş dirençli (cluster-robust) standart hatalar veya panel dirençli standart hatalar kullanılır (Cameron ve Trivedi, 2015). Bu nedenle kısa panel verilerde standart olarak daha tutarlı tahminlerin elde edilmesi için birim üzerinden kümeleme yapılarak kümelenmiş dirençli standart hataların hesaplanması gerekir. Örneklem ortalaması panel NegBin tahmini yapılırken iki farklı varyans tahmin edicisi opsiyonu olan **vce(robust)** veya **vce(bootstrap)** kullanılarak birim üzerinden kümelenmiş

dirençli standart hatalar elde edilir. Söz konusu standart hata opsiyonlarına dair komutlar ve karşılaştırma tablosu aşağıda yer almaktadır.

```
qui xtnbreg patent arge makale startup gdp, pa nolog irr
estimates store nb_pa_temel
qui xtnbreg patent arge makale startup gdp, pa nolog irr vce(robust)
estimates store nb_pa_robust
qui xtnbreg patent arge makale startup gdp, pa nolog irr vce(boot)
estimates store nb_pa_bootstrap
estimates table nb_pa_temel nb_pa_robust nb_pa_bootstrap , b se t p
```

Tablo 15.21. Örneklem Ortalaması Panel NegBin Modele Ait Dirençli Tahminciler

Variable	nb_pa_temel	nb_pa_robust	nb_pa_bootstrap
arge	27.49981	27.49981	27.49981
	.3875239	3.0211036	10.318974
	70.96	9.10	2.66
	0.0000	0.0000	0.0077
makale	-4.108e-06	-4.108e-06	-4.108e-06
	2.244e-06	1.841e-06	6.149e-06
	-1.83	-2.23	-0.67
	0.0671	0.0257	0.5040
startup	.00768241	.00768241	.00768241
	.00493461	.01185886	.01098325
	1.56	0.65	0.70
	0.1195	0.5171	0.4843
gdp	.37396492	.37396492	.37396492
	.28318059	.37241527	.60269769
	1.32	1.00	0.62
	0.1866	0.3153	0.5349
_cons	5.2682611	5.2682611	5.2682611
	.17563453	.34933826	1.7211702
	30.00	15.08	3.06
	0.0000	0.0000	0.0022

legend: tahmin katsayısı/standart hata/ z değeri /p değeri

Örneklem büyüklüğünün az olduğu durumlarda grup içi korelasyon matrisinin doğru tanımlanması tahmin edilen parametrelerin etkinliğini kesin biçimde arttırmaktadır (Wang ve Carey 2003). Bu nedenle örneklem ortalaması panel NegBin tahmini yapılırken, farklı korelasyon yapılarının modele eklenmesine izin veren ve daha önceden açıklanan `corr()` opsiyonu kullanılır.

```
quietly xtnbreg patent arge makale startup gdp, pa corr(ind)
estimates store nb_pa_ind
quietly xtnbreg patent arge makale startup gdp, pa corr(exc)
estimates store nb_pa_exc
quietly xtnbreg patent arge makale startup gdp, pa corr(ar1)
estimates store nb_pa_ar1
quietly xtnbreg patent arge makale startup gdp, pa corr(ar2)
estimates store nb_pa_ar2
estimates table nb_pa_ind nb_pa_exc nb_pa_ar1 nb_pa_ar2, b se t p
```

Tablo 15.22. Örneklem Ortalaması Panel NegBin Modele Ait Farklı Corr() Yapılarına Ait Tahminler

Variable	nb_pa_ind	nb_pa_exc	nb_pa_ar1	nb_pa_ar2
arge	19.63612	27.49981	18.28651	17.345152
	.3900388	.38752389	.87831022	.89272174
	50.34	70.96	20.82	19.43
	0.0000	0.0000	0.0000	0.0000
makale	-.00002345	-4.108e-06	-2.904e-06	-3.391e-06
	1.775e-06	2.244e-06	1.826e-06	1.784e-06
	-13.21	-1.83	-1.59	-1.90
	0.0000	0.0671	0.1117	0.0573
startup	.02544675	.00768241	.00538123	.00516863
	.011641	.00493461	.0039395	.00381578
	2.19	1.56	1.37	1.35
	0.0288	0.1195	0.1719	0.1756
gdp	-.93646623	.37396492	.23478022	.22533743
	1.0312555	.28318059	.23870375	.23276098
	-0.91	1.32	0.98	0.97
	0.3638	0.1866	0.3253	0.3330
_cons	6.4259786	5.2682611	6.5389191	6.7053469
	.11553759	.17563453	.20310367	.19972759
	55.62	30.00	32.19	33.57
	0.0000	0.0000	0.0000	0.0000

legend: tahmin katsayısı/standart hata/ z değeri /p değeri

Örneklem ortalaması panel NegBin tahminlerinde genelleştirilmiş tahmin denklemleri kullanıldığı ve benzerlik fonksiyonu tanımlanmadığı için Akaike Bilgi Kriteri (AIC) kullanılamamaktadır. Bu nedenle Quasi-likelihood Bilgi Kriteri (QIC) yardımıyla hem uygun modelin, hem de uygun korelasyon yapısının seçimine karar verilebilir. QIC değerlerinin hesaplanması GEE üzerinden yapılmaktadır. Cui ve Feng (2008) QIC'nin herhangi veri yayılım parametresine bağlı olarak elde edilmesine olanak tanıyacak şekilde geliştirmişlerdir. Öncelikle yayılım parametresini dikkate alan ($\alpha = 2.504$) model ile dikkate almayan default ($\alpha = 1$) modelin model uyum iyilikleri Tablo 15.23'te karşılaştırılmıştır. Bunun için daha önceden açıklandığı üzere `qic` komutu kullanılmıştır.

```
qic patent arge makale startup gdp, i(id) t(time) fam(nb) link(log) corr(exc)
```

```
qic patent arge makale startup gdp, i(id) t(time) fam(nb 2.504) link(log) corr(exc)
```


Tablo 15.23. Yayılım Parametresine Bağlı NegBin Modeli QIC Sonuçları

Yayılım Parametresi ($\alpha = 1$)		Yayılım Parametresi ($\alpha=2.504$)	
Corr =	exc	Corr =	exc
Family =	nb	Family =	nb 2.504
Link =	log	Link =	log
p =	5	p =	5
Trace =	168.281	Trace =	67.257
QIC =	14426.995	QIC =	6260.080
QIC_u =	14100.432	QIC_u =	6135.566

Tablodan görüleceği üzere Yayılım Parametresine ($\alpha = 2.504$) bağlı QIC sonuçların model uyum iyiliğinin daha yüksek olduğu söylenebilir. Farklı korelasyon yapılarının modele eklenmesine izin veren ve daha önceden açıklanan `corr()` opsiyonu kullanılarak Yayılım Parametresine ($\alpha = 2.504$) bağlı QIC sonuçları aşağıdaki komutlarla elde edilir.

```
qic patent arge makale startup gdp, i(id) t(time) fam(nb 2.504) link(log) corr(ind)
qic patent arge makale startup gdp, i(id) t(time) fam(nb 2.504) link(log) corr(exc)
qic patent arge makale startup gdp, i(id) t(time) fam(nb 2.504) link(log) corr(ar1)
qic patent arge makale startup gdp, i(id) t(time) fam(nb 2.504) link(log) corr(ar2)
```

Tablo 15.24. Farklı Korelasyon Yapılarına Bağlı Olarak NegBin Modeli ($\alpha = 2.504$) QIC Sonuçları

Corr(ind)	Corr(exc)	Corr(ar1)	Corr(ar2)
Corr = ind	Corr = exc	Corr = ar1	Corr = ar2
Family = nb 2.504	Family = nb 2.504	Family = nb 2.504	Family = nb 2.504
Link = log	Link = log	Link = log	Link = log
p = 5	p = 5	p = 5	p = 5
Trace = 74.022	Trace = 67.257	Trace = 92.377	Trace = 90.735
QIC = 6174.937	QIC = 6260.080	QIC = 6265.224	QIC = 6267.344
QIC_u = 6036.894	QIC_u = 6135.566	QIC_u = 6090.469	QIC_u = 6095.874

Birimler arası farklı korelasyon yapılarından en düşük QIC değeri alan modelin tercih edilmesi gerekir. Veri setimize en uygun korelasyon yapısının en düşük QIC değeri olan 6174.937 ile `corr(ind)` yani gözlemler arası korelasyonun 1'e eşit olduğu modelin olduğu söylenebilir. Eğer gözlemler arası korelasyon yapısı bağımsız ise örneklem ortalaması panel NegBin ile modeli aynı sonucu verir. Bir başka ifade ile `nbreg` komutu ile `xtnbreg`, `corr(independence)` eş tahminler üretir.

Bu bölümde kullanılan Stata komutları aşağıda toplu olarak yer almaktadır.

```
xtnbreg patent arge makale startup gdp ,pa nolog
xtgee patent arge makale startup gdp, family(nbinomial) link(log) corr(exchangeable)
nbreg patent arge makale startup gdp, nolog
xtgee patent arge makale startup gdp, family(nbinomial 2.503918) link(log)
corr(exchangeable) nolog
xtnbreg patent arge makale startup gdp ,pa nolog irr
xtgee patent arge makale startup gdp, family(nbinomial 2.504) link(log)
corr(exchangeable) nolog eform
```

```

qui xtnbreg patent arge makale startup gdp ,pa nolog
margins, dydx(*) atmean noatlegend
margins, dydx(*)
margins, eyex(*)
margins, eydx(*)
qui xtnbreg patent arge makale startup gdp, pa nolog irr
estimates store nb_pa_temel
qui xtnbreg patent arge makale startup gdp, pa nolog irr vce(robust)
estimates store nb_pa_robust
qui xtnbreg patent arge makale startup gdp, pa nolog irr vce(boot)
estimates store nb_pa_bootstrap
estimates table nb_pa_temel nb_pa_robust nb_pa_bootstrap , b se t p
quietly xtnbreg patent arge makale startup gdp, pa corr(ind)
estimates store nb_pa_ind
quietly xtnbreg patent arge makale startup gdp, pa corr(exc)
estimates store nb_pa_exc
quietly xtnbreg patent arge makale startup gdp, pa corr(ar1)
estimates store nb_pa_ar1
quietly xtnbreg patent arge makale startup gdp, pa corr(ar2)
estimates store nb_pa_ar2
estimates table nb_pa_ind nb_pa_exc nb_pa_ar1 nb_pa_ar2, b se t p
qic patent arge makale startup gdp, i(id) t(time) fam(nb) link(log) corr(exc)
qic patent arge makale startup gdp, i(id) t(time) fam(nb 2.504) link(log) corr(exc)
qic patent arge makale startup gdp, i(id) t(time) fam(nb 2.504) link(log) corr(ind)
qic patent arge makale startup gdp, i(id) t(time) fam(nb 2.504) link(log) corr(exc)
qic patent arge makale startup gdp, i(id) t(time) fam(nb 2.504) link(log) corr(ar1)
qic patent arge makale startup gdp, i(id) t(time) fam(nb 2.504) link(log) corr(ar2)

```

15.2.2. Sabit Etkiler Panel Negatif Binom Modeli

Sabit etkiler panel NegBin modelinde temel varsayım Denklem (15.37)'da gösterildiği gibi μ_{it} 'nin gamma dağılıma sahip olduğudur.

$$\mu_{it} \sim \Gamma(\lambda_{it}, \theta_i) \quad (15.37)$$

Bu varsayımdan hareketle y_{it} 'nin olasılık fonksiyonu aşağıdaki gibi olur ve burada θ_i bireysel etkileri temsil ederken, $\lambda_{it} = \exp(x'_{it}\beta)$ 'dır (Winkelmann, 2008: 227):

$$f(y_{it}) = \frac{\Gamma(y_{it} + \lambda_{it})}{\Gamma(\lambda_{it})\Gamma(y_{it} + 1)} \left(\frac{1}{1 + \theta_i}\right)^{\lambda_{it}} \left(\frac{\theta_i}{1 + \theta_i}\right)^{y_{it}} \quad (15.38)$$

Koşullu sabit etkiler NegBin log olabilirlik aşağıdaki gibidir:

$$\mathcal{L} = \sum_i [\log \Gamma(\sum_t \lambda_{it}) + \log \Gamma(\sum_t y_{it} + 1) - \log \Gamma(\sum_t \lambda_{it} + \sum_t y_{it})] + \sum_i \sum_t [\log \Gamma(y_{it} + \lambda_{it}) - \log \Gamma(\lambda_{it}) - \log \Gamma(y_{it} + 1)] \quad (15.39)$$

15.2.2.1. Sabit Etkiler Panel Negatif Binom Model Stata Uygulaması

Sabit etkiler Negbin modelinin tahmini için kullanılan Stata komutunun temel yapısı aşağıdaki gibidir:

```
xtnbreg depvar [indepvars] [if] [in] [weight] , fe [FE_options]
```

Komut yapısı hakkında detaylı bilgi için `help xtnbreg` komutu yardımıyla açılan bilgilendirme ekranından detaylara ulaşabilirsiniz. Patent sayısı ile Ar-Ge harcamalarının GSMH içindeki payı, yayınlanan akademik makale sayısı, kurulan yeni start-up işletme sayısı arasındaki ilişkinin sabit etkiler NegBin model tahminlerini elde etmek için aşağıdaki komut kullanılır.

Tablo 15.25. Sabit Etkiler NegBin Model Tahmin Sonuçları

```
xtnbreg patent arge makale startup gdp ,fe nolog irr
```

kullanılan opsiyonlar:

fe : sabit etkiler

nolog : iterasyonları göstermemesi için

irr : incidence-rate ratio hesaplanması için

```
Conditional FE negative binomial regression      Number of obs      =          680
Group variable: id                             Number of groups    =           38
                                                Obs per group:
                                                min =             15
                                                avg  =            17.9
                                                max  =             18
                                                Wald chi2(4)       =          378.61
Log likelihood = -4974.8637                     Prob > chi2         =           0.0000
```

patent	IRR	Std. Err.	z	P> z	[95% Conf. Interval]
-----+-----					
arge	161.11	42.59565	19.22	0.000	95.95649 270.502
makale	.9999986	2.09e-06	-0.68	0.499	.9999945 1.000003
startup	.9899348	.0048399	-2.07	0.039	.980494 .9994666
gdp	1.375551	.4413497	0.99	0.320	.7334422 2.579807
_cons	4.994707	.4272795	18.80	0.000	4.223697 5.90646
-----+-----					

Note: _cons estimates baseline incidence rate (conditional on zero random effects).

Tablo 15.25'te yer alan model tahmin sonuçlarına bakıldığında Ar-Ge harcamalarının GSMH oranındaki bir birimlik artışın patent sayıları üzerinde istatistiksel olarak anlamlı ve %160'lık (161-1=160) bir artışa, start-up sayısındaki bir birimlik artışın patent sayıları üzerinde %5 seviyesinde istatistiksel olarak anlamlı ve %1'lik (0.99-1=-0.01) azalışa neden olduğu görülmektedir. Diğer değişkenlere dair tahmin edilen katsayılar istatistiksel olarak anlamlı değildir.

Eğer zaman boyutu çok uzun değilse, poisson ya da negatif binom gibi doğrusal olmayan modellerde sabit etkiler ve rassal etkiler tahmincileri için marjinal etkiler anlamlı bilgiler sunmaz. Dikkat edilirse Tablo 15.25.a ve Tablo 15.25.b'de yer alan marjinal etkilere dair tahmin sonuçları ile Tablo 15.27'de birinci sütunda yer alan sabit etkiler NegBin tahmin sonuçları aynı değerleri vermektedir. Bu nedenle sabit etkiler panel NegBin modellerinde mar-

jinal etkiler yerine esneklikler ve yarı esnekliklerin hesaplanması önerilir. Sabit etkiler panel NegBin modeline ait esneklikler ile yarı esneklikler sırasıyla Tablo 15.25.c ve Tablo 15.25.d'de gösterilmektedir.

Tablo 15.25.a. Sabit Etkiler Panel NegBin: Ortalamaya Göre Marjinal Etkiler

```
qui xtnbreg patent arge makale startup gdp ,fe nolog
margins, dydx(*) atmean noatlegend
```

Conditional marginal effects				Number of obs	=	680	
Model VCE : OIM							
Expression : Linear prediction, predict()							
dy/dx w.r.t. : arge makale startup gdp							

		Delta-method					
		dy/dx	Std. Err.	z	P> z	[95% Conf. Interval]	

arge		5.082087	.2643887	19.22	0.000	4.563895	5.600279
makale		-1.41e-06	2.09e-06	-0.68	0.499	-5.50e-06	2.68e-06
startup		-.0101162	.0048892	-2.07	0.039	-.0196987	-.0005336
gdp		.3188541	.3208532	0.99	0.320	-.3100066	.9477147

Tablo 15.25.b. Sabit Etkiler Panel NegBin: Ortalama Marjinal Etkiler

```
qui xtnbreg patent arge makale startup gdp ,fe nolog
margins, dydx(*)
```

Average marginal effects				Number of obs	=	680	
Model VCE : OIM							
Expression : Linear prediction, predict()							
dy/dx w.r.t. : arge makale startup gdp							

		Delta-method					
		dy/dx	Std. Err.	z	P> z	[95% Conf. Interval]	

arge		5.082087	.2643887	19.22	0.000	4.563895	5.600279
makale		-1.41e-06	2.09e-06	-0.68	0.499	-5.50e-06	2.68e-06
startup		-.0101162	.0048892	-2.07	0.039	-.0196987	-.0005336
gdp		.3188541	.3208532	0.99	0.320	-.3100066	.9477147

Tablo 15.25.c. Sabit Etkiler Panel NegBin: Esneklikler

qui xtnbreg patent arge makale startup gdp ,fe nolog
margins, eyex(*)

Average marginal effects			Number of obs		=		680
Model VCE : OIM							
Expression : Linear prediction, predict()							
ey/ex w.r.t. : arge makale startup gdp							

		Delta-method					
		ey/ex	Std. Err.	z	P> z	[95% Conf. Interval]	

arge		.3011307	.0164171	18.34	0.000	.2689538	.3333076
makale		-.0102475	.0156029	-0.66	0.511	-.0408287	.0203337
startup		-.0339247	.0165524	-2.05	0.040	-.0663667	-.0014827
gdp		.0045264	.0045294	1.00	0.318	-.004351	.0134038

Sabit etkiler panel NegBin modeline ait esnekliklerin sunulduğu Tablo 15.25.c, Ar-Ge harcamalarında %1'lik artış yaşandığında patent sayısının %0.30 artacağını, start-up firma sayısında %1'lik artış yaşandığında ise patent sayısının %0.03 azalacağını göstermektedir.

Tablo 15.25.d. Sabit Etkiler Panel NegBin: Yarı Esneklikler

qui xtnbreg patent arge makale startup gdp ,fe nolog
margins, eydx(*)

Average marginal effects				Number of obs	=	680
Model VCE : OIM						
Expression : Linear prediction, predict()						
ey/dx w.r.t. : arge makale startup gdp						

		Delta-method				
		ey/dx	Std. Err.	z	P> z	[95% Conf. Interval]

arge		2.333529	.1455813	16.03	0.000	2.048195 2.618863
makale		-.00000064	9.60e-07	-0.67	0.500	-2.53e-06 1.23e-06
startup		-.004645	.0022527	-2.06	0.039	-.0090602 -.0002298
gdp		.1464074	.1473815	0.99	0.321	-.142455 .4352698

Sabit etkiler panel NegBin modeline ait yarı esnekliklerin hesaplandığı Tablo 15.25.d'ye göre Ar-Ge harcamalarında bir birim artış yaşandığında patent sayısı %233 artmakta, start-up şirket sayısı bir adet arttığında ise patent sayısı% 0.04 azalmaktadır.

Sabit etkiler NegBin modelinde tahmin edilen katsayıların sıfırdan farklı olup olmadığı Wald testi yardımıyla yapılır. Wald test istatistiğinin hesaplanan 378.61 değeri için p değeri (olasılık değeri) 0.0000<0.05 olduğundan modelde kullanılan katsayıların bir bütün olarak anlamlı olduğu sonucuna ulaşılır.

Ayrıca sabit etkiler panel NegBin modellerinde Wald testine alternatif olarak olabilirlik oranı (LR) testi de yapılabilir. Bu durumda `lrmodel` opsiyonu kullanılır. Söz konusu komut ve tahmin sonuçları aşağıda yer almaktadır. Tablo 15.26'dan görüleceği üzere LR test sonuçları da Wald test sonuçları ile paralellik göstermektedir.

Tablo 15.26. Sabit Etkiler NegBin Model LR Test Sonuçları

```
xtnbreg patent arge makale startup gdp ,fe nolog irr lrmodel
fe      : sabit etkiler
nolog   : iterasyonları göstermemesi için
irr      : incidence-rate ratio hesaplanması için
lrmodel : olabilirlik oranı (LR) testi için
```

Conditional FE negative binomial regression	Number of obs	=	680
Group variable: id	Number of groups	=	38
	Obs per group:		
	min	=	15
	avg	=	17.9
	max	=	18
	LR chi2(4)	=	227.56
Log likelihood = -4974.8637	Prob > chi2	=	0.0000

(model tahmin sonuçları dışlandı)

Kısa panel verilerde standart olarak daha tutarlı tahminlerin elde edilmesi için birim üzerinden kümeleme yapılarak kümelenmiş dirençli standart hataların hesaplanması gerekir. Sabit etkiler NegBin tahmini yapılırken iki farklı varyans tahmin edicisi opsiyonu olan `vce(jackknife)` veya `vce(bootstrap)` kullanılarak birim üzerinden kümelenmiş dirençli standart hatalar elde edilir. Söz konusu standart hata opsiyonlarına dair komutlar ve karşılaştırma tablosu aşağıda yer almaktadır.

```
qui xtnbreg patent arge makale startup gdp, fe nolog
estat ic
estimates store nb_fe_temel
qui xtnbreg patent arge makale startup gdp, fe nolog vce(boot)
estat ic
estimates store nb_fe_bootstrap
qui xtnbreg patent arge makale startup gdp, fe nolog vce(jack)
estat ic
estimates store nb_fe_jackknife
estimates table nb_fe_temel nb_fe_bootstrap nb_fe_jackknife, b se t p
```

Tablo 15.27. Sabit Etkiler NegBin Modeli İçin Dirençli Tahminciler

Variable	nb_fe_temel	nb_fe_bootstrap	nb_fe_jackknife
arge	5.0820871	5.0820871	5.0820871
	.2643887	.9017648	.7445605
	19.22	5.64	6.83
	0.0000	0.0000	0.0000
makale	-1.409e-06	-1.409e-06	-1.409e-06
	2.087e-06	6.011e-06	6.760e-06

	-0.68	-0.23	-0.21
	0.4995	0.8146	0.8363
startup	-.01011616	-.01011616	-.01011616
	.00488916	.00508575	.00480129
	-2.07	-1.99	-2.11
	0.0385	0.0467	0.0436
gdp	.31885408	.31885408	.31885408
	.32085315	.2938147	.25927685
	0.99	1.09	1.23
	0.3203	0.2778	0.2283
_cons	1.6083787	1.6083787	1.6083787
	.08554646	.39172644	.5007042
	18.80	4.11	3.21
	0.0000	0.0000	0.0031
Wald	378.61	37.87	20.29
p-value	0.0000	0.0000	0.0000

 legend: tahmin katsayısı/standart hata/ z değeri /p değeri

Dirençli model tahmin sonuçlarının yer aldığı Tablo15.27'ye bakıldığında temel model ile dirençli modellere ait standart hatalarda değişimin sınırlı kaldığı görülmektedir. Bu durumda kurulan modeller arasında seçim yapılmasında model uyum iyiliği olarak kullanılan Wald test değerinden yararlanılabilir. Ancak verilen örnekte her üç model için katsayıların toplu olarak anlamlı olduğu görülmektedir. Ayrıca model sonuçlarından elde edilen AIC ve BIC değerleri her üç model için de aynı değerleri vermektedir. Dolayısıyla bu aşamada istatistiksel olarak model seçiminin yapılması uygun değildir. Bu durumda araştırmacının önsel olarak veri seti yapısından kaynaklı olası bilgiler altında düzeltilmiş standart hataların oluşturulduğu modellerden birini seçebilir.

Bu bölümde kullanılan Stata komutları aşağıda toplu olarak yer almaktadır.

```
xtnbreg patent arge makale startup gdp ,fe nolog irr
qui xtnbreg patent arge makale startup gdp ,fe nolog
margins, dydx(*) atmean noatlegend
margins, dydx(*)
margins, eyex(*)
margins, eydx(*)
xtnbreg patent arge makale startup gdp ,fe nolog irr lrmodel
qui xtnbreg patent arge makale startup gdp, fe nolog
estat ic
estimates store nb_fe_temel
qui xtnbreg patent arge makale startup gdp, fe nolog vce(boot)
estat ic
estimates store nb_fe_bootstrap
qui xtnbreg patent arge makale startup gdp, fe nolog vce(jack)
estat ic
estimates store nb_fe_jackknife
estimates table nb_fe_temel nb_fe_bootstrap nb_fe_jackknife, b se t p
```

15.2.3 Rassal Etkiler Panel Negatif Binom Modeli

Rassal etkiler panel NegBin modelinde temel varsayımlar aşağıdaki denklemlerde gösterildiği gibidir (Winkelmann, 2008: 229):

$$\mu_{it} \sim \Gamma(v_i \lambda_{it}, \theta_i) \quad (15.40.a)$$

$$v_i \lambda_{it} = \exp(x'_{it} \beta + \eta_{it}) \quad (15.40.b)$$

$$\frac{1}{1+\theta_i} \sim \text{beta}(r, s) \quad (15.40.c)$$

ve rassal etkiler NegBin olasılık fonksiyonu Denklem (15.41)'de gösterildiği gibi olur:

$$f(y_{it}) = \prod_t \frac{\Gamma(y_{it} + \lambda_{it})}{\Gamma(\lambda_{it}) \Gamma(y_{it} + 1)} \left(\frac{1}{1+v_i} \right)^{\lambda_{it}} \left(\frac{v_i}{1+v_i} \right)^{y_{it}} \quad (15.41)$$

Buradan hareketle rassal etkiler NegBin modeline ait koşullu log olabilirlik fonksiyonu aşağıdaki gibidir:

$$\mathcal{L} = \sum_i [\log \Gamma(r + s) + \log \Gamma(r + \sum_t \lambda_{it}) + \log \Gamma(s + \sum_t y_{it} + 1)] - \sum_i [\log \Gamma(r) + \log \Gamma(s) + \log \Gamma(r + s + \sum_t \lambda_{it} + \sum_t y_{it})] + \sum_i \sum_t [\log \Gamma(y_{it} + \lambda_{it}) - \log \Gamma(\lambda_{it}) - \log \Gamma(y_{it} + 1)] \quad (15.42)$$

15.2.3.1. Rassal Etkiler Panel Negatif Binom Model Stata Uygulaması

Rassal etkiler NegBin modelinin tahmini için kullanılan Stata komutunun temel yapısı aşağıdaki gibidir:

xtnbreg depvar [indepvars] [if] [in] [weight] , re [RE_options]

Komut yapısı hakkında detaylı bilgi için **help xtnbreg** komutu yardımıyla açılan bilgilendirme ekranından detaylara ulaşabilirsiniz. Patent sayısı ile Ar-Ge harcamalarının GSMH içindeki payı, yayınlanan akademik makale sayısı, kurulan yeni start-up işletme sayısı arasındaki ilişkinin rassal etkiler NegBin model tahminlerini elde etmek için aşağıdaki komut kullanılır.

Tablo 15.28. Rassal Etkiler NegBin Model Tahmin Sonuçları

xtnbreg patent arge makale startup gdp ,re nolog irr

kullanılan opsiyonlar:

re : rassal etkiler

nolog : iterasyonları göstermemesi için

irr : incidence-rate ratio hesaplanması için

Random-effects negative binomial regression	Number of obs	=	680
Group variable: id	Number of groups	=	38
Random effects u_i ~ Beta	Obs per group:		
	min	=	15
	avg	=	17.9
	max	=	18
	Wald chi2(4)	=	390.70
Log likelihood = -5452.6932	Prob > chi2	=	0.0000

patent	IRR	Std. Err.	z	P> z	[95% Conf. Interval]	
arge	166.7513	43.70046	19.52	0.000	99.76886	278.7041
makale	.9999987	2.05e-06	-0.65	0.513	.9999946	1.000003
startup	.9895206	.0048223	-2.16	0.031	.9801141	.9990174
gdp	1.380466	.4427029	1.01	0.315	.7362969	2.588202
_cons	4.981906	.4246523	18.84	0.000	4.215412	5.887773
/ln_r	-.7127347	.1898918			-1.084916	-.3405536
/ln_s	3.329787	.3035753			2.734791	3.924784
r	.4903015	.0931043			.3379302	.7113764
s	27.9324	8.479586			15.40652	50.64213

Note: Estimates are transformed only in the first equation.

Note: _cons estimates baseline incidence rate (conditional on zero random effects).

LR test vs. pooled: chibar2(01) = 2800.45

Prob >= chibar2 = 0.000

Tablo 15.28'de yer alan model tahmin sonuçlarına bakıldığında Ar-Ge harcamalarının GSMH oranındaki bir birimlik artışın patent sayıları üzerinde istatistiksel olarak anlamlı ve %66'lık ($1.66-1=0.66$) bir artışa neden olduğu görülmektedir. Start-up sayısındaki bir birimlik artışın patent sayıları üzerinde %5 seviyesinde istatistiksel olarak anlamlı ve %1.1'lik ($0.989-1=-0.011$) azalışa neden olduğu söylenebilir. Ancak diğer değişkenlerin tahmin edilen katsayıları istatistiksel olarak anlamlı değildir.

Sabit etkiler NegBin ile rassal etkiler NegBin modellerindeki temel ayrım " θ " yayılım parametresi üzerinden yapılmaktadır. Sabit etkiler NegBin modelinde yayılım parametresi aynı gruptaki tüm elementler için aynı iken " θ ", rassal etkiler NegBin modelinde yayılım parametresinin " θ_i " gruptan gruba rassal dağıldığı ve Denklem (15.40.c)'de gösterilen $\frac{1}{1+\theta_i} \sim \text{beta}(r, s)$ koşulunun sağlandığı varsayılır. Tablo 15.28'de yer alan r ve s parametresi söz konusu $\frac{1}{1+\theta_i} \sim \text{beta}(r, s)$ 'de gösterilen r ve s parametreleri temsil ederken /ln_r ve /ln_s sırasıyla r ve s'nin logaritmik değerleri olan $\ln(r)$ ve $\ln(s)$ 'yi temsil etmektedir. Örnek üzerinden yayılım parametresine bağlı olarak oluşturulan $\frac{1}{1+\theta_i}$ 'nin; ortalaması 0.49, varyansı 27.93 olan beta dağılıma sahip olduğu söylenebilir.

Tablo 15.28'de rassal etkiler NegBin modeli ile havuzlanmış NegBin modelinin karşılaştırmasına olanak tanıyan LR testi, rassal etkiler NegBin tahminlerinin altında hesaplanmaktadır. Tablo 15.28'in en altındaki satırda yer alan LR testi panel tahminlerin havuzlanmış tahminlerden istatistiksel olarak farklı olduğunu, bu nedenle havuzlanmış NegBin modelinin tercih edilmemesi gerektiğini göstermektedir.

Eğer zaman boyutu çok uzun değilse, poisson ya da negatif binom gibi doğrusal olmayan modellerde sabit etkiler ve rassal etkiler tahminçileri için marjinal etkiler anlamlı bilgiler sunmaz (Hilbe, 2014). Bu nedenle rassal etkiler panel NegBin modellerinde marjinal etkiler yerine esneklikler ve yarı esnekliklerin hesaplanması önerilir. Rassal etkiler panel NegBin

Rassal etkiler NegBin modelinde tahmin edilen katsayılarının sıfırdan farklı olup olmadığı Wald testi yardımıyla yapılır. Wald test istatistiğinin hesaplanan 390.70 değeri için p değeri (olasılık değeri) $0.0000 < 0.05$ olduğundan modelde kullanılan katsayıların bir bütün olarak anlamlı olduğu sonucuna ulaşılır. Ayrıca rassal etkiler NegBin modellerinde Wald testine alternatif olarak olabirlik oranı (LR) testi de yapılabilir. Bu durumda **lrmodel** opsiyonu kullanılır. Söz konusu komut ve tahmin sonuçları aşağıda yer almaktadır. Tablo 15.29'dan görüleceği üzere LR test sonuçları da Wald test sonuçları ile paralellik göstermektedir.

Tablo 15.29. Rassal Etkiler NegBin Model LR Test Sonuçları

```
xtnbreg patent arge makale startup gdp ,re nolog irr lrmodel
```

re: rassal etkiler

nolog: iterasyonları göstermemesi için

irr : incidence-rate ratio hesaplanması için

lrmodel: olabirlik oranı (LR) testi için

Random-effects negative binomial regression	Number of obs	=	680
Group variable: id	Number of groups	=	38
Random effects u_i ~ Beta	Obs per group:		
	min	=	15
	avg	=	17.9
	max	=	18
	LR chi2(4)	=	235.04
Log likelihood = -5452.6932	Prob > chi2	=	0.0000

(model tahmin sonuçları dışlandı)

Kısa panel verilerde standart olarak daha tutarlı tahminlerin elde edilmesi için birim üzerinden kümeleme yapılarak kümelenmiş dirençli standart hataların hesaplanması gerekir. Rassal etkiler NegBin tahmini yapılırken iki farklı varyans tahmin edicisi opsiyonu olan **vce(jackknife)** veya **vce(bootstrap)** kullanılarak birim üzerinden kümelenmiş dirençli standart hatalar elde edilir. Söz konusu standart hata opsiyonlarına dair komutlar ve karşılaştırma tablosu aşağıda yer almaktadır.

```
qui xtnbreg patent arge makale startup gdp, re nolog
estat ic
estimates store nb_re_temel
qui xtnbreg patent arge makale startup gdp, re nolog vce(boot)
estat ic
estimates store nb_re_bootstrap
qui xtnbreg patent arge makale startup gdp, re nolog vce(jack)
estat ic
estimates store nb_re_jackknife
estimates table nb_re_temel nb_re_bootstrap nb_re_jackknife, b se t p
```

Tablo 15.30. Rassal Etkiler NegBin Modeli İçin Dirençli Tahminciler

Variable	nb_re_temel	nb_re_bootstrap	nb_re_jackknife
patent			
arge	5.1165033	5.1165033	5.1165033
	.2620697	.8897399	1.0517904
	19.52	5.75	4.86
	0.0000	0.0000	0.0000
makale	-1.339e-06	-1.339e-06	-1.339e-06
	2.049e-06	6.402e-06	6.781e-06
	-0.65	-0.21	-0.20
	0.5134	0.8344	0.8446
startup	-.01053468	-.01053468	-.01053468
	.00487334	.00534817	.00516202
	-2.16	-1.97	-2.04
	0.0306	0.0489	0.0485
gdp	.3224209	.3224209	.3224209
	.32069095	.34739422	.27952462
	1.01	0.93	1.15
	0.3147	0.3533	0.2561
_cons	1.6058125	1.6058125	1.6058125
	.08523893	.49091163	.51387706
	18.84	3.27	3.12
	0.0000	0.0011	0.0035
/ln_r	-.71273475	-.71273475	-.71273475
	.18989184	.20716694	.2323707
	-3.75	-3.43	-3.07
	0.0002	0.0006	0.0040
/ln_s	3.3297873	3.3297874	3.3297873
	.30357527	.39056331	.60297008
	10.97	8.53	5.52
	0.0000	0.0000	0.0000
Wald	390.70	40.40	7.82
p-value	0.0000	0.0000	0.0000

legend: tahmin katsayısı/standart hata/ z değeri /p değeri

Dirençli model tahmin sonuçlarının yer aldığı Tablo 15.30'a bakıldığında temel model ile dirençli modellere ait standart hatalarda değişimin sınırlı kaldığı görülmektedir. Bu durumda kurulan modeller arasında seçim yapılmasında model uyum iyiliği olarak kullanılan Wald test değerinden yararlanılabilir. Ancak örnekte her üç model içinde katsayıların toplu olarak anlamlı olduğu görülmektedir. Ayrıca model sonuçlarından elde edilen AIC ve BIC değerleri her üç model için de aynı değerleri vermektedir. Dolayısıyla bu aşamada istatistiksel olarak model seçiminin yapılması mümkün değildir. Bu durumda araştırmacı önsel olarak veri setinin yapısından kaynaklı olası bilgiler altında düzeltilmiş standart hataların oluşturulduğu modellerden birini seçebilir.

15.2.4. Model Seçimi

Teorik açıdan bakıldığında basit olarak Panel Poisson ve Panel NegBin modelleri arasındaki temel ayrım aşırı yayılım parametresi üzerinden yapılmaktadır. Aşırı yayılım durumu veri setlerinde sıklıkla karşılaşılan bir durum olmakla birlikte aşırı yayılım durumunda poisson modellerinin kullanılması standart hataların ciddi derecede düşük tahmin edilmesine ve regresyon parametrelerinden yanıltıcı çıkarımların yapılmasına neden olabilir (Hinde ve Demetrio, 1998).

Ancak birçok durumda olduğu gibi koşullu ortalamaların tahmin edilmesi amaçlanıyorsa bu durumda ortalama tahmin edicileri sözde aşırı yayılımı dikkate alan tahmincilerden daha güvenli bir yöntemdir. Eğer araştırmacı tarafından olayların olasılıklarının hesaplanması istenmiyorsa aşırı yayılım ciddi bir problem yaratmamaktadır.

NegBin tahmin edicileri ile rassal etkiler tahmin edicileri aşırı yayılım nedeniyle dağılımlar üzerinde çok güçlü varsayımlara dayanmaktadır ki bu durum söz konusu tahmin ediciler için yapılan en önemli eleştiridir. Wooldridge (1999) sabit etkiler poisson tahmin edicilerinin önemli ölçüde tutarlı sonuçlar ürettiğini belirtmektedir. Tüm bu belirtilen nedenlerden dolayı standart hataların uygun şekilde düzeltildiği dirençli sabit etkiler poisson model tahmin edicilerinin tercih edilmesi önerilmektedir.

Bu bölümde kullanılan Stata komutları aşağıda toplu olarak yer almaktadır.

```
xtnbreg patent arge makale startup gdp ,re nolog irr
qui xtnbreg patent arge makale startup gdp ,re
margins, eyex(*)
margins, eydx(*)
xtnbreg patent arge makale startup gdp ,re nolog irr lrmodel
qui xtnbreg patent arge makale startup gdp, re nolog
estat ic
estimates store nb_re_temel
qui xtnbreg patent arge makale startup gdp, re nolog vce(boot)
estat ic
estimates store nb_re_bootstrap
qui xtnbreg patent arge makale startup gdp, re nolog vce(jack)
estat ic
estimates store nb_re_jackknife
estimates table nb_re_temel nb_re_bootstrap nb_re_jackknife, b se t p
```

15.3. KAYNAKLAR

- Cameron, A. C., & Trivedi, P. K. (2009). *Microeconometrics using Stata*. Vol. 5. College Station, TX: Stata press.
- Cameron, A. C., & Trivedi, P. K. (2013). *Regression analysis of count data* (Vol. 53). Cambridge university press.
- Cameron, A. C., & Trivedi, P. K. (2015). *Count panel data*. The Oxford Handbook of Panel Data içinde Editör: Baltagi, Badi H., Oxford University Press.
- Cui, James (2007). QIC program and model selection in GEE analyses. *The Stata Journal*, 7(2): 209-220. <https://doi.org/10.1177%2F1536867X0700700205>

- Jisheng Cui & Liyun Feng (2008). Correlation Structure and Model Selection for Negative Binomial Distribution in GEE. *Communications in Statistics - Simulation and Computation*, 38:1, 190-197. <https://doi.org/10.1080/03610910802446985>
- Greene, William H. (2008). *Econometric analysis*. 6th ed., Upper Saddle River, N.J.: Prentice Hall.
- Hilbe, J. M. (2011). *Negative binomial regression*. Cambridge University Press.
- Hilbe, J. M. (2014). *Modeling count data*. Cambridge University Press.
- J. Hinde, & G.B. Demetrio (1998). Overdispersion: Models and estimation *Computational Statistics & Data Analysis*, 27: 151-170. [https://doi.org/10.1016/S0167-9473\(98\)00007-3](https://doi.org/10.1016/S0167-9473(98)00007-3)
- Liang, K. Y. & Zeger, S. L. (1986). A comparison of two bias-corrected covariance estimators for generalized estimating equations. *Biometrika*, 73:1, 13-22. <https://doi.org/10.1093/biomet/73.1.13>
- Long, J. & J. Freese (2014): *Regression models for categorical dependent variables using Stata*. Stata Press, 3rd. edition.
- McCullagh, P., & J. A. Nelder. (1989). *Generalized Linear Models*. 2nd ed. London: Chapman & Hall.
- Rabe-Hesketh, S., & Skrondal, A. (2012). *Multilevel and longitudinal modeling using Stata*. STATA press.
- StataCorp. 2019. *Stata: Release 16. Statistical Software*. College Station, TX: StataCorp LLC.
- Wedderburn, R. W. M. (1974). Quasilikelihood functions, generalized linear models, and the Gauss-Newton method. *Biometrika* 61:3, 439-447. <https://doi.org/10.1093/biomet/61.3.439>
- Winkelmann, R. (2008). *Econometric analysis of count data*. Springer Springer, Berlin, Heidelberg. <https://doi.org/10.1007/978-3-540-78389-3>
- Wooldridge, J. M. (1999). Distribution-free estimation of some nonlinear panel data models. *Journal of Econometrics*, 90:1, 77-97. [https://doi.org/10.1016/S0304-4076\(98\)00033-5](https://doi.org/10.1016/S0304-4076(98)00033-5)
- Y.-G. Wang & V. Carey, (2003). Working correlation structure misspecification, estimation and covariate design: implications for generalized estimating equations performance. *Biometrika*, 90(1): 29-41. <https://doi.org/10.1093/biomet/90.1.29>

BÖLÜM 16

YAŞAM ANALİZLERİ VE STATA UYGULAMALARI

Arzu Kökcen Eryavuz¹

Zaman ve olay olmak üzere iki önemli unsuru bulunan yaşam analizleri (sağkalım analizleri-survival analysis), belirlenen zaman aralığında tanımlanan bir olayın başlamasından, gerçekleşmesine kadar geçen süre ile ilgilenen temel bir araştırma yöntemidir. Olayların oluşumunu ve zamanlamasını göz önünde bulunduran yaşam analizleri, sosyolojide olay geçmişi analizi (Tuma and Hannan 1984), sağlık alanında yaşam çözümlemesi (Cox and Oakes 1984), mühendislikte ise güvenilirlik kuramı (Barlow and Proschan 1975) olarak adlandırılmaktadır (Hutchison, 1988: 206).

Başarısızlık analizi (failure time analysis) olarak da ifade edilen yaşam analizlerinde tanımlanan olayı; bir hastalığın ortaya çıkması, hastalık nedeniyle ölümün gerçekleşmesi, depremler, bir makinanın ya da makine parçalarının bozulması, enflasyon, banka kredileri, işsizlik, çalışma, grevler, firmaların hayatta kalması ya da kapanması, eğitim, evlilik, doğum, boşanma gibi örnekler oluşturmaktadır.

Herhangi bir olay için başarısızlığın sadece bir defa gerçekleştiği yaşam analizlerin de amaç; farklı zamanlarda yaşam olasılığı tahminlerinin elde edilmesi, yaşam süresinin dağılımının belirlenmesi, farklı grupların yaşam süresi dağılımlarının karşılaştırılması ve başarısızlık süresini etkileyen faktörlerin belirlenmesidir (Tuncer, 2014: 4).

Yaşam analizleri 17.yy'da demografide ve aktüeryal bilimde mortalite (ölüm) tablolarının kullanılmasıyla bilinen en eski istatistiksel disiplinlerden biridir. İkinci dünya savaşı kadar bir araştırma yöntemi olarak ortaya konulamayan yaşam çözümlemeleri, savaş sırasında askeri teçhizatın güvenilirliğine veya başarısızlık süresine olan ilginin artması sonucu ortaya çıkmıştır (Dickman, 2014: 1; Smith ve Smith, 1972: 1). Savaş sonunda ölüm oranlarının araştırılmasından başarısızlık zamanının araştırılmasına kadar uzanan bu yeni istatistiksel yöntemler, özellikle elektronik endüstrisi alanı olmak üzere özel sektörde yayılarak kullanılmaya başlamıştır (Singh ve Mukhopadhyay, 2011: 145). Özellikle 20.yy'ın ikinci yarısından itibaren büyük gelişme gösteren yaşam analizleri, başta tıp alanı olmak üzere, mühendislik, biyoistatistik, iktisat, eğitim, demografi ve sosyoloji gibi birçok alanda kullanılmıştır. Yaşam analizlerinin kullanımı yaygınlaştıkça,

¹ Dr., Araştırmacı. E-mail: arzukokcen@yahoo.com

parametrik olmayan ve parametrik yaklaşımlar geliştirilmiştir. Bu bölüm içerisinde parametrik olmayan yaşam analizi modellerine değinilmiştir.

Araştırmacı tarafından belirlenen olayın başlangıç zamanı ile başarısızlığı (ölümü) arasında geçen yaşam süresi ya da başarısızlık süresi; yıl, ay, gün gibi belirli bir süre olarak tanımlanmakta ve T ile gösterilmektedir. Yaşam süresi T 'yi belirlemek için üç temel unsura ihtiyaç durulmaktadır (Le, 1997: 3):

- zamanın başlangıç noktası veya tanımlanan olayın başlangıç noktası,
- ilgilenilen olayın sonlanması ve
- zaman geçişi için ölçme skalası.

Sıfırdan büyük ve pozitif değerli olan gözlenen yaşam ya da başarısızlık süresi, yaşam analizlerinde kullanılan bağımlı değişkendir.

Yaşam analizinin avantajları; araştırılan konuların örneği için betimleyici olması, deneklerin temsili nüfusu için tahmini olması (anakütleden çekilen örneklem için), karşılaştırmalı bir yöntem olarak objektif ve yüksek bir doğruluk sunması olarak sıralanabilir (Danacica ve Babucea, 2010: 440).

16.1. Yaşam Analizinde Kullanılan Sansürlü Veri Türleri

Özelliklerinden dolayı nicel verilere kıyasla farklı bir istatistiksel analiz gerektiren yaşam verileri çoğu kez sansürlü (censored) olmaktadır. Sansür, bir çalışmada olay tüm bireyler için gözlenemediğinde, başarısızlık (ölüm) meydana gelmediğinde gerçekleşir. Sansür birçok farklı nedenden dolayı oluşmaktadır: *Araştırmalarda olay gerçekleşmeden önce veri toplama işleminin bitmesi veya bir veya daha fazla kişinin araştırmadan ayrılması durumunda genellikle sansür meydana gelir* (Malacane ve diğerleri, 1997: 4). Basit ortalamanın hesaplanamadığı ve standart sapmanın bulunamadığı sansürlü verilerle analiz gerçekleştirmek için sıradan istatistiksel yöntemleri kullanmak mümkün değildir (Justine, 2013: 8). Yaşam analizi verilerine özgü olan sansürün uygulanabileceği birkaç farklı biçim vardır ve bir çalışmada yer alan sansür türünü tanımlamak önemlidir.

Uygulamada en çok karşılaşılan sağdan sansür (right censored), başlangıçta belirlenen çalışma periyodunda ilgilenilen olayın gözlemlenemediği ya da belli bir süre sonunda herhangi bir bilginin çalışmada yer alan bir bireyden alınamadığı durumlarda ortaya çıkmaktadır (Yazıcı ve diğerleri, 2017: 4360). i . birey için sansür zamanı t_i ve yaşam süresi T_i olarak tanımlanırsa, her bir birey için sansür göstergesi,

$$\delta_i = \begin{cases} 1, & T_i > t_i \quad \text{gözlem sansürlü} \\ 0, & T_i \leq t_i \quad \text{gözlem sansürsüz} \end{cases} \quad i = 1, 2, \dots, n \quad (16.1)$$

olarak tanımlanır (Steele, 2005: 5).

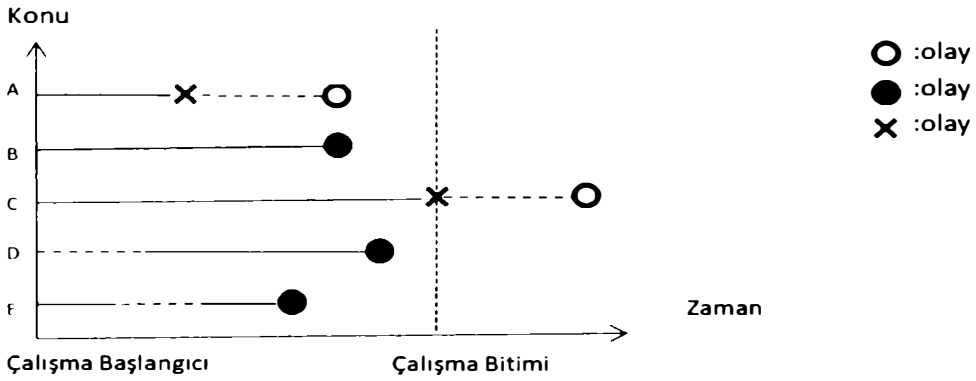
Birey için tanımlanan olayın, yaşam sürdürme zamanı (iyi belirlenmiş bir başlangıç zamanı ile tanımlanan durumun ortaya çıktığı zaman arasında geçen süre) başlamadan önce ortaya çıktığı

durumlarda soldan sansür (left censored) ile karşılaşmaktadır ve başlangıç zamanı kesin olarak bilinmemektedir (Göz Çekçeki, 2007:9). Örneğin, bir kanser hastalığında, ilk tümörün alınmasından üç ay sonra kanserin tekrarlayıp tekrarlanmadığı araştırılıyor olsun. Bazı hastalarda tekrarlama süresi belirlenen yaşam sürdürme süresi başlamadan (üç aydan önce) ortaya çıkabilir. Bu durumda tekrarlama süresi soldan sansürlenmiştir (Collet, 2003: 2).

Aralıklı sansür (interval censored), bireyin zaman içerisindeki iki nokta arasında başarısızlığa sahip olduğu bilindiğinde ancak olayın ortaya çıkma zamanının kesin olarak bilinmediği durumlarda karşılaşılan sansür tipidir (Allison, 2010: 415). Örneğin, ilk tümörü alınan kanser hastasının ilk üç aylık kontrolünde tümör tekrarlamamışken altıncı ayda yapılan kontrolde tümör tekrarlamış ise hastalığın tekrarlanma süresi üç ile altıncı aylar arasında olacaktır ve tekrarlama süresi aralıklı sansürlenmiş olacaktır (Collet, 2003: 2).

Tanımlanan sansür türleri Şekil 16.1'de gösterilmiştir:

Şekil 16.1. Sansürlü ve Sansürsüz Veri Örneği



A bireyi, çalışma henüz sona ermeden farklı bir nedenden dolayı çalışmadan çekilmiştir ya da bir sebepten dolayı birey takip edilememiştir. Bu nedenle A bireyine ait yaşam süresi (tesadüfi-rasgele- random) sağdan sansürlüdür. Çünkü sansür araştırmacının kontrolünde değildir. Çalışmanın başında izlenmeye başlanan B bireyinin başarısızlık durumu kesindir. Belirlenen olay çalışma süresi içerisinde gerçekleştiği için bu bireye ait veri sansürsüzdür. C bireyi, çalışmanın başından çalışmanın sonuna kadar belirlenen olayı yaşamamış olup, çalışmanın bitiminden sonra olayı yaşamıştır. Yaşam süresi çalışma süresinden uzun olduğu için bu bireye ait veri (sabit-fixed) sağdan sansürlüdür. D bireyi çalışmaya geç katılmış olsa da, belirlenen olay çalışma süresi bitmeden gerçekleşmiştir. Bu nedenle de bireye ait veri sansürsüzdür. E bireyi çalışmanın başından itibaren belli bir süre gözlenmiş fakat bir süre takip edilememiş ve daha sonra tekrar geri dönmüştür. İzleme süresi içerisinde belirlenen olay gerçekleşmiş olsa da zamanı tam olarak bilmediği için bireye ait veri aralıklı sansüre sahiptir.

Bu bölümde Stata uygulamaları için Hacettepe Üniversitesi Nüfus Etütleri Enstitüsü tarafından gerçekleştirilen 2013 Türkiye Nüfus ve Sağlık Araştırması (TNSA) verileri kullanılmıştır. Sonuçları hızlı bir şekilde elde edebilmek ve kısıtlamaları en aza indirgeyebilmek için veri seti küçültülerek sadece Akdeniz bölgesinde yaşayan kadınların ilk evlilik yaşı için uygulamalar yapılacaktır. Kadının ilk evlilik yaşı ile kadının babasının eğitim durumu, kardeş sayısı ve çalışma durumu arasındaki ilişki için yapılacak uygulamada; babanın eğitim seviyesi (eğitimi yok, ilkök, ortaokul, lise ve yükseköğrenim) ve kardeş sayısı (1-2, 3-4, 5 ve üzeri) kategorik değişken olarak alınmıştır. Kadının çalışma durumu, 12 yaşından anketin yapıldığı tarihe kadar, gelir getirsin veya getirmesin en az 6 ay süreyle bir işte çalışıp çalışmadığını göstermektedir. Veri seti yaşam sürelerini (evlilik yaşı) ve sansür durumunu özetleyen değişkenleri içermektedir. Kadınların ilk evlilik yaşına kadar geçen zaman *yaşam süresi* olarak tanımlanmıştır. Kadınların medeni durumunun evli, boşanmış, dul ve eşiyle ayrı yaşamaya başlamış olması başarısızlık olarak belirtilmişken, anketin yapıldığı zamana kadar hâlâ bekar olan gözlemler sansürlü olarak tanımlanmıştır. Verilerde eksik gözlem (missing value) ve soldan sansür yoktur. İncelenen birey için değeri zamanla değişebilen açıklayıcı bir değişkenin olmadığı (zamanla değişen ortak değişkenlerin olmaması), tüm açıklayıcı değişkenlerin sabit bir değerinin olduğu varsayılacaktır. Öncelikle çalışmada kullanılacak değişkenlere ait bilgiler ve tanımlayıcı istatistikler elde edilmektedir:

Tablo 16.1. evlilikyaşı.dta Veri Seti İçeriği

```
use "DOSYAKONUMU\evlilikyaşı.dta"
```

```
describe
```

variable name	storage type	display format	value label	variable label
sansur	float	%9.0g		1=Evlendi, 0=Sansür
calisma	float	%9.0g		1=Çalışıyor, 0=Çalışmıyor
baegt	float	%9.0g		Baba Eğitim Seviyesi
egt_yok	byte	%8.0g		Eğitimi Yok
egt_ilkokul	byte	%8.0g		İlkokul Mezunu
egt_ortaokul	byte	%8.0g		Ortaokul Mezunu
egt_lise	byte	%8.0g		Lise Mezunu
egt_yukse	byte	%8.0g		Yükseköğrenim Mezunu
evl_yas	float	%9.0g		Evlilik Yaşı
krds	float	%9.0g		Kardeş Sayısı
krds1	byte	%8.0g		1-2
krds2	byte	%8.0g		3-4
krds3	byte	%8.0g		5 ve üzeri

Tablo 16.2. Değişkenlere Ait Tanımlayıcı İstatistikler

```
sum evl_yas sansur calisma egt_yok egt_ilkokul egt_ortaokul egt_lise egt_yuksek krds1
krds2 krds3
```

Variable	Obs	Mean	Std. Dev.	Min	Max
-----+-----					
evl_yas	1250	20.8808	5.012014	12	49
sansur	1250	.7552	.4301406	0	1
calisma	1250	.52	.4997998	0	1
egt_yok	1250	.2208	.4149519	0	1
egt_ilkokul	1250	.596	.4908939	0	1
-----+-----					
egt_ortaokul	1250	.0816	.2738639	0	1
egt_lise	1250	.064	.2448509	0	1
egt_yuksek	1250	.0376	.1903029	0	1
krds1	1250	.0936	.291388	0	1
krds2	1250	.3344	.4719691	0	1
-----+-----					
krds3	1250	.572	.4949869	0	1

Tablo 16.2 incelendiğinde, Akdeniz bölgesinde yaşayan kadınların ilk evlenme yaşı yaklaşık olarak ortalama 21'dir ve %75.52'si evlidir (sansür değişkeni). Kadınların %52'si gelir getirsin veya getirmesin herhangi bir işte çalışmıştır. %22.08'nin babasının hiç eğitimi yokken, %3.76'sının babası yükseköğrenim görmüştür. Kadınların %9.4'ü 1 ya da 2 kardeşe sahipken, %57.2'sinin 5'ten fazla kardeşi vardır.

Yaşam süresi analizi tahminleri için Stata komutu olarak **st** kullanılmaktadır. Yaşam süresi analizlerinin yapılabilmesi için öncelikle yaşam süresinin ve sansürün tanımlanması gerekmektedir. Düzenleme için gereken tek komut ise **stset**'tir:

```
stset yaşam süresi, failure (sansür==1)
```

failure (sansür==1) başarısızlık olayını belirtir. Eğer belirtilmezse, her kaydın bir başarısızlıkla biteceği varsayılır. Uygulamada yaşam süresi olarak kadının ilk evlilik yaşı (evl_yas) kullanılırken, kadının evlenmiş olması başarısızlıktır ve 1'e eşittir.

Tablo 16.3. Evlilik Yaşına Ait Yaşam Süresi Bilgileri

```
stset evl_yas, failure(sansur==1)
```

```
failure event:  sansur == 1
obs. time interval:  (0, evl_yas]
exit on or before:  failure
```

```
-----
1250 total observations
0 exclusions
-----
```

```

1250 observations remaining, representing
  944 failures in single-record/single-failure data
26101 total analysis time at risk and under observation
      at risk from t =      0
    earliest observed entry t =      0
      last observed exit t =     49

```

Tablo 16.3 incelendiğinde; toplam gözlem sayısı 1250 iken 944 başarısız (en az 1 kez evlenmiş) kadın vardır. Tüm kadınlar arasındaki süre toplamı (evlilik yaşları) ise 26101'dir.

Stset komutunun kullanımından sonra, verilerde bir dizi yeni değişken oluştuğuna ve hepsinin önünde “_” olduğuna dikkat edilmelidir.

Tablo 16.4. Elde Edilen Yeni Değişkenler

de _*

variable name	storage type	display format	value label	variable label

_st	byte	%8.0g		
_d	byte	%8.0g		
_t	byte	%10.0g		
_t0	byte	%10.0g		

Tüm yaşam analizi (**st**) komutları Tablo 16.4'te belirtilen değişkenleri kullanır, çünkü yaşam sürelerine ilişkin tüm bilgiler söz konusu dört değişken içinde yer alır. Yeni değişkenlere ait tanımlayıcı istatistikler ise **su _*** komutu ile elde edilebilir.

Tablo 16.5. Yeni Değişkenlere Ait Tanımlayıcı İstatistikler

su _*

Variable	Obs	Mean	Std. Dev.	Min	Max
-----+-----					
_st	1250	1	0	1	1
_d	1250	.7552	.4301406	0	1
_t	1250	20.8808	5.012014	12	49
_t0	1250	0	0	0	0

_st değişkeni, veri yaşam analizlerine dâhil edilmişse 1, hariç tutulmuşsa 0'dır.

_d değişkeni, 0 ve 1 değerlerini alan sansürleme göstergesidir ve başarısız olan değişkene karşılık gelir.

_t değişkeni, çalışma zamanına karşılık gelen süre değişkenidir (evlilik yaşı).

Son olarak, `_t0`, her kadın için çalışmaya giriş tarihini, yani olay riski altında ilk kez ne zaman gözlemlendiklerini kaydeden bir değişkendir. Bu veri kümesi için tüm durumlar, giriş anından itibaren gözlemlenmiştir, $t = 0$ olarak ifade edilmektedir.

Tüm bu bilgiler Tablo 16.6'da gösterilen **stsum** komutu kullanılarak doğrudan elde edilebilir.

Tablo 16.6. Yaşam Süresine Ait İstatistikler

stsum

```
failure _d: sansur == 1
analysis time _t: evl_yas
```

	incidence time at risk	no. of rate	no. of subjects	Survival time		
				25%	50%	75%
total	26101	.0361672	1250	18	21	25

Olay oranı $944/26101=0.0361672$ 'dir. Çalışmanın başlangıcından anketin yapıldığı tarihe kadar medyan yaşam süresi 21 yıldır. 25. yüzdelik dilim 18 yıldır ve kadınların %25'inin yaşam süresi (evlilik yaşı) 18 yıldan azdır. Kadınların %25'i 18, %50'si 21 ve %75'i 25 yaşında evlenmiştir. **stdes** komutu ile Tablo 16.7'de gösterildiği gibi özet bilgiler ve medyan yaşam süreleri elde edilebilir

Tablo 16.7. Yaşam Süresine Ait Özet Bilgiler

stdes

```
failure _d: sansur == 1
analysis time _t: evl_yas
```

Category	total	per subject			
		mean	min	median	max
no. of subjects	1250				
no. of records	1250	1	1	1	1
(first) entry time		0	0	0	0
(final) exit time		20.8808	12	20	49
subjects with gap	0				
time on gap if gap	0				
time at risk	26101	20.8808	12	20	49
failures	944	.7552	0	1	1

stsum ve **stdes** komutlarıyla elde edilen medyan yaşam süresi değerleri birbirinden farklıdır. **stdes** ile elde edilen medyan yaşam süresi ham verilerden türetilmektedir ve sansür göz ardı edilmektedir. **stsum**'dan elde edilen Tablo 16.6'da medyan yaşam süresi 21 iken, **stdes** ile elde

edilen Tablo 16.7’de medyan yaşam süresi 20’dir. Yapılan araştırmada sansürlü veriler olduğu için, sansürü de dikkate alan Tablo 16.6’daki medyan yaşam süresi dikkate alınmalıdır.

16.2. Yaşam Analizinde Kullanılan Fonksiyonlar

Dağılımı simetrik olmayan ve sansürlenmiş verilerden oluşan yaşam verilerinin analizinde; yaşam (sağkalım) fonksiyonu, olasılık yoğunluk fonksiyonu ve tehlike (hazard) fonksiyonu olmak üzere birbiriyle ilişkili üç temel fonksiyon kullanılmaktadır.

Yaşam süresini ifade eden T rastlantı değişkeninin, t ’ye eşit ya da daha büyük olma olasılığını veren yaşam fonksiyonu $S(t)$; $f(t)$ ve $F(t)$ verilen bir dağılımın sırasıyla olasılık yoğunluk fonksiyonu ve birikimli dağılım fonksiyonu olmak üzere,

$$F(t) = P(T < t) = \int_0^t f(u) du \quad (16.2)$$

ve

$$S(t) = P(T \geq t) = 1 - F(t) \quad (16.3)$$

biçiminde tanımlanır (Collet, 2003: 11) .

Birikimli hayatta kalma oranı olarak da bilinen yaşam fonksiyonu t zamanının özellikleriyle artmayan bir fonksiyondur:

$$S(t) = \begin{cases} 1, & t = 0 \\ 0, & t = \infty \end{cases} \quad (16.4)$$

t ’nin sıfır olması durumunda hayatta kalma olasılığı 1’dir ve t sonsuza giderken hayatta kalma olasılığı 0’dır (Lee ve Wang, 2003: 8).

Uygulamada sansürlenmiş gözlem yoksa yaşam fonksiyonu, t ’den daha uzun süre hayatta kalan bireylerin oranı olarak tahmin edilir:

$$\hat{S}(t) = \frac{\text{t zamanından daha fazla yaşayan birey sayısı}}{\text{toplam birey sayısı}} \quad (16.5)$$

ve $\hat{S}(t)$, fonksiyonun bir tahminini gösterir (Lee ve Wang, 2003: 9).

Yaşam analizleri uygulamasında T negatif olamaz ve olasılık yoğunluk fonksiyonu,

$$f(t) = \frac{dF(t)}{dt} = -\frac{dS(t)}{dt} \quad (16.6)$$

olmak üzere yaşam analizlerinde ölüm eğrisi olarak da adlandırılır. Tanım olarak $f(t)$ aslında mutlak ani ölüm oranını göstermektedir (Elandt-Johnson, Johnson, 1999: 51). Sansürlenmiş gözlem olmaması durumunda olasılık yoğunluk fonksiyonu,

$$\hat{S}(t) = \frac{\text{t zamanıyla başlayan aralıkta başarısız olan birey sayısı}}{\text{bireylerin toplam sayısı} \times \text{aralık genişliği}} \quad (16.7)$$

olarak tahminlenir.

t zamanında hayatta kaldığı bilinen bir birimin, $[t, t + \delta]$ zaman aralığında yaşamının sona ermesi (başarısızlık) olasılığı tehlike (hazard) fonksiyonu olarak tanımlanır. Özellikle yaşam verisinin modellenmesinde kullanılan tehlike fonksiyonu; güvenilirlikte koşullu başarısızlık oranı, demografide mortalite gücü, stokastik süreçlerde yoğunluk fonksiyonu, ekonomide Mill's oranının tersi olarak da bilinmektedir (Klein ve Moeschberger, 2003: 27).

