

T.C.
MARMARA ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ
EKONOMETRİ ANABİLİM DALI
EKONOMETRİ BİLİM DALI

**İÇ GÖÇ SONRASI BİREYLERİN İŞGÜCÜNE KATILIMINI ETKİLEYEN
FAKTÖRLERİN SEMİPARAMETRİK LOGİT MODELLERİ İLE
İNCELENMESİ**

Yüksek Lisans Tezi

Mustafa Saygın ARAZ

İstanbul,2018

T.C.
MARMARA ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ
EKONOMETRİ ANABİLİM DALI
EKONOMETRİ BİLİM DALI

**İÇ GÖÇ SONRASI BİREYLERİN İŞGÜCÜNE KATILIMINI ETKİLEYEN
FAKTÖRLERİN SEMİPARAMETRİK LOGİT MODELLERİ İLE
İNCELENMESİ**

Yüksek Lisans Tezi

Mustafa Saygın ARAZ

Danışmanı: Prof. Dr. Selahattin GÜRİŞ

İstanbul,2018



T.C.
MARMARA ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ MÜDÜRLÜĞÜ

TEZ ONAY BELGESİ

EKONOMETRİ Anabilim Dalı EKONOMETRİ Bilim Dalı TEZLİ YÜKSEK LİSANS öğrencisi MUSTAFA SAYGIN ARAZ'ın İÇ GÖÇ SONRASI BİREYLERİN İŞGÜCÜNE KATILIMINI ETKİLEYEN FAKTÖRLERİN SEMİPARAMETRİK LOGİT MODELLERİ İLE İNCELENMESİ adlı tez çalışması, Enstitümüz Yönetim Kurulunun 5.07.2018 tarih ve 2018-18/31 sayılı kararıyla oluşturulan jüri tarafından oy birliği / oy çokluğu ile Yüksek Lisans Tezi olarak kabul edilmiştir.

Tez Savunma Tarihi 02.08.2018

Öğretim Üyesi Adı Soyadı

İmzası

Öğretim Üyesi Adı Soyadı	İmzası
1. Tez Danışmanı Prof. Dr. SELAHATTİN GÜRİŞ	
2. Jüri Üyesi Dr. Öğr. Üyesi TURGUT ÜN	
3. Jüri Üyesi Dr. Öğr. Üyesi ÖZLEM GÖKTAŞ	

GENEL BİLGİLER

İsim ve Soyadı : Mustafa Saygın ARAZ
Anabilim Dalı : Ekonometri
Programı : Ekonometri
Tez Danışmanı : Prof. Dr. Selahattin GÜRİŞ
Tez Türü ve Tarihi : Yüksek Lisans – Ağustos 2018
Anahtar Kelimeler : İç göç, İşgücü, Semiparametrik Logit Modeli, Nonparametrik Logit Modeli

ÖZET

İÇ GÖÇ SONRASI BİREYLERİN İŞGÜCÜNE KATILIMINI ETKİLEYEN FAKTÖRLERİN SEMİPARAMETRİK LOGİT MODELLERİ İLE İNCELENMESİ

İç göç konusu Türkiye için geçmişte olduğu gibi günümüzde de oldukça öneme sahip bir konudur. Türkiye'nin Cumhuriyet tarihi boyunca değişen ekonomik ve siyasi koşullar bireylerin ülke içinde bir yerden başka bir yere göç etmesine neden olmaktadır. Göç eden bireylerin mekânsal hareketliliği işgücü piyasasını da etkilemektedir. Göçün işgücü piyasası üzerindeki etkisinin incelenmesi göç eden bireylerin işgücüne katılım biçimlerinin tespit edilmesi ile mümkündür. Bireylerin işgücüne katılmasını etkileyen bir dizi faktör bulunmaktadır. Bu çalışmanın amacı 2011 yılında Türkiye'de iç göç sürecine katılan bireylerin işgücüne katılımını etkileyen faktörleri tespit etmektir. Araştırmada Türkiye İstatistik Kurumu (TÜİK)'e ait 2011 yılı Nüfus ve Konut Araştırması (NKA) mikro verisinden yararlanılmıştır. Ekonometrik analiz yöntemi olarak parametrik ve semiparametrik logit modelleri kullanılmış ve bu modeller arasında karşılaştırma yapılmıştır. Araştırma sonucuna göre semiparametrik logit modelinin bireylerin işgücüne katılmasını etkileyen faktörleri parametrik logit modeline göre daha iyi açıkladığı görülmüştür.

GENERAL KNOWLEDGE

Name and Surname : Mustafa Saygın ARAZ
Field : Econometrics
Programme : Econometrics
Supervisor : Prof. Dr. Selahattin GÜRİŞ
Degree Awarded and Date : Master – August 2018
Keywords : Internal Migration, Labour Force, Semiparametric Logit Model, Nonparametrik Logit Model

ABSTRACT

INVESTIGATION OF THE FACTORS AFFECTING LABOR PARTICIPATION AFTER INTERNAL MIGRATION WITH THE SEMIPARAMETRIC LOGIT MODELS

The internal migration issue is also an issue having great importance nowadays as it was in the past. The changing economic and political conditions through the history of the Republic of Turkey internally cause the individuals to migrate to some place from some place. The spatial mobility of migrated individuals also affects the labor market. Researching the effect of migration on the labor market is possible by determining the participation features of individuals to the labor market. There are a series of factors affecting the participation of individuals to the labor market. This study aims to determine the factors affecting the participation of individuals participated in the internal migration process to the labor market in Turkey in 2011. The year 2011 Population and Housing Research (PHR) micro data belonging to the Turkish Statistical Institute (TSI) was benefited from the research. The parametric and semiparametric logit models were used as the econometric analysis method, and a comparison was made between those models. It was seen that the semiparametric logit model clarified the factors affecting the participation of individuals to the labor better than the parametric logit model as a result of the research.

İÇİNDEKİLER

Sayfa No.

TABLO LİSTESİ	V
ŞEKİL LİSTESİ	VI
KISALTMALAR	VII
GİRİŞ	1
1.PARAMETRİK LOGİT MODELİ	3
1.1 Parametrik Logit Modeli'nin Tanımı ve Yapısı.....	3
1.2 Parametrik Logit Modeli'ne İlişkin Tahmin Yöntemi.....	5
1.3 Parametrik Logit Modeli'nin Katsayılarının Anlamlılığı.....	6
1.4 Parametrik Logit Modeli'nin Uyum İyiliği.....	8
1.5 Parametrik Logit Modeli'nin Katsayı Yorumu.....	12
2.NONPARAMETRİK LOGİT MODELİ	14
2.1 Nonparametrik Regresyonda Düzgünleştirme Kavramı ve Düzgünleştirici.....	14
2.1.1 Kernel (Çekirdek) Tahminci.....	15
2.1.2 Local Linear (Yerel Doğrusal) Tahminci.....	17
2.1.3 Smoothing Spline Tahminci.....	18
2.1.4 Local Polynomial Tahminci.....	21
2.2 Nonparametrik Regresyon'da Düzgünleştirme Parametresinin Seçimi.....	23
2.3 Nonparametrik Logit Modeli Tanımı ve Yapısı.....	26
2.4 Nonparametrik Logit Modeli'ne İlişkin Tahmin Yöntemi.....	27
2.5 Nonparametrik Logit Modeli'nin Katsayı Yorumu.....	30
3.SEMİPARAMETRİK LOGİT MODELİ	30
3.1 Semiparametrik Logit Modeli'nin Tanımı ve Yapısı.....	30
3.2 Semiparametrik Logit Modeli'nin Tahmini.....	33
3.2.1 Profil Benzerlik Tahmin Yöntemi.....	33
3.2.2 Semiparametrik Maksimum Benzerlik Tahmini.....	35
3.2.3 Genelleştirilmiş Speackman Tahmini.....	37
3.2.4 Backfitting Tahmin Yöntemi.....	38
3.2.5 Maksimum Score Tahmini.....	40
3.3 Semiparametrik Logit Modeli'nin Anlamlılık Testi.....	41
3.3.1 Benzerlik Oranı Testi.....	42
3.3.2 Kısıtlı F-Testi.....	43
3.3.3 Bootstrap İle Doğrusal Olmama Testi.....	44
3.4 Semiparametrik Logit Modeli'nin Uyum İyiliği.....	45
4. UYGULAMA: İÇ GÖÇ SONRASI BİREYLERİN İŞGÜCÜNE KATILIMINI ETKİLEYEN FAKTÖRLER	49
4.1 Araştırmanın Konusu ve Amacı.....	49
4.2 Araştırmanın Kapsamı ve Kısıtları.....	49
4.3 Araştırmanın Yöntemi.....	50
4.4 Araştırmada Kullanılan Verinin Yapısı.....	50

4.5 Verilerin Analizi.....	51
4.5.1 Demografik Bilgilere İlişkin Frekanslar	51
4.5.2 Göç Etme Nedenine İlişkin Frekanslar	52
4.5 Göçmenlerin İşgücüne Katılımını Etkileyen Faktörlerin Logit Modeli İle Analizi: Parametrik ve Semiparametrik Logit Modeller İçin Uygulama	52
SONUÇ	68
KAYNAKÇA.....	72
EKLER	76

TABLO LİSTESİ

Sayfa No.

TABLO 1 : Logit Modeli için Gerçek ve Tahmini Değerlere Ait Çapraz Tablo	10
TABLO 2 : Kernel Tablosu	17
TABLO 3 : Genelleştirilmiş Speckman Tahmini İçin İterasyon Adımları.....	38
TABLO 4 : Backfitting Tahmini İçin İterasyon Adımları	39
TABLO 5 : Parametrik Logit Modeli'ne İlişkin Tahmin Sonuçları	54
TABLO 6 : Genel Model İçin Uyum Kriterleri	55
TABLO 7 : Semiparametrik Logit Modeli'ne İlişkin Tahmin Sonuçları	56
TABLO 8 : Kadın İçin Parametrik Logit Modeli'ne İlişkin Tahmin Sonuçları	60
TABLO 9 : Kadına Ait Model İçin Uyum Kriterleri	60
TABLO 10 : Kadın İçin Semiparametrik Logit Modeli'ne İlişkin Tahmin Sonuçları	61
TABLO 11 : Erkek İçin Parametrik Logit Modeli'ne İlişkin Tahmin Sonuçları	64
TABLO 12 : Erkeğe Ait Model İçin Uyum Kriterleri	65
TABLO 13 : Erkek İçin Semiparametrik Logit Modeli'ne İlişkin Tahmin Sonuçları ..	58

ŞEKİL LİSTESİ

Sayfa No.

- ŞEKİL 1 : Genel Model İçin Eğitim ve Yaş'a Ait Düzgünleştirme Fonksiyonu 58
ŞEKİL 2 : Kadın İçin Eğitim ve Yaş'a Ait Düzgünleştirme Fonksiyonu 63
ŞEKİL 3 : Erkek İçin Eğitim ve Yaş'a Ait Düzgünleştirme Fonksiyonu..... 67

KISALTMALAR

DOM	Doğrusal Olasılık Modeli
EKK	En Küçük Kareler
LR	Benzerlik Oranı Testi
LM	Langrange Çarpanı Testi
NW	Nadaraya-Watson Tahmincisi
LL	Yerel Doğrusal Tahminci
RSS	Kalıntı Kareleri Toplamı
MSE	Ortalama Hata Kare
IMSE	Bütünleşik Ortalama Hata Kare
MSFE	Tahmini Ortalama Hata Kare
CV	Çapraz Geçerlilik
GCV	Genelleştirilmiş Çapraz Geçerlilik
GLM	Genelleştirilmiş Doğrusal Model
GAM	Genelleştirilmiş Toplamsal Model
IRLS	Yeniden Ağırlıklandırılmış En Küçük Kareler
TÜİK	Türkiye İstatistik Kurumu
NKA	Nüfus ve Konut Araştırması
SİM	Tek Endeks / İndeks Modeli

GİRİŞ

Her ne sebeple gerçekleşirse gerçekleşsin göç hareketleri toplumun sosyal ve ekonomik yapısında değişimlere neden olmaktadır. Bu değişimler ekonominin önemli alt yapı unsurlarından biri olan işgücü piyasasını da etkilemektedir. Göç olgusunu anlamaya ve etkilerini açıklamaya çalışırken göçü, göç eden bireyden ayrı düşünemeyeceğimiz gibi emek gücü de bireyin ayrılmaz bir parçasıdır. Dolayısıyla göçün işgücü piyasasına etkisinin anlaşılması göç eden bireylerin bu piyasadaki konumunun- işgücünde olup olmadığı, işgücünde olanların ise işsiz mi yoksa istihdamda mı olduğu- tespit edilmesi ile söz konusu olmaktadır. Göç eden bireylerin işgücüne katılmasını etkileyen/engelleyen bir dizi faktör bulunmaktadır. Bu faktörler bireye ait özellikler (cinsiyet, eğitim, yaş vb.) olabileceği gibi göç edilen yere ait özellikler de olabilmektedir. Bu çalışma; göç eden bireylerin işgücüne katılmasını etkileyen faktörleri, yani göçün işgücü piyasası üzerindeki etkisini incelemeyi amaçlamaktadır.

Araştırmada ekonometrik analiz yöntemi olarak nitel tercih modellerinden biri olan logit modeli kullanılmıştır. Bağımlı değişkenin nitel olarak tanımlandığı bu modeller, bağımlı değişkenin aldığı değere bağlı olarak ikili ya da çoklu yapıda olabilmektedir. Araştırma konusu itibariyle göç eden bireylerin işgücünde olup olmadığı inceleneceğinden bu çalışmada ikili tercih yapısı kullanılarak parametrik ve semiparametrik logit model tahminleri karşılaştırılmıştır.

Parametrik logit modelinde bağımlı değişken ve bağımsız değişkenler arasındaki ilişkinin varlığı belirli bir matematiksel forma bağlıdır ve tahmin edilen katsayılar doğrusal ilişkiyi varsayar. Semiparametrik logit modelde ise bazı değişkenler modele parametrik olmayan (nonparametrik) bileşen halinde eklenebilmektedir. Değişkenlerin yalnızca parametrik olarak modele dahil edilmesi kimi zaman göçün çok boyutlu yapısının anlaşılmasını güçleştirmektedir. Çünkü göç konusunun incelenmesi ekonomik nedenlerin dışında sosyolojik ve siyasal faktörleri de göz önünde bulundurmayı gerektirmektedir. Bu nedenle parametrik yöntemler göçün etkilerinin incelenmesinde önemli ölçüde bilgi sağlamakla birlikte bu noktaları göz ardı ettiğinden elde edilen sonuçlar konunun tam olarak anlaşılmasında yetersiz kalmaktadır. Bu

durumun sonucu olarak gerçekte parametrik olmayan ilişkinin modele parametrik olarak yansıtılmasından kaynaklı tanımlama hatası ile karşılaşmaktadır. Semiparametrik logit modeli ise parametrik ve parametrik olmayan ilişkileri birlikte modellemeye olanak sağladığından bahsi geçen sakıncaların ortadan kaldırılması için alternatif bir yöntem olarak karşımıza çıkmaktadır.