Birimin t zamanında hayatta kaldığını varsayarak anlık başarısızlık oranını veren tehlike fonksiyonu $\lambda(t)$,

$$\lambda(t) = \lim_{\delta \rightarrow 0} \frac{\Pr(t \leq T \leq t + \delta \mid t \leq T)}{\delta} \quad (16.8)$$

ya da

$$\lambda(t)dt = \Pr(t \leq T \leq t + dt \mid t \leq T) \quad (16.9)$$

olarak tanımlanır (Le, 1997: 8). Tehlike fonksiyonu bir olasılık fonksiyonu olmadığı için zamana göre değişiklik gösterebilir ya da sabit kalabilir (Olça, 2014: 10).

Uygulamada sansürlenmiş gözlemler bulunmadığında ise tehlike fonksiyonu:

$$\hat{\lambda}(t) = \frac{\text{Bu aralıkta birim zaman başına başarısız olan sayısı}}{t \text{ zaman noktasında yaşayan birey sayısı}} \quad (16.10)$$

olarak hesaplanır (Lee ve Wang, 2003: 12).

Olasılık yoğunluk fonksiyonu tanımını kullanarak tehlike fonksiyonu,

$$\lambda(t) = \frac{f(t)}{S(t)} = -\frac{d}{dt} \log S(t) \quad (16.11)$$

şeklinde yazılabilir. $S(0)=1$ 'i kullanarak eşitlik yeniden düzenlenirse yaşam fonksiyonu ile tehlike fonksiyonu arasındaki ilişki,

$$\begin{aligned} S(t) &= \exp \left[- \int_0^t \lambda(u) du \right] \\ &= \exp[-\Lambda(t)] \end{aligned} \quad (16.12)$$

elde edilir ve $\Lambda(t) = \int_0^t \lambda(u) du$ birikimli tehlike fonksiyonu (cumulative hazard function) olarak adlandırılır. Eşitlik (16.12)'den elde edilen $S(t)$ eşitlik (16.11)'de yerine yazılırsa olasılık yoğunluk fonksiyonu, tehlike fonksiyonu cinsinden

$$f(t) = \lambda(t) \exp[-\Lambda(t)] \quad (16.13)$$

olarak elde edilir (Kalbfleisch ve Prentice, 2002: 7). Birikimli tehlike fonksiyonu artan ve sağdan sürekli bir fonksiyon olma ile $\lim \Lambda(t) = \infty$ özelliklerini sağlamaktadır (Jerenz, 2008: 74).

16.3. Yaşam Analizi Yöntemleri

Yaşam analizinin;

- Yaşam verileri ile yaşam ve tehlike fonksiyonlarının tahmin edilmesi ve yorumlanması,
- Yaşam ve tehlike fonksiyonlarının karşılaştırılması
- Açıklayıcı değişkenlerin yaşam süresi ile olan ilişkisinin değerlendirilmesi

olmak üzere 3 temel hedefi bulunmaktadır (Kleinbaum ve Klein, 2005: 15). İlk iki amaca, bireysel yaşam fonksiyonlarının karşılaştırılması yoluyla yaşam verilerini alt gruplara ayırabilmek için ihtiyaç duyulmaktadır (Jerenz, 2008: 72).

Yaşam analizlerinde ilgilenilen durumun özelliğine göre, parametrik olmayan, yarı parametrik ve parametrik analiz yöntemi olmak üzere üç farklı yöntem vardır. Dağılım varsayımlarının uygunluklarının tespiti gerçekleştirilerek yaşam verilerinin dağılımının bilinmesi durumunda parametrik analiz yöntemleri, herhangi bir dağılıma uygunluk göstermeyen yaşam verileri için parametrik olmayan analiz yöntemleri kullanılır (Bilgi, 2009: 14). Bu bölümde parametrik olmayan yaşam analizi yöntemleri incelenecektir.

16.3.1. Parametrik Olmayan Analiz Yöntemleri

Yaşam sürelerine ait uygun teorik dağılımlar bilinmediğinde parametrik analiz yöntemlerine göre daha kullanışlı olan parametrik olmayan analiz yöntemlerinden en önemlileri Yaşam Tabloları Analizi (LT), Kaplan-Meier (KM) yöntemi ve yarı-parametrik Cox regresyon yöntemidir.

16.3.1.1. Yaşam Tablosu (LT) Analizleri

Yaşam süresi verilerini belirli zaman aralıklarına göre frekans tablolarına dönüştürerek analiz eden yaşam tablosu (LT) yöntemi, ilk kez 17. yüzyılda Edmund Halley tarafından bir kentte oturanların yaşam sürelerinin belirlenmesinde kullanılmıştır (Aksakoğlu, 2013: 341). Berkson ile Gage'in 1950'de yaşam fonksiyonunu tahmin etmek için kullandığı yaşam tablosu yöntemi Cutler ile Ederer (1958) tarafından genişletilmiş ve geliştirilmiştir. Gehan (1969) ise çalışmasında yaşam fonksiyonu ile birlikte yoğunluk ve tehlike fonksiyonlarını da tahmin etmek için yaşam tablosu analiz yöntemini geliştirerek kullanmıştır (Tekin, 2001: 27).

Toplam veri sayısının oldukça yüksek olduğu ($n \geq 100$) çalışmalarda kullanılması önerilen LT analizlerinde:

- 1- Belirlenen zaman aralıkları içerisinde ayrılan veya takip edilemeyen izlemlerin, zaman aralığının tam ortasında ayrıldığı,
- 2- Bir zaman aralığındaki yaşam, önceki tüm zaman aralıklarına bağımlı iken o zaman aralığındaki yaşam olasılığı diğerlerinden bağımsız olmak üzere iki varsayımdan faydalanılır (Aksakoğlu, 2013: 344).

Her bir zaman aralığı içinde ayrı ayrı yorumlama yapmayı sağlayan yaşam tablolarında, yaşam fonksiyonunun LT tahmin edicisi; n_j , t_j ile t_{j+1} aralığında risk altındaki birey sayısı, d_j j. aralıkta ölen kişilerin sayısı olmak üzere

$$\hat{S}(t) = \prod_{j=1}^k \frac{n_j' - d_j}{n_j'} \quad t_k' \leq t \leq t_{k+1}' \quad k = 2, 3, \dots, m \quad (16.14)$$

ile elde edilir. $(n_i' - d_i / 2)\tau_i$ aralığında, ortalama yaşam süresi ve τ_i , i. zaman aralığının uzunluğu olmak üzere bu aralıktaki tehlike fonksiyonunun yaşam tablo tahmini

$$\lambda(t) = \frac{d_i}{(n_i' - d_i / 2)\tau_i} \quad (16.15)$$

olarak elde edilir (Collet, 2003: 17-29).

Stata'da evlilik yaşını frekans tablolarına dönüştürerek analiz eden yaşam tablosu için temel olarak **ltable** komutu kullanılır.

Stata Tahmini (ltable)

Komutun temel yapısı aşağıda belirtilmektedir.

ltable timevar [deadvar] [if] [in] [weight] [, options]

timevar modelde kullanılan başarısızlık veya sansürleme zamanını tanımlarken **deadvar** sansürü ifade eden değişkeni tanımlamaktadır.

Tablo 16.8. Evlilik Yaşına Ait Yaşam Tablosu

ltable evl_yas sansur

Interval		Beg. Total	Deaths	Lost	Survival	Std. Error	[95% Conf. Int.]	
12	13	1250	3	0	0.9976	0.0014	0.9926	0.9992
13	14	1247	7	0	0.9920	0.0025	0.9852	0.9957
14	15	1240	18	0	0.9776	0.0042	0.9677	0.9845
15	16	1222	24	36	0.9581	0.0057	0.9454	0.9679
16	17	1162	73	37	0.8969	0.0087	0.8784	0.9128
17	18	1052	78	33	0.8294	0.0109	0.8067	0.8496
18	19	941	116	30	0.7255	0.0131	0.6988	0.7503
19	20	795	103	26	0.6299	0.0144	0.6010	0.6574
20	21	666	87	20	0.5464	0.0150	0.5165	0.5753
21	22	559	104	18	0.4431	0.0152	0.4130	0.4726
22	23	437	60	11	0.3815	0.0150	0.3520	0.4109
23	24	366	60	19	0.3173	0.0146	0.2888	0.3460
24	25	287	42	14	0.2697	0.0141	0.2423	0.2977
25	26	231	51	14	0.2083	0.0133	0.1829	0.2349
26	27	166	22	7	0.1801	0.0128	0.1558	0.2058
27	28	137	25	9	0.1461	0.0120	0.1235	0.1706
28	29	103	16	4	0.1230	0.0114	0.1016	0.1464
29	30	83	14	5	0.1016	0.0108	0.0817	0.1239
30	31	64	5	3	0.0934	0.0105	0.0742	0.1153

31	32	56	11	2	0.0748	0.0098	0.0571	0.0955
32	33	43	5	0	0.0661	0.0094	0.0492	0.0861
33	34	38	5	1	0.0573	0.0089	0.0415	0.0765
34	35	32	3	0	0.0519	0.0086	0.0368	0.0706
35	36	29	4	3	0.0443	0.0082	0.0303	0.0623
36	37	22	3	1	0.0382	0.0078	0.0250	0.0555
37	38	18	2	2	0.0337	0.0075	0.0212	0.0506
39	40	14	0	2	0.0337	0.0075	0.0212	0.0506
41	42	12	0	1	0.0337	0.0075	0.0212	0.0506
42	43	11	3	0	0.0245	0.0071	0.0133	0.0414
44	45	8	0	1	0.0245	0.0071	0.0133	0.0414
46	47	7	0	1	0.0245	0.0071	0.0133	0.0414
47	48	6	0	1	0.0245	0.0071	0.0133	0.0414
48	49	5	0	4	0.0245	0.0071	0.0133	0.0414
49	50	1	0	1	0.0245	0.0071	0.0133	0.0414

Tablo 16.8'de ilk sütun gözlem süresindeki zaman aralıklarını göstermektedir (interval). Zaman aralıkları araştırmacılar tarafından belirlenebilir. Örneğin, 5'er yıllık zaman aralığı oluşturulmak istenirse `ltable evl_yas sansur, survival intervals(5)` komutu zaman aralığı en sona eklenerek kullanılabilir. Her aralığın başında o aralık boyunca izlenen birey sayısı 2. sütunda (Beg. Total) yer alırken, aralıktaki başarısız birey (death) sayısı 3. sütunda yer almaktadır. 4. sütun da aralıktaki sansürlü veri sayısını göstermektedir. Aralığın başlangıcında yaşayan bireyler için tahmin edilen yaşam olasılığı (survival) 5. sütunda, aralığa ait standart hatalar 6. sütunda ve %95 güven aralıkları son iki sütunda yer almaktadır. Tabloyu örnek olarak sadece bir gruba göre yorumlayacak olursak; 15-16 yaş aralığında izlenen 1222 kadın yer almaktadır. Belirtilen aralıkta evlenerek başarısız olan kadın sayısı 24 ve sansürlü gözlem sayısı 36'dır. Aralığın başlangıcında yaşayan bireyler için tahmin edilen yaşam olasılığı ise 0.9581'dir.

Değişkenlerin düzeylerine göre yaşam tablosunda birikimli başarısızlık ve tehlike oranları da elde edilebilir. Üç tabloyu da bir arada görmek için;

`ltable evl_yas sansur, survival failure hazard`

komutu kullanılabilir.

Sigortacılıkta, tıbbi ve medikal araştırmalarda, demografide, nüfus hareketlerinde kullanılan yaşam tabloları, yaşam süresi verilerini zaman aralıklarına göre tablolastırarak ele almaktadır. Yaşam fonksiyonları bu aralıklara göre hesaplanmaktadır. Yaşam tabloları bize, gözlemde çıkan, gözlemde kaybolan veya takip edilemeyen bireyler (yani sansürlü veri sayısı) hakkında bilgi verdiği gibi, aralıkların başlangıcındaki yaşayan bireyler için tahmin edilen yaşam olasılıklarını da görmemize yardımcı olur.

16.3.1.2. Kaplan Meier (KM) Yöntemi

Küçük ve orta büyüklükteki örneklerle ait hayatta kalma süresiyle ilgili verileri belirli zaman aralıklarına bölmeden yaşam ve ölüm fonksiyonunu tahmin etmek için kullanılan KM yöntemi, 1958 yılında Kaplan ve Meier tarafından geliştirilmiştir. Sansürlenmiş veriyi göz ardı etmemesi KM yönteminin en önemli özelliğidir.

N kişinin başarısızlık zamanları $t_1, t_2, \dots, t_n$ ve bu başarısızlık zamanları arasındaki ilişki $t_1 \leq t_2 \leq \dots \leq t_n$ olarak tanımlandığında, t_i 'ye gelmeden hemen önce başarısız olan birey sayısı n_i , t_i zamanındaki başarısızlık sayısı d_i olmak üzere, yaşam fonksiyonunun KM tahmin edicisi,

$$\hat{S}(t) = \prod_{t_i \leq t} \frac{n_i'}{n_i} \quad n_i' = n_i - d_i \quad (16.16)$$

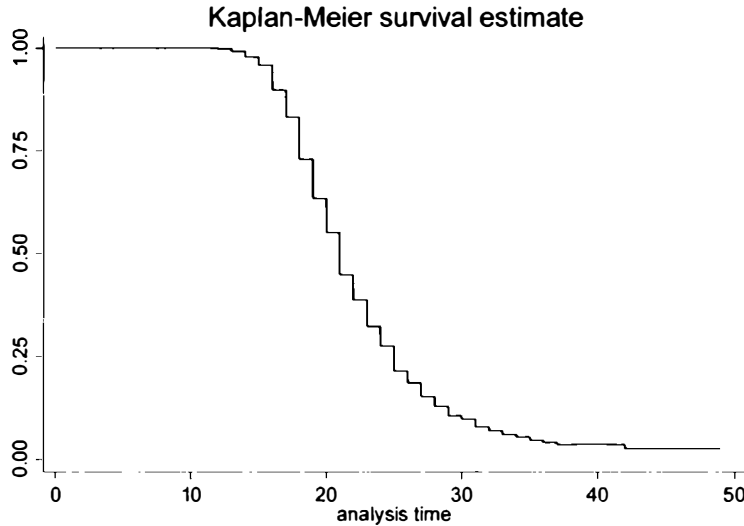
şeklinde formüle edilmektedir (Kaplan ve Meier, 1958: 463). Sansürlenmiş gözlem olmadığı zaman, n_i , t_i zamanında hayatta olanların sayısına eşit olur. KM yaşam fonksiyonu tahmini, her başarısızlık zamanında azalan bir basamak fonksiyonu olmaktadır ve sadece başarısızlıklar üzerinden yaşam olasılığı hesaplanmaktadır (İnceoğlu, 2013: 49). Zaman aralıklarına bölünmemiş yaşam verisi için tehlike fonksiyonu; bir başarısızlık zamanındaki başarısızlık sayısının, o zaman noktasında riskli olan birey sayısına oranı ile de tahminlenmektedir (Parlak, 2004: 11).

Parametrik olmayan analiz yöntemlerinde dikkat edilmesi gereken konu, yaşam fonksiyonu $0 < S(t) < 1$ aralığında yer alan bir olasılık fonksiyonu iken, tehlike fonksiyonu $0 < \lambda(t) < \infty$ aralığında yer alan bir oran fonksiyonudur (Çekçeki, 2007: 16).

Aşağıda evlilik yaşı ve açıklayıcı değişkenleri için KM yaşam olasılıklarına ait grafiklerin elde edilmesinde kullanılan komutlar ve yorumları verilmiştir.

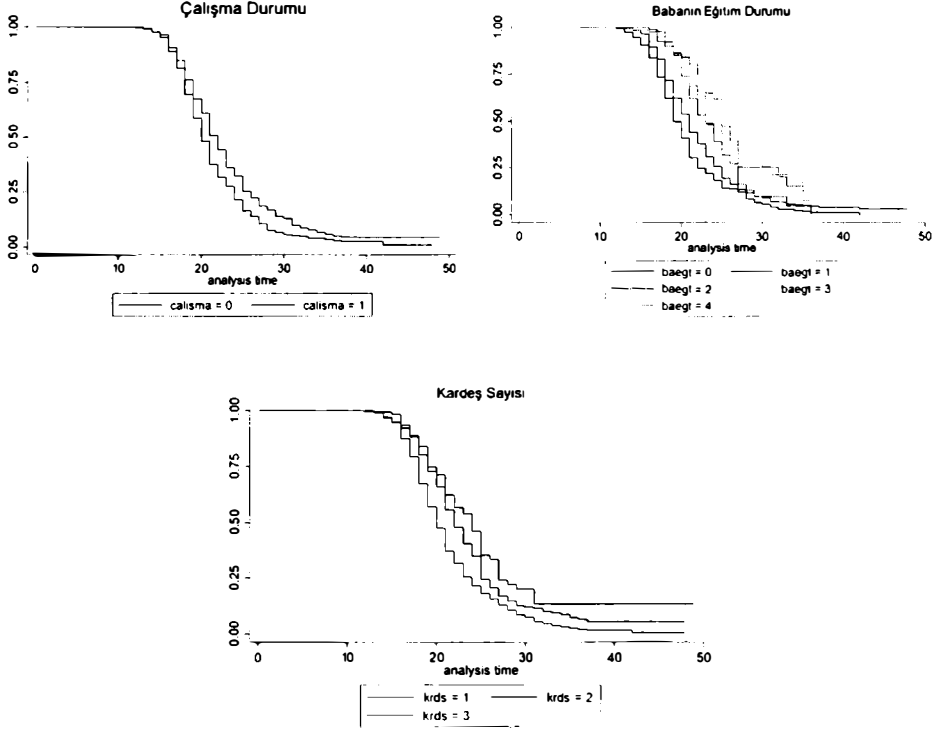
Şekil 16.2. Evlilik Yaşına Ait KM Yaşam Olasılığı

sts graph



Şekil 16.3. Değişkenlere Göre Evlilik Yaşına Ait KM Yaşam Olasılıkları

```
sts graph, by(calisma) title(Çalışma Durumu)
sts graph, by( baegt ) title(Babanın Eğitim Durumu)
sts graph, by( krds ) title(Kardeş Sayısı)
```



Şekil 16.3'te her bir açıklayıcı değişken sırasıyla dikkate alındığında elde edilen KM eğrileri; 12 yaşından anketin yapıldığı tarihe kadar (2013) gelir getirsin veya getirmesin en az 6 ay süreyle bir işte çalışan, babası yükseköğrenim görmüş ve 1 ya da 2 kardeşi olan kadınların evlilik yaşının daha geç olduğunu göstermektedir. Herhangi bir işte çalışmayan kadınların medyan evlilik yaşı 20'dir. Babası ortaokul ve lise mezunu olan kadınların medyan evlilik yaşı aynıdır. 5 ve daha fazla kardeşi olan kadınların medyan evlilik yaşı 20'dir.

KM yöntemiyle elde edilen iki gruba ait yaşam eğrilerinin karşılaştırılmasında Log-rank testi ve Breslow-Wilcoxon testi yaygın olarak kullanılmaktadır ve skorlamaya dayanmaktadır.

Parametrik olmayan Log-rank testinde sıfır hipotezi, iki gruptaki bireylere ait yaşam fonksiyonları arasında fark olmadığını göstermektedir. Testte, i.grup için başarısız olan birimlerin, gözlenen sayısı ile beklenen sayısı arasındaki farklılık araştırılmaktadır. Ağırlıklar tüm zaman sürecinde aynıdır ve tehlike oranları birbiri ile oransal olduğu zaman test optimum güce sahip olmaktadır (İnceoğlu, 2013: 55). Eğer O_1 ve O_2 iki grupta gözlenen başarısızlık sayıları, E_1 ve E_2 beklenen başarısızlık sayıları ise karşılaştırma için;

$$T = \frac{(O_1 - E_1)^2}{E_1} + \frac{(O_2 - E_2)^2}{E_2} \quad (16.17)$$

istatistiği kullanılır ve T, 1 serbestlik dereceli χ^2 dağılışı gösterir (Özdemir, 1994: 20).

İki ve ya daha fazla yaşam dağılımını karşılaştırmak için kullanılan Breslow-Wilcoxon testinde her bir başarısızlık zamanında risk altındaki birim sayılarını ağırlık alarak yaşam dağılımları karşılaştırılır. Breslow-Wilcoxon testi tehlike hızı sabit olmadığında kullanılabilen güçlü bir testtir. Logrank testi erken ya da geç başarısızlığa aynı önemi verirken, Wilcoxon testleri erken başarısızlığa daha fazla önem vermektedir (İnceoğlu, 2013: 55).

Yaşam eğrilerinin karşılaştırılmasında kullanılan Logrank ve Wilcoxon testleri için Stata'da aşağıdaki komutlar kullanılır:

```
sts test değişken, logrank
sts test değişken, wilcoxon
```

Tablo 16.9. Kadının Çalışma Durumu İçin Logrank ve Wilcoxon Test Sonuçları

```
sts test calisma, logrank
```

```
      failure _d:  sansur == 1
      analysis time _t:  evl_yas
```

Log-rank test for equality of survivor functions

calisma	Events observed	Events expected
0	427	364.42
1	517	579.58
Total	944	944.00

chi2(1) =	20.34
Pr>chi2 =	0.0000

```
sts test calisma, wilcoxon
```

```
      failure _d:  sansur == 1
      analysis time _t:  evl_yas
```

Wilcoxon (Breslow) test for equality of survivor functions

calisma	Events observed	Events expected	Sum of ranks
0	427	364.42	43636
1	517	579.58	-43636
Total	944	944.00	0

chi2(1) =	16.21
Pr>chi2 =	0.0001

Tablo 16.9’da yer alan Log-rank ve Wilcoxon test istatistikleri incelendiğinde, yaşam süresi yönünden evlilik yaşı ile kadının 12 yaşından anketin yapıldığı tarihe kadar (2013) gelir getirsin veya getirmesin en az 6 ay süreyle bir işte çalışıp çalışmaması arasındaki fark, her iki testte de istatistiksel olarak anlamlı bulunmuştur (20.34 ve 16.21; $p=0.0000$ ve $p=0.0001$). Çalışan kadınlar ile çalışmayan kadınlar arasında evlilik yaşı açısından fark vardır.

Tablo 16.10. Babanın Eğitim Durumu İçin Logrank ve Wilcoxon Test Sonuçları

sts test baegt , logrank

```
failure _d: sansur == 1
analysis time _t: evl_yas
```

Log-rank test for equality of survivor functions

baegt	Events observed	Events expected
0	256	190.89
1	561	545.25
2	60	83.95
3	41	75.99
4	26	47.92
Total	944	944.00
chi2(4) =		64.20
Pr>chi2 =		0.0000

sts test baegt , wilcoxon

```
failure _d: sansur == 1
analysis time _t: evl_yas
```

Wilcoxon (Breslow) test for equality of survivor functions

baegt	Events observed	Events expected	Sum of ranks
0	256	190.89	57267
1	561	545.25	11208
2	60	83.95	-25863
3	41	75.99	-26201
4	26	47.92	-16411
Total	944	944.00	0
chi2(4) =		82.72	
Pr>chi2 =		0.0000	

Tablo 16.10'daki Log-rank ve Wilcoxon test istatistikleri incelendiğinde, hesaplanan iki test içinde babanın eğitim seviyeleri arasında yaşam sürelerinin dağılımı açısından sıfır hipotezi reddedilerek anlamlı farklılık bulunmuştur (64.20 ve 82.72; $p=0.0000$ ve $p=0.0000$). Babanın eğitim seviyesinin tüm düzeyleri arasında kadının evlilik yaşı açısından fark vardır.

Tablo 16.11. Sahip Olunan Kardeş Sayısı İçin Logrank ve Wilcoxon Test Sonuçları

sts test krds , logrank

```
failure _d: sansur == 1
analysis time _t: evl_yas
Log-rank test for equality of survivor functions
```

krds	Events observed	Events expected
1	63	95.59
2	270	328.56
3	611	519.85
Total	944	944.00

```
chi2(2) = 43.22
Pr>chi2 = 0.0000
```

sts test krds ,wilcoxon

```
failure _d: sansur == 1
analysis time _t: evl_yas
Wilcoxon (Breslow) test for equality of survivor functions
```

krds	Events observed	Events expected	Sum of ranks
1	63	95.59	-20895
2	270	328.56	-52077
3	611	519.85	72972
Total	944	944.00	0

```
chi2(2) = 46.03
Pr>chi2 = 0.0000
```

Tablo 16.11'de yer alan Log-rank ve Wilcoxon test istatistikleri incelendiğinde, sahip olunan kardeş sayıları arasındaki fark her iki testte de istatistiksel olarak anlamlıdır (43.22 ve 46.03; $p=0.0000$ ve $p=0.0000$). Kadınların sahip olduğu kardeş sayılarının düzeyleri arasında evlilik yaşı açısından fark bulunmuştur.

16.3.1.3. Cox Regresyon Modeli

Yaşam süresinde belirli bir dağılım tanımlanmadan, tehlike oranı ile değişkenler arasında parametrik ilişki içeren yaşam verisinin çözümlemesinde kullanılan yöntemler, yarı parametrik yöntemlerdir.

Yaşam analizlerinde en çok kullanılan ve yaşam süresi ile açıklayıcı değişkenler arasındaki sebep-sonuç ilişkisini, yaşam süresinde belirli bir dağılım tanımlamadan açıklayan orantılı tehlike-ler regresyon (proportional hazards regression) modeli 1927 yılında David R. Cox tarafından geliştirilen yarı parametrik bir yöntemdir ve Cox regresyon modeli olarak bilinmektedir.

Cox regresyon modelinde, başlangıç tehlike fonksiyonu regresyonun parametrik olmayan kısmını oluştururken, β katsayısı parametrik kısmı oluşturmaktadır. Bu yüzden yarı-parametrik bir model olarak tanımlanmaktadır (Kleinbaum ve Klein, 2005: 261).

Daha genel bir kullanıma sahip olan Cox regresyon modeli başarısızlık risklerinin belirlenmesi açısından önemli bir istatistiksel yöntemdir. Yaşam süresini etkilediği düşünülen açıklayıcı değişkenler modeli çarpımsal olarak etkilemektedir ve değişkenlerin etkileri eşanlı olarak açıklanmaktadır.

Cox regresyon modeli eş zamanlı gözlem olmadığı varsayımı altında,

z : Açıklayıcı değişkenler (eş değişken, kovaryant) vektörü

β : $p \times 1$ boyutlu bilinmeyen regresyon katsayıları vektörü

$\lambda_0(t)$: formu hakkında herhangi bir varsayım yapılmadan $z=0$ olan bir birimin temel tehlike fonksiyonu (baseline hazard function) olmak üzere;

$$\lambda(t; z) = \exp(z\beta)\lambda_0(t) \quad (16.18)$$

biçiminde ifade edilir (Cox, 1972: 189). Tehlike fonksiyonu, $\lambda_0(t)$ ve $\exp(z\beta)$ olmak üzere iki fonksiyondan oluşmaktadır. $\lambda_0(t)$ fonksiyonu, tehlike fonksiyonunun yaşam süresinin bir fonksiyonu olarak nasıl değiştiğini ifade ederken, diğer fonksiyon olan $\exp(z\beta)$ açıklayıcı değişkenlerin tehlike fonksiyonunu nasıl etkilediğini göstermektedir (Hosmer ve Lemeshow, 1999: 90).

Doğrusal bileşenleri sabit terim içermeyen Cox regresyon modeli,

$$\lambda_i(t) = \exp(\beta_1 z_{1i} + \beta_2 z_{2i} + \dots + \beta_p z_{pi}) \lambda_0(t) \quad (16.19)$$

olarak yeniden yazılırsa, tehlike oranının logaritması için doğrusal model elde edilir (Cox, 1972: 189)

z verildiğinde T için yaşam fonksiyonu,

$$S(t, z) = \exp \left\{ - \int_0^t \lambda_0(u) \exp[z\beta] du \right\} \quad (16.20)$$

ve yoğunluk fonksiyonu

$$f(t, z) = \lambda(t; z) s(t, z) \quad (16.21)$$

olarak elde edilir (Kalbfleisch ve Prentice, 2002: 96).

Her bir başarısızlık zamanında eş zamanlı gözlem olmadığı varsayımı altında bir bireyin başarısız olduğu varsayılın. $t_{(i)}$, i . sıradaki başarısızlık zamanı olmak üzere k tane sıralı başarısızlık zamanı $t_1 < t_2 < \dots < t_k$ olmak üzere, $t_{(i)}$ zamanında risk altında kalan bireylerin kümesi (risk kümesi) $R(t_{(i)})$ ile tanımlansın. Buna göre risk oranı

$$\exp\{z_i\beta\} / \sum_{l \in R(t_{(i)})} \exp\{z_l\beta\} \quad (16.22)$$

biçiminde elde edilir (Cox, 1972: 191).

k , birbirinden farklı başarısızlık süreleri sayısı olarak alındığında, Cox regresyon modeli için olabilirlik fonksiyonu aşağıdaki gibi tanımlanır (Kalbfleisch ve Prentice, 2002: 132):

$$L(\beta) = \prod_{i=1}^k \frac{\exp(z_i\beta)}{\sum_{l \in R(t(i))} \exp\{z_l\beta\}} \quad (16.23)$$

Eşitlik (16.23)'ün maksimize edilmesiyle açıklayıcı değişkenlerin bilinmeyen katsayılarına ait tahminler gerçekleştirilir.

k yaşam süresinin gözlemlenmesinden oluşan verilerin, sansüre sahip bireyleri içerdiği varsayılın:

$$\delta_i = \begin{cases} 1, & i.\text{birey sağdan sansürlü ise} \\ 0, & \text{değilse} \end{cases} \quad (16.24)$$

olmak üzere olabilirlik fonksiyonu,

$$L(\beta) = \prod_{i=1}^k \left\{ \frac{\exp(z_i\beta)}{\sum_{l \in R(t(i))} \exp\{z_l\beta\}} \right\}^{\delta_i} \quad (16.25)$$

şeklinde dir. Eşitlik (16.25)'in logaritması alınarak log-olabilirlik fonksiyonu,

$$\log L(\beta) = \sum_{i=1}^k \delta_i \left\{ \beta' z_i - \log \sum_{l \in R(t_{(i)})} \exp(\beta' z_l) \right\} \quad (16.26)$$

tanımlanır. Parametre tahmincisi β , $L(\beta)$ 'nın log fonksiyonu maksimize edilerek elde edilir ve böylece parametre tahminleri $\frac{d \log L(\beta)}{d\beta} = 0$ olmak üzere analitik yöntemler yerine, nümerik (sayısal) yöntemler kullanılarak çözülür (Hidayat ve diğerleri, 2014: 202).

Açıklayıcı değişkenlerin, bağımlı değişken olan yaşam süresi üzerindeki etkilerinin araştırıldığı Cox regresyon modelinin kullanılabilmesi için orantılı tehlikeler varsayımının sağlanması gerekmektedir. Orantılı tehlikeler varsayımı, tehlike fonksiyonunda yer alan kovaryantların (açıklayıcı değişken) zamandan bağımsız olması yani farklı regresyon kovaryantlarına sahip iki birey ya da grup için tehlike fonksiyonu oranının zamanla değişmemesi olarak tanımlanmaktadır (Wilson, 2013: 1). İki farklı grubu karşılaştırmak için; z_1 ilk birime ait açıklayıcı değişkenler vektörü, z_2 ikinci birime ait açıklayıcı değişkenler vektörü olmak üzere tehlike oranı

$$HR = \frac{\lambda(t, z_1)}{\lambda(t, z_2)} = \frac{\lambda_0(t) \exp(\beta z_1)}{\lambda_0(t) \exp(\beta z_2)} = \frac{\exp(\beta z_1)}{\exp(\beta z_2)} = \exp[\beta(z_1 - z_2)] \quad (16.27)$$

şeklinde ve varsayım, farklı açıklayıcı değişken setleri ile tanımlanan iki birim için tehlike oranının yalnızca açıklayıcı değişkenlerin değerlerine bağlı olduğunu ve zamana bağlı olmadığını göstermektedir. (Borucka, 2014: 86). Tehlike oranının 1'den büyük olması olayın meydana gelme ihtimalinin yüksek olduğu anlamına gelirken, 1'den küçük olması olayın meydana gelme ihtimalinin düşük, tehlike oranının bire eşit olması durumlar arasında farklılık olmadığını ifade etmektedir. Tehlike oranının zamanla sabit kalmaması durumunda ise değişkenin zamanla değişen bir etkisi olduğunu göstermektedir (Bellera ve diğerleri, 2010: 3; George ve diğerleri, 2014: 689).

β katsayılarının,

$$H_0: \beta_1 = \beta_2 = \dots = \beta_k = 0$$

$$H_0: \beta_1 \neq \beta_2 \neq \dots \neq \beta_k \neq 0 \quad (16.28)$$

hipotezlerini test etmek için

- En çok olabilirlik tahmin edicilerinin normal olduğu varsayımına dayanan Wald Testi,
- Kategorik değişkenlerin iki ve daha fazla düzeyinin olduğu ve modelde birden çok açıklayıcı değişken bulunduğu kullanılan Olabilirlik Oran Test İstatistiği,
- Logaritmik benzerlik istatistiklerinden yararlanılarak hesaplanan Skor Testi

öne çıkmaktadır (Özdamar, 2013: 477).

n toplam gözlem sayısı, p bilinmeyen parametre sayısı, k önceden belirlenmiş sabit (2 olarak alınmaktadır) ve L modelin olabilirlik fonksiyonunu göstermek üzere yaşam modellerinde yaşam verisine ait en uygun olan modele karar verebilmek için aşağıdaki Akaike bilgi kriteri (AIC) ve Bayesci bilgi kriteri (BIC) kullanılmaktadır (Klein ve Moeschberger, 2003: 277).

$$AIC = -2 \log L + kp \quad (16.29)$$

$$BIC = 2 \log L + p \ln(n)$$

Amacı, yaşam süresine etki eden faktörleri belirlemek olan Cox regresyon modelinin en önemli varsayımı tehlike oranlarının zaman boyunca değişmeyerek sabit kalmasıdır (Terzi ve diğerleri, 2005: 771). Fakat yaşam verilerine ait süre uzun olduğunda, tehlike oranı zamanla değişerek sabit kalmadığı için orantısız tehlikelere neden olmaktadır (Ata, 2010: 22). Orantılı tehlikeler varsayımı, log-log yaşam eğrileri, Arjas grafikleri, Schoenfeld artıkları ile yaşam süresinin rankı arasındaki korelasyon testi ve en çok kullanılan zamana bağlı değişkenlerin modele eklenmesiyle araştırılabilir (Karasoy ve diğerleri, 2015: 62). Bir faktör için tehlikelerin oransallığı varsayımı sağlanamıyorsa aşağıdaki analiz seçenekleri kullanılabilir (Schemper, 1992: 456):

- Oransal olmayan tehlikelerle modelin bir değişken ile tabakalandırılması,
- Bölünmüş zaman dilimleri için ayrı modeller kullanılması,
- Modelde zamana bağlı eş değişken terimlerinin kullanılması.

Cox regresyon modelinin tahminlenmesi için Stata komutu olarak `stcox` kullanılmaktadır. Komutun temel yapısı aşağıda belirtilmektedir.

stcox [varlist] [if] [in] [, options]

varlist modelde kullanılan değişkenleri tanımlamaktadır. Komut yapısı hakkında detaylı bilgi için **help stcox** komutu yardımıyla açılan bilgilendirme ekranından detaylara ulaşabilirsiniz. Başarısızlığı (kadının evlenmesi) etkileyen faktörlerin belirlenmesi için tüm değişkenler dahil edilerek çoklu Cox regresyon modeli Stata’da Tablo 16.12’deki gibi elde edilir. Burada Cox regresyon modeline ait katsayıları görmek için komutun sonunda yer alan options kısmına **nohr** eklenmelidir.

Tablo 16.12. Cox Regresyon Modeline Ait Sonuçlar

```
stcox calisma egt_ilkokul egt_ortaokul egt_lise egt_yuksek krds2 krds3 , nohr
```

```
failure _d: sansur == 1
analysis time _t: evl_yas
```

```
Iteration 0: log likelihood = -5887.6129
Iteration 1: log likelihood = -5849.5483
Iteration 2: log likelihood = -5848.6833
Iteration 3: log likelihood = -5848.6815
Refining estimates:
Iteration 0: log likelihood = -5848.6815
```

Cox regression -- Breslow method for ties

No. of subjects =	1250	Number of obs =	1250
No. of failures =	944		
Time at risk =	26101		
		LR chi2(7) =	77.86
Log likelihood =	-5848.6815	Prob > chi2 =	0.0000

	_t	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
calisma		-.1781119	.0673757	-2.64	0.008	-.3101658 -.046058
egt_ilkokul		-.1719584	.0792379	-2.17	0.030	-.3272618 -.016655
egt_ortaokul		-.4878425	.1488341	-3.28	0.001	-.7795519 -.196133
egt_lise		-.7196831	.1756003	-4.10	0.000	-1.063853 -.3755128
egt_yuksek		-.6585859	.2154931	-3.06	0.002	-1.080945 -.2362271
krds2		.1591006	.1411436	1.13	0.260	-.1175357 .4357369
krds3		.3385248	.1382027	2.45	0.014	.0676526 .6093971

Stata, Cox regresyon modelini tahminlerken otomatik olarak Breslow yöntemini kullanır. Eş zamanlı gözlemlerin olduğu durumlarda Cox regresyon modelleri için Efron (1977), Breslow (1974) veya Kesin (exact) yaklaşımları da kullanılabilir (Komutun sonuna hangi yaklaşım kullanılmak isteniyorsa onun ismi yazılır).

$\hat{\beta}$ parametresinin pozitif değerli elde edilmesi katsayıların referans sınıfa göre daha fazla riskli (zamansal olarak daha kısa sürede olayın gerçekleşmesi durumu), negatif değerli elde edilmesi,

referans sınıfa göre daha az riskli (zamansal olarak daha uzun sürede olayın gerçekleşmesi durumu) olduğu durumu ifade etmektedir. Evlilik yaşı için, araştırılan değişkenin katsayısının 1'den büyük olması, o özelliğe sahip bireylerin daha erken sürede evlendikleri ya da evlilik kararı aldıkları söylenebilir. Eğer katsayı 1'den küçük ise, kadının evlenme olasılığı referans sınıfa göre azalmakta ve evlenmek için geçen süre artmaktadır. $\hat{\beta}$ parametresinin değerinin 1 olması durumunda ise açıklayıcı değişkenin etkisinin olmadığı söylenir. Tablo 16.12 incelendiğinde, kadının 3-4 kardeşe sahip olması dışında tüm değişkenlerin ilk evlilik yaşını etkileyen önemli risk faktörleri olduğu %95 güven düzeyinde söylenebilir. Gelir getirsin veya getirmesin herhangi bir işte çalışan kadınlar daha az erken evlenme riskine sahiptir. Kadının babasının eğitim seviyesi arttıkça erken evlenme riski azalırken, kardeş sayısı arttıkça risk artmaktadır.

Tehlike oranlarının (göreceli riskler) elde edilerek yorumlanabilmesi için komutun sonunda yer alan **nohr** kaldırılmalı ve modelde değişen varyans problemine karşı dirençli standart hataları üretmek için **robust** eklenmelidir.

Tablo 16.13. Cox Regresyon Modelinde Tehlike Oranları

```
stcox calisma egt_ilkokul egt_ortaokul egt_lise egt_yuksek krds2 krds3, robust
```

```
failure _d:  sansur == 1
analysis time _t:  evl_yas
```

```
Iteration 0:  log pseudolikelihood = -5887.6129
Iteration 1:  log pseudolikelihood = -5849.5483
Iteration 2:  log pseudolikelihood = -5848.6833
Iteration 3:  log pseudolikelihood = -5848.6815
Refining estimates:
Iteration 0:  log pseudolikelihood = -5848.6815
```

```
Cox regression -- Breslow method for ties
```

```
No. of subjects      =          1250          Number of obs      =          1250
No. of failures      =           944
Time at risk         =          26101

                                Wald chi2(7)      =          94.54
Log pseudolikelihood = -5848.6815              Prob > chi2       =          0.0000
```

		Robust				
	_t	Haz. Ratio	Std. Err.	z	P> z	[95% Conf. Interval]
calisma		.8368488	.0543363	-2.74	0.006	.7368495 .9504191
egt_ilkokul		.8420142	.066775	-2.17	0.030	.7208018 .9836101
egt_ortaokul		.6139496	.0758003	-3.95	0.000	.4819927 .7820328
egt_lise		.4869065	.0739767	-4.74	0.000	.3615104 .6557985
egt_yuksek		.5175827	.0866473	-3.93	0.000	.3728054 .7185837
krds2		1.172456	.1510746	1.23	0.217	.9107864 1.509303
krds3		1.402877	.1803307	2.63	0.008	1.090443 1.804828

Tehlike oranının ($\exp(\beta)$) değeri incelenen düzeyin, referans sınıfa göre kaç kat (ya da % ne kadar) daha riskli olduğunu göstermektedir (Ata, 2010: 67). Referans sınıfın tehlike oranı daima 1'dir.

Tablo 16.13'e göre kadının çalışma durumuna ait kategorik değişken için tahmin edilen tehlike oranı 0.84'tür. Diğer tüm değişkenler sabit tutulduğunda, kadının gelir getirsin ya da getirmesin bir işte çalışıyor olması durumunda, kadınların %16'sı (0.84-1) evlilik kararını ertelemektedir. Kadının babasının eğitim seviyesi arttıkça erken evlenme riski azalmaktadır. Referans sınıfa göre, $\exp(\beta) < 1$ olması hâlinde, $1/\exp(\beta)$ değeri bulunup referans sınıf da benzer şekilde tersten yorumlanabilir (Şimşek, 2013: 53). Buna göre; babası hiçbir eğitim almamış olan kadın, babası ilköğretim eğitimi almış olanlara göre 1.19 (1/0.8420), babası ortaokul eğitimi almış olanlara göre 1.63 (1/0.6139), babası lise eğitimi almış olanlara göre 2.05 (1/0.4869) ve babası yükseköğrenim görmüş olanlara göre 1.93 (1/0.5176) kat daha fazla erken evlenme riskine sahiptir. 5 ya da daha fazla kardeşe sahip olan kadınlar, 1 ya da 2 kardeşi olan kadınlara göre 1.40 kat (%40.29) daha fazla erken evlenme riskine sahiptir.

Cox regresyon analizinde herhangi bir dağılım varsayımının olmaması, kolay uygulanabilir olmasına rağmen, orantılı tehlikeler varsayımı önemlidir ve bu nedenle varsayımın incelenmesi gerekmektedir. Orantılı tehlikeler varsayımı, değişken düzeyleri (kategoriler) fazla olduğunda yaşam eğrileri yerine Schoenfeld artıkları ile yaşam sürelerinin rankı arasındaki korelasyon testi kullanılarak incelenebilir. "Orantılı tehlike varsayımı sağlanmaktadır" biçiminde kurulan sıfır hipotezi ile gerçekleştirilen teste göre, varsayımın sağlanabilmesi için korelasyonun sıfıra yakın olması istenmektedir (Ata ve diğerleri, 2007: 62-68). Schoenfeld artıkları ile yaşam sürelerinin rankı arasındaki korelasyon testi için Stata komutu olarak **estat phtest** kullanılmaktadır. Komutun temel yapısı aşağıda belirtilmektedir:

estat phtest, [, phtest_options]

Değişkenlerin tüm kategorilerine ait test sonuçlarını görmek için komutun sonunda yer alan options kısmına **detail** eklenmelidir.

Tablo 16.14. Schoenfeld Artıkları Orantılı Tehlike Testi Sonuçları

estat phtest, detail

Test of proportional-hazards assumption

Time: Time

	rho	chi2	df	Prob>chi2
calisma	0.00233	0.00	1	0.9441
egt_ilkokul	0.04970	2.46	1	0.1168
egt_ortaokul	0.11027	8.68	1	0.0032
egt_lise	0.08047	4.92	1	0.0266
egt_yukse	0.09694	5.94	1	0.0148
krds2	0.02851	0.65	1	0.4192
krds3	0.01838	0.29	1	0.5901
global test		17.25	7	0.0158

Tablo 16.14'teki Schoenfeld artıkları ile yaşam sürelerinin rankı arasındaki korelasyon testine ait sonuçlar incelendiğinde; $p < 0.05$ olan katsayılar için orantılı tehlikeler varsayımı sağlanamamaktadır. Modelin tamamı için de Schoenfeld artıklarının anlamlılığının $p < 0.05$ ($p = 0.0158$) olması nedeniyle orantılı tehlike (hazard) varsayımının sağlanamadığı görülmektedir.

Temel varsayımın sağlanamadığı durumunda Cox regresyon modeline ait sonuçları kullanmak yanıltıcı olacaktır. Bu nedenle, zamana bağlı açıklayıcı değişkenli (genişletilmiş) Cox regresyon modeli ya da tabakalandırılmış Cox regresyon modeli çözümlenmesi yapılabileceği gibi parametrik modellerin kullanılması daha uygun olabilir. Orantısız tehlikelere sahip değişkenlerin zamanın bir fonksiyonu ile çarpılarak modele dâhil edilmesiyle elde edilen genişletilmiş Cox regresyon modeli, varsayımı sağlamayan birden fazla değişken olduğunda ve zamanın fonksiyonunu belirlemek için kesin bir bilgi olmadığında kullanılamamaktadır (Hosta, 2016: 62). Tabakalandırılmış Cox regresyon modeli ise varsayımı sağlamayan değişkene göre tabakalandırma yapılarak uygulanmaktadır. Orantısız tehlikelere sahip değişkenler ile çok fazla tabakalandırma yapılması bilgi kaybına neden olabileceği için dikkatli olunmalıdır.

Araştırılan konunun belirlenen zaman aralığında başlamasından ve izlem sürecinde sonuca ulaşmasına kadar geçen süre ile ilgilenen yaşam analizlerinin, zaman ve olay olmak üzere iki unsuru bulunmaktadır. Akdeniz bölgesinde yaşayan kadınların ilk evlilik yaşını etkileyen üç faktörün incelenmesinde, kadınların ilk evlilik yaşına kadar geçen zaman, yaşam süresi olarak tanımlanmıştır. Kadınların medeni durumunun evli, boşanmış, dul ve eşiyle ayrı yaşamaya başlamış olması başarısızlık olarak belirtilmişken, anketin yapıldığı zamana kadar hâlâ bekâr olan gözlemler sansürlü olarak tanımlanmıştır. Evlilik yaşını etkileyen faktörleri araştırırken yaşam analizi yöntemlerinin en önemli özelliği, medeni durumu bekar olan kadınları da hesaplamalara dahil etmesidir. Çalışmanın başlangıcından anketin yapıldığı tarihe kadar (2013) medyan yaşam süresi 21 yıldır. Yaşam süresi verilerini belirli zaman aralıklarına göre frekans tablolarına dönüştürerek analiz eden yaşam tablosu yöntemine göre, kadınların en çok 18-19 yaş aralığında evlendiği görülmüştür. Gözlem süresini belirli aralıklara bölmeyen ve tam yaşam olasılığının elde edildiği KM yöntemine göre; 12 yaşından anketin yapıldığı tarihe kadar (2013) gelir getirsin veya getirmesin en az 6 ay süreyle bir işte çalışan, babası yükseköğrenim görmüş ve 1 ya da 2 kardeşi olan kadınlar daha geç evlenmektedir. Log-rank ve Wilcoxon test istatistikleri incelendiğinde, yaşam süresi yönünden evlilik yaşı değişkenlerin her bir kategorisinde farklılık göstermektedir. Cox regresyon analizinde herhangi bir dağılım varsayımının olmaması ve kolay uygulanabilir olmasına rağmen orantılı tehlikeler varsayımı önemlidir. Varsayım, tehlike oranının zamana karşı sabit olmasını ifade etmektedir. Bu nedenle de öncelikle varsayımın incelenmesi gerekmektedir. Modelin tamamı için de Schoenfeld artıkları ile elde edilen test sonucuna göre orantılı tehlike (hazard) varsayımının sağlanamadığı görülmektedir. Fakat burada varsayım göz ardı edilerek yorumlar yapılmıştır.

Stata Komutları

```
use "DOSYAKONUMU\evlilikyaşı1.dta"
```

```
describe
```

```
sum evl_yas sansur calisma egt_yok egt_ilkokul egt_ortaokul egt_lise egt_yuksek krds1  
krds2 krds3
```

```
stset evl_yas, failure(sansur==1)
de _*
su _*
stsum
stdes
ltable evl_yas sansur
ltable evl_yas sansur, survival failure hazard
sts graph
sts graph, by(calisma) title(Çalışma Durumu)
sts graph, by( baegt ) title(Babanın Eğitim Durumu)
sts graph, by( krds ) title(Kardeş Sayısı)
sts test calisma, logrank
sts test calisma, wilcoxon
sts test baegt , logrank
sts test baegt , wilcoxon
sts test krds , logrank
sts test krds ,wilcoxon
stcox calisma egt_ilkokul egt_ortaokul egt_lise egt_yuksek krds2 krds3 , nohr
stcox calisma egt_ilkokul egt_ortaokul egt_lise egt_yuksek krds2 krds3, robust
estat phtest, detail
```

16.4. Kaynaklar

- Enstitüsü, H.Ü.N.E. (2014). 2013 Türkiye Nüfus ve Sağlık Araştırması. *Hacettepe Üniversitesi Nüfus Etütleri Enstitüsü, TC Kalkınma Bakanlığı ve TÜBİTAK, Ankara, Türkiye, 1.*
- A.Özdemir, A. (1994). *Hayatta Kalabilme Analizi Yöntemleri*. Ankara: Ankara Üniversitesi Fen Bilimleri Enstitüsü.
- Akyol, M. (2011). *Yaşam Çözümlemesine Yeni Bir Yaklaşım: MARS*. Ankara: Ankara Üniversitesi Sağlık Bilimleri Enstitüsü.
- Allison, P. D. (2010). Survival Analysis. G. R. Hancock, ve R. O. Mueller içinde, *The Reviewer's Guide to Quantitative Methods in the Social Sciences* (s. 413-424). New York: Routledge.
- Ata, N., ve Sözer, M. T. (2007). Cox Regression Models With Nonproportional Hazards Applied To Lung Cancer Survival Data. *Hacettepe Journal of Mathematics and Statistics*, 36(2): 157-167.
- Ata, N. (2010). *Orantısız Tehlikeler İçin Yaşam Modelleri* (Yayınlanmış Doktora Tezi). Ankara: Hacettepe Üniversitesi Fen Bilimleri Enstitüsü.
- Ata, N., Karasoy, D. S., Sözer, M. T. (2007) Orantılı tehlike varsayımının incelenmesinde kullanılan yöntemler ve bir uygulama. *Eskişehir Osmangazi Üniversitesi Mühendislik Mimarlık Fakültesi Dergisi*, 20(1): 57-80
- Bellera, C. A., MacGrogan, G., Debled, M., Tunon de Lara, C., Brouste, V., ve Mathoulin-Pelissier, S. (2010). Variables with Time-Varying Effects and the Cox model: Some Statistical Concepts

- Illustrated with a Prognostic Factor Study in Breast Cancer. *BMC Medical Research Methodology*. 10(20): 1-12.
- Bilgi, S. (2009). *Sağkalım Analizinde Kullanılan İstatistiksel Yöntemler ve Aktüerya Alanında Bir Uygulama* (Yayınlanmamış Yüksek Lisans Tezi). İzmir: Ege Üniversitesi Fen Bilimleri Enstitüsü.
- Borucka, J. (2014). Extensions of Cox model for Non-proportional Hazards Purpose. *Econometrics (Ekonometria)*. 3(45): 85-101.
- Collet, D. (2003). *Modelling Survival Data in Medical Research* (Cilt 2nd.Edition). Florida: Chapman and Hall/CRC.
- Cox, D. (1972). Regression Models and Life-Tables. *Journal of the Royal Statistical Society*. 34(2): 187-220.
- Danacica, D.-E., ve Babucea, A.-G. (2010). Using Survival Analysis in Economics. http://anale.feaa.uaic.ro/anale/resurse/sta_danacica_babucea.pdf, 439-450.
- Dickman, P. (2014). *Survival Analysis, Overview*. Kasım 23, 2016 tarihinde <http://www.pauldickman.com>: <http://www.pauldickman.com/survival/handouts/>
- George, B., Seals, S., ve Aban, I. (2014). Survival Analysis and Regression Models. *Journal of Nuclear Cardiology*. 21(4): 686-694.
- Çekçeki, Ö. (2007). *Orantılı Hazard Modelinin Zamana Bağlı Değişkenlerle Genişletilmesi ve Çocuk Suçluluğu Üzerine Bir Uygulama* (Yayınlanmamış Yüksek Lisans Tezi). Ankara: Gazi Üniversitesi Fen Bilimleri Enstitüsü .
- Hidayat, R., Sumarno, H., ve Nugrahani, E. H. (2014). Survival Analysis in Modeling the Birth Interval of the First Child in Indonesia. *Open Journal of Statistics*. 4, 198-206.
- Hosmer, D. W., ve Lemeshow, S. (1999). *Applied Survival Analysis: Regression Modelling of Time to Event Data*. USA: A Wiley-Interscience Publication.
- Hutchison, D. (1988). Event History and Survival Analysis in The Social Sciences. *Quality ve Quantity*. 22(3): 255-278.
- İnceoğlu, F. (2013). *Sağkalım Analizi Yöntemleri ve Karaciğer Nakli Verileri İle Bir Uygulama* (Yayınlanmamış Yüksek Lisans Tezi). Malatya: İnönü Üniversitesi Sağlık Bilimleri Enstitüsü.
- Jerenz, A. (2008). *Revenue Management and Survival Analysis in the Automobile Industry*. Germany: Gabler Edition Wissenschaft.
- Justine, N. (2013). *Application of Survival Analysis Methods to Study Under-Five Child Mortality in Uganda*. (S. a. School of Mathematics, Dü.) University Of Kwazulu-Natal Master Thesis.
- Kalbfleisch, J. D., ve Prentice, R. L. (2002). *The Statistical Analysis of Failure Time Data* (Cilt Second Edition). USA: John Wiley ve Sons,Inc.
- Kaplan, E. L., ve Meier, P. (1958). Nonparametric Estimation from Incomplete Observations. *Journal of the American Statistical Association*. 53, 457-481.
- Klein, J. P., ve Moeschberger, M. L. (2003). *Survival Analysis Techniques for Censored and Truncated Data* (Cilt Second Edition). New York: Springer.
- Kleinbaum, D. G., ve Klein, M. (2005). *Survival Analysis A Self-Learning Text* (Cilt Second edition). USA: Springer.
- Le, C. T. (1997). *Applied Survival Analysis* . USA: John Wiley ve Sons,Inc.
- Malacane, L., Murphy, S., ve Collins, L. (1997). *Determining The Risk Set for Hazard Functions when Survival Methods Are Applied to Social Science and Prevention Data*. Pennsylvania State University. University Park, PA:The Methodology Center.

- Olça, K. (2014). *Yaşam Verilerinin Çözümlemesinde Uygun Regresyon Modeli Seçme Yöntemleri* (Yayınlanmamış Yüksek Lisans Tezi). İstanbul: İstanbul Üniversitesi Sağlık Bilimleri Enstitüsü.
- Özdamar, K., (2013). *SPSS ile Biyoistatistik*, 9.Basım. Eskisehir: Kaan kitabevi.
- Parlak, Y. (2004). *Yaşam Analizinde Grup Ardıışık Test Yöntemlerinin Karşılaştırılması*. Ankara: Hacettepe Üniversitesi Fen Bilimleri Enstitüsü.
- Schemper, M. (1992). Cox Analysis of Survival Data with Non-Proportional Hazard Functions. *Journal of the Royal Statistical Society*, 41(4): 455-465.
- Singh, R., ve Mukhopadhyay, K. (2011). Survival analysis in clinical trials: Basics and must know areas. *Perspectives in Clinical Research*, 2(4): 145-148.
- Smith, T., ve Smith, B. (1972). *Survival Analysis And The Application Of Cox's Proportional Hazards Modeling Using SAS*. Haziran 01, 2017 tarihinde [www2.sas.com: http://www2.sas.com/proceedings/sugi26/p244-26.pdf](http://www2.sas.com/proceedings/sugi26/p244-26.pdf)
- Şimşek, G. (2013). *Bireysel Emeklilik Süresini Etkileyen Faktörlerin Yaşam Çözümlemesi İle İncelenmesi*. Ankara: Hacettepe Üniversitesi Aktüarya Bilimleri Bölümü Yüksek Lisans Tezi.
- Tekin, A. Ö. (2001). *Survival Models and An Application*. İzmir: Dokuz Eylül Üniversitesi Fen Bilimleri Enstitüsü.
- Terzi, Y., Cengiz, A., ve Bek, Y. (2005). Cox Regresyon Modelinde Oransal Hazard Varsayımının Artıklarla İncelenmesi ve Akciger Kanseri Hastaları Üzerinde Uygulanması. *Türkiye Klinikleri Journal of Medical Sciences*, 25(6): 770-775.
- Tuncer, N. (2014). *Yaşam Çözümlemesinde Aykırı Değerler*. Ankara: Hacettepe Üniversitesi İstatistik Bölümü Yüksek Lisans Tezi.
- Wilson, M. G. (2013). *Assessing Model Adequacy in Proportional Hazards Regression*. SAS Global Forum.
- Yazıcı Güvercin, A. C., Tekindal, M. A., Kaymaz, Ö., ve Güvercin, C. H. (2017). Comparison of Nonparametric, Semiparametric and Parametric Survival Analysis Methods in Right Censored Medical Data. *Biomedical Research*, 28(10): 4360-4366.

BÖLÜM 17

POLİTİKA ETKİ ANALİZİNDE MİKROEKONOMETRİK YÖNTEMLER VE STA- TA UYGULAMALARI

Eleftherios Giovanis¹

Öznur Özdamar Giovanis²

Ekonometri ya da uygulamalı iktisat alanında *etki analizi* (impact assessment) yöntemleri olarak bilinen, özellikle uygulamadaki politika ve programların etkisinin ölçülmesine imkân tanıyan yöntemler günümüzde sıklıkla kullanılmaktadır. Örneğin kamu politikalarının belli hedef çıktılarını iyileştirmek için uygulandığı düşünüldüğünde, etki analizi yöntemleri belirlenen hedef çıktılara ulaşmada politikanın ne kadar etkin ve başarılı olduğunu ortaya koymak için kullanılan yöntemler bütünüdür. Bu yöntemlerin öncelikli amacı uygulanan politika veya programların *neden-sel etkisini* (causal effect) tahmin etmektir.

İlgili yazında etki analizi yöntemleri **Nitel** ve **Nicel** yöntemler olarak sınıflandırılmakta olup, **Nicel** yöntemler “**Program Sonrası (Ex-Post) Değerlendirme**” ve “**Program Öncesi (Ex-Ante) Değerlendirme**” yöntemleri olarak iki sınıfa ayrılmaktadır. **Program Öncesi Değerlendirme** yöntemleri, *mikrosimülasyon* yöntemi gibi politikanın hükümet ya da karar alıcılar tarafından uygulanmasından önce yani politika tamamlanmadan önce yapılan tahmin yöntemlerini kapsamaktadır. Politikanın tamamlanmasından sonra elde edilen verilere dayalı ölçme yapan **Program Sonrası Değerlendirme** yöntemleri ise *deneysel* ve *deneysel olmayan* yöntemler olarak ikiye ayrılmaktadır. *Deneysel yöntemler*, *rastgele dağıtım* (random assignment) yöntemi ile toplanan verileri kullanmaktadır. Deneysel yöntemlerde politikanın uygulandığı bireylerin rastgele seçim yoluyla belirlenmesi ve anketler yoluyla bu bireylerden veri toplanması çok maliyetli olduğundan, araştırmacılar deneysel olmayan (*hazırda var olan, gözlemsel*) verileri kullanan yöntemlerle çalışmayı tercih etmektedirler. Politika etki analizinde deneysel olmayan yöntemler arasında literatürde en çok kullanılan yöntemler “**Farkların-Farkı (Difference in Differences)**”, “**Eğilim Skoru Eşleştirme (Propensity Score Matching)**” ve “**Regresyon Süreksizliği Analizi (Regression Discontinuity Analysis)**” olduğundan, kitabın 17. bölümünde söz konusu yöntemlerin teorik çerçevesi ve Stata 15 ekonometrik programı ile yapılan uygulamaları anlatılacaktır. Bu

¹ Dr., Manchester Metropolitan University, Department of Economics, Policy and International Business. Adnan Menderes Üniversitesi, Nazilli İİBF, Maliye Bölümü. E-mail: giovanis95@gmail.com

² Doç. Dr., İzmir Bakırçay Üniversitesi, İktisadi ve İdari Bilimler Fakültesi, İktisat Bölümü. E-mail: oznurozdamar@gmail.com

yöntemler için genellikle çok dönemli ve gözlem sayısı yüksek veri setlerine ihtiyaç duyulduğundan, mikro temelli hanehalkı ve firma anketlerinden elde edilen veri setleri ile çalışılmaktadır.

17.1. FARKLARIN FARKI YÖNTEMİ

17.1.1. Farkların Farkı Yönteminin Teorik Çerçevesi

Farkların Farkı (FF) yöntemi, nicel yöntemlerin alt sınıflandırması olan “*program sonrası değerlendirme*” yöntemlerinden “*deneysel olmayan yöntemler*” grubuna girer. Çünkü FF yöntemi gözlenebilir hazır verilerle, politika tamamlandıktan sonra etki analizi yapmaktadır. Analizde hem politika öncesi hem de politika sonrası dönem verilerine ihtiyaç duyulduğundan, kullanılan veri seti türü havuzlanmış yatay kesit ya da panel veri setidir.

FF yöntemi, hedeflenen çıktı açısından politika öncesi ve politika sonrası dönemde *müdahale grubu* (politikanın uygulandığı ya da politikadan etkilenen bireyler) ve *kontrol grubu* (politikanın uygulanmadığı ya da politikadan etkilenmeyen bireyler) arasındaki göreceli farklılıkları ortaya koyarak politikanın nedensel etkisini tahmin etmektedir. Örneğin hükümetin ortalama istihdam oranı düşük olan A bölgesinde bireylerin istihdamını artırmayı amaçladığı bir politika uyguladığını düşünelim. Burada hükümetin iyileştirmeyi amaçladığı hedef çıktı, bireylerin istihdam durumudur. Politika ile birlikte işsiz bireylerin istihdam edilmesi hedeflenmektedir. Politika A bölgesinde uygulandığından, A bölgesinde yaşayan bireyler *müdahale grubunu* oluşturmaktadır. Politikadan yararlanmayan B bölgesindeki bireyler de *kontrol grubunu* oluşturmaktadır. FF yöntemi, politika öncesi dönemde müdahale grubu ve kontrol grubu arasında hedeflenen çıktı açısından farklılıklar olmasına izin vermektedir. Örneğin politika A bölgesinde uygulandığı için, politika öncesi dönemde A bölgesindeki ortalama istihdam oranının B bölgesindeki ortalama istihdam oranından daha düşük olması beklenen bir durumdur. Politika uygulandıktan sonra da bu iki grup arasında ortalama istihdam oranı açısından farklılıklar oluşabilecektir. Politika sonrasında, politikanın etkinliği açısından pozitif ve negatif olmak üzere iki durumla karşılaşılabilir. Pozitif durum politikanın etkin olduğu yani hedeflediği amaca ulaştığı durumdur. Negatif durum ise politikanın etkin olmadığı durumdur. Politika etkinse, politika sonrası dönemde A bölgesi istihdam oranı B bölgesi istihdam oranını geçmiş ya da B bölgesi istihdam oranına yaklaşılarak aralarındaki istihdam oranı farkı kapanmış olabilir. Politika etkin değilse aralarındaki fark daha da açılmış olabilir. FF yöntemi olası pozitif durumun politikadan kaynaklandığını anlamak için politika dışında istihdam oranını etkileyebilecek diğer faktörleri regresyonda kontrol altında tutmaktadır. FF yöntemi politika dışında istihdam oranını etkileyebilecek faktörleri regresyonda kontrol ederek, politika sonrası ve öncesi dönem için müdahale ve kontrol grubu arasındaki ortalama istihdam oranı *farklarının farkını* tahmin eder. Diğer bir ifade ile politika sonrasında bu iki grup arasındaki ortalama istihdam oranı farkının, politika öncesi dönem ortalama istihdam oranı farkına göre ne kadar değiştiğini tahmin ederek politikanın amaca ulaşmada etkin olup olmadığını test etmektedir. Yöntem bu nedenle “*farkların farkı*” olarak adlandırılmaktadır. Politika sonrası dönemde politika öncesi döneme kıyasla bu iki grup arasındaki ortalama istihdam oranı farkı azalmış ise, politikanın etkin olduğu söylenebilir.

Günümüzde özellikle kamu politika ve müdahalelerinin etkisini ölçmek için uygulamalı iktisat başta olmak üzere sosyal bilimlerde kullanılan FF yönteminin ilk uygulamaları sağlık bilimleri alanında olmuştur. Epidemiyoloji biliminin babası olarak adlandırılan John Snow'un FF yöntemini (o dönemde bu adla adlandırılmasa da) kullanarak yaptığı çalışmalarına kadar (Snow, 1855; Snow, 1856) kolera hastalığının en önemli nedeninin hava kirliliği olduğu düşünülürken, John Snow kolera'nın kirli suyuyla bulaşan bir hastalık olduğunu kanıtlamıştır. Analizlerde 1849 ve 1854 yıllarında içme suyu kirli olan Londra'nın Soutwark ilçesinin 12 semti kontrol, Southwark ve Lambert ilçelerinin 16 semti de müdahale grubu olarak ele alınmıştır. 1849 ve 1854'de içme suyu kirli olan Southwark'ın yanı sıra 1854'de içme suyu temiz olan Lambert ilçesi semtlerinin müdahale grubunda olması bu iki grubu farklılaştırmaktadır. Özetle 1849'da (müdahale öncesi dönemde) sadece içme suyu kirli olan bir müdahale ve kontrol grubu varken, 1854'de (müdahale sonrası dönemde) müdahale grubunda temiz içme suyuna erişimi olan semtler de yer almıştır. Böylece 1849 yılında müdahale ve kontrol grubu arasında koleradan kaynaklı ölüm oranı farkı tahmin edilirken, 1854'te aynı fark elde edilmiş ve koleradan ölüm oranı farklarının ne kadar değiştiği (farkların farkı) yani temiz su kullanımının etkisi ölçülmüştür.