Tez çalışması dört bölümden oluşmaktadır. İlk üç bölümde parametrik, nonparametrik ve semiparametrik logit modeli tanıtılmış ve bu modeller için tahmin, çıkarsama ve sonuçların yorumlanmasına ilişkin açıklamalarda bulunulmuştur. Buna ilave olarak birinci bölümde logit modellerinin doğasının anlaşılması ve sonuçların açıklanmasında önemli rol oynayan logit, odds oranı ve marjinal etkilerden bahsedilmiştir.

İkinci bölümde parametrik olmayan yönteme ait kavramlar incelenmiştir. Bu bölümde parametrik olmayan tahminler için düzgünleştirme kavramı ve düzgünleştiriciler ele alınmış parametrik olmayan tahminlerde sıkça kullanılan Kernel Fonksiyonları tanıtılmıştır. Kernel Fonksiyonlarında kullanılan düzgünleştirme parametresinin seçimine ilişkin açıklamalara da bu bölümde yer verilmiştir.

Üçüncü bölüm parametrik ve parametrik olmayan yöntemlerin birleştirildiği semiparametrik yönteme ilişkin açıklamaları içermektedir.

Dördüncü bölüm göç eden bireylerin işgücüne katılmasını etkileyen faktörlerin tespit edilmesi için parametrik ve semiparametrik logit modeli tahminlerinin karşılaştırıldığı uygulama bölümüdür. Sonuç kısmı ise araştırma bulgularından hareketle konuya ait tartışmanın ve önerilerin sunulduğu kısımdır.

1. PARAMETRİK LOGİT MODELİ

1.1 PARAMETRİK LOGİT MODELİ'NİN TANIMI VE YAPISI

Ekonometrik regresyon modellerinde bağımlı değişken her zaman nicel diğer bir deyişle sürekli bir yapıya sahip olmayabilir. Kimi durumlarda bağımlı değişken nitel yapıdadır. Örneğin “Evet” ve “Hayır” gibi ifade edilebilir. Bu durumda bağımlı değişken $Y_i = 1$ ve $Y_i = 0$ şeklinde sayısallaştırılarak regresyon analizleri yapılmaktadır. Bu tarz ikili değer alan regresyonlar genel itibariyle ikili tepki modelleri olarak adlandırılmaktadır. Regresyon analizlerinde amaç X_i veriyken Y_i 'nin koşullu beklenen değerini $E(Y_i/X_i)$ tahmin etmektir. İkili tepki modelleri için ise bu durum bağımlı değişkenin iki seçenekli olması -sıfır ve bir değerini alması- nedeniyle X_i veriyken bir olayın gerçekleşmesinin koşullu olasılığı $P(Y_i = 1/X_i)$ şeklinde yorumlanabilir¹.

Bilinen en temel ikili tepki modeli Doğrusal Olasılık Modeli (DOM)'dir. Bu modelde bağımlı değişken Y_i 'nin koşullu beklenen değeri modelde yer alan bağımsız değişkenlerin X_i doğrusal bir fonksiyonu ile açıklanmaktadır. DOM yapısı gereği doğrusal bir fonksiyona sahiptir ve modele ait parametreler (β_i) de olayın gerçekleşmesinin koşullu olasılığı ile doğrusal ilişki içerisindedir. Bu da bir dizi tahmin sorununa yol açmaktadır. Bu sorunlardan ilki bağımlı değişkenin tahmini değerinin yani $P(Y_i = 1/X_i)$ olasılığının bir çok durumda 0 ve 1 aralığının dışına çıkmasıdır². İkinci sorun ise olasılığın X_i 'e doğrusal bağlı olarak artışı nedeniyle X_i 'in marjinal etkisinin sabit olduğunun varsayılmasıdır. Sonuç olarak DOM bazı durumlarda dağılım sorunu nedeni ile doğru sonuç vermez ve genellikle bu model yerine logit veya probit modelleri kullanılır. Sözü edilen sorunların giderilmesi için logit ve probit modeller olasılık için belirli farklı fonksiyonlara sahiptirlerdir. Logit model lojistik birikimli dağılım esasına dayanır. Lojistik birikimli dağılım fonksiyonu;

¹ Damodar N. Gujarati, **Basic Econometrics**, Fourth Edition, New York: McGraw-Hill, 2003, s.583.

² Gangadharrao Soundaryarao Maddala, **Introduction to Econometrics**, Second Edition, New York: Macmillan Publishing Company, 1992, s.324.

$$P_i = E(Y_i = 1/X_i) = \frac{1}{1 + e^{-Z_i}}$$

olarak ifade edilir. Burada P_i bir olayın gerçekleşme veya belirli bir tercihin seçilme olasılığını, $Z_i = \beta_1 + \beta_2 X_i$ ise modelde yer alan sabit terim ve açıklayıcı değişkenlere ilişkin katsayıları temsil eder.

Lojistik birikimli dağılım fonksiyonundan anlaşılacağı gibi X_i 'ler ile olasılık değeri arasında doğrusal olmayan bir ilişki bulunmaktadır. X_i azalırken olasılık değeri 0'a, X_i arttıkça değerler 1'e daha yavaş şekilde yaklaşır. Bu durumda Z_i , $-\infty$ ve $+\infty$ arasında değer alırken olasılıklar da $0 \leq P(Y_i = 1/X_i) \leq 1$ koşulunu sağlamış olur³.

P_i , olayın gerçekleşme olasılığı veri iken gerçekleşmeme olasılığı $1 - P_i$ ise şu şekilde tanımlanır;

$$1 - P_i = \frac{e^{-Z_i}}{1 + e^{-Z_i}} = \frac{1}{1 + e^{Z_i}}$$

Bu iki değer birbirine oranlanması ile bahis oranı şu şekilde ifade edilir;

$$\frac{P_i}{1 - P_i} = \frac{1 + e^{Z_i}}{1 + e^{-Z_i}} = e^{Z_i}$$

Bahis oranının logaritması alındığında Logit model yapısı,

$$L_i = \ln\left(\frac{P_i}{1 - P_i}\right) = Z_i = \beta_1 + \beta_2 X_i + u_i$$

olacaktır. Burada L_i logaritmik bahis oranı, diğer bir ifade ile Logit olarak adlandırılır. Yukarıdaki modelde görüldüğü gibi modele ait parametreler Logit ile doğrusal ilişki içerisindeyken olayın gerçekleşme olasılığı P_i ile aralarında doğrusal olmayan bir ilişki söz konusudur.

³ Selahattin Gürış ve Ebru Çağlayan, **Ekonometri: Temel Kavramlar**, 3. Basım, İstanbul: Der Yayınları, 2010, s.667.

1.2 PARAMETRİK LOGİT MODELİ'NE İLİŞKİN TAHMİN YÖNTEMİ

Klasik En Küçük Kareler (EKK) tahmini açıklayıcı değişkenlerin, Y_i 'nin koşullu beklenen değeri üzerinde $E(Y_i/X_i)$ doğrusal bir etkiye sahip olduğu varsayımına dayanmaktadır. Logit Model'de açıklayıcı değişkenler L_i yani Logit değeri ile doğrusal ilişki içerisinde olduğundan gruplanmış verinin varlığında L_i değerleri hesaplanabilecek ve Logit Model EKK yöntemi ile tahmin edilecektir.

Logit Model'in EKK yöntemi ile tahmin edilmesi durumunda tahminciler sapmasızdır fakat etkinliklerini kaybederler. EKK yöntemiyle tahmin edilen model artıkları değişen varyanslı olacağından bu problemi gidermek için Tartılı En Küçük Kareler yöntemi kullanılacaktır⁴.

Tartılı En Küçük Kareler yöntemi ile tahmin edildiğinde parametreler etkin olsa dahi doğrusal olmayacaktır. Her ne kadar değişen varyanslı standart hatalar ile EKK tahmini veriyi açıklamakta kullanışlı olsa da Logit Model'in tahmininde en kullanışlı yöntem Maksimum Benzerlik tahmin yöntemidir⁵. Maksimum Benzerlik tahmini X_i verildiğinde Y_i 'nin dağılımına dayalı olduğundan $Var(Y_i/X_i)$ 'teki değişen varyans kendiliğinden açıklanır⁶.

Maksimum Benzerlik yönteminin uygulanmasında benzerlik fonksiyonundan faydalanılır. Logit model için benzerlik fonksiyonu γ , şu şekilde tanımlanır;

$$\gamma(\beta) = \prod_{i=1}^n F(x_i' \beta)^{y_i} (1 - F(x_i' \beta))^{1-y_i}$$

Benzerlik fonksiyonunun logaritması alınırsa logaritmik benzerlik fonksiyonu L ,

$$\log \gamma(\beta) = L(\beta) = \sum_{i=1}^n y_i \log(F(x_i' \beta)) + (1 - y_i) \log(1 - F(x_i' \beta))$$

olarak elde edilir. Burada $F(\cdot)$ logistik birikimli dağılım fonksiyonu, x_i' açıklayıcı değişkenler vektörü, β ise bilinmeyen katsayılar vektörüdür. Maksimum Benzerlik

⁴ Maddala, s.324.

⁵ A. Colin Cameron ve Pravin K. Trivedi, **Microeconometrics**, New York: Cambridge University Press, 2005, s.471.

⁶ Jeffery M. Wooldridge, **Introductory Econometrics a Modern Approach**, Fourth Edition, Mason, Ohio: South Western Cengage Learning, 2009, s.578.

yöntemi, logaritmik benzerlik fonksiyonunu maksimum yapacak $\hat{\beta}$ katsayılar vektörünün tahminine dayanmaktadır. Her bir β_i için logaritmik benzerlik fonksiyonunun birinci dereceden türevi alınır,

$$\sum_{i=1}^n (y_i - F(x_i\beta))x_i' = 0$$

denklemin elde edilir. Bu denklemin sıfıra eşitlenmesi ile $\hat{\beta}$ katsayılar vektörünün tahminine ulaşırlar.

Bu yöntemin uygulanmasında denklemlerin çözümü için iteratif bir yöntem olan Newton-Rapshon algoritmasından faydalanılır⁷. Maksimum benzerlik yöntemi ile tahmin edilen parametreler genellikle etkin, asimtotik tutarlı ve asimtotik normal tahmincilerdir.

1.3 PARAMETRİK LOGİT MODELİ'NİN KATSAYILARININ ANLAMLILIĞI

Parametrik Logit modelde tahmin edilen katsayıların anlamlılığını test etmek için birden fazla yöntem mevcuttur. Her bir katsayının tek tek anlamlılığı için t-testi ile araştırma yapılabileceği gibi çoklu katsayıların test edilmesinde de belirli yöntemler kullanılır.

Karmaşık hipotezleri test ederken birkaç katsayının eşitliği eş zamanlı olarak veya iki katsayının eşitliği test edilebilir. Bu tür hipotezler Wald, Benzerlik Oranı (LR) ve Lagrange Çarpanı (LM) ile test edilir⁸. Literatürde LM testi logit ve probit modelleri için oldukça sınırlıdır⁹. Bu nedenle bu bölümde bu test tipine yer verilmeyecektir.

Wald testi en basit haliyle tek bir katsayı değerini test eder. Yöntemin çalışma prensibi kısıtlar olmaksızın regresyon tahminine dayanır.

⁷ Christian Gourieroux, **Econometrics of Qualitative Dependent Variables**, Paul B. Klassen(çev.), Second Edition, New York: Cambridge University Press, 1991, s.16.

⁸ Scott Long, **Regression Models for Categorical and Limited Variables: Advanced Quantitative Techniques in the Social Sciences 7**, USA: Sage Publications, 1997, s.87.

⁹ William H. Greene, **Econometric Analysis**, Seventh Edition, California: Pearson Education, 2012, s.115.

$$P(Y_i = 1/X_i) = F(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k)$$

Tek bir katsayının test edilmesi durumunu incelemek üzere k sayıda parametreye sahip logit modelde $H_0 = \beta_1 = 0$ hipotezi test edilmek istenirse Wald istatistiği;

$$W = \frac{(\widehat{\beta}_1 - \beta^*)^2}{\widehat{\sigma}_{\widehat{\beta}_1}^2} = \left(\frac{\widehat{\beta}_1 - \beta^*}{\widehat{\sigma}_{\widehat{\beta}_1}} \right)^2$$

olarak hesaplanır. $\beta^* = \beta_1 = 0$ eşitliği için W , logit modelin Maksimum Benzerlik yöntemiyle tahmin edilen katsayı değerinin aynı katsayı için elde edilen standart hataya bölümünün karesine eşit olacaktır. Bu haliyle W değeri Z istatistik değerinin karesine eşittir¹⁰.

$$Z = \frac{\widehat{\beta}_1}{\widehat{\sigma}_{\widehat{\beta}_1}} = \left(\frac{\widehat{\beta}_1}{\widehat{\sigma}_{\widehat{\beta}_1}} \right)^2 = W$$

Wald testi bu fikirden hareketle çoklu kısıtları test etmek için de kullanılır. Bu durumda test istatistiğini şu şekilde genellemek mümkündür;

$$W = \sum_{k=1}^K \frac{\widehat{\beta}_k^2}{\widehat{\sigma}_{\widehat{\beta}_k}^2} = \sum_{k=1}^K \left(\frac{\widehat{\beta}_k}{\widehat{\sigma}_{\widehat{\beta}_k}} \right)^2 = \sum_{k=1}^K Z_{\widehat{\beta}_k}^2$$

Burada test edilen hipotez $H_0 = \beta_1 = \beta_2 = \dots = \beta_k = 0$ şeklinde olacak ve W değeri p serbestlik dereceli Ki-kare dağılımına uyacaktır. Burada p kısıtlanan parametre sayısıdır.

Benzerlik Oranı (LR), Parametrik Logit modelde çoklu kısıtların test edilmesi amacıyla kullanılan bir başka yöntemdir. Kısıtlı ve kısıtsız regresyon modellerinin tahmin edilmesi temeline dayanan Benzerlik Oranı testi için test istatistiği şu şekildedir;

$$LR = 2(L_{UR} - L_R)$$

¹⁰ Russell Davidson ve James G. MacKinnon. "Convenient Specification Tests for Logit and Probit Models", **Journal of Econometrics**, Cilt.25, Sayı.3, 1984, s.242.

Burada L_{UR} kısıtsız modele ait, L_R ise kısıtlı model için hesaplanan log-benzerlik değeridir. Bu iki logaritmik benzerlik değerinin farkı sıfır hipotezi altında asimtotik ki-kare dağılımıdır¹¹.

1.4 PARAMETRİK LOGİT MODELİ'NİN UYUM İYİLİĞİ

Model uyum iyiliği, gözlenen verinin tahmin edilen regresyon doğrusuna ne kadar yakın olduğunu gösteren özetleyici bir istatistiktir. Parametrik regresyon modellerinde belirlilik katsayısı R^2 model uyum iyiliği ölçüsü olarak kullanılmaktadır. Model uyum iyiliği ölçüsü olan R^2 , $[0,1]$ aralığında değer alır. Bağımlı değişkeni iki seçenektен oluşan modeller için hesaplanan R^2 istatistiği bazı durumlarda $[0,1]$ aralığının dışına çıkabilmektedir. Bu nedenle belirlilik katsayısı doğrusal modeller için iyi bir uyum iyiliği ölçüsü olmasına rağmen doğrusal olasılık, logit ve probit modeller için zayıf bir ölçüdür¹².