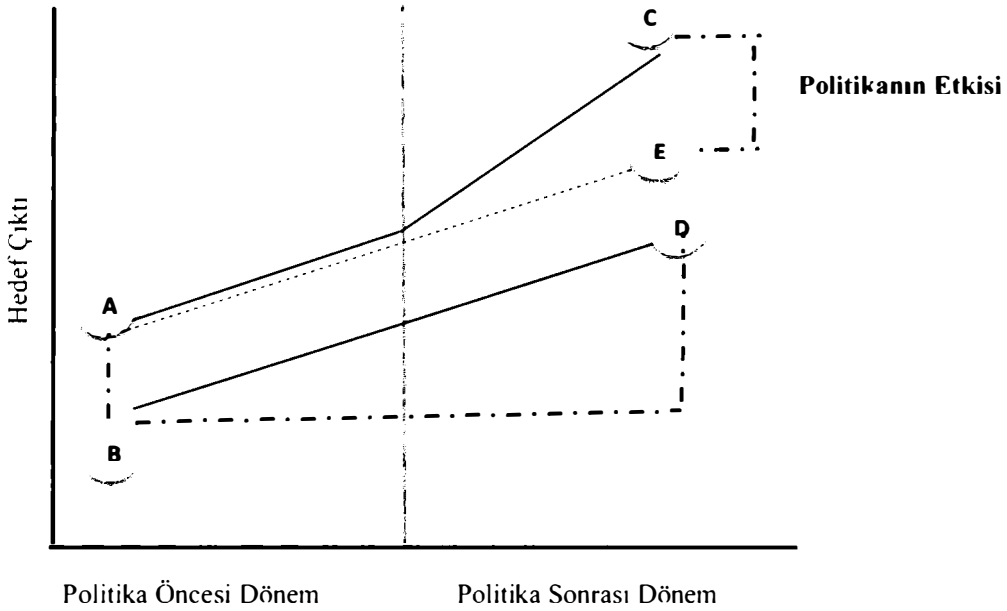
Zamanla sağlık alanı dışında da kullanılmaya başlanan FF yöntemi ile sosyal bilimlerde ilk çalışmalardan olan Rose (1952) hükümet kanunlarındaki değişikliğin etkisini analiz etmiştir. Simon (1966) ise fiyatlarda yapılan değişikliğin likör satışlarına etkisini incelemiş ve fiyat değişikliğinin yaşandığı dönem sonrasında, değişiklik öncesi döneme göre iki eyalet arasındaki ortalama likör satış farklarının nasıl değiştiğini tahmin etmiştir.

Bu çalışmaların öncülük ettiği literatürde FF yöntemi çok sıklıkla kullanılan bir yöntem hâlini almıştır. Yapılan tüm bu çalışmalarda elde edilen sonuçların doğruluğu tek bir varsayımın geçerliliğine bağlıdır. Bu varsayım politikanın uygulanmaması durumunda politika öncesinde hedef çıktı açısından müdahale ve kontrol grupları arasındaki farkın politika sonrası dönemde de aynı kalacağını belirten *paralel eğilim* (parallel trend) varsayımdır. Eğer paralel eğilim varsayımı geçerli değilse yani politika öncesi dönemde müdahale ve kontrol gruplarının ortalama hedef çıktı değerleri paralel eğilim sergilemezse, politika sonrası oluşacak çıktı farkının sadece politikadan kaynaklandığı söylenemez. Bu farklılık hedef çıktıyı politika dışında etkileme potansiyeli olan diğer faktörlere atfedilir ve politikanın etkisini görmek mümkün olmaz. Paralel eğilim varsayımının sağlandığı durumda FF yöntemi, politika sonrası dönemde müdahale ve kontrol grupları arasındaki kalıcı farklılıklardan kaynaklı yanlılığı veya müdahale grubu için hedef çıktıyı etkileyebilecek diğer faktörlerden kaynaklı yanlılığı ortadan kaldırmaktadır. Paralel eğilim varsayımı ve politikanın etkisi basit olarak Şekil 17.1.1'de gösterilmiştir.

Şekil 17.1.1'de A ve B noktaları sırasıyla müdahale ve kontrol gruplarının politika öncesi dönemde hedef çıktı değerlerinin ortalamasını göstermektedir. Görüldüğü üzere politika öncesi dönemde hedef çıktı açısından aralarında fark vardır ve politika yılına kadar hedef çıktındaki değişim paralellik göstermektedir. Bu durum paralel eğilim varsayımının geçerli olduğuna işarettir. Politikanın uygulanması ile birlikte kontrol grubunun ortalama hedef çıktı değeri D düzeyine gelirken, müdahale grubunun C düzeyine geldiği görülmektedir. Diğer bir ifadeyle, hedef çıktı açısından politika öncesi fark (A-B) iken politika sonrasında (C-D) olmuştur. Bu iki farkın yani *farkların farkının* (C-E) arasındaki uzaklığa eşit olduğu görülmektedir. (C-E) farkı, politikanın etkisini göstermektedir. Politika uygulanmamış olsaydı, müdahale grubunun ortalama hedef çıktı

değerinin politika sonrası dönemde E noktası ile gösterileceği ve politika öncesi dönemdeki paralel eğilimin bozulmayacağı söylenebilir.

Şekil 17.1.1. FF yöntemiyle elde edilen politika etkisinin grafiksel gösterimi³



Grafiksel gösterimin yanı sıra politikanın etkisi (17.1.1) nolu eşitlik ile de gösterilebilir. ATE (average treatment effect), *ortalama politika etkisi* ya da *ortalama müdahale etkisi* olarak adlandırılır. Politika sonrası farktan politika öncesi farkın çıkarılması ile bulunur. D bir kukla değişken olup 1 değerini politikadan etkilenen müdahale grubu için, 0 değerini kontrol grubu için almaktadır, $d \in \{0, 1\}$.

$$ATE = \{E[Y_1|D = 1] - E[Y_0|D = 1]\} - \{E[Y_1|D = 0] - E[Y_0|D = 0]\} \quad (17.1.1)$$

Burada,

$E[Y_1|D = 1]$; politika sonrası dönemde müdahale grubunun ortalama hedef çıktı değerini,

$E[Y_1|D = 0]$; politika sonrası dönemde kontrol grubunun ortalama hedef çıktı değerini,

$E[Y_0|D = 1]$; politika öncesi dönemde müdahale grubunun ortalama hedef çıktı değerini,

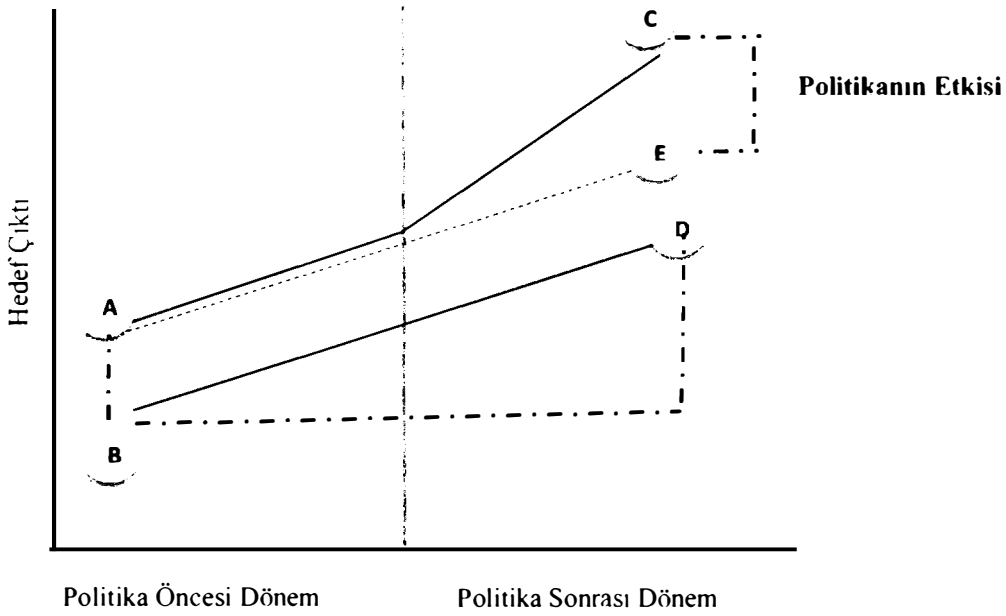
$E[Y_0|D = 0]$; politika öncesi dönemde kontrol grubunun ortalama hedef çıktı değerini

göstermektedir.

³Şekil 17.1.1 Columbia Üniversitesi, Halk Sağlığı Araştırma Yöntemlerinin web sayfasından alınmış olup ufak notasyon değişikliklerinin yanı sıra türkçeleştirilmiştir. Erişim Tarihi: 5 Haziran 2020. Erişim Linki; <https://www.publichealth.columbia.edu/research/population-health-methods/difference-difference-estimation>

değerinin politika sonrası dönemde E noktası ile gösterileceği ve politika öncesi dönemdeki paralel eğilimin bozulmayacağı söylenebilir.

Şekil 17.1.1. FF yöntemiyle elde edilen politika etkisinin grafiksel gösterimi³



Grafiksel gösterimin yanı sıra politikanın etkisi (17.1.1) nolu eşitlik ile de gösterilebilir. ATE (average treatment effect), *ortalama politika etkisi* ya da *ortalama müdahale etkisi* olarak adlandırılır. Politika sonrası farktan politika öncesi farkın çıkarılması ile bulunur. D bir kukla değişken olup 1 değerini politikadan etkilenen müdahale grubu için, 0 değerini kontrol grubu için almaktadır, $d \in \{0, 1\}$.

$$ATE = \{E[Y_1|D = 1] - E[Y_0|D = 1]\} - \{E[Y_1|D = 0] - E[Y_0|D = 0]\} \quad (17.1.1)$$

Burada,

$E[Y_1|D = 1]$; politika sonrası dönemde müdahale grubunun ortalama hedef çıktı değerini,

$E[Y_1|D = 0]$; politika sonrası dönemde kontrol grubunun ortalama hedef çıktı değerini,

$E[Y_0|D = 1]$; politika öncesi dönemde müdahale grubunun ortalama hedef çıktı değerini,

$E[Y_0|D = 0]$; politika öncesi dönemde kontrol grubunun ortalama hedef çıktı değerini

göstermektedir.

³Şekil 17.1.1 Columbia Üniversitesi, Halk Sağlığı Araştırma Yöntemlerinin web sayfasından alınmış olup ufak notasyon değişikliklerinin yanı sıra türkçeleştirilmiştir. Erişim Tarihi: 5 Haziran 2020. Erişim Linki; <https://www.publichealth.columbia.edu/research/population-health-methods/difference-difference-estimation>

Hedef çıktı değerinin simgelendiği Y değişkenin alt indisi politika öncesi dönem için 0, politika sonrası dönem için 1 değerini almaktadır. Politika dışında hedef çıktıyı etkileyebilecek diğer faktörlerin kontrol edilerek politikanın etkisinin tahmin edilmesinde (17.1.2) nolu eşitlikteki ekonometrik model kullanılmaktadır.

$$y_{it} = \beta_0 + \beta_1 \text{müdahale}_{it} + \beta_2 \text{dönem}_{it} + \beta_3 (\text{müdahale}_{it} * \text{dönem}_{it}) + \delta' X_{it} + \varepsilon_{it} \quad (17.1.2)$$

(17.1.2) nolu denklemde y_{it} politikanın üzerinde etkisi merak edilen hedef çıktıyı göstermektedir. FF yönteminin uygulandığı örneklem havuzlanmış yatay kesit ya da panel veri setinden elde edildiğinden, değişkenler t zamanında yaşayan i bireyine ait gözlemleri içermektedir. müdahale_{it} değişkeni, politikadan etkilenen grup için 1, kontrol grubu için 0 değerini almaktadır. dönem_{it} değişkeni politika öncesi dönem için 0, politika sonrası dönem için 1 değerini alan bir kukla değişkendir. X , kontrol değişkenleri vektörüdür ve politika dışında bağımlı değişkeni etkileyebilecek diğer açıklayıcı değişkenleri içerir. ε_{it} , özdeş ve bağımsız olduğu varsayılan hata terimini göstermektedir. Modelde $(\text{müdahale}_{it} * \text{dönem}_{it})$ değişkeninin katsayısı olan β_3 politikanın etkisini gösterir ve bu katsayı FF tahmincisi olarak adlandırılır. $(\text{müdahale}_{it} * \text{dönem}_{it})$ değişkeni müdahale ve dönem değişkenlerinin çarpımından oluşan bir etkileşim değişkenidir. Tablo 17.1.1'in ilk iki satırı politika dışında hedef çıktıyı etkileyebilecek diğer değişkenler kontrol altındayken müdahale ve kontrol grupları için politika öncesi ve sonrası dönemde bağımlı değişkenin ortalama değerini ve her bir grup için dönemler arası farkı (değişimi) göstermektedir. Üçüncü satır ise bağımlı değişkenin ortalaması açısından gruplar arası farkın politika öncesinde, politika sonrasında ve bir dönemden diğerine aldığı değeri göstermektedir. Özetle β_3 , FF tahmincisi olup, politikanın etkisini ifade etmektedir.

Tablo 17.1.1. Politika Öncesi ve Sonrası Dönemde Hedef Çıktı Düzeyleri ve FF Tahmincisi⁴

	POLİTİKA ÖNCESİ	POLİTİKA SONRASI	Değişim (Fark)
KONTROL	β_0	$\beta_0 + \beta_2$	β_2
MÜDAHALE	$\beta_0 + \beta_1$	$\beta_0 + \beta_1 + \beta_2 + \beta_3$	$\beta_2 + \beta_3$
Değişim (Fark)	β_1	$\beta_1 + \beta_3$	β_3

17.1.2. Farkların Farkı Yöntemi: Stata Uygulaması

COVID-19 dünyada tüm ülkelerin savaştığı küresel bir pandemi olarak varlığını pek çok alanda hissettirmiştir. Hükümetler virüsün yayılmasını önlemek için pek çok uygulamayı hayata geçirmiştir. Alışveriş merkezleri, restoranlar ve eğlence mekânlarının kapatılması, spor müsabakalarının seyircisiz yapılması, okulların uzaktan eğitime geçmesi ve çalışma hayatında olabildiğince esnek çalışma uygulamaları bunlara örnektir. İstihdam göstergeleri açısından istenmeyen sonuçlar doğuran, dünyanın pek çok ülkesinde işsizliğin artmasına neden olan COVID-19'un bireylerin yaşam standardını daha da kötü etkilememesi için hükümetler ayrıca istihdam destek politikalarını hayata geçirmiştir. Örneğin Birleşik Krallık Mart 2020'de işsizliği önlemek için,

⁴Tablo 17.1.1, Wooldridge, J. M. (2013:457) kaynağındaki tablodan esinlenerek hazırlanmış olup (17.1.2) nolu denklemdeki katsayılar kullanılarak incelenen konuya uyarlanmıştır.

“İzindeyken İşini Koruma” (Furlough Job Retention) programını başlatmıştır. Bu program sayesinde çalışanlar geçici olarak işlerine devam edemeseler de işten çıkarılmamaktadır. Program kapsamında 2,500 £’a kadar her bir çalışanın aylık maaşının %80’i ile işverenler tarafından ödenen sağlık ve emeklilik primlerinin hükümet tarafından karşılanması kararlaştırılmıştır. 2020 Temmuz sonuna kadar bu şartlarla devam eden programda, Ağustos itibariyle işverenlerin sağlık ve emeklilik prim ödemelerini kendilerinin yapmalarına, aylık ücretin ise Eylül’de %70, Ekim’de %60’ının devlet tarafından ödenmesine karar verilmiştir (HM Government, 2020).

Bu bölümde yapılan FF uygulamasının amacı “İzindeyken İşini Koruma” politikasının bireylerin “gelecekteki finansal durum algıları” üzerine etkisini analiz etmektir. Program kapsamında ücretlerin yüzde yüzü ödenmediğinden ve giderek azalan oranlarda ödeneceğinden, bireyler gelecek dönem finansal durumları hakkında kötümser olabilirler. Gelecek dönemlerdeki finansal durumları hakkında kaygı duymalarına neden olabilecek bir diğer faktör 2020 yılı Ekim sonunda programın uygulanmaya devam edip etmeyeceğinin belli olmamasıdır. Ayrıca Ağustos itibariyle sigorta maliyetlerinin tekrar işverenler tarafından ödenecek olması işten çıkarımları beraberinde getirebileceğinden, çalışanların finansal durum algılarının olumsuz yönde etkilenmesi olasıdır. Diğer yandan pek çok iş yerinin kapatıldığı bu süreçte böyle bir uygulamanın en azından gelecek dönemde işlerini koruyabilme şansı verdiği düşüncesiyle bireyler, gelecek dönem finansal durumları hakkında iyimser bir algı içerisinde olabilirler. “İzindeyken İşini Koruma” politikasının bu ihtimallerden hangisine sebebiyet vereceği, bireylerin finansal durum algıları konusunda etkisinin pozitif mi negatif mi olacağı (17.1.3) nolu model ile analiz edilmiştir.

$$gfd_{it} = \beta_0 + \beta_1 mud_{it} + \beta_2 covd_{it} + \beta_3 (mudcovd_{it}) + \delta' X_{it} + \theta_t + \varepsilon_{it} \quad (17.1.3)$$

(17.1.3) nolu modelin tahmininde Birleşik Krallık Hanehalkı Panel Anketi (UK Household Longitudinal Study (UKHLS)) kullanılmıştır. Anket, Essex Üniversitesi’nde bulunan Sosyal ve Ekonomik Araştırma Enstitüsü tarafından 2009 yılından itibaren düzenli aralıklarla yapılmaktadır. 2020 yılına kadar 9 dalga (wave) hâlinde yayınlanmış olan anket, Nisan 2020’de 10. dalga olarak adlandırılmasa da COVID-19’a ilişkin soruları kapsayan özel bir anket şeklinde yayımlanmıştır. Analizde, 7., 8., ve 9. dalgaların kapsadığı 2016-2019 yılları politika öncesi dönem, 2020 yılı politika sonrası dönem olarak ele alınmıştır. Program öncesi döneme ait veriler ve son UKHLS anket verilerinden veri birleştirme (data merging) teknikleri kullanılarak 2016-2020 yıllarını kapsayan Covid.dta⁵ veri seti elde edilmiştir. Bu bağlamda modelde covd değişkeni politika öncesi dönem olan 2016-2019 yılları için 0, politika sonrası dönem olan 2020 yılı için 1 değerini almaktadır. gfd, bireylerin gelecek dönemdeki finansal durum algılarını göstermektedir. Bu değişkeni ankette temsil eden soru iki cevaplı olup, gelecek dönem finansal durumlarının çok kötüye gideceğini düşünen bireyler için 1, iyiye gideceğini düşünenler için 0 değerini almaktadır⁶. mud değişkeni “İzindeyken İşini Koruma” politikasından yararlananlar (müdahale grubu) için 1, politikadan yararlanmayanlar (kontrol grubu) için 0 değerini almaktadır.

⁵Covid.dta veri setinin orijinalinde değişkenlerin isimleri İngilizce olup, anlatımı kolaylaştırmak adına değişken adları türkçeleştirilmiştir. Covid.dta veri setine karekod ile ulaşabilirsiniz.

⁶Analizde incelenen bağımlı değişken iki durumlu bir kukla değişken olmakla birlikte, FF yöntemi bağımlı değişkenin sıralı (ordered) bir nitel değişken veya sürekli değerler alan nicel bir değişken olduğu durumda da uygulanabilir.

β_3 katsayısı FF tahmincisi olup, “İzindeyken İşini Koruma” politikasının bireylerin gelecek dönem finansal durumlarına yönelik algıları (*gfd*) üzerine etkisini, yani politikanın etkisini ölçmektedir. Bu nedenle *mudcovid* değişkeni FF değişkeni ya da politika değişkeni olarak adlandırılmaktadır. Bağımlı değişkeni etkileyebilecek yaş, cinsiyet, istihdam durumu gibi sosyo-ekonomik faktörler X vektörü içerisinde kontrol edilmiştir. *yas* yıl cinsinden nicel bir değişken olmakla birlikte, *cinsiyet* kadınlar için 0, erkekler için 1 değerini alan bir kukla değişkendir. *istihdam*, üç seçeneekli kategorik bir değişken olup, işçi/memur olarak çalışanlar için 1, kendi hesabına çalışanlar için 2, hem kendi hesabına hem de işçi/memur olarak çalışanlar için 3 değerini almaktadır. θ_t yıl sabit etkilerini, ε_{it} özdeş ve bağımsız olduğu varsayılan hata terimini göstermektedir.

17.1.2.1. Analizde Kullanılan Değişkenlerin Tanımları ve Tanımlayıcı İstatistikler

Analiz öncesi “describe” komutu ile **Covid.dta** veri setinde yer alan değişkenlerin tanımlarının sunulması ve “sum” komutu ile tanımlayıcı istatistiklerin gösterilmesi önemlidir. Değişkenlerin tanımları ve tanımlayıcı istatistikleri sırasıyla Tablo 17.1.2 ve Tablo 17.1.3’te gösterilmiştir. **Covid.dta** veri setinde gözlem sayısı (obs) 10,842’dir⁷. *fkimlik* değişkeni veri setinde yer alan her bir bireye ait fert kimlik numaralarıdır.

Tablo 17.1.2. Değişkenlerin Tanımları
describe

Contains data from C:\Users\Asus\Desktop\Covid.dta				
obs:	10,842			
vars:	7			9 Sep 2020 22:57
size:	238,524			

variable name	storage type	display format	value label	variable label
cinsiyet	byte	%8.0g	cinsiyet1	Ferdin Cinsiyeti (Kadın=0, Erkek=1)
istihdam	byte	%50.0g	istihdam1	İstihdam durumu
mud	float	%9.0g		Müdahale (müdahale =1, kontrol=0)
gfd	float	%9.0g		Gelecekteki finansal durum algısı
yil	float	%9.0g		Anket Yılı
yas	float	%9.0g		Ferdin bitirdiği yaş
fkimlik	float	%9.0g		Ferdin kimlik numarası

⁷Stata, Türkçe dilinde çalışan bir program olmadığından ondalık ayracını nokta (.) binlik ayracını ise virgül (,) olarak göstermektedir. Bu bölümde Stata çıktıları yorumlanırken orijinaline sadık kalınarak sayısal ifadeler türkçeleştirilmemiş olup, binlik ayracı olarak virgül ve ondalık ayracı olarak nokta kullanılmıştır.

Tablo 17.1.3. Değişkenlerin Tanımlayıcı İstatistikleri

sum

Variable	Obs	Mean	Std. Dev.	Min	Max
cinsiyet	10,842	.4002029	.4899619	0	1
istihdam	10,842	1.085778	.3955554	1	3
mud	10,842	.1915698	.3935545	0	1
gfd	10,622	.1322726	.3388029	0	1
yil	10,842	2017.827	1.486051	2016	2020
yas	10,842	41.74885	12.02185	16	78
fkimlik	10,842	38901.93	5506.525	29436	48418

(17.1.3) nolu modelin analiz edilebilmesi için **Covid.dta** veri setinde yer alan *yil* ve *mud* değişkenleri kullanılarak *covd* ve *mudcovd* değişkenlerinin oluşturulması gerekmektedir. *covd* değişkeni 2020 yılı öncesi yıllar için 0, 2020 yılı için 1 değerini almakta olup [17.1.1]⁸ nolu komut dizini ile oluşturulmuştur. *mudcovd* olarak ifade edilen FF değişkeni ise [17.1.2] nolu komutla elde edilmiştir.

```
gen covd=.
```

```
replace covd=1 if yil==2020
```

```
replace covd=0 if yil<2020 [17.1.1]
```

```
gen mudcovd=mud*covd [17.1.2]
```

(17.1.3) nolu denklem sırasıyla Doğrusal Olasılık Modeli (DOM)⁹ ve Logit Modeli ile analiz edilmiştir. Bağımlı değişken bir kukla değişken olduğu için bu yöntemlerin uygulanması tercih edilmiştir¹⁰.

17.1.2.2. Doğrusal Olasılık Modeli (DOM) Temelinde Farkların Farkı (FF) Analizi

DOM temelinde yapılan analizde [17.1.3] nolu komut kullanılmış olup analiz sonuçları Tablo 17.1.4'te gösterilmiştir. Değişen varyansın varlığı altında dirençli standart hatalar elde etmek için *cluster(fkimlik)* opsiyonu kullanılmıştır¹¹.

```
xi: reg gfd mud covd mudcovd cinsiyet yas i.istihdam i.yil , cluster(fkimlik) [17.1.3]
```

⁸Kitap bölümünde tüm denklem ve kullanılan Stata komutları başta bölüm numarası gelecek şekilde numaralandırılmış olup, karışıklığa neden olmamak için denklemlerin numaraları yuvarlak parantez (), komutların numaraları ise köşeli parantez [] içine alınmıştır.

⁹DOM En Küçük Kareler Yöntemi (EKKY) ile analiz edilmiştir.

¹⁰Probit modeli de uygulanabilecek bir diğer model olup, kitap bölümünde sadece Logit modelinin uygulanması tercih edilmiştir. Probit modeli ile analizi gerçekleştirmek için ilgili komutlarda **logit** yazan yere **probit** yazılması yeterli olacaktır.

¹¹*cluster (fkimlik)* yerine *robust* opsiyonu da değişen varyansa karşı dirençli standart hatalar elde etmek için kullanılabilir.

Tablo 17.1.4. Doğrusal Olasılık Modelinin Tahmini

```

xi: reg gfd mud covd mudcovd cinsiyet yas i.istihdam i.yil , cluster(fkimlik)
i.istihdam      _Iistihdam_1-3      (naturally coded; _Iistihdam_1 omitted)
i.yil           _Iyil_2016-2020     (naturally coded; _Iyil_2016 omitted)
note: _Iyil_2020 omitted because of collinearity

```

```

Linear regression              Number of obs      =      10,622
                              F(10, 2789)          =      17.65
                              Prob > F              =      0.0000
                              R-squared              =      0.0268
                              Root MSE           =      .33439

```

(Std. Err. adjusted for 2,790 clusters in fkimlik)

		Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]	
gfd							
mud		-.0166647	.0104219	-1.60	0.110	-.0371001	.0037706
covd		.029918	.0093901	3.19	0.001	.0115057	.0483303
mudcovd		.2283853	.022869	9.99	0.000	.1835434	.2732271
cinsiyet		-.0134803	.0080588	-1.67	0.094	-.0292821	.0023214
yas		.001118	.0003168	3.53	0.000	.0004967	.0017393
_Iistihdam_2		.0891623	.0514175	1.73	0.083	-.0116578	.1899825
_Iistihdam_3		.0660354	.0251775	2.62	0.009	.0166671	.1154038
_Iyil_2017		.0183897	.0081064	2.27	0.023	.0024946	.0342847
_Iyil_2018		.0193225	.0079945	2.42	0.016	.0036467	.0349982
_Iyil_2019		.0117159	.0211575	0.55	0.580	-.0297699	.0532018
_Iyil_2020		0	(omitted)				
_cons		.0624185	.0143043	4.36	0.000	.0343704	.0904665

Tablo 17.1.4'te gösterilen sonuçlarda *mud* değişkenine ait katsayı istatistiksel olarak anlamsız çıktığından ($p=0.110$) politika öncesi dönemde bu iki grubun *gfd* açısından birbirine benzer olduğu söylenebilir. Tablo 17.1.1'den hatırlanacağı üzere müdahale (*mud*) değişkeninin katsayısı (β_1), politika öncesinde bağımlı değişkenin (*gfd*) ortalama değeri açısından bu iki grup arasında bir farkın olup olmadığını göstermektedir. İlerleyen bölümlerde paralel eğilim varsayımının geçerli olup olmadığı Stata 15 komutlarıyla test edilecek olsa da (*mud*) değişkeninin katsayısının anlamsız bulunmasının, politika öncesi dönemde müdahale ve kontrol gruplarına ait ortalama *gfd* değerlerinin paralel bir eğilim izlediklerine işaret ettiğini söylemek yerinde olacaktır.

covd değişkeninin katsayısı ise bağımlı değişkenin ortalama değerinin politika öncesi dönemden politika sonrası döneme sadece kontrol grubu için ne kadar değiştiğini göstermektedir. *covd* değişkeninin katsayısı pozitif ve istatistiksel olarak anlamlı olduğundan ($p=0.001$), kontrol grubundaki bireylerin finansal durum algılarının politika sonrası dönemde, politika öncesi döneme göre %2.9 daha da kötüleştiği söylenebilir. *gfd* değişkeni gelecek dönem finansal durumlarının çok kötüye gideceğini düşünen bireyler için 1, iyiye gideceğini düşünenler için 0 değerini aldığından pozitif bir katsayının elde edilmesi algıda artan kötüleşmeye işaret etmektedir. FF değişkeninin (*mudcovd*) katsayısının ise pozitif ve istatistiksel olarak anlamlı olduğu görülmektedir ($p=0.000$). Bu katsayı politikanın etkisini göstermekte olup, müdahale grubu için gelecek dönem

finansal duruma ilişkin kaygı artışının kontrol grubundaki bireylerin kaygı artışından %22.8 daha fazla olduğuna işaret etmektedir. Özetle “İzindeyken İşini Koruma” politikasının, politikadan yararlananlarda daha kötümser bir algı yarattığı belirlenmiştir. “İzindeyken İşini Koruma” politikasına COVID-19 nedeni ile iş yapamayan sektörlerdeki işverenler başvurmakta, doğal olarak bu politikadan bu sektörlerde çalışanlar yararlanmaktadır. Kontrol grubunda yer alan bireyler ise izine ayrılmak zorunda kalmadan çalışmaya devam edebilen bireylerdir. Örneğin süpermarketlerde, eczanelerde, hastanelerde, eğitim kurumlarında çalışanlar bu dönemde izne ayrılmadan işlerine devam edebilmişlerdir. Her ne kadar kontrol grubundaki bireyler de durumdan kötü etkilenip gelecek kaygısı taşısalar da FF analizi müdahale grubundakilerin gelecekteki finansal durumları hakkında daha kötümser olduklarını göstermektedir.

Kontrol değişkenlerinden yaş, cinsiyet, istihdam durumuna ilişkin elde edilen tahminler incelendiğinde, artan her bir yıl yaş ile birlikte bireylerin gelecek dönem finansal durumları hakkındaki düşüncelerinin %0.11 oranında daha olumsuz olduğu görülmektedir. Kadınlar için 0, erkekler için 1 değerini alan cinsiyet değişkeninin katsayısı ise erkeklerin gelecek dönem finansal durumları konusunda kadınlardan %1.34 oranında daha iyimser olduğunu göstermektedir. *gfd* değişkeni gelecek dönem finansal durumlarının çok kötüye gideceğini düşünen bireyler için 1, iyiye gideceğini düşünenler için 0 değerini aldığından negatif bir katsayının elde edilmesi erkeklerin gelecek dönemdeki finansal durumlarına ilişkin algılarının referans kategori olan kadınlara kıyasla daha olumlu olduğunu göstermektedir. London School of Economics bünyesinde faaliyet gösteren Ekonomik Performans Merkezi'nin (Center for Economic Performance) son dönem araştırmaları, Covid-19 tecritinin kadın çalışanların yoğun olduğu sektörleri daha olumsuz etkilediğini ve kadınların erkeklere kıyasla daha fazla iş kaybı yaşadığını ortaya koymuştur (Hupkau ve Petrongolo, 2020). Bu sonuç incelenilen örnekte elde edilen bulguları destekler niteliktedir. Kendi hesabına çalışanlar ile hem kendi hesabına hem de işçi/memur olarak çalışanların sadece işçi/memur olarak çalışanlara göre gelecek dönem finansal durum algılarının sırasıyla %8.91 ve %6.60 oranlarında daha olumsuz olduğu da elde edilen diğer bulgular arasındadır. Restoran, kafe gibi işletmelerde kendi hesabına çalışan bireylerin bu dönemde faaliyetlerini zorunlu olarak durdurmaları memurlar gibi devlet güvencesinde işleri olan bireylere nazaran daha çok gelecek kaygısı taşımalarına neden olmuş olabilir.

17.1.2.3. Logit Modeli Temelinde Farkların Farkı (FF) Analizi

Logit modeli temelinde FF analizi [17.1.4] nolu komut ile gerçekleştirilmiştir.

xi: logit gfd mud covd mudcovd cinsiyet yas i.istihdam i.yil, cluster(fkimlik) [17.1.4]

Logit modeli ile yapılan analizin sonuçları Tablo 17.1.5'te gösterilmiştir. [17.1.4] nolu komut çalıştırıldıktan sonra [17.1.5] komut ile ayrıca marjinal etkiler elde edilmiş ve marjinal etki tahminleri Tablo 17.1.6'da sunulmuştur.

margins, dydx(*) [17.1.5]

Tablo 17.1.5'te görüldüğü gibi *mudcovd* değişkeninin pozitif olan katsayısı, müdahale grubu için gelecek dönem finansal duruma ilişkin kaygı artışının kontrol grubundaki kaygı artışından daha fazla olduğuna işaret etmektedir.

Tablo 17.1.5. Logit Modelinin Tahmini

```
xi: logit gfd mud covd mudcovd cinsiyet yas i.istihdam i.yil , cluster(fkimlik)
```

```
i.istihdam      _Iistihdam_1-3      (naturally coded; _Iistihdam_1 omitted)
i.yil           _Iyil_2016-2020     (naturally coded; _Iyil_2016 omitted)
note: _Iyil_2020 omitted because of collinearity
Iteration 0:    log pseudolikelihood = -4149.8474
Iteration 1:    log pseudolikelihood = -4061.8613
Iteration 2:    log pseudolikelihood = -4035.8071
Iteration 3:    log pseudolikelihood = -4035.7846
Iteration 4:    log pseudolikelihood = -4035.7846
```

```
Logistic regression                                Number of obs      =      10,622
                                                    Wald chi2(10)      =      256.15
                                                    Prob > chi2        =      0.0000
Log pseudolikelihood = -4035.7846                Pseudo R2          =      0.0275
                                                    (Std. Err. adjusted for 2,790 clusters in fkimlik)
```

		Coef.	Robust Std. Err.	z	P> z	[95% Conf. Interval]	
gfd							
mud		-.1677989	.1091105	-1.54	0.124	-.3816516	.0460537
covd		.2858279	.0893428	3.20	0.001	.1107192	.4609365
mudcovd		1.385206	.1431488	9.68	0.000	1.104639	1.665772
cinsiyet		-.1212359	.0734489	-1.65	0.099	-.2651931	.0227213
yas		.0100197	.0028257	3.55	0.000	.0044813	.015558
_Iistihdam_2		.6480452	.3069258	2.11	0.035	.0464817	1.249609
_Iistihdam_3		.4922519	.163968	3.00	0.003	.1708805	.8136233
_Iyil_2017		.1854222	.0814324	2.28	0.023	.0258176	.3450268
_Iyil_2018		.1935723	.0803094	2.41	0.016	.0361687	.3509759
_Iyil_2019		.1206536	.2112812	0.57	0.568	-.29345	.5347572
_Iyil_2020		0	(omitted)				
_cons		-2.532364	.1369474	-18.49	0.000	-2.800776	-2.263952

Tablo 17.1.6'da gösterilen *mudcovd* değişkenine ait marjinal etki tahminleri, politikadan yararlanan bireylerin kontrol grubundaki bireylere göre %15.46 daha yüksek olasılıkla gelecekteki finansal durumları hakkında kötümser olduklarını göstermektedir. Analizde kontrol edilen bireysel karakteristikler için DOM analizinden elde edilen sonuçların çok benzerine ulaşılmıştır. Artan her bir yıl yaş ile birlikte bireylerin gelecek dönem finansal durum algılarının %0.11 oranında kötüleştiği, erkeklerin kadınlara kıyasla gelecek dönem finansal durum algılarının %1.35 oranında daha iyimser olduğu bulunmuştur. Kendi hesabına çalışanlar ile hem kendi hesabına hem de işçi/memur olarak çalışanların sadece işçi/memur olarak çalışanlara göre gelecek dönem finansal durum algılarının sırasıyla %7.23 ve %5.49 oranında daha kötümser olduğu elde edilen diğer bulgulardandır. *gfd* değişkeni gelecek dönem finansal durumlarının çok kötüye gideceğini düşünen bireyler için 1, iyiye gideceğini düşünenler için 0 değerini aldığından pozitif katsayıların elde edilmesi, işçi/memur olarak çalışanlara göre diğer kategorilerdeki bireylerin gelecekteki finansal durumları hakkında daha olumsuz düşündüklerini göstermektedir.

Tablo 17.1.6. Logit Modelinin Marjinal Etkiler Tahmini**margins, dydx(*)**

Average marginal effects	Number of obs = 10,622					
Model VCE : Robust						
Expression : Pr(gfd), predict()						
dy/dx w.r.t. : mud covd mudcovd cinsiyet yas _Iistihdam_2 _Iistihdam_3 _Iyil_2017 _Iyil_2018 _Iyil_2019 _Iyil_2020						

		Delta-method				
		dy/dx	Std. Err.	z	P> z	[95% Conf. Interval]

mud	- .0187389	.0122035	-1.54	0.125	- .0426573	.0051795
covd	.0319198	.009919	3.22	0.001	.0124788	.0513607
mudcovd	.1546926	.0161238	9.59	0.000	.1230904	.1862947
cinsiyet	- .013539	.0081743	-1.66	0.098	- .0295603	.0024823
yas	.0011189	.0003188	3.51	0.000	.000494	.0017438
_Iistihdam_2	.0723703	.0342437	2.11	0.035	.0052539	.1394867
_Iistihdam_3	.0549721	.0183245	3.00	0.003	.0190568	.0908874
_Iyil_2017	.020707	.0091027	2.27	0.023	.0028659	.038548
_Iyil_2018	.0216172	.0089794	2.41	0.016	.0040179	.0392164
_Iyil_2019	.013474	.0235884	0.57	0.568	- .0327585	.0597064
_Iyil_2020	0	(omitted)				

Özet olarak hem DOM hem de logit modeli kullanılarak yapılan analizlerle müdahale ve kontrol grupları arasında gelecek dönem finansal durum algıları açısından bir farklılığın olduğu tespit edilmiştir. 17.1.1 nolu bölümde bu farklılığın uygulanan politikadan kaynaklandığını söyleyebilmenin yolunun *paralel eğilim varsayımının geçerli olmasına* bağlı olduğu belirtilmişti. Bu nedenle paralel eğilim varsayımının geçerli olup olmadığı, yani COVID-19 tecriti öncesi bağımlı değişkenin ortalama değerinin müdahale ve kontrol grupları için paralellik gösterip göstermediği test edilmelidir.

17.1.2.4. Paralel Eğilim Varsayımı Testleri

Paralel eğilim varsayımı *LAGS* (Gecikmeli Değerler) *Testi* ve *Grafiksel Test* olmak üzere iki farklı yoldan test edilebilmektedir. Ancak LAGS testi politika dışında hedef çıktıyı etkileyebilecek diğer faktörleri kontrol altında tuttuğundan daha güvenilir sonuçlar vermektedir.

17.1.2.4.1. LAGS (Gecikmeli Değerler) Testi

Şekil 17.1.1'den hatırlanacağı üzere politikanın uygulandığı yıla kadar bağımlı değişkenin ortalama değeri müdahale ve kontrol grupları için paralel eğimlerle hareket etmekte, politikanın uygulandığı yıl itibarıyla farklı eğimlere sahip olmaktadır. Bu durum sadece politikanın uygulandığı yılda politikadan kaynaklı olarak müdahale ve kontrol grubu arasında farklılık yaşanmış olduğunu göstermektedir. Bu nedenle politikanın uygulandığı yılın 2020 olduğu kabul edilerek oluşturulan *mudcovd* değişkenin katsayısı anlamlı çıkmıştır. Bu mantıktan hareketle, veri setindeki diğer yılların herhangi birinin (2016, 2017, 2018 veya 2019) politika yılıymış gibi kabul edilmesiyle oluşturulan farklı yıllara ait FF değişkenlerinin katsayılarının anlamsız çıkması gerekmektedir.

Sonuçları Tablo 17.1.4, Tablo 17.1.5 ve Tablo 17.1.6'da gösterilen analizler, politikanın asıl uygulandığı yıl olan 2020 göz önünde bulundurularak elde edilen *mudcovid* değişkeninin (2020 yılı FF değişkeni) katsayısını tahmin etmek için yapılmıştır. LAGS testinde politikanın uygulandığı yılın 2020 değil de öncesindeki yıllar olduğu varsayılarak farklı yıllara ait gecikmeli FF değişkenleri oluşturulmakta ve bağımlı değişkeni *gfd* olan bir model içerisinde bu değişkenler analiz edilerek katsayılarının ortak anlamlılığı sınanmaktadır. Sadece 2020 yılı FF değişkeninin katsayısı anlamlı, diğer yıllara ait gecikmeli FF değişkenlerinin katsayıları anlamsız bulunursa paralel eğilim varsayımının geçerli olduğu kabul edilmektedir. Bu nedenle LAGS testinin **ilk adımında** 2016-2020 arası her bir yılı temsil eden kukla değişkenleri **[17.1.6]** nolu komut ile elde edilmiştir.

tab yıl, gen(t)

[17.1.6]

Elde edilen yıl kukla değişkenleri sırasıyla, *t1*, *t2*, *t3*, *t4* ve *t5*'tir. Örneğin *t1* 2016 yılı için 1, diğer yıllar için 0 değerini almaktadır. Aynı şekilde *t5*, 2020 yılı için 1 diğer yıllar için 0 değerini almaktadır. **İkinci adımda** elde edilen her bir yıl kukla değişkeni ile *mud* değişkeninin etkileşimi yani gecikmeli yıllara ait FF değişkenleri **[17.1.7]** nolu komut dizini ile oluşturulur.

gen mudt1=mud*t1

gen mudt2=mud*t2

gen mudt3=mud*t3

gen mudt4=mud*t4

gen mudt5=mud*t5

[17.1.7]

mudt1, *mudt2*, *mudt3* ve *mudt4* gecikmeli FF değişkenleridir. *mudt5* ise 2020 yılına ait FF değişkenidir. LAGS testi F istatistiği temelli bir test olup, aynı modelde bağımsız değişkenler olarak yer alan gecikmeli FF değişkenlerinin ortak anlamlılıkları sınanmaktadır¹². Testin sıfır hipotezi gecikmeli FF değişkenlerin katsayılarının sıfıra eşit olduğunu, alternatif hipotezi ise sıfırdan farklı olduklarını belirtmektedir. Paralel eğilim varsayımının geçerli olması için sıfır hipotezinin reddedilmemesi¹³, 2020 yılı FF değişkeni olan *mudt5*'in katsayısının ise reddedilmesi (sıfırdan farklı olarak anlamlı bulunması) gerekmektedir. DOM temelli LAGS testi için **[17.1.8]** nolu komut dizini kullanılmış olup testin sonuçları Tablo 17.1.7'de gösterilmiştir.

xi:reg gfd mud t2 t3 t4 t5 mudt2 mudt3 mudt4 mudt5 yas cinsiyet i.istihdam, cluster(fkimlik)

test mudt2=mudt3=mudt4=0

[17.1.8]

¹² 2020 sonrası anket verisi olmadığından FF değişkeninin sadece geçmiş dönem değerleri (lags) anlamlılık testine tabi tutulmuştur. Farklı politika analizlerinde eğer politika sonrası dönem birden fazla yılı kapsıyorsa FF değişkeninin gelecek dönem değerleri (leads) de sınanmalıdır. Testin orijinal adı bu nedenle "Lags&Leads Test" olup, burada politika sonrası yaşanan sadece 2020 yılı olduğundan Lags testi uygulanmıştır.

¹³ Çoklu bağlantı problemine neden olabileceğinden *mudt1* değişkeni analize alınmamıştır.

Tablo 17.1.7. DOM Temelinde LAGS Testi

```
xi:reg gfd mud t2 t3 t4 t5 mudt2 mudt3 mudt4 mudt5 yas cinsiyet i.istihdam, cluster(fkimlik)
i.istihdam      _Iistihdam_1-3      (naturally coded; _Iistihdam_1 omitted)
Linear regression
```

		Number of obs	=	10,622
		F(13, 2789)	=	14.29
		Prob > F	=	0.0000
		R-squared	=	0.0270
		Root MSE	=	.3344

(Std. Err. adjusted for 2,790 clusters in fkimlik)

	gfd	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]
	mud	-.030048	.0144435	-2.08	0.038	-.0583689 -.0017271
	t2	.0142455	.0091009	1.57	0.118	-.0035997 .0320908
	t3	.0148164	.0089425	1.66	0.098	-.0027181 .0323509
	t4	.020512	.0244102	0.84	0.401	-.0273519 .0683759
	t5	.0273709	.0096798	2.83	0.005	.0083906 .0463511
	mudt2	.0218268	.0199664	1.09	0.274	-.0173235 .0609772
	mudt3	.023433	.0199506	1.17	0.240	-.0156864 .0625524
	mudt4	-.0527852	.0431591	-1.22	0.221	-.1374121 .0318418
	mudt5	.2417644	.024724	9.78	0.000	.1932853 .2902435
	yas	.0011214	.0003169	3.54	0.000	.0005001 .0017428
	cinsiyet	-.0134079	.0080571	-1.66	0.096	-.0292064 .0023906
	_Iistihdam_2	.0891481	.0514143	1.73	0.083	-.0116658 .189962
	_Iistihdam_3	.065909	.0251955	2.62	0.009	.0165053 .1153127
	_cons	.064799	.0144411	4.49	0.000	.0364827 .0931152

```
test mudt2=mudt3=mudt4=0
```

- ```
(1) mudt2 - mudt3 = 0
(2) mudt2 - mudt4 = 0
(3) mudt2 = 0
```

```
F(3, 2789) = 1.48
Prob > F = 0.2166
```

Tablo 17.1.7’de görüldüğü gibi *mudt5*’in katsayısı istatistiksel olarak anlamlıdır. Gecikmeli FF değişkenlerinin ortak olarak anlamlı olmadıklarını söyleyen sıfır hipotezi de reddedilmemiştir ( $p=0.2166$ ). Bu nedenle paralel eğilim varsayımının geçerli olduğunu söylemek mümkündür. Alternatif olarak Logit Modeli temelinde yapılan ve benzer sonuçlar elde edilen LAGS testi [17.1.9] nolu komut dizini ile gerçekleştirilmiş ve sonuçları Tablo 17.1.8’de sunulmuştur.

```
xi:logit gfd mud t2 t3 t4 t5 mudt2 mudt3 mudt4 mudt5 yas cinsiyet i.istihdam,
cluster(fkimlik)
```

```
test mudt2=mudt3=mudt4=0
```

[17.1.9]



**Tablo 17.1.8.** Logit Modeli Temelinde LAGS Testi

```

xi:logit gfd mud t2 t3 t4 t5 mudt2 mudt3 mudt4 mudt5 yas cinsiyet i.istihdam, cluster(fkimlik)
i.istihdam _Iistihdam_1-3 (naturally coded; _Iistihdam_1 omitted)
Iteration 0: log pseudolikelihood = -4149.8474
Iteration 1: log pseudolikelihood = -4060.2523
Iteration 2: log pseudolikelihood = -4033.9354
Iteration 3: log pseudolikelihood = -4033.909
Iteration 4: log pseudolikelihood = -4033.909
Logistic regression Number of obs = 10,622
 Wald chi2(13) = 259.18
 Prob > chi2 = 0.0000
Log pseudolikelihood = -4033.909 Pseudo R2 = 0.0279
 (Std. Err. adjusted for 2,790 clusters in fkimlik)

```

|              |     | Robust    |           |        |       |                      |           |
|--------------|-----|-----------|-----------|--------|-------|----------------------|-----------|
|              | gfd | Coef.     | Std. Err. | z      | P> z  | [95% Conf. Interval] |           |
| mud          |     | -.3510566 | .1852312  | -1.90  | 0.058 | -.7141031            | .0119899  |
| t2           |     | .1396787  | .0890213  | 1.57   | 0.117 | -.0347998            | .3141573  |
| t3           |     | .1446938  | .087385   | 1.66   | 0.098 | -.0265775            | .3159652  |
| t4           |     | .1969591  | .2220561  | 0.89   | 0.375 | -.2382628            | .6321811  |
| t5           |     | .2561476  | .0912778  | 2.81   | 0.005 | .0772463             | .4350488  |
| mudt2        |     | .2736018  | .2214545  | 1.24   | 0.217 | -.160441             | .7076446  |
| mudt3        |     | .2882794  | .222932   | 1.29   | 0.196 | -.1486593            | .7252181  |
| mudt4        |     | -.7494528 | .7804389  | -0.96  | 0.337 | -2.279085            | .7801795  |
| mudt5        |     | 1.568422  | .2033241  | 7.71   | 0.000 | 1.169914             | 1.96693   |
| yas          |     | .0100521  | .0028259  | 3.56   | 0.000 | .0045135             | .0155907  |
| cinsiyet     |     | -.1205643 | .0734364  | -1.64  | 0.101 | -.264497             | .0233685  |
| _Iistihdam_2 |     | .6478782  | .3067854  | 2.11   | 0.035 | .0465899             | 1.249166  |
| _Iistihdam_3 |     | .4912438  | .1641274  | 2.99   | 0.003 | .16956               | .8129276  |
| _cons        |     | -2.504267 | .137791   | -18.17 | 0.000 | -2.774332            | -2.234201 |

```

test mudt2=mudt3=mudt4=0
(1) [gfd]mudt2 - [gfd]mudt3 = 0
(2) [gfd]mudt2 - [gfd]mudt4 = 0
(3) [gfd]mudt2 = 0
 chi2(3) = 3.70
 Prob > chi2 = 0.2952

```

Logit modeli temelli LAGS testi ile de sıfır hipotezi reddedilemediğinden ( $p=0.2952$ ), gecikmeli FF değişkenlerinin ortak olarak anlamsız oldukları ortaya koyulmuştur. Bu sonuç paralel eğilim varsayımının sağlandığını göstermektedir. Varsayımının gerçekleşmiş olması ise Tablo 17.1.4, Tablo 17.1.5 ve Tablo 17.1.6'da sunulan tahminlerin geçerli olduğunu, “İzindeyken İşini Korumaya” politikasından yararlanan bireylerin kontrol grubundaki bireylere nazaran gelecekteki finansal durumları hakkında daha kötümser olduklarını ifade etmektedir.

#### 17.1.2.4.2. Grafiksel Test

Paralel eğilim varsayımının geçerli olup olmadığı grafiksel gösterim yoluyla da test edilebilir. Ancak daha önce belirtildiği gibi, LAGS testi bağımlı değişkeni etkileyebilecek diğer faktörlerin kullanılan modelde kontrol edilmesine imkân tanıdığından daha güvenilirdir. Grafiksel test ise, paralel eğilim varsayımın geçerliliği konusunda sadece kabaca fikir vermektedir. Grafiksel test için [17.1.10] nolu komut dizini kullanılmıştır.

**label define mud 1 "Müdahale" 0 "Kontrol"**

**label values mud mud1**

**lgraph gfd yıl , by(mud) error(ci(95)) stat(mean) ytitle("Ortalama")**

**xtitle("Yıl") xline(2019) ylab(, nogrid) scheme(s2mono)**

[17.1.10]<sup>14</sup>

[17.1.10] nolu komut dizininin ilk iki satırında *mud* değişkeni kullanılarak, Şekil 17.1.2’de gösterilen “müdahale (politikadan yararlanan)” ve “kontrol (politikadan yararlanmayan)” grupları isimlendirilmiş ve tanımlanmıştır. 3. satırında **lgraph** komutu ile politika öncesi ve sonrası dönemde müdahale ve kontrol grupları için, gelecek dönem finansal durum algısı (*gfd*) değişkeninin aldığı ortalama değerler çizilmiştir. Grafiğin X eksenı “yıl” değişkeninin aldığı değerleri, Y eksenı *gfd*’nin ortalama değerlerini göstermektedir. Veri seti çok dönemli bir mikro veri seti olduğundan, her bir yıl için *gfd*’nin ortalaması **stat(mean)** komutu ile elde edilmiştir. **xline(2019)** komutu ile 2019 yılı bitimini gösteren bir eşik çizgisi dikey olarak X eksenine çizilmiştir. Politika 2020 yılında uygulandığından politikanın etkisi eşik çizgisinin sağında görülmektedir. **ylab(, nogrid)** komutu her bir bireyin her bir yıl için aldığı *gfd* değerini grafikte noktalar (grid) hâlinde göstermemek için kullanılmıştır. 10000’den fazla gözlem ile çalışıldığından her bir gözlem için grafikte bir noktanın yer alması, grafiğin açık bir şekilde yorumlanmasına engel olmaktadır. **scheme(s2mono)** komutu grafikte ortalama değerleri birleştiren düz ve kesikli çizgilerin rengi (siyah) ve grafiğin arka fonu ile ilgili yapılmış olan seçim için kullanılmıştır<sup>15</sup>.

Şekil 17.1.2’de görüldüğü gibi müdahale grubu için 2020 yılı ortalama *gfd* değeri 2019 yılı ortalama *gfd* değerine göre çok daha yüksek bir değer aldığından, politika ile birlikte gelecek dönem finansal durum algısında olumsuz yönde bir artış yaşandığı söylenebilir. Kontrol grubu için böyle bir değişiklik ise söz konusu değildir. Politika öncesi yıllardaki durum incelendiğinde, müdahale ve kontrol gruplarına ait *ortalama gfd* değişkenlerinin paralel bir eğimle hareket ettiklerinden bahsedilebilir. İstisnai olarak sadece 2018 yılı değerinden 2019 yılı değerine geçişte paralellik biraz bozulmuş gibi görünmektedir. Ancak daha önce belirtildiği gibi grafiksel testte politika dışında hedef çıktığı etkileyebilecek diğer faktörler (istihdam, yaş ve cinsiyet) kontrol edilmediğinden, sadece grafiksel testin bulgularına güvenilmemeli ve mutlaka LAGS testi ile bir sınamaya yapılmalıdır. 17.1.2.4.1 nolu bölümde yapılan LAGS testlerinde 2019 yılı FF değişkeninin (*mudt4*) katsayısı anlamsız bulunduğu ve politika öncesi tüm yıllara ait FF değişkenlerinin ortak anlamlılıkları reddedildiğinden, paralel eğilim varsayımının ve dolayısıyla politika etkisine ilişkin elde edilen bulgunun hâlâ geçerli olduğu söylenebilir. Tabii ki grafiksel gösterimde en

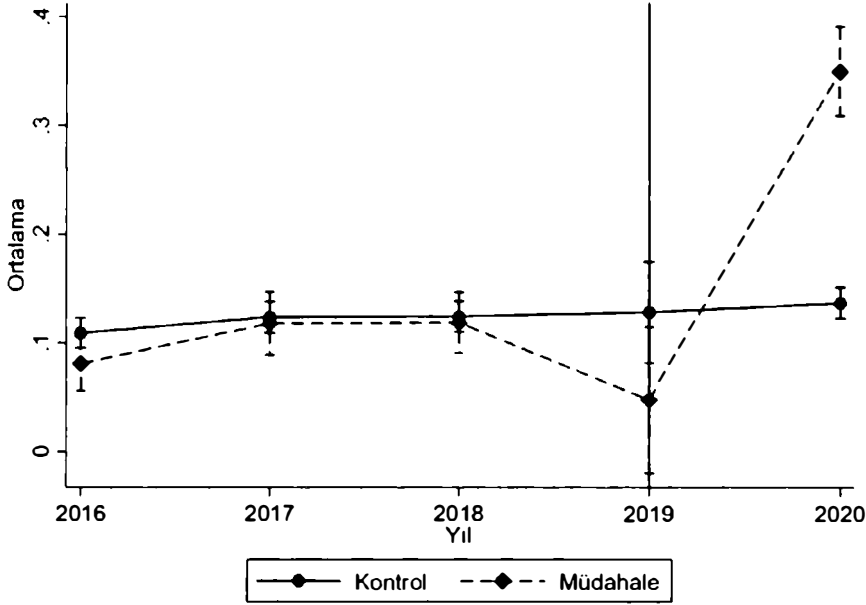
<sup>14</sup> Şekil 17.1.2’deki grafiğin elde edilmesinde kullanılan **lgraph** komutu indirilmeden çalışmayabilir. Stata

15’de komut penceresine **ssc install lgraph** yazarak bu komut indirilmelidir.

<sup>15</sup> Farklı tasarımlar için bknz: <https://www.stata.com/manuals/l3/g-4schemes2.pdf>

arzu edilen durum, politika öncesi yıllarda iki gruba ait hedef çıktı değerlerinin tümüyle paralellik göstermesidir.

**Şekil 17.1.2.** Politika Öncesi ve Sonrası Dönemde Müdahale ve Kontrol Grupları için Ortalama *gfd* Değerleri



### 17.1.3. Farkların Farkı Analizinde Kullanılan Stata Komutları

17.1 nolu bölümde kullanılan Stata komutları aşağıda toplu olarak yer almaktadır.

```
clear all
set more off
global in_dir "DOSYAKONUMU\"
use "${in_dir}Covid.dta", clear
ssc install lgraph
describe
sum
gen covd=.
replace covd=1 if yil==2020
replace covd=0 if yil<2020
gen mudcovd=mud*covd
xi: reg gfd mud covd mudcovd cinsiyet yas i.istihdam i.yil , cluster(fkimlik)
xi: logit gfd mud covd mudcovd cinsiyet yas i.istihdam i.yil , cluster(fkimlik)
margins, dydx(*)
tab yil, gen(t)
gen mudt1=mud*t1
gen mudt2=mud*t2
gen mudt3=mud*t3
gen mudt4=mud*t4
gen mudt5=mud*t5
xi: reg gfd mud t2 t3 t4 t5 mudt2 mudt3 mudt4 mudt5 yas cinsiyet i.istihdam,
cluster(fkimlik)
```

```

test mudt2=mudt3=mudt4=0
xi:logit gfd mud t2 t3 t4 t5 mudt2 mudt3 mudt4 mudt5 yas cinsiyet i.istihdam,
cluster(fkimlik)
test mudt2=mudt3=mudt4=0
label define mud 1 "Müdahale" 0 "Kontrol"
label values mud mud1
lgraph gfd yil , by(mud) error(ci(95)) stat(mean) ytitle("Ortalama") xtitle("Yıl")
xline(2019) ylab(, nogrid) scheme(s2mono)

```

## 17.2. EĞİLİM SKORU EŞLEŞTİRME YÖNTEMİ

### 17.2.1. Eğilim Skoru Eşleştirme Yönteminin Teorik Çerçevesi

17.1. nolu bölümde anlatılan Farkların Farkı (FF) yöntemiyle bir politikanın etkisinin, *ortalama müdahale etkisi* (average treatment effect-ATE) tahmin edilerek ölçüldüğü belirtilmişti. Eğilim Skoru Eşleştirme (ESE) yöntemi ise sadece politikaya maruz kalanlar yani sadece *müdahale görenler üzerinden ortalama müdahale etkisini* (average treatment on the treated-ATT) tahmin ederek politikanın etkisini ölçmektedir. ATT, politikadan yararlanmaya hak kazanan ve politikadan yararlanan bireylerin ortalama hedef çıktısı ile bu bireyler politikadan yararlanmamış olsaydı elde edilecek ortalama hedef çıktı arasındaki farkı göstermektedir. (17.2.1) nolu denklemde gösterildiği gibi,  $D_i$ 'nin 1 olması bireylerin politikadan yararlanma hakkının olduğunu göstermektedir.  $E(Y_i^T | D_i = 1)$ ; bu bireylerin politikadan yararlandıkları durumda elde edecekleri ortalama hedef çıktı değeri iken,  $E(Y_i^C | D_i = 1)$ ; politikadan yararlanma hakkı olan aynı bireylerin politikadan yararlanmadığı durumdaki ortalama hedef çıktı düzeyini gösterir. T (treated); politikadan yararlanılmış, C (counterfactual); politikadan yararlanılmamış olunan durumu simgelemektedir.

$$ATT = E(Y_i^T | D_i = 1) - E(Y_i^C | D_i = 1) \quad (17.2.1)$$

Ancak politikadan yararlanmaya hak kazanan bireyler (müdahale grubu) aynı anda hem politikadan yararlanmayı hem de yararlanmamayı seçemeyeceğinden, politikadan yararlandıkları durumda sahip olacakları hedef çıktı değerleri gözlemlenebilirken yararlanmadıkları durumdaki hedef çıktı değerleri (karşıt gerçeklik-counterfactual) gözlemlenemeyecektir. Böyle bir durumda politikanın etkisi de hesaplanamayacaktır. *Seçim yanlılığı* gibi bir problem olmasaydı, politikadan yararlanma hakkı olmayan bireylere (kontrol grubu) zaten politika etki etmeyeceğinden, eşitlik (17.2.2)'de gösterildiği gibi müdahale grubunun politikadan yararlanmadığı durumdaki ortalama hedef çıktı düzeyi kontrol grubundaki bireylerin hedef çıktı ortalamasına eşit olabilecekti.

$$E(Y_i^C | D_i = 1) = E(Y_i^C | D_i = 0) \quad (17.2.2)$$

Böylece politikadan yararlanma hakkı olmayan kontrol grubu bireylerinin ortalama hedef çıktı değeri olan  $E(Y_i^C | D_i = 0)$  gözlemlenebilir ve ölçülebilir olduğundan, ATT (17.2.3) nolu denklemdeki gibi rahatlıkla hesaplanabilirdi.

$$ATT = E(Y_i^T | D_i = 1) - E(Y_i^C | D_i = 0) \quad (17.2.3)$$

(17.2.2) nolu denklemde olduğu gibi, politikadan yararlanmayı hak eden bireylerin politikadan yararlanmadığı durumdaki hedef çıktıların, politikadan yararlanma hakkı olmayan bireylerin hedef çıktılarına eşit olabilmesi için bu iki grubun politika dışında hedef çıktıyı etkileyebilecek özellikler (dış faktörler) açısından çok benzer olması gerekmektedir. Aksi halde seçim yanlılığı ortaya çıkabileceğinden (17.2.2) nolu eşitlik geçerli olmayacaktır. Politika analizlerinde çok sıklıkla karşılaşılan ve içsellik probleminin bir türü olan seçim yanlılığı problemi (17.2.2) nolu eşitliğin yazılabilmesini engellemektedir. Örneğin hedef çıktının ( $Y_i$ ) gelecek dönem finansal durum algısı olduğu 17.1.2 nolu bölümdeki Stata uygulamasında, “İzindeyken İşini Koruma” programından<sup>16</sup> COVID-19 nedeniyle işine gidemeyen bireyler yararlanmaktadır. Programa katılma hakkı olmayanlar ise yaptığı işin türü nedeniyle COVID-19 sürecinin istihdam durumlarını etkilemediği bireylerdir. Programdan yararlanma hakkı olan bireylerin yararlanmadığı durumdaki ortalama finansal durum algıları  $E(Y_i^C | D_i = 1)$  ile programdan yararlanma hakkı olmayan bireylerin ortalama finansal durum algılarının  $E(Y_i^C | D_i = 0)$  eşit olabilmesi için, seçim yanlılığı probleminin olmaması gerekmektedir. Bunun için politikadan yararlanmaya hak sahibi olma durumu ile finansal durum algısı değişkeni arasındaki ilişkinin politika dışında diğer hiç bir faktörden etkilenmemesi gerekmektedir. Örneğin programa hâlihazırda düşük ücretli işlerde çalışanların istihdam edildiği firmalar başvurursa ve tersine kontrol grubu daha iyi ücretli bireylerden oluşursa, bireylerin gelecekteki finansal durumları hakkında kötümser olması politikadan değil düşük ücretli olmaktan kaynaklanabilir. Bu durum seçim yanlılığı nedeniyle başka bir faktörün etkisinin politikanın etkisiymiş gibi algılanması yani etkinin politikaya atfedilmesi problemidir. COVID-19’un istihdam açısından evden (uzaktan) esnek çalışmaya ayak uyduramayan iş türlerini etkilediği düşünüldüğünde, yaş ve yetenekleri uzaktan çalışmaya uygun olmayan bireylerin programdan yararlanması ve işçilerin yaş ve yetenekleri nedeniyle gelecek dönem finansal durum algılarının olumsuz olma ihtimali de bu duruma örnektir.