Doğrusal modelin tersine parametrik logit modelinde uyum iyiliğini ölçmek için tek bir en iyi ölçü ve yaklaşım yoktur. Birden çok yaklaşım ve uyum iyiliği ölçüsü söz konusudur. Doğrusal olmayan modeller için kullanılan bu uyum iyiliği ölçüleri genellikle *Pseudo* – R^2 olarak adlandırılmaktadır¹³. Parametrik logit modeline ait uyum iyiliği ölçüsü olan R^2 için doğrudan veya dolaylı şekilde modelde sadece sabit parametrenin açıklayıcı olarak yer aldığı duruma bağlı karşılaştırma yapılır¹⁴. Tüm modele ait benzerlik oranı değeri L_1 ve kesim parametresi hariç tüm parametrelerin sıfır olduğu durumda elde edilen benzerlik oranı değeri L_0 iken logit modeli için önerilen ilk uyum iyiliği ölçüsü,

$$Pseudo - R^2 = 1 - \frac{1}{1 + 2(\ln L1 - \ln L0)/N}$$

olarak ifade edilir. N burada gözlem sayısını ifade etmektedir.

¹¹ Badi H. Baltagi, **Econometrics**, Fourth Edition, Berlin: Springer, 2008, s.26.

¹² Ebru Çağlayan ve Melek Astar, “Logit ve Probit Modellerinde Uyum İyiliği Ölçüleri”, **Trakya Üniversitesi Sosyal Bilimler Dergisi**, Cilt.12, Sayı.2, (Aralık 2010), s.3.

¹³ Cameron ve Trivedi, s.473.

¹⁴ Marno Verbeek, **A Guide to Modern Econometrics**, Second Edition, England: John Wiley & Sons, 2004, s.194.

Parametrik logit modeli için benzerlik oranlarını kullanarak geliştirilen bir başka alternatif R^2 değeri McFadden tarafından,

$$McFadden(R^2) = \frac{(\ln L0 - \ln L1)}{(\ln L0)} = 1 - \frac{(\ln L1)}{(\ln L0)}$$

olarak önerilmiştir. Burada $\ln L1 \geq \ln L0$ olmak koşulu ile tüm değişkenlerin bağımlı değişkeni açıklama başarısı incelenmektedir. Eğer tüm eğim parametrelerinin tahmini sıfıra eşitse $\ln L1 = \ln L0$ bir başka ifade ile $\ln L1 / \ln L0 = 1$ olacak ve R^2 değeri “0” olarak elde edilecektir. R^2 değerinin 0’a eşit olması modelde yer alan açıklayıcı değişkenlerin birlikte açıklama gücünün olmadığı anlamına gelmektedir.

Parametrik logit model için bir başka uyum iyiliği ölçüsü Maddala tarafından,

$$R^2_{ML} = 1 - \left(\frac{\ln L0}{\ln L1} \right)^{2/n}$$

olarak önerilmiştir.

Cragg ve Uhler’in yine benzerlik oranına dayalı önerdiği bir başka ölçü ise aşağıdaki gibidir;

$$R^2_{C\&U} = \frac{1 - (\ln L0 - \ln L1)^{2/n}}{1 - (\ln L0)^{2/n}}$$

Efron, ikili tepki modelleri için açıklanan varyans yüzdesinden hareketle,

$$R^2_{Efron} = 1 - \frac{\sum_{i=1}^n (y_i - \hat{\pi}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

formülünü önermiştir. $\hat{\pi}_i = P(Y_i = 1/X_i)$, $Y_i = 1$ olduğu durum için tahmin edilmiş olasılık değerleridir.

McKelvey and Zavoina’nın önerdiği R^2 ölçüsü gizli değişkene ait $y^* = x\beta + \varepsilon$ modeli için,

$$R^2_{M\&Z} = \frac{\widehat{Var}(\hat{y}^*)}{\widehat{Var}(y^*)} = \frac{\widehat{Var}(\hat{y}^*)}{\widehat{Var}(\hat{y}^*) + Var(\varepsilon)}$$

olarak tanımlanmıştır. Bu ölçü doğrusal regresyon modeli için hesaplanan R^2 değerinden iki açıdan farklılık göstermektedir¹⁵. Birincisi gizli değişkenin varyansı yerine gözlenen y_i 'leri kullanır. İkincisi ise hata varyansını sabit varsayar ve tahmin edilen varyans yerine logit modelinde $Var(\varepsilon) = \pi^2/3$ değerini kullanır. \hat{y}^* için ise varyans şu şekilde hesaplanır;

$$\widehat{Var}(\hat{y}^*) = \hat{\beta}' \widehat{Var}(x) \hat{\beta}$$

Formülde yer alan $\widehat{Var}(x)$, açıklayıcı değişkenlerin kovaryans matris tahminidir.

Model uyumunu değerlendirmek için bir başka yaklaşım bağımlı değişkene ait tahmin grup üyeliği ile gözlem grup üyeliğinin karşılaştırılmasıdır. Bu karşılaştırma logit modelinde bağımlı değişkenin doğru tahmin edilip edilmediğini dikkate alır. Logit modelinde bağımlı değişken iki seçenekli olduğundan y_i 'nin tahmin edilen değeri \hat{y}_i 'da iki seçenekli olmalıdır. $y_i = 1$ olup olmadığını belirlemek modelden elde edilen olasılık tahminlerini değerlendirmekle mümkündür. Bir tahmin kuralı olarak $\hat{P} = F(x_i' \beta) > 0,5$ olduğunda $\hat{y}_i = 1$ 'e, olasılık değeri $\hat{P} = F(x_i' \beta) \leq 0,5$ durumunda ise $\hat{y}_i = 0$ 'a eşittir. (y_i, \hat{y}_i) her bir gözlem çifti için (1,0) veya (0,1)'e eşitse yanlış belirlenme söz konusudur. Tablo 1 bağımlı değişkenin gözlenen değerleri için doğru ve yanlış belirlenen tahminlerin sayısını gösterir.

Tablo 1
Logit Modeli İçin Gerçek ve Tahmini Değerlere Ait Çapraz Tablo

	$\hat{y}_i = 0$	$\hat{y}_i = 1$	<i>Toplam</i>
$y_i = 0$	n_{00}	n_{01}	N_0
$y_i = 1$	n_{10}	n_{11}	N_1
<i>Toplam</i>	n_0	n_1	N

¹⁵ Long, s.105.

Tablo 1’de yer alan n_{11} , bağımlı değişkenin gerçek değeri 1 iken doğru tahmin edilmiş gözlemlerin sayısını ve n_{10} gerçek değer 1 iken 0 olarak tahmin edilmiş gözlemlerin sayısını ifade eder. $N_1 = n_{10} + n_{11}$ ve $n_1 = n_{01} + n_{11}$ şeklinde tanımlanır. Bu tabloyu temel alarak birkaç uyum iyiliği ölçüsü hesaplayabiliriz. Örneğin,

$$wr_1 = \frac{n_{01} + n_{10}}{N}$$

genel olarak yanlış belirlenmiş tahminlerin oranını verir. Doğru tahminler genel oranı ise,

$$1 - wr_1 = \frac{n_{00} + n_{11}}{N}$$

olarak gösterilebilir. Bu eşitlik başarı oranı olarak bilinir¹⁶. Her ne kadar başarı oranı model uyum iyiliği için kullanışlı bir ölçü gibi gözüksün de bu durum yanıltıcı olabilir. Çünkü bağımsız değişkenler veri iken gözlemlenen durumların en büyük yüzdesi ile seçilmiş olan tepki kategorisinin yaklaşık %50’sini doğru tahmin etmek mümkündür. Bu nedenle Tablo 1’deki doğru sınıflamaya ait en büyük sıra marjinalini hesaba katmak gerekir¹⁷. $\hat{P} = \frac{N_1}{N} > 1/2$ ve $\hat{P} = \frac{N_1}{N} \leq 1/2$ iken $1 - \hat{P} = 1 - \frac{N_1}{N} = \frac{N_0}{N}$ olarak kabul edilir ve yanlış tahminlerin oranı böylece,

$$wr_0 = \begin{cases} 1 - \hat{P}, & \hat{P} > 0,5 \\ \hat{P}, & \hat{P} \leq 0,5 \end{cases}$$

olarak ifade edilir. Sıra marjinalinin hesaba katılması ile uyum iyiliği ölçüsü,

$$R_p^2 = 1 - \frac{wr_1}{wr_0}$$

olarak elde edilir.

¹⁶ Verbeek, s.196.

¹⁷ Maddala, s.108.

1.5 PARAMETRİK LOGİT MODELİ'NİN KATSAYI YORUMU

Parametrik Logit modelde bağımsız değişkenlerin etkileri logaritmik bahis oranı yani Logit, odds oranı ve marjinal etkiler kavramı üzerinden farklı şekilde yorumlanmaktadır. Logaritmik bahis oranı aşağıdaki gibi tanımlandığında,

$$\ln\left(\frac{P_i}{1-P_i}\right) = \beta_1 + \beta_2 X_i$$

bağımsız değişkenler logaritmik bahis oranı üzerinde doğrusal ve toplamsal bir etkiye sahiptir¹⁸. Bu durumda her bir bağımlı değişkenin logaritmik bahis oranı üzerindeki etkisini yorumlamak çoklu doğrusal regresyon analiziyle aynıdır.

Bir diğer yorum logaritmik bahis oranı yerine bağımsız değişkenlerin odds üzerindeki etkilerini logit regresyon katsayılarından gelen dönüşüm ile yorumlamaktır. Bu dönüşüm logaritmik bahis oranından,

$$\ln\left(\frac{P_i}{1-P_i}\right) = \beta_1 + \beta_2 X_i$$

$$e^{\ln(P_i/1-P_i)} = e^{\beta_1 + \beta_2 X_i}$$

$$(P_i/1 - P_i) = e^{\beta_1} * e^{\beta_2 X_i}$$

olarak elde edilir. Logaritmik kısmın antilogaritması alındığında sol taraftaki denklem odds eşitliğine dönüşür. Bu durumda odds eşitliği sağ taraf için toplamsal yerine çarpımsal etkilerin katsayıların üssünün yorumlanmasına bağlı olarak değişir. Burada $P_i/1 - P_i$ ifadesi $Y_i = 1$ ve $Y_i = 0$ için nispi olasılıkları ölçer ve odds oranı veya relative risk olarak adlandırılır¹⁹. İstihdama katılımı incelerken $Y_i = 1$ istihdamda ve $Y_i = 0$ işsiz olmayı ifade etsin. X_j 'nin katsayısı 0,7 iken hesaplanan odds oranı $e^{(0.7)}$ yaklaşık 2'ye eşit olacaktır. Bunun anlamı X_j 'de bir birimlik değişimin istihdamda olmayı işsizliğe göre iki kat değiştirdiği (arttırdığı / azalttığı) yönündedir.

¹⁸ Fred C. Pampel, **Logistic Regression: A Primer**, USA: Sage, 2000, s.19.

¹⁹ Cameron ve Trivedi, s.470.

Logit model için hesaplanan marjinal etkiler katsayıların doğrudan olasılıklar üzerindeki etkisini göstermektedir. X_{ij} sürekli bir değişken olarak kabul edildiğinde marjinal etkiler,

$$\frac{\partial P_i}{\partial X_{ij}} = P_i(1 - P_i)\beta_j$$

olarak elde edilir. Marjinal etkiler yorumlanırken parametrik logit modele ait katsayıların işaretlerine bakılır. Modeldeki katsayıların işareti marjinal etkilerin de işaretini verir²⁰. Bu durum relative risk değerini yorumlarken de geçerlidir.

²⁰ Jan Salomon Cramer, **Logit Models From Economics And Other Fields**, New York: Cambridge University Press, 2003, s.13.

2. NONPARAMETRİK LOGİT MODELİ

Parametrik regresyon modelleri, bağımlı değişken ile bağımsız değişken arasındaki ilişkinin genellikle doğrusal ve aynı zamanda belirli bir matematiksel kalıba uygun olduğu varsayımına dayanmaktadır. Nonparametrik (parametrik olmayan) regresyonda ise ilişkinin şeklinin bilinmediği ya da bilinen belirli bir matematiksel kalıba uymadığı durumlar geçerlidir.

Parametrik regresyonda geçerli olan fonksiyonel biçime ait varsayımlar bu yöntem için dezavantaj oluşturmaktadır. Bu yöntem için modelin doğru şekilde belirlenmemesi tutarsız tahminlere neden olmaktadır. Parametrik regresyonun aksine nonparametrik yöntem gözlem dağılımlarına ilişkin herhangi bir varsayım gerektirmediği ve katı istatistiksel varsayımlara uymadığı için tercih edilmektedir.

2.1 NONPARAMETRİK REGRESYON'DA DÜZGÜNLEŞTİRME KAVRAMI ve DÜZGÜNLEŞTİRİCİ

Regresyon analizinde amaç bilinmeyen tepki fonksiyonu için makul bir yaklaşım üretmektir. Nonparametrik yöntem için ise bu durum bilinmeyen regresyon fonksiyonu yardımıyla gözlemsel hataları azaltarak y 'nin x 'e olan bağımlılığını önemli ölçüde açıklayabilmektir. Bir eğriyi yaklaştırma yöntemine yaygın olarak düzgünleştirme denir²¹.

Düzgünleştirme fikrinin esası verileri bir eğriye uydurmak ve daha basit fonksiyonların birleşimi olabilen esnek fonksiyonları kullanmaktır²². Parametrik olmayan regresyon modeli,

$$y = m(x) + \varepsilon$$

şeklinde tanımlandığında, $m(\cdot)$ bilinmeyen bir düzgünleştirme fonksiyonunu ifade etmektedir. $m(\cdot)$ 'i düzgünleştirme fonksiyonu olarak varsayarsak $m(\cdot)$ 'i tahmin etmenin birden fazla yolu olduğu ortaya çıkar.

²¹ Wolfgang Hardle, **Applied Nonparametric Regression**, United Kingdom: Cambridge University Press, 1990, s.3.

²² Ebru Çağlayan, **Nonparametrik Regresyon Modelleri**, İstanbul: Derin Yayınları, 2012, s.23.

Bu yöntemlere parametrik olmayan regresyon kestirimcileri, bir başka ifade ile $m(x)$ regresyon kestirimcisine (tahmincisine) düzgünleştirici denir. Parametrik olmayan yöntemde düzgünleştirme işlemi düzgünleştiriciler vasıtasıyla gerçekleştirilmektedir. Düzgünleştirme yapılırken yoğunluk fonksiyonu tahmincilerinden yararlanır. Bu tahminler genellikle Kernel (çekirdek) tahminci, Local Polynomial ve Local Linear gibi yerel regresyon tahmincileri ile Spline Smoothing yöntemi kullanılarak yapılmaktadır.

2.1.1 Kernel (Çekirdek) Tahmincisi

Temel olarak parametrik olmayan regresyon modeli; bağımlı ve bağımsız değişkenlerin ağırlıklı ortalamalarını alarak $m(\cdot)$ regresyon fonksiyonunu tahmin etmeye çalışır. Kernel nonparametrik tahmin yöntemlerinde kullanılan ağırlıklandırma fonksiyonu olup bu ağırlıkların kullanılmasıyla kernel tahmincisi,

$$\hat{m}(x) = \frac{\sum_{i=1}^n k\left(\frac{x_i - x}{h}\right) y_i}{\sum_{i=1}^n k\left(\frac{x_i - x}{h}\right)}$$

şeklinde tanımlanmaktadır. Kernel tahmincisi için ağırlıklar,

$$w_i(x) = \frac{k\left(\frac{x_i - x}{h}\right)}{\sum_{j=1}^n k\left(\frac{x_j - x}{h}\right)}$$

olur. $k(\cdot)$ kernel fonksiyonu h ise bant genişliği (bandwidth) diğer bir ifade ile düzgünleştirme parametresi olarak tanımlanmaktadır. Ağırlıklar ile yeniden yazıldığında parametrik olmayan regresyon tahmincisi için eşitlik,

$$\hat{m}(x) = \sum_{i=1}^n w_i(x) y_i$$

şeklinde olacaktır. Bu eşitlik bize regresyon tahmincisinin, açıklanan değişkenin yerel olarak ağırlıklı ortalaması olduğunu göstermektedir. Esasen bir kernel fonksiyonu sıfır etrafında simetrik dağılımlı ve

$$k(u) = k(-u)$$

$$\int_{-\infty}^{+\infty} k(u) du = 1$$

özelliklerine sahip bir olasılık yoğunluk fonksiyonudur. Burada u için $u = \left(\frac{x_i - x}{h}\right)$ eşitliği geçerlidir. Üniform yoğunluk $k_0(u)$ bilinen kernel fonksiyonunun özel bir durumudur. Üniform kernel,

$$k_0(u) = \begin{cases} 1/2, & |u| \leq 1 \\ 0, & d.d. \end{cases}$$

şeklinde tanımlanır.

Kernel analizlerinde üniform kernel yoğunluğu kullanılabileceği gibi bunun yerine herhangi bir kernel fonksiyonuyla değiştirilerek düzgünleştiriciyi genellemek mümkündür. Kernel regresyon tahmincisi, Nadaraya-Watson (NW) tahmincisi veya Local (yerel) sabit tahminci olarak da bilinir. Kernel yoğunluk fonksiyonu tahmini için düzgünleştirme parametresi (h) düzgünleştirme işleminde önemli rol oynamaktadır. Öyle ki bant genişliği seçimi kernel tahmincisinin seçiminden daha önemlidir²³.

Düzgünleştirme işlemi h 'ın aldığı değerlere göre değişkenlik göstermektedir. h 'ın daha büyük değerleri $\hat{m}(x)$ tahmin sonuçlarını x 'de düzgünleştirirken daha küçük h değerleri tahmin sonuçlarını daha fazla düzensizleştirir. Düzgünleştirme parametresi için $h \rightarrow 0$ ve $h \rightarrow \infty$ iki uç durumu göstermektedir. $h \rightarrow 0$ iken $\hat{m}(x_i) \rightarrow y_i$ olacaktır. Bu durum $\hat{m}(x_i)$ için tahmin sonuçlarının y_i 'ye yakınsayacağı yani verilerin enterpolasyonunun elde edileceği anlamına gelmektedir²⁴. Bunun aksine $h \rightarrow \infty$ olduğu durumda tüm x 'ler için $\hat{m}(x_i) \rightarrow \bar{y}$ tahmin sonuçları örnek ortalamasına eşit olacaktır. Bu durumda parametrik olmayan şartlı beklenti fonksiyonu sabit bir fonksiyon olacaktır.

Kernel analizlerinde üniform yoğunluğu sürekli tahmin üretmek için iyi bir kernel tercihi değildir. Bunun yerine sürekli dağılımlar elde etmek için sürekli bir kernel fonksiyonuna ihtiyaç vardır. En çok kullanılan iki sürekli dağılım Epanechnikov kernel veya Normal veya Gaussian kernel fonksiyonları kullanılmaktadır. Bu fonksiyonlar ile diğer alternatif kernel fonksiyonlarına Tablo 2'de yer verilmiştir.

²³ Matt P. Wand ve M. Chris Jones, "Kernel Smoothing", 1995, Monographs on Statistics and Applied Probability Chapman and Hall, http://compdiag.molgen.mpg.de/docs/talk_05_01_04_stefanie.pdf (12 Ocak 2017), s.15.

²⁴ Wolfgang Hardle ve Diğerleri, **Nonparametric and Semiparametric Models**, Berlin: Springer, 2004, s.90.

Tablo 2
Kernel Tablosu

Kernel	Eşitlik
Uniform	$k_0(u) = \frac{1}{2}I(u \leq 1)$
Epanechnikov	$k_1(u) = \frac{3}{4}(1 - u^2) I(u \leq 1)$
Biweight	$k_2(u) = \frac{15}{16}(1 - u^2)^2 I(u \leq 1)$
Triweight	$k_3(u) = \frac{25}{32}(1 - u^2)^3 I(u \leq 1)$
Gaussian	$k_\phi(u) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right)$

2.1.2 Local Linear (Yerel Doğrusal) Tahminci

NW tahmincisi sık sık yerel sabit tahminci olarak adlandırılır ve $m(x)$ için şartlı beklenti fonksiyonu yerel olarak sabit kabul edilir. $\hat{m}(x)$ 'e ait minimizasyon problemi,

$$\hat{m}(x) = \underset{\alpha}{\operatorname{argmin}} \sum_{i=1}^n k\left(\frac{x_i - x}{h}\right) (y_i - \alpha)^2$$

bu durumu açıkça göstermektedir. Bu belirli bir bölgede yalnızca y 'nin ağırlıklı regresyonudur. Ağırlıklar olmaksızın bu tahmin problemi örnek ortalamasına indirgenir. NW tahmincisi bu yerel ortalamayı genelleştirir ve yerel yaklaşımlar ile parametrik olmayan tahminci için alternatif bir şartlı beklenti fonksiyonu elde edilebilir. Böyle bir çok yerel yaklaşım mümkündür. Popüler bir tercih yerel doğrusal (LL) yaklaşımı kullanmaktır. $\hat{m}(x)$ ' i yerel olarak sabit görmek yerine, yerel doğrusal yaklaşımı bölgesel doğrusal fonksiyon ile şartlı beklenti fonksiyonuna yaklaşır ve bu yerel

yaklaşımın yerel olarak ağırlıklandırılmış en küçük kareler ile tahminini yapar ve her x için minimizasyon problemini,

$$\{\hat{\alpha}(x), \hat{\beta}(x)\} = \underset{\alpha, \beta}{\operatorname{argmin}} \sum_{i=1}^n k\left(\frac{x_i - x}{h}\right) (y_i - \alpha - \beta(x_i - x))^2$$

olacak şekilde çözer. Yerel doğrusal tahminci $\hat{m}(x)$ 'in tahmin edilen kesim parametresi,

$$\hat{m}(x) = \hat{\alpha}(x)$$

ve regresyon türevinin $\nabla \hat{m}(x)$ yerel doğrusal tahmincisi için eğim katsayısı,

$$\nabla \hat{m}(x) = \hat{\beta}(x)$$

olarak tahmin edilir.

Matrislerle ifade edecek olursak her x kümesi için tahmin sonuçları,

$$z_i(x) = \begin{pmatrix} 1 \\ x_i - x \end{pmatrix}$$

ve

$$k_i(x) = k\left(\frac{x_i - x}{h}\right)$$

$$\begin{aligned} \begin{pmatrix} \hat{\alpha}(x) \\ \hat{\beta}(x) \end{pmatrix} &= \left(\sum_{i=1}^n k_i(x) z_i(x) z_i(x)' \right)^{-1} \sum_{i=1}^n k_i(x) z_i(x) y_i \\ &= (Z' K Z)^{-1} Z' K y \end{aligned}$$

olacak şekilde elde edilir. Ve K ,

$$K = \operatorname{diag}\{k_1(x), k_2(x) \dots, k_n(x)\}$$

olarak tanımlanır.

2.1.3 Smoothing Spline Tahminci

Spline modeller en küçük kareler yöntemi ile tahmin edildiğinden doğrusal regresyon modellerinin özelliklerini paylaşır²⁵. Spline düzgünleştirme yöntemi $m(\cdot)$

²⁵ Luke Keele, *Semiparametric Regression for the Social Sciences*, England: John Wiley & Sons, 2008, s.64.

fonksiyonu için uyum iyiliği ölçüsü olarak kalıntı kareleri toplamını (RSS) dikkate alır. Bu yüzden spline tahmin-doğrusal regresyon modelleri gibi- y ile $m(x)$, parametrik olmayan tahmin arasındaki farkın kareler toplamını minimize eder. Parametrik olmayan regresyon için kalıntı kareleri toplamı,

$$RSS = \sum_{i=1}^n (y - m(x))^2$$

olarak tanımlanır.

Spline düzeltme yönteminin esası cezalı en küçük kareler toplamının minimizasyonuna dayanmaktadır²⁶. Bu minimizasyon probleminde $m(x)$ 'i tahmin etmek için önemli bir husus çok fazla parametrenin kullanılmasıdır. Bu nedenle spline düzeltme yöntemi parametrik olmayan uyumu tahmin etmek için kullanılan parametrelerin sayısına bir ceza eklemektedir. Bu RSS'yi minimize ederken aynı zamanda kullanılan yerel parametrelerin sayısı için bir ceza terimi eklemek anlamına gelmektedir. Spline model için kullanılan ceza,

$$\lambda \int [m^{(t)}(x)]^2 dx$$

şeklinde yazılabilir. Bu ceza iki kısımdan oluşur. λ düzgünleştirme parametresi olarak adlandırılır ve minimizasyon işlemi sırasında düzgünleştiriciye verilen ağırlığı kontrol eder²⁷. İkinci kısım ise $m(x)$ fonksiyonunun integralinden oluşur. Burada $m^{(t)}(x)$ gösterimi fonksiyonun t. dereceden türevini ifade eder. λ büyük olduğunda pürüzsüzlüğe bir değer yerleştirilir ve olası tahminciler t. dereceden yüksek türevler ile cezalandırılır. $\lambda = \infty$ olduğu durumda t. dereceden polynomial regresyon veri için uyum üretir²⁸.

Ceza teriminin eklenmesiyle birlikte kalıntı kareleri toplamına ait eşitlik,

²⁶ Dursun Aydın ve Diğerleri, "Yarı Parametrik Modellerde Splayn Düzeltme ile Tahmin ve Çıkarımlar", **Anadolu Üniversitesi Bilim ve Teknoloji Dergisi**, Cilt.8, Sayı.1, (2007), s.42.

²⁷ Hardle ve Diğerleri, s.103.

²⁸ Randall L. Eubank. **Nonparametric Regression And Spline Smoothing**, Second Edition, New York: Marcel Dekker, 1999, s.227.

$$RSS(m, \lambda) = \sum_{i=1}^n (y_i - m(x_i))^2 + \lambda \int [m^{(t)}(x)]^2 dx$$

olur. Bu eşitlikte t . dereceden türev ifadesi olan $m^{(t)}(x)$ terimi bir eğrinin değişim oranını ölçer. $m^{(t)}(x)$ için büyük bir değer yüksek eğim anlamına gelirken küçük bir değer ise eğimin düşük olduğunu ifade eder. $m^{(t)}(x)$ teriminin karesinin integralini almak parametrik olmayan tahminler için pürüzsüzlüğün bir ölçüsünü verir ve böylece polynomial modelden yola çıkarak ne kadar sınırlama getirileceği incelenmiş olur.

Büyük olduğunda $m(x)$ daha pürüzlü küçük olduğunda ise daha pürüzsüzdür. Pürüzsüzlük ile uyum iyiliği arasındaki dengeyi düzenleyen λ parametresi için çok küçük ve çok büyük değerler pürüzsüzlük miktarı üzerinde yorumlanırken orta değerler veriye uygulanan pürüzsüzlük miktarı üzerinde yorumlanabilir bir etkiye sahip değildir. Bu problemi aşmak için λ 'yı serbestlik derecesinin bir yaklaşımı olarak dönüştürebiliriz. Çünkü λ 'yı parametrik olmayan tahmin için kullanılan yerel parametrelerin sayısını kontrol etmek olarak inceleyebiliriz.

Smoothing spline için hesaplanan serbestlik derecesi sıradan EKK ile aynı mantığa sahiptir. Sıradan EKK için serbestlik derecesi doğrusal modelin şapka matrisinden hesaplanır. Doğrusal model için şapka matrisi,

$$H = X(X'X)^{-1}X'$$

olurken serbestlik derecesi ise şapka matrisin izine $tr(H)$ eşittir. Spline model için de serbestlik derecesi modele ait şapka matrisinin izine eşittir. Bunun için şapka matrisini spline modele genellemek gerekmektedir. Şapka matrisini spline modele genellemek amacıyla ceza matrisini yazarsak,

$$\int [m^{(t)}(x)]^2 dx = \beta' D \beta$$

şeklinde olacaktır. Ceza matrisinde yer alan D matrisi,

$$D = \begin{bmatrix} 0_{txt} & 0_{txk} \\ 0_{kxt} & I_{kxk} \end{bmatrix}$$

olacaktır. Bu matriste k düğüm sayısını gösterir. Ceza matrisiyle birlikte spline regresyon modeli,

$$RSS(m, \lambda) = \|y - x\beta\|^2 + \lambda\beta'D\beta$$

olarak yazılır. Spline modele ait şapka veya düzgünleştirme matrisi cezaların eklenmesiyle birlikte,

$$S_\lambda = X(X'X + \lambda^{2p}D)^{-1}X'$$

olacaktır. Burada p temel fonksiyonların polinom derecesini göstermektedir. Sabit bir değer olarak λ^{2p} ve D matrisinin çarpımı ceza terimini göstermektedir. Doğrusal regresyon modelinde olduğu gibi spline model için de serbestlik derecesi şapka matrisi olan S_λ 'nın izine eşittir. Fakat doğrusal model ile spline modele ait serbestlik derecesi arasında bir fark vardır. Doğrusal model için hesaplanan serbestlik derecesi tam sayı iken spline model için ceza teriminden kaynaklı olarak serbestlik derecesi bir tam sayı olmayacaktır.