Rubin (1977) *koşullu bağımsızlık varsayımı* (conditional independence assumption-CIA) ile politikadan yararlanma statüsü ve hedef çıktı arasındaki ilişkinin, bireylerin benzer karakteristikler ve özellikler sergilemesi koşulu ile bağımsız olabileceğini kanıtlamıştır. Bireyler farklı özellikler taşırsa (bir grubun yaşça büyük ve uzaktan çalışmada daha yeteneksiz olması gibi), bu özellikler politikadan yararlanma statüsünü değiştirmek yoluyla hedef çıktıyı etkileyecektir. Bu varsayım (17.2.4) nolu gösterimle ifade edilebilir.

$$Y_i^T, Y_i^C \perp D_i | X_i = x \quad (17.2.4)$$

CIA temelli çalışan ESE yöntemi bahsi geçen karakteristikler, özellikler veya dış faktörler açısından birbirine çok benzer (karşılaştırılabilir) kontrol ve müdahale gruplarını elde etmek ve bu yolla politikanın etkisini ölçmek için kullanılır. Eşleştirme sonrası  $g$  kontrol grubundaki,  $i$  müdahale grubundaki bireyleri göstermek üzere *müdahale görenler üzerinden politikanın ortalama etkisi* (average treatment on the treated-ATT) (17.2.6) nolu eşitlikteki gibi hesaplanabilir.

$$E(Y_i^C | D_i = 1, X_i = x) = E(Y_i^C | D_i = 0, X_i = x) \quad (17.2.5)$$

$$\hat{\tau}_{ATT} = E(Y_i^T | D_i = 1, X_i = x) - E(Y_i^C | D_i = 0, X_i = x) \quad (17.2.6)$$

<sup>16</sup> “İzindeyken İşini Koruma” programının detaylı anlatımı 17.1.2 nolu bölümde yapılmıştır.

Eşleştirmede ideal olanı müdahale grubundaki her bir gözlemin kontrol grubundaki bir gözlemle eşleşmesidir. Böylece müdahale ve kontrol gruplarındaki bireyler özellik ve karakteristikler bakımından aynı olacağından, politika sonrasında bu iki grup arasında ortalama hedef çıktı açısından gerçekleşecek fark, politikanın etkisi olarak değerlendirilebilir. Ancak bazen müdahale grubundaki bireylerin kontrol grubunda eşleri bulunmayabilir ve böylece her iki grupta da eşleşmeden dışarıda kalan bireyler olabilir. Bu durum “*ortak destek alanının* (common support)” dışında kalmak olarak adlandırılır ve eşleşmeyen bireyler politika analizinde kullanılacak örnekleme alınamayacağından gözlem kaybı yaşanabilecektir. ESE yönteminin çok gözlemlili mikro verilere uygulanmasının nedeni de budur.

Rosenbaum ve Rubin (1983) çalışması takip edilerek, ESE yöntemi **üç aşamada** gerçekleştirilmektedir. **İlk aşamada** politikadan yararlanma durumunu gösteren kukla değişkeni bağımlı değişken, müdahale öncesi bireysel karakteristik ve özellikler bağımsız değişkenler olmak üzere probit ya da logit analizi gerçekleştirilir. Böylece her bir bireysel karakteristik için politikadan yararlanma durumu ile ilişkisini gösteren katsayı tahmin edilmiş olur. **İkinci aşamada** kontrol ve müdahale gruplarındaki her bir birey için politikadan yararlanma olasılığı (eğilim skor-propensity score) tahmin edilir ve son aşamada bu olasılık değerleri kullanılarak kontrol ve müdahale grubundaki bireyler eşleştirilir. Eşleştirme sonrası (17.2.6) nolu eşitlik ile gösterilen ATT hesaplanır. ATT, *eşleştirme tahmincisi* (matching estimator) olarak da adlandırılır.

**Üçüncü aşamada** gerçekleştirilen eşleştirme farklı tekniklerle yapılabilmektedir. En bilinen eşleştirme türü ‘*en yakın komşu eşleştirmesi* (nearest neighbour matching)’dir. Bu metod önce müdahale ve kontrol grubundaki bireyleri rassal olarak sıralayarak müdahale grubundan bir birey seçmekte ve bu bireyi eğilim skoruna en yakın skora sahip kontrol grubu bireylerinden bir veya daha fazlası ile eşleştirmektedir. Eğer müdahale grubundaki bir birey kontrol grubundan eğilim skoruna en yakın skora sahip tek bir birey ile eşleştirilirse bu tür komşu eşleştirmesinin özel adı *bire-bir eşleştirme* (one-to-one matching)’dir. Bire-bir eşleştirme yapmak tercih edilmezse eğilim skoruna yakınlığına göre müdahale grubundan bir birey, kontrol grubundan birden fazla komşusu ile eşleştirilebilir. Örneğin müdahale grubundaki bireyin eğilim skoruna en yakın iki veya üç bireyin kontrol grubundan seçildiği durumda, bu eşleştirme yöntemi *bire-çok eşleştirme* (many-to-one matching) olarak adlandırılmaktadır.

En yakın komşu eşleştirmesi (EYKE) herhangi bir eşleştirmede kullanılan bir kontrol grubu bireyini başka bir eşleştirmede kullanılmak üzere “*yerine koyma*” (replacement)” veya “*yerine koymama*” (no replacement) yolu ile uygulanabilir. Yerine koyma ile yapılan en yakın komşu eşleştirmesinde, daha önce bir eşleştirmede kullanılan kontrol grubu üyesi tekrar başka bir eşleştirmede kullanılmak üzere yeniden seçilebilir. Smith ve Todd (2005), yerine koyma yoluyla EYKE yapmanın eşleştirme kalitesini yükselteceğini ve seçim yanlılığını azaltacağını belirtmiştir. Seçim yanlılığını azaltmak benzer karakteristiklere sahip bireylerin müdahale ve kontrol grubunda olması anlamına geldiğinden, yerine koyma yolu ile karakteristikler açısından benzer bireylerin eşleşme şansı artmaktadır. Yerine koyma yoluyla EYKE yapmanın tek dezavantajı politikanın etkisini gösteren ATT tahmincisinin varyansının yüksek çıkmasına neden olmasıdır. ATT tahmincisinin varyansı Lechner (2001)’in önerdiği şekliyle eşitlik (17.2.7)’deki gibi yazıldığında, kontrol grubundaki birey  $g$ ’nin farklı eşleşmelerde kullanıldığı sayı ( $w_g$ ) arttıkça varyansın büyüdüğü görülmektedir.  $N_1$  eşleşen müdahale grubu birey sayısını temsil etmektedir.

$$Var(\hat{\tau}_{ATT}) = \frac{1}{N_1} Var(Y(1)|D = 1) + \frac{(\sum_{g \in I_0} (w_g)^2)}{(N_1)^2} Var(Y(0)|D = 0) \quad (17.2.7)$$

Yerine koyma ya da yerine koymama yolu ile EYKE uygulamasında karşılaşılabilecek en genel risk, en yakın komşunun çok uzakta olduğu durumdur. Böyle bir durumda eşleştirme performansı iyi olmayacaktır. Bu nedenle uygulamalarda eğer EYKE ile çalışılacaksa bir *kaliper* veya *eğilim aralığı* (caliper) kullanılmaktadır. Bu aralık, müdahale ve kontrol bireyinin eğilim skorları arasındaki fark için belirlenen maksimum tolerans düzeyidir. Rosenbaum ve Rubin (1983) bu aralığı 0.25 olarak belirlemiş olup, literatürdeki pek çok yayın bu aralığa yakın değer seçmek ile birlikte daha küçük tolerans aralıklarının daha kaliteli sonuçlar verdiğini belirtmektedir. Belli bir eğilim aralığında en yakın komşu eşleştirmesi yapılmakta olup bu eşleştirme bazen *kaliper* ya da *eğilim aralığı eşleştirmesi* (caliper matching) olarak da adlandırılmaktadır. Böylece sadece en yakın komşu seçilmemekte, kontrol grubunda belli bir aralığa düşen en yakın komşu ile eşleştirme yapılmış olmaktadır.

Dehejia ve Wahba (2002), eğilim aralığı ya da kaliper eşleştirmesinin bir türü olan *yarıçap eşleştirmesini* (radius matching) geliştirmiştir. Bu yöntemde müdahale grubundaki bireyin sadece belli bir eğilim aralığı içindeki en yakın komşu ile değil, o aralıkta bulunan tüm uygun kontrol grubu bireyleri ile eşleşme şansı olmaktadır. Yöntem bu şekilde iyi eşleşmeler varsa daha çok kontrol grubu üyesinin eşleşmeye katılmasını sağlayarak hatalı eşleşme riskini de önlemektedir.

ESE yöntemlerinden en yaygın olarak kullanılan bir diğer eşleştirme türü *kernel eşleştirmesi* (kernel matching)'dir. *Çekirdek eşleştirmesi* olarak da adlandırılır. Daha önce bahsedilen eşleştirmelerden farklı olarak parametrik olmayan bir eşleştirme türüdür. Eşleştirmede her bir müdahale bireyi için birden fazla kontrol bireyi, müdahale gözleminden uzaklaştıkça azalan bir ağırlık alacak şekilde kullanılmaktadır. Yakın olanlar daha yüksek ağırlık almakta yani ağırlıklar müdahale grubu bireylerinin kontrol grubu bireyleri arasındaki eğilim skoru farkı ile ters orantılı olmaktadır (Becker ve Ichino, 2002). Ağırlık, (17.2.8) nolu eşitlikte tanımlanmıştır (Eren, 2007).

$$g_{ij} = \frac{K\left(\frac{p_i - p_j}{h}\right)}{\sum_{j \in C(i)} K\left(\frac{p_i - p_j}{h}\right)} \quad (17.2.8)$$

$K(\cdot)$ , Gaussian ve Epanechnikov gibi kernel fonksiyon türlerinden birini simgelerken,  $h$  bant genişliğidir. Bant genişliği, *düzeltilme parametresi* (smoothing parameter) olarak da adlandırılmakta olup kernel fonksiyonunun şeklini dolayısıyla bir ağırlık verilebilecek kontrol grubu gözlemlerinin düştüğü aralığı belirler. Küçük bant genişlikli kernel fonksiyonlarında müdahale grubu bireylerinin eğilim skorlarına yakın değerdeki eğilim skorlu kontrol grubu bireyleri eşleştirmeye dahil edilirken, bant genişliği büyük değerler alan kernel fonksiyonlarında eğilim skoru açısından müdahale grubu bireylerine daha az benzeyen kontrol grubu üyeleri eşleştirmede kullanılır. Çok küçük bant genişliği eşleştirmede kullanılacak yeterli sayıda kontrol grubu bireyi bulamama riskini doğururken, çok büyük bant genişlikleri müdahale grubu üyelerine bireysel karakteristikler açısından çok benzemeyen kontrol grubu bireylerinin eşleştirmeye dahil olmasına, böylece kaliteli bir eşleştirmenin yapılamamasına sebebiyet verir. Bu nedenle kernel eşleştirmesinde önemli

olan en uygun bant genişliğine karar vermektir<sup>17</sup>. Fonksiyon türünün ne olduğunun çok önemli olmadığı literatürde belirtilmekle birlikte (DiNardo ve Tobias, 2001), bant genişliğinin önemi vurgulanmıştır. Yüksek değerdeki bant genişliğinin düzeltmeyi başarı ile yaparak ATT tahmincisinin varyansını düşürdüğü, ancak seçim yanlılığının artmasına sebep olduğu belirtilmiştir (Caliendo ve Kopeinig, 2008).

Görüldüğü gibi ESE yöntemleri çok çeşitli olmakla birlikte, eşleştirme eğilim skoru temelli yapılmadan önce literatürde en yaygın kullanılan eşleştirme türü *mahalanobis metrik eşleştirmesiydi* (mahalanobis metric matching) (D'Agostino, 1998). Bu eşleştirme eğilim skoru yerine mahalanobis metrik uzaklığı temelinde uygulanmaktadır. Yöntemde müdahale grubu bireyleri ile kontrol grubu bireyleri arasındaki (17.2.9) nolu eşitlik ile gösterilen Mahalanobis metrik uzaklığı hesaplanmaktadır (Rubin, 1980).

$$d(i, g) = (X_i - X_g)'C^{-1}(X_i - X_g) \quad (17.2.9)$$

$X_i$  ve  $X_g$  sırasıyla müdahale grubu bireyi  $i$  ve kontrol grubu bireyi  $g$  için eşleştirmede kullanılan bağımsız değişkenler matrisidir.  $C$ , tüm örneklemede eşleştirmede kullanılan bağımsız değişkenlerin kovaryans matrisidir. Her bir müdahale ve kontrol grubu bireyi, hesaplanan en küçük mahalanobis uzaklığına göre eşleştirilir. Ayrıca karma bir yöntem olarak Rosenbaum ve Rubin (1985), müdahale ve kontrol bireyleri için logit modeli ile eğilim skorlarını tahmin ederek bunları eşleştirmede kullanılan bağımsız değişkenler seti içerisine almış ve bu şekilde hesapladığı mahalanobis metrik uzaklığını eşleştirmede kullanmıştır. Bu yöntem *mahalanobis metrik eşleştirmesi ile eğilim skorlarının birleştirilmesi* olarak adlandırılmaktadır.

Mahalanobis metrik eşleştirmesi ile ESE yönteminin bir arada kullanıldığı bir diğer karma yöntem, *bir eğilim aralığında mahalanobis metrik eşleştirmesidir*. Burada rassal olarak sıralanan müdahale grubu bireylerinin ilki için belirlenen eğilim aralığında (kaliper) yer alan kontrol grubu üyelerinden bir alt küme oluşturulmaktadır. Ardından, bu alt küme içerisinde en küçük mahalanobis uzaklığında yer alan kontrol bireyi ile müdahale bireyi eşleştirilmektedir.

ESE yöntemlerinin tek, ama önemli dezavantajı, kontrol ve müdahale grupları arasındaki farkların yalnızca gözlemlenebilir karakteristiklerden kaynaklandığını varsaymasıdır. Bu nedenle sadece gözlemlenebilir karakteristikler temelinde eşleştirme yapılabilir. Ancak bu iki grup arasında gözlemlenemeyen karakteristiklerden kaynaklı farklar olabilir ve bu farklar politikadan yararlanma durumu ve hedef çıktı arasındaki ilişkiyi etkileyebilir. Bu yüzden ESE, gözlemlenemeyen karakteristiklerden kaynaklı seçim yanlılığı probleminde çözüm bulamamaktadır.

Kullanılan veri seti tek dönemlik yatay kesit veri seti olduğunda, Farkların Farkı Yöntemi ile politikanın etkisi tahmin edilemeyeceğinden, ESE politikanın etkisinin hesaplanmasında tek başına kullanılan bir yöntem olarak karşımıza çıkmaktadır. Ancak çok dönemli havuzlanmış yatay kesit veri seti veya panel veri setinin mevcut olması durumunda ESE ve FF yöntemlerini bir arada uygulamanın büyük avantajı vardır. Böyle bir durumda politikanın etkisi ESE ile ölçülmekte, seçim yanlılığı azalmış bir örneklem FF analizinde kullanılmak üzere ESE ile elde edilmektedir. Böylece ESE ile gözlemlenebilir değişkenlerden kaynaklı heterojenlik giderilebilirken,

<sup>17</sup> Kernel fonksiyonu ve bant genişliği hakkında daha detaylı bilgi için bkz. Cameron ve Trivedi (2005).



gözlemlenemeyen karakteristiklerden kaynaklı heterojenlikler de FF ile giderilmiş olmaktadır. Heckman vd. (1998) çalışması FF-ESE yöntemlerinin bir arada kullanılmasını öneren ilk çalışmalardan olup, bu bölümde iki yöntemin bir arada kullanılması anlatılmayacak olsa da literatürde iki yöntemin bir arada kullanıldığı pek çok çalışmaya yer almaktadır.

### 17.2.2. Eğilim Skoru Eşleştirme Yöntemi: Stata Uygulaması

Bu bölümde, 17.1.2. nolu bölümde kullanılan **Covid.dta** veri seti ile “İzindeyken İşini Koruma” politikasının bireylerin “gelecekteki finansal durum algıları (*gfd*)” üzerine etkisi farklı ESE yöntemleri ile analiz edilmiştir<sup>18</sup>. Analizler Stata 15 programında *psmatch2*<sup>19</sup> komutu kullanılarak gerçekleştirilmiştir.

17.1.2 nolu bölümde anlatıldığı gibi *mud* “İzindeyken İşini Koruma” programından yararlananlar için 1, yararlanmayanlar için 0 değerini almaktadır. ESE ile programa katılma durumu ve hedef çıktı (*gfd*) arasındaki ilişkiyi etkileyerek seçim yanlılığına neden olabilecek bireysel karakteristikler açısından müdahale ve kontrol grubu benzer kılınarak ATT hesaplanmıştır<sup>20</sup>. Bu değişkenler yaş, cinsiyet ve istihdam durumu olarak seçilmiştir<sup>21</sup>. Bu bölümde sırasıyla en yakın komşu eşleştirmesi, eğilim aralığı (kaliper) eşleştirmesi, yarıçap eşleştirmesi, kernel eşleştirmesi ve mahalnobis metrik eşleştirmesinin uygulamalarına yer verilmiştir. Çünkü en iyi eşleştirmeyi hangi yöntemin yaptığı konusunda genel bir kanı olmadığından, literatürdeki çalışmalar eşleştirme kalitesi yüksek birden fazla eşleştirme yöntemiyle elde edilen tüm sonuçları raporlamayı tercih etmektedir. Büyük örneklerle çalışıldığında, her bir eşleştirme yöntemi ile elde edilen sonuçlar genelde benzerlik göstermektedir (Smith, 2000). Ancak küçük örneklerle çalışıldığında ya da müdahale ve kontrol grubu üyelerinin eğilim skorları dağılımı benzerlik göstermediğinde eşleştirme yöntemleri farklı sonuçlar ortaya koyabilmektedir. Bu gibi durumlarda sonuçların sağlamlığından emin olmak için birden fazla eşleştirme yöntemi ile çalışılmaktadır (Caliendo ve Kopeinig, 2008). Farklı sonuçlar ortaya koyan eşleştirme yöntemlerinin olduğu durumda ise her bir eşleştirmenin sağlamlığı *pstest2* komutu ile sınanmaktadır.

#### 17.2.2.1. Yerine Koymama Yoluyla En Yakın Komşu Eşleştirmesi

17.2.1 nolu bölümde detayları ile bahsedildiği gibi en yakın komşu eşleştirmesinin (EYKE) en temel uygulaması *bire-bir eşleştirme*dir. Bu eşleştirme [17.2.1] nolu komut ile gerçekleştirilmiştir<sup>22</sup>.

<sup>18</sup> Covid.dta veri setindeki değişken tanımları ve bu değişkenlere ait tanımlayıcı istatistikler 17.1.2. nolu bölümde bulunabilir.

<sup>19</sup> Detaylı bilgi için bkz; <http://repec.org/bocode/p/psmatch2.html>

<sup>20</sup> Tamamen deneysel yöntemler ile toplanmış verilerde, programdan veya politikadan yararlananlar gözlenebilir veya gözlenemeyen bireysel karakteristiklerden bağımsız ve rassal olarak belirlendiğinden, seçim yanlılığı ortaya çıkmayacak böylece eşleştirme yapmaya gerek kalmadan politikanın etkisi ölçülebilecektir.

<sup>21</sup> Kitap bölümü ders kitabı niteliğinde olduğundan ve ESE'nin ampirik uygulamasını öğretmeyi hedeflediğinden, basitliği sağlamak adına sadece bu üç değişken açısından eşleştirme sağlanmış olup, *gfd* ve programa katılımı etkileme yoluyla seçim yanlılığına neden olabilecek farklı değişkenlerde eşleştirmede kullanılan modellere eklenebilirdi. Bir politikanın etkisinin analizine odaklanılmış akademik makalelerde bu değişkenler daha dikkatli seçilmelidir. Ancak bazen seçim yanlılığına neden olabilecek ama gözlenemeyen değişkenlerin de olduğu unutulmamalıdır.

<sup>22</sup> Stata'da *psmatch2* ile ilk kez çalışılıyorsa, komut penceresine “*ssc install psmatch2*” yazarak komut indirilmelidir.

**xi:psmatch2 mud cinsiyet yas i.istihdam if yıl==2020 , noreplacement caliper (0.1) out(gfd)**  
**[17.2.1]**

**noreplacement** opsiyonu EYKE'nin yerine koymama yoluyla yapıldığını göstermektedir. **out(gfd)** yada **outcome(gfd)** olarak da yazılabilecek opsiyon ile politikanın etkisi merak edilen hedef çıktı değişkeni (*gfd*) komutta belirtilmiştir. Ayrıca belli bir eğilim aralığında (kaliper) en yakın komşu eşleştirmesi ya da diğer adıyla eğilim aralığı eşleştirmesi yapılmak isteniyorsa **caliper( )** ile hangi aralıkta eşleştirme yapılmak istendiği eklenmelidir. 17.2.1. nolu bölümde anlatıldığı gibi Rosenbaum ve Rubin (1983) bu aralığı 0.25 olarak belirlemiş olup, literatürdeki pek çok yayın daha küçük aralık değerlerinin daha kaliteli sonuçlar verdiğini belirtmektedir. Bu örnekte eğilim aralığı 0.1 olarak seçilmiştir. Tablo 17.2.1 'de yerine koymama yoluyla yapılan bire-bir EYKE sonuçları gösterilmiştir.

**Tablo 17.2.1. Yerine Koymama Yoluyla Bire-Bir En Yakın Komşu Eşleştirmesi**

**xi:psmatch2 mud cinsiyet yas i.istihdam if yıl==2020 , noreplacement caliper (0.1) out(gfd)**

i.istihdam                    \_Iistihdam\_1-3                    (naturally coded; \_Iistihdam\_1 omitted)

|                             |               |   |        |
|-----------------------------|---------------|---|--------|
| Probit regression           | Number of obs | = | 2,745  |
|                             | LR chi2(4)    | = | 7.37   |
|                             | Prob > chi2   | = | 0.1176 |
| Log likelihood = -1337.4261 | Pseudo R2     | = | 0.0027 |

| mud          | Coef.     | Std. Err. | z      | P> z  | [95% Conf. Interval] |           |
|--------------|-----------|-----------|--------|-------|----------------------|-----------|
| cinsiyet     | .0194901  | .0562678  | 0.35   | 0.729 | -.0907929            | .129773   |
| yas          | .0048751  | .0022808  | 2.14   | 0.033 | .0004049             | .0093453  |
| _Iistihdam_2 | -.456244  | .3790538  | -1.20  | 0.229 | -1.199176            | .2866878  |
| _Iistihdam_3 | .1163664  | .1367073  | 0.85   | 0.395 | -.151575             | .3843078  |
| _cons        | -1.086806 | .1009236  | -10.77 | 0.000 | -1.284613            | -.8889993 |

| Variable | Sample    | Treated    | Controls   | Difference | S.E.       | T-stat |
|----------|-----------|------------|------------|------------|------------|--------|
| gfd      | Unmatched | .349809886 | .136547995 | .213261891 | .018079978 | 11.80  |
|          | ATT       | .349809886 | .129277567 | .220532319 | .025448653 | 8.67   |

Note: S.E. does not take into account that the propensity score is estimated.

|            |           |       |
|------------|-----------|-------|
|            | psmatch2: |       |
| psmatch2:  | Common    |       |
| Treatment  | support   |       |
| assignment | On suppor | Total |
| Untreated  | 2,219     | 2,219 |
| Treated    | 526       | 526   |
| Total      | 2,745     | 2,745 |

Tablo 17.2.1’de gösterilen probit modeli analiz sonuçları eşleştirmenin ilk aşama sonuçlarıdır. Politikadan yararlanma durumunu gösteren *mud* değişkeninin bağımlı değişken, *cinsiyet*, *yas* ve *istihdamın* bağımsız değişken olduğu probit modeli analiziyle her bir bireysel karakteristiğin politikadan yararlanma durumu ile ilişkisini gösteren katsayı tahmin edilmiştir<sup>23</sup>. Probit modeli analizi ile elde edilen katsayıları kullanarak Stata 15 arka planda *pscore* adlı yeni bir değişken oluşturmaktadır. *pscore*, her bir bireyin eşleştirilmesinde kullanılacak eğilim skorlarını gösteren değişkendir. Her bir bireye ait eğilim skorları ile eşleştirme yapıldıktan sonra tahmin edilen ATT (difference) değeri politikanın etkisini göstermektedir. Tablo 17.2.1’de görüldüğü gibi ATT (difference) değeri 0.220532319 olarak elde edilmiş olup, istatistiksel olarak anlamlıdır ( $t=8.67$ ). Sonuç “İzindeyken İşini Koruma” politikasından yararlanan bireylerin, yararlanmayan bireylere göre %22 daha yüksek olasılıkla gelecek dönem finansal durumları hakkında kötümser olduklarını ortaya koymaktadır.

Tablo 17.2.1’deki son bulgular, eşleştirme yapılmadan önce elde var olan genel örneklemden kaç tane müdahale grubu bireyinin, kaç tane kontrol grubu bireyinin eşleştirme sonucu elde edilen örnekleimde yani ortak destek (common support) alanında yer aldığını göstermektedir. Sonuçlara göre bu eşleştirme yöntemi dışarıda eşleşmeyen birey bırakmamış olup, tüm bireyler sorunsuz bir şekilde eşleşmiştir. Genel örnekleimde yer alan 2219 müdahale grubu bireyinin ve 526 kontrol grubu bireyinin hepsinin ortak destek alanında olduğu (on support) görülmektedir. Eşleştirilmeyen bireyin olması durumunda ortak destek alanında olanların sayısı (common support) toplam sayıdan (total) daha az olmaktadır. Böyle bir durumda ilgili tablo ortak destek alanının dışında kalanların sayısını da göstermektedir. Tüm bireylerin eşleştirilmiş olması ve eşleştirilmemiş bireyin kalmaması, eşleştirmenin kalitesini gösteren bir durum değildir. Tüm bireylerin eşleştirilmiş olması, elimizde olan genel örneklemdaki müdahale ve kontrol grubu bireylerinin eşleştirme yapılmadan önce yaş, cinsiyet ve istihdam durumu açısından zaten birbirine benzer olduğunu göstermektedir. Bu durum eşleştirme öncesi elde bulunan örneklemin eşleştirme yapılmadan da kullanılabileceğine işaret etmektedir. Bu nedenle burada, 17.1.2 nolu bölümde FF yöntemi ile elde edilen ve Tablo 17.1.4’te gösterilen politika etkisi sonucuna (%22.8) çok yakın bir sonuç elde edilmiştir. Bazen eşleştirme öncesi elde bulunan örneklemden pek çok birey eşleştirme sonrası ortak destek alanında kendisine yer bulamayabilir. Bu durum da eşleştirme kalitesinin iyi veya kötü olduğuna işaret etmez. Özetle Tablo 17.2.1’deki son bulgular kaç bireyin eşleştirme sonrası ortak alanda olduğunu göstermekten başka bir sonuca işaret etmemektedir. Eşleştirmenin kaliteli olup olmadığını, eşleştirme sonrası elde edilen örnekleimde müdahale ve kontrol grubundaki bireylerin birbirlerine ne kadar benzediğini test etmek için [17.2.1] nolu eşleştirme komutu sonrası [17.2.2] nolu **pstest2** komutu çalıştırılmıştır. Test sonuçları Tablo 17.2.2’de gösterilmektedir.

<sup>23</sup> Stata 15 varsayılan model olarak probit modelini kullanmakta olup, 17.2.2. nolu bölümde detayları ile anlatıldığı gibi probit modeli yerine logit modeli de bu aşamada kullanılabilir. Logit modelini kullanmak tercih edilirse ilgili komutu **xi:psmatch2 mud cinsiyet yas i.istihdam if yıl==2020 , noreplacement caliper (0.1) out(gfd) logit** olarak çalıştırmak yeterli olacaktır.

## xi:pstest2 cinsiyet yas i.istihdam

[17.2.2]

Tablo 17.2.2. Yerine Koymama Yoluyla Bire-Bir En Yakın Komşu Eşleştirmesi pstest2 Sonuçları

xi:pstest2 cinsiyet yas i.istihdam

i.istihdam \_Iistihdam\_1-3 (naturally coded; \_Iistihdam\_1 omitted)

| Variable     | Sample    | Mean    |         | %reduct |       | t-test |       |
|--------------|-----------|---------|---------|---------|-------|--------|-------|
|              |           | Treated | Control | %bias   | bias  | t      | p> t  |
| cinsiyet     | Unmatched | .41255  | .40063  | 2.4     |       | 0.50   | 0.617 |
|              | Matched   | .41255  | .42205  | -1.9    | 20.2  | -0.31  | 0.755 |
| yas          | Unmatched | 42.787  | 41.483  | 10.5    |       | 2.24   | 0.025 |
|              | Matched   | 42.787  | 42.981  | -1.6    | 85.1  | -0.24  | 0.812 |
| _Iistihdam_2 | Unmatched | .0038   | .00856  | -6.1    |       | -1.13  | 0.260 |
|              | Matched   | .0038   | .0038   | 0.0     | 100.0 | -0.00  | 1.000 |
| _Iistihdam_3 | Unmatched | .04753  | .0374   | 5.0     |       | 1.07   | 0.283 |
|              | Matched   | .04753  | .03802  | 4.7     | 6.1   | 0.76   | 0.447 |

Tablo 17.2.2’de eşleştirmede kullanılan bireysel karakteristikleri temsil eden değişkenlerin müdahale (treated) ve kontrol (control) grupları için ortalama değerleri (mean) rapor edilmiştir. *cinsiyet* değişkeni veri setinde kadınlar için 0, erkekler için 1 değerini aldığından, eşleşmemiş (unmatched) örneklemdaki müdahale grubunun %41.255’inin, kontrol grubunun ise %40.063’ünün erkeklerden oluştuğu görülmektedir<sup>24</sup>. Bu sonuç referans kategori olan kadınlar için genelleştirilirse, eşleşmemiş örneklemdaki müdahale grubunun %58.745’ini<sup>25</sup>, kontrol grubunun ise %59.937<sup>26</sup>’sini kadınların oluşturduğu söylenebilir. Eşleşmemiş örneklemdaki *cinsiyet* değişkeni açısından seçim yanlılığı (bias) sadece %2.4 ile ölçülmüştür. Bu sonuç kadın ve erkeğin müdahale ve kontrol gruplarındaki dağılımının eşleştirme yapılmamış örneklemden zaten benzer olduğunu göstermektedir. Benzer olduğunu söyleyen sıfır hipotezi 0.5 değerindeki t istatistiği ve p=0.617 değeri ile reddedilmemiştir. *cinsiyet* değişkeni için elde edilen bu sonuçlara göre eşleştirmenin yapılmadığı durumda da müdahale ve kontrol grubu bireylerinin kadın ve erkek dağılımı açısından benzer olduğu görülse de, diğer tüm bireysel karakteristikler açısından bu durum geçerli değilse eşleştirme yapılmasına ihtiyaç vardır. Örneğin eşleştirme yapılmamış örneklemden müdahale ve kontrol gruplarındaki bireylerin ortalama yaşları arasındaki benzerliğin 2.24 değerindeki t istatistiği ve p=0.025 değeri ile reddedildiği görülmektedir. Eşleşmiş örneklemden (matched) ise ortalama yaşlara göre bu iki grubun benzer olduklarını söyleyen sıfır hipotezi reddedilmemiştir (t=-0.24 ve p=0.812). Eşleştirme sayesinde yaşlar açısından bu iki grup birbirine

<sup>24</sup> Stata, Türkçe dilinde çalışan bir program olmadığından ondalık ayracı nokta (.) binlik ayracı ise virgül (,) olarak kullanılmaktadır. Bu bölümde Stata çıktıları yorumlanırken orijinaline sadık kalınarak sayısal ifadeler türkçeleştirilmemiş olup, binlik ayracı olarak virgül ve ondalık ayracı olarak nokta kullanılmıştır.

<sup>25</sup> 100-41.255=58.745

<sup>26</sup> 100-40.063=59.937

benzer kılınmış ve seçim yanlılığı %85 azalmıştır. Eşleşmiş örnekleme müdahale ve kontrol grubunun ortalama yaşları sırasıyla 42.7 ve 42.9 olup neredeyse aynıdır. Cinsiyet gibi istihdamın kategorileri açısından da eşleştirme öncesi elde bulunan örnekleme benzer kontrol ve müdahale grupları olmakla birlikte, eşleştirme sonrası bu iki grup özellikle kendi hesabına çalışma kategorisi (**\_listihdam\_2**) açısından birebir aynı olmuştur. Hem kendi hesabına hem de memur/işçi olarak çalışma kategorisi (**\_listihdam\_3**) açısından eşleştirme sonrası seçim yanlılığının azaldığı görülmektedir<sup>27</sup>. Özetle, t istatistiği ve p değerlerine göre eşleşmemiş örnekleme (unmatched) sadece *cinsiyet* ve *istihdam durumu* değişkenleri açısından ama eşleştirilmiş örnekleme (matched) yaş dâhil tüm bireysel karakteristikler açısından müdahale ve kontrol gruplarının benzer oldukları reddedilmemektedir. t istatistik değerleri ve p değerlerinin yanı sıra bunu görmenin bir diğer yolu %seçim yanlılığı (%bias) değerine bakmaktır. Rosenbaum ve Rubin (1985), %10'un altında olan %seçim yanlılığı değerinin eşleştirmenin iyi yapıldığına işaret ettiğini belirtmiştir. Kullanılan değişkenler açısından eşleştirme sonrası %seçim yanlılığı değerlerinin hepsi %10'un altındadır. t istatistiği, değişkenlerin normal dağıldığını varsayması nedeniyle hassas bir gösterge olup, alternatif olarak %seçim yanlılığı değerine bakılmasında fayda vardır.

Eğer müdahale ve kontrol grubu bireyleri *cinsiyet* ve *istihdam durumu* gibi yaş açısından da eşleştirme yapılamamış örnekleme birbirine benzer olsaydı, eşleştirme yapılmamış örneklem analizler için kullanılabilirdi. Bu bağlamda bireysel karakteristiklerin sadece biri açısından bile benzemeyen müdahale ve kontrol grubunun olduğu örnekleme kullanmak yerine eşleştirme yapılmış örnekleme kullanmak önem taşımaktadır. Araştırmacı için önemli olan eldeki bireysel karakteristiklerin her biri için birbirine çok benzeyen kontrol ve müdahale grupları ile çalışmaktır.

**pstest2** komutu ile sınama yapmak dışında iyi bir eşleştirme yapılıp yapılmadığının kontrolü grafiksel inceleme yoluyla da yapılabilir. Örneğin yerine koymama yoluyla yapılan bire-bir EYKE grafiksel çizim ile incelenmek istendiğinde öncelikle 17.2.2.1 nolu bölümün başında kullanılan ve aşağıda tekrar verilen [17.2.1] nolu komut ile eşleştirme yapılır.

**xi:psmatch2 mud cinsiyet yas i.istihdam if yil==2020 , noreplacement caliper (0.1) out(gfd)**  
Eşleştirme sonrası eşleşen bireylerin ayırt edilmesini sağlayan bir değişken oluşturulmalıdır. [17.2.3] nolu komut dizisi kullanılarak *eyke\_norep* isimli bir değişken oluşturulmuştur.

**gen match\_eyke=\_n1**

**replace match\_eyke=\_id if match\_eyke==.**

**duplicates tag match\_eyke, gen(eyke\_norep)**

[17.2.3]

Oluşturulan *eyke\_norep* değişkeni yardımıyla, eşleştirme öncesi ve sonrası müdahale ve kontrol grubu bireylerinin eğilim skoru dağılımlarını gösteren grafikler [17.2.4] nolu komut dizisi ile çizilmiş ve elde edilen grafikler Şekil 17.2.1'de gösterilmiştir.

**twoway (kdensity \_pscore if \_treated==1, lcolor(black)) (kdensity \_pscore if \_treated==0 & eyke\_norep>0, lpattern(dash) lcolor(black) ytitle("Kdensity Eğilim Skoru")),  
legend( label( 1 "müdahale") label( 2 "kontrol" ) ) xtitle("Eğilim Skoru")  
graphregion(color(white)) saving("\${in\_dir}eykenorepsonra.gph", replace)**

<sup>27</sup>Tahmin edileceği üzere çoklu doğrusal bağlantı nedeniyle cinsiyet ve istihdam değişkenlerinin temel kategorileri analizden dışlanmıştır.

```

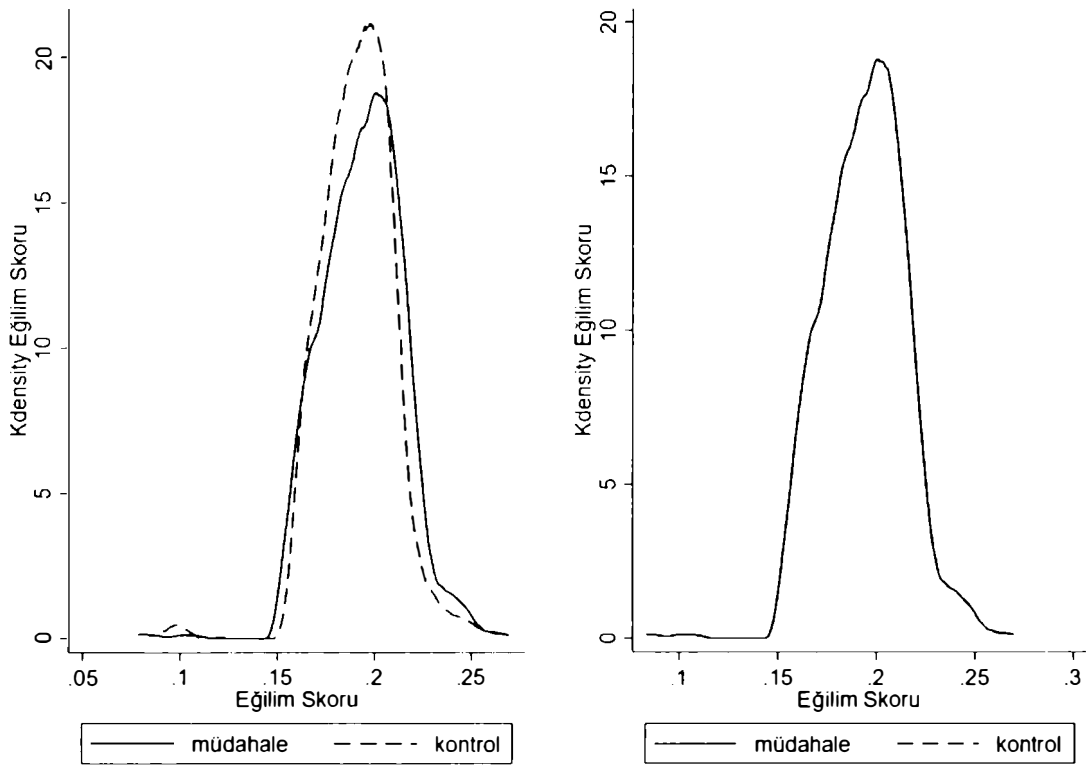
twoway (kdensity _pscore if _treated==1, lcolor(black)) (kdensity _pscore if _treated==0 ,
lpattern(dash) lcolor(black) ytitle("Kdensity Eğilim Skoru")), legend(label(1 "
müdahale") label(2 "kontrol")) xtitle("Eğilim Skoru") graphregion(color(white))
saving("${in_dir}eykenorepönçe.gph", replace)
graph combine "${in_dir}eykenorepönçe.gph" "${in_dir}eykenorepsonra.gph"

```

[17.2.4]

Şeklin sol tarafında, eşleştirme öncesi müdahale ve kontrol grubu bireylerinin eğilim skoru dağılımlarını gösteren grafik yer alırken, sağ tarafında eşleştirme sonrası durumu gösteren grafik yer almaktadır. Eşleştirme sonrası müdahale ve kontrol grubu üyelerinin eğilim skorlarının neredeyse birebir benzerlik gösterdiği, eşleştirme öncesinde ise aralarında farklılıklar olduğu görülmektedir.

Şekil 17.2.1 Eşleştirme Kalitesinin Grafiksel Olarak İncelenmesi



17.2.1 nolu bölümde bahsedildiği gibi *yerine koymama* yoluyla yapılan EYKE'de bir kez kullanılan kontrol grubu birey veya bireylerinin başka bir eşleştirmede kullanılmasına izin verilmemektedir. Gözlem sayısı az olan örneklerle çalışıldığında yerine koymama yolunun tercih edilmemesi, *yerine koyma* yoluyla kontrol grubundan eşleştirmeye birey seçilmesi önerilmektedir. Bu durum eşleştirme kalitesini yükselterek seçim yanlılığını azaltmaktadır. Ancak 17.2.1 nolu

bölümde detayları ile bahsedildiği gibi bu durumun ATT tahmincisinin varyansını yükseltmesi gibi bir riski de bulunmaktadır. Varyans ve yanlışlık arasında hep bir seçim (trade-off) yapılmak durumunda kalınacağından, veri setinin özelliğine göre *yerine koyma* ya da *yerine koymama* yöntemlerinden hangisinin tercih edileceğine iyi karar verilmelidir. Gözlem sayısının az olduğu ve müdahale ile kontrol gruplarının eğilim skorlarının dağılımında dengesizlik olduğu durumda *yerine koymama* yönteminin uygulanması önerilmemektedir. Örneğin müdahale grubu üyelerinin çoğu yüksek, kontrol grubu bireyleri ise düşük eğilim skorlarına sahipse yerine koyma yöntemi ile EYKE'nin yapılması tavsiye edilmektedir (Smith ve Todd, 2005).

### 17.2.2.2. Yerine Koyma Yoluyla En Yakın Komşu Eşleştirmesi

Yerine koyma yoluyla bire-bir EYKE yaparken, [17.2.1] nolu komutundan **noreplacement** opsiyonu çıkarılır. Diğer bir ifade ile eşleştirmede [17.2.5] nolu komut kullanılmıştır ve eşleştirme sonuçları Tablo 17.2.3'te gösterilmiştir.

**xi:psmatch2 mud cinsiyet yas i.istihdam if yil==2020 , caliper (0.1) out(gfd) [17.2.5]**

**Tablo 17.2.3.** Yerine Koyma Yoluyla Bire-Bir En Yakın Komşu Eşleştirmesi

```
xi:psmatch2 mud cinsiyet yas i.istihdam if yil==2020 , caliper (0.1) out(gfd)
i.istihdam _Iistihdam_1-3 (naturally coded; _Iistihdam_1 omitted)
Probit regression Number of obs = 2,745
 LR chi2(4) = 7.37
 Prob > chi2 = 0.1176
Log likelihood = -1337.4261 Pseudo R2 = 0.0027
```

| mud          | Coef.     | Std. Err. | z      | P> z  | [95% Conf. Interval] |           |
|--------------|-----------|-----------|--------|-------|----------------------|-----------|
| cinsiyet     | .0194901  | .0562678  | 0.35   | 0.729 | -.0907929            | .129773   |
| yas          | .0048751  | .0022808  | 2.14   | 0.033 | .0004049             | .0093453  |
| _Iistihdam_2 | -.456244  | .3790538  | -1.20  | 0.229 | -1.199176            | .2866878  |
| _Iistihdam_3 | .1163664  | .1367073  | 0.85   | 0.395 | -.151575             | .3843078  |
| _cons        | -1.086806 | .1009236  | -10.77 | 0.000 | -1.284613            | -.8889993 |

| Variable | Sample    | Treated    | Controls   | Difference | S.E.       | T-stat |
|----------|-----------|------------|------------|------------|------------|--------|
| gfd      | Unmatched | .349809886 | .136547995 | .213261891 | .018079978 | 11.80  |
|          | ATT       | .349809886 | .134980989 | .214828897 | .04856086  | 4.42   |

Note: S.E. does not take into account that the propensity score is estimated.

|            |           |       |
|------------|-----------|-------|
|            | psmatch2: |       |
| psmatch2:  | Common    |       |
| Treatment  | support   |       |
| assignment | On suppor |       |
|            | Total     |       |
| Untreated  | 2,219     | 2,219 |
| Treated    | 526       | 526   |
| Total      | 2,745     | 2,745 |

Bu analizde ATT (difference) değeri 0.214828897 olarak elde edilmiş olup, istatistiksel olarak anlamlıdır ( $t=4.42$ ). Sonuç “İzindeyken İşini Koruma” politikasından yararlanan bireylerin, yararlanmayan bireylere göre %21.48 daha yüksek olasılıkla gelecek dönem finansal durumları hakkında kötümser olduklarını ortaya koymaktadır. Tablo 17.2.1’deki yerine koymama yoluyla bire-bir EYKE ile ulaşılan politika etkisine çok yakın bir sonuç elde edilmiştir. Yerine koyma yoluyla yapılan bire-bir EYKE’nin kalitesi [17.2.6] nolu komutla sınanabilir.

### xi:pstest2 cinsiyet yas i.istihdam

[17.2.6]

**Tablo 17.2.4.** Yerine Koyma Yoluyla En Yakın Komşu Eşleştirmesi pstest2 Sonuçları

xi:pstest2 cinsiyet yas i.istihdam

i.istihdam \_Iistihdam\_1-3 (naturally coded; \_Iistihdam\_1 omitted)

| Variable     | Sample    | Mean    |         | %reduct |       | t-test |       |
|--------------|-----------|---------|---------|---------|-------|--------|-------|
|              |           | Treated | Control | %bias   | bias  | t      | p> t  |
| cinsiyet     | Unmatched | .41255  | .40063  | 2.4     |       | 0.50   | 0.617 |
|              | Matched   | .41255  | .42395  | -2.3    | 4.3   | -0.37  | 0.708 |
| yas          | Unmatched | 42.787  | 41.483  | 10.5    |       | 2.24   | 0.025 |
|              | Matched   | 42.787  | 42.876  | -0.7    | 93.1  | -0.11  | 0.912 |
| _Iistihdam_2 | Unmatched | .0038   | .00856  | -6.1    |       | -1.13  | 0.260 |
|              | Matched   | .0038   | .0038   | 0.0     | 100.0 | -0.00  | 1.000 |
| _Iistihdam_3 | Unmatched | .04753  | .0374   | 5.0     |       | 1.07   | 0.283 |
|              | Matched   | .04753  | .04183  | 2.8     | 43.7  | 0.45   | 0.655 |

Tablo 17.2.4’de gösterildiği gibi eşleştirme yapılmadan önce cinsiyet ve istihdam açısından birbirine benzer ancak yaş değişkeni açısından benzer olmayan müdahale ve kontrol grupları olduğu görülmektedir. Ancak eşleştirme sonrası yaş dâhil tüm değişkenler açısından birbirine çok benzer müdahale ve kontrol grupları elde edilmiştir. Eşleştirmenin kalitesini test etmenin bir diğer yolu grafiksel gösterimdir. Şekil 17.2.1’de yerine koymama yoluyla yapılan bire-bir EYKE için çizilmiş grafik gibi burada da [17.2.5] nolu komut çalıştırıldıktan sonra benzer komutlarla ilgili grafik elde edilebilir<sup>28</sup>.

Eğer birden fazla komşu kontrol grubundan seçilerek yerine koyma yoluyla EYKE yapılmak isteniyorsa, [17.2.5] nolu komuta **neighbor(integer k>1)** opsiyonu eklenmelidir. Örneğin 2 komşu için **neighbor(2)** yazılmalıdır. Kontrol grubu bireylerinin gözlem sayısının çok olduğu durumda, daha güvenilir eşleştirmeler için birden fazla komşunun seçilmesi önerilmektedir (Caliendo ve Kopeinig, 2008). [17.2.7] nolu komut ile en yakın 2 komşu kontrol grubundan eşleş-

<sup>28</sup> **match\_eyke, eyke\_norep, eykenorepönce.gph, eykenorepsonra.gph** gibi değişken ve grafik isimleri yerine koymama yoluyla EYKE’nin grafiksel testi için yazılan komutlarda bir kere kullanılmış olduğundan yerine koyma yoluyla EYKE ve diğer eşleştirme türleri için yapılacak grafiksel testlerde yeni değişken ve grafik isimlerinin kullanılması gerekmektedir.





temleriyle tahmin edilen değerlere çok yakındır (Bknz: Tablolar 17.2.1, 17.2.3). [17.2.8] nolu komut ile bu eşleştirmenin kalitesi test edilmiş, sonuçlar Tablo 17.2.6'da gösterilmiştir. Eşleştirme yapılmadan önce cinsiyet ve istihdam açısından birbirine benzer ancak yaş değişkeni açısından benzer olmayan müdahale ve kontrol grupları olduğu görülmektedir. Ancak eşleştirme sonrası yaş dâhil tüm değişkenler açısından birbirine çok benzer müdahale ve kontrol grupları elde edilmiştir.

**xi:pstest2 cinsiyet yas i.istihdam**

[17.2.8]

**Tablo 17.2.6.** Yerine Koyma Yoluyla İkiye-Bir EYKE pstest2 Sonuçları

**xi:pstest2 cinsiyet yas i.istihdam**

i.istihdam \_Iistihdam\_1-3 (naturally coded; \_Iistihdam\_1 omitted)

| Variable     | Sample    | Mean    |         | %reduct |       | t-test |       |
|--------------|-----------|---------|---------|---------|-------|--------|-------|
|              |           | Treated | Control | %bias   | bias  | t      | p> t  |
| cinsiyet     | Unmatched | .41255  | .40063  | 2.4     |       | 0.50   | 0.617 |
|              | Matched   | .41255  | .4249   | -2.5    | -3.7  | -0.41  | 0.685 |
| yas          | Unmatched | 42.787  | 41.483  | 10.5    |       | 2.24   | 0.025 |
|              | Matched   | 42.787  | 42.916  | -1.0    | 90.1  | -0.16  | 0.873 |
| _Iistihdam_2 | Unmatched | .0038   | .00856  | -6.1    |       | -1.13  | 0.260 |
|              | Matched   | .0038   | .0038   | 0.0     | 100.0 | 0.00   | 1.000 |
| _Iistihdam_3 | Unmatched | .04753  | .0374   | 5.0     |       | 1.07   | 0.283 |
|              | Matched   | .04753  | .03992  | 3.8     | 24.9  | 0.60   | 0.547 |

### 17.2.2.3. Yarıçap Eşleştirmesi

17.2.1 nolu bölümde yarıçap eşleştirmesi anlatılırken bahsedildiği gibi bu eşleştirme türü, müdahale grubundaki bireye sadece belli bir eğilim aralığı içindeki en yakın komşu ile değil, o aralıkta bulunan tüm uygun kontrol grubu bireyleri ile eşleşme şansı tanımaktadır. Yarıçap eşleştirmesi ve eşleştirmenin kalite testi sırasıyla [17.2.9] ve [17.2.10] nolu komutlarla gerçekleştirilmiştir. Eğilim aralığı olarak 0.1 seçilmiştir. Daha önceki eşleştirme türlerinde bahsedildiği gibi farklı eğilim aralıklarının seçilmesi de mümkündür. Tablo 17.2.7'de sunulan eşleştirme sonuçlarına bakıldığında, ATT (difference) değerinin 0.213963827 olarak elde edildiği ve istatistiksel olarak anlamlı ( $t=9.70$ ) olduğu görülmektedir. Özetle, "İzindeyken İşini Koruma" politikasından yararlanan bireylerin, yararlanmayan bireylere göre %21.39 daha yüksek olasılıkla gelecek dönem finansal durumları hakkında kötümser oldukları bulunmuştur.

**xi:psmatch2 mud cinsiyet yas i.istihdam if yıl==2020 , radius caliper(0.1) out(gfd)** [17.2.9]

**xi:pstest2 cinsiyet yas i.istihdam**

[17.2.10]

**Tablo 17.2.7. Yarıçap Eşleştirmesi**

**xi:psmatch2 mud cinsiyet yas i.istihdam if yıl==2020 , radius caliper(0.1) out(gfd)**

i.istihdam      \_Iistihdam\_1-3      (naturally coded; \_Iistihdam\_1 omitted)  
 Probit regression      Number of obs      =      2,745  
                                          LR chi2(4)      =      7.37  
                                          Prob > chi2      =      0.1176  
 Log likelihood = -1337.4261      Pseudo R2      =      0.0027

| mud          | Coef.     | Std. Err. | z      | P> z  | [95% Conf. Interval] |           |
|--------------|-----------|-----------|--------|-------|----------------------|-----------|
| cinsiyet     | .0194901  | .0562678  | 0.35   | 0.729 | -.0907929            | .129773   |
| yas          | .0048751  | .0022808  | 2.14   | 0.033 | .0004049             | .0093453  |
| _Iistihdam_2 | -.456244  | .3790538  | -1.20  | 0.229 | -1.199176            | .2866878  |
| _Iistihdam_3 | .1163664  | .1367073  | 0.85   | 0.395 | -.151575             | .3843078  |
| _cons        | -1.086806 | .1009236  | -10.77 | 0.000 | -1.284613            | -.8889993 |

| Variable | Sample    | Treated    | Controls   | Difference | S.E.       | T-stat |
|----------|-----------|------------|------------|------------|------------|--------|
| gfd      | Unmatched | .349809886 | .136547995 | .213261891 | .018079978 | 11.80  |
|          | ATT       | .349809886 | .135846059 | .213963827 | .022056692 | 9.70   |

Note: S.E. does not take into account that the propensity score is estimated.

|            |           |       |
|------------|-----------|-------|
|            | psmatch2: |       |
| psmatch2:  | Common    |       |
| Treatment  | support   |       |
| assignment | On suppor | Total |
| Untreated  | 2,219     | 2,219 |
| Treated    | 526       | 526   |
| Total      | 2,745     | 2,745 |

Tablo 17.2.8’de yarıçap eşleştirmesi test sonuçlarına bakıldığında ise, yerine koyma ve yerine koymama yoluyla yapılan en yakın komşu eşleştirmelerine kıyasla daha zayıf bir eşleştirme yapıldığı söylenebilir (Bknz: Tablolar 17.2.2, 17.2.4, 17.2.6). Çünkü cinsiyet ve istihdam durumu değişkenleri açısından eşleşmiş (matched) örnekleme müdahale ve kontrol grubu bireylerinin birbirine benzediğini söyleyen sıfır hipotezleri reddedilmemişken (sırasıyla  $p=0.674$ ,  $p=0.853$  ve  $p=0.407$ ), aynı benzerlik yaş değişkeni açısından yakalanamamıştır. Eşleşmiş örnekleme (matched) ortalama yaşlara göre bu iki grubun benzer olduklarını söyleyen sıfır hipotezi reddedilmiştir ( $p=0.081$ ).

**Tablo 17.2.8.** Yarıçap Eşleştirmesi ptest2 Sonuçları**xi:pstest2 cinsiyet yas i.istihdam****i.istihdam**                      **\_Iistihdam\_1-3**                      (naturally coded; **\_Iistihdam\_1** omitted)

| Variable     | Sample    | Mean    |         | %reduct |      | t-test |       |
|--------------|-----------|---------|---------|---------|------|--------|-------|
|              |           | Treated | Control | %bias   | bias | t      | p> t  |
| cinsiyet     | Unmatched | .41255  | .40063  | 2.4     |      | 0.50   | 0.617 |
|              | Matched   | .41255  | .3998   | 2.6     | -7.0 | 0.42   | 0.674 |
| yas          | Unmatched | 42.787  | 41.483  | 10.5    |      | 2.24   | 0.025 |
|              | Matched   | 42.787  | 41.447  | 10.8    | -2.8 | 1.74   | 0.081 |
| _Iistihdam_2 | Unmatched | .0038   | .00856  | -6.1    |      | -1.13  | 0.260 |
|              | Matched   | .0038   | .00454  | -0.9    | 84.6 | -0.18  | 0.853 |
| _Iistihdam_3 | Unmatched | .04753  | .0374   | 5.0     |      | 1.07   | 0.283 |
|              | Matched   | .04753  | .03722  | 5.1     | -1.8 | 0.83   | 0.407 |

Literatürde eşleştirme yöntemlerinin arasında en iyi eşleştirmeyi hangi yöntemin yaptığı konusunda genel bir kanı olmadığından ve kullanılan örnekleme göre eşleştirme yöntemlerinin yaptığı eşleştirme kalitesi değişebildiğinden, birden fazla eşleştirme yöntemleriyle eşleştirme yapılması ve eşleştirme kalitesi yüksek yöntemlerin elde etmiş olduğu ATT değerlerinin politika sonucu olarak kabul edilmesi gerekmektedir. Aslında ele alınan bu örnekte yarıçap eşleştirmesi ile elde edilen eşleştirmede yalnızca yaş açısından müdahale ve kontrol grupları arasında benzerlik %10 düzeyinde ( $p=0.081$ ) reddedilmiş olup, marjinde bir red söz konusu olduğundan ATT değeri (0.213963827) EYKE yöntemleri ile elde edilen ATT değerlerinden çok da farklı bulunmamıştır. Ancak benzer ATT değerlerine ulaşmak her zaman söz konusu olmayabileceğinden, eşleştirmede kullanılan tüm değişkenler açısından çok benzer müdahale ve kontrol grubu elde edilmesini sağlayan eşleştirme yöntemlerinin ortaya koyduğu politika etkisi sonuçları kabul edilmelidir.

#### 17.2.2.4. Mahalanobis Metrik Eşleştirmesi

Mahalanobis metrik eşleştirmesinin en temel hâli [17.2.11] nolu komut çalıştırılarak yapılmaktadır. [17.2.12] nolu komut ise eşleştirmenin kalitesini ölçen test komutudur. Eşleştirme ve test sonuçları sırasıyla Tablo 17.2.9 ve Tablo 17.2.10'da gösterilmiştir.

**xi:psmatch2 mud if yıl==2020 , mahalanobis(cinsiyet yas i.istihdam ) out(gfd)** [17.2.11]

**xi:pstest2 cinsiyet yas i.istihdam** [17.2.12]

**Tablo 17.2.9.** Mahalanobis Metrik Eşleştirmesi

```
xi:psmatch2 mud if yıl==2020 , mahalanobis(cinsiyet yas i.istihdam) out(gfd)
```

```
i.istihdam _Iistihdam_1-3 (naturally coded; _Iistihdam_1 omitted)
```

| Variable | Sample    | Treated    | Controls   | Difference | S.E.       | T-stat |
|----------|-----------|------------|------------|------------|------------|--------|
| gfd      | Unmatched | .349809886 | .136547995 | .213261891 | .018079978 | 11.80  |
|          | ATT       | .349809886 | .136882129 | .212927757 | .044996648 | 4.73   |

Note: S.E. does not take into account that the propensity score is estimated.

|            |           |       |
|------------|-----------|-------|
|            | psmatch2: |       |
| psmatch2:  | Common    |       |
| Treatment  | support   |       |
| assignment | On suppor | Total |
| Untreated  | 2,219     | 2,219 |
| Treated    | 526       | 526   |
| Total      | 2,745     | 2,745 |

**Tablo 17.2.10.** Mahalanobis Metrik Eşleştirmesi pstest2 Sonuçları

```
xi:pstest2 cinsiyet yas i.istihdam
```

```
i.istihdam _Iistihdam_1-3 (naturally coded; _Iistihdam_1 omitted)
```

| Variable     | Sample    | Mean    |         | %reduct |       | t-test |       |
|--------------|-----------|---------|---------|---------|-------|--------|-------|
|              |           | Treated | Control | %bias   | bias  | t      | p> t  |
| cinsiyet     | Unmatched | .41255  | .40063  | 2.4     |       | 0.50   | 0.617 |
|              | Matched   | .41255  | .41255  | 0.0     | 100.0 | -0.00  | 1.000 |
| yas          | Unmatched | 42.787  | 41.483  | 10.5    |       | 2.24   | 0.025 |
|              | Matched   | 42.787  | 42.8    | -0.1    | 99.0  | -0.02  | 0.987 |
| _Iistihdam_2 | Unmatched | .0038   | .00856  | -6.1    |       | -1.13  | 0.260 |
|              | Matched   | .0038   | .0038   | 0.0     | 100.0 | 0.00   | 1.000 |
| _Iistihdam_3 | Unmatched | .04753  | .0374   | 5.0     |       | 1.07   | 0.283 |
|              | Matched   | .04753  | .04753  | 0.0     | 100.0 | 0.00   | 1.000 |

17.2.1 nolu bölümde detayları ile bahsedildiği gibi, [17.2.13] nolu komutla karma bir yöntem olan *belli bir eğilim aralığında Mahalanobis metrik eşleştirmesi* de uygulanabilir. Belli bir eği-

lim aralığında Mahalanobis metrik eşleştirmesi ve test sonuçları sırasıyla Tablo 17.2.11 ve Tablo 17.2.12'de gösterilmiştir.

**xi:psmatch2 mud if yıl==2020 , mahalanobis(cinsiyet yas i.istihdam) caliper(0.1) out(gfd)** [17.2.13]

**xi:pstest2 cinsiyet yas i.istihdam** [17.2.14]

**Tablo 17.2.11.** Belli bir Eğilim Aralığında Mahalanobis Metrik Eşleştirmesi

**xi:psmatch2 mud if yıl==2020 , mahalanobis(cinsiyet yas i.istihdam) caliper(0.1) out(gfd)**

| i.istihdam | _Iistihdam_1-3 | (naturally coded; _Iistihdam_1 omitted) |            |            |            |        |
|------------|----------------|-----------------------------------------|------------|------------|------------|--------|
| Variable   | Sample         | Treated                                 | Controls   | Difference | S.E.       | T-stat |
| gfd        | Unmatched      | .349809886                              | .136547995 | .213261891 | .018079978 | 11.80  |
|            | ATT            | .351145038                              | .13740458  | .213740458 | .045382061 | 4.71   |

Note: S.E. does not take into account that the propensity score is estimated.

|            |                  |           |       |
|------------|------------------|-----------|-------|
| psmatch2:  | psmatch2: Common |           |       |
| Treatment  | support          |           |       |
| assignment | Off suppo        | On suppor | Total |
| Untreated  | 0                | 2,219     | 2,219 |
| Treated    | 2                | 524       | 526   |
| Total      | 2                | 2,743     | 2,745 |

**Tablo 17.2.12.** Belli bir Eğilim Aralığında Mahalanobis Metrik Eşleştirmesi **pstest2** Sonuçları

**xi:pstest2 cinsiyet yas i.istihdam**

**i.istihdam** **\_Iistihdam\_1-3** (naturally coded; \_Iistihdam\_1 omitted)

| Variable     | Sample    | Mean    |         | %reduct |       | t-test |       |
|--------------|-----------|---------|---------|---------|-------|--------|-------|
|              |           | Treated | Control | %bias   | bias  | t      | p> t  |
| cinsiyet     | Unmatched | .41255  | .40063  | 2.4     |       | 0.50   | 0.617 |
|              | Matched   | .41412  | .41412  | 0.0     | 100.0 | -0.00  | 1.000 |
| yas          | Unmatched | 42.787  | 41.483  | 10.5    |       | 2.24   | 0.025 |
|              | Matched   | 42.803  | 42.811  | -0.1    | 99.4  | -0.01  | 0.992 |
| _Iistihdam_2 | Unmatched | .0038   | .00856  | -6.1    |       | -1.13  | 0.260 |
|              | Matched   | .00191  | .00191  | 0.0     | 100.0 | 0.00   | 1.000 |
| _Iistihdam_3 | Unmatched | .04753  | .0374   | 5.0     |       | 1.07   | 0.283 |
|              | Matched   | .0458   | .0458   | 0.0     | 100.0 | 0.00   | 1.000 |

Sırasıyla Tablo 17.2.9 ve Tablo 17.2.11’de görüleceği üzere Mahalanobis metrik eşleştirmesinde ATT değeri 0.212927757, belli bir eğilim aralığında Mahalanobis metrik eşleştirmesinde ise 0.213740458 olarak bulunmuştur. Elde edilen sonuçlara göre “İzindeyken İşini Koruma” politikasından yararlanan bireylerin, yararlanmayan bireylere göre sırasıyla %21.29 ve %21.37 daha yüksek olasılıkla gelecek dönem finansal durumları hakkında kötümser oldukları söylenebilir. Bu bulgular daha önce yapılan eşleştirme yöntemleri ile ulaşılan ATT değerleri ile benzerdir (Bknz: Tablolar 17.2.1, 17.2.3, 17.2.5, 17.2.7, 17.2.9).

Sırasıyla Tablo 17.2.10 ve Tablo 17.2.12’de sunulan eşleştirme kalitesi test sonuçlarına göre ise bireysel karakteristikler açısından eşleştirmenin başarılı bir şekilde yapıldığı görülmektedir. Örneğin Tablo 17.2.10’daki sonuçlara göre eşleşmiş örnekleme (matched) cinsiyet, yaş ve istihdam durumu kategorileri açısından elde edilen müdahale ve kontrol gruplarının benzer olduklarını söyleyen sıfır hipotezleri sırasıyla  $p=1.000$ ,  $p=0.987$ ,  $p=1.000$  ve  $p=1.000$  olasılık değerleriyle reddedilmemiştir. 17.2.12’de ise bu hipotezler sırasıyla  $p=1.000$ ,  $p=0.992$ ,  $p=1.000$  ve  $p=1.000$  olasılık değerleriyle reddedilmemiştir.

Yarıçap eşleştirmesi haricinde şu ana kadar uygulanmış EYKE ve Mahalanobis metrik eşleştirmesi yöntemleriyle kaliteli eşleştirmelerin yapıldığı, hatta Mahalanobis metrik eşleştirmesinde bireysel karakteristiklerin ortalama değerlerinin eşleştirme sonrası müdahale ve kontrol grupları için birebir aynı olduğu görülmektedir. Örneğin Tablo 17.2.12’de sunulan belli bir eğilim aralığında Mahalanobis metrik eşleştirmesi sonuçlarına göre, cinsiyet açısından eşleştirme sonrası iki grupta da bireylerin %41’inin erkeklerden oluştuğu ve ortalama yaşın iki grupta da 42.8 olduğu görülmektedir.

#### 17.2.2.5. Kernel Eşleştirmesi

Kernel eşleştirmesinin en temel hâli [17.2.15] nolu komut çalıştırılarak yapılmıştır. [17.2.16] nolu komutla da eşleştirme kalitesinin testi gerçekleştirilmiştir.