2.1.4 Local (Yerel) Polynomial Tahminci

Parametrik olmayan regresyonda düzgünleştirme yapılırken yerel ortalama kestirimcisini kullanmak yerine yerel regresyonları kullanarak çözüm elde edebiliriz. Yerel yaklaşımlar ile parametrik olmayan tahminci için alternatif bir şartlı beklenti fonksiyonu elde etmek istenirse LL yaklaşımı kullanılabileceği gibi Local Polynomial yaklaşımdan da faydalanmak mümkündür. LL yaklaşımı parametrik olmayan yöntem için doğrusal uyum sağlarken Local Polynomial yaklaşım doğrusal model yerine polynomial regresyon modelini kullanmayı önerir. Açıklayıcı değişkenlerin p. dereceden polinom olduğu polynomial regresyon,

$$y = \beta_0 + \beta_1x + \beta_2x^2 + \dots + \beta_px^p + \varepsilon$$

olarak tanımlanır. Bu yöntemde kernel fonksiyonları kullanılmaktadır ve özel noktalarda p. dereceden polinomların tahmini ile yerel olarak elde edilen tahminlerden yoğunluk fonksiyonu elde edilmektedir²⁹. Burada düzgünleştirme parametresi ile belirlenen her bir gözlem aralığı için x ile y ilişkilendirilir ve yerel tahmin $x = x_0$ gözlem aralığında y için öngörülen değerdir. Bu işlem her bir gözlem aralığı için tekrarlanır. Her gözlem aralığından elde edilen değerleri birleştirmek $m(x)$ 'in

²⁹ Çağlayan, s.80.

parametrik olmayan tahminini üretir. İteratif tahmin süreci parametrik olmayan uyum için tahminleri şu şekilde üretir; öncelikle her x için bant genişliği içerisinde bir dizi ağırlık hesaplanır. Bu ağırlıklar kernel tahmincisi için kullanılanlarla aynı şekilde hesaplanır. Yani merkez noktasından her gözlemin ölçekli ve tartılı mesafesi hesaplanır ve merkez noktasına verilen en büyük ağırlıkla her gözlem için simetrik ağırlık seti üretmek amacıyla bu mesafelere bir kernel fonksiyon uygulanır. Sonra yerel polynomial regresyon modeli,

$$\frac{y_i}{w_i} = \beta_0 + \beta_1 \frac{x_i}{w_i} + \beta_2 \frac{x_i^2}{w_i} + \dots + \beta_p \frac{x_i^p}{w_i} + \frac{\varepsilon_i}{w_i}$$

için w_i , ağırlıkları kullanılarak ağırlıklı kalıntı kareleri toplamı $\sum_{i=1}^n (w_i \varepsilon_i)^2$ minimize edilir. Bu süreç ağırlıklar olmaksızın çözülebilir. Ağırlıksız çözüm LOESS ağırlıklar ile yapılan çözüm ise LOWESS olarak adlandırılmaktadır³⁰.

Bu tahmin sürecini matrisler ile göstermek istersek, Kernel ağırlıklarının eklenmesiyle polynomial regresyon için kalıntı kareleri minimizasyonu,

$$\hat{m}(x) = \min_{\beta} \sum_{i=1}^n \{y_i - \beta_0 - \beta_1(x_i - x) - \dots - \beta_p(x_i - x)^p\}^2 K_h(x_i - x)$$

şeklinde olacaktır. Burada $\beta = (\beta_0, \beta_1, \dots, \beta_p)^T$ katsayılar vektörünü ifade eder. Polinomun derecesi “1” olduğunda bu eşitlik doğrusal bir uyuma sahip olacağından LL tahmincisi halini alır. Katsayıları tahmin etmek için ağırlıklar $K_h(x_i - x)$ ile ağırlıklı en küçük kareler tahmincisinin matris gösterim,

$$X = \begin{pmatrix} 1 & (x_1 - x) & (x_1 - x)^2 & \dots & (x_1 - x)^p \\ 1 & (x_2 - x) & (x_2 - x)^2 & \dots & (x_2 - x)^p \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & (x_n - x) & (x_n - x)^2 & \dots & (x_n - x)^p \end{pmatrix}, Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}$$

³⁰ Keele, s.28.

$$W = \begin{pmatrix} K_h(x - x_1) & 0 & \dots & 0 \\ 0 & K_h(x - x_2) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & K_h(x - x_n) \end{pmatrix}$$

olacaktır. Matrisler yardımıyla ağırlıklı en küçük kareler tahmincisi için,

$$\hat{\beta}(x) = (X^T W X)^{-1} (X^T W Y)$$

formülü β' 'yi minimize eder.

2.2 NONPARAMETRİK REGRESYON'DA DÜZGÜNLEŞTİRME PARAMETRESİ'NİN SEÇİMİ

Nonparametrik regresyon tahminlerinde düzgünleştirme parametresinin seçimi oldukça önemlidir. h 'in seçimi yoğunluk fonksiyonu tahminlerinde ne kadar düzgünleştirme yapılacağını belirlemektedir. h azaldığında tahminler daha pürüzlü tahminler elde edilmesine neden olacaktır. h arttıkça tahminler daha pürüzsüz hale gelir. Bu durum tahmin varyansı azalsa da sapma pahasına gerçekleşir³¹. Düzgünleştirme parametresi h 'in seçimi, sapma ve varyans arasındaki bu özelliği dengelemeyi amaçlamaktadır.

Parametrik olmayan modellerde en önemli sorun verilerin ne kadar pürüzsüzleştirileceğidir. Bunun için düzgünleştirme parametresinin belirlenmesi gerekmektedir. Düzgünleştirme parametresinin seçimine ilişkin birden fazla yöntem vardır. Bu yöntemleri uygulamak kimi zaman zor ve karmaşık sonuçlara neden olduğundan uygun pürüzsüzleştirme seviyesini belirleyebilmek için daha basit bir yöntem olan otomatik seçiciler kullanılmaktadır³². Burada bazı zorluklara rağmen geniş bir uygulama alanına sahip ve çeşitli yöntemlere uyarlanabilir olması nedeniyle genel bir otomatik düzgünleştirme seçicisi olan çapraz geçerlilik yöntemi anlatılacaktır.

³¹ Bruce E. Hansen, "Econometrics", 2017, University of Wisconsin Department of Economics, <http://www.ssc.wisc.edu/~bhansen/econometrics/EconometricsKindle.pdf> (17 Şubat 2017), s.491.

³² Adrian Bowman ve Ludger Evers, "Nonparametric Smoothing", 2015, School of Mathematics & Statistics The University of Glasgow, <https://warwick.ac.uk/fac/sci/statistics/apts/students/resources-1415/smoothing-preliminary-material.pdf> (15 Şubat 2018), s.21

Parametrik regresyonda olduğu gibi parametrik olmayan regresyon için de temel sorun elde edilen tahminler ile $m(x)$ arasındaki tutarsızlıktır. Ortalama Hata Kare (MSE) bu tutarsızlığı gösterir. İdeal olarak h 'ı seçerken kullanılacak ölçüler MSE temeline dayanmaktadır. Düzgünleştirme parametresi h ile birlikte ele alındığında MSE,

$$MSE_n(x, h) = E(\hat{m}(x, h) - m(x))^2$$

şeklinde tanımlanır. Bu eşitlik varyans ve sapmanın karesine eşittir. MSE ile ilişkili olan bir diğer ölçü Bütünleşik Ortalama Hata Karedir (IMSE). IMSE,

$$IMSE_n(x, h) = \int_{-\infty}^{+\infty} MSE_n(x, h) f(x) dx$$

$$IMSE_n(x, h) = \int_{-\infty}^{+\infty} E(\hat{m}(x, h) - m(x))^2 f(x) dx$$

olarak ifade edilir. Burada $f(x)$, x_i 'in marjinal yoğunluğu olup parametrik olmayan regresyon için kullanılan ağırlık fonksiyonudur.

Tahmini Ortalama Hata Kare (MSFE) ve $IMSE_n(h)$ nonparametrik regresyon için uyumun eşdeğer ölçüsüdür³³. MSFE,

$$MSFE_n(h) = \sigma^2 + IMSE_n(h)$$

varyans ve $IMSE_n(h)$ toplamına eşittir.

Optimum bant genişliği h , $IMSE_n(h)$ 'yi veya $MSFE_n(h)$ 'ı minimize eden değerdir. Burada bu değerler bilinmemekle birlikte $MSFE_n(h)$ 'ı örnekten elde edilen kalıntı kareleri ortalaması ile tahmin edebiliriz. Bunun için parametrik olmayan regresyona ait kalıntıların hesaplanması gerekmektedir. Parametrik olmayan regresyon için kalıntılar,

$$\tilde{e}_i(h) = y_i - \tilde{m}_{-i}(x_i, h)$$

olarak hesaplanır. $\tilde{m}_{-i}(x_i, h)$, veriden (X_i, Y_i) gözlem çiftinin dışlanması ile bir regresyon denkleminde türetilir. i 'nci verinin dışlanması ile kullanılan regresyon denkleminde elde edilen i 'nci değerin tahmini i 'nci veri için gelecekteki pseudodata

³³ Hansen, s.492.

(pseudo veri) olarak varsayılır. Bu yöntem gelecekte veri tahminlerinin tahmin hata değerini analiz etmeye olanak sağlar. Bir veri yerine birden fazla veriyi dışlayan ve parametrik olmayan kalıntıların elde edilmesinde kullanılan bu yöntem leave-one-out olarak adlandırılır³⁴. Leave-one-out yöntemiyle tahmin edilen kalıntıları kullanarak düzgünleştirme parametresinin seçimi Çapraz Geçerlilik (*CV*) olarak bilinir.

Tahmincinin düzgünleştirme parametresine olan bağımlılığını gösteren *CV* kriteri tahmin edilen kalıntı karelerinden,

$$CV(h) = \frac{1}{n} \sum_{i=1}^n \tilde{e}_i(h)^2$$

olarak elde edilir. Optimum düzgünleştirme parametresi *CV(h)* 'ı minimize eden değerdir. Düzgünleştirme parametresi *h*'ın *CV(h)*'ı minimize eden değeri için,

$$E\{CV(h)\} = MSFE_{n-1}(h) = IMSE_{n-1}(h) + \sigma^2$$

eşitlikleri geçerlidir. Burada $IMSE_{n-1}(h)$, parametrik olmayan tahmincinin kullanılmasıyla $n - 1$ gözlemden elde edilen bütünleşik ortalama hata kareyi ifade etmektedir. Şayet n büyük ise $IMSE_{n-1}(h)$ ile $IMSE_n(h)$ neredeyse aynı olacaktır. Bu nedenle $CV(h)$, $IMSE_{n-1}(h) + \sigma^2$ 'nin bir tahmincisi olarak esasen yansızdır. Eşitlikteki ikinci terim olan σ^2 düzgünleştirme parametresinden bağımsız bir sabit olduğundan h 'ın seçim problemi ile alakasızdır. Bu sebeple $CV(h)$ 'ı, $IMSE$ 'nin tahmincisi olarak ve h 'ın $CV(h)$ 'ı minimize eden değerini $IMSE_n(h)$ 'ı minimize eden değer olarak kabul edebiliriz. Bununla birlikte çoğu kez çapraz geçerlilik yerine etkin bir hesaplama biçimine sahip olan Genelleştirilmiş Çapraz Geçerlilik (*GCV*) yaklaşımı kullanılmaktadır. *GCV* seçicisi,

$$GCV(h) = n \frac{\sum_{i=1}^n (y_i - \tilde{m}_i(x_i, h))^2}{tr(I - S)^2}$$

olarak hesaplanmaktadır. Eşitliğin paydasında yer alan ifade nonparametrik regresyondan elde edilen serbestlik derecesidir.

³⁴ Kunio Takezawa, **Introduction to Nonparametric Regression**, New Jersey: John Wiley & Sons Publication, 2006, s.115-116.

2.3 NONPARAMETRİK LOGİT MODEL'İ TANIMI ve YAPISI

Logit regresyon modeli için Y_i 'nin koşullu beklenen değeri,

$$E(Y_i/X_i) = P(Y_i = 1/X_i) = G(x_i'\beta + \varepsilon)$$

olarak ifade edilmektedir. Burada $x_i'\beta$, X 'in tüm bileşenlerinin doğrusal bir kombinasyonudur. Bir kişinin çoklu özelliklerini bir sayıya toplar ve bu nedenle indeks fonksiyonu olarak adlandırılır. β , bilinmeyen katsayı vektörüdür. $G(\cdot)$, $[0,1]$ aralığında gerçel doğruyu işaret eden herhangi bir sürekli fonksiyonu simgeler. $G(\cdot)$ aynı zamanda $x_i'\beta$ 'yi koşullu beklenti olan $E(Y_i/X_i)$ ile bağladığı için bağlantı fonksiyonu (link fonksiyonu) olarak adlandırılır.

Parametrik yöntem $G(\cdot)$ 'nin bilinen bir parametrik form olduğunu nonparametrik yöntem ise $G(\cdot)$ 'nin bilinmeyen düzgün bir fonksiyon olduğunu varsayar. Parametrik olmayan regresyonda $G(\cdot)$ 'nin herhangi bir varsayıma sahip olmaması yanı sıra açıklayıcı değişkenler arasındaki ilişki de bilinmemektedir. Parametrik logit modelde $G(x_i'\beta)$ tüm X 'ler için lojistik kümülatif dağılım fonksiyonudur ve

$$E(Y_i/X_i) = P(Y_i = 1/X_i) = \frac{e^{(x_i'\beta)}}{1 + e^{(x_i'\beta)}} = \frac{1}{1 + e^{-(x_i'\beta)}}$$

olarak tanımlanır. Lojistik regresyon, geliştirilmiş doğrusal model (GLM) olarak adlandırılan geniş bir model sınıfının belirli bir örneğidir. Geliştirilmiş doğrusal model için doğrusal tahminci;

$$\eta = x_i'\beta$$

şeklinde tanımlanmaktadır. Geliştirilmiş doğrusal regresyon bağlamında doğrusal bir modelin hedef değişkenini elde etmeye yarayan logit dönüşümü gibi bir dönüştürme fonksiyonuna bağlantı fonksiyonu denir³⁵. Logit model için lojistik bağlantı fonksiyonu,

$$G(P) = \log\left(\frac{P}{1-P}\right) = \eta$$

şeklinde tanımlanır.

³⁵ Takezawa, s.421-422.

Parametrik logit modeli katı istatistiksel sınırlamalar getirdiğinden daha esnek regresyon incelemeleri gerekmektedir. Logit modelini genelleştirmenin bir yolu tek indeks modelidir (SIM). Tek indeks modeli ile $x_i'\beta$ indeksinin doğrusal formu korunurken $G(\cdot)$ fonksiyonunun seçilmiş bir düzgünleştirme fonksiyonu olan $g(\cdot)$ ile veriden elde edilmesine izin verilir³⁶. Bu durumda logit dönüşümü sonucu doğrusal model yerine toplamsal model olarak ifade edilirken parametrik olmayan logit modeli,

$$\log\left(\frac{P}{1-P}\right) = \sum_{j=1}^d g_j(x_j)$$

olarak tanımlanmaktadır. Burada g_j parametrik olmayan bir düzgünleştirme fonksiyonudur. Bilinen bir bağlantı fonksiyonu olan $G(\cdot)$ ile bu model genelleştirilmiş toplamsal model (GAM) olarak adlandırılır. Genelleştirilmiş toplamsal model bilinen bağlantı fonksiyonu ile X 'in her bileşeninin bilinmeyen nonparametrik fonksiyonların toplamı olarak yazılabileceğini varsayar³⁷. Böylelikle her bir değişkenin doğrusal olmayan fonksiyonlarına izin veren GAM toplamsallık özelliğini koruyarak standart bir doğrusal modeli genelleştirilmiş toplamsal modele genişletir.