**xi:psmatch2 mud cinsiyet yas i.istihdam if yıl==2020 , kernel bwidth(0.1) out(gfd)**

**[17.2.15]**

**xi:pstest2 cinsiyet yas i.istihdam**

**[17.2.16]**

Tablo 17.2.13’te gösterildiği gibi kernel eşleştirmesi sonrası ATT (difference) değeri 0.215333635 olarak bulunmuştur. Bu sonuç “İzindeyken İşini Koruma” politikasından yararlanan bireylerin, yararlanmayan bireylere göre %21.53 daha yüksek olasılıkla gelecek dönem finansal durumları hakkında kötümser olduklarını göstermektedir. Elde edilen tahmin değeri, diğer eşleştirme yöntemleriyle elde edilen değerlere çok yakındır (Bknz: Tablolar 17.2.1, 17.2.3, 17.2.5, 17.2.7, 17.2.9, 17.2.11).

**Tablo 17.2.13. Kernel Eşleştirmesi**

```

xi:psmatch2 mud cinsiyet yas i.istihdam if yıl==2020 , kernel bwidth(0.1) out(gfd)
i.istihdam _Iistihdam_1-3 (naturally coded; _Iistihdam_1 omitted)
Probit regression Number of obs = 2,745
 LR chi2(4) = 7.37
 Prob > chi2 = 0.1176
Log likelihood = -1337.4261 Pseudo R2 = 0.0027

```

| mud          | Coef.     | Std. Err. | z      | P> z  | [95% Conf. Interval] |           |
|--------------|-----------|-----------|--------|-------|----------------------|-----------|
| cinsiyet     | .0194901  | .0562678  | 0.35   | 0.729 | -.0907929            | .129773   |
| yas          | .0048751  | .0022808  | 2.14   | 0.033 | .0004049             | .0093453  |
| _Iistihdam_2 | -.456244  | .3790538  | -1.20  | 0.229 | -1.199176            | .2866878  |
| _Iistihdam_3 | .1163664  | .1367073  | 0.85   | 0.395 | -.151575             | .3843078  |
| _cons        | -1.086806 | .1009236  | -10.77 | 0.000 | -1.284613            | -.8889993 |

| Variable | Sample    | Treated    | Controls   | Difference | S.E.       | T-stat |
|----------|-----------|------------|------------|------------|------------|--------|
| gfd      | Unmatched | .349809886 | .136547995 | .213261891 | .018079978 | 11.80  |
|          | ATT       | .349809886 | .134476251 | .215333635 | .02206365  | 9.76   |

Note: S.E. does not take into account that the propensity score is estimated.

|            |           |       |
|------------|-----------|-------|
| psmatch2:  |           |       |
| psmatch2:  | Common    |       |
| Treatment  | support   |       |
| assignment | On suppor | Total |
| Untreated  | 2,219     | 2,219 |
| Treated    | 526       | 526   |
| Total      | 2,745     | 2,745 |

Tablo 17.2.14'te sunulan kernel eşleştirmesi test sonuçları, yaş dışında diğer değişkenler açısından eşleştirmenin başarılı olduğunu göstermektedir. Eşleşmiş örnekleme yaş değişkeni açısından müdahale ve kontrol gruplarındaki bireylerin birbirine çok benzer olduğu %10 anlamlılık düzeyinde reddedilmiştir ( $t=1.67$ ,  $p=0.094$ ). Aslında bireysel karakteristikler arasında sadece yaş değişkeni açısından müdahale ve kontrol gruplarının benzerliği reddedildiğinden ve marjinde bir red ( $p=0.094$ ) söz konusu olduğundan, ATT değeri (0.215333635) eşleştirme kalitesi yüksek bulunan EYKE ve Mahalanobis metrik eşleştirmesi yöntemleri ile elde edilen ATT değerlerinden çok da farklı bulunmamıştır. Yarıçap eşleştirmesi örneğinde de buna benzer bir sonucun elde edildiği hatırlanacaktır. Ancak benzer ATT değerlerine ulaşmak her zaman söz konusu olmayabileceğinden, eşleştirmede kullanılan tüm değişkenler açısından çok benzer müdahale ve kontrol grubu elde edilmesini sağlayan eşleştirme yöntemlerinin ortaya koyduğu politika etkisi sonuçları kabul edilmelidir.



**Tablo 17.2.14.** Kernel Eşleştirmesi ptest2 Sonuçları

**xi:ptest2 cinsiyet yas i.istihdam**  
**i.istihdam \_Iistihdam\_1-3** (naturally coded; \_Iistihdam\_1 omitted)

| Variable     | Sample    | Mean    |         | %reduct |       | t-test |       |
|--------------|-----------|---------|---------|---------|-------|--------|-------|
|              |           | Treated | Control | %bias   | bias  | t      | p> t  |
| cinsiyet     | Unmatched | .41255  | .40063  | 2.4     |       | 0.50   | 0.617 |
|              | Matched   | .41255  | .39967  | 2.6     | -8.1  | 0.43   | 0.671 |
| yas          | Unmatched | 42.787  | 41.483  | 10.5    |       | 2.24   | 0.025 |
|              | Matched   | 42.787  | 41.511  | 10.2    | 2.2   | 1.67   | 0.094 |
| _Iistihdam_2 | Unmatched | .0038   | .00856  | -6.1    |       | -1.13  | 0.260 |
|              | Matched   | .0038   | .00184  | 2.5     | 58.9  | 0.60   | 0.550 |
| _Iistihdam_3 | Unmatched | .04753  | .0374   | 5.0     |       | 1.07   | 0.283 |
|              | Matched   | .04753  | .03232  | 7.5     | -50.2 | 1.26   | 0.208 |

En temel kernel eşleştirmesinin, eşleştirmede tam başarılı olamadığı ya da başarısız olduğu durumlarda mahalanobis metrik uzaklığı temelinde de kernel eşleştirmesi yapılabilmektedir. 17.2.1 nolu bölümde anlatıldığı gibi kernel eşleştirmesinde her bir müdahale bireyi için birden fazla kontrol bireyi, müdahale gözleminden uzaklaştıkça azalan bir ağırlık alacak şekilde kullanılmaktadır. Yani kernel eşleştirmesinin en büyük avantajı farklı ağırlıklar vererek neredeyse tüm kontrol grubu bireylerini eşleştirmeye dâhil etmesi ve böylece ATT varyansını küçük bulmasıdır. Ancak çok sayıda kontrol grubu bireyinin eşleştirmeye dâhil edilmesi kötü eşleştirmelere de sebep olabilmektedir. Böyle bir durumda mahalanobis metrik uzaklığı temelinde kernel eşleştirmesi tercih edilmektedir. Mahalanobis metrik uzaklığı temelinde kernel eşleştirmesi [17.2.17] nolu komut ile yapılmıştır. [17.2.18] nolu komutla da eşleştirme kalitesinin testi gerçekleştirilmiştir.

**xi:psmatch2 mud cinsiyet yas i.istihdam if yıl=2020, kernel bwidth(0.1)**

**mahalanobis(cinsiyet yas i.istihdam ) out(gfd)** [17.2.17]

**xi:ptest2 cinsiyet yas i.istihdam** [17.2.18]

Tablo 17.2.15'te gösterilen sonuçlarda ATT değeri 0.220416698 olarak bulunmuştur. Sonuç olarak "İzindeyken İşini Koruma" politikasından yararlanan bireylerin, yararlanmayan bireylere göre %22.04 daha yüksek olasılıkla gelecek dönem finansal durumları hakkında kötümser oldukları söylenebilir. Elde edilen tahmin değeri, diğer eşleştirme yöntemleriyle tahmin edilen değerlere çok yakındır (Bknz: Tablolar 17.2.1, 17.2.3, 17.2.5, 17.2.7, 17.2.9, 17.2.11, 17.2.13).

**Tablo 17.2.15.** Mahalanobis Metrik Uzaklığı Temelinde Kernel Eşleştirmesi

```

xi:psmatch2 mud cinsiyet yas i.istihdam if yıl==2020, kernel bwidth(0.1)
mahalanobis(cinsiyet yas i.istihdam) out(gfd)
i.istihdam _Iistihdam_1-3 (naturally coded; _Iistihdam_1 omitted)
Probit regression
Number of obs = 2,745
LR chi2(4) = 7.37
Prob > chi2 = 0.1176
Pseudo R2 = 0.0027
Log likelihood = -1337.4261

```

| mud          | Coef.     | Std. Err. | z      | P> z  | [95% Conf. Interval] |           |
|--------------|-----------|-----------|--------|-------|----------------------|-----------|
| cinsiyet     | .0194901  | .0562678  | 0.35   | 0.729 | -.0907929            | .129773   |
| yas          | .0048751  | .0022808  | 2.14   | 0.033 | .0004049             | .0093453  |
| _Iistihdam_2 | -.456244  | .3790538  | -1.20  | 0.229 | -1.199176            | .2866878  |
| _Iistihdam_3 | .1163664  | .1367073  | 0.85   | 0.395 | -.151575             | .3843078  |
| _cons        | -1.086806 | .1009236  | -10.77 | 0.000 | -1.284613            | -.8889993 |

| Variable | Sample    | Treated    | Controls   | Difference | S.E.       | T-stat |
|----------|-----------|------------|------------|------------|------------|--------|
| gfd      | Unmatched | .349809886 | .136547995 | .213261891 | .018079978 | 11.80  |
|          | ATT       | .351145038 | .13072834  | .220416698 | .022283259 | 9.89   |

Note: S.E. does not take into account that the propensity score is estimated.

|            |                  |           |       |
|------------|------------------|-----------|-------|
| psmatch2:  | psmatch2: Common |           |       |
| Treatment  | support          |           |       |
| assignment | Off suppo        | On suppor | Total |
| Untreated  | 0                | 2,219     | 2,219 |
| Treated    | 2                | 524       | 526   |
| Total      | 2                | 2,743     | 2,745 |

**Tablo 17.2.16.** Mahalanobis Metrik Uzaklığı Temelinde Kernel Eşleştirmesi pstest2 Sonuçları

```

xi:pstest2 cinsiyet yas i.istihdam
i.istihdam _Iistihdam_1-3 (naturally coded; _Iistihdam_1 omitted)

```

| Variable     | Sample    | Mean    |         | %reduct |       | t-test |       |
|--------------|-----------|---------|---------|---------|-------|--------|-------|
|              |           | Treated | Control | %bias   | bias  | t      | p> t  |
| cinsiyet     | Unmatched | .41255  | .40063  | 2.4     |       | 0.50   | 0.617 |
|              | Matched   | .41412  | .41412  | 0.0     | 100.0 | -0.00  | 1.000 |
| yas          | Unmatched | 42.787  | 41.483  | 10.5    |       | 2.24   | 0.025 |
|              | Matched   | 42.803  | 42.773  | 0.2     | 97.7  | 0.04   | 0.970 |
| _Iistihdam_2 | Unmatched | .0038   | .00856  | -6.1    |       | -1.13  | 0.260 |
|              | Matched   | .00191  | .00191  | 0.0     | 100.0 | 0.00   | 1.000 |
| _Iistihdam_3 | Unmatched | .04753  | .0374   | 5.0     |       | 1.07   | 0.283 |
|              | Matched   | .0458   | .0458   | 0.0     | 100.0 | 0.00   | 1.000 |

Ayrıca Tablo 17.2.16'da görüldüğü gibi Mahalanobis metrik uzaklığı temelindeki kernel eşleştirmesi, Tablo 17.2.14'te sunulan temel kernel eşleştirmesinden daha başarılı bir eşleştirme ortaya koymuştur. Eşleştirme sonrası tüm bireysel karakteristikler açısından kontrol ve müdahale grubu neredeyse birebir aynı değerleri almaktadır.

Hem en temel kernel eşleştirmesi hem de Mahalanobis metrik uzaklığı temelindeki kernel eşleştirmesi yapılırken, [17.2.15] ve [17.2.17] nolu komutlarda görülebileceği gibi **bwidth( )** opsiyonu yer almaktadır. **bwidth**, Türkçesi 17.2.1 nolu bölümde detayları ile anlatılan bant genişliği ya da düzeltme parametresi olarak adlandırılan **bandwidth** kelimesinin kısaltmasıdır. [17.2.15] ve [17.2.17] nolu komutlarda analizin nasıl yapıldığını göstermek amacıyla ihtiyari olarak 0.1 değeri seçilmiştir. Küçük bant genişlikli kernel fonksiyonlarında müdahale grubu bireylerinin eğilim skorlarına yakın eğilim skorlu kontrol grubu bireyleri eşleştirmeye dâhil edilirken, bant genişliği büyük değerler alan kernel fonksiyonlarında müdahale grubu bireylerine daha az benzeyen kontrol grubu üyeleri eşleştirmede kullanılır. Bu iki durumun da dezavantajları olabileceğinden en uygun yol herhangi bir bant genişliğine karar verdikten sonra **pstest2** komutu ile eşleştirmenin ne kadar doğru yapıldığına bakmaktır. Ayrıca Stata 15'te [17.2.19] ve [17.2.20] nolu **kmatch** komutlarıyla kernel eşleştirmesi yapılarak otomatik bant genişliği belirleme avantajından da yararlanılabilmektedir (Jann, 2017)<sup>30</sup>.

**kmatch md mud cinsiyet yas i.istihdam (gfd) if yıl=2020 , att**

[17.2.19]

**kmatch ps mud cinsiyet yas i.istihdam (gfd) if yıl=2020 , att**

[17.2.20]

**ps**; eğilim skorları temelinde, **md**; mahalanobis metrik uzaklığı temelinde kernel eşleştirmesi yapıldığını göstermektedir.

Sırasıyla Tablo 17.2.17 ve Tablo 17.2.18'de **kmatch md** ve **kmatch ps** komutları ile yapılan eşleştirmelerin sonuçları rapor edilmiştir. ATT tahminlerinin 0.217342 ve 0.2130992 bulunduğu bu iki eşleştirmede elde edilen sonuçlar Tablo 17.2.13 ve 17.2.15'te gösterilen sonuçlarla benzerdir.

<sup>30</sup> [17.2.19] ve [17.2.20] nolu komutların çalıştırılabilmesi için Stata 15'te komut penceresine **ssc install kmatch** yazılarak **kmatch** komutu indirilmelidir.

**Tablo 17.2.17.** Mahalanobis Metrik Uzaklığı Temelinde Kernel Eşleştirmesi

([17.2.19] nolu kmatch komutu ile)

**kmatch md mud cinsiyet yas i.istihdam (gfd) if yil==2020 , att**

(computing bandwidth ... done)

|                                       |               |   |       |
|---------------------------------------|---------------|---|-------|
| Multivariate-distance kernel matching | Number of obs | = | 2,745 |
|                                       | Kernel        | = | epan  |

Treatment : mud = 1

Metric : mahalanobis

Covariates : cinsiyet yas i.istihdam

Matching statistics

|         | Yes | Matched<br>No | Total | Used | Controls<br>Unused | Total | Bandwidth<br>.6913545 |
|---------|-----|---------------|-------|------|--------------------|-------|-----------------------|
| Treated | 526 | 0             | 526   | 2202 | 17                 | 2219  |                       |

Treatment-effects estimation

| gfd | Coef.   | Std. Err. | t    | P> t  | [95% Conf. Interval] |
|-----|---------|-----------|------|-------|----------------------|
| ATT | .217342 | .0220943  | 9.84 | 0.000 | .1740188 .2606652    |

**Tablo 17.2.18.** Eğilim Skorları Temelinde Kernel Eşleştirmesi

([17.2.20] nolu kmatch komutu ile)

**kmatch ps mud cinsiyet yas i.istihdam (gfd) if yil==2020 , att**

(computing bandwidth ... done)

|                                  |               |   |       |
|----------------------------------|---------------|---|-------|
| Propensity-score kernel matching | Number of obs | = | 2,745 |
|                                  | Kernel        | = | epan  |

Treatment : mud = 1

Covariates : cinsiyet yas i.istihdam

PS model : logit (pr)

Matching statistics

|         | Yes | Matched<br>No | Total | Used | Controls<br>Unused | Total | Bandwidth<br>.0019383 |
|---------|-----|---------------|-------|------|--------------------|-------|-----------------------|
| Treated | 526 | 0             | 526   | 2197 | 22                 | 2219  |                       |

Treatment-effects estimation

| gfd | Coef.    | Std. Err. | t    | P> t  | [95% Conf. Interval] |
|-----|----------|-----------|------|-------|----------------------|
| ATT | .2130992 | .0223575  | 9.53 | 0.000 | .1692599 .2569385    |

### 17.2.3. Eğilim Skoru Eşleştirme Yönteminde Kullanılan Stata Komutları

17.2 nolu bölümde kullanılan Stata komutları aşağıda toplu olarak yer almaktadır.

```
clear all
set more off
global in_dir "DOSYAKONUMU\"
use "${in_dir}Covid.dta", clear
ssc install psmatch2
ssc install pstest2
ssc install kmatch
xi:psmatch2 mud cinsiyet yas i.istihdam if yil==2020 , noreplacement caliper (0.1)
out(gfd)
xi:pstest2 cinsiyet yas i.istihdam
xi:psmatch2 mud cinsiyet yas i.istihdam if yil==2020 , noreplacement caliper (0.1)
out(gfd)
gen match_eyke=_n1
replace match_eyke=_id if match_eyke==.
duplicates tag match_eyke, gen(eyke_norep)

tway (kdensity _pscore if _treated==1, lcolor(black)) (kdensity _pscore if
_treated==0 & eyke_norep>0, lpattern(dash) lcolor(black) ytitle("Kdensity Eğilim
Skoru")), legend(label(1 "müdahale") label(2 "kontrol")) xtitle("Eğilim Skoru")
graphregion(color(white)) saving("${in_dir}eykenorepsonra.gph", replace)

tway (kdensity _pscore if _treated==1, lcolor(black)) (kdensity _pscore if
_treated==0 , lpattern(dash) lcolor(black) ytitle("Kdensity Eğilim Skoru")),
legend(label(1 "müdahale") label(2 "kontrol")) xtitle("Eğilim Skoru")
graphregion(color(white)) saving("${in_dir}eykenorepönce.gph", replace)
graph combine "${in_dir}eykenorepönce.gph" "${in_dir}eykenorepsonra.gph"

xi:psmatch2 mud cinsiyet yas i.istihdam if yil==2020 , caliper (0.1) out(gfd)
xi:pstest2 cinsiyet yas i.istihdam
xi:psmatch2 mud cinsiyet yas i.istihdam if yil==2020 , neighbor(2)caliper(0.1)
out(gfd)
xi:pstest2 cinsiyet yas i.istihdam
xi:psmatch2 mud cinsiyet yas i.istihdam if yil==2020 , radius caliper(0.1) out(gfd)
xi:pstest2 cinsiyet yas i.istihdam
xi:psmatch2 mud if yil==2020 , mahalanobis(cinsiyet yas i.istihdam) out(gfd)
xi:pstest2 cinsiyet yas i.istihdam
xi:psmatch2 mud if yil==2020 , mahalanobis(cinsiyet yas i.istihdam) caliper(0.1)
out(gfd)
xi:pstest2 cinsiyet yas i.istihdam
xi:psmatch2 mud cinsiyet yas i.istihdam if yil==2020 , kernel bwidth(0.1)
out(gfd)
xi:pstest2 cinsiyet yas i.istihdam
xi:psmatch2 mud cinsiyet yas i.istihdam if yil==2020 , kernel mahalanobis(cinsiyet
yas i.istihdam) bwidth(0.1) out(gfd)
xi:pstest2 cinsiyet yas i.istihdam
kmatch md mud cinsiyet yas i.istihdam (gfd) if yil==2020 , att
kmatch ps mud cinsiyet yas i.istihdam (gfd) if yil==2020 , att
```

### 17.3. REGRESYON SÜREKSİZLİĞİ ANALİZİ

#### 17.3.1. Regresyon Süreksizliği Analizinin Teorik Çerçevesi

Regresyon Süreksizliği Analizi veya diğer adıyla Regresyon Süreksizliği Tasarımı (Regression Discontinuity Design-RDD), Farkların Farkı (FF) veya Eğilim Skoru Eşleştirme (ESE) yöntemleri gibi deneysel olmayan politika analiz yöntemlerindendir. Regresyon Süreksizliği Analizi (RSA) deneysel yöntem olan rastgele seçim yöntemi ile elde edilen sonuçlara en yakın sonuçların elde edildiği tasarım olarak kabul edilir. Bu bağlamda yarı-deneysel (quasi experimental) yöntem olarak da adlandırılır. Uygulanan bir politikadan yararlanma veya bir müdahaleden etkilendirme olasılığı belli değişkenlerin süreksiz (kesikli) fonksiyonudur. Örneğin devletin uygulamaya koyduğu bir istihdam destek programından yararlanma koşulu son 6 aydır işsiz olmak veya belli bir yaşın üzerinde olmak ise, burada politikadan yararlanma bitirilen yaşın ya da işsiz kalınan sürenin bir fonksiyonudur. Politikadan yararlanmanın belli koşulları olduğundan yani sadece belli kriterleri karşılayan bireyler politikadan ya da müdahaleden etkilendiğinden önceden belirlenmiş bir **seçilim kriteri (eşik değeri)** ile örneklemdeki bireyler müdahale ve kontrol grubu olarak ikiye ayrılırlar. Müdahale ve kontrol grubunu ayıran seçim kriteri ya da eşik değerinin araştırmacının elinde olması, yani araştırmacı tarafından bilinmesi şarttır. Devlet tarafından uygulanan politikalar genelde kanunlarla yürürlüğe girdiğinden, politikadan yararlanma ya da seçim kriterinin ne olduğu kanunun belirttiği şartlardan anlaşılabilmektedir. Örneğin Bronzini ve Iachini (2014) çalışması, 2003 yılında İtalya'nın Emilia-Romagna bölgesinde 2002/7 sayılı Bölgesel Kanun'un 4. maddesi ile uygulamaya koyulan "Endüstriyel Araştırma, İnovasyon ve Teknoloji Transferi" adlı Ar-Ge destek programının firmaların yatırım düzeyine etkisini analiz etmiştir. Programdan yararlanan gruba (müdahale grubuna) girebilmek için desteğe başvuran firmanın sunduğu projenin 100 üzerinden en az 75 puan alması koşulu getirilmiştir. Burada belirlenen 75 puan, seçim kriteri yani programdan yararlanma kriteri ya da eşik değeridir.

Bazen araştırmacılara politika yapıcılar tarafından bir eşik değeri bilgisi verilmeyebilir. Diğer bir deyişle, politikadan yararlanma kriteri olmayabilir. Bu durumlarda RSA politikanın etkinliğinin belirlenmesinde kullanılabilecek bir yöntem değildir. Diğer yandan RSA sadece politika analizinde kullanılan bir yöntem de değildir. Herhangi bir uygulamanın etkinliği de RSA ile tahmin edilebildiğinden, eşik değeri politika uygulamasında olduğu gibi araştırmacıya politika yapıcılar ya da kanunlar tarafından sunulmayabilir. Bu durumda araştırmacı eşik değerini farklı çalışma ve bulgulardan elde etmelidir. Örneğin Angrist ve Lavy (1999) sınıf mevcudunun öğrencilerin sınav sonuçları üzerindeki etkisini analiz ettiği çalışmalarında eşik değerini, 12. yüzyıl filozoflarından Musa bin Meymun (Maimonides) tarafından önerilen bir sınıfta olması gereken en çok öğrenci sayısından yola çıkarak belirlemişlerdir. Musa bin Meymun bir sınıfta öğrencilerin başarısı için en fazla 40 öğrencinin olması gerektiğini, 40'dan fazla öğrenci varsa o sınıfa en az 2 öğretmenin tayin edilmesi gerektiğini belirtmiştir. Bu öneriden yola çıkarak Angrist ve Lavy (1999) çalışmalarında eşik değerini 40 olarak belirlemiştir. RSA yöntemini ilk kullanan Thistlethwaite ve Campbell (1960) çalışması ise öğrencilerin okulda başarı ödülü almalarının gelecekteki akademik çalışmaları üzerindeki etkisini ölçerken eşik değerini belli bir sınav sonucu değeri olarak ele

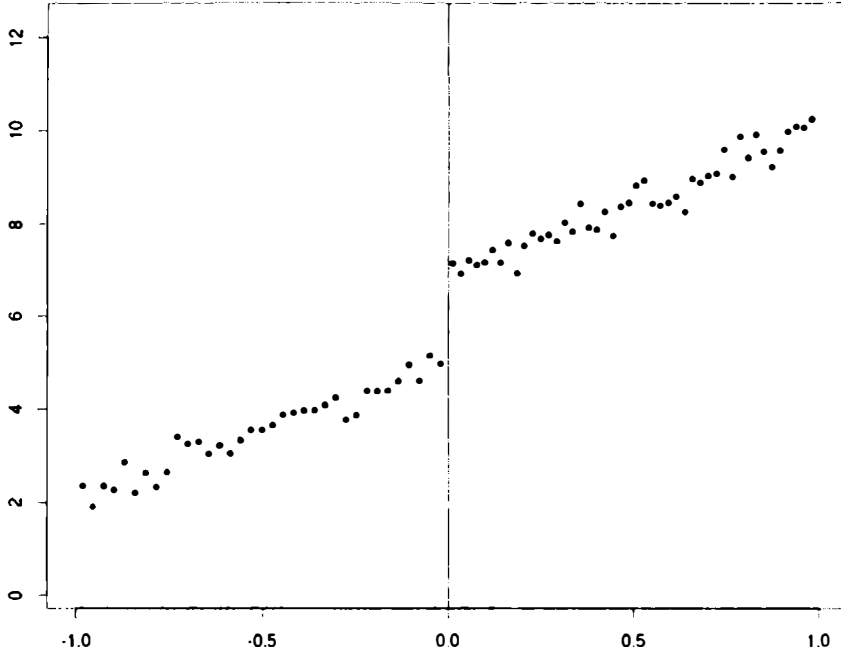
almıştır. Thistlethwaite ve Campbell (1960) çalışmasından itibaren RSA ilgili yazında sıklıkla kullanılan bir yöntem hâline gelmiştir<sup>31</sup>.

Özetle bir eşik değeri elde edilebilirse, RSA herhangi bir politika, müdahale veya uygulamanın nedensel etkilerini analiz etmek için çok kullanışlı bir yöntemdir. RSA temelde eşik değerinin biraz altında kalan bireylerle (politikadan yararlanamayan), biraz üstünde kalan bireylerin (politikadan yararlanan) birbirine benzer olduğunu, eşik değerinden uzaklaşıldıkça benzerlik derecesinin azaldığı varsayımıyla çalışır. Bu nedenle eşik değerin çok üzerinde veya çok altında kalan bireyler analizde kullanılan örnekleme alınmaz. Sadece eşik değerinin hemen üzerinde ve hemen altında olan bireylerin analize dâhil edilmesi seçim yanlılığı problemine<sup>32</sup> de çözüm sunmaktadır. Bu bağlamda RSA'nin en önemli avantajı politika dışında hedef çıktıyı etkileyebilecek diğer faktörler açısından çok benzer kontrol ve müdahale gruplarının elde edilmesidir. Regresyon süresizliği analizinin nasıl çalıştığı Şekil 17.3.1'de özetlenmiştir. Y ekseninde politikanın etkisi merak edilen hedef çıktının ortalama değerleri yer almaktadır. X ekseninde ise *marj değişkeni* ya da *zorlayıcı değişken* (forcing variable veya running variable) olarak adlandırılan değişken yer almaktadır. X ekseninde dikey bir çizgi ile gösterilen eşik değeri noktasında *marj değişkeni* 0 değerini almaktadır. RSA için kullanılan örnekleme bilindiği üzere politikadan ucu ucuna yararlanmaya hak kazanan grup (müdahale grubu) ile yararlanma hakkını ucu ucuna kaçıran grup (kontrol grubu) bulunmaktadır. Politikadan yararlanma koşulu olan *eşik değer* ile *programdan yararlanmayı belirleyen değişkenin* bireyler için aldığı değer arasındaki fark *marj değişkeni* ile temsil edilmektedir. Örneğin istihdam destek programının ücretler üzerine etkisinin araştırıldığı RSA'da bireylerin ücret düzeyi, politikanın etkisinin merak edildiği hedef çıktıdır ve bireylerin ortalama ücret değerleri Y ekseninde gösterilmektedir. Program şartlarında 30 yaş ve üzerindeki bireylerin programdan yararlanma hakkının olduğunun belirtilmesi durumunda RSA'nın *eşik değeri* (seçilim kriteri) 30'dur. *Programdan yararlanmayı belirleyen değişken* (seçilim değişkeni) yaşıdır. Marj değişkeni ise örneklemedeki bireylerin yaşlarından eşik değerinin çıkarılması ile bulunur. Bu örnekte marj değişkeni, programdan yararlanmayı belirleyen değişken yaş olduğundan yaş marjı olarak adlandırılabilir. X ekseninde yaş marjı değişkeninin aldığı değerler gösterilirken, eşik değer noktasında bu değişken 0 değerini almaktadır. Çünkü eşik değer noktasından dikey olarak yukarıya çizilen çizgi üzerinde 30 yaşındaki bireylerin ortalama ücret değerleri yer almaktadır. Aynı mantıkla 30,5 yaş yani 30 yıl 6 ayı tamamlamış bireyler için yaş marjı 0.5 ve 31 yaşındaki bireyler için yaş marjı 1'dir. Tersine X ekseninde eşik değerin sol tarafı çok az fark ile 30 yaşın altında kalıp programdan yararlanamayan bireylerin yaş marjlarını göstermektedir. Örneğin -0.5, 29 yıl 6 ayı tamamlayan bireylerin yaş marjı iken, -1.0 örneklemede 29 yaşındaki bireylerin yaş marjını temsil eden değerdir. Şekil 17.3.1'de görüldüğü gibi eşik çizgisinin sağında marj değişkeni ile hedef çıktı arasında doğrusal ve artan bir ilişki söz konusudur ve çizginin sağında doğrusal ilişkide ani bir sıçrama yaşanmıştır. Bu ani sıçrama politikanın etkisi olarak yorumlanmaktadır.

<sup>31</sup> İlgili yazında yapılmış çalışmaları Imbens ve Lemieux (2008); Lee ve Lemieux (2010) detayları ile incelemiştir.

<sup>32</sup> Seçim yanlılığı problemi 17.2.1 nolu bölümde detayları ile anlatılmıştır.

Şekil 17.3.1. RSA'nın Şekilsel Anlatımı



Politikadan yararlanma durumu marj değişkeninin deterministik ve süreksiz bir fonksiyonudur. Her bir bireyin politikadan yararlanma durumu  $D_i$  ile ifade edilirse, marj değişkenine  $(X_i - c)$  göre bireylerin politikadan yararlanıp yararlanmadığı (17.3.1) nolu eşitliklerdeki gibi ifade edilebilir. Şekil 17.3.1 anlatılırken kullanılan örnek için özelleştirilirse,  $X_i - c$  ile gösterilen marj değişkeni (yaş marjı) bireylerin tamamladığı yaşın eşik değeri ( $c$ ) olan 30'dan farkını göstermektedir. Bu durumda RSA'ya göre eşik değerinin hemen altı ve hemen üstündeki yani belli bir bant genişliği (bandwidth) içerisinde yer alan bireylerin çok benzer oldukları düşünülmektedir<sup>33</sup>. Şekil 17.3.1'de bant genişliği 1'dir çünkü yaş marjı, eşik değerinin sağında ve solunda sırasıyla (0.0)-(1.0) ve (0.0)-(-1.0) arasında değişen değerleri almaktadır.

$$D_i = \begin{cases} 1 & \text{eğer } X_i - c \geq 0 \\ 0 & \text{eğer } X_i - c < 0 \end{cases} \quad (17.3.1)$$

$D_i$  politikadan yararlanan bireyler için 1 yararlanmayanlar için 0 değerini alan bir kukla değişken olup,  $Y_i(1)$  ve  $Y_i(0)$  sırasıyla politikadan yararlanan ve yararlanmayan grup için potansiyel hedef çıktı değerlerini göstermektedir. Politikadan yararlananlar için ortalama hedef çıktı değeri;  $E[Y_i(1) | X_i]$ , eşik değerinin sağında gözlemlenebilirken, politikadan yararlanmayanların ortalama çıktı değeri;  $E[Y_i(0) | X_i]$ , eşik değerinin solunda gözlemlenmektedir. RSA ile bulunan politika etkisi, ortalama hedef çıktı değerleri arasındaki farkın politikadan yararlanma olasılığına bölümü ile elde edilir ve etki *yerel ortalama müdahale etkisi* (LATE-local average treatment

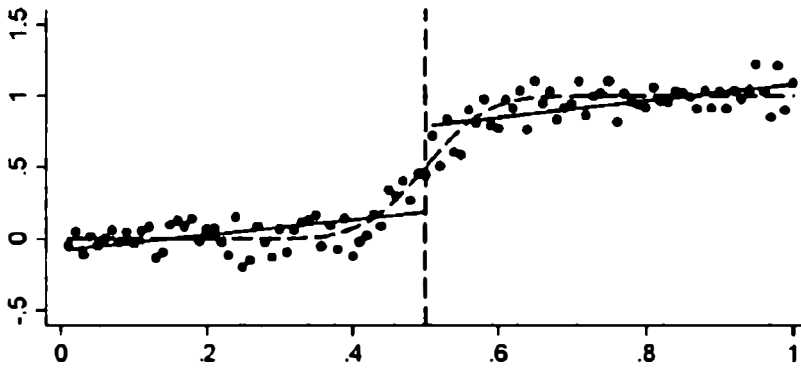
<sup>33</sup> Ancak örnekleme dahil edilen bireylere ait  $X_i$  değerlerinin, eşik değerinin ( $c$ ) ne kadar altında ya da üstünde olmasına ilişkin net bir kural yoktur.  $-h \leq X_i - c \leq h$  gösteriminde olduğu gibi eksi veya artı değerler alabilecek bant genişliğinin ( $h$ ) optimal seçimi için belli yöntemler vardır (Imbens ve Lemieux, 2008; Imbens ve Kalyanaraman, 2009).



effect) olarak adlandırılır. Burada vurgulanması gereken nokta, bu etkinin sadece eşik değerine yeterli derecede yakın olan bireyler için geçerli olduğudur.

Politikadan yararlanma durumunu belirleyen değişken ile hedef çıktı arasında doğrusal bir ilişkinin olup olmadığı da önem taşımaktadır. Örneğin Şekil 17.3.1, istihdam destek politikasından yararlanma durumunu belirleyen yaş değişkeni ile ücretler arasında doğrusal bir ilişkinin olduğu varsayımı ile çizilmiş olup, elde edilen sıçrama ya da politikanın etkinliği yanlış fonksiyon kullanmaktan kaynaklanabilir. Şekil 17.3.2’de görüldüğü gibi doğrusal ilişki olduğu varsayımıyla yapılan bir tahminde sıçrama yaşansa dahi, serpilim çizimindeki noktaların davranışı ilişkinin doğrusal olmadığını göstermektedir. Çünkü eşik değeri etrafında belirgin bir yükselme ve süreksizlik söz konusu değildir. Diğer bir ifade ile keskin bir artıştan ziyade kademeli bir artış yaşanmıştır. Bu durum politikanın etkin olmadığına işaretler. Yaş marjı ve ücret arasında doğrusal bir ilişkinin olduğu varsayımı altında politikanın etkin olduğu bulgusu, ilişki doğrusal değilse geçerli değildir. Bu nedenle marj değişkeninin farklı dereceden polinomları ile analizin gerçekleştirilmesi gerekebilir.

Şekil 17.3.2. Doğrusal İlişkinin Olmama Durumu



Diğer taraftan yüksek dereceden polinomun kullanıldığı bir regresyon süreksizliği analizi önemli bir dezavantajı beraberinde getirmektedir. Çünkü böyle bir durumda eşik değeri etrafındaki etkiyi hesaplamak amacı ile yapılan analizde eşik değerine uzak veriler kullanılmaktadır. Bu gibi nedenlerle Gelman ve Imbens (2019) doğrusal ya da en fazla ikinci dereceden polinomların kullanılmasını önermektedir.

Regresyon süreksizliği analizinde tahmin edilen ekonometrik model (17.3.2) nolu denklem ile gösterilmiştir.

$$Y_i = \beta_0 + \beta_1(X_i - c)^p + \beta_2(D_i * (X_i - c)^p) + \beta_{RSA}D_i + \varepsilon_i \quad (17.3.2)$$

$p=1,2,\dots,k$  olmak üzere 1’den k’ya kadar kullanılabilecek polinomun derecesini gösterirken,  $\beta_{RSA}$  politikanın etkinliğini gösteren katsayıdır. RSA modelinde marj değişkeni ile politikadan yararlanma durumunu gösteren  $D_i$ ’nin çarpımı olan etkileşim terimi de yer almaktadır.

Regresyon süreksizliği analizi pek çok yöntem gibi belli varsayımlar temelinde çalışan bir yöntemdir. **Birinci varsayım** daha önce verilen örneklerde de açıklandığı gibi, eşik değeri gibi kesin

bir kuralın olması, politikadan yararlananlarla yararlanmayanların bu eşik değerine göre belirleniyor olduğu varsayımdır. Ancak bazı politika uygulamalarında politikadan yararlananlarla yararlanmayanlar her zaman bu keskin hatlarla ayrılmamış olabilir. Örneğin politikadan yararlanma koşulunu sağlasa dahi kendi isteğiyle yararlanmayan ya da koşulu sağlamasa dahi politikadan yararlandırılan bireylerin olduğu politikalar gerçek hayatta karşımıza çıkabilmektedir. Bu durum göz önünde bulundurularak, RSA; *keskin* (sharp) ve *bulanık* (fuzzy) regresyon süreksizliği analizi olmak üzere iki farklı şekilde uygulanmaktadır. Keskin RSA, politikadan yararlanma durumunu belirleyen koşuldan taviz verilmeyerek koşulu sağlayanların politikadan yararlandığı, koşulu sağlamayanların ise politikadan yararlanmadığı durumda uygulanmaktadır. Diğer bir deyişle eşik değer altındaki kalan bireylerin kontrol grubunu, eşik değer üstünde kalan bireylerin ise müdahale grubunu oluşturduğu örneklem ile analiz gerçekleştirilmektedir. Koşuldan taviz verildiği durumda ise politikaların etkinliği değerlendirilirken bulanık RSA uygulanmaktadır. Aşık (2018), Türkiye’de erken emeklilik teşviklerinin genç erkek ve kadınların işgücü arzı üzerine etkilerini incelediği çalışmasında bulanık regresyon süreksizliği yöntemini kullanmış olup, bu politikanın değerlendirilmesinde neden bu yöntemin tercih edildiğini çok açık bir şekilde ifade etmiştir. Erken emeklilik teşviğinden yararlanmak bir zorunluluk olmadığı için, bazı bireyler sigortalılık süresi, yaş, prim gün sayısı gibi emeklilik koşullarını sağlasalar dahi politikadan yararlanmamayı, yani emekli olmamayı tercih edebilirler. Bu nedenle müdahale grubunda yer almama durumu bireyler arasında rassallık göstermeyeceğinden ve böylece yanlış tahminler elde edilebileceğinden Aşık (2018) bulanık RSA yöntemini kullanmıştır. Diğer bir ifade ile erken emekli olma koşullarını yerine getirmek asıl ilgilenilen bağımlı değişken olan işgücü arzını doğrudan etkileyen faktör olmamakta, sadece emekli olma kararını etkilemektedir. İşgücü arzını etkileyen asıl faktörün emekli olma durumu olmasından dolayı, içsel olan emekli olma durumu yerine dışsal olan politika şartlarını sağlama (müdahale grubuna girme) değişkeni araç değişken (instrumental variable) olarak kullanılmıştır. Aşık (2018) politikadan yararlanma durumunu gösteren değişken, yani tüm koşulları sağlayarak emekli olanların 1, olmayanların 0 değerini aldığı  $D_i$  değişkeni yerine müdahale grubuna girme durumunu gösteren değişkeni ( $Z_i$ ) araç değişken olarak kullanmıştır.  $Z_i$ , politikadan yararlanmak için şartları sağlayan bireylerin emekli olup olmadıklarına bakılmaksızın 1, şartları sağlamayanlar için 0 değerini alan değişkendir. Bulanık RSA’da araç değişken kullanıldığından iki aşamalı tahmin yapılmaktadır. İki aşamalı tahmin yöntemlerinden en çok kullanılanı İki Aşamalı En Küçük Kareler Yöntemi (2AEKKY)’dir. Bulanık RSA’da 2AEKKY ile analiz edilen modeller (17.3.3) ve (17.3.4) nolu denklemler ile gösterilmiştir.

$$D_i = \alpha_0 + \alpha_1(X_i - c)^p + \alpha_2(Z_i * (X_i - c)^p) + \alpha_3Z_i + \vartheta_i \quad (17.3.3)$$

$$Y_i = \gamma_0 + \gamma_1(X_i - c)^p + \gamma_2(\widehat{D}_i * (X_i - c)^p) + \gamma_{BRSA}\widehat{D}_i + \mu_i \quad (17.3.4)$$

Analizde  $D_i$  değişkeni yerine  $Z_i$  araç değişkeni olarak kullanılmış olup, (17.3.4) nolu modelde (17.3.3) nolu modelden elde edilen  $D_i$ ’nin tahmin değerleri ( $\widehat{D}_i$ ) kullanılmıştır.  $\gamma_{BRSA}$  katsayısı ise politikanın etkisini gösteren katsayıdır. Araç değişkeni yaklaşımının geçerli olabilmesi için seçilen araç değişkeninin zayıf olmaması ve  $D_i$  ile korelasyonunun yüksek olması gerekmektedir.

RSA'nın dayandığı **ikinci varsayım**, *seçilim değişkeni* yani bireylerin *politikadan yararlanıp yararlanamayacaklarını belirleyen* değişken ( $X_i$ ) ile hedef çıktı ( $Y_i$ ) arasında basit regresyon ilişkisinin olmasıdır. **Üçüncü varsayım** eşik değerde oluşan süreksizliğin sadece politikadan kaynaklanması gerekliliğidir. Politika dışında süreksizliğe ya da sıçramaya neden olabilecek başka faktör olmamalıdır. Bu bağlamda birbirine benzer kontrol ve müdahale gruplarının sadece seçilim kriteri açısından değil, diğer faktörler açısından da benzer oldukları varsayılır. **Dördüncü varsayım** ise bir bireyin politikadan yararlanmasının diğer bireylerin hedef çıktılarını etkilememesi gerektiğidir (Rubin, 1977). *Etkilememe prensibi* (no interference principle) olarak bilinen prensip, RSA uygulamalarında sağlanıp sağlanmadığı sıklıkla göz ardı edilen bir varsayımdır. Örneğin bir okuldaki hafta sonu kurslarının öğrenci başarısına etkisinin analiz edildiği bir çalışmada, kursa giden yani programdan yararlanan öğrencilerin diğer öğrencilerin başarısını etkileyecek şekilde onlarla iletişim hâlinde olması (kursta gördüklerini anlatarak ders çalıştırma gibi) RSA sonuçlarının güvenilirliğini olumsuz yönde etkileyebilmektedir. **Beşinci ve son varsayım** ise, süreksizliğin analiz edildiği eşik değeri bandının dışında müdahale ve kontrol gruplarına ait regresyon doğrularının paralel olması gerekliliğidir. Bu durum, kontrol ve müdahale gruplarının birbirine benzeyen bireylerden oluştuğu anlamına gelir ki, böylece kontrol grubu geçerli bir karşıt gerçeklik (counterfactual) tahmininde kullanılabilir<sup>34</sup>.

Maalesef politika etkinlik analizlerinde çoğunlukla bu varsayımlar göz ardı edilmekte ya da dikkate alınmamaktadır. Bu durumlarda RSA sonuçlarının iç geçerliliğinden<sup>35</sup> (internal validity) şüphe duyulması gerekmektedir (Cattaneo vd., 2020).

### 17.3.2. Regresyon Süreksizliği Analizi: STATA Uygulaması

1 Temmuz 2008 tarihinde 4447 sayılı İşsizlik Sigortası Kanunu'na, 5763 sayılı “İş Kanunu ve Bazı Kanunlarda Değişiklik Yapılması Hakkında Kanun” ile Geçici Madde 7 eklenmiştir (Resmi Gazete, 26 Mayıs 2008). Bu madde ile Türkiye’de ilk defa tüm yaşlardaki kadınlar ve 30 yaşın altındaki genç erkeklerin istihdamını teşvik etmek amacıyla işverenler tarafından ödenecek sigorta primlerinin devlet tarafından ödenmesi kararlaştırılmıştır. Bu çalışmada Türkiye İstatistik Kurumu (TÜİK), Gelir ve Yaşam Koşulları Anketi (GYKA), 2009<sup>36</sup> yılı yatay kesit veri seti kullanılarak 2008 yılı politikasının 30 yaşın altındaki genç erkeklerin **kayıtlı istihdam durumu** üzerine etkisi RSA ile analiz edilmiştir. Söz konusu politika, işverenlere ait sigorta primlerinin devlet tarafından ödeneceğini garanti ettiğinden politikanın kayıtlı istihdama yani bireylerin sigortalı olarak istihdam edilmelerine katkı sunup sunmadığı araştırılmıştır. 2009 yılına ait verilerin kullanılmasının nedeni, politikanın Temmuz 2008’de uygulanmış olmasına rağmen 2008 yılı GYKA verilerinin daha önceki aylarda toplanmış olmasıdır. Her yaşta kadının desteklenmesi

<sup>34</sup> Karşıt gerçeklik, bir politikadan yararlanan bireyin hedef çıktısının, politikadan yararlanmasaydı ne olabileceği ile ilgili yapılan tahmindir. Birey aynı anda hem politikadan yararlanıp hem de yararlanamayacağından, karşıt gerçeklik gözlemlenemez. Ancak ve ancak politikadan yararlanan bireylere çok benzer özellikler gösteren kontrol grupları ile tahmin edilebilir.

<sup>35</sup> İç geçerlilik, karşıt gerçekliğin doğru tahminine imkân tanıyan bir kontrol grubunun analizlerde kullanılmasıdır.

<sup>36</sup> Kitap bölümü olarak hazırlanan bu çalışmada ilgili veri seti kitabın diğer tüm bölümlerinde kullanılan veri setleri gibi okuyucularla paylaşılacağından örneklem büyüklüğü olarak veri dosyasının çok yer kaplamaması için tek yıllık örneklem kullanılmıştır. Okuyucular <https://web.tuik.gov.tr/microdata-request-system/> adresinden anketin çok yıllık panel veri versiyonunu talep edebilirler.

politika kapsamına alındığından, kadınlar açısından belli bir yaş aralığında politikadan yararlanmayan bir grup olmadığından, politikanın sadece genç erkekler üzerine etkisi **keskin RSA** ile analiz edilmiştir. 17.3.1 nolu bölümde detayları ile bahsedildiği gibi politikadan yararlanma durumunu belirleyen koşul belli olduğundan (30 yaş altında genç erkek olmak) ve sadece koşulu sağlayanlar politikadan yararlandığından **keskin RSA** gerçekleştirilmiştir. Çünkü politikadan yararlanma durumu yaş gibi bir bireysel karakteristiğin deterministik fonksiyonudur (Thistlethwaite ve Campbell, 1960; Imbens ve Lemieux, 2008; Lee ve Lemieux, 2010).

### 17.3.2.1. Analizde Kullanılan Değişkenlerin Tanımları, Tanımlayıcı İstatistikler ve RSA'nın Grafikselsel Gösterimi

RSA analizi öncesi “**describe**” komutu ile analizde kullanılan **rsa.dta**<sup>37</sup> veri setinde yer alan değişkenlerin tanımlarının sunulması ve “**sum**” komutu ile tanımlayıcı istatistiklerin gösterilmesi önemlidir. Değişkenlerin tanımları ve tanımlayıcı istatistikleri sırasıyla Tablo 17.3.1 ve Tablo 17.3.2’de gösterilmiştir. **rsa.dta** veri setinde gözlem sayısı (obs) 4,299’dur<sup>38</sup>. *f.kimlik* değişkeni veri setinde yer alan her bir bireye ait fert kimlik numaralarıdır. *vil* anket yılını göstermekte olup 2009 yılına ait yatay kesit veri seti kullanıldığından *vil* değişkeni tüm bireyler için 2009 değerini almaktadır. *yas* bireylerin tamamladığı yaşı göstermekte olup bu değişken keskin RSA için kullanılan (17.3.2) nolu modeldeki seçim değişkeni olan  $X_i$ ’dir. Politikanın etkisinin analiz edileceği hedef çıktı kayıtlı istihdam durumunu gösteren *kayıtlı\_ist* değişkenidir. (17.3.2) nolu modelde bu değişken  $Y_i$  ile gösterilmiştir. *kayıtlı\_ist*, bireyler bir sosyal güvenlik kuruluşuna kayıtlı olarak çalışıyorlarsa 1, sosyal güvenlik kuruluşuna kayıtlı değillerse 0 değerini almaktadır<sup>39</sup>. Politikadan yararlanma sınır yaşı 30 olduğundan, regresyon süresizliği analizinin eşik değeri 30 yaştır. 30 ve üstü yaştaki bireyler politikadan yararlanamazken 30 yaşın altındakiler politikadan yararlanmaktadır. Bant genişliği 3 olarak seçilmiş olup, analizde kullanılacak örnekleme 27-33 yaş arası erkekler dâhil edilmiştir. Spesifik olarak 27-29 yaş aralığındakiler programdan yararlanırken, 30-33 yaş aralığındakiler programdan yararlanamamaktadır<sup>40</sup>. Sigorta prim teşviklerinin kayıtlı istihdam üzerinde nedensel etkilerini ortaya koymaya çalışan analiz, *rassal seçime dayalı kontrollü deneylerde* (randomised controlled trials) olduğu gibi 30 yaşının biraz üstünde ve biraz altında olan bireylerin benzer bireysel karakteristikler taşıdığı ve birbirleriyle yer değiştirebilir

<sup>37</sup> **rsa.dta** veri setine karekod ile ulaşabilirsiniz.

<sup>38</sup> Stata, Türkçe dilinde çalışan bir program olmadığından ondalık ayracını nokta (.) binlik ayracını ise virgül (,) olarak kullanmaktadır. Bu bölümde STATA çıktıları yorumlanırken orijinaline sadık kalınarak sayısal ifadeler Türkçeleştirilmemiş olup, binlik ayracı olarak virgül ve ondalık ayracı olarak nokta kullanılmıştır.

<sup>39</sup> Bu bölümde incelenen STATA uygulamasında kullanılan TÜİK, Gelir ve Yaşam Koşulları Anketi (GYKA), 2009 Yatay Kesit Veri Seti’nde aşağıdaki komutlar kullanılarak kayıtlı istihdamı temsil eden *kayıtlı\_ist* değişkeni elde edilmiştir. Öncelikle orijinal veri setinde *fi190* olarak yer alan değişken *sosyal\_guvenlik* olarak adlandırılmış ve daha sonra *sosyal\_guvenlik* değişkeni yardımıyla *kayıtlı\_ist* değişkeni oluşturulmuştur.

```
rename fi190 sosyal_guvenlik
label var sosyal_guvenlik "SGK'ya kayıtlılık durumu"
label define sosyal_guvenlik1 1 "Kayıtlı" 2 "Kayıtlı Değil"
label values sosyal_guvenlik sosyal_guvenlik1
gen kayitli_ist=sosyal_guvenlik
replace kayitli_ist=0 if kayitli_ist==2
```

<sup>40</sup> Bant genişliğinin kaç olacağının seçimi araştırmacının özgürlüğündedir. Ancak 17.3.1 nolu bölümde anlatıldığı gibi eşik değerden çok uzaklaşmayacak şekilde bant genişliği belirlenmelidir. Çünkü RSA’ya göre eşik değerinin hemen altı ve hemen üstünde yer alan bireylerin çok benzer oldukları düşünülmektedir.

olduğu (*interchangeable*) varsayımıyla çalışmaktadır. Bu sayede politikadan yararlanmayı sağlayan yaş ile politikanın etkisinin merak edildiği kayıtlı istihdam durumu arasındaki ilişkide elde edilecek süreksizliğin sadece politikadan kaynaklandığı söylenebilmektedir.

**Tablo 17.3.1. Değişkenlerin Tanımları**

**describe**

Contains data from C:\Users\Asus\Desktop\rsa.dta

obs: 4,299

vars: 10

19 Sep 2020 21:58

size: 141,867

| variable name   | storage type | display format | value label                                              |
|-----------------|--------------|----------------|----------------------------------------------------------|
| yil             | int          | %8.0g          | Anket Yılı                                               |
| yas             | int          | %8.0g          | Bitirilen Yaş                                            |
| sosyal_guvenlik | byte         | %16.0g         | SGK'ya kayıtlılık durumu<br>(kayıtlı=1, kayıtlı değil=2) |
| fkimlik         | float        | %9.0g          | Fert kimlik numarası                                     |
| kayitli_ist     | float        | %9.0g          | Kayıtlı İstihdam Durumu<br>(kayıtlı=1, kayıtlı değil=0)  |
| mudahale        | float        | %9.0g          | yas<30 için mudahale=1,<br>yas>30 için mudahale=0        |
| yas_marj        | float        | %9.0g          | yas-30                                                   |
| yasmarj2        | float        | %9.0g          | yas_marj*yas_marj                                        |
| mudym           | float        | %9.0g          | mudahale*yas_marj                                        |
| mudym2          | float        | %9.0g          | mudym*mudym                                              |

**Tablo 17.3.2. Değişkenlerin Tanımlayıcı İstatistikleri**

**sum**

| Variable     | Obs   | Mean      | Std. Dev. | Min  | Max  |
|--------------|-------|-----------|-----------|------|------|
| yil          | 4,299 | 2009      | 0         | 2009 | 2009 |
| yas          | 4,299 | 29.61061  | 1.866543  | 27   | 33   |
| sosyal_guv~k | 4,299 | 1.391254  | .4880878  | 1    | 2    |
| fkimlik      | 4,299 | 2150      | 1241.159  | 1    | 4299 |
| kayitli_ist  | 4,299 | .6087462  | .4880878  | 0    | 1    |
| mudahale     | 4,299 | .5094208  | .4999694  | 0    | 1    |
| yas_marj     | 4,299 | -.3893929 | 1.866543  | -3   | 3    |
| yasmarj2     | 4,299 | 3.634799  | 3.360205  | 0    | 9    |
| mudym        | 4,299 | -1.002559 | 1.147049  | -3   | 0    |
| mudym2       | 4,299 | 2.32054   | 3.288879  | 0    | 9    |

(17.3.2) nolu modelde politikadan yararlanma durumunu gösteren  $D_i \in \{0, 1\}$ , veri setinde *mudahale* değişkenine karşılık gelmekte olup, bu değişken eşik değeri olan 30 yaşın altında olan yani politikadan yararlanma hakkına sahip bireyler için 1, eşik değerinin üstündeki bireyler için 0 değerini almaktadır<sup>41</sup>.

*yas\_marj* değişkeni, (17.3.2) nolu denklemde  $(X_i - c)^p$  ile gösterilen marj değişkeninin 1. dereceden polinomudur ( $p=1$  iken). Marj değişkeni olarak adlandırılan bu değişken bireylerin bitirdiği yaşın eşik değeri olan 30 yaştan farkını göstermektedir<sup>42</sup>. Bant genişliği 3 olduğundan, Tablo 17.3.2'de görüldüğü gibi -3 ile +3 arasında değerler almaktadır. 17.3.2 nolu bölümde bahsedildiği gibi Gelman ve Imbens (2019) doğrusal (birinci dereceden polinomun kullanıldığı) RSA ya da en fazla ikinci dereceden polinomun kullanıldığı (karesel veya kuadratik) RSA yapılmasını önermektedir. Bu bölümde incelenen örnekte birinci ve ikinci dereceden polinomların kullanıldığı RSA yapılmıştır. Marj değişkeninin ikinci dereceden polinomu ise *yasmarj2* değişkenidir<sup>43</sup>. (17.3.2) nolu denklemde  $(D_i * (X_i - c)^p)$  değişkeni politikadan yararlanma durumunu gösteren değişken ( $D_i$ ) ile marj değişkeninin  $p$  derecesindeki polinomunun etkileşimidir. İncelenen örnekte politikadan yararlanma durumunu gösteren değişken *mudahale* olduğundan ve marj değişkeninin hem birinci hem de ikinci dereceden polinomu RSA'da kullanıldığından, sırasıyla *mudahale* değişkeninin *yas\_marj* ve *yasmarj2* değişkenleri ile etkileşimini gösteren *mudym* ve *mudym2* değişkenleri analize dahil edilmiştir<sup>44</sup>. İncelenen örnekteki değişken isimleri kullanılarak (17.3.2) nolu denklem, marj değişkeninin sadece 1. derecede polinomunun kullanıldığı durumda (17.3.5) nolu model olarak ve 2. dereceden polinomun da dahil edildiği durumda (17.3.6) nolu model olarak tekrar yazılabilir.

$$Y_i = \beta_0 + \beta_1 \text{yas\_marj}_i + \beta_2 (\text{mudym}_i) + \beta_{RSA} \text{mudahale}_i + \varepsilon_i \quad (17.3.5)$$

$$Y_i = \beta_0 + \beta_1 \text{yas\_marj}_i + \beta_2 (\text{mudym}_i) + \beta_3 (\text{yasmarj2}_i) + \beta_4 (\text{mudym2}_i) + \beta_{RSA} \text{mudahale}_i + \varepsilon_i \quad (17.3.6)$$

(17.3.5) ve (17.3.6) nolu modellerin analizine geçmeden önce RSA'nın grafiksel incelemesinin yapılması önemlidir. Şekil 17.3.3'te marj değişkeninin sadece doğrusal polinomunun kullanıldığı RSA'nın grafiksel gösterimi yer almaktadır. Bu grafiği elde etmek için [17.3.1] nolu komut ça-

<sup>41</sup> TÜİK'ten alınan orijinal veri setinde *mudahale* değişkeni yer almadığından bu değişken aşağıdaki komut dizisi ile oluşturulmuştur. Değişken 27 yaş ve üstü bireyler ile 29 yaş ve altındaki bireyler için 1, 30 ve üstü yaştaki bireyler ile 33 ve altı yaştaki bireyler için 0 değerini almaktadır.

```
rename fk070 yas
```

```
label var yas "Bitirilen Yaş"
```

```
gen mudahale=.
```

```
replace mudahale=1 if yas>=27 & yas<=29
```

```
replace mudahale=0 if yas>=30 & yas<=33
```

<sup>42</sup> *yas\_marj* değişkeni bireylerin yaşından eşik değeri olan 30'un çıkarılması ile her bir birey için aşağıdaki komut ile elde edilmiştir.

```
gen yas_marj=yas-30
```

<sup>43</sup> *yasmarj2* değişkeni *yas\_marj* değişkeninin karesidir ve *gen yasmarj2=yas\_marj\*yas\_marj* komutu ile elde edilmiştir.

<sup>44</sup> *mudym* ve *mudym2* değişkenleri sırasıyla aşağıdaki komutlarla elde edilmiştir

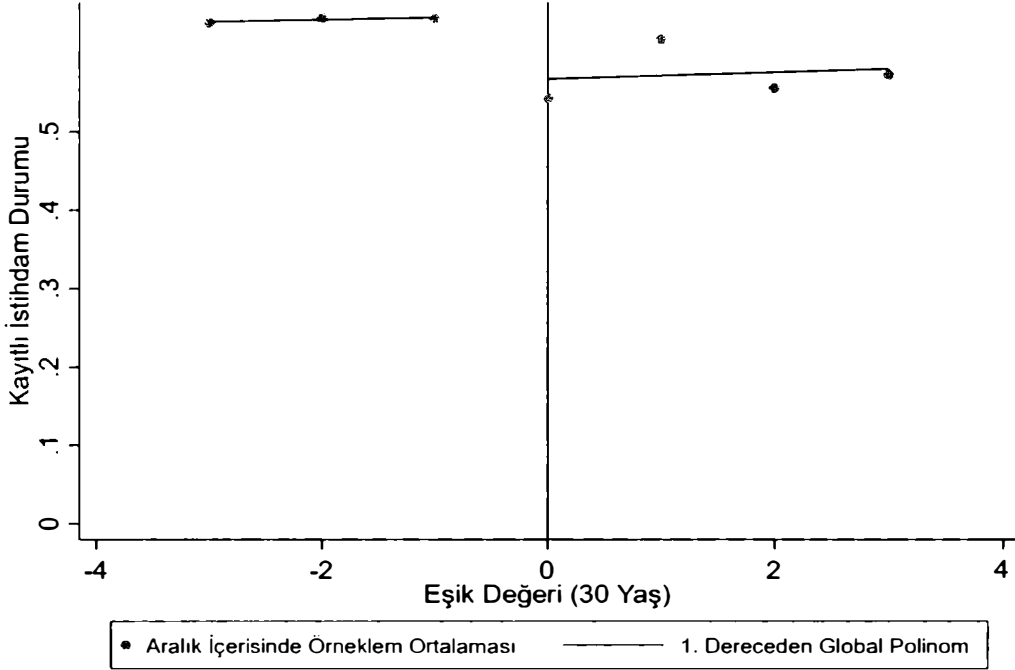
```
gen mudym=mudahale*yas_marj
```

```
gen mudym2=mudahale*yasmarj2
```

lıştırılmalıdır. Komuttaki **p(1)** polinomun 1. dereceden olduğunu göstermektedir. Bu yüzden grafik doğrusal RSA'yı yansıtmaktadır.

```
rdplot kayitli_ist yas_marj if abs(yas_marj)<=3, p(1) graph_options(xtitle("Eşik Değeri
(30 Yaş)") ytitle("Kayıtlı İstihdam Durumu") ylabel(0(0.1)0.5)) [17.3.1]
```

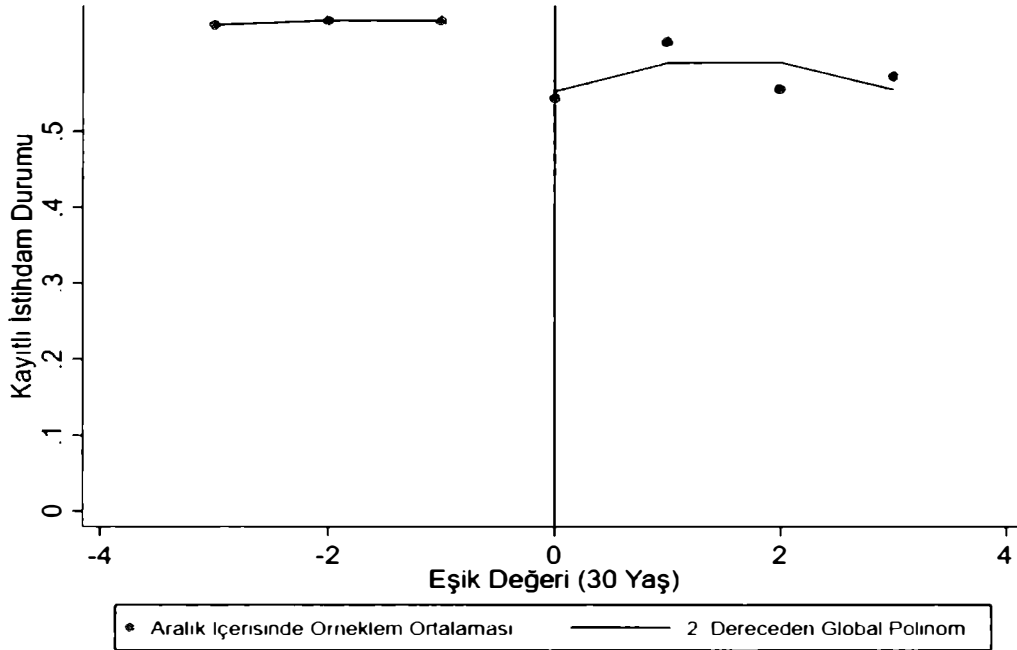
Şekil 17.3.3. Doğrusal RSA



Şekil 17.3.3'te Y eksenini hedef çıktı olan kayıtlı istihdam durumunun (*kayitli\_ist*) ortalama değerlerini gösterirken, X eksenini bireylerin bitirdiği yaşın eşik değeri olan 30'dan farkını gösteren marj değişkeninin (*yas\_marj*) aldığı değerleri göstermektedir. X ekseninde marj değişkeni 30 yaşındaki bireyler için 0 değerini almaktadır. Şekilde hedef çıktının marj değişkeni etrafındaki dağılımına bakılmaktadır. Burada politikanın müdahale grubu için kayıtlı istihdam üzerinde artırıcı bir etkiye sahip olduğunu düşünebilmek için eşik çizgisinin solunda belli bir trend izleyen hedef çıktının ortalama değerlerinin eşik çizgisinin sağında bir süreksizlik göstermesi (sıçrama yapması) gerekmektedir. Şekil 17.3.3'te görüldüğü gibi kayıtlı istihdam doğrusu eşik değerinden sonra bir süreksizlik göstermiş, aşağı doğru azalan bir sıçrama gerçekleştirmiştir. Bu durum 30 yaşının üstündeki bireylerin politikadan yararlananlardan daha az olasılıkla kayıtlı olarak istihdam edildiklerini göstermektedir. Diğer bir ifade ile 30 yaş altı bireylerin daha yüksek olasılıkla sigortalı bir işte çalıştıkları bulgusuna ulaşılmıştır. Şekil 17.3.4'te ise marj değişkeninin ikinci dereceden polinomunun kullanıldığı kuadratik (karesel) RSA grafiği gösterilmiştir. Bu grafik [17.3.2] nolu komut ile elde edilmiştir.

```
rdplot kayitli_ist yas_marj if abs(yas_marj)<=3, p(2) graph_options(xtitle("Eşik Değeri
(30 Yaş)") ytitle("Kayıtlı İstihdam Durumu") ylabel(0(0.1)0.5)) [17.3.2]
```

Şekil 17.3.4 Kuadratik RSA



Şekil 17.3.4'te eşik değerinden çizilen çizginin soluna bakıldığında marj değişkeni ile kayıtlı istihdam durumu arasında doğrusal bir ilişki varmış gibi görünse de, çizginin sağına bakıldığında doğrusal olmayan kuadratik bir ilişkinin varlığından bahsetmek güç görünmektedir. RSA modelinde kullanılan marj değişkeninin polinom derecesine karar verilmesi analizin doğru gerçekleştirilmesinde önemlidir. Grafıksel gösterim ilişkinin derecesini anlamakta sadece kabaca bir fikir vermekte olup, 17.3.2.3 nolu bölümde doğrusal RSA ile kuadratik RSA arasında seçim yapmakta kullanılan *Bilgi Kriterleri* (Information Criteria –IC) yaklaşımı anlatılmıştır.