2.4 NONPARAMETRİK LOGİT MODELİ'NE İLİŞKİN TAHMİN YÖNTEMİ

Genelleştirilmiş doğrusal model olarak da tanımlanan parametrik logit modelin tahmini benzerlik fonksiyonun maksimizasyonuna dayanmaktadır. Bu model tahmininde Newton-Rapshon algoritması kullanılarak en çok benzerlik tahmini iteratif olarak elde edilmektedir.

Parametrik logit model ve daha genel olarak kanonik üstel modellerin tahmini için Newton-Rapshon, Fisher skorlama ve iteratif yeniden ağırlıklandırılmış en küçük

³⁶ Hardle ve Diğerleri, s.15.

³⁷ Trevor Hastie ve Robert Tibshirani, **Generalized Additive Models**, London: Chapman and Hall, 1990, s166-167.

kareler (IRLS) yöntemi aynıdır³⁸. IRLS, genelleştirilmiş doğrusal modelin tahminini ağırlıklı en küçük karelerin yinelenen uygulamasına indirgediğinden nonparametrik logit modelin tahmin edilmesi için oldukça uygundur.

Parametrik logit model için IRLS tahmin prosedürü, aynı şekilde nonparametrik logit modelin tahmininde kullanılan local scoring algoritmasının da temelini oluşturmaktadır³⁹. Bu tahmin yöntemi uygulanırken backfitting algoritmasından yararlanır. Backfitting yönteminin arkasındaki temel düşünce, toplamsal modelden kısmi kalıntıların tekrarlı olarak düzgünleştirilmesi yoluyla toplamsal modeldeki her bir parametrik olmayan bileşeni tahmin etmektir.

Nonparametrik regresyon modelini,

$$\eta_i = y_i = \alpha + \sum_{j=1}^m g_j(x_{ji}) + \varepsilon_i$$

olarak gösterirsek bu modelden kısmi kalıntıları,

$$e_p^j = y - \hat{\alpha} - \sum_{k \neq j} \hat{g}_k$$

şeklinde elde ederiz. Burada j. düzgünleştirme terimine ilişkin kısmi kalıntılar j. açıklayıcı değişken hariç tepki değişkeninden mevcut tüm model terimlerinin tahmininin çıkarılmasıyla elde edilir. Bu şekilde modelde yer alan tüm açıklayıcı değişkenler için kısmi kalıntılar hesaplanır. Düzgünleştirme işlemi sırasında herhangi bir nonparametrik yöntem kullanılabilir. Backfitting algoritması tekrarlı bir yöntem olduğundan uygulanabilmesi için başlangıç değerlerine ihtiyaç duyulmaktadır. Tahmin prosedürü şu adımlardan oluşmaktadır,

1. Adım : Başlangıç değerleri $\hat{\alpha} = n^{-1} \sum_{i=1}^n Y_i = \bar{Y}$ ve tüm x, j için $\hat{g}_j(x) = 0$ olarak belirlenir.

2. Adım: $j = 1, \dots, m$ olacak şekilde kısmi artıklar hesaplanır.

3. Adım: Elde edilen kısmi artıklar x_j 'de düzgünleştirilir.

³⁸ David Ruppert, Matt P. Wand ve Raymond J. Carroll, **Semiparametric Regression**, New York: Cambridge University Press, 2003, s.211.

³⁹ Hardle, s.355.

4. Adım: x_j değişkeni düzgünleştirilmiş x_j tahminleri ile değiştirilir.

5. Adım: 2-4 arasındaki adımlar her x için 2'den j 'ye kadar tekrarlanır.

6. Adım: Kalıntı kareleri toplamı aşağıdaki şekilde hesaplanır.

$$RSS = \sum_{y=1}^m \left[\left(y_i - \hat{\alpha} + \sum_{j=1}^m \hat{g}_j(x_{ji}) \right) \right]^2$$

7. Adım: Kalıntı kareleri toplamındaki değişim belirli bir tolerans seviyesinde ise model yakınsamış demektir ve algoritma durur. RSS 'deki değişim belirli bir tolerans seviyesinde değilse tolerans seviyesine gelene kadar algoritma devam eder.

Bu yöntem, sırayla her bir açıklayıcı değişken için-diğer değişkenleri sabit tutup- tahminleri tekrar tekrar güncelleyerek birden çok açıklayıcı değişkeni içeren nonparametrik modeli tahmin eder. Backfitting algoritması nonparametrik logit modeli tahmin etmek amacıyla local scoring algoritması ile birleştirildiğinde local scoring algoritmasına ait tahmin prosedürü aşağıdaki gibi olacaktır⁴⁰.

1.Adım : Başlangıç değerleri $\hat{\alpha} = \log(\bar{y} / (1 - \bar{y}))$, burada $\bar{y} = ortalama(y_i)$ yani örnek oranıdır ve tüm x, j için $\hat{g}_j(x) = 0$ olarak hesaplanır.

2.Adım: $\hat{\eta}_i = \hat{\alpha} + \sum_{j=1}^m \hat{g}_j(x_{ji})$ ve $\hat{p}_i = 1 / (1 + \exp(-\hat{\eta}_i))$ şeklinde tanımlanır.

Tekrarlama Adımı:

a) Hedef değişkeni oluşturulur,

$$z_i = \hat{\eta}_i + \frac{(y_i - \hat{p}_i)}{\hat{p}_i(1 - \hat{p}_i)}$$

b) Ağırlıklar oluşturulur, $w_i = \hat{p}_i(1 - \hat{p}_i)$

c) Ağırlıklandırılmış backfitting algoritması kullanılarak w_i ağırlıkları ile toplamsal modelin hedef değişkenleri tahmin edilir.

3. Adım: Fonksiyonlardaki değişim önceden belirlenmiş bir tolerans seviyesinin altına düşene kadar 2. adım devam eder.

⁴⁰ Trevor Hastie, Robert Tibshirani ve Jerome Friedman, **The Elements of Statistical Learning: Data Mining, Inference and Prediction**, Second Edition, Berlin: Springer, 2008, s.300.

2.5 NONPARAMETRİK LOGİT MODELİ'NİN KATSAYI YORUMU

Nonparametrik ve semiparametrik logit modelin kurulması ve tahmin edilmesinde toplamsal modeller önemli yer tutmaktadır. Toplamsal modellerin doğrusal regresyon modelini genelleştirmek için çok kullanışlı olduğu ve marjinal değişimin, başka bir deyişle diğer değişkenler sabit tutulduğunda herhangi bir değişkenin ortalama fonksiyonu olan $m(\cdot)$ üzerindeki etkisini yorumlamaya izin verdiği kanıtlanmıştır⁴¹.

Toplamsal model yardımıyla tahmin edilen nonparametrik logit modele ilişkin katsayı yorumları parametrik modelle benzerlik göstermektedir. Fakat burada zorluk nonparametrik regresyon modelinden elde edilen düzgünleştirme parametresinin yorumlanmasıdır. Bu nedenle nonparametrik logit model için tahmin edilen sonuçlar genellikle grafikler yardımıyla yorumlanmaktadır.

Parametrik logit modeli yorumlarken bahsi geçen kavramlar (logit, odds, marjinal etkiler) nonparametrik logit modelin yorumlanması için de geçerlidir. Nonparametrik logit modele ait sonuçlardan elde edilen grafikler bu kavramlar yardımıyla kolaylıkla yorumlanabilmektedir.

⁴¹ Hardle vd. s.211.

3. SEMİPARAMETRİK LOGİT MODELİ

İkili tepki modelleri için parametrik ve parametrik olmayan tahmin yöntemleri anlatıldı. Parametrik yöntemin tahmin edilmesi ve yorumlanması kolay olsa da bu yöntem çok fazla varsayıma ihtiyaç duyduğundan sonuçların güvenilirliği giderek azalmaktadır. Parametrik olmayan yaklaşım ise hiçbir varsayımın bulunmayışından ötürü kullanışlı gibi gözükse de modelde yer alan açıklayıcı değişken sayısındaki artış çözümü zorlaştırmaktadır.

Parametrik olmayan tahminler oldukça esnektir ancak modele birkaç açıklayıcı değişken katıldığında modele ait istatistiksel hassasiyet azalır. Değişken sayısındaki artışın yol açtığı bir diğer problem ise boyutsallıktır. Araştırmacılar standart parametrik yöntemden daha fazla esneklik sağlayan fakat boyutsallık probleminin de üstesinden gelmeyi amaçlayan modeller ve tahminler geliştirmeye çalışmaktadır. Bu yöntemler genellikle parametrik ve parametrik olmayan yöntemlerin özelliklerinin birleştirilmesi ile oluşan semiparametrik yöntem olarak ifade edilmektedir.

Nonparametrik logit model için kurulan modeller yalnızca sürekli değişkenleri içermektedir. Semiparametrik logit modelin avantajı modele sürekli değişkenlerin yanı sıra kategorik değişkenlerin de dahil edilebilmesinin mümkün olması-parametrik bir şekilde-sonuçların kolay yorumlanabilmesi ve bir modelin parçalarının belirlenme olasılığının bulunmasıdır.

3.1 SEMİPARAMETRİK LOGİT MODELİ'NİN TANIMI ve YAPISI

Bağımlı değişkenin ikili değer aldığı modeller için parametrik logit modeli tahmin kolaylığı ve yorumlama açısından sağladığı avantaj nedeniyle ekonometrik çalışmalarda yaygın kullanıma sahiptir. Öte yandan semiparametrik logit model ekonometrik modeller oluşturmak için cazip bir uzlaşmayı temsil eder. Parametrik ve parametrik olmayan yaklaşımlar arasında orta seviyede varsayım yaparak

semiparametrik modeller parametrik modeldeki nispi yanlış belirlenme riskini azaltır ve daha önce bahsi geçen nonparametrik metotlara ait sakıncaları önler⁴².

Semiparametrik modeller, parametrik ve parametrik olmayan modellerin bileşenlerini birleştirerek parametrik yöntemin kolay yorumlanabilirliğini ve nonparametrik modelin esnekliğini muhafaza eder. Bunun yanı sıra semiparametrik logit model, değişkenler arasındaki fonksiyonel yapının parametrik yaklaşımdaki gibi doğrusal olduğu ve fonksiyonel yapının bilinmemesi durumunda hata dağılımı $G(\cdot)$ 'nin modelden tahmin edileceğini varsayar⁴³. Bu durumda modele ait kısımları doğrusal indeks olarak incelemek isteyebiliriz. Bu analitik nedenlerden ve ekonomik teoriden kaynaklı olabilir.

Açıklayıcı olması adına modele ait bileşenleri iki kısımda inceleyebiliriz. Modele ait değişkenlerin d -boyutlu vektörünü $Z = (Z_1, \dots, Z_p)^T$ ve $X = (X_1, \dots, X_q)^T$ olarak ayırdığımızda açıklayıcı değişkenlerin tepki değişkeni Y üzerindeki regresyonu,

$$E(Y_i/Z_i, X_i) = \beta^T Z_i + m(X_i)$$

olacak şekilde ifade edilecektir. Burada $m(\cdot)$, X vektörünün bilinmeyen çok değişkenli bir fonksiyonudur. Bu model tamamen parametrik kısım olan $\beta^T Z_i$ ve parametrik olmayan kısım olan $m(\cdot)$ 'in toplamından oluşan kısmi doğrusal model olarak adlandırılmaktadır. Kısmi doğrusal modeli bağlantı fonksiyonu ile genelleştirdiğimizde semiparametrik logit model şu şekilde ifade edilir,

$$E(Y_i/X_i) = P(Y_i = 1/X_i) = G(\beta^T Z_i + m(X_i))$$

Bu model için $G(\cdot)$ bilinen bir fonksiyon olup daha önce bahsedildiği gibi lojistik bağlantı fonksiyonudur. Burada β bilinmeyen regresyon katsayılarına ait katsayı vektörü $m(\cdot)$ ise bilinmeyen düzleştirme fonksiyonudur. Semiparametrik logit model için tepki değişkeni Y_i açıklayıcı değişkenler çifti olan (X, Z) 'ye bağlı olup parametrik bileşen olan Z üzerinde doğrusal olduğu ve parametrik olmayan bileşen için doğrusal olmadığı varsayımı geçerlidir⁴⁴.

⁴² Jean Loic Berthet ve Valentin Patilea, "A Review of Some Semiparametric Regression Models with Applications to Scoring", Proceedings of the **XIth International Symposium in Applied Stochastic Models and Data Analysis**. France: Brest, 2005, s.594.

⁴³ Tuğba Dayıoğlu, "Parametrik ve Semiparametrik Logit Modelleri İle Hanehalkı Yoksulluğunun Analizi", (Doktora Tezi, Marmara Üniversitesi SBE, 2011), s.72.

⁴⁴ Hardle, s.233.

3.2 SEMİPARAMETRİK LOGİT MODELİ'NİN TAHMİNİ

Semiparametrik logit modelde β parametrelerinin ve $m(\cdot)$ 'in tahmininde hem parametrik hem de parametrik olmayan regresyon yöntemlerinin kombinasyonu gerekmektedir. Modelde yer alan β ve $m(\cdot)$ için tahmin prosedürü iki gruba ayrılmaktadır.

Birincisi profil benzerlik tipi ikinci grup ise backfitting temelli tahmin prosedürüdür. İlk grup profil benzerlik ve genelleştirilmiş Speckman yaklaşımını kapsar. Her iki yöntem de kısmi doğrusal model için özdeştir. Backfitting tipi yaklaşım parametrik ve parametrik olmayan tahminler arasındaki değişime dayanmaktadır.

Backfitting yöntemi her ne kadar nonparametrik logit model tahmini için kullanışlı gibi gözükse de semiparametrik logit modelin iki toplamsal bileşeni olduğundan backfitting uygulaması bu model tahminleri için de geçerlidir. Bunun dışında Semiparametrik Maksimum Benzerlik ve Maksimum Score Yöntemi semiparametrik logit modelin tahmini için kullanılabilir diğer yöntemleridir.

3.2.1 Profil Benzerlik Yöntemi

Profil benzerlik yöntemi Y 'nin koşullu dağılımını veren U ve X 'in parametrik olduğu varsayımına dayanmaktadır. Bu yöntem β 'yı sabit tutar ve sabit tutulan β 'ya bağlı olarak parametrik olmayan $m_\beta(\cdot)$ fonksiyonunu tahmin etmeye başlar. $m_\beta(\cdot)$ için elde edilen tahmin sonuçları β 'nin profil benzerliğini oluşturmak için kullanılır. Elde edilen β tahminleri asimtotik normal dağılıma sahip ve asimtotik etkindir.

Parametrik olmayan $m(\cdot)$ fonksiyonu sürekli olarak $m(\cdot) = m_\beta(\cdot)$ ile tahmin edilebilir. Bu durumda profil benzerlik fonksiyonu,

$$\ell(Y, \mu_\beta, \varphi) = \sum_{i=1}^n \ell\{Y_i, G(\beta^T U_i + m_\beta(X_i)), \varphi\}$$

şeklinde tanımlanır. Burada Y ve μ_β tüm gözlemler için n boyutlu vektörleri simgeler ve

$$Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}, \quad \mu_\beta = \begin{pmatrix} G(\beta^T U_1 + m_\beta(X_1)) \\ \vdots \\ G(\beta^T U_n + m_\beta(X_n)) \end{pmatrix}$$

olarak tanımlanır. Profil benzerlik fonksiyonunun maksimizasyonu β için elde edilmek istenen tahminleri üretir. Modelin parametrik olmayan kısmı için amaç fonksiyonu,

$$\ell_H(Y, \mu_{m_\beta(T)}, \varphi) = \sum_{i=1}^n K_H(x - X_i) \ell\{Y_i, G(\beta^T U_i + m_\beta(x)), \varphi\}$$

şeklinde tanımlanır. Düzgünleştirilmiş veya lokal benzerlik X noktasında $m_\beta(X_i)$ 'i tahmin etmek için maksimize edilir. Burada $\mu_{m_\beta(T)}$, $G(\beta^T U_i + m_\beta(x))$ bileşenlerine ait vektörü simgeler. Yerel ağırlık $K_H(X - X_i)$, K çok boyutlu çekirdek fonksiyonu ve H bant genişliği matrisi ile kernel ağırlıklarını ifade eder.

Semiparametrik logit model tahmini için profil benzerlik yöntemi, logaritmik benzerlik fonksiyonunun doğrusal olmayan yapısı nedeniyle β ve $m(\cdot)$ tahminlerini üretmek için tekrarlı algoritmalara ihtiyaç duyar. Bu nedenle bu yöntemin tahmininde Newton-Raphson tipi algoritmadan faydalanılır. Tahmin edilecek parametreler vektörünü,

$$\theta = \text{vec}(\beta_i, \sigma, \rho), i = 1, 2, \dots, n$$

olacak şekilde tanımlarsak Newton-Raphson algoritması,

$$\theta^{yeni} = \theta^{eski} - \mathcal{H}_\ell^{-1} \mathcal{D}_\ell$$

şeklindeki yenileme adımlarını kullanır. Burada \mathcal{H}_ℓ , logaritmik benzerlik fonksiyonu için Hessian matrisi, \mathcal{D}_ℓ ise Gradient vektörü ifade eder. ℓ'_i ve ℓ''_i sırasıyla benzerlik fonksiyonunun birinci ve ikinci türevini simgelemek koşuluyla semiparametrik logit model için profil benzerlik tahmini şu adımlardan oluşur,⁴⁵

1. Adım : Modeldeki tüm açıklayıcı değişkenler için gradient vektör \mathcal{D}_ℓ (ve \mathcal{H}_ℓ Hessian matris) hesaplanır.

2. Adım :

$$\check{u}_j = u_j - \frac{\sum_{i=1}^n \ell''_i(\beta^T u_i + m_j) K_h(X_i - X_j) u_i}{\sum_{i=1}^n \ell''_i(\beta^T u_i + m_j) K_h(X_i - X_j)} \quad j = 1, 2, \dots, n$$

⁴⁵ Irene Bertsek ve Marlene Müller, "Productivity Effects of IT-Outsourcing: Semiparametric Evidence for German Companies", Stephan Sperlich, Wolfgang Hardle ve Gökhan Aydın (Ed.), **The Art of Semiparametrics** içinde (130-154), Berlin: Springer Science & Business Media, 2006, s.141-142.

değerleri hesaplanır.

3. **Adım :** \mathcal{D}_ℓ matrisinde nonparametrik değişken için değerler $\beta^T \check{u}_i + m_i$ ile değiştirilir ve daha sonra yeniden parametrik olmayan terim \check{u}_i ile değiştirilerek gradient tahmini $\check{\mathcal{D}}_\ell$ (ve Hessian $\check{\mathcal{H}}_\ell$) elde edilir. θ için güncelleme adımı $\check{\mathcal{D}}_\ell$ (ve $\check{\mathcal{H}}_\ell$) kullanarak iteratif optimizasyon prosedürü ile gerçekleştirilir.
4. **Adım :** m_j için güncelleştirme adımı,

$$m_j^{yeni} = m_j^{eski} - \frac{\sum_{i=1}^n \ell'_i(\beta^T u_i + m_j) K_h(X_i - X_j) u_i}{\sum_{i=1}^n \ell''_i(\beta^T u_i + m_j) K_h(X_i - X_j)} \quad j = 1, 2, \dots, n$$

olacak şekilde hesaplanır. Tahmin adımları β ve $m(\cdot)$ için tahminler yakınsayınca kadar devam eder.

3.2.2 Semiparametrik Maksimum Benzerlik Tahmini

Daha önce de belirtildiği üzere bağımlı değişkenin 0 ve 1 gibi iki değer aldığı durumda en çok benzerlik problemini çözmek için logaritmik benzerlik fonksiyonu olan,

$$\log \ell(\beta) = L(\beta) = \sum_{i=1}^n y_i \log(G(x_i' \beta)) + (1 - y_i) \log(1 - G(x_i' \beta))$$

eşitliği kullanılır. Bu durumda $G(x_i' \beta) = P(Y = 1/X)$ tahmini, G biliniyorsa β 'nin asimptotik etkin tahmincisi en çok benzerlik tahmin yöntemiyle elde edilebilir. Semiparametrik ilişkinin varlığı söz konusu olduğunda G bilinmez. Semiparametrik yöntemde G için herhangi bir varsayım yapılmadığından en çok benzerlik yöntemi için logaritmik benzerlik fonksiyonunda yer alan G yerine geçebilecek G_N gibi bir tahminciyle değiştirilmesi gerekmektedir.

Klein ve Spady β 'nin bir tahmincisi olarak $G(x_i' \beta)$ 'nin yerine $G_N(x_i' \beta)$ şeklinde tanımlanan ve $G(x_i' \beta)$ 'nin parametrik olmayan bir tahmininin quasi-benzerlik fonksiyonunu maksimize ederek elde eden b_{KS} 'yi önermektedir⁴⁶. Semiparametrik logit model tahmini için böyle bir G_N tahminci parametrik olmayan yöntemlerin kullanılması

⁴⁶ Roger W. Klein ve Richard H. Spady. "An Efficient Semiparametric Estimator for Binary Response Models", *Econometrica: Journal of the Econometric Society*, Cilt.61, Sayı.2, (Mart,1993), s.388.

ile elde edilebilecektir. G_N , y koşullu $x b_{ks}$ yoğunluğunun parametrik olmayan çekirdek tahmini tarafından şu şekilde hesaplanır,

$$P_N = N^{-1} \sum_{n=1}^N y_n$$

P_N , bağımlı değişkenin “1” olarak tanımlandığı durum için verilen cevapların oranını gösterir.

$$G_N(v) = \frac{P_N g_N(v/y = 1)}{P_N g_N(v/y = 1) + (1 - P_N) g_N(v/y = 0)}$$

$g_N(. / y)$ ifadesi $g(. / y)$ ’nin çekirdek tahminini ve $x\beta$ ’nin koşullu yoğunluğudur. Bu tahminde yer alan terimler,

$$g_N(v/y = 1) = (NP_N h_N)^{-1} \sum_{n=1}^N y_n K\{(v - x_n b_{ks})/h_N\}$$

ve

$$g_N(v/y = 0) = (N(1 - P_N) h_N)^{-1} \sum_{n=1}^N (1 - y_n) K\{(v - x_n b_{ks})/h_N\}$$

olarak tanımlanmaktadır. Tahmin için kullanılan bu eşitliklerde K kernel fonksiyonunu h_N ise düzgünleştirme parametresini ifade etmektedir. Elde edilen G_N tahmincisi benzerlik fonksiyonunda yerine konulursa modele ait parametreler vektörü olan β ’yı tahmin etmiş oluruz. Bilinmeyen parametre vektörünün tahmin edilmesinden sonra bilinmeyen G fonksiyonu tahmin edilir.

Bu tahminler gözlemlerin bağımlı değişkende “1” olarak tanımlanan düzeye ait olma olasılıklarını vermektedir⁴⁷. Tahmin işlemi sırasında öncelikle $x_i' \beta$ değerleri hesaplanır. Daha sonra bağımlı değişkenin $x_i' \beta$ üzerine tek değişkenli parametrik olmayan regresyon uygulaması ile olasılıkların tahminine ulaşılarak tahmin işlemi tamamlanmış olur.

⁴⁷ Özge Akkuş, Serdar Demir ve Durdu Karasoy, “İki Düzeyli Bağımlı Değişken Modelinin Yarı Parametrik Tahmini”, *İstatistikçiler Dergisi*, Sayı.1, s.138.

3.2.3 Genelleştirilmiş Speckman Tahmini

Profil benzerlik tahmini, kısmi doğrusal model durumunda, diğer bir deyişle G bağlantı fonksiyonu tanımlı olduğunda ve normal dağılıma sahip Y_i için türetilmesi kolaydır. Burada benzerlik fonksiyonunun birinci dereceden türevi,

$$\ell'_i = -(Y_i - U_i^T \beta - m_j) / \sigma^2$$

ve ikinci dereceden türevi,

$$\ell''_i = -1 / \sigma^2$$

olarak ele alındığında bu öğeler tarafından tanımlanan daha basit bir S düzgünleştirme matrisi,

$$(S)_{ij} = \frac{K_h(X_i - X_j)}{\sum_{i=1}^n K_h(X_i - X_j)}$$

elde edilir. Bu durumda m_j için güncelleştirme adımı,

$$m_j^{yeni} = \frac{\sum_{i=1}^n (Y_i - U_i^T \beta) K_h(X_i - X_j)}{\sum_{i=1}^n K_h(X_i - X_j)}$$

şeklinde basitleştirilir. Bunu matris notasyonuyla,

$$m^{yeni} = S(Y - U\beta)$$

şeklinde ifade edebiliriz. Y için vektör gösterimi aynı olmak koşulu ile nonparametrik fonksiyon tahmini,

$$m^{yeni} = (m_1^{yeni}, m_2^{yeni}, \dots, m_n^{yeni})$$

olacaktır. $\tilde{U} = (I - S)U$ ve $\tilde{Y} = (I - S)Y$ eşitlikleri kullanılarak β parametresi için tahminci,

$$\beta^{yeni} = (\tilde{U}^T \tilde{U})^{-1} \tilde{U}^T \tilde{Y}$$

olacak şekilde elde edilir. Bu tahmin için herhangi bir iterasyona gerek duyulmadan parametreler tahmin edilebilir.

Semiparametrik logit model tahmini için Speckman yaklaşımı genelleştirilmiş doğrusal model tahmininde kullanılan IRLS yöntemiyle birleştirilir. Genelleştirilmiş doğrusal model her iterasyon adımında ayarlanmış bağımlı değişkene ağırlıklı en küçük

kareler uygulanarak elde edilir. Bu durumdan hareketle IRLS ayarlanmış bağımlı değişkene ağırlıklı kısmi doğrusal bir uyumla değiştirilerek semiparametrik logit model tahmini gerçekleştirilecektir. Ayarlanmış bağımlı değişken aşağıdaki gibi tanımlanmaktadır,

$$Z = U\beta + m - W^{-1}v$$

Burada $v = (\ell'_1, \ell'_2, \dots, \ell'_n)$ benzerlik fonksiyonunun birinci dereceden türev vektörü $W = \text{diag}(\ell''_1, \ell''_2, \dots, \ell''_n)$ ise ikinci dereceden türev matrisin köşe elemanlarını ifade eder. Bu durumda düzgünleştirme matrisini yeniden tanımlarsak,

$$(S)_{ij} = \frac{\ell''_i K_h(X_i - X_j)}{\sum_{i=1}^n \ell''_i K_h(X_i - X_j)}$$

olacaktır. $\tilde{U} = (I - S)U$ ve $\tilde{Z} = \tilde{U}\beta + m - W^{-1}v$ eşitlikleri kullanılarak matris gösterimindeki her bir iterasyon adımı için tahminleri aşağıda Tablo 3'deki gibi özetlemek mümkündür.

Tablo 3
Genelleştirilmiş Speckman Tahmini İçin İterasyon Adımları

<p>β tahmini için ayarlama adımı</p> $\beta^{yeni} = (\tilde{U}^T W \tilde{U})^{-1} \tilde{U}^T W \tilde{Z}$ <p>m tahmini için ayarlama adımı</p> $m^{yeni} = S(Z - U\beta)$
--

3.2.4 Backfitting Tahmin Yöntemi

Backfitting yöntemi başlangıçta toplamsal modelin tahmini için tekrarlı bir algoritma olarak önerilmiştir. Yöntemin ana fikri toplamsal bileşenleri kısmi kalıntılar üzerinde ayrı ayrı tahmin etmektir.

Ekonometrik modeller için açıklayıcı değişkenlerin birbiriyle ilişkisiz olduğu durumda çoklu parametrik olmayan terimlerin uyumu basittir. Fakat açıklayıcı değişkenler arasında korelasyon olmaması durumu oldukça nadir görülmektedir.

Dolayısıyla açıklayıcı değişkenler arasındaki kovaryansları açıklayan toplamsal ya da semiparametrik modelin tahmini için bir yonteme ihtiyaç duyulmaktadır.

Backfitting algoritması semiparametrik model tahmini için parametrik ve parametrik olmayan terimler arasındaki korelasyonu hesaba katacak şekilde tasarlanmıştır. Bu yontem kısmi regresyon fonksiyonlarını kullanmaktadır.

Backfitting yontemi kısmi doğrusal regresyon modellerine doğrudan uygulanabilirken semiparametrik regresyon tahmini için tekrarlı algoritmaları kullanır. Bu nedenle semiparametrik modelin tahmininde başlangıç deęerlerine ihtiyaç duyulmaktadır. Bir kez daha semiparametrik logit modeli,

$$P(Y_i = 1/X_i) = G(\beta^T U_i + m(X_i))$$

olarak tanımlarsak backfitting algoritması için başlangıç deęerleri,

$$\beta = 0 \text{ ve } m(X_i) = G^{-1}(Y_i)$$

olarak belirlenir. İkili baęımlı deęişken için,

$$m_i = G^{-1}((Y_i + 0,5)/2)$$

olacak şekilde düzeltilir.

Bu yontem için düzgünleştirme matrisi olan S , W matrisi ve v vektörü yanı sıra ayarlama deęişkeni Speackman yaklaşımında olduęu gibi tanımlanır. Tahmin yontemine ilişkin her bir iterasyon adımı için tahminçiler Tablo 4'de özetlenmiştir.