RSA'da bir diğer önemli nokta daha önce bahsedildiği gibi eşik değerinin biraz üstünde ve altında bir bant genişliği seçmenin yanı sıra bu bant genişliğinde yer alan gözlemlerin eşik değerine göre konumlarının ağırlıklandırılmasını sağlayan kernel (çekirdek) fonksiyon türlerinden hangisinin analizde kullanılması gerektiğine karar verilmesidir. Lee ve Lemieux (2010) "*Regression Discontinuity Designs in Economics*" adlı çalışmalarında dikdörtgen (*rectangular* veya *uniform*), üçgen (*triangular*) ve epanechnikov gibi kernel fonksiyon türlerinin kullanılabileceğinden ancak basit regresyon analizi temelinde çalışan dikdörtgen kernel fonksiyonunun en etkin fonksiyon türü olmasından dolayı kullanılmasının önemini vurgulamışlardır. Örneğin üçgen kernel fonksiyonu kullanımında, bant içerisindeki gözlemler eşik değerine mutlak değer olarak yakınlıklarına göre ağırlıklandırılmakta ve ağırlıklandırılmış regresyon tahmini yapılmaktadır. Dikdörtgen fonksiyonu kullanıldığında ise bant içerisindeki her bir gözleme eşit ağırlık verilerek standart regresyon analizi gerçekleştirilmektedir. Lee ve Lemieux (2010) üçgen fonksiyon gibi sofistike bir kernel fonksiyon türünün kullanılmasından ziyade farklı bant genişlikleri ile birden fazla sayıda dikdörtgen fonksiyon kullanılarak standart regresyon analizi yapılmasını önermiştir. Kitap



bölümünde bu öneriler dikkate alınarak dikdörtgen kernel fonksiyonu temelinde RSA'nın Stata 15 programı kullanılarak uygulaması anlatılmıştır.

### 17.3.2.2. Dikdörtgen Kernel Fonksiyonu Temelinde Doğrusal RSA

Bu bölümde ilk olarak marj değişkeninin polinom derecesinin 1 (doğrusal) olarak ele alındığı (17.3.5) nolu ekonometrik modelin tahmini, dikdörtgen kernel fonksiyonu temelinde RSA ile **üç değişik yoldan** farklı Stata komutları kullanılarak gerçekleştirilmiştir. **Birinci yol** için [17.3.3] nolu komut kullanılmıştır.

**reg kayitli\_ist mudahale yas\_marj mudym if yas\_marj >= -3 & yas\_marj <= 3** [17.3.3]

Bant genişliği eşik değerinin sağında ve solunda mutlak değer olarak 3'tür. [17.3.3] nolu komutta bant genişliğinin 3 olduğu **yas\_marj >= -3 & yas\_marj <= 3** opsiyonu ile tanımlanmıştır. Tablo 17.3.3, [17.3.3] nolu komut ile gerçekleştirilen dikdörtgen kernel fonksiyonu temelinde doğrusal RSA sonuçlarını göstermektedir.

**Tablo 17.3.3.** Dikdörtgen Kernel Fonksiyonu Temelinde Doğrusal RSA ([17.3.3] komutu ile)

| reg kayitli_ist mudahale yas_marj mudym if yas_marj >= -3 & yas_marj <= 3 |            |           |            |                       |                      |          |
|---------------------------------------------------------------------------|------------|-----------|------------|-----------------------|----------------------|----------|
| Source                                                                    | SS         | df        | MS         | Number of obs = 4,299 |                      |          |
| Model                                                                     | 5.40858304 | 3         | 1.80286101 | F(3, 4295)            | = 7.60               |          |
| Residual                                                                  | 1018.50256 | 4,295     | .237136801 | Prob > F              | = 0.0000             |          |
|                                                                           |            |           |            | R-squared             | = 0.0053             |          |
|                                                                           |            |           |            | Adj R-squared         | = 0.0046             |          |
| Total                                                                     | 1023.91114 | 4,298     | .238229675 | Root MSE              | = .48697             |          |
| kayitli_ist                                                               | Coef.      | Std. Err. | t          | P> t                  | [95% Conf. Interval] |          |
| mudahale                                                                  | .0812002   | .03151    | 2.58       | 0.010                 | .0194244             | .142976  |
| yas_marj                                                                  | .0043065   | .010034   | 0.43       | 0.668                 | -.0153654            | .0239784 |
| mudym                                                                     | -.0016532  | .0161069  | -0.10      | 0.918                 | -.0332311            | .0299247 |
| _cons                                                                     | .5674007   | .0164233  | 34.55      | 0.000                 | .5352025             | .5995989 |

(17.3.5) nolu denklemde görüldüğü gibi  $\beta_{RSA}$  yani politikanın etkisini gösteren katsayı *mudahale* değişkeninin katsayısıdır. Bu katsayı uygulanan politika ile birlikte 27-29 yaşlarındaki genç erkeklerin 30-33 yaş arası erkeklerden ortalamada %8.12 daha fazla olasılıkla sigortalı olarak istihdam edildiklerini göstermektedir. **İkinci yol** olarak eşik değerinin önce solunda bulunan müdahale grubu için regresyon analizi, daha sonra sağında bulunan kontrol grubu için regresyon analizi gerçekleştirilerek elde edilen sabit terimlerin farkı alındığında (0.64860 - 0.56740=0.08120) %8.12 olan politikanın etkisine ulaşılabilir. Çünkü politikanın etkisi marj değişkeninin (*yas\_marj*) sıfır olduğu noktada ortalama hedef çıktı değerleri arasındaki farktır. (17.3.5) nolu modelin **ikinci yoldan** analizi için [17.3.4] nolu komut dizisi kullanılmıştır.

\*\*\*\* Eşik değerinin solunda bulunan müdahale grubu için regresyon analizi

```
reg kayitli_ist yas_marj if yas_marj >= -3 & yas_marj < 0
```

```
matrix katsayi_sol = e(b)
```

```
local sabitterim_sol = katsayi_sol[1, 2]
```

\*\*\*\* Eşik değerinin sağında bulunan kontrol grubu için regresyon analizi

```
reg kayitli_ist yas_marj if yas_marj >= 0 & yas_marj <= 3
```

```
matrix katsayi_sag = e(b)
```

```
local sabitterim_sag = katsayi_sag[1, 2]
```

\*\*\*\* Politikanın etkisini gösteren sabit terimler farkı

```
local fark = `sabitterim_sol' - `sabitterim_sag'
```

```
display "RSA tahmini `fark'"
```

[17.3.4]

**Tablo 17.3.4.** Dikdörtgen Kernel Fonksiyonu Temelinde Doğrusal RSA ([17.3.4] komutu ile)

```
reg kayitli_ist yas_marj if yas_marj >= -3 & yas_marj < 0
```

| Source   | SS         | df    | MS         | Number of obs | = |         |
|----------|------------|-------|------------|---------------|---|---------|
| Model    | .010516462 | 1     | .010516462 | F(1, 2188)    | = | 2,190   |
| Residual | 502.468479 | 2,188 | .229647385 | Prob > F      | = | 0.05    |
| Total    | 502.478995 | 2,189 | .22954728  | R-squared     | = | 0.8306  |
|          |            |       |            | Adj R-squared | = | 0.0000  |
|          |            |       |            | Root MSE      | = | -0.0004 |

| kayitli_ist | Coef.    | Std. Err. | t     | P> t  | [95% Conf. Interval] |
|-------------|----------|-----------|-------|-------|----------------------|
| yas_marj    | .0026533 | .0123991  | 0.21  | 0.831 | -.0216619 .0269686   |
| _cons       | .6486009 | .0264634  | 24.51 | 0.000 | .5967048 .700497     |

```
matrix katsayi_sol = e(b)
local sabitterim_sol = katsayi_sol[1, 2]
reg kayitli_ist yas_marj if yas_marj >= 0 & yas_marj <= 3
```

| Source   | SS         | df    | MS         | Number of obs | = |         |
|----------|------------|-------|------------|---------------|---|---------|
| Model    | .043681857 | 1     | .043681857 | F(1, 2107)    | = | 2,109   |
| Residual | 516.03408  | 2,107 | .244914134 | Prob > F      | = | 0.18    |
| Total    | 516.077762 | 2,108 | .244818673 | R-squared     | = | 0.6728  |
|          |            |       |            | Adj R-squared | = | 0.0001  |
|          |            |       |            | Root MSE      | = | -0.0004 |

| kayitli_ist | Coef.    | Std. Err. | t     | P> t  | [95% Conf. Interval] |
|-------------|----------|-----------|-------|-------|----------------------|
| yas_marj    | .0043065 | .0101972  | 0.42  | 0.673 | -.0156912 .0243042   |
| _cons       | .5674007 | .0166905  | 34.00 | 0.000 | .5346691 .6001322    |

```
matrix katsayi_sag = e(b)
local sabitterim_sag = katsayi_sag[1, 2]
local fark = `sabit_sol' - `sabit_sag'
display "RSA tahmini `fark'"
RSA tahmini .0812002066797439
```

**Üçüncü yol** olarak aynı sonuca **[17.3.5]** nolu komut kullanılarak ulaşılabilir. Tablo 17.3.3 ve Tablo 17.3.4'te sonuçları gösterilen ilk iki yol, RSA'nın mantığı çerçevesinde adım adım yazılmış komutlar olup, **[17.3.5]** nolu komut bu adımları arka planda uygulayarak nihai sonuca ulaştırmaktadır. Ancak bu bölümde incelenen politika örneğinin tersine, bu komut müdahale grubunun belli bir eşik değerinin üzerindeki bireylerden oluştuğu politikaların etkisini analiz etmek için dizayn edildiğinden, politika etkisini gösteren katsayıyı eksi değerli üretmektedir. İncelediğimiz politikada 27-29 yaş arasındaki bireyler politikadan yararlandığından müdahale grubu eşik değerinin soluna düşmekte olup, **rdrobust** her zaman eşik değerinin sağındaki grubu müdahale grubu olarak varsayıp onlar için etki sonucunu vermektedir. Bu nedenle, **[17.3.5]** nolu komut çalıştırıldığında elde edilen -0.0812 değerindeki katsayı, kontrol grubunun (30-33 yaş aralığındaki erkeklerin) %8.12 daha az olasılıkla kayıtlı olarak istihdam edileceği şeklinde yorumlanmalıdır.

**xi:rdrobust kayitli\_ist yas\_marj , kernel(uniform) p(1) h(3)** **[17.3.5]**

**h(3)** bant genişliğinin 3, **p(1)** polinom derecesinin 1 olduğunu, **kernel(uniform)** dikdörtgen kernel fonksiyonu kullanıldığını göstermektedir. **[17.3.5]** nolu komutla gerçekleştirilen dikdörtgen kernel fonksiyonu temelinde RSA sonuçları Tablo 17.3.5'te gösterilmiştir.

**Tablo 17.3.5.** Dikdörtgen Kernel Fonksiyonu Temelinde Doğrusal RSA ([17.3.5] komutu ile)

**xi:rdrobust kayitli\_ist yas\_marj , kernel(uniform) p(1) h(3)**

Sharp RD estimates using local polynomial regression.

|                                     |       |       |                      |  |
|-------------------------------------|-------|-------|----------------------|--|
| Cutoff c = 0   Left of c Right of c |       |       | Number of obs = 4299 |  |
| -----+-----                         |       |       | BW type = Manual     |  |
| Number of obs                       | 2190  | 2109  | Kernel = Uniform     |  |
| Eff. Number of obs                  | 2190  | 2109  | VCE method = NN      |  |
| Order est. (p)                      | 1     | 1     |                      |  |
| Order bias (q)                      | 2     | 2     |                      |  |
| BW est. (h)                         | 3.000 | 3.000 |                      |  |
| BW bias (b)                         | 3.000 | 3.000 |                      |  |
| rho (h/b)                           | 1.000 | 1.000 |                      |  |

Outcome: kayitli\_ist. Running variable: yas\_marj.

| Method       | Coef.  | Std. Err. | z       | P> z  | [95% Conf. Interval] |         |
|--------------|--------|-----------|---------|-------|----------------------|---------|
| Conventional | -.0812 | .0313     | -2.5941 | 0.009 | -.142551             | -.01985 |
| Robust       | -      | -         | -1.0777 | 0.281 | -.241069             | .070014 |

Görüldüğü gibi **[17.3.3]**, **[17.3.4]**, **[17.3.5]** komutları kullanılarak 3 farklı yoldan, (17.3.5) nolu modelin tahmini dikdörtgen kernel fonksiyonu temelinde doğrusal RSA ile gerçekleştirilmiş ve politika etkisinin %8.12 olduğuna ulaşılmıştır. Özetle uygulanan politika sayesinde 27-29 yaşlarındaki genç erkeklerin 30-33 yaşlarındaki erkeklerden ortalamada %8.12 daha fazla olasılıkla sigortalı olarak istihdam edildikleri söylenebilir. Ancak elde edilen bu politika etkisi sonucunun sağlamlılığı (stability) sınanmalıdır. Diğer bir ifade ile *yerel politika değişmezliğinin* (local po-

licy invariance) gerçekleştiğinden emin olunmalıdır. Yerel politika değişmezliği, kabul edilen eşik değerinin biraz altında veya üstünde yer alan bireyler için politika etkisinin eşik değerdeki çok ufak bir değişimden etkilenmeyeceğini ve aynı kalacağını kabul eder. Örneğin incelenen örnekte eşik değer 30 yaş olup, 29.5 yaşındaki bireyler için hesaplanan politika etkisinin eşik değer 29.5'a çekilmesi ile değişmemesi yerel politika değişmezliğidir. Dong ve Lewbel (2015) tarafından geliştirilmiş **TED** (Treatment Effect Derivative-Müdahale Etkisi Türevi) testi ile elde edilen politika etkisinin sağlam olup olmadığı ve eşik değerdeki ufak bir değişimin politikanın etkisini değiştirip değiştirmediği sınıanabilmektedir. [17.3.6] nolu **ted** komutu ile Stata 15, politika etkisinin analizinde kullanılan orijinal eşik değerinde çok küçük sayılarla dahi farklılaşmaya izin vererek orijinal eşik değeri ile hesaplanan politikanın etkisinin sağlamlılığını ortaya koymaktadır. [17.3.6] nolu komut ile elde edilen sonuçlar Tablo 17.3.6'da sunulmuştur.

**ted kayitli\_ist yas\_marj mudahale, model(sharp) h(3) c(0) m(1) k(uniform) vce(robust)**

**[17.3.6]**

Komuttaki **model(sharp)** keskin RSA'nın yapıldığını göstermektedir. Daha önce [17.3.5] nolu komutta belirtildiği gibi **h(3)** bant genişliğinin 3 olduğunu, **k(uniform)** RSA uygulamasında kernel fonksiyonu türü olarak dikdörtgen kernel fonksiyonunun kullanıldığını göstermektedir. **c(0)** ise eşik değerinde marj değişkeninin aldığı değerin 0 olduğunu belirtmektedir.

Tablo 17.3.6'da sunulan çıktılarda görüldüğü gibi ilk olarak hedef çıktı (*kayitli\_ist*) üzerinde *mudahale* (**\_T**) değişkeninin katsayısı ile gösterilen politikanın etkisi -0.0812 (-%8.12) olarak bulunmuştur. [17.3.5] nolu **rdrobust** komutu gibi, **ted** komutu da her zaman eşik değerin sağındaki grubun müdahale grubu olduğu varsayılarak tasarlandığından politika etkisini gösteren katsayının eksi değerli hesaplanmış olması normaldir. Bu nedenle elde edilen -0.0812 değerindeki katsayı, kontrol grubunun (30-33 yaşlarındaki erkeklerin) %8.12 daha az olasılıkla kayıtlı olarak istihdam edildiği şeklinde yorumlanmalıdır. Tablo 17.3.6'da "**Test of Significance of LATE**" başlığı altında ise politika etkisinin (*yerel ortalama müdahale etkisi*-local average treatment effect (LATE)) anlamlı olup olmadığına ilişkin sonuç yer almaktadır. Politika etkisini gösteren katsayının (-0.0812) anlamlı olmadığını söyleyen sıfır hipotezi %5 anlamlılık düzeyinde reddedilmiştir ( $p=0.010$ ).

"**Test of significance for TED**" başlığı altında hesaplanan katsayı ise *mudahale* ve *yas\_marj* değişkeninin etkileşiminin (**\_T\_x\_1**), yani (17.3.5) nolu modelde *mudym* olarak gösterilen değişkenin katsayısıdır. Marj değişkenindeki ufak bir değişimin politika etkisinde yaratacağı değişimi (marjinal etkiyi) gösteren bu katsayının (0.0016532) anlamlı olması, marj değişkeninin aldığı değerler eşik değerin değişmesi ile değişeceğinden, eşik değerdeki ufak bir değişimin politikanın etkisini değiştireceği anlamına gelmektedir. Ancak sonuçlarda görüldüğü gibi bu katsayının anlamlı olmadığını söyleyen sıfır hipotezi  $p=0.918$  olasılık değeri ile reddedilmemiştir. Özetle dikdörtgen kernel fonksiyonu temelli doğrusal RSA analizinin ortaya koyduğu politika etkisinin sağlam olduğu söylenebilir.

**Tablo 17.3.6.** Dikdörtgen Kernel Fonksiyonu Temelinde Doğrusal RSA için TED Testi

```

ted kayitli_ist yas_marj mudahale, model(sharp) h(3) c(0) m(1) k(uniform)
vce(robust)
(4,299 real changes made)

***** DISCONTINUITY IN THE OUTCOME *****

(sum of wgt is 1.000000000000075)
Linear regression Number of obs = 4,299
 F(3, 4295) = 7.59
 Prob > F = 0.0000
 R-squared = 0.0053
 Root MSE = .48697

```

| kayitli_ist | Coef.     | Robust Std. Err. | t     | P> t  | [95% Conf. Interval] |           |
|-------------|-----------|------------------|-------|-------|----------------------|-----------|
| _x_1        | .0026533  | .0124082         | 0.21  | 0.831 | -.0216731            | .0269798  |
| _T          | -.0812002 | .0313102         | -2.59 | 0.010 | -.1425843            | -.0198161 |
| _T_x_1      | .0016532  | .0160873         | 0.10  | 0.918 | -.0298863            | .0331926  |
| _cons       | .6486009  | .0264543         | 24.52 | 0.000 | .5967368             | .700465   |

```

(2,190 missing values generated)
(2,109 missing values generated)
(0 real changes made)

***** Test of significance for LATE *****
LATE: _b[_T]

```

| kayitli_ist | Coef.     | Std. Err. | z     | P> z  | [95% Conf. Interval] |           |
|-------------|-----------|-----------|-------|-------|----------------------|-----------|
| LATE        | -.0812002 | .0313102  | -2.59 | 0.010 | -.142567             | -.0198334 |

```

***** Test of significance for TED *****
TED: _b[_T_x_1]

```

| kayitli_ist | Coef.    | Std. Err. | z    | P> z  | [95% Conf. Interval] |          |
|-------------|----------|-----------|------|-------|----------------------|----------|
| TED         | .0016532 | .0160873  | 0.10 | 0.918 | -.0298774            | .0331837 |

TED testinin yanı sıra, RSA sonuçlarının sağlamlığını test etmenin bir diğer yolu farklı bant genişliklerinde analizi tekrarlamaktır. Örnek olarak Tablo 17.3.7'de, bant genişliğinin 2 yıl olarak alındığı dikdörtgen kernel fonksiyonu temelli doğrusal RSA sonuçları verilmiştir. Bu sonuçların elde edilmesi için [17.3.7] nolu komut kullanılmıştır.

**reg kayitli\_ist mudahale yas\_marj mudym if yas\_marj>= -2 & yas\_marj <= 2 [17.3.7]**

Tablo 17.3.7’de görüldüğü gibi politikanın etkisini gösteren *mudahale* değişkeninin katsayısı 0.079 bulunmuştur. Bu sonuç 2008 yılı istihdam teşvik politikası sayesinde 27-29 yaşlarındaki bireylerin sigortalı olarak istihdam edilme olasılığının 30-33 yaşlarındaki bireylere göre %7.9 daha yüksek olduğunu göstermektedir. Bant genişliğinin 3 yıl olduğu analizlerde politika etkisi %8.12 bulunmuş olup, bant genişliğinin 2 yıl olduğu durumda çok farklı bir sonuç elde edilmiştir. Özetle politikanın müdahale grubu bireylerinin kayıtlı olarak istihdam edilme olasılığını artırdığı bulgusunun sağlam olduğu söylenebilir.

**Tablo 17.3.7.** Dikdörtgen Kernel Fonksiyonu Temelinde Doğrusal RSA  
(Bant Genişliği:2 Yıl)

`reg kayitli_ist mudahale yas_marj mudym if yas_marj>= -2 & yas_marj <= 2`

| Source   | SS         | df    | MS         | Number of obs | = | 3,254  |
|----------|------------|-------|------------|---------------|---|--------|
| Model    | 4.31805686 | 3     | 1.43935229 | F(3, 3250)    | = | 6.05   |
| Residual | 772.815625 | 3,250 | .237789423 | Prob > F      | = | 0.0004 |
|          |            |       |            | R-squared     | = | 0.0056 |
|          |            |       |            | Adj R-squared | = | 0.0046 |
| Total    | 777.133682 | 3,253 | .238897535 | Root MSE      | = | .48764 |

| kayitli_ist | Coef.     | Std. Err. | t     | P> t  | [95% Conf. Interval] |
|-------------|-----------|-----------|-------|-------|----------------------|
| mudahale    | .0799269  | .0432154  | 1.85  | 0.064 | -.0048053 .1646591   |
| yas_marj    | .0091289  | .0143894  | 0.63  | 0.526 | -.0190843 .0373421   |
| mudym       | -.0097067 | .0292127  | -0.33 | 0.740 | -.0669839 .0475704   |
| _cons       | .5644505  | .0176118  | 32.05 | 0.000 | .5299193 .5989818    |

### 17.3.2.3. Dikdörtgen Kernel Fonksiyonu Temelinde Kuadratik RSA

Marj değişkeninin polinom derecesinin 2 olarak varsayıldığı (17.3.6) nolu model, dikdörtgen kernel fonksiyonu temelinde RSA ile üç farklı yoldan tahmin edilebilir. **Birinci yol** için [17.3.8] nolu komut çalıştırılmış ve elde edilen sonuçlar Tablo 17.3.8’de gösterilmiştir.

`reg kayitli_ist mudahale yas_marj yasmarj2 mudym mudym2 if yas_marj >= -3 & yas_marj <= 3` [17.3.8]

Tablo 17.3.8’de görüldüğü gibi politikanın etkisini gösteren katsayı (0.0855273) istatistiksel olarak anlamlı değildir ( $p=0.288$ ). Bu sonuç istihdam teşvik politikasının müdahale grubu olan 27-29 yaş arası genç erkeklerin sigortalı olarak istihdam edilmelerinde etkin olmadığını göstermektedir.

**Tablo 17.3.8.** Dikdörtgen Kernel Fonksiyonu Temelinde Kuadratik RSA  
([17.3.8] nolu komut ile)

**reg kayitli\_ist mudahale yas\_marj yasmarj2 mudym mudym2 if yas\_marj >= -3 & yas\_marj <= 3**

| Source   | SS         | df    | MS         | Number of obs | = | 4,299  |
|----------|------------|-------|------------|---------------|---|--------|
| Model    | 6.08335165 | 5     | 1.21667033 | F(5, 4293)    | = | 5.13   |
| Residual | 1017.82779 | 4,293 | .237090098 | Prob > F      | = | 0.0001 |
|          |            |       |            | R-squared     | = | 0.0059 |
|          |            |       |            | Adj R-squared | = | 0.0048 |
| Total    | 1023.91114 | 4,298 | .238229675 | Root MSE      | = | .48692 |

| kayitli_ist | Coef.     | Std. Err. | t     | P> t  | [95% Conf. Interval] |
|-------------|-----------|-----------|-------|-------|----------------------|
| mudahale    | .0855273  | .0805162  | 1.06  | 0.288 | -.072326 .2433807    |
| yas_marj    | .0555241  | .0320842  | 1.73  | 0.084 | -.0073774 .1184257   |
| yasmarj2    | -.0182552 | .0108621  | -1.68 | 0.093 | -.0395505 .0030401   |
| mudym       | -.0659417 | .095583   | -0.69 | 0.490 | -.2533337 .1214504   |
| mudym2      | .0149753  | .0248685  | 0.60  | 0.547 | -.0337798 .0637305   |
| _cons       | .5522903  | .0187219  | 29.50 | 0.000 | .5155858 .5889948    |

İkinci yoldan analiz için ise [17.3.9] nolu komut dizini çalıştırılarak Tablo 17.3.9'da gösterilen sonuçlar elde edilmiştir. Eşik değerinin önce solunda bulunan müdahale grubu için regresyon analizi, daha sonra sağında bulunan kontrol grubu için regresyon analizi gerçekleştirilerek elde edilen sabit terimlerin farkı alındığında ( $0.6378177 - 0.5522903 = 0.0855274$ ) %8.55 olan politikanın etkisine ulaşılmıştır. Birinci yol ile elde edilen sonucun aynısı olup, ikinci yolun tek dezavantajı politikanın etkisini gösteren katsayının anlamlı olup olmadığını Tablo 17.3.9'da gösterilen Stata çıktısından anlaşılamamasıdır. Ancak birinci yol zaten politikanın etkisini gösteren katsayının anlamsız olduğunu ortaya koyduğundan, burada ilave bir yorum yapmaya gerek yoktur.

\*\*\*\* Eşik değerinin solunda bulunan müdahale grubu için regresyon analizi

**reg kayitli\_ist yas\_marj yasmarj2 if yas\_marj >= -3 & yas\_marj < 0**

**matrix katsayi\_sol = e(b)**

**local sabitterim\_sol = katsayi\_sol[1, 3]**

\*\*\*\* Eşik değerinin solunda bulunan kontrol grubu için regresyon analizi

**reg kayitli\_ist yas\_marj yasmarj2 if yas\_marj >= 0 & yas\_marj <= 3**

**matrix katsayi\_sag = e(b)**

**local sabitterim\_sag = katsayi\_sag[1, 3]**

\*\*\*\* Politikanın etkisini gösteren sabit terimler farkı

**local fark = `sabitterim\_sol' - `sabitterim\_sag'**

**display "RSA tahmircisi `fark'"**

**[17.3.9]**

**Tablo 17.3.9.** Dikdörtgen Kernel Fonksiyonu Temelinde Kuadratik RSA  
([17.3.9] nolu komut ile)

```
reg kayitli_ist yas_marj yasmarj2 if yas_marj >= -3 & yas_marj < 0
```

| Source   | SS         | df    | MS         | Number of obs | = | 2,190   |
|----------|------------|-------|------------|---------------|---|---------|
|          |            |       |            | F(2, 2187)    | = | 0.03    |
| Model    | .015612879 | 2     | .00780644  | Prob > F      | = | 0.9666  |
| Residual | 502.463383 | 2,187 | .229750061 | R-squared     | = | 0.0000  |
|          |            |       |            | Adj R-squared | = | -0.0009 |
| Total    | 502.478995 | 2,189 | .22954728  | Root MSE      | = | .47932  |

| kayitli_ist | Coef.     | Std. Err. | t     | P> t  | [95% Conf. Interval] |
|-------------|-----------|-----------|-------|-------|----------------------|
| yas_marj    | -.0104175 | .0886326  | -0.12 | 0.906 | -.1842305 .1633954   |
| yasmarj2    | -.0032799 | .0220219  | -0.15 | 0.882 | -.046466 .0399062    |
| _cons       | .6378177  | .0770876  | 8.27  | 0.000 | .486645 .7889903     |

```
matrix katsayi_sol = e(b)
```

```
local sabitterim_sol = katsayi_sol[1, 3]
```

```
reg kayitli_ist yas_marj yasmarj2 if yas_marj >=0 & yas_marj <= 3
```

| Source   | SS         | df    | MS         | Number of obs | = | 2,109  |
|----------|------------|-------|------------|---------------|---|--------|
|          |            |       |            | F(2, 2106)    | = | 1.46   |
| Model    | .713354046 | 2     | .356677023 | Prob > F      | = | 0.2330 |
| Residual | 515.364408 | 2,106 | .244712444 | R-squared     | = | 0.0014 |
|          |            |       |            | Adj R-squared | = | 0.0004 |
| Total    | 516.077762 | 2,108 | .244818673 | Root MSE      | = | .49468 |

| kayitli_ist | Coef.     | Std. Err. | t     | P> t  | [95% Conf. Interval] |
|-------------|-----------|-----------|-------|-------|----------------------|
| yas_marj    | .0555241  | .0325958  | 1.70  | 0.089 | -.0083992 .1194475   |
| yasmarj2    | -.0182552 | .0110353  | -1.65 | 0.098 | -.0398964 .003386    |
| _cons       | .5522903  | .0190204  | 29.04 | 0.000 | .5149895 .5895911    |

```
matrix katsayi_sag = e(b)
```

```
local sabitterim_sag = katsayi_sag[1, 3]
```

```
local fark = `sabitterim_sol' - `sabitterim_sag'
```

```
display "RSA tahmini `fark'"
```

```
RSA tahmini .0855273489446241
```

Son olarak üçüncü yoldan dikdörtgen kernel fonksiyonu temelli kuadratik RSA için [17.3.10] nolu **rdrobust** komutu çalıştırılarak Tablo 17.3.10'da gösterilen sonuçlar elde edilmiştir.

**xi:rdrobust kayitli\_ist yas\_marj , kernel(uniform) p(2) h(3)**

**[17.3.10]**

Aynı şekilde üçüncü yoldan yapılan dikdörtgen kernel fonksiyonu temelli RSA'da politikanın etkisi 0.08553 (%8.55) olarak tahmin edilmiştir ancak Tablo 17.3.10'da sunulduğu gibi ilgili katsayı istatistiksel olarak anlamsızdır ( $p=0.281$ ).



**Tablo 17.3.10.** Dikdörtgen Kernel Fonksiyonu Temelinde Kuadratik RSA

([17.3.10] nolu komut ile)

**xi:rdrobust kayitli\_ist yas\_marj , kernel(uniform) p(2) h(3)**

Preparing data.

Computing variance-covariance matrix.

Computing RD estimates.

Estimation completed.

Sharp RD estimates using local polynomial regression.

|                                     |       |       |                 |         |
|-------------------------------------|-------|-------|-----------------|---------|
| Cutoff c = 0   Left of c Right of c |       |       | Number of obs = | 4299    |
| -----+-----                         |       |       | NN matches =    | 3       |
| Number of obs                       | 2190  | 2109  | BW type =       | Manual  |
| Order loc. poly. (p)                | 2     | 2     | Kernel type =   | Uniform |
| Order bias (q)                      | 3     | 3     |                 |         |
| BW loc. poly. (h)                   | 3.000 | 3.000 |                 |         |
| BW bias (b)                         | 3.000 | 3.000 |                 |         |
| rho (h/b)                           | 1.000 | 1.000 |                 |         |

Outcome: kayitli\_ist. Running variable: yas\_marj.

| Method       | Coef.   | Std. Err. | z       | P> z  | [95% Conf. Interval] |          |
|--------------|---------|-----------|---------|-------|----------------------|----------|
| -----+-----  |         |           |         |       |                      |          |
| Conventional | -.08553 | .0793     | -1.0786 | 0.281 | -.240945             | .06989   |
| Robust       | -       | -         | -2.6811 | 0.007 | -.173606             | -.026974 |
| -----+-----  |         |           |         |       |                      |          |

Üç farklı yoldan gerçekleştirilen kuadratik regresyon süreksizliği analizlerinin tersine doğrusal RSA ile elde edilen sonuçlarda politikanın etkin olduğu sonucuna ulaşılmış olup, doğrusal ve kuadratik RSA sonuçlarının birbirinden farklılaştığı görülmektedir. Böyle bir durumda Bilgi Kriterleri (Information Criteria–IC) yaklaşımı takip edilerek bilgi kriterlerinin tavsiyesi doğrultusunda kuadratik RSA ve doğrusal RSA arasında seçim yapılması gerekmektedir.

[17.3.11] nolu komut dizini kullanılarak kuadratik RSA ve doğrusal RSA arasında seçim yapılmasını sağlayan bilgi kriterleri elde edilmiştir.

```
reg kayitli_ist mudahale yas_marj mudym if yas_marj >= -3 & yas_marj <= 3
estat ic
reg kayitli_ist mudahale yas_marj yasmarj2 mudym mudym2 if yas_marj >= -3 &
yas_marj <= 3
estat ic
```

**[17.3.11]**

Elde edilen bilgi kriteri değerleri Tablo 17.3.9’gösterilmiştir.

**Tablo 17.3.11.** Dikdörtgen Kernel Fonksiyonu Temelinde Doğrusal ve Kuadratik RSA Karşılaştırılması (Bilgi Kriterleri Yaklaşımı ile)

```
reg kayitli_ist mudahale yas_marj mudym if yas_marj >= -3 & yas_marj <= 3
```

| Source   | SS         | df    | MS         | Number of obs | = | 4,299  |
|----------|------------|-------|------------|---------------|---|--------|
|          |            |       |            | F(3, 4295)    | = | 7.60   |
| Model    | 5.40858304 | 3     | 1.80286101 | Prob > F      | = | 0.0000 |
| Residual | 1018.50256 | 4,295 | .237136801 | R-squared     | = | 0.0053 |
|          |            |       |            | Adj R-squared | = | 0.0046 |
| Total    | 1023.91114 | 4,298 | .238229675 | Root MSE      | = | .48697 |

| kayitli_ist | Coef.     | Std. Err. | t     | P> t  | [95% Conf. Interval] |
|-------------|-----------|-----------|-------|-------|----------------------|
| mudahale    | .0812002  | .03151    | 2.58  | 0.010 | .0194244 .142976     |
| yas_marj    | .0043065  | .010034   | 0.43  | 0.668 | -.0153654 .0239784   |
| mudym       | -.0016532 | .0161069  | -0.10 | 0.918 | -.0332311 .0299247   |
| _cons       | .5674007  | .0164233  | 34.55 | 0.000 | .5352025 .5995989    |

estat ic

Akaike's information criterion and Bayesian information criterion

| Model | Obs   | ll(null)  | ll(model) | df | AIC      | BIC      |
|-------|-------|-----------|-----------|----|----------|----------|
| .     | 4,299 | -3016.016 | -3004.631 | 4  | 6017.263 | 6042.728 |

Note: N=Obs used in calculating BIC; see [R] BIC note.

```
reg kayitli_ist mudahale yas_marj yasmarj2 mudym mudym2 if yas_marj >= -3 & yas_marj <= 3
```

| Source   | SS         | df    | MS         | Number of obs | = | 4,299  |
|----------|------------|-------|------------|---------------|---|--------|
|          |            |       |            | F(5, 4293)    | = | 5.13   |
| Model    | 6.08335165 | 5     | 1.21667033 | Prob > F      | = | 0.0001 |
| Residual | 1017.82779 | 4,293 | .237090098 | R-squared     | = | 0.0059 |
|          |            |       |            | Adj R-squared | = | 0.0048 |
| Total    | 1023.91114 | 4,298 | .238229675 | Root MSE      | = | .48692 |

| kayitli_ist | Coef.     | Std. Err. | t     | P> t  | [95% Conf. Interval] |
|-------------|-----------|-----------|-------|-------|----------------------|
| mudahale    | .0855273  | .0805162  | 1.06  | 0.288 | -.072326 .2433807    |
| yas_marj    | .0555241  | .0320842  | 1.73  | 0.084 | -.0073774 .1184257   |
| yasmarj2    | -.0182552 | .0108621  | -1.68 | 0.093 | -.0395505 .0030401   |
| mudym       | -.0659417 | .095583   | -0.69 | 0.490 | -.2533337 .1214504   |
| mudym2      | .0149753  | .0248685  | 0.60  | 0.547 | -.0337798 .0637305   |
| _cons       | .5522903  | .0187219  | 29.50 | 0.000 | .5155858 .5889948    |

estat ic

Akaike's information criterion and Bayesian information criterion

| Model | Obs   | ll(null)  | ll(model) | df | AIC      | BIC      |
|-------|-------|-----------|-----------|----|----------|----------|
| .     | 4,299 | -3016.016 | -3003.207 | 6  | 6018.414 | 6056.611 |

Note: N=Obs used in calculating BIC; see [R] BIC note.

İlk olarak doğrusal regresyon süreksizliği analizi yapılarak çalıştırılan “estat ic” komutu ile AIC<sup>45</sup> ve BIC<sup>46</sup> değerleri sırasıyla 6017.263 ve 6042.728 olarak hesaplanmıştır. Kuadratik RSA sonrası “estat ic” komutu ile elde edilen AIC ve BIC değerleri ise 6018.414 ve 6056.611’dir. Doğrusal RSA sonrası elde edilen bilgi kriterleri daha küçük sayısal değerler aldığından, *Bilgi Kriterleri* yaklaşımına göre doğrusal RSA’nın kuadratik RSA’ya tercih edilmesi gerekmektedir. Bu bağlamda doğrusal RSA ile yapılan analiz sonuçlarının geçerli olduğunu ve istihdam teşvik politikasının müdahale grubu bireyleri üzerinde kayıtlı istihdamı artırıcı pozitif etkilerinin olduğunu söylemek yerinde olacaktır.

### 17.3.3. Regresyon Süreksizliği Analizi Yönteminde Kullanılan Stata Komutları

17.3 nolu bölümde kullanılan Stata komutları aşağıda toplu olarak yer almaktadır.

```
clear all
set more off
use "DOSYAKONUMU\rsa.dta" , clear
ssc install rdrobust
ssc install rdplot
describe
sum
rdplot kayitli_ist yas_marj if abs(yas_marj)<=3, p(1) graph_options(xtitle("Eşik
Değeri (30 Yaş)") ytitle("Kayıtlı İstihdam Durumu") ylabel(0(0.1)0.5))

rdplot kayitli_ist yas_marj if abs(yas_marj)<=3, p(2) graph_options(xtitle("Eşik
Değeri (30 Yaş)") ytitle("Kayıtlı İstihdam Durumu") ylabel(0(0.1)0.5))

reg kayitli_ist mudahale yas_marj mudym if yas_marj >= -3 & yas_marj <= 3

reg kayitli_ist yas_marj if yas_marj >= -3 & yas_marj < 0
matrix katsayi_sol = e(b)
local sabitterim_sol = katsayi_sol[1, 2]
reg kayitli_ist yas_marj if yas_marj >= 0 & yas_marj <= 3
matrix katsayi_sag = e(b)
local sabitterim_sag = katsayi_sag[1, 2]
local fark = `sabitterim_sol' - `sabitterim_sag'
display "RSA tahmincisi `fark'"

xi:rdrobust kayitli_ist yas_marj , kernel(uniform) p(1) h(3)

ted kayitli_ist yas_marj mudahale, model(sharp) h(3) c(0) m(1) k(uniform)
vce(robust)

reg kayitli_ist mudahale yas_marj mudym if yas_marj>= -2 & yas_marj <= 2

reg kayitli_ist mudahale yas_marj yasmarj2 mudym mudym2 if yas_marj >= -3 & yas_marj
<= 3
reg kayitli_ist yas_marj yasmarj2 if yas_marj >= -3 & yas_marj < 0
matrix katsayi_sol = e(b)
```

<sup>45</sup> Akaike Bilgi Kriteri (Akaike's Information Criterion)

<sup>46</sup> Bayes Bilgi Kriteri (Bayesian Information Criterion)

```

local sabitterim_sol = katsayi_sol[1, 3]
reg kayitli_ist yas_marj yasmarj2 if yas_marj >= 0 & yas_marj <= 3
matrix katsayi_sag = e(b)
local sabitterim_sag = katsayi_sag[1, 3]
local fark = `sabitterim_sol' - `sabitterim_sag'
display "RSA tahmincisi `fark'"

xi:rdrobust kayitli_ist yas_marj , kernel(uniform) p(2) h(3)

reg kayitli_ist mudahale yas_marj mudym if yas_marj >= -3 & yas_marj <= 3
estat ic
reg kayitli_ist mudahale yas_marj yasmarj2 mudym mudym2 if yas_marj >= -3 & yas_marj
<= 3
estat ic

```

## 17.4. BÖLÜM DEĞERLENDİRMESİ

Farkların Farkı (FF), Eğilim Skoru Eşleştirme (ESE) ve Regresyon Süreksizliği Analizi (RSA) yöntemlerinin incelendiği 17. bölümde, uygulanan politikaların etkinliğinin Stata 15 programı ile nasıl analiz edileceğine temel bir bakış açısı geliştirilmiştir. Sonuç olarak rassal atama ya da rassal seçim yöntemi ile tam deneysel analizlerin yapılamadığı durumlarda FF, ESE ve RSA yöntemleri politikaların nedensel etkisini ortaya koymakta kullanılan yarı deneysel yöntemler olarak karşımıza çıkmaktadır. Birinin diğerine üstünlüğü olmayan bu yöntemlerin hangisinin kullanılacağı her zaman araştırma sorusuna ve elde bulunan veri setinin yapısına bağlıdır. Bazı politikaların etkinliği için FF en uygun yöntem olabilirken, katılım şartları idari kurallar ve kanunlar ile net bir şekilde belirlenmiş politikaların etkinlik analizlerinde RSA'nın daha çok tercih edildiği görülmektedir. Bazen FF-ESE gibi bu yöntemlerin karma bir şekilde kullanılmasının daha sağlam sonuçlara sebebiyet verdiği literatürde sıklıkla belirtilmektedir.

## 17.5. KAYNAKLAR

- Abraham, S. ve Sun, L. (2018). Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects. Papers 1804.05785, arXiv.org.
- Angrist, J. D. ve Lavy, V. (1999). Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement. *The Quarterly Journal of Economics*, 114(2), 533-575
- Angrist, J.D. ve Pischke, J.S. (2008). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.
- Aşık, G. (2018). *Impact of Early Retirement Incentives on Labor Supply of Young Men and Women: Evidence From Turkey*. UMR Développement et Sociétés, Université Paris 1 Panthéon-Sorbonne, Institut de Recherche pour le Développement.
- Becker, S.O. ve Ichino, A. (2002) Estimation of Average Treatment Effects Based on Propensity Scores. *Stata Journal* 2(4): 358–377.
- Bronzini, R., ve Iachini, E. (2014). Are incentives for R&D effective? Evidence from a regression discontinuity approach. *American Economic Journal: Economic Policy*, 6(4), 100-134.
- Caliendo, M., ve Kopeinig, S. (2008). Some Practical Guidance for The Implementation of Propensity Score Matching. *Journal of Economic Surveys*, 22(1), 31-72.

- Cattaneo, M. D., Keele, L., Titiunik, R., ve Vazquez-Bare, G. (2020). Extrapolating Treatment Effects in Multi-cutoff Regression Discontinuity Designs. *Journal of the American Statistical Association*, 1-12. <https://doi.org/10.1080/01621459.2020.1751646>
- D'Agostino Jr, R. B. (1998). Propensity Score Methods for Bias Reduction in The Comparison of A Treatment to A Non-Randomized Control Group. *Statistics in Medicine*, 17(19), 2265-2281.
- Dehejia, R. H. ve Wahba, S. (2002). Propensity Score-Matching Methods for Nonexperimental Causal Studies. *Review of Economics and Statistics*, 84(1), 151-161.
- DiNardo, J. ve Tobias, J. (2001) Nonparametric Density and Regression Estimation. *Journal of Economic Perspectives* 15(4): 11-28.
- Dong, Y. ve Lewbel, A. (2015). Identifying The Effect of Changing The Policy Threshold in Regression Discontinuity Models. *Review of Economics and Statistics*, 97(5), 1081-1092.
- Eren, O. (2007). Measuring the Union–Nonunion Wage Gap Using Propensity Score Matching. *Industrial Relations: A Journal of Economy and Society*, 46(4), 766-780.
- Fan J., ve Gijbels I. (1996). *Local Polynomial Modelling and Its Applications*. London; New York and Melbourne: Chapman and Hall
- Gelman, A., ve Imbens, G. (2019). Why High-Order Polynomials Should Not Be Used in Regression Discontinuity Designs. *Journal of Business & Economic Statistics*, 37(3), 447-456.
- Heckman, J.J., Ichimura, H. ve Todd, P. (1998). Matching as An Econometric Evaluation Estimator. *Review of Economic Studies*, 65(2), 261-294.
- HM Government (2020). Coronavirus: Claim For Your Employees' Wages Through The Coronavirus Job Retention Scheme A Step By Step Guide For Employers. Erişim Tarihi: 16.06.2020  
[https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/894161/Coronavirus\\_Job\\_Retention\\_Scheme\\_step\\_by\\_step\\_guide\\_for\\_employers.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/894161/Coronavirus_Job_Retention_Scheme_step_by_step_guide_for_employers.pdf)
- Hupkau, C. ve Petrongolo, B. (2020). Work, Care and Gender during the Covid-19 Crisis. Center for Economic Performance Report, No: 002. London Schhol of Economics, UK. Erişim Tarihi: 20.06.2020 <http://cep.lse.ac.uk/pubs/download/cepcovid-19-002.pdf>
- Imbens, G.W. ve Lemieux, T. (2008). Regression Discontinuity Designs: A Guide to Practice. *Journal of Econometrics*, 142(2), 615-635.
- Jann, B. (2017). kmatch: Kernel Matching with Automatic Bandwidth Selection. United Kingdom Stata Users' Group Meetings 2017 11, Stata Users Group.
- Lechner, M. (2011). The Estimation of Causal Effects by Difference-in-Difference Methods. *Foundations and Trends in Econometrics*, 4(3), 165-224.
- Lee, D. ve Lemieux, T. (2010). Regression Discontinuity Designs in Economics. *Journal of Economic Literature, American Economic Association*, 48(2) 281-355.
- Resmi Gazete ( 26.05.2008). T.C. 5763 sayılı, İş Kanunu ve Bazı Kanunlarda Değişiklik Yapılması Hakkında Kanun. *Ankara: (26887 Sayılı Resmi Gazete)*. Erişim Tarihi: 10.07.2020  
<https://www.resmigazete.gov.tr/eskiler/2008/05/20080526.htm>.
- Rose, A. M. (1952). Needed Research on the Mediation of Labour Disputes. *Personal Psychology*, 5, 187-200.
- Rosenbaum, P.R ve Rubin, B. (1983). The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika*, 70(1), 41-55
- Rosenbaum, P. ve Rubin, D. (1985) Constructing A Control Group Using Multivariate Matched Sampling Methods That Incorporate The Propensity Score. *The American Statistician* 39(1): 33-38.

- Rubin, D.B. (1977). Assignment to Treatment Group On The Basis Of Covariate. *Journal of Educational Statistics*, 2, 1-26.
- Rubin, D. B. (1980). Bias Reduction Using Mahalanobis-Metric Matching. *Biometrics*, 293-298.
- Simon, J. L. (1966). The Price Elasticity of Liquor in The U.S. and a Simple Method of Determination. *Econometrica*, 34, 193-205.
- Smith, J. (2000) A Critical Survey of Empirical Methods for Evaluating Active Labor Market Policies. *Schweizerische Zeitschrift fuer Volkswirtschaft und Statistik* 136(3): 1–22.
- Smith, J., Todd, P. 2005. Does Matching Overcome LaLonde's Critique of Nonexperimental Estimators? *Journal of Econometrics*, 125, 305-353.
- Snow, J. (1855). *On The Mode of Communication of Cholera*. London: John Churchill, 2nd edition, 1855. <http://archive.org/details/b28985266>. Erişim Tarihi: 02.07.2020
- Snow, J. (1856). Cholera and the Water Supply in the South District of London in 1854. *Journal of Public Health and Sanitary Review*, 2,239-257.
- <http://www.ph.ucla.edu/epi/snow/cholerawatersouthlondon.html> Erişim Tarihi: 02.07.2020
- Thistlethwaite, D.L. and Campbell, D.T. (1960). Regression-Discontinuity Analysis: An Alternative to the Ex-Post Facto Experiment. *Journal of Educational Psychology*, 51(6), 309-317.
- Wooldridge, J. M. (2013). *Introductory Econometrics: A Modern Approach (5th edition)*. South-Western/Cengage Learning.

# BÖLÜM 18

## YAPISAL EŞİTLİK MODELLERİ VE STATA UYGULAMASI<sup>1</sup>

Hatice Hicret ÖZKOÇ<sup>2</sup>

---

İngilizce adıyla Structural Equation Models (SEM) olarak da anılan Yapısal Eşitlik Modelleri (YEM), ölçülebilen değişkenler ile ölçülemeyen değişkenlerin nedensel ve ilişkisel (korelasyona bağlı) olarak tanımlanması üzerine dayalı istatistiksel bir yaklaşımdır. YEM'in temelinde, kuramsal olarak oluşturulan yapısal modellerin hipotez testleri yoluyla çözümlenmesi yatmaktadır. Bu yapısal modeller değişkenler arasındaki nedensellik bağlarına dayalıdır. Nedensellik bağları ise regresyon denklemleri şeklindeki eşitlikler ile tanımlanır ve şematik gösterimlerle daha anlaşılır ve kavramsal hâle getirilebilmektedir. Bu nedenle YEM ile yapılan çalışmalarda modellerin ve bulguların gösteriminde bilhassa şemalar tercih edilmektedir.

YEM, çok değişkenli istatistiksel bir yöntem olmasının yanı sıra kullandığı farklı tahmin yöntemleri ve uyum iyiliği ölçütleri ile de zengin bir yöntemsel içeriğe sahiptir. Son yıllarda özellikle sosyal bilimler alanında gözlenen amprik çalışmaların önemli bir kısmında YEM ve ona bağlı analizlerin yapıldığı görülmektedir. Özellikle davranışsal çalışmalarda sıklıkla başvurulanan bu yöntem her ne kadar hazır paket programları ile uygulama açısından kolaylaştırılmış olsa da yöntemin ayrıntılı olarak bilinmemesi hâlinde bu uygulamalar tamamen yanlış sonuçlar verebilmektedir. Bu nedenle, bölümde öncelikle YEM teorisi ele alınacak, ayrıntılara ve dikkat edilmesi gereken hususlara değinilecek ardından son kısımda örnek veri setleri ve modeller üzerinden Stata uygulaması gerçekleştirilecektir.

### 18.1. YAPISAL EŞİTLİK MODELLERİNDE YER ALAN TEMEL KAVRAMLAR

Sosyal bilimler yapısı gereği çoğu zaman doğrudan gözlenemeyen yapılarla ilgilenmektedir. Gözlenemeyen ve açıklanmaya çalışılan bu yapılara YEM'de **gizil (gizli, latent)**<sup>3</sup> **değişken** adı verilmek-

---

<sup>1</sup> Bu bölümün teorik kısımları yazar tarafından hazırlanmış olan “Yapısal Eşitlik Modeli: Sağlık Sektörüne Uygulanması” başlıklı doktora tezinden yararlanılarak oluşturulmuştur.

<sup>2</sup> Doç.Dr. Muğla Sıtkı Koçman Üniversitesi, İktisadi ve İdari Bilimler Fakültesi, İşletme Bölümü. E-mail: hatice.ozkoc@imu.edu.tr

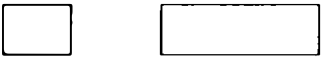
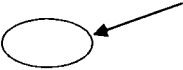
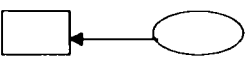
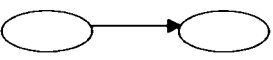
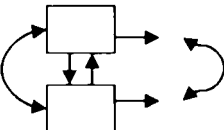
<sup>3</sup> İngilizce literatürde yer alan “latent” kavramı yerli yazında “gizil” ve “gizli” değişken olarak karşılık bulmasına rağmen “gizil” değişken kullanımının daha fazla kabul gördüğü bilinmektedir. Bu nedenle kitabın bu bölümünde “gizil” kavramının kullanımı benimsenmiş ve böylece yerli yazından da faydalanacak olan araştırmacılar için bir dil birliğinin sağlanması düşünülmüştür. YEM literatürü incelendiğinde terimlerin İngilizce karşılıklarının kullanımının da kabul gördüğü görülmektedir. Bu yüzden bölüm içerisinde bazı kavramların İngilizce karşılıkları parantez içerisinde verilecektir.

tedir. Ekonomideki kapitalizm ya da yolsuzluk, pazarlama alanında algılanan hizmet kalitesi ya da memnuniyet, psikolojide bağıllık ya da endişe söz konusu gizil yapılara farklı disiplinlerden verilebilecek örneklerdendir. Pratik hayatta doğrudan gözlenemeyen bu yapıların doğrudan **gözlenebilen değişkenlerle** kavramlaştırılması ve bu değişkenler aracılığıyla tanımlanıp ölçülmesi gerekmektedir. Özellikle psikoloji ve işletme bilimlerinde kullanılan ölçeklerde yer alan her bir madde gözlenen değişkenlere karşılık gelirken, bunlar aracılığıyla ölçüldüğü düşünülen yapılar (farklı kaynaklarda boyut ya da faktör olarak isimlendirilebilmektedir) ise gizil değişkenleri ifade etmektedir.

YEM'in şematik gösterimleri gerek farklı paket programlarda (Lisrel, Amos, Jasp ve Stata gibi) modelin tahminleme süreçlerinde gerekse sonuçların raporlaştırılmasında sıklıkla kullanılmaktadır. Bu modellerin çiziminde evrensel olarak kabul görmüş semboller Tablo 18.1'de verilmektedir. Bu semboller yardımıyla modellerin çizimi ve çözümü, modellerin denklemsel yazımına kıyasla daha kolay olabilmektedir.

Yapısal Eşitlik Modellerinde oluşturulan modeller temel olarak dört sınıfta toplanmaktadır: Path (yol) analizi modelleri, Doğrulayıcı faktör analizi modelleri, Yapısal regresyon modelleri ve Gizil değişken değişim modelleri (Raykov ve Marcoulides, 2006: 3-4). Bir başka yaklaşım ise YEM modellerini iki kısımda incelemektedir: Ölçüm Modeli ve Yapısal Model. Ölçüm Modeli ölçüme dair özellikleri değerlendirirken yapısal model ise gizil değişkenler arasındaki nedensel etkileri tanımlamak için kullanılmaktadır (Koufteros, 1999: 476).

**Tablo 18.1.** Yapısal Eşitlik Modellerinde Kullanılan Başlıca Semboller

| Semboller                                                                           | Açıklama                                                              |
|-------------------------------------------------------------------------------------|-----------------------------------------------------------------------|
|  | Gözlenen değişkenler ( $x, y$ )                                       |
|  | Gizil değişkenler (boyut, faktör, yapı) ( $\xi, \eta$ )               |
|  | Gizil değişkendeki hata                                               |
|  | Gözlenen değişkendeki hata                                            |
|  | Gözlenen bir değişkenin gizil değişken üzerindeki regresyon katsayısı |
|  | Gizil değişkenler arasındaki nedensel ilişki                          |
|  | Çift yönlü oklar; değişkenler arasındaki korelasyon ya da kovaryans   |



### 18.1.1. Ölçme Modeli

YEM içerisinde yer alan iki modelden ilki olan ölçme modeli, gizil ve gözlenen değişkenler arasındaki bağlantıyı temsil eden yapısal denklemler olup bu denklemler stokastik karakterdedir. Bağımlı ve bağımsız değişken için ayrı ayrı oluşturulan ölçme modellerinde bağımsız ve bağımlı değişkenlere ilişkin ölçüm modelinin açılımı

$$x = \Lambda_x \xi + \delta \quad (18.1)$$

$$y = \Lambda_y \eta + \varepsilon \quad (18.2)$$

olarak yazılmaktadır (Long, 1983: 21). Her iki ölçüm modeline ilişkin varsayımlar, kullanılan notasyon ve tanımlar Tablo 18.2'de verilmektedir (Hair vd., 1998: 646-647).

**Tablo 18.2. Ölçüm Modelleri**

| Ölçüm modeli için yapısal eşitlikler                                                  | Notasyon             | Tanım                                          |
|---------------------------------------------------------------------------------------|----------------------|------------------------------------------------|
| $x = \Lambda_x \xi + \delta$                                                          | $y$                  | $\eta$ 'nin gözlenen değişkenleri              |
| $y = \Lambda_y \eta + \varepsilon$                                                    | $x$                  | $\xi$ 'nin gözlenen değişkenleri               |
| <b>Varsayımları</b>                                                                   | $\varepsilon$        | y için ölçme hataları                          |
|                                                                                       | $\delta$             | x için ölçme hataları                          |
|                                                                                       | $\Lambda_y$          | y'yi $\eta$ ile bağlayan katsayılar            |
|                                                                                       | $\lambda_x$          | x 'i $\xi$ ile bağlayan katsayılar             |
|                                                                                       | $\Theta_\varepsilon$ | $E(\varepsilon\varepsilon')$ kovaryans matrisi |
|                                                                                       | $\Theta_\delta$      | $E(\delta\delta')$ kovaryans matrisi           |
| $E(\eta) = 0$                                                                         |                      |                                                |
| $E(\xi) = 0$                                                                          |                      |                                                |
| $E(\varepsilon) = 0$                                                                  |                      |                                                |
| $E(\delta) = 0$                                                                       |                      |                                                |
| $Cov(\xi, \varepsilon) = 0, Cov(\eta, \varepsilon) = 0, Cov(\delta, \varepsilon) = 0$ |                      |                                                |
| $Cov(\xi, \delta) = 0, Cov(\eta, \delta) = 0$                                         |                      |                                                |

Gizil değişkenlerin gözlenen değişkenler üstündeki regresyon katsayıları olan  $\lambda_i$  'ler ise gizil değişkenin bir birim değişmesi durumunda gözlenen değişkendeki beklenen değişme miktarını ifade etmektedir ve path katsayıları olarak da isimlendirilmektedir.

### 18.1.2. Yapısal Model

Yapısal Eşitlik Modellerinde yer alan bir diğer kısım ise yapısal modeldir. Bir yapısal model, ikişer ikişer tüm gizil değişkenler arasındaki nedensel ilişkileri özetleyen yapısal denklemleri kapsamaktadır. Bu anlamda yapısal modelin parametrik açılımı

$$\eta = B\eta + \Gamma\xi + \varsigma \quad (18.3)$$

biçiminde yazılmaktadır (Bollen, 1989: 13; Tabachnick ve Fidel, 1996: 721). Tablo 18.3'de bir yapısal modelde kullanılan notasyon, varsayım ve tanımlar verilmektedir.

**Tablo 18.3.** Yapısal Modelde Kullanılan Gösterimler

|                                                                                                                                                                                                                 | Notasyon | Tanım                                   |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------|-----------------------------------------|
| <b>Gizil değişken için yapısal model</b><br>$\eta = B\eta + \Gamma \xi + \zeta$<br><b>Varsayımları</b><br>$E(\eta) = 0$<br>$E(\xi) = 0$<br>$E(\zeta) = 0$<br>$(I - B)$ tekil değildir.<br>$Cov(\xi, \zeta) = 0$ | $\eta$   | Gizil içsel değişken                    |
|                                                                                                                                                                                                                 | $\xi$    | Gizil dışsal değişken                   |
|                                                                                                                                                                                                                 | $\zeta$  | Eşitliklerdeki gizil hatalar            |
|                                                                                                                                                                                                                 | $B$      | Gizil içsel değişkenin katsayı matrisi  |
|                                                                                                                                                                                                                 | $\Gamma$ | Gizil dışsal değişkenin katsayı matrisi |
|                                                                                                                                                                                                                 | $\Phi$   | $E(\xi\xi')$ Kovaryans matrisi          |
|                                                                                                                                                                                                                 | $\Psi$   | $E(\zeta\zeta')$ Kovaryans matrisi      |

### 18.1.3. Toplam, Doğrudan ve Dolaylı Etkiler

Doğrudan etkiler YEM modellerinde  $B, \Gamma, \Lambda_x$  ve  $\Lambda_y$  matrislerinden bulunmaktadır. Ancak YEM’de sadece doğrudan etkilerin incelenmesi sapmalı sonuçlar verebilmektedir. Bu nedenle toplam etkilerin incelenmesi gerekmektedir. Toplam etki, doğrudan ve dolaylı etkilerin toplamı olarak ifade edilmektedir. Gizil değişkenli YEM için etkilerin ayrıştırılması özet bir biçimde Tablo 18.4’te verilmiştir.

**Tablo 18. 4.**  $\xi$  ve  $\eta$ ’nin  $\eta, y$  ve  $x$  Üstündeki Doğrudan, Dolaylı ve Toplam Etkileri

|                    |          | $\eta$ üzerinde                | $y$ üzerinde                         | $x$ üzerinde |
|--------------------|----------|--------------------------------|--------------------------------------|--------------|
|                    | Doğrudan | $\Gamma$                       | 0                                    | $\Lambda_x$  |
| $\xi$ ’nin etkisi  | Dolaylı  | $(I - B)^{-1} \Gamma - \Gamma$ | $\Lambda_y (I - B)^{-1} \Gamma$      | 0            |
|                    | Toplam   | $(I - B)^{-1} \Gamma$          | $\Lambda_y (I - B)^{-1} \Gamma$      | $\Lambda_x$  |
|                    | Doğrudan | $B$                            | $\Lambda_y$                          | 0            |
| $\eta$ ’nin etkisi | Dolaylı  | $(I - B)^{-1} - I - B$         | $\Lambda_y (I - B)^{-1} - \Lambda_y$ | 0            |
|                    | Toplam   | $(I - B)^{-1} - I$             | $\Lambda_y (I - B)^{-1}$             | 0            |

**Kaynak:** Bollen, 1989: 382.

## 18.2. YAPISAL EŞİTLİK MODELLERİNİN TAHMİNİ

Genel bir yapısal eşitlik modelinde temel hipotez

$$\Sigma = \Sigma(\theta) \quad (18.4)$$

şeklinde kurulmaktadır. Burada  $\Sigma$ ,  $y$  ve  $x$ 'in anakütle kovaryans matrisi,  $\Sigma(\theta)$  modelin  $\theta$ 'daki serbest parametrelerinin bir fonksiyonu olarak yazılan kovaryans matrisidir. (18.4) eşitliği kovaryans matrisinin her bir elemanının model parametrelerinin bir ya da daha fazla bir fonksiyonu olduğunu belirtmektedir.  $\Sigma(\theta)$  matrisi üç kısımdan oluşmaktadır: (1)  $y$ 'nin kovaryans matrisi, (2)  $x$  ile  $y$ 'nin kovaryans matrisi, (3)  $x$ 'in kovaryans matrisi (Bollen, 1989: 85; Long, 1987: 24). YEM'de tahmin süreçleri yapısal parametrelerin gözlenen değişkenlerin kovaryans matrisiyle olan ilişkisine dayanılarak yapılmaktadır.

$$\Sigma(\theta) = \begin{bmatrix} (I - B)^{-1} (\Gamma \Phi \Gamma' + \Psi) (I - B)^{-1'} & (I - B)^{-1'} \Gamma \Phi \\ \Phi \Gamma' (I - B)^{-1} & \Phi \end{bmatrix} \quad (18.5)$$

Eğer yapısal eşitlik modeli doğru ise ve popülasyon parametreleri biliniyorsa, bu durumda  $\Sigma = \Sigma(\theta)$  olur. Yapısal parametrelere göre  $\Sigma$  matrisi ise

$$\Sigma(\theta) = \begin{bmatrix} \phi_{11} + \psi_{11} & \phi_{12} \\ \phi_{12} & \phi_{22} \end{bmatrix} \quad (18.6)$$

şeklinde dir. Modelin doğru olduğu ve anakütle parametrelerinin bilindiği varsayıldığında, eşitlik (18.5)'teki her eleman eşitlik (18.6)'nın elemanlarına karşılık gelir. Uygulamalarda çoğunlukla popülasyon kovaryansları, varyansları ve parametreleri bilinmez. Bu nedenle de örnek tahminlerinin kovaryans matrisine dayanılarak bilinmeyen parametrelerin tahmini elde edilmeye çalışılır (Bollen, 1989: 104-105).  $\Sigma(\theta)$ 'deki model parametrelerini tahmin etmek amacıyla örneklem tahmini  $S$  elde edilmeli ve  $F(S, \Sigma(\theta))$  fonksiyonunun skaler hatasını verecek uyum fonksiyonunun seçilmesi gerekmektedir. Uyum fonksiyonların bir takım özellikleri vardır (Kaplan, 2000: 24, Bollen, 1989: 106):  $F(S, \Sigma(\theta))$  skalerdir,  $F(S, \Sigma(\theta)) \geq 0$ ,  $F(S, \Sigma(\theta)) = 0 \Leftrightarrow \Sigma(\theta) = 0$ ,  $F(S, \Sigma(\theta))$ ,  $S$  ve  $\Sigma(\theta)$ 'de süreklidir.

YEM'e ilişkin literatürde en sık kullanılan tahmin yöntemleri şunlardır: En çok olabilirlik yöntemi (EÇO), Ağırlıklandırılmamış en küçük kareler (EKK), Genelleştirilmiş en küçük kareler (GEKK), Ağırlıklı en küçük kareler (AEKK). Olağan En Küçük Karelerin (OEKK) ardışık modeller için kullanılması uygundur. Ardışık olmayan modeller için İki Aşamalı En Küçük Kareler (2AEKK) çok yaygın biçimde kullanılan bir tahmin sürecidir (Kaplan, 2000: 24-25; Bollen, 1989: 106; Çelik ve Yılmaz, 2013: 24-25).

### 18.2.1. En Çok Olabilirlik Tahmin Yöntemi (EÇO)<sup>4</sup>

Yöntem Jöreskog (1973) tarafından YEM için uyarlanmıştır ve bu alanda en yaygın kullanılan uyum fonksiyonu hâline gelmiştir (Kaplan, 2000: 25; Şimşek, 2007: 61). YEM analizlerinin gerçekleştirildiği paket programlarda varsayılan (default) olarak seçili bu yöntemin önemli bir kısıtı, çok değişkenli güçlü bir normallik varsayımına dayanmasıdır. Dağılıma ilişkin varsayımların ihlal edildiği durumlarda ciddi yanlış sonuçlar ortaya çıkmaktadır. Örneğin  $\chi^2$  istatistiği tarafından yitirmekte ve parametrelerin standart hatalarını olumsuz etkilemektedir. Ancak öte yandan normallik varsayımının bozulduğu durumlarda oldukça dirençli (robust) olduğu da görülmektedir (Çelik ve Yılmaz, 2013: 26).

EÇO, YEM katsayılarının elde edilmesi amacıyla En Çok Olabilirlik Uyum Fonksiyonunu ( $F_{EÇO}$ ) en küçüklenmeye çalışılarak  $\hat{\theta}$  tahminlerini bulmaya çalışmaktadır (Bollen, 1989: 107).

$$F_{EB} = \log|\Sigma(\theta)| + tr(S\Sigma^{-1}(\theta)) - \log|S| - (p + q) \quad (18.7)$$

( $F_{EÇO}$ ) uyum fonksiyonu  $y$  ve  $x$ 'in çok değişkenli normal dağıldıkları varsayımından yararlanılarak türetilmektedir.  $F_{EÇO}$ 'nin bulunmasında biri  $y$  ve  $x$ 'in çok değişkenli normal dağılımından, diğeri örnek kovaryans matrisi  $S$ 'nin Wishart dağılımlı olmasından kaynaklanan iki alternatif yöntem kullanılmaktadır.

EÇO tahmincilerinin büyük örnekler için sağlanan bazı asimptotik özellikleri bulunmaktadır. Bunlardan ilki EÇO parametre tahminleri ve standart hataları sapmasızdır. İkinci özellik bu tahmincilerin tutarlı olmalarıdır. Üçüncü olarak tutarlı tahminciler içinde en küçük varyansa sahip olan tahminciler olduklarından etkindirler. Son olarak elde edilen tahminlerin dağılımı örneklem büyüklüğünün artmasıyla normal dağılıma yakınsar yani asimptotik normal dağılımlıdır. Bu yakınsama sonucunda tahmin edilen parametrelerin standart hataları, büyük örneklemelerde  $z$  dağılımına yakınsar (Bollen, 1989: 108) Aynı zamanda EÇO yöntemiyle elde edilen tahminciler ölçek değişmezlik ve ölçek bağımsızlık özelliklerini taşımaktadırlar.

### 18.2.2. Ağırlıksız En Küçük Kareler

Ağırlıksız en küçük kareler yöntemi, modele ilişkin kovaryans matrisinin gözlenen değişkenlerden elde edilen matrinden farkı olan artık değerler matrisinin diyagonal değerlerine ait değerleri minimize eden bir fonksiyondur. Ağırlıksız en küçük kareler için uyum fonksiyonu aşağıdaki gibidir:

$$F_{EKK} = \left(\frac{1}{2}\right) tr \left[ (S - \Sigma(\theta))^2 \right] \quad (18.8)$$

Burada  $S$  gözlenen kovaryans matrisi,  $\Sigma(\theta)$  modele ilişkin tahmini kovaryans matrisi ve  $\theta$  parametrelerin  $(t \times 1)$  boyutlu vektörüdür. Bu tahmin yöntemi, dağılıma ilişkin varsayımlar gerektirmediği ve tahmin sürecinin kolay olması açısından diğer tahmin yöntemlerine göre daha avantajlıdır (Mulaik, 2009). Bu yöntemin dezavantajı ise, veriler çok değişkenli normal

<sup>4</sup> Kitabın önceki bölümlerinde tahmin yöntemleri ayrıntılı olarak ele alındığı için bu kısımda sadece YEM açısından değinilecektir.

dağılmadığında ve çözümlenen matris, kovaryans matrisi olmadığında, model uyumu hakkındaki olasılıksal çıkarsamaların yeterli olmamasıdır. Tahminciler ölçek değişmezlik ve ölçek bağımsızlık özelliklerini taşımamaktadırlar.

### 18.2.3. Genelleştirilmiş En Küçük Kareler Yöntemi

Genelleştirilmiş En Küçük Kareler (GEKK) tahmincisi ilk olarak Aitken (1935) tarafından ortaya çıkarılmış daha sonra Jöreskog ve Goldberger (1972) tarafından YEM için geliştirilmiştir. GEKK, varyans ve kovaryanslara göre artıklar matrisinin elemanlarını ağırlıklandırır. GEKK uyum fonksiyonunun genel formu,

$$F_{GEKK} = \left( \frac{1}{2} \right) tr \{ (S - \Sigma(\theta)) W^{-1} \}^2 \quad (18.9)$$

biçimindedir. Burada  $tr$  matrisin izi,  $S$  gözlenen kovaryans matrisi,  $\Sigma(\theta)$  modele ilişkin tahmini kovaryans matrisi,  $\theta$  parametre vektörü ve  $W^{-1}$  artıkların  $p \times p$  boyutlu ağırlık matrisidir. GEKK ağırlık matrisi olarak örneklem kovaryans matrisini kullanır ve  $W^{-1} = S^{-1}$  alınırsa:

$$\begin{aligned} F_{GEKK} &= \left( \frac{1}{2} \right) tr \{ (S - \Sigma(\theta)) S^{-1} \}^2 \\ &= \left( \frac{1}{2} \right) tr \{ (I - \Sigma(\theta) S^{-1}) \}^2 \end{aligned} \quad (18.10)$$

elde edilir. GEKK uyum fonksiyonu genel doğrusal olmayan regresyon modeli ile kovaryans yapısı modelinin benzerliği temel alınarak geliştirilmiştir (Browne, 1982: 79; Bollen, 1989: 113; Kaplan, 2000: 30).

$F_{GEKK}$  'de  $F_{E\check{C}O}$  gibi ölçek değişmezlik ve ölçek bağımsızlık gibi istenen özelliklerine sahiptir. Buna karşın  $F_{GEKK}$  'in bazı sınırlılıkları da bulunmaktadır. Bunlardan en önemlisi gözlenen değişkenlerin dağılımının çok kalın veya çok ince kuyruklu olması durumunda anlamlılık testlerinin yapılmasında asimptotik standart hatalara ve  $\chi^2$  testine güvenilmemesidir (Browne, 1974: 12; Bollen, 1989: 108; Çelik ve Yılmaz, 2013: 28; Yuan ve Bentler, 1998: 365).

### 18.2.4. Ağırlıklı En Küçük Kareler Yöntemi

Varyansların homojen olmaması veya hatalarda otokorelasyon olması hâlinde sonuçlar anakütlede gözlenen durumdan sapma göstermekte ve model yanlış değerlendirilmektedir. Bu durumda Ağırlıklı En Küçük Kareler Yöntemi (AEKK) sonuçlarını kullanmak yerinde olacaktır. Yöntemin bir başka üstünlüğü ise değişkenlerin bir kısmı sıralı ve diğerleri sürekliyse, sürekli değişkenlerin dağılımı normal dağılımdan sapmalar gösteriyorsa ve model iki düzeyli değişkenleri içeriyorsa kullanılabilir. EÇÖ yönteminden farklı olarak ise bu yöntemde verilerin analiz edilmesinde ham veriye ihtiyaç duyulmaktadır (Çelik ve Yılmaz, 2013: 29). AEKK yöntemi için uyum fonksiyonu aşağıdaki gibidir:

$$F_{GLS} = \left( \frac{1}{2} \right) \text{tr} \left( \left\{ \left[ S - \Sigma(\theta)W^{-1} \right] \right\}^2 \right) \quad (18.11)$$

$W$  matrisi örneklem varyans ve kovaryanslarının (veya korelasyonlarının) analiz edildiği, asimptotik kovaryans matrisinin tutarlı bir tahminidir. AEKK metodu çeşitli avantajlara sahip olmakla birlikte bazı dezavantajlara da sahiptir. Gözlenen değişkenlerin artmasıyla hızla büyüyen ağırlık matrisi, AEKK için bir kısıtlamadır. Eğer gözlenen değişkenlerin dağılımı normal dağılımdan önemli bir sapma göstermez ise EÇO yönteminin kullanılması tercih edilmektedir. Çok değişkenli normalliğin sağlanmaması durumunda ise diyagonal olarak ağırlıklandırılmış en küçük kareler ve eliptik dağılımlar gibi alternatif yöntemler geliştirilmiştir. Aynı zamanda değişken sayısı çok olduğu zaman tahminlenen katsayıların asimptotik kovaryans matrislerinin hesaplanması çok zaman alır ve bilgisayarda hesaplama yapılıyorsa çok fazla yer kaplar. Bu durumda alternatif bir yaklaşım sadece asimptotik varyansların hesaplandığı diyagonal olarak Ağırlıklandırılmış En Küçük Karelerdir.

### 18.2.5. Standartlaştırılmış ve Standartlaştırılmamış Katsayılar

Gözlenen değişkenlerin kovaryans matrisinin analizi sonucunda değişkenlerin ölçeklendirildiği birimlere bağlı olan standartlaştırılmamış katsayılar elde edilmektedir. Ancak aynı bağımlı değişken üzerinde etkili olan bir veya daha fazla değişken farklı ölçüm birimleri ile ölçülmüş ise bu durumda etkileri standartlaştırılmamış katsayıları kullanarak açıklamak oldukça zor olmaktadır. Bu gibi durumlarda standartlaştırılmış katsayılar farklı açıklayıcı değişkenlerin göreceli etkilerini karşılaştırmada kullanılabilir. Bu katsayılar

$$\begin{aligned} \hat{\lambda}_{ij}^s &= \hat{\lambda}_{ij} \left( \frac{\hat{\sigma}_{ij}}{\hat{\sigma}_{ii}} \right)^{1/2} \\ \hat{\beta}_{ij}^s &= \hat{\beta}_{ij} \left( \frac{\hat{\sigma}_{ij}}{\hat{\sigma}_{ii}} \right)^{1/2} \\ \hat{\gamma}_{ij}^s &= \hat{\gamma}_{ij} \left( \frac{\hat{\sigma}_{ij}}{\hat{\sigma}_{ii}} \right)^{1/2} \end{aligned} \quad (18.12)$$

şeklinde hesaplanmaktadır.  $s$ , katsayının standartlaştırılmış değerini ifade etmektedir.  $i$  bağımlı değişkeni,  $j$  ise bağımsız değişkeni göstermektedir.  $\hat{\sigma}_{ii}$  ve  $\hat{\sigma}_{ij}$  ise sırasıyla  $i$ 'inci ve  $j$ 'inci değişkenin model tarafından tahmin edilen varyanslarını göstermektedir. Standartlaştırılmış katsayı, diğer değişkenler sabit tutulduğunda, bağımsız değişkendeki bir standart sapmalı değişimin bağımlı değişkenin standart sapmasında meydana getirdiği etkiyi göstermektedir. Özellikle veri toplama tekniği olarak anket yöntemi tercih edildiğinde ve anket içerisinde farklı ölçeklerin kullanılması durumunda bu katsayıların yorumlanması yerinde olacaktır.

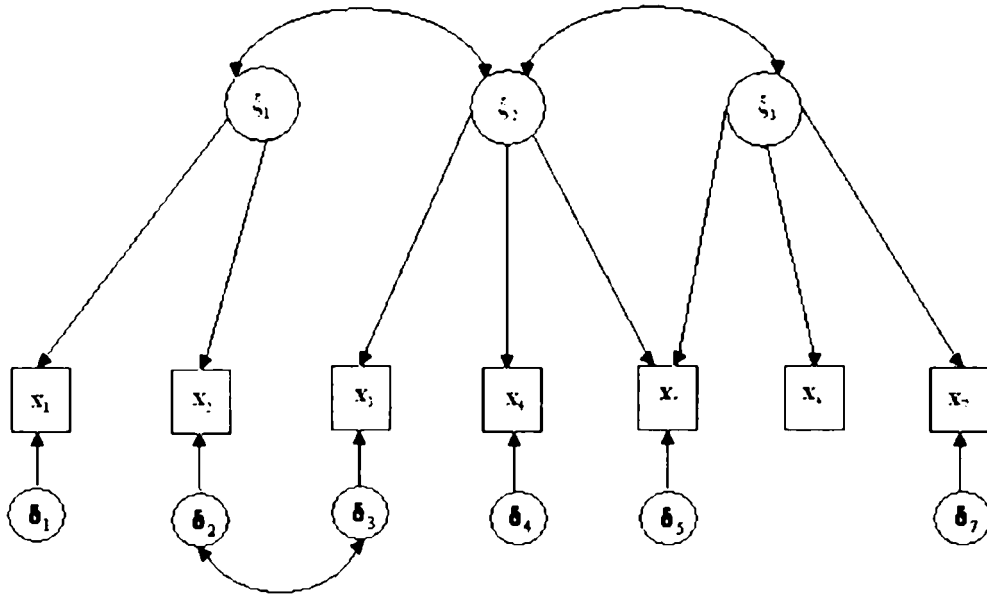
### 18.3. DOĞRULAYICI FAKTÖR ANALİZİ (DFA)

Açıklayıcı Faktör Analizi (AFA), gözlenen değişkenler arasındaki kovaryans veya korelasyonları göreceli olarak daha az sayıdaki gizil değişkenler aracılığıyla açıklamayı amaçlayan bir boyut indirgeme yöntemidir ve boyutları keşfedici bir yanı vardır. AFA'nde ilgili yapı ya da faktörlerin

sayısı, gözlenen ve gözlenemeyen değişkenler arasındaki ilişki hakkında önsel herhangi bir beklenti veya varsayım gerekmemektedir. AFA araştırmacının beklentilerine gerek duymaz ve analizler bu beklentilerle şekillenmez (Şimşek, 2007: 123; Bollen, 2002: 605). Doğrulayıcı Faktör Analizi (DFA) ise 1970'lerde ilk olarak Jöreskog tarafından geliştirilmiştir. DFA net bir şekilde ve doğrudan varsayımsal bir faktör modelinin veri ile olan uyumunu test etmektedir. Farklı bir ifade ile tanım koşullarını gerçekleyen önsel bir yapının var olan veri ile uyumlu olup olmadığını belirlemeye çalışır. DFA, AFA'ndeki sınırlamaları ortadan kaldırmakta ve araştırmacının bir takım kısıtları modele dahil etmesine olanak sağlamaktadır (Şimşek, 2007: 125). Bu kısıtlar şöyledir: (1) hangi ortak faktör çiftinin (/çiftlerinin) korelasyonlu olduğu, (2) hangi gözlenen değişkenlerin hangi ortak faktörlerden etkilendiği, (3) hangi gözlenen değişkenlerin bir özgün faktörden etkilendiği, (4) hangi özgün faktör çiftinin korelasyonlu olduğu.

Şekil 18.1 ile bir örneği verilen DFA'da örnek verisinin koyulan kısıtlarla tutarlı olup olmadığının veya başka bir ifade ile veri setinin üretilen modeli sağlayıp sağlamadığının belirlenmesi için istatistiksel testler uygulanmaktadır. Bu bakımdan analiz "doğrulayıcı" olarak adlandırılmaktadır.

**Şekil 18.1.** Doğrulayıcı Faktör Modeli Örneği



**Kaynak:** Şimşek, 2007: 126.

DFA modeli Eşitlik (18.13)'teki gibi gösterilebilir; gözlenen değişkenler ( $y$ ), gizil faktörler ( $\eta$ ) ve tek varyans ( $\varepsilon$ ) olmak üzere

$$y = \Lambda_y \eta + \varepsilon \quad (18.13)$$

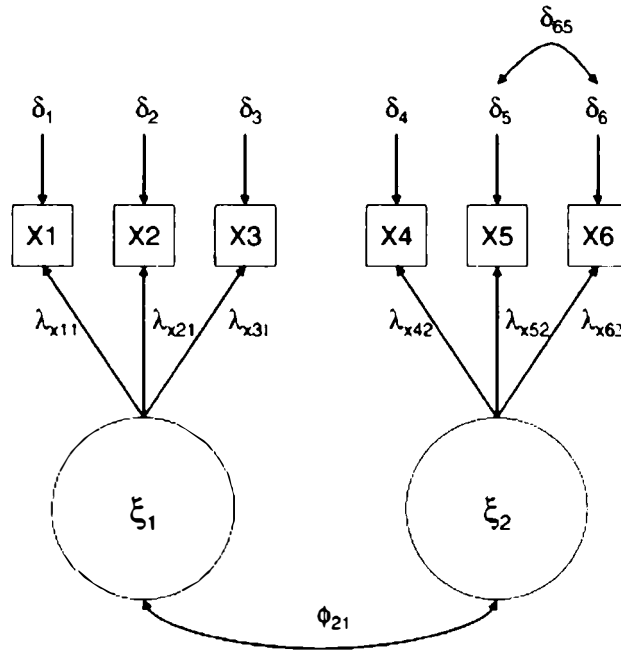
$y$ 'nin kovaryans matrisi;

$$\Sigma = \Lambda_y \Psi \Lambda_y' + \Theta_{\varepsilon} \quad (18.14)$$

dir.  $\Sigma$ ,  $p$  tane gösterge değişkenin  $p \times p$  boyutlu simetrik kovaryans matrisidir.  $\Lambda_x, \Lambda$  faktör yüklerinin  $p \times m$  boyutlu matrisi iken  $\Psi$ , faktör korelasyonlarının  $(1 \times 1)m \times m$  boyutlu simetrik matrisidir.  $\Theta_\epsilon, \epsilon$  varyanslarının  $p \times p$  boyutlu diyagonal matrisidir (Jöreskog ve Sörbom, 1996: 123-124; Bollen, 1989: 233-234; Çelik ve Yılmaz, 2013: 45).

DFA’da gizil değişkenler içsel ya da dışsal olabilir. DFA modellerinde dışsal değişkenler  $x$  ile gösterilebilir ve bunlar bağımsız ve açıklayıcı değişkenlere karşılık gelmektedir. Benzer biçimde, içsel değişkenler ise  $y$ ’ye eşittir ve bağımlı değişken olarak düşünülebilir (Çelik ve Yılmaz, 2013: 45). İki faktörlü DFA modeli için dışsal ve içsel gizil değişken gösterimi Şekil 18.2 ve Şekil 18.3’te verilmektedir.

**Şekil 18.2.** İki Faktörlü DFA Modeli İçin Dışsal Gizil Değişken Gösterimi

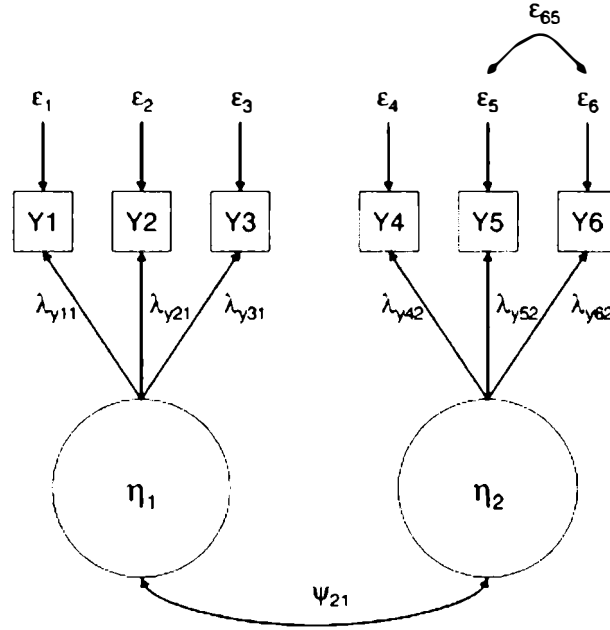


| Adı         | Parametre   | Matris          | Türü                | Tanımlama                           |
|-------------|-------------|-----------------|---------------------|-------------------------------------|
| Lambda-x    | $\lambda_x$ | $\Lambda_x$     | Regresyon           | Faktör yükleri                      |
| Theta delta | $\delta$    | $\Theta_\delta$ | Varyans - kovaryans | Hata varyansları ve kovaryansları   |
| Phi         | $\phi$      | $\Phi$          | Varyans - kovaryans | Faktör varyansları ve kovaryansları |
| Tau- x      | $\tau_x$    |                 | Ortalama vektör     | Gösterge kesişimler                 |
| Kappa       | $\kappa$    |                 | Ortalama vektör     | Gizil ortalamalar                   |
| Xi (Ksi)    | $\xi$       |                 | Vektör              | Dışsal değişkenlerin isimleri       |

**Kaynak:** Brown, 2006: 55



Şekil 18.3. İki Faktörlü DFA Modeli İçin İçsel Gizil Değişken Gösterimi



| Adı         | Parametre     | Matris               | Türü                | Tanımlama                           |
|-------------|---------------|----------------------|---------------------|-------------------------------------|
| Lambda-y    | $\lambda_y$   | $\Lambda_y$          | Regresyon           | Faktör yükleri                      |
| Theta delta | $\varepsilon$ | $\Theta_\varepsilon$ | Varyans - kovaryans | Hata varyansları ve kovaryansları   |
| Phi         | $\psi$        | $\Psi$               | Varyans - kovaryans | Faktör varyansları ve kovaryansları |
| Tau- y      | $\tau_y$      |                      | Ortalama vektör     | Gösterge kesişimler                 |
| Alfa        | $\alpha$      |                      | Ortalama vektör     | Gizil ortalamalar                   |
| Eta         | $\eta$        |                      | Vektör              | İçsel değişkenlerin isimleri        |

**Kaynak:** Brown, 2006: 56.

DFA modellerinde gösterge değişkenleri, faktörlerden gelen tek yönlü oklar gözlenen ölçümler üzerinde gizil boyutların doğrudan etkilerini göstermektedir.  $\Theta$ , gösterge hata varyansları kovaryanslarının matrislerini gösterirken çift yönlü oklar ise kovaryansları (korelasyonları) ifade etmek için kullanılmaktadır. Şekil 18.2. ve Şekil 18.3'te yer alan eğri oklar, faktörler arasındaki kovaryansları ve gösterge değişkenler arasındaki hata kovaryanslarını göstermektedir (Brown, 2006: 57).

#### 18.4. MODEL KURMA AŞAMALARI VE UYUM ÖLÇÜTLERİ

Bir yapısal eşitlik modelinin kurulma aşamasının ilk ve en önemli adımı teorik yapıdır. Teorinin desteklediği şekilde oluşturulan modelin şemasının kurulması ve bu şema üzerinde araştırılacak olan ilişkilere dair hipotezlerin oklar yardımıyla belirlenmesi ise ikinci aşamayı oluşturmaktadır. Oluşturulan bu modelin ve ona bağlı hipotezlerin test edilmesi için uygun örneklemin belirlen-

mesi ise takip edilmesi gereken bir sonraki aşamadır. Uygun örneklemden toplanan veriler kullanılarak model parametreleri tahmin edilir ve test edilir. Ortaya çıkan model uyum ölçütleri değerlendirilerek model bir bütün olarak reddedilir veya reddedilemez ya da model üzerinde modifikasyon(lar) yapılarak model iyileştirilmeye çalışılır. Uygun modele ulaşılması hâlinde ise hipotezlerin neticeleri belirtilerek bulgular yorumlanır.

YEM’de parametreler tahmin edildikten sonra modelden elde edilen kovaryans matrisinin örnekleme ait kovaryans matrisi ile olan uyumunun değerlendirilmesi için diğer çok değişkenli yöntemlerden farklı olarak genel kabul görmüş tek bir hipotez testi ya da bir ölçüt yoktur. Bunun yerine YEM’de birbirinin yerine geçemeyen çok sayıda test ve ölçüt geliştirilmiştir (Jöreskog ve Sörbom, 1993: 115; Şimşek, 2007: 161). Tablo 18.5’te literatürde sıklıkla kullanılan uyum ölçütleri verilmiştir.

**Tablo 18.5. Uyum Ölçütleri ve Sınırları**

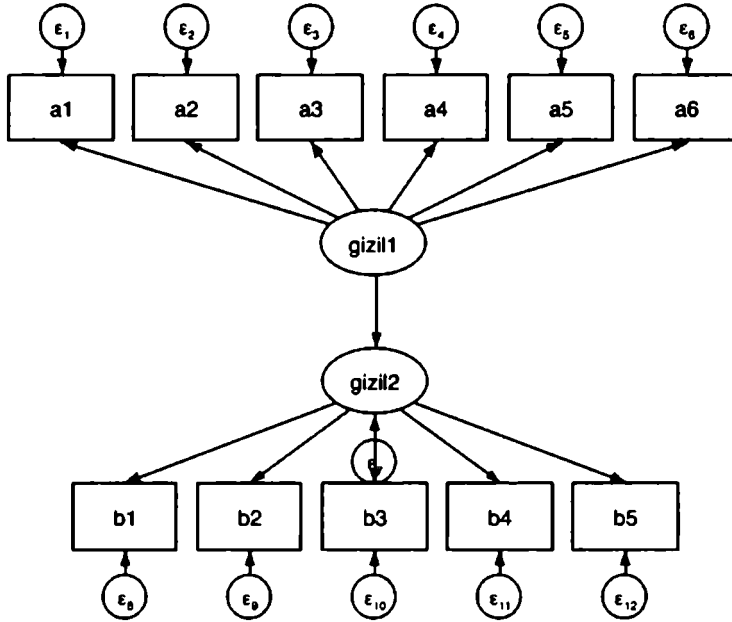
| Uyum Ölçütleri | İyi Uyum                   | Kabul Edilebilir Sınır      |
|----------------|----------------------------|-----------------------------|
| $\chi^2$       | $0 \leq \chi^2 \leq 2sd$   | $2sd \leq \chi^2 \leq 3sd$  |
| p değeri       | $0.05 \leq p \leq 1.00$    | $0.01 \leq p \leq 0.05$     |
| $\chi^2/sd$    | $0 \leq \chi^2/sd \leq 2$  | $2 \leq \chi^2/sd \leq 3$   |
| RMSEA          | $0 \leq RMSEA \leq 0.05$   | $0.05 \leq RMSEA \leq 0.08$ |
| SRMR           | $0 \leq SRMR \leq 0.05$    | $0.05 \leq SRMR \leq 0.10$  |
| NFI            | $0.95 \leq NFI \leq 1.00$  | $0.90 \leq NFI \leq 0.95$   |
| NNFI           | $0.97 \leq NNFI \leq 1.00$ | $0.95 \leq NNFI \leq 0.97$  |
| CFI            | $0.97 \leq CFI \leq 1.00$  | $0.95 \leq CFI \leq 0.97$   |
| GFI            | $0.95 \leq GFI \leq 1.00$  | $0.90 \leq GFI \leq 0.95$   |

### 18.5. UYGULAMALI ÖRNEK-1

Bu bölümde ilk olarak ölçüm modeline ilişkin Stata uygulama örneklerine yer verilecektir. Uygulamalar hem Stata komutları üzerinden ele alınacak hem de önceki sayfalarda bahsedildiği üzere YEM çalışmalarında büyük bir yer tutan şematik gösterimler üzerinden anlatılacaktır. Bu amaçla kullanılacak ilk veri seti 265 gözlemlili bir anket verisi olan yem\_veri.dta<sup>5</sup> verisidir. İki gizil değişkenin yer aldığı yapısal modelde toplam 11 gösterge değişken bulunmaktadır. Bunların altısı ilk gizil değişkene dair gösterge değişkenler iken kalan beşi ise ikinci gizil değişkenin göstergeleridir. Şekil 18.4’te yer alan modelden de görüldüğü üzere gizil değişkenler arasında nedensel bir ilişki tanımlanmıştır.

<sup>5</sup> Veriye erişmek için iç kapak sayfasında verilen karekod ya da linki kullanınız. Veri setinden de görüleceği üzere, veri girişinde her bir sütunda bir gösterge değişken olacak şekilde bir yapı benimsenir.

Şekil 18.4. Örnek 1 için Oluşturulan Teorik Model



Oluşturulan model iki aşamalı olarak çözülecektir. İlk olarak ölçüm modeli test edilip onaylandıktan sonra yapısal modelin testine geçilecektir. Ölçüm modelinin Stata yardımıyla analiz edilmesi için kullanılabilir ve aynı sonuçları veren alternatif komut yapıları Tablo 18.6’da verilmektedir<sup>6</sup>. Komut satırlarında yer alan “**covstruct**” ifadesi kovaryanslara kısıtlama koymak için, “**latent**” ifadesi ise gizil değişkenlerin isimlerinin belirlenmesi amacıyla kullanılmaktadır. “**cov**” ifadesi ise gizil değişkenler arasında belirlenen kovaryans yapısını tanımlamaktadır.

Tablo 18.6. Örnek 1 için Ölçüm Modeli

*Alternatif 1:*

```
sem (gizil1 -> a1,) (gizil1 -> a2,) (gizil1 -> a3,) (gizil1 -> a4,) (gizil1 ->
a5,) (gizil1 -> a6,) (gizil2 -> b1,) (gizil2 -> b2,) (gizil2 -> b3,) (gizil2 ->
b4,) (gizil2 -> b5,), covstruct(_lexogenous, diagonal) latent(gizil1 gizil2) cov(
gizil2*gizil1) nocapslatent
```

*Alternatif 2:*

```
sem (gizil1 -> a1 a2 a3 a4 a5 a6) (gizil2 -> b1 b2 b3 b4 b5),
covstruct(_lexogenous, diagonal) latent(gizil1 gizil2) cov(gizil2*gizil1)
nocapslatent
```

<sup>6</sup> Veri setinin çok değişkenli normalliği sağladığı varsayılmış ve bu nedenle tahmin yöntemi ile ilgili bir seçim yapılmamıştır. Stata varsayılan tahmin yöntemi olarak EÇO seçmiş olup uygulamalar bu tahmin yöntemi üzerinden yapılmıştır. Stata’da EÇO dışında kayıp verilerle EÇO ve normallik varsayımının sağlanmaması durumunda kullanılabilir Asymptotic Distribution Free tahmin seçenekleri yer almaktadır. Daha geniş tahmin yöntemleri için kullanılabilir bir başka program ise Lisrel programıdır. Lisrel aynı zamanda Stata’dan farklı olarak, veri setinin çok değişkenli normalliğini sinayabilmektedir.

*Alternatif 3:*

```
sem (a1 a2 a3 a4 a5 a6 <-gizil1) (b1 b2 b3 b4 b5 <- gizil2), covstruct(_lexogenous,
diagonal) latent(gizil1 gizil2) cov(gizil2*gizil1) nocapslatent
```

Endogenous variables

Measurement: a1 a2 a3 a4 a5 a6 b1 b2 b3 b4 b5

Exogenous variables

Latent:gizil1 gizil2

Structural equation model Number of obs=265

Estimation method = ml

Log likelihood = -3822.257

( 1) [a1]gizil1 = 1

( 2) [b1]gizil2 = 1

|             |        | Coef.     | OIM<br>Std. Err. | z     | P> z  |
|-------------|--------|-----------|------------------|-------|-------|
| -----       |        |           |                  |       |       |
| Measurement |        |           |                  |       |       |
| a1 <-       | gizil1 | 1         | (constrained)    |       |       |
|             | _cons  | 4.437736  | .0605214         | 73.33 | 0.000 |
| -----       |        |           |                  |       |       |
| a2 <-       | gizil1 | -.2443299 | .1052202         | -2.32 | 0.020 |
|             | _cons  | 2.630189  | .0807369         | 32.58 | 0.000 |
| -----       |        |           |                  |       |       |
| a3 <-       | gizil1 | 1.022377  | .0792249         | 12.90 | 0.000 |
|             | _cons  | 4.067925  | .0694562         | 58.57 | 0.000 |
| -----       |        |           |                  |       |       |
| a4 <-       | gizil1 | 1.068601  | .075272          | 14.20 | 0.000 |
|             | _cons  | 4.211321  | .0670983         | 62.76 | 0.000 |
| -----       |        |           |                  |       |       |
| a5 <-       | gizil1 | 1.063405  | .0651621         | 16.32 | 0.000 |
|             | _cons  | 4.30566   | .0606411         | 71.00 | 0.000 |
| -----       |        |           |                  |       |       |
| a6 <-       | gizil1 | .453515   | .1029389         | 4.41  | 0.000 |
|             | _cons  | 3.588679  | .0800646         | 44.82 | 0.000 |
| -----       |        |           |                  |       |       |
| b1 <-       | gizil2 | 1         | (constrained)    |       |       |
|             | _cons  | 3.867925  | .0739832         | 52.28 | 0.000 |
| -----       |        |           |                  |       |       |
| b2 <-       | gizil2 | 1.011356  | .048738          | 20.75 | 0.000 |
|             | _cons  | 3.830189  | .0759755         | 50.41 | 0.000 |
| -----       |        |           |                  |       |       |
| b3 <-       | gizil2 | .852132   | .0537329         | 15.86 | 0.000 |
|             | _cons  | 3.879245  | .0720088         | 53.87 | 0.000 |
| -----       |        |           |                  |       |       |

```

b4 <- |
 gizil2 | .8587035 .0602426 14.25 0.000
 _cons | 3.69434 .0766179 48.22 0.000
-----+-----
b5 <- |
 gizil2 | .7484092 .0567455 13.19 0.000
 _cons | 3.818868 .0699189 54.62 0.000
-----+-----
 var(e.a1)| .3163266 .0364155
 var(e.a2)| 1.688329 .1470782
 var(e.a3)| .5944693 .0604772
 var(e.a4)| .4459004 .0483995
 var(e.a5)| .2345664 .0326727
 var(e.a6)| 1.564161 .1373111
 var(e.b1)| .2907171 .0402254
 var(e.b2)| .3434013 .0437038
 var(e.b3)| .5319597 .0545068
 var(e.b4)| .7004515 .070512
 var(e.b5)| .6458906 .063422
 var(gizil1)| .6543249 .0832249
 var(gizil2)| 1.159763 .1273609
-----+-----
cov(gizil1,gizil2)| .5550335 .0736713 7.53 0.000
-----+-----
LR test of model vs. saturated: chi2(43) = 161.34, Prob > chi2 = 0.0000

```

Model çıktısı incelendiğinde ilk olarak gizil değişkenlerin hangi isimlerle belirtildiğine yer verildiği görülmektedir. Tahmin yöntemi olarak EÇO yönteminin seçildiği modelin çıktısında “OIM” ifadesi göze çarpmaktadır. Bu ifade “observed information matrix” ifadesinin kısaltılması olup kullanılan tekniği göstermektedir. EÇO yöntemi için diğer kullanılabilir teknikler “EIM (expected information matrix”, “OPG (outer product of the gradients”, “robust”, “bootstrap” ve “jackknife” teknikleridir. Kullanılan tahmin yöntemi ve tekniğinin verilmesinin ardından model çıktısında ölçüm modeline ilişkin katsayılara ait sonuçlara yer verilmektedir. Her bir gösterge, değişkenin gizil değişkenlerle olan ilişkilerine ait katsayılara bakıldığından, tümünün istatistiksel olarak anlamlı olduğu görülmektedir. Bu durumda ölçme modelinde “a” olarak kodlanmış olan gösterge değişkenlerin “gizil1” isimli gizil değişken ile anlamlı bir ilişkide olduğu söylenebilir. Benzer durum “b” ile kodlanmış gösterge değişkenlerinde de geçerli olup bu değişkenlerin “gizil2” gizil değişkeni ile ilişkilerinin de anlamlı olduğu görülmektedir. Ölçüm model sonuçlarında katsayıların ardından her bir gösterge değişkeninin hata varyansları ve son olarak gizil değişkenler arasındaki kovaryans değeri yer almaktadır. İki gizil değişken arasında çıkan kovaryans 0.555 olarak bulunmuştur. Önceki kısımlarda da belirtildiği üzere yapısal eşitlik modelleri kovaryans matrisinin analizini içermektedir. Ayrıca elde edilen katsayının sıfırdan farklı olması iki gizil değişkenin bağımsız olmadığını da göstermektedir.

Model çıktısının en altında yer alan ve “Model uyumludur” şeklinde ifade edilen sıfır hipotezini test etmek amacıyla hesaplanan Ki-Kare testine ilişkin prob değerinin 0.0000 olması modelin iyi uyum sağlamadığı anlamına gelmektedir. Bir başka ifadeyle gözlenen değişkenlerin kovaryans matrisi ve ortalama vektörünün anakütle kovaryans matrisine ve ortalama vektörüne eşit olmadığı



İstenirse bir üçüncü alternatif olarak SEM Builder ekranında sol tarafta yer alan sembollerin doğrudan kullanımı yardımıyla da ölçüm modeli oluşturulabilir. Bunun için gösterge değişken sayısı kadar dikdörtgensel sembol seçilip boş alana yerleştirilir. Her biri tek tek tıklanarak ekranda çıkan “variable” alanında hangi gösterge değişkeni ifade ettiği seçilerek tanımlanır. Gizil değişken için ise oval sembolden yararlanılır. Her bir gizil değişken için bir oval sembol seçilir ve boş alana yerleştirilir. Bu esnada ekranda çıkan “name” alanına gizil değişkenin ismi yazılır. Böylece oluşturulan gizil değişkenler ilişkili olduğu gösterge değişkenlerle tek yönlü oklar yardımıyla bağlanır.

Şekil 18.6. SEM Builder Ölçüm Modeli Menüsü

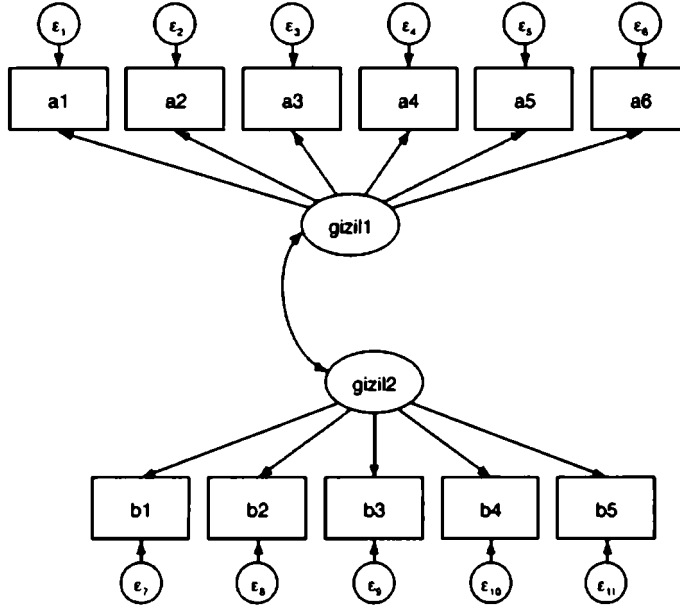
Diğer gizil değişkene ilişkin yapının belirlenmesinin ve gerek duyulursa gizil değişkenler arasındaki kovaryansın (çift yönlü ok) eklenmesinin ardından oluşan SEM Builder ekranı Şekil 18.7’de verilmektedir. Şemada gösterge değişkenlere bağlı olarak görülen dairesel yapılar ise ölçme hatalarını ifade etmektedir.

Oluşturulan iki gizil değişkenli bu ölçüm modelinin tahmin edilmesi için SEM Builder ekranında yer alan  butonu tıklanmalıdır. Böylece ölçüm modelinin tahminine ilişkin tercihlerin belirlendiği alt ekran açılmaktadır. Bu ekranda kullanılacak tahmin yöntemi seçilmekle birlikte aynı zamanda sonuçların seçilen bir değişkenin kategorilerine göre ayrı ayrı verilmesine ilişkin tercihler de yapılabilmektedir.

Tahminlenen ölçüm modeline ilişkin elde edilen standartlaştırılmamış katsayılar Şekil 18.8’de gösterilmektedir. Şayet standartlaştırılmış katsayıların yer aldığı sonuçlara ilişkin görsel istenilirse bu durumda SEM Builder ekranında aşağıdaki menü yolu takip edilir:

View > SEM Builder > View > Standardized estimates

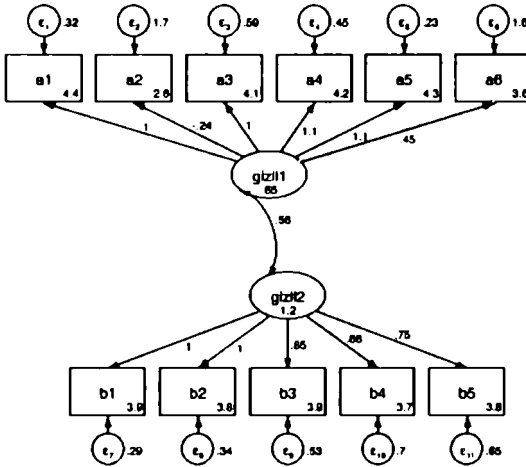
Şekil 18.7. SEM Builder Ölçüm Modeli



Modele ilişkin testlere ve ölçütlere ilişkin sonuçlara ulaşmak istenirse yine aşağıda verilen menü yolu takip edilebilir.

Statistics > SEM (structural equation modeling)

Şekil 18.8. Ölçüm Modeli Standartlaştırılmamış Sonuçları



| Details       |          |
|---------------|----------|
| ▼ Coefficient |          |
| B             | .453515  |
| SE            | .1029389 |
| z             | 4.405671 |
| P             | .0000105 |
| CI-LB         | .2517584 |
| CI-UB         | .6552717 |
| ▼ Std. Coef   |          |
| B             | .2814654 |
| SE            | .0602099 |
| z             | 4.674734 |
| P             | 2.94e-06 |
| CI-LB         | .1634561 |
| CI-UB         | .3994747 |

Şematik olarak elde edilen ölçüm modeli sonuçlarının yer aldığı Şekil 18.8'de gösterge değişkenler (*a1*-*a6*, *b1*-*b5*) ile gizil değişkenler (*gizil1* ve *gizil2*) arasında elde edilen modellere ilişkin katsayılar yer almaktadır. Her bir tek yönlü ok regresyon katsayısı gibi yorumlanmaktayken gizil değişkenler arasında mevcut olan çift yönlü ok ise bu değişkenler arasındaki ilişkiyi vermektedir. Şekil 18.8'de yer alan katsayılar dikkat edilirse, önceki sayfalarda verilmiş olan Stata model çıktısı (Tablo 18.6)



ile uyushmaktadır. SEM Builder ile elde edilen bu şematik gösterimde her hangi bir ok tıklandığında ekranın sağ tarafında ilgili katsayıya ilişkin tüm bilgiler görülebilmektedir. Katsayının büyüklüğü, standart hatası, anlamlılığı ve standartlaştırılmış katsayı değerlerine buradan ulaşmak mümkün olmaktadır. Yani bir anlamda söz konusu şematik gösterim model çıktısının görsel sunumudur.

Oluşturulan ölçüm modelinin elde edilen uyum ölçütlerine bakarak modelin uygunluğuna dair karar verilebilmektedir. Şayet uyum ölçütleri istenilen sınırların dışında kalırsa bu durumda önerilen düzeltme indeksleri kullanılarak model revize edilir ve yeniden tahmin edilir. Aşağıda Tablo 18.7 ile verilen sonuçlar dikkate alındığında, modelden elde edilen kovaryans matrisi ile örnek kovaryans matrisinin uyumuna ilişkin testlerden Ki-Kare Testi sonuçlarına göre uyum sağlanmamış olduğu görülmektedir. Tablo 18.5'te verilen sınırlar dikkate alınarak benzer şekilde RMSEA ölçütünün değerinin 0.102 olduğu ve yine kabul edilebilir sınırlarda bulunmadığı görülmektedir. CFI ölçütüne göre de model bu hâliyle kabul edilebilir bulunmazken bir tek SRMR ölçütü uyarınca kabul edilebilir sınırlar içerisinde. Model uyum ölçütlerinin ortaya koyduğu bu sonuç itibarıyla her ne kadar gösterge değişkenler ile gizil değişkenler arasında elde edilen katsayılar istatistiksel olarak anlamlı bulunmuş olsa da modelde düzeltme yapılması gerekmektedir. Çünkü bu hâliyle model ile örneklem varyans-kovaryans yapıları uyushmamaktadır. Bu durumda önceki bölümlerde ele alınmış olan düzeltme indekslerinden yararlanılmalıdır.

**Tablo 18.7.** Ölçüm Modeline İlişkin Uyum İyiliği Ölçütleri

estat gof, stats(all)

| Fit statistic        | Value    | Description                              |
|----------------------|----------|------------------------------------------|
| Likelihood ratio     |          |                                          |
| chi2_ms(43)          | 161.343  | model vs. saturated                      |
| p > chi2             | 0.000    |                                          |
| chi2_bs(55)          | 1652.752 | baseline vs. saturated                   |
| p > chi2             | 0.000    |                                          |
| Population error     |          |                                          |
| RMSEA                | 0.102    | Root mean squared error of approximation |
| 90% CI, lower bound  | 0.085    |                                          |
| upper bound          | 0.119    |                                          |
| pclose               | 0.000    | Probability RMSEA <= 0.05                |
| Information criteria |          |                                          |
| AIC                  | 7712.514 | Akaike's information criterion           |
| BIC                  | 7834.225 | Bayesian information criterion           |
| Baseline comparison  |          |                                          |
| CFI                  | 0.926    | Comparative fit index                    |
| TLI                  | 0.905    | Tucker-Lewis index                       |
| Size of residuals    |          |                                          |
| SRMR                 | 0.055    | Standardized root mean squared residual  |
| CD                   | 0.987    | Coefficient of determination             |

Aşağıda komut yapısı ve sonuçları verilmekte olan düzeltme indeksleri temel olarak iki kısımdan oluşmaktadır. Bunlardan ilki gözlenen değişkenler ile gizil değişkenler arasındaki ilişkileri gösterirken, ikincisi gözlenen değişkenlere ait hata terimleri arasındaki ilişkileri içermektedir. İlk kısımda yer alan düzeltme önerileri bir gizil değişken ve onunla ilişkili olmayan gösterge değişkenin ilişkilendirilmesi sonucunda Ki-Kare değerindeki düşüşü göstermektedir. YEM'nin temel varsayımlarından biri hata terimlerinin bağımsız olmasıdır. İkinci kısımda yer alan düzeltme önerileri ise hata terimleri arasındaki ilişkiyi göstermektedir. Bu durumda kuramsal yapı bozulmakta ve gizil değişkenlerin bağımsızlığı sağlanmamaktadır. Oysa ölçüm modelinde yer alan gizil değişkenlerin ilişkili olması durumunda gösterge değişkenlerinin istenildiği şekilde ayrışması da mümkün olmayacaktır. Bu durum klasik regresyon modelinin varsayımları içerisinde de yer alan çoklu bağlantı sorununu ortaya çıkaracaktır. Stata programı ile elde edilen ölçüm modeline ilişkin önerilen düzeltme indekslerini elde etmek için kullanılacak komut yapısı ve sonuçları aşağıda verilmektedir.

**Tablo 18.8.** Ölçüm Modeline İlişkin Düzeltme İndeksleri

**estat mindices**

Modification indices

|                | MI     | df | P>MI | EPC       | Standard EPC |
|----------------|--------|----|------|-----------|--------------|
| Measurement    |        |    |      |           |              |
| a3 <-          |        |    |      |           |              |
| gizil2         | 10.027 | 1  | 0.00 | .2203879  | .2099124     |
| a5 <-          |        |    |      |           |              |
| gizil2         | 4.288  | 1  | 0.04 | -.1119508 | -.1221298    |
| a6 <-          |        |    |      |           |              |
| gizil2         | 3.852  | 1  | 0.05 | .2030281  | .1677558     |
| cov(e.a1,e.b1) | 10.124 | 1  | 0.00 | .0809545  | .266955      |
| cov(e.a1,e.b4) | 5.211  | 1  | 0.02 | -.0782953 | -.1663331    |
| cov(e.a2,e.a4) | 4.549  | 1  | 0.03 | -.1267628 | -.146098     |
| cov(e.a2,e.a6) | 5.193  | 1  | 0.02 | .2289255  | .140872      |
| cov(e.a3,e.a6) | 9.194  | 1  | 0.00 | .1942736  | .201469      |
| cov(e.a3,e.b4) | 4.437  | 1  | 0.04 | .0935423  | .144962      |
| cov(e.a4,e.a5) | 4.206  | 1  | 0.04 | .0752802  | .2327712     |
| cov(e.a4,e.b1) | 3.935  | 1  | 0.05 | .0583639  | .1621027     |
| cov(e.a5,e.b1) | 5.689  | 1  | 0.02 | -.0566161 | -.2168064    |
| cov(e.b1,e.b2) | 41.290 | 1  | 0.00 | .2706709  | .8566537     |
| cov(e.b1,e.b4) | 7.445  | 1  | 0.01 | -.1113893 | -.246842     |
| cov(e.b1,e.b5) | 22.122 | 1  | 0.00 | -.177865  | -.4104645    |
| cov(e.b2,e.b4) | 13.147 | 1  | 0.00 | -.1530427 | -.312049     |
| cov(e.b2,e.b5) | 4.337  | 1  | 0.04 | -.0816845 | -.173444     |
| cov(e.b3,e.b4) | 11.293 | 1  | 0.00 | .1485217  | .2433109     |
| cov(e.b4,e.b5) | 46.029 | 1  | 0.00 | .3164086  | .4704135     |

1.KISIM

2.KISIM

EPC = beklenen parametre değişimi (expected parameter change)

Tablo 18.8’de verilen sonuçlara bakıldığında MI, modifikasyon katsayısına karşılık gelmektedir ve ilk sütunda ifade edilen ilişkinin modele eklenmesi hâlinde uyum iyiliği ölçütleri içerisinde yer alan Ki-Kare değerinde oluşacak değişimi göstermektedir. Buna göre en büyük değişimi verecek olan katsayıya karşılık gelen ilişkiden başlanarak modelde revizyona gidilmelidir. Örneğin yukarıda verilen tablodaki sonuçlara göre “a3” gösterge değişkeni ile “gizil2” gizil değişkeni arasında bir ilişki söz konusudur. Bu durumda ya bu gösterge değişken modelden çıkarılmalıdır ya da bu önerilen ilişki modele eklenmeli ve “a3” ile “gizil2” arasında tek yönlü bir ok çizilmelidir. İlk durumun tercih edilmesi hâlinde elbette bir bilgi kaybı olacaktır ancak değişkenler arasındaki bağımsızlık korunmaya devam edecektir. İkinci kısımda yer alan düzeltme indeksleri ise gösterge değişkenlerin hata terimleri arasındaki önerilen ilişkileri ifade etmektedir. Örneğin “b4” ve “b5” gösterge değişkenlerinin hata varyansları birbiri ile yüksek derecede ilişkili bulunmuş ve aralarında karşılıklı çift yönlü ok yardımıyla bir ilişkinin eklenmesi hâlinde modelin Ki-Kare uyum iyiliği ölçütünde önemli derecede bir iyileşme sağlanacaktır. Düzeltme indekslerinden yararlanılarak modelde bir revizyon yapılırken dikkat edilmesi gereken önemli hususlardan biri adım adım ilerlemek olmalıdır. Tüm önerilen düzeltmelerin birden yapılmasından ziyade her bir düzeltmeden sonra model yeniden tahminlenmeli ve uyum iyiliğindeki sonuçlar kontrol edilmelidir.

Ölçüm modelinde düzeltme indekslerinin kullanımı sonucunda kabul edilebilir modelin elde edilmesinin ardından<sup>7</sup> bu defa gizil değişkenler arasındaki nedensel ilişki yapısını gösteren yapısal modelin tahmin sürecine geçilir. Şekil 18.4 ile verilen yapısal modelin Stata yardımıyla elde edilmesi amacıyla yazılacak komut yapısı, ölçüm modelinde kullanılan komuta benzemekle birlikte, gizil değişkenler arasındaki ilişkiyi de ifade eden (gizil1 -> gizil2,) terimin komuta eklenmesiyle oluşturulmaktadır. Bu komut yapısında dikkat edilirse ölçüm modelinde de ifade edilen ilişkiler (gösterge değişkenlerin gizil değişkenlerle olan bağlantıları) yer almaya devam etmektedir.

### Tablo 18.8. Örnek 1 İçin Yapısal Model

```
sem (gizil1 -> a1,) (gizil1 -> a2,) (gizil1 -> a3,) (gizil1 -> a4,) (gizil1 ->
a5,) (gizil1 -> a6,) (gizil1 -> gizil2,) (gizil2 -> b1,) (gizil2 -> b2,) (gizil2
-> b3,) (gizil2 -> b4,) (gizil2 -> b5,), latent(gizil1 gizil2) nocapslatent
```

Endogenous variables

Measurement: a1 a2 a3 a4 a5 a6 b1 b2 b3 b4 b5

Latent: gizil2

Exogenous variables

Latent: gizil1

Structural equation model

Number of obs = 265

Estimation method = ml

Log likelihood = -3822.257

( 1) [b1]gizil2 = 1

( 2) [a1]gizil1 = 1

<sup>7</sup> Bu hâliyle model, metin içinde de değinildiği üzere, henüz uygun hâle gelmemiş sadece uygun hâle gelmesi için önerilen düzeltmelere değinilmiştir. Kitabın bu bölümünde, ele alınan uygulamalarda uygun modelin elde edilmesi amacından ziyade Stata yardımıyla ölçüm ve yapısal model analizlerinin nasıl gerçekleştirildiğinin anlatımı benimsenmiştir. İteratif tekniklerle uygun modelin elde edilmesi süreci kitabın amacının dışına çıkacağından araştırmacıların bu noktada doğrudan YEM başlığı ile yazılmış kaynaklardan yararlanması yerinde olacaktır.

|             |                 | OIM       |       |       |                      |           |
|-------------|-----------------|-----------|-------|-------|----------------------|-----------|
|             | Coef.           | Std. Err. | z     | P> z  | [95% Conf. Interval] |           |
| <hr/>       |                 |           |       |       |                      |           |
| Structural  |                 |           |       |       |                      | 1.KISIM   |
| gizil2 <-   |                 |           |       |       |                      |           |
| gizil1      | .8482537        | .086665   | 9.79  | 0.000 | .6783935             | 1.018114  |
| <hr/>       |                 |           |       |       |                      |           |
| Measurement |                 |           |       |       |                      | 2.KISIM   |
| a1 <-       |                 |           |       |       |                      |           |
| gizil1      | 1 (constrained) |           |       |       |                      |           |
| _cons       | 4.437736        | .0605214  | 73.33 | 0.000 | 4.319116             | 4.556356  |
| <hr/>       |                 |           |       |       |                      |           |
| a2 <-       |                 |           |       |       |                      |           |
| gizil1      | -.2443299       | .1052202  | -2.32 | 0.020 | -.4505576            | -.0381022 |
| _cons       | 2.630189        | .0807369  | 32.58 | 0.000 | 2.471947             | 2.78843   |
| <hr/>       |                 |           |       |       |                      |           |
| a3 <-       |                 |           |       |       |                      |           |
| gizil1      | 1.022377        | .0792249  | 12.90 | 0.000 | .8670988             | 1.177655  |
| _cons       | 4.067925        | .0694562  | 58.57 | 0.000 | 3.931793             | 4.204056  |
| <hr/>       |                 |           |       |       |                      |           |
| a4 <-       |                 |           |       |       |                      |           |
| gizil1      | 1.068601        | .075272   | 14.20 | 0.000 | .9210707             | 1.216131  |
| _cons       | 4.211321        | .0670983  | 62.76 | 0.000 | 4.07981              | 4.342831  |
| <hr/>       |                 |           |       |       |                      |           |
| a5 <-       |                 |           |       |       |                      |           |
| gizil1      | 1.063405        | .0651621  | 16.32 | 0.000 | .9356893             | 1.19112   |
| _cons       | 4.30566         | .0606411  | 71.00 | 0.000 | 4.186806             | 4.424515  |
| <hr/>       |                 |           |       |       |                      |           |
| a6 <-       |                 |           |       |       |                      |           |
| gizil1      | .453515         | .1029389  | 4.41  | 0.000 | .2517584             | .6552717  |
| _cons       | 3.588679        | .0800646  | 44.82 | 0.000 | 3.431756             | 3.745603  |
| <hr/>       |                 |           |       |       |                      |           |
| b1 <-       |                 |           |       |       |                      |           |
| gizil2      | 1 (constrained) |           |       |       |                      |           |
| _cons       | 3.867925        | .0739832  | 52.28 | 0.000 | 3.72292              | 4.012929  |
| <hr/>       |                 |           |       |       |                      |           |
| b2 <-       |                 |           |       |       |                      |           |
| gizil2      | 1.011356        | .048738   | 20.75 | 0.000 | .9158312             | 1.106881  |
| _cons       | 3.830189        | .0759755  | 50.41 | 0.000 | 3.681279             | 3.979098  |
| <hr/>       |                 |           |       |       |                      |           |
| b3 <-       |                 |           |       |       |                      |           |
| gizil2      | .852132         | .0537329  | 15.86 | 0.000 | .7468174             | .9574466  |
| _cons       | 3.879245        | .0720088  | 53.87 | 0.000 | 3.738111             | 4.02038   |
| <hr/>       |                 |           |       |       |                      |           |
| b4 <-       |                 |           |       |       |                      |           |
| gizil2      | .8587035        | .0602426  | 14.25 | 0.000 | .7406302             | .9767769  |
| _cons       | 3.69434         | .0766179  | 48.22 | 0.000 | 3.544171             | 3.844508  |
| <hr/>       |                 |           |       |       |                      |           |
| b5 <-       |                 |           |       |       |                      |           |
| gizil2      | .7484092        | .0567455  | 13.19 | 0.000 | .6371901             | .8596283  |
| _cons       | 3.818868        | .0699189  | 54.62 | 0.000 | 3.681829             | 3.955906  |
| <hr/>       |                 |           |       |       |                      |           |
| var(e.a1)   | .3163266        | .0364155  |       |       | .2524325             | .396393   |

|               |          |          |          |          |
|---------------|----------|----------|----------|----------|
| var(e.a2)     | 1.688329 | .1470782 | 1.423328 | 2.002669 |
| var(e.a3)     | .5944693 | .0604772 | .4870057 | .7256461 |
| var(e.a4)     | .4459004 | .0483995 | .3604505 | .5516073 |
| var(e.a5)     | .2345664 | .0326727 | .1785263 | .3081977 |
| var(e.a6)     | 1.564161 | .1373111 | 1.316916 | 1.857825 |
| var(e.b1)     | .2907171 | .0402254 | .221663  | .3812835 |
| var(e.b2)     | .3434013 | .0437038 | .267591  | .440689  |
| var(e.b3)     | .5319597 | .0545068 | .4351721 | .6502739 |
| var(e.b4)     | .7004515 | .070512  | .5750302 | .8532288 |
| var(e.b5)     | .6458906 | .063422  | .5328154 | .7829628 |
| var(e.gizil2) | .6889542 | .0833658 | .5434906 | .8733508 |
| var(gizil1)   | .6543249 | .0832249 | .5099498 | .8395751 |

-----  
 LR test of model vs. saturated: chi2(43) = 161.34, Prob > chi2 = 0.0000

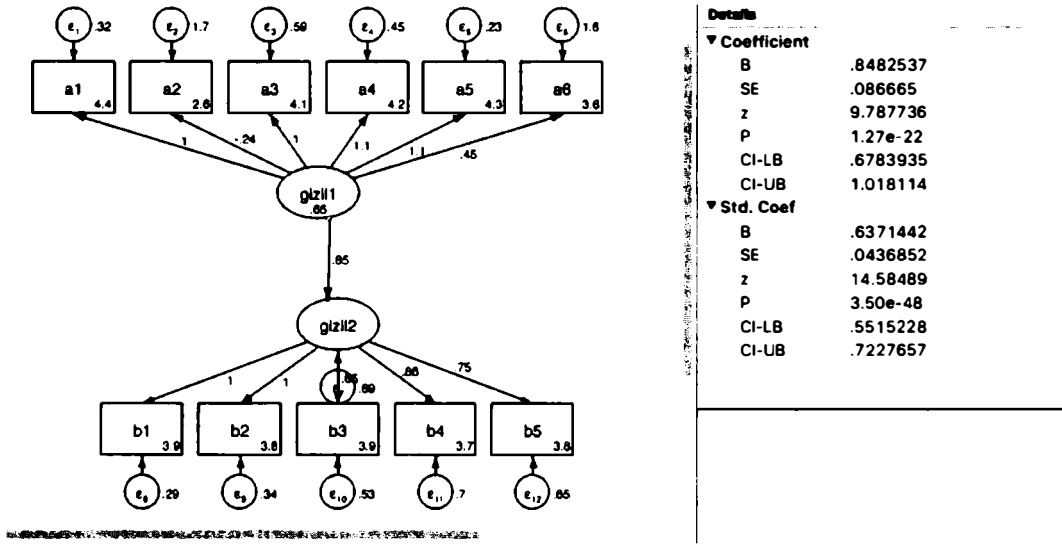
Tablo 18.8’de elde edilen bu model çıktısında ilk dikkat çeken nokta, gizil değişkenlere bağlı gösterge değişkenlerden, ilk değişkenlerin katsayılarının 1 ile sabitlenmesidir. Bu sayede gizil değişkene, gösterge değişkenle aynı varyans tanımlanmış olmaktadır. Yapısal model çıktısı, ölçüm modelinden farklı olarak hem ölçüm modeli sonuçlarını (2.Kısım ile verilen) hem de yapısal model sonuçlarını (1.Kısım ile verilen) bir arada sunmaktadır. Ölçüm modeli önceki kısımlarda ayrıntılı olarak ele alındığından burada sadece yapısal kısım ile ilgilenilmektedir. Buna göre, Tablo 18.8’de görüleceği üzere, “gizil1” değişkeninden “gizil2” değişkenine tanımlanan tek yönlü ilişkinin istatistiksel olarak anlamlı ve katsayısının da 0.8482537 olduğu görülmektedir. Bu durumda gizil değişkenler arasında bir nedensellik ilişkisinin olduğu ifade edilebilmektedir. Yine ölçüm modelinin tahmininde olduğu gibi, yapısal modelde de EÇO yöntemi ve OIM tekniği model tahmini için kullanılmıştır. Model çıktısının son kısmında ise ölçüm modeli sonuçlarında da verilmiş olan gösterge değişkenlerin hatalarının varyansları yer almaktadır.

Yapısal modelin elde edilmesinin ardından yine bu modele ilişkin uyum ölçütlerinin incelenmesi yerinde olacaktır. Tablo 18.8’in sonunda yer alan Ki-Kare istatistiği mevcut yapısal modelin uyum iyiliğini sağlamadığını göstermektedir (Prob > chi2 = 0.0000). Ancak diğer uyum ölçütleri model çıktısında doğrudan yer almamaktadır. Onlara ait değerleri görebilmek amacıyla ölçüm modelinde ifade edildiği gibi “**estat gof, stats(all)**” komutu kullanılmalıdır. Şayet model kabul edilebilir bulunmazsa bu durumda yine düzeltme indekslerinden yararlanılır. Yapısal modelde düzeltme indekslerinin ve uyum ölçütlerinin elde edilmesi süreci ölçüm modelinde olduğu gibidir ve aynı komutlar yardımıyla gerçekleştirilir.

Yapısal modelin şematik gösterimi Şekil 18.9 ile verilmektedir. Dikkat edilirse burada yer alan sonuçlar, yukarıda yer alan Tablo 18.8’deki sonuçların aynısıdır. SEM Builder sonuç ekranında modelde yer alan herhangi bir sembol seçili hâle getirildiğinde, ölçüm modelinde olduğu üzere, ekranın sağ tarafında ilgili sembole ilişkin detaylar verilmektedir. Böylece oklarla gösterilen katsayıların anlamlılıkları kolaylıkla bu ekrandan da görülebilmektedir. Örneğin Şekil 18.9’da gizil1’den gizil2’ye doğru olan ok, gizil1 ve gizil2 değişkenleri arasındaki nedensel ilişkiyi göstermektedir ve bu ilişkiye ait katsayı istatistiksel olarak anlamlı bulunmuştur. Aynı zamanda standartlaştırılmış katsayılar da kolaylıkla bu kısımdan okunabilmektedir. Aynı (gizil) bağımlı değişken üzerinde etkili olan bir veya daha fazla gösterge değişken, farklı ölçüm birimleri ile ölçülmüş ise, bu durumda etkileri standartlaştırılmamış katsayıları kullanarak açıklamak oldukça

zor olmaktadır. Bu nedenle standartlaştırılmış katsayılar, farklı gösterge (açıklayıcı) değişkenlerin göreceli etkilerini karşılaştırmada oldukça kullanışlıdır.

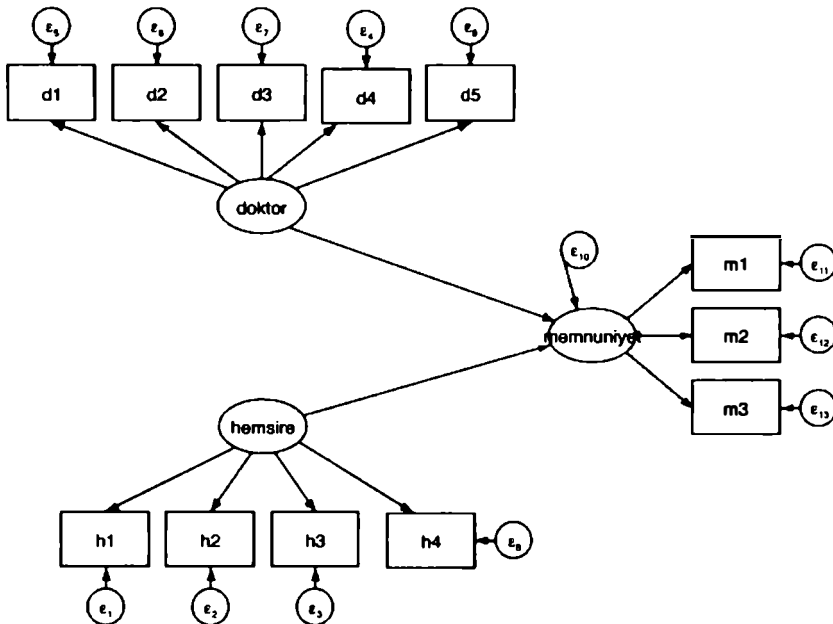
Şekil 18.9. Yapısal Model Standartlaştırılmamış Sonuçları



## 18.6. UYGULAMALI ÖRNEK-2

256 yatan hasta üzerinde gerçekleştirilen bir çalışmada hastaların hemşire ve doktorlardan memnuniyetlerinin genel memnuniyet düzeylerini nasıl etkilediği araştırılmak istenmiştir ve memnuniyet.dta<sup>8</sup> isimli veri dosyası oluşturulmuştur. Kuramsal temel dikkate alınarak model aşağıdaki gibi oluşturulmuştur.

Şekil 18.10. Kuramsal Model



<sup>8</sup> Veriye erişmek için link'i kullanınız.

Bir önceki uygulama örneğinde verildiği üzere öncelikle ölçüm modeli test edilecek ve uyum iyiliği ölçütlerine göre gerekirse düzeltme indeksleri yardımıyla revize edilecektir. Kabul edilebilir ölçüm modeline ulaşılmasının ardından ise yapısal model test edilecek ve model uygunluğunun sağlanması hâlinde dolaylı etkilere nasıl bakılabileceği gösterilecektir. Dikkat edilirse bir önceki uygulamada dolaylı etkilerden bahsedilmemişti. Bir önceki uygulamada iki gizil değişken ile çalışılmış olması nedeniyle aralarında test edilen ilişki doğrudan bir ilişki yapısıydı. Ancak Şekil 18.10'dan da görüldüğü üzere bu yapısal modelde üç gizil değişken bulunmaktadır ve şayet "hemşire" gizil değişkeninden "doktor" gizil değişkeni yönüne bir nedensellik modele eklenecek olursa bu durumda "memnuniyet" üzerinde dolaylı etki ortaya çıkacaktır.

Uygulamada kullanılan veri setinde kayıp gözlemler olması nedeniyle ölçüm ve yapısal model tahmininde tahmin yöntemi olarak kayıp gözlemlili EÇO seçilmiştir. Bir önceki uygulamada belirtildiği gibi burada da veri setinin çok değişkenli normalliği sağladığı varsayılmıştır. Stata tek tek değişkenlerin normalliğini sınavabilirken çoklu normallik sınavasını gerçekleştirememektedir.

Şekil 18.10 ile verilen yapısal modelin elde edilmesi için kullanılan komut yapısı ve sonuçları Tablo 18.9 ile verilmektedir. Dikkat edilirse bir önceki örnekten farklı olarak bu komut yapısında yöntem olarak MLMV (eksik gözlemlili EÇO) kullanılmıştır. Yine komut yapısı aynı zamanda ölçüm modelini de içermektedir. Buna göre "h" ile isimlendirilmiş gösterge değişkenler "hemşire" gizil değişkeni, "d" ile olanlar "doktor" gizil değişkeni ve "m" ile olanlar ise "memnuniyet" gizil değişkenini açıklamaktadırlar.

**Tablo 18.9.** Örnek 2 İçin Yapısal Model

```
sem (hemsire -> h1,) (hemsire -> h2,) (hemsire -> h3,) (hemsire -> h4,) (hemsire
-> memnuniyet,) (doktor -> d1,) (doktor -> d2,) (doktor -> d3,) (doktor -> d4,)
(doktor -> d5,) (doktor -> memnuniyet,) (memnuniyet -> m1,) (memnuniyet -> m2,)
(memnuniyet -> m3,), covstruct(_lexogenous, diagonal) method(mlmv) latent(hemsire
doktor memnuniyet) nocapslatent
```

Endogenous variables

Measurement: h1 h2 h3 h4 d1 d2 d3 d4 d5 m1 m2 m3

Latent: memnuniyet

Exogenous variables

Latent: hemsire doktor

Structural equation model Number of obs = 298

Estimation method = mlmv

Log likelihood = -3809.4021

( 1) [m1]memnuniyet = 1

( 2) [h1]hemsire = 1

( 3) [d1]doktor = 1

|               |  | OIM   |           |   |      |                      |
|---------------|--|-------|-----------|---|------|----------------------|
|               |  | Coef. | Std. Err. | z | P> z | [95% Conf. Interval] |
| Structural    |  |       |           |   |      |                      |
| memnuniyet <- |  |       |           |   |      | 1.KISIM              |





|                   |          |          |          |          |
|-------------------|----------|----------|----------|----------|
| var(e.h1)         | .1110187 | .0106979 | .0919122 | .1340969 |
| var(e.h2)         | .0330754 | .0076289 | .0210463 | .0519799 |
| var(e.h3)         | .1126531 | .0126299 | .09043   | .1403375 |
| var(e.h4)         | .2477685 | .0229383 | .2066533 | .2970639 |
| var(e.d1)         | .3827722 | .0432797 | .3066873 | .4777326 |
| var(e.d2)         | .3019033 | .0335468 | .2428199 | .3753631 |
| var(e.d3)         | .2347609 | .027913  | .1859595 | .2963693 |
| var(e.d4)         | 1.612652 | .1348612 | 1.368855 | 1.899871 |
| var(e.d5)         | .6644131 | .0660321 | .5468174 | .8072983 |
| var(e.m1)         | 1.501671 | .1521224 | 1.23125  | 1.831484 |
| var(e.m2)         | .5820053 | .0731797 | .4548826 | .7446541 |
| var(e.m3)         | .2574413 | .0389609 | .1913632 | .3463364 |
| var(e.memnuniyet) | .3768631 | .1014251 | .2223821 | .6386564 |
| var(hemsire)      | .4123924 | .0422807 | .3373191 | .5041741 |
| var(doktor)       | .5421974 | .0751569 | .4132071 | .7114545 |

-----  
 LR test of model vs. saturated: chi2(52) = 202.56, Prob > chi2 = 0.0000

Tablo 18.9 model çıktısında öncelikle her gizil değişken için varyans hesaplamasında kullanılmak amacıyla bir gösterge değişkenin katsayısı 1 ile sabitlenmiştir. Ardından 1. kısımda yapısal model sonuçları ve 2. kısımda ise ölçüm modeli sonuçlarına yer verilmektedir. İlk kısımda yer alan sonuçlara göre “hemşire” gizil değişkeninden “memnuniyet” gizil değişkenine tanımlanan ilişki istatistiksel olarak anlamsız bulunmuştur. Buna karşın “doktor” değişkeninin “memnuniyet” üzerinde pozitif yönlü (0.5878931) anlamlı bir etkisinin olduğu görülmektedir.

İkinci kısımda yer alan ölçüm modeli sonuçlarına göre her bir gösterge değişkenin (h1-h4, d1-d5, m1-m3) gizil değişken üzerinde anlamlı ve pozitif yönlü bir etkisi olduğu görülmektedir. Model çıktısının son kısmında ise yapısal model içerisinde yer alan tüm regresyon modellerine ilişkin hata terimlerinin varyansları yer almaktadır.

Elde edilen model çıktısında her ne kadar tüm katsayılar anlamlı olarak bulunmuş olsa da modelin uyumunun sınanması ve gerekli olması hâlinde düzeltilmesi yerinde olacaktır. Tahminlenen yapısal modelin aşağıda Tablo 18.10 ile sonuçları verilen uyum iyiliği indekslerine (Ki-Kare, RMSEA, CFI) bakıldığında modelin Tablo 18.5 uyarınca kabul edilebilir olmadığı anlaşılmaktadır. Bu amaçla oluşturulan yapısal model için düzeltme indekslerinden yararlanılmalıdır.

**Tablo 18.10.** Örnek 2 için Yapısal Modele İlişkin Uyum İyiliği Sonuçları

estat gof, stats(all)

| Fit statistic    | Value    | Description            |
|------------------|----------|------------------------|
| -----            |          |                        |
| Likelihood ratio |          |                        |
| chi2_ms(52)      | 202.556  | model vs. saturated    |
| p > chi2         | 0.000    |                        |
| chi2_bs(66)      | 2040.145 | baseline vs. saturated |
| p > chi2         | 0.000    |                        |

|                      |  |          |                                          |
|----------------------|--|----------|------------------------------------------|
| -----+-----          |  |          |                                          |
| Population error     |  |          |                                          |
| RMSEA                |  | 0.099    | Root mean squared error of approximation |
| 90% CI, lower bound  |  | 0.084    |                                          |
| upper bound          |  | 0.113    |                                          |
| pclose               |  | 0.000    | Probability RMSEA <= 0.05 *              |
| -----+-----          |  |          |                                          |
| Information criteria |  |          |                                          |
| AIC                  |  | 7694.804 | Akaike's information criterion           |
| BIC                  |  | 7835.294 | Bayesian information criterion           |
| -----+-----          |  |          |                                          |
| Baseline comparison  |  |          |                                          |
| CFI                  |  | 0.924    | Comparative fit index                    |
| TLI                  |  | 0.903    | Tucker-Lewis index                       |
| -----+-----          |  |          |                                          |
| Size of residuals    |  |          |                                          |
| CD                   |  | 0.995    | Coefficient of determination             |

Düzeltilme indeksleri için Stata programında “**estat mindices**” komutundan yararlanıldığı önceki kısımlarda belirtilmişti. Yapısal model için elde edilen düzeltme indekslerinin aşağıda Tablo 18.11 ile verilen çıktılarına bakıldığında üç kısımdan oluştuğu görülmektedir. İlk kısımda yapısal modelde önerilen düzeltmelere yer verilmektedir. Örneğin memnuniyet ile “h4” değişkeni arasında bir ilişki önerilmekte ve bunun yapılması neticesinde uyum iyiliği değerlerinden Ki-Kare değerinin 9.623 azalacağı anlaşılmaktadır. İkinci kısımda yer alan düzeltme indeksleri ise ölçüm modelinde yer alan gösterge değişkenlerle ilgilidir. Burada verilen ilk düzeltme önerisi “h1” ve “h2” değişkenleri arasında bir ilişki tanımlanmasıdır. Bu kısımda dikkat edilirse “h4” ve “d4” değişkenlerinin pek çok değişkenle ilişkili olduğu görülmektedir. Son kısımda ise hata terimleri arasında tespit edilen kovaryanslar verilmiştir. Bu kısım söz konusu hata terimlerinin birbirleri ile karşılıklı ilişkide olmasını gösteren kısımdır.

**Tablo 18.11.** Örnek 2 İçin Yapısal Modele İlişkin Düzeltme İndeksleri

**estat mindices**

Modification indices

|               |    | Standard |    |      |           |           |
|---------------|----|----------|----|------|-----------|-----------|
|               |    | MI       | df | P>MI | EPC       | EPC       |
| Structural    |    |          |    |      |           |           |
| memnuniyet <- |    |          |    |      |           |           |
|               | h4 | 9.623    | 1  | 0.00 | -.3130686 | -.3717886 |

1.KISIM

|             |            |        |   |      |           |           |
|-------------|------------|--------|---|------|-----------|-----------|
|             | d2         | 5.787  | 1 | 0.02 | .2495019  | .2751754  |
|             | m2         | 7.441  | 1 | 0.01 | .4210156  | .5660157  |
|             | m3         | 5.830  | 1 | 0.02 | -.6624776 | -.6405584 |
| -----+----- |            |        |   |      |           |           |
| Measurement |            |        |   |      |           |           |
| h1 <-       |            |        |   |      |           |           |
|             | h2         | 14.739 | 1 | 0.00 | 1.265314  | 1.292356  |
|             | h4         | 14.073 | 1 | 0.00 | -.1814693 | -.2252386 |
| -----+----- |            |        |   |      |           |           |
| h2 <-       |            |        |   |      |           |           |
|             | h1         | 14.739 | 1 | 0.00 | .3769709  | .369083   |
|             | h3         | 14.784 | 1 | 0.00 | -.4411938 | -.4767564 |
| -----+----- |            |        |   |      |           |           |
| h3 <-       |            |        |   |      |           |           |
|             | h2         | 14.784 | 1 | 0.00 | -1.502682 | -1.390592 |
|             | h4         | 16.862 | 1 | 0.00 | .2131062  | .239654   |
| -----+----- |            |        |   |      |           |           |
| h4 <-       |            |        |   |      |           |           |
|             | h1         | 14.073 | 1 | 0.00 | -.4049984 | -.3262975 |
|             | h3         | 16.862 | 1 | 0.00 | .4687044  | .4167835  |
|             | d1         | 14.124 | 1 | 0.00 | .1190787  | .1275372  |
|             | d3         | 9.499  | 1 | 0.00 | .1224562  | .1052507  |
|             | d5         | 7.936  | 1 | 0.00 | .077332   | .0958198  |
|             | m2         | 6.294  | 1 | 0.01 | -.0755014 | -.085473  |
|             | doktor     | 8.629  | 1 | 0.00 | .1322985  | .1084857  |
| -----+----- |            |        |   |      |           |           |
| d1 <-       |            |        |   |      |           |           |
|             | h4         | 8.216  | 1 | 0.00 | .1307259  | .122056   |
| -----+----- |            |        |   |      |           |           |
| d2 <-       |            |        |   |      |           |           |
|             | m3         | 8.858  | 1 | 0.00 | .1724948  | .1512265  |
|             | memnuniyet | 5.987  | 1 | 0.01 | .1993754  | .1807739  |
| -----+----- |            |        |   |      |           |           |
| d3 <-       |            |        |   |      |           |           |
|             | h4         | 6.298  | 1 | 0.01 | .0917563  | .1067559  |
|             | d5         | 6.536  | 1 | 0.01 | .1337975  | .1928856  |
|             | m2         | 5.660  | 1 | 0.02 | -.0874467 | -.115179  |
| -----+----- |            |        |   |      |           |           |
| d4 <-       |            |        |   |      |           |           |
|             | h3         | 3.939  | 1 | 0.05 | -.1848204 | -.1113273 |
|             | h4         | 8.062  | 1 | 0.00 | -.2360345 | -.1598879 |
|             | m1         | 12.139 | 1 | 0.00 | .1950533  | .2118735  |
|             | m2         | 8.128  | 1 | 0.00 | .2292252  | .1757828  |
|             | m3         | 5.921  | 1 | 0.01 | -.2801827 | -.1545299 |
| -----+----- |            |        |   |      |           |           |
| d5 <-       |            |        |   |      |           |           |
|             | h1         | 8.678  | 1 | 0.00 | .2079511  | .1352153  |
|             | h2         | 6.827  | 1 | 0.01 | .1806888  | .1199995  |
|             | h3         | 4.818  | 1 | 0.03 | .1404079  | .1007644  |
|             | h4         | 9.095  | 1 | 0.00 | .1721571  | .1389405  |
|             | d3         | 6.536  | 1 | 0.01 | .3786695  | .2626687  |
|             | hemsire    | 7.603  | 1 | 0.01 | .2236033  | .1290555  |

2.KISIM

|                        |        |        |   |      |           |           |
|------------------------|--------|--------|---|------|-----------|-----------|
| -----+-----            |        |        |   |      |           |           |
| m1 <-                  |        |        |   |      |           |           |
|                        | d4     | 14.208 | 1 | 0.00 | .2253861  | .2074932  |
|                        | m2     | 5.830  | 1 | 0.02 | .3740082  | .2640413  |
|                        | m3     | 7.441  | 1 | 0.01 | -.7457415 | -.3786485 |
| -----+-----            |        |        |   |      |           |           |
| m2 <-                  |        |        |   |      |           |           |
|                        | h4     | 6.715  | 1 | 0.01 | -.156517  | -.1382572 |
|                        | d3     | 11.110 | 1 | 0.00 | -.3028099 | -.2299007 |
|                        | d4     | 9.021  | 1 | 0.00 | .1175372  | .1532714  |
|                        | m1     | 5.830  | 1 | 0.02 | .1449551  | .2053254  |
|                        | doktor | 5.715  | 1 | 0.02 | -.32887   | -.2382145 |
| -----+-----            |        |        |   |      |           |           |
| m3 <-                  |        |        |   |      |           |           |
|                        | d2     | 9.279  | 1 | 0.00 | .1832013  | .2089666  |
|                        | d3     | 8.124  | 1 | 0.00 | .1925252  | .203236   |
|                        | d4     | 16.507 | 1 | 0.00 | -.113431  | -.205665  |
|                        | m1     | 7.441  | 1 | 0.01 | -.1278474 | -.2517932 |
| -----+-----            |        |        |   |      |           |           |
| cov(e.h1,e.h2)         |        | 14.739 | 1 | 0.00 | .0418508  | .6906422  |
| 3.KISIM                |        |        |   |      |           |           |
| cov(e.h1,e.h4)         |        | 14.073 | 1 | 0.00 | -.0449624 | -.2710992 |
| cov(e.h2,e.h3)         |        | 14.784 | 1 | 0.00 | -.0497018 | -.814232  |
| cov(e.h3,e.h4)         |        | 16.862 | 1 | 0.00 | .052801   | .316044   |
| cov(e.h4,e.d1)         |        | 8.873  | 1 | 0.00 | .0636141  | .2065668  |
| cov(e.h4,e.d4)         |        | 7.840  | 1 | 0.01 | -.1094932 | -.1732183 |
| cov(e.h4,e.m2)         |        | 11.743 | 1 | 0.00 | -.0900571 | -.2371545 |
| cov(e.h4,e.memnuniyet) |        | 9.623  | 1 | 0.00 | -.0775685 | -.2538463 |
| cov(e.d1,e.m3)         |        | 4.106  | 1 | 0.04 | -.0519358 | -.1654464 |
| cov(e.d2,e.m3)         |        | 7.131  | 1 | 0.01 | .060293   | .2162689  |
| cov(e.d2,e.memnuniyet) |        | 5.787  | 1 | 0.02 | .0753255  | .223314   |
| cov(e.d3,e.d5)         |        | 6.536  | 1 | 0.01 | .0888968  | .2250889  |
| cov(e.d3,e.m2)         |        | 7.605  | 1 | 0.01 | -.0793051 | -.2145483 |
| cov(e.d3,e.m3)         |        | 5.979  | 1 | 0.01 | .0503302  | .2047277  |
| cov(e.d4,e.m1)         |        | 14.182 | 1 | 0.00 | .3726644  | .239475   |
| cov(e.d4,e.m2)         |        | 11.647 | 1 | 0.00 | .2191038  | .2261602  |
| cov(e.d4,e.m3)         |        | 19.164 | 1 | 0.00 | -.1999232 | -.31028   |
| cov(e.m1,e.m2)         |        | 5.830  | 1 | 0.02 | .2176748  | .2328398  |
| cov(e.m1,e.m3)         |        | 7.441  | 1 | 0.01 | -.1919847 | -.3087736 |
| cov(e.m2,e.memnuniyet) |        | 7.441  | 1 | 0.01 | .2450333  | .5232026  |
| cov(e.m3,e.memnuniyet) |        | 5.830  | 1 | 0.02 | -.1705492 | -.5475435 |
| cov(hemsire,doktor)    |        | 43.250 | 1 | 0.00 | .2006446  | .42432    |
| -----+-----            |        |        |   |      |           |           |

EPC = Beklenen parametre değişimi (expected parameter change)

Tablo 18.11’de verilen düzeltme indeks önerilerine göre sırasıyla gerekli revizyonlar yapıldığında oluşan model tahmininin yine kabul edilebilir bulunmaması neticesinde bir kez daha düzeltme indekslerinden yararlanılmıştır ve Tablo 18.12 ile verilen nihai modele ulaşılmıştır. Bu model ilk modelden farklı olarak, “h4” ve “d4” değişkenlerini içermemektedir.



|                                                                        |  |          |               |       |       |          |          |
|------------------------------------------------------------------------|--|----------|---------------|-------|-------|----------|----------|
| d5 <-                                                                  |  |          |               |       |       |          |          |
| doktor                                                                 |  | 1.017988 | .0953069      | 10.68 | 0.000 | .8311897 | 1.204786 |
| _cons                                                                  |  | 1.65007  | .0645893      | 25.55 | 0.000 | 1.523477 | 1.776662 |
| -----                                                                  |  |          |               |       |       |          |          |
| m1 <-                                                                  |  |          |               |       |       |          |          |
| memnuniyet                                                             |  | 1        | (constrained) |       |       |          |          |
| _cons                                                                  |  | 2.146798 | .085189       | 25.20 | 0.000 | 1.97983  | 2.313765 |
| -----                                                                  |  |          |               |       |       |          |          |
| m2 <-                                                                  |  |          |               |       |       |          |          |
| memnuniyet                                                             |  | .8920146 | .1287812      | 6.93  | 0.000 | .639608  | 1.144421 |
| _cons                                                                  |  | 2.006711 | .0588475      | 34.10 | 0.000 | 1.891372 | 2.12205  |
| -----                                                                  |  |          |               |       |       |          |          |
| m3 <-                                                                  |  |          |               |       |       |          |          |
| memnuniyet                                                             |  | .7148514 | .1135489      | 6.30  | 0.000 | .4922997 | .9374032 |
| _cons                                                                  |  | 1.590135 | .042649       | 37.28 | 0.000 | 1.506545 | 1.673726 |
| -----                                                                  |  |          |               |       |       |          |          |
| var(e.h1)                                                              |  | .1053631 | .0112791      |       |       | .0854215 | .12996   |
| HATA VARYANSLARI                                                       |  |          |               |       |       |          |          |
| var(e.h2)                                                              |  | .0301846 | .0092449      |       |       | .0165608 | .055016  |
| var(e.h3)                                                              |  | .1232031 | .0134266      |       |       | .099508  | .1525404 |
| var(e.d1)                                                              |  | .3804935 | .0435106      |       |       | .3040954 | .4760853 |
| var(e.d2)                                                              |  | .2998999 | .0337441      |       |       | .2405476 | .3738967 |
| var(e.d3)                                                              |  | .2325244 | .0281323      |       |       | .183436  | .2947492 |
| var(e.d5)                                                              |  | .6737041 | .0666634      |       |       | .5549353 | .8178923 |
| var(e.m1)                                                              |  | 1.517537 | .1530501      |       |       | 1.245351 | 1.849211 |
| var(e.m2)                                                              |  | .5906545 | .0732892      |       |       | .4631432 | .7532718 |
| var(e.m3)                                                              |  | .2497252 | .0393698      |       |       | .1833447 | .3401389 |
| var(e.memnuniyet)                                                      |  | .3673784 | .1007127      |       |       | .2146675 | .6287254 |
| var(hemsire)                                                           |  | .418048  | .0426246      |       |       | .3423235 | .5105233 |
| var(doktor)                                                            |  | .5444761 | .075445       |       |       | .4149851 | .714373  |
| -----                                                                  |  |          |               |       |       |          |          |
| LR test of model vs. saturated: chi2(33) = 93.79, Prob > chi2 = 0.0000 |  |          |               |       |       |          |          |

Yukarıda verilen yapısal model sonuçlarına göre “hemşire”lerden duyulan memnuniyetin genel “memnuniyet” üzerinde anlamlı bir etkisi bulunmazken “doktor”ların “memnuniyeti” pozitif yönlü olarak etkilediği görülmektedir. Ölçüm modeli sonuçlarına göre ise her bir gösterge değişkenin (h, d ve m) gizil değişkenler üzerinde pozitif yönde bir etkisinin bulunduğu söylenebilmektedir. Düzeltme indeksleri uyarınca elde edilen bu modelin uyum iyiliği sonuçları Tablo 18.13’te verilmektedir.

**Tablo 18.13.** Örnek 2 için Nihai Yapısal Modele İlişkin Uyum İyiliği Sonuçları

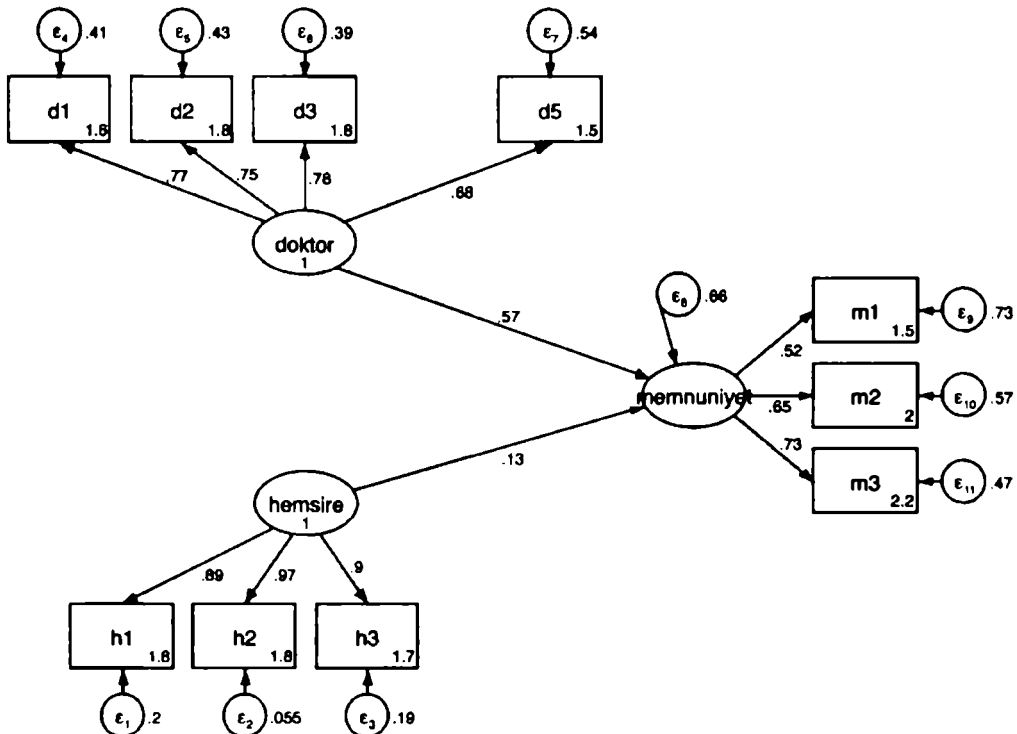
estat gof, stats(all)

| Fit statistic    | Value    | Description            |
|------------------|----------|------------------------|
| -----            |          |                        |
| Likelihood ratio |          |                        |
| chi2_ms(33)      | 93.785   | model vs. saturated    |
| p > chi2         | 0.000    |                        |
| chi2_bs(45)      | 1589.656 | baseline vs. saturated |
| p > chi2         | 0.000    |                        |
| -----            |          |                        |

|                      |          |                                          |  |
|----------------------|----------|------------------------------------------|--|
| Population error     |          |                                          |  |
| RMSEA                | 0.079    | Root mean squared error of approximation |  |
| 90% CI, lower bound  | 0.060    |                                          |  |
| upper bound          | 0.098    |                                          |  |
| pclose               | 0.007    | Probability RMSEA <= 0.05                |  |
| <hr/>                |          |                                          |  |
| Information criteria |          |                                          |  |
| AIC                  | 6250.882 | Akaike's information criterion           |  |
| BIC                  | 6369.189 | Bayesian information criterion           |  |
| <hr/>                |          |                                          |  |
| Baseline comparison  |          |                                          |  |
| CFI                  | 0.961    | Comparative fit index                    |  |
| TLI                  | 0.946    | Tucker-Lewis index                       |  |
| <hr/>                |          |                                          |  |
| Size of residuals    |          |                                          |  |
| CD                   | 0.994    | Coefficient of determination             |  |

Tablo 18.5'te sınırları verilen uyum iyiliği ölçütleri (RMSEA, CFI ve  $\chi^2/sd$ ) ve yukarıda yer alan Tablo 18.13 sonuçlarına göre modelin kabul edilebilir bir model olduğu görülmektedir. Model tahmin sonuçlarına ilişkin SEM Builder ile elde edilen şematik gösterim ise Şekil 18.11'de verilmektedir. Sonuçlar incelendiğinde hemşirelerden duyulan memnuniyetin genel memnuniyet üzerindeki etkisi anlamlı bulunmazken, doktorlardan duyulan memnuniyetin genel memnuniyeti anlamlı bir şekilde etkilediği görülmektedir. Doktora duyulan memnuniyetin artması (0.57) genel memnuniyet düzeyini de arttırmaktadır.

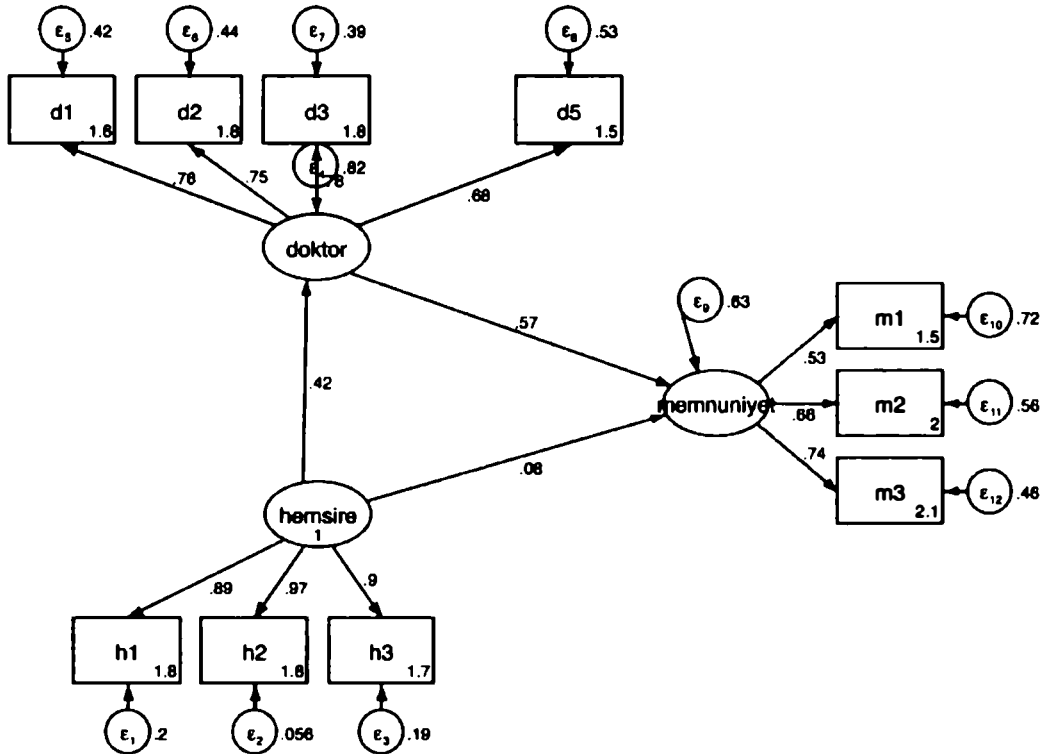
Şekil 18.11. Yapısal Model Sonuçları



Yapısal modelde yer alan gizil dışsal değişkenler arasında tanımlanacak bir ilişki, genel memnuniyet düzeyi (gizil içsel değişken) üzerinde dolaylı bir etkinin oluşmasına neden olacaktır. Oysa ki yukarıda tahmin sonuçları verilen yapısal modelde, genel memnuniyet üzerinde sadece doğrudan etkiler tanımlanmıştır. “Hemşire” gizil değişkeninden “doktor” gizil değişkenine doğru tanımlanacak olan bir ilişki neticesinde, “hemşire” gizil değişkeninin genel “memnuniyet” üzerinde “doktor” gizil değişkeni aracılığıyla oluşan bir dolaylı etkisi daha ortaya çıkacaktır. Bu şekilde dolaylı etkilerin yer aldığı bir model elde edilmek istenirse bu durumda Stata komutuna bu defa (hemsire -> doktor, ) ifadesi eklenir ve model bu hâliyle çalıştırılır.

```
sem (hemsire -> h1,) (hemsire -> h2,) (hemsire -> h3,) (hemsire -> doktor,)
(hemsire -> memnuniyet,) (doktor -> d1,) (doktor -> d2,) (doktor -> d3,)
(doktor -> d5,) (doktor -> memnuniyet,) (memnuniyet -> m1,) (memnuniyet -> m2,)
(memnuniyet -> m3,), method(mlmv) latent(hemsire doktor memnuniyet) nocapslatent
```

Şekil 18.12. Dolaylı Etkinin Yer Aldığı Yapısal Modelin Standartlaştırılmış Sonuçları



Elde edilen modelde dikkat edilirse “hemşire” değişkeninden hem “memnuniyet” değişkenine hem de “doktor” değişkenine bir ok (tek yönlü bir ilişki) çıkmaktadır. Aynı zamanda “doktor” değişkeninden “memnuniyet” değişkenine de bir ok bulunduğu için “hemşire” değişkeninin “doktor” aracılığı ile “memnuniyet” üzerinde dolaylı bir etkisi ortaya çıkmıştır. 18.1.3. başlığı altında ele alınan ve memnuniyet üzerinde var olan bu doğrudan, dolaylı ve toplam etkileri ayırtmak için kullanılacak Stata komut yapısı ve sonuçları aşağıda yer alan Tablo 18.14 ile verilmektedir. Sonuçlar içerisinde yer alan doğrudan etkiler model sonuçlarında yer alan katsayılarla



eşit olmaktadır, toplam etkiler ise doğrudan ve dolaylı etkilerin toplamından oluşmaktadır. Hemşire gizil değişkeninin genel memnuniyet üzerinde doktor gizil değişkeni aracılığıyla olan dolaylı etkisinin (0.48) istatistiksel olarak anlamlı olduğu görülmektedir. Oysa doğrudan etkisinin anlamlı olmadığı yine model sonuçlarından görülmektedir.

**Tablo 18.14.** Örnek 2 İçin Nihai Yapısal Modele İlişkin Etki Ayırıştırması

**estat teffects**

**Direct effects**

|               |  | OIM      |           |      |       |                      |
|---------------|--|----------|-----------|------|-------|----------------------|
|               |  | Coef.    | Std. Err. | z    | P> z  | [95% Conf. Interval] |
| Structural    |  |          |           |      |       |                      |
| doktor <-     |  |          |           |      |       |                      |
| hemsire       |  | .4776119 | .0731351  | 6.53 | 0.000 | .3342697 .620954     |
| memnuniyet <- |  |          |           |      |       |                      |
| doktor        |  | .5898073 | .1094897  | 5.39 | 0.000 | .3752114 .8044033    |
| hemsire       |  | .094774  | .084846   | 1.12 | 0.264 | -.071521 .2610691    |

**Indirect effects**

|               |  | OIM         |           |      |       |                      |
|---------------|--|-------------|-----------|------|-------|----------------------|
|               |  | Coef.       | Std. Err. | z    | P> z  | [95% Conf. Interval] |
| memnuniyet <- |  |             |           |      |       |                      |
| doktor        |  | 0 (no path) |           |      |       |                      |
| hemsire       |  | .281699     | .0650442  | 4.33 | 0.000 | .1542147 .4091833    |

**Total effects**

|               |  | OIM      |           |      |       |                      |
|---------------|--|----------|-----------|------|-------|----------------------|
|               |  | Coef.    | Std. Err. | z    | P> z  | [95% Conf. Interval] |
| Structural    |  |          |           |      |       |                      |
| doktor <-     |  |          |           |      |       |                      |
| hemsire       |  | .4776119 | .0731351  | 6.53 | 0.000 | .3342697 .620954     |
| memnuniyet <- |  |          |           |      |       |                      |
| doktor        |  | .5898073 | .1094897  | 5.39 | 0.000 | .3752114 .8044033    |
| hemsire       |  | .376473  | .0917468  | 4.10 | 0.000 | .1966525 .5562935    |

## 18.7. ÖZET

Bu bölümde sosyal bilimlerde son yıllarda sıklıkla kullanılan Yapısal Eşitlik Modelleri tanıtılmaya ve Stata ile çözümünün anlatılmasına çalışılmıştır. YEM, kullanım yerlerinin çeşitliliğine karşın çözümünün karmaşıklığı nedeniyle kimi zaman korkup kaçılan kimi zaman ise istatistiksel alt yapısı dikkate alınmadan analizlerinin yapıp sapmalı sonuçların elde edildiği birçok de-

ğişkenli yöntemdir. Bu karmaşıklığına karşın model kurma aşamalarının titizlikle takip edilmesi, gerekli varsayım ve ölçüt değerlerinin kontrol edilmesi neticesinde oldukça kullanışlı bilgiler sunmaktadır. Üstelik bunu gerçekleştirirken karmaşık ilişkileri bir bütün hâlinde, eşanlı olarak analiz etmeyi ve tahminlemeyi sağlamaktadır.

YEM, varsayım gereklilikleri dikkate alındığında geniş bir yelpazede tahmin yöntemlerinin kullanılabilceği bir analizdir. En Çok Olabilirlik Yöntemi her ne kadar en sık kullanılan tahmin yöntemi olsa da varsayımlara ve ölçek yapısına bağlı olarak Diyagonal Ağırlıklandırılmış En Küçük Kareler Yöntemi gibi yöntemlere de başvurulmaktadır. Günümüzde pek çok paket programda uygulanabilen YEM için alternatif tahmin yöntemlerinin en fazla yer aldığı paket program Lisrel'dir. Ancak Lisrel'in kullanımının zorluğu nedeniyle özellikle Amos programı daha tercih edilir hâle gelmiştir. Son yıllarda Stata içerisinde de YEM'e yer verilmesi nedeniyle Stata programı kullanıcılar için iyi bir alternatif hâline gelmektedir.

### Stata Komutları

*İlk örnek için kullanılan komutlar:*

```
sem (gizil1 -> a1,) (gizil1 -> a2,) (gizil1 -> a3,) (gizil1 -> a4,) (gizil1 ->
a5,) (gizil1 -> a6,) (gizil2 -> b1,) (gizil2 -> b2,) (gizil2 -> b3,) (gizil2 ->
b4,) (gizil2 -> b5,), covstruct(_lexogenous, diagonal) latent(gizil1 gizil2)
ov(gizil2*gizil1) nocapslatent

sem (gizil1 -> a1 a2 a3 a4 a5 a6) (gizil2 -> b1 b2 b3 b4 b5), covst-
ruct(_lexogenous, diagonal) latent(gizil1 gizil2) cov(gizil2*gizil1) nocapslatent

sem (a1 a2 a3 a4 a5 a6 <-gizil1) (b1 b2 b3 b4 b5 <- gizil2), covstruct(_lexogenous,
diagonal) latent(gizil1 gizil2) cov(gizil2*gizil1) nocapslatent

sem (a1 a2 a3 a4 a5 a6 <-gizil1) (b1 b2 b3 b4 b5 <- gizil2), vce(sbentler)
covstruct(_lexogenous, diagonal) latent(gizil1 gizil2) cov(gizil2*gizil1)
nocapslatent

estat gof, stats(all)

estat mindices

sem (gizil1 -> a1,) (gizil1 -> a2,) (gizil1 -> a3,) (gizil1 -> a4,) (gizil1 ->
a5,) (gizil1 -> a6,) (gizil1 -> gizil2,) (gizil2 -> b1,) (gizil2 -> b2,) (gizil2
-> b3,) (gizil2 -> b4,) (gizil2 -> b5,), latent(gizil1 gizil2) nocapslatent
```

*İkinci örnek için kullanılan komutlar:*

```
sem (hemsire -> h1,) (hemsire -> h2,) (hemsire -> h3,) (hemsire -> h4,) (hemsire
-> memnuniyet,) (doktor -> d1,) (doktor -> d2,) (doktor -> d3,) (doktor -> d4,)
(doktor -> d5,) (doktor -> memnuniyet,) (memnuniyet -> m1,) (memnuniyet -> m2,)
(memnuniyet -> m3,), covstruct(_lexogenous, diagonal) method(mlmv) latent(hemsire
doktor memnuniyet) nocapslatent

estat gof, stats(all)

estat mindices
```

```

sem (hemsire -> h1,) (hemsire -> h2,) (hemsire -> h3,) (hemsire -> memnuniyet,)
(doktor -> d1,) (doktor -> d2,) (doktor -> d3,) (doktor -> d5,) (doktor ->
memnuniyet,) (memnuniyet -> m1,) (memnuniyet -> m2,) (memnuniyet -> m3,),
covstruct(_lexogenous, diagonal) method(mlmv) latent(hemsire doktor memnuniyet)
ocapslatent
estat gof, stats(all)

sem (hemsire -> h1,) (hemsire -> h2,) (hemsire -> h3,) (hemsire -> doktor,) (hem-
sire -> memnuniyet,) (doktor -> d1,) (doktor -> d2,) (doktor -> d3,) (doktor ->
d5,) (doktor -> memnuniyet,) (memnuniyet -> m1,) (memnuniyet -> m2,) (memnuniyet
-> m3,), method(mlmv) latent(hemsire doktor memnuniyet) nocapslatent

estat teffects

```

## 18.8. KAYNAKLAR

- Aitken, A. M. (1935). Acute Iodine Burn. *The British Medical Journal*, 358-358.
- Bollen, K.A. (1989). *Structural Equations with Latent Variables*. Wiley.
- Browne, M.W. (1982). *Covariance Structures. Topics in Multivariate Analysis* (72- 141). Cambridge: Cambridge University Pres.
- Brown, T.A. (2006). *Confirmatory Factor Analysis for Applied Research*. New York : The Guilford Press.
- Çelik, H. E., & Yılmaz, V. (2013). Yapısal eşitlik modellemesi temel kavramlar uygulamalar programlama. *Ankara: Anı Yayıncılık*.
- Hair, F.J., Anderson, E. R.,Tatham,L.R. ve Black, C.W. (1998). *Multivariate Data Analysis*. New Jersey: Prentice Hall.
- Jöreskog, K. (1973). G.(1973). A general method for estimating a linear structural equation system. *AS Goldberger & OD Duncan (Eds.). Structural equation models in the social sciences*, 85-112.
- Jöreskog, K. G., & Goldberger, A. S. (1972). Factor analysis by generalized least squares. *Psychometrika*, 37(3), 243-260.
- Jöreskog, K.G. ve Sörbom, D. (1993). *LISREL 8: Structural Equation Modeling with The SIMPLIS Command Language*. Chicago: Lawrence Earlbaum Ass.
- Jöreskog, K.G. ve Sörbom, D. (1996). *LISREL 8: Users Reference Guide*. Scientific Software, Inc.
- Kaplan, D. (2000). *Structural Equation Modeling: Foundation and Extensions*. Sage Publications.
- Koufteros, X. A. (1999). Testing a model of pull production: a paradigm for manufacturing research using structural equation modeling. *Journal of operations Management*, 17(4), 467-488.
- Long, J. S. (1983). *Confirmatory Factor Analysis: A Preface to LISREL*. California: Sage Publications.
- Mulaik, S. A. (2009). *Linear causal modeling with structural equations*. CRC press.
- Raykov, T., & Marcoulides, G. A. (2006). *A first course in structural equation modeling* (2nd ed.). Lawrence Erlbaum Associates Publishers.
- Satorra, A., & Bentler, P. M. (1994). Corrections to test statistics and standard errors in covariance structure analysis.
- Şimşek, G.G. (2007). *Latent Değişkenli Yapısal Denklemlerine İlişkin Bir Uygulama*. Yayınlanmamış Doktora Tezi. İstanbul : Marmara Üniversitesi Sosyal Bilimler Enstitüsü.
- Tabachnick, B.G. ve Fidel, L.S. (1996). *Using Multivariate Statistics*. Prentice Hall.
- Yuan, K. ve Bentler, P.M. (1998). Structural Equation Modeling with Robust Covariance. *Sociological Methodology*, 28(1):363-396.



# DİZİN

---

## A

Akaike Bilgi Kriteri, 64, 130, 189, 200, 365, 390, 397, 402, 497  
Alternatif Spesifik Değişkenler, 163, 168  
Anakütle, 19, 20, 23, 33, 50, 51, 179, 182  
Anderson CC LR Testi, 288  
Araç Değişkenler Yöntemi , 269, 271, 281, 296  
Araç Değişken Uygunluğu, 271  
Aralıklı Sansür , 407  
Ardışık Cevap Modelleri, 141  
Aşırı Belirlenmiş Model, 294, 297  
Aşırı Yayılım, 183, 187, 189, 190, 191, 192, 193, 195, 196, 200, 201, 206, 207, 215, 351, 352, 374, 380, 382, 403  
Aykırı Değerler, 277

## B

Bahis Oranı, 310, 311  
Bant Genişliği, 453, 454, 473, 478, 482, 484, 486, 487, 489, 490, 491, 492  
Başarısızlık Süresi, 405, 406, 423  
Bayes Bilgi Kriteri, 130, 189, 200, 397, 402  
Bayesyen Ekonometri, 316, 318, 322  
Benzersizlik Parametresi, 164, 172  
BHHH Algoritması, 201  
Bir Eğilim Aralığında Mahalanobis Metrik Eşleştirmesi, 454, 467, 468, 469  
Bire-Bir Eşleştirme, 452, 455  
Bire-Çok Eşleştirme, 452  
Birey-Spesifik Değişkenler, 164, 168  
Birikimli Dağılım Fonksiyonu , 57, 58, 105, 332, 340  
Birikimli Fark Oranı, 108  
Birikimli Logit Model, 106, 108, 109, 133, 134  
Birikimli Normal Dağılım Fonksiyonu , 58, 106  
Birikimli Olasılık, 111, 331, 340  
Birikimli Standart Normal Dağılımı , 58, 109  
Birim Etki, 301, 302, 303, 304, 305, 307, 308, 309, 315, 320, 321, 322, 329, 330, 332  
Bootstrap, 339, 360, 373, 379, 388, 396, 401  
Brant Testi, 129, 130, 131, 133, 138  
Breslow-Wilcoxon Testi, 418, 419, 420, 421

Budanmış Normal Dağılım, 220, 221

Budanmış Regresyon Modeli, 222

Bulanık Regresyon Süreksizliği Analizi, 480

## C

Chow Testi

Cox Regresyon, 414, 422, 423, 424, 425, 426, 427, 428

## Ç

Çekirdek Eşleştirmesi, 453

Çift Eşik Modeli, 343, 344, 345, 346, 351

Çok Durumlu Logit Modeli, 315, 316, 317, 318, 319, 320, 321, 322, 323, 324, 325

Çok Durumlu Modeller, 315

Çok Durumlu Nitel Değişken, 161

Çok Durumlu Probit Modeli, 94, 95, 96, 97, 316

Çoklu Tercih Modelleri, 3, 161

## D

Delta Yöntemi, 353

Değişen Varyans, 44, 45, 48, 49, 59, 183, 277, 290

Deneyisel Olmayan Yöntemler, 433, 434, 476

Deneyisel Yöntemler, 433, 455, 498

Dışlanan Açıklayıcı Değişken, 270

Dışsal (Exogenous) Switching Regresyon Modeli, 255

Dikdörtgen Kernel Fonksiyonu, 486, 487, 488, 489, 490, 491, 492, 493, 494, 495, 496

Dirençli Standart Hata, 49, 59, 192, 338, 339, 292, 360, 361, 372, 373, 374, 379, 380, 388, 396

Dirençli Tahminci, 208, 277, 281, 374, 380, 397

Doğrulamalı Faktör Analizi, 508, 509, 510, 511

Doğrusal Araç Değişkenli Regresyon Modeli , 271

Doğrusal En İyi Sapmasız Tahminci , 309, 270

Doğrusal Olmayan Regresyon Modeli, 184, 188, 304, 356, 367, 371, 377, 399

Doğrusal Regresyon Modeli, 43, 50, 63, 179, 182, 184, 187, 304, 269, 271, 281, 296, 364, 367

Double Hurdle Modeli, 343, 344

Durağan Olmayan , 361

Düzeltilme Parametresi, 453, 473

- E**  
 Egzojen Açıklayıcı Değişken , 269  
 Eğilim Aralığı, 453, 454, 455, 456, 464  
 Eğilim Skoru, 433, 450, 452, 453, 454, 455, 459, 460, 475, 476, 498  
 Eksik Belirlenmiş Model, 282, 288, 294  
 Eksik Yayılım , 183, 188, 351, 374  
 En Çok Benzetimli Olabilirlik, 333, 340  
 En Çok Olabilirlik Yöntemi , 51, 58, 59, 105, 107, 110, 117, 124, 165, 183, 189, 194, 303, 304, 318, 322, 325, 330, 331, 340, 364, 368, 382, 505, 506, 507, 508, 513, 515, 523, 525, 536  
 En Küçük Kareler Yöntemi, 45, 270, 440, 480  
 En Yakın Komşu Eşleştirmesi, 452, 453, 455, 456, 458, 461, 462, 465  
 Endojen Açıklayıcı Değişken , 270  
 Epanechnikov Tipi Kernel Fonksiyonu, 453, 486  
 Eş Dağılımı  
 Eş Yayılım , 183, 192, 215, 352, 360, 372, 382  
 Eşanlı Denklem Yanlılığı, 270  
 Eşik Değer, 58, 104, 105, 111, 117, 124, 133, 220, 223, 331, 332, 336, 339, 340, 342, 476, 477, 478, 479, 480, 481, 482, 483, 484, 485, 486, 487, 488, 489, 490, 493, 497  
 Eşik Parametresi  
 Eşit Varyans  
 Eşleştirme Tahmincisi, 452  
 Etki Analizi Yöntemleri, 433, 434  
 Etkilememe Prensibi, 481  
 Etkileşim Terimi, 479
- F**  
 Faktör Değişken  
 Fark Oram, 51, 53, 54, 107, 108, 109, 110, 117, 118, 119, 145, 153, 154, 155, 310, 311, 312, 345, 353  
 Farkların Farkı, 434, 435, 437, 440, 442, 449, 450, 454, 476, 498
- G**  
 Gama Dağılımı, 193  
 Gauss-Hermite Kareleme Yöntemi , 304, 332, 333, 336, 340, 341  
 Gauss-Hermite Tümlüvi  
 Gaussian Tipi Kernel Fonksiyonu, 453  
 Genelleştirilmiş Doğrusal Karma Modeller, 315, 318, 320  
 Genelleştirilmiş Doğrusal Modeller  
 Genelleştirilmiş Logit Modeli, 109, 130, 133, 134, 135, 137  
 Genelleştirilmiş Momentler Metodu, 292, 296  
 Genelleştirilmiş Tahmin Denklemleri  
 GHK Simülasyon Yöntemi, 316  
 Gizli Değişken , 301, 331 501
- Görelî Fark Oranı  
 Görelî Risk Oranı  
 Gözlemlenen Bilgi Matrisi , 338, 339  
 Gradyan vektörü, 201  
 Gumbel Dağılımı, 73  
 Gumbel'in Uç Değerler Dağılımı, 163
- H**  
 Hansen J İstatistiği, 288, 296  
 Hausman Testi, 307, 309  
 Havuzlanmış Model, 348, 377, 399  
 Havuzlanmış Yatay Kesit Veri Seti, 1, 2, 434, 437, 454  
 Heckman Seçim Modeli, 219, 233  
 Hessian matrisi  
 Heterojenlik , 317, 321, 329, 330  
 Hiyerarşik Yanıt Değişkeni, 141  
 Huber White Sandwich Tahmincisi, 338, 277
- I**  
 Incidence Rate Ratio, 186, 197, 204, 353, 370, 385
- İ**  
 İç Geçerlilik, 481  
 İçsel (Endogenous) Switching Regresyon Modeli, 256  
 İçsellik Problemi, 269, 296  
 İki Aşamalı En Küçük Kareler Yöntemi , 269, 279, 282, 296, 505  
 İki Durumlu Bağımlı Değişkenli Model, 63, 301, 302, 304  
 İki Durumlu Nitel Değişken, 301, 304, 310  
 İki Durumlu Sonuç  
 İkili Cevap Değişken  
 İkili Değişken, 43  
 İkili Tercih Modelleri, 161  
 İlişkisiz Alternatiflerin Bağımsızlığı, 74, 75, 76, 77, 81, 82, 95, 162, 315  
 İndirgenmiş Katsayı  
 İndirgenmiş Yapısal Katsayılar, 282  
 İndirgenmiş Yapısal Denklem, 282, 283, 286
- J**  
 Jackknife, 373, 379, 396, 401
- K**  
 Kaliper Eşleştirmesi, 453  
 Kanonik Korelasyon, 288  
 Kaplan-Meier , 415, 416, 417, 418, 428  
 Kapsayıcı Değer, 165, 175  
 Kareleme Noktası , 333, 336, 341  
 Karma Logit Model, 315, 318  
 Karma Modeller, 315, 320  
 Karşıt Gerçeklik, 450, 481

Kategorik Değişken, 4, 43, 71, 103, 109, 110, 111, 116, 166, 329, 355, 359, 408, 424, 427, 439

Kernel Eşleştirmesi, 453, 455, 469, 470, 471, 472, 473, 474

Kesikli Değişim

Kesikli Tercih Modelleri, 161

Keskin Regresyon Süreksizliği Analizi, 480, 482, 490

Kesme Noktası , 331

Kontrol Grubu, 434, 435, 436, 437, 438, 441, 442, 443, 444, 447, 448, 450, 451, 452, 453, 454, 455, 457, 458, 459, 460, 461, 462, 464, 465, 466, 467, 470, 471, 473, 476, 477, 480, 481, 487, 488, 489, 490, 493

Koşullu Bağımsızlık Varsayımı, 451

Koşullu Olasılık, 43, 44, 46, 142, 143, 144, 303, 332

Koşulsuz Olasılık, 331, 332, 340

Köşe Çözümü

Kuadratik Moment Koşulu, 293

Kukla Değişken, 46, 47, 50, 55, 56, 61, 63, 186, 283, 284, 436, 437, 438, 439, 440, 445, 452, 478

Kümelenmiş Dirençli Standart Hatalar

**L**

Lagrange Çarpanı Testi

Logaritmik-Doğrusal Model

Logit Model, 43, 46, 49, 51, 52, 53, 54, 55, 59, 61, 64, 65, 66, 67, 203, 204, 206, 208, 209, 302, 303, 304, 305, 307, 308, 309, 310, 311, 312

Logaritmik Olabilirlik Fonksiyonu, 107, 183, 184, 194, 201, 304

Log-Rank Testi, 418, 419, 420, 421, 428, 429

LR Testi, 52, 117, 124, 130, 189, 191, 196, 307, 308, 377, 378, 384, 396, 399, 401

**M**

Mahalanobis Metrik Eşleştirmesi, 454, 455, 466, 467, 469, 470

Mahalanobis Metrik Uzaklığı , 454, 471, 472, 473, 474

Marj Değişkeni, 477, 478, 479, 484, 485, 486, 487, 490, 492

Marjinal Etki, 51, 53, 55, 56, 59, 60, 62, 110, 111, 117, 119, 125, 137, 144, 145, 153, 156, 157, 184, 185, 186, 197, 198, 205, 209, 210, 305, 310, 311, 336, 337, 338, 342, 344

Markov Zinciri Monte Carlo Analizi, 322

MCMC, 322

Metropolish-Hastings algoritması, 323

Modifikasyon İndeksleri

MSM Simülasyon Yöntemi, 316

Müdahale Görenler Üzerinden Ortalama Müdahale Etkisi, 450

Müdahale Grubu, 434, 435, 436, 438, 441, 442, 448, 450, 451, 452, 453, 454, 457, 458, 461, 464, 473, 476, 477, 480, 485, 487, 488, 489, 490, 492, 493, 497

## N

Negatif Binom Dağılımı, 206

Negatif Binom Regresyon Modeli, 180, 189, 190, 191, 192, 198, 200, 208, 214, 215, 381

Negbin-1 Modeli, 194, 381, 382

Negbin-2 Modeli, 193, 194, 382

Newton-Raphson Yöntemi, 184, 206, 320

Nicel Değişken, 55, 161, 434, 439

Nitel Değişken, 2, 43, 161, 166, 301, 303, 304, 315, 334, 438

Nitel Tercih Modelleri, 148, 161, 348

## O

Olabilirlik Fonksiyonu, 107, 303, 304

Olabilirlik Oran Testi, 52, 59, 307, 308, 336

Olasılık Kütle Fonksiyonu

Olasılık Yoğunluk Fonksiyonu, 57, 412, 413

Orantılı Tehlikeler Varsayımı, 423, 424, 427, 428

Orantısız Oran Varsayımı, 108, 109, 118

Ortak Destek Alanı, 452, 457

Ortalama Müdahale Etkisi, 257, 436, 450, 478, 490

Ortalamaya Göre Marjinal Etkiler, 110, 125, 356

## Ö

Ölçme Hataları, 270, 279

Ölçme Modeli, 348, 382, 503, 515, 517

Örneklem Seçim Yanlılığı, 233, 234, 235, 257, 258

## P

Panel Çok Durumlu Probit Model, 315, 316

Panel Çok Durumlu Logit Model, 315

Panel Negatif Binom, 381

Panel Poisson Model, 347

Panel Sayma Veri, 348

Panel Sıralı Modeller, 329, 330, 331

Paralel Eğilim, 435, 436, 441, 444, 445, 446, 447, 448

Paralellik Varsayımı, 118, 130, 131, 133, 135, 137

Parametrik Olmayan Eşleştirme, 453

Path Katsayısı, 503

Pearson İstatistiği, 187

Poisson Dağılımı, 182, 183, 184, 190, 201, 202

Poisson Regresyon Modeli, 180, 182, 189, 190, 191, 192, 200, 204, 215

Probit Model, 43, 46, 57, 58, 59, 60, 61, 63, 64, 65, 66, 67, 302, 303, 304, 305, 315, 316

Program Öncesi Değerlendirme , 433

Program Sonrası Değerlendirme, 433, 434

**R**

Rassal Etkiler, 301, 302, 303, 304, 307, 309, 316, 318, 320, 329, 330, 340  
 Rassal Etkiler Çok Durumlu Logit Model, 317, 318, 320  
 Rassal Etkiler Çok Durumlu Probit Model, 316  
 Rassal Etkiler İki Durumlu Modeller, 302  
 Rassal Etkiler Logit Modeli, 303, 305, 308, 309, 310, 311  
 Rassal Etkiler Modeli, 301, 302, 303, 304  
 Rassal Etkiler Probit Modeli, 303, 304, 305  
 Rassal Etkiler Sıralı Logit Model, 331, 340, 341, 342, 344  
 Rassal Etkiler Sıralı Probit Model, 331, 332, 333, 335, 336, 337, 338, 339  
 Rassal Seçime Dayalı Kontrollü Deneyler  
 Rastlantısal Parametre Problemi , 330  
 Referans Sınıf, 6, 10, 11, 14, 48, 54, 55, 60  
 Regresyon Süreksizliği Analizi, 433, 476, 477, 479, 480, 481, 482, 495, 497, 498  
 Regresyon Süreksizliği Tasarımı, 476  
 Rejim, 253

**S**

Sabit Etkiler, 301, 304, 305, 307, 329, 330, 335  
 Sabit Etkiler Modeli, 301, 304, 305, 307  
 Sabit Etkiler İki Durumlu Modeller, 304  
 Sabit Etkiler Logit Model, 305, 307, 309  
 Sabit Etkiler Panel Negatif Binom Regresyon Modeli, 392  
 Sabit Etkiler Panel Poisson Regresyon Modeli , 367  
 Sağdan Sansür , 220, 406, 407, 423  
 Sansürlü Normal Dağılım, 222  
 Sapma İstatistiği , 188  
 Sayma Değişkeni, 179, 184, 195, 200, 201, 206  
 Saymalı Modeller, 180, 184, 186, 187, 188, 190, 200, 206, 211, 215  
 Schoenfeld Artıkları , 424, 427, 428  
 Seçim Kriteri, 63, 64, 65, 66  
 Seçim Yanlılığı, 450, 451, 452, 454, 455, 458, 459, 460, 477  
 Shea'nin Kısmi Belirlilik Katsayısı, 292  
 Sıfır Yığılmalı Negatif Binom Regresyon Modeli, 200, 206  
 Sıfır Yığılmalı Poisson Regresyon Modeli, 201  
 Sıfır Yığılmalı Sayma Modelleri, 200, 215  
 Sıralı Logit Modeli, 106, 108, 10, 111, 115, 117, 118, 129, 130  
 Sıralı Probit Modeli, 105, 109, 110, 123, 124, 125, 130  
 SML Simülasyon Yöntemi  
 Soldan Sansür, 407, 408  
 Süre Modelleri, 329  
 Switching Regresyon, 253, 255, 256, 257, 258, 262, 263, 264, 265, 266

**T**

Tam Belirlenmiş Model, 282, 292, 294, 297  
 Tam Bilgiye Dayalı En Çok Olabilirlik Tahmincisi , 165  
 Tehlike Fonksiyonu , 413, 414, 415, 417, 422, 423  
 Tehlike Oranı, 418, 422, 423, 424, 426, 427, 428  
 Temel Sınıf, 10  
 Temel Tehlike Fonksiyonu , 422  
 Ters Mills Oranı, 223, 235  
 Tesadüfî Fayda Teorisi, 72, 94, 163  
 Tobit Modeli, 219, 223, 225, 226, 227, 233, 235

**U**

Uyum İyiliği , 63, 117, 187, 188, 189, 200, 374, 380, 501, 521, 523, 525, 527, 528, 532, 533  
 Uyum İyiliği İstatistikleri, 187, 214, 374, 390, 391, 397

**Ü**

Üçgen Kernel Fonksiyonu, 486

**V**

Vuong Testi, 202, 203, 204, 206, 207, 208, 215

**W**

Wald Testi, 129, 130, 135, 137, 191, 192, 196, 308

**Y**

Yakınsama testi, 323  
 Yapısal Denklem, 281, 282, 283, 503  
 Yapısal Eşitlik Modeli, 501, 502, 503, 504, 505, 506, 507, 512, 520, 521, 535, 536  
 Yapısal Model, 502, 503, 504, 521, 523, 524, 525, 527, 528, 531, 532, 533  
 Yarı DeneySEL Yöntem, 498  
 Yarıçap Eşleştirmesi, 453, 455, 464, 465, 466, 469, 470  
 Yaşam Süresi , 405, 406, 407, 408, 409, 411, 412, 414, 415, 416, 420, 422, 423, 424, 428  
 Yaşam Fonksiyonu, 412, 413, 414, 416, 417, 418, 422  
 Yaşam Tablosu , 414  
 Yayılım Parametresi , 183, 193, 365, 384, 399, 403  
 Yerel Ortalama Müdahale Etkisi, 478, 490  
 Yerine Koyma Yoluyla En Yakın Komşu Eşleştirmesi, 452, 453, 461, 462, 463, 464, 465  
 Yerine Koymama Yoluyla En Yakın Komşu Eşleştirmesi, 452, 453, 455, 456, 458, 459, 460, 461, 462, 463  
 Yoğunlaştırılmış Olabilirlik Fonksiyonu , 368  
 Yuvalanmış Logit Modeli, 3, 162, 163, 164, 165, 166, 167, 169, 170, 171, 172, 173, 175, 177

**Z**

Zorlayıcı Değişken, 477



# Uygulamalarla MİKRO EKONOMETRİ

EDİTÖRLER

PROF. DR. ŞENAY ÜÇDOĞRUK BİRECİKLİ - PROF. DR. SEDA ŞENGÜL

Mikroekonometrik yöntemler; birçok alanda yatay kesit ve panel veri setlerinin yaygın olarak toplanması, ekonometrik yöntemlerin standart yazılım paket programlarının bir parçası hâline gelmesi ve bu yöntemlerin kolay programlanmasını sağlayan yeni yazılımların geliştirilmiş olması nedeniyle hızla gelişmektedir.

*Uygulamalarla Mikroekonometri* kitabı, kapsadığı 18 ana bölümle mikroekonometrinin hemen hemen tüm güncel konularının bir araya getirildiği, Türkçede yayınlanmış oldukça kapsamlı bir kitaptır. Kitapta; mikroekonometrik yöntemlerin önce teorisi, kısa, sezgisel ve uygulamaya dayalı anlatılmakta, sonra gerçek veri setleri ile Stata'da uygulamalar yapılarak yöntemlerin uygulama yoluyla kolayca anlaşılmasına çalışılmaktadır. Kitaptan yararlanan araştırmacılar ve öğrenciler, her bölümde kullanılan veri seti ve program dosyalarına (do dosyaları) önsöz sonunda verilen karekodla erişebilecek, kitaptaki tüm uygulamaları yapabileceklerdir.

*Uygulamalarla Mikroekonometri* kitabı; ekonometri, ekonomi, işletme, psikoloji, sağlık, eğitim vb. tüm sosyal bilim dallarındaki araştırmacıların, öğrencilerin hem araştırmalarında hem de lisans ve lisansüstü derslerinde yararlanabilecekleri bir kaynak kitaptır.



AKADEMİK YAYINCILIK

[www.nobelyayin.com](http://www.nobelyayin.com)

nobelyayin | nobelkitap | nobelcocuk | nobelyasam | nobelcocukyayin



Kapak Tasarım Aslı Boz



ISBN 978-625-439-171-2