Tablo 4

Backfitting Tahmini İçin İterasyon Adımları

β tahmini için ayarlama adımı

$$\beta^{yeni} = (U^T W \tilde{U})^{-1} U^T W \tilde{Z}$$

m tahmini için ayarlama adımı

$$m^{yeni} = S(Z - U\beta)$$

3.2.5 Maksimum Score Tahmini

Semiparametrik logit model için tahminciler genellikle indeks modeline dayanır. Y_i , i . bireyin tepkisini tanımlayan indeks olup y^* gizli değişkeni tarafından belirlenir. $y^* = x_i' \beta + u_i$ olduğu düşünüldüğünde model,

$$y_i = \begin{cases} 1, & 1(x_i' \beta + u_i > 0) \\ 0, & \text{diğer durumlar} \end{cases}$$

olacaktır. Burada $1(\cdot)$ gösterge fonksiyonunu simgeler ve parantez içindeki ifade doğruysa $1(\cdot) = 1$ olarak tanımlanır. Bu model için u_i 'nin dağılımının bilinmediği varsayılmakta ve u_i 'nin bilinmeyen formda hemen hemen tesadüfi bir sabit varyansa sahip olması sağlanmaktadır⁴⁸.

Hata terimi bilinmediğinden $u_i = 0$ olarak belirlenmiş ve tahmin edilen y_i değeri $1(x_i' \beta > 0)$ olarak tanımlanmıştır. Bu şartlar altında doğru tahmin edilen y_i 'lerin sayısının bir skoru,

$$S_N(\beta) = \sum_{i=1}^n y_i 1(x_i' \beta > 0) + (1 - y_i) 1(x_i' \beta \leq 0)$$

şeklinde olacaktır. $1(x_i' \beta > 0)$ olduğunda $y_i = 1$ veya $1(x_i' \beta \leq 0)$ ise $y_i = 0$ olduğu durumda doğru tahminler gerçekleşir.

Burada bir tahmin problemi olarak $1(x_i' \beta > 0)$, β için türevlenebilir nitelikte olmadığından medyan regresyonun ikili yanıt modeline uygulanması önerilmiştir. Tahmin için, y_i ve $Medyan(y_i/x_i)$ arasındaki mutlak farkların toplamının minimize edilmesi gerekmektedir. $Medyan(u_i/x_i) = 0$ ise $Medyan(y^*/x_i) = x_i' \beta$, dolayısıyla $Medyan(y_i/x_i) = 1(x_i' \beta > 0)$ olacaktır. Bu düzenlemeler ile $S_N(\beta)$ için maksimum score tahminine ilişkin amaç fonksiyonunu yeniden yazarsak,

$$S_{ms}(\beta) = \sum_{i=1}^n (2y_i - 1) 1(x_i' \beta > 0) + N - \sum_{i=1}^n y_i$$

⁴⁸ Michael Gerfin, "Parametric and Semi-Parametric Estimation of the Binary Response Model of Labour Market Participation", **Journal of Applied Econometrics**, Cilt.11, Sayı.3, (Mayıs-Haziran, 1996), s.325.

olacaktır. Burada ikinci terim β içermediğinden göz ardı edilebilir. $S_N(\beta)$, β 'nin süreksiz bir fonksiyonu olan basamak fonksiyondur. Düzgünleştirilmiş maksimum skor tahmin edicisi bu gösterge fonksiyonunun iki kez türevlenebilir olan bir fonksiyon ile değiştirilmesi sonucunda elde edilir⁴⁹. Düzgünleştirilmiş maksimum skor tahmini için amaç fonksiyonu,

$$S_{sms}(\beta) = \sum_{i=1}^n (2y_i - 1)K(x_i' \beta / h_n)$$

şeklinde olacaktır. Maksimum skor tahmincisi, karmaşık bir sınırlayıcı dağılıma sahiptir ve amaç fonksiyonu kesikli olduğu için türetilmesi güçtür. Bu nedenle düzgünleştirilmiş maksimum score yöntemi ile türevlenebilir bir fonksiyon elde edilirken süreksiz fonksiyonları da düzgünleştirmiş oluruz.

Amaç fonksiyonunda yer alan K çekirdek tahmin edicisi değil genel gösterim açısından parametrik olmayan bir yoğunluk fonksiyonunu ifade etmektedir. Fakat parametrik olmayan bir yöntem olduğundan K çekirdek yoğunluğu olarak da düşünülebilir.

3.3 SEMİPARAMETRİK LOGİT MODELİ'NİN ANLAMLILIK TESTİ

Semiparametrik logit model için model anlamlılığı testi iki farklı durumu yansıtmaktadır. Bunlardan ilki açıklayıcı değişkenlerin bağımlı değişken üzerinde sıfırdan farklı anlamlı bir etkisinin olup olmadığını incelerken ikinci yaklaşım bir değişken için parametrik olmayan tahminlerin parametrik bir terimin kullanımından farklı bir model uyumu geliştirip geliştirmediğini belirler.

Semiparametrik modelde yer alan parametrik değişkenlerin bağımlı değişken üzerindeki anlamlılığı t testi ile incelenirken parametrik olmayan değişkenler için herhangi bir anlamlılık testi bulunmamaktadır⁵⁰. Bu nedenle nonparametrik değişken için kısıtlı F -testi yardımı ile model karşılaştırması yapılarak bu değişkenin anlamlılığı test edilebilecektir.

⁴⁹ Joel L. Horowitz, **Semiparametric and Nonparametric Methods in Econometric**, New York: Springer, 2009, s.109.

⁵⁰ Elif Öztürk, "Parametrik Olmayan ve Yarı Parametrik Panel Veri Modelleri: Çevresel Kuznets Eğrisinin Analizi", (Doktora Tezi, Marmara Üniversitesi SBE, 2011), s.118.

Bir diğ er ay rım test istatisti ğ ine iliřkin da ğ ılım ın bilinip bilinmemesi üzerinedir. Benzerlik oran ı temeline dayanan testler için da ğ ılım ın řekli yaklařık χ^2 da ğ ılım ını izlemeye devam ederken semiparametrik logit model tahminlerinde kullanılacak dü z ğ ünleřtirme parametresi test istatisti ğ ine iliřkin da ğ ılım problemi ile karřılařılmasına neden olur. Parametrik olmayan terimlerin tahmininde otomatik dü zeltme teknikleri kullanılıyorsa, test istatisti ğ i da ğ ılım ının yaklařık olarak χ^2 da ğ ılım ını izledi ğ i dü ř ünölür, ancak ne derece yakın oldu ğ u tam olarak bilinmemektedir⁵¹. Bu duruma alternatif olarak herhangi bir da ğ ılım varsayımına dayanmayan ve yeniden ö rnekleme yöntemi olan bootstrap yöntemi ile bir p-de ğ eri hesaplanarak test sı namas ı gerç ekleřtirilebilir.

3.3.1 Benzerlik Oranı Testi

Sapma, genelleřtirilmiř modellerde kalıntı kareleri toplam ı rolü üstlenir ve modellerin karřılařtırılması yanı sıra model uyum iyili ğ i ölçütünün de ğ erlendirilmesi için kullanılır.

Parametrik logit model için asimtotik da ğ ılım teorisi iyi bilinmektedir. Parametrik logit model için yuvalanmıř modellerin test edilmesi sapmalar aracılı ğ ıyla gerç ekleřtiriliyorsa test istatisti ğ i asimtotik χ^2 da ğ ılım ına sahip olur. Da ğ ılım teorisinin geliřtirilmemiř olmasıyla birlikte parametrik olmayan ve toplamsal modeller için sapma, modelleri de ğ erlendirmek için hala bir araç olarak kullanılabilir⁵². Bö yle bir yaklařım izlendi ğ inde χ^2 da ğ ılım ı karřılařtırma modelleri için gayri resmi bir řekilde referans olarak kullanacaktır. Bu durumda yapılan analizlere ait sonuçlar dikkatle de ğ erlendirilmelidir. Ö rne ğ in sı fır hipotezini güvenle reddedilmek için tahmin edilen p olasılıklarının mümkün oldu ğ unca küçük olması gerekmektedir⁵³.

Semiparametrik logit modeli,

$$\hat{\mu}_i = G(Z_i^T \hat{\beta} + \hat{m}(X_i))$$

⁵¹ Simon Wood, **Generalized Additive Models: An Introduction With R**, Boca Raton: Chapman & Hall/CRC, 2006, s.200.

⁵² Hastie ve Tibshirani, s. 155.

⁵³ Münevver Turanlı ve Seda Ba ğ datlı, "Semiparametrik Regresyon", **Ö neri Dergisi**, Cilt.9, Sayı.35, (Ocak 2011), s.212.

ve parametrik logit modeli,

$$\tilde{\mu}_i = G(Z_i^T \tilde{\beta} + X^T \tilde{\gamma} + \tilde{\gamma}_0)$$

olarak tanımlarsak, sapma veya benzerlik oranı istatistiği,

$$LR = D(Y_i, \tilde{\mu}_i, \varphi) - D(Y_i, \hat{\mu}_i, \varphi)$$

olacaktır. Bu test problemi için bir test istatistiği, parametrik benzerlik oranı testine bağlı olarak semiparametrik teste genelleştirilmektedir. Temel modele bağlı olarak gerçekleştirilecek test prosedürü için nonparametrik değişkenin test edilmesinde hipotezler,

H₀: Doğrusal fonksiyon

H_A: Herhangi bir düzgünleştirme fonksiyonu

şeklinde kurulur. Testin gerçekleştirilmesi için öncelikle temel alınacak semiparametrik logit modelin tahmin edilmesi gerekmektedir. Daha sonra temel modelde düzgünleştirme fonksiyonu olarak yer alan değişkenin doğrusal şekilde ele alındığı yeni bir model tahmini gerekmektedir.

Tahmin sonrası her iki modelin sapma değerlerinin karşılaştırıldığı test istatistiği hesaplanarak modellere ilişkin değerlendirme yapılır. İki model arasındaki serbestlik derecesi farkına eşit serbestlik dereceli test istatistiği, bir χ^2 dağılımına sahiptir. Sıfır hipotezi altında test edilecek değişken için hipotez sınaması ilgili değişkenin modele doğrusal fonksiyon şeklinde dahil edilmesi gerektiğini ifade ederken alternatif hipotez değişkenin düzgünleştirme fonksiyonu olarak modele alınması gerektiğini ifade etmektedir.

3.3.2 Kısıtlı F-Testi

Bu test nonparametrik logit ve semiparametrik logit regresyon modelleri için hipotezlerin test edilmesine imkân sağlar. Parametrik terim için R matrisinden elde edilen standart hatalar hipotez testleri için çıkarımda bulunmaya izin verir. Parametrik olmayan terim için kısıtlı F testi açıklayıcı değişkenin bağımlı değişken üzerinde

sıfırdan farklı anlamlı bir etkisinin olup olmadığını araştırır. F testi kalıntı kareleri toplamına bağlı olup test istatistiği,

$$F = \frac{(RSS_0 - RSS_1)/(tr(R) - 1)}{RSS_1/df_{res}}$$

olarak tanımlanır. Bu test için önce semiparametrik logit model tahmini yapılır daha sonra parametrik olmayan değişkenin modelden çıkarılması ile kısıtlı model tahmin edilir. RSS_0 kısıtlı modele ait kalıntı kareleri toplamı, RSS_1 ise semiparametrik logit modelin kalıntı kareleri toplamını ifade eder.

Kısıtlı modelin kalıntı kareleri toplamı ile kısıtsız modelin kalıntı kareleri karşılaştırılarak parametrik olmayan değişkenin anlamlı olup olmadığına karar verilir. Test istatistiğinin paydasında yer alan serbestlik derecesi df_{res} , kısıtlı modele ait serbestlik derecesinden kısıtsız modelin serbestlik derecesinin farkı alınarak elde edilir. Test istatistiği df_{res} ve kısıtsız modele ait serbestlik derecesi ile yaklaşık F dağılımına sahiptir.

3.3.3 Bootstrap Yöntemi ile Doğrusal Olmama Testi

Parametrik olmayan ve semiparametrik regresyon modelleri için çıkarımın bir bileşeni parametrik bir alternatife karşı spline veya yerel polinom tahminini test etmektir.

Smoothing spline için test istatistiğinin dağılımı yalnızca yaklaşık değerdir ve otomatik düzeltme yöntemleri kullanıldığında test istatistiğinin dağılımı hakkında çok az bilgi elde edilir. Aşırı parametrelendirilmiş kübik splinlerin kullanımı p-değerine ilişkin tahminlerin doğru olmasını sağlayan bir yöntemdir.

Bootstrap, semiparametrik modele göre parametrik bir null modelin sınanmasına da olanak tanır. Karşılaştırılan iki modelin test istatistiğini bootstrap yöntemi kullanarak herhangi bir dağılım varsayımına dayanmayan bir p-değerinin hesaplanması ile test edebiliriz. Bootstrap yöntem ile elde edilecek p-değeri için artıkların yeniden örnekleme yöntemi kullanılarak hesaplanması gerekmektedir. Kalıntıların yeniden örnekleme yöntemi kullanılarak hesaplanması ile semiparametrik modeli test ettiğimiz null modelin

hata yapısına yaklaşmayı hedefleriz. Test istatistiğinin hesaplanması için bootstrap algoritma adımları aşağıdaki gibidir,

- 1- Yalnızca sabit terimin yer aldığı parametrik bir model (null model) tahmin edilir
- 2- Semiparametrik model tahmin edilir
- 3- Modeller arasındaki benzerlik oranı testinden test istatistiği olan t değeri hesaplanır
- 4- Null modelinden kalıntılar ε hesaplanır
- 5- Yeniden örnekleme ile kalıntılarda ε^* dönüştürme yapılır
- 6- Bu dönüştürme için $Y^* = \hat{Y} + \varepsilon^*$ modeli kullanılır ve model artıkları hesaplanır
- 7- Bağımlı değişken olarak Y^* 'ın kullanımı ile parametrik ve semiparametrik model tahmini yapılır
- 8- t^* değerleri hesaplanır ve test istatistiği Y^* 'ı tahmin eden modellerden hesaplanır
- 9- Bu işlem B defa tekrar edilir ve p değeri,

$$\frac{[1 + \#(t^* \geq t)]}{(B + 1)}$$

olarak hesaplanır. Burada $\#(t^* \geq t)$, t^* 'ın orijinal t istatistiğine eşit veya bu sınırın üzerine çıkma sayısıdır. Algoritma, benzerlik oranı testinden p -değeri ile karşılaştırılabilen kolayca yorumlanabilir bir p -değerini döndürür.

Bootstrap testi beklendiği gibi daha katı olma eğilimindedir. Doğrusal olmama durumu için hipotez testi kritik ise, standart testi bootstrap teste karşı kontrol etmek iyi bir uygulamadır. Bununla birlikte, genel teşhis için standart benzerlik oranı testi yararlı bir rehber olmaya devam etmektedir.

3.4 SEMİPARAMETRİK LOGİT MODELİ'NİN UYUM İYİLİĞİ

Ekonometrik uygulamalarda geliştirilmiş modeller ile çalışılırken sıradan doğrusal modellerin kalıntı karelerine benzer bir şekilde yorumlanabilecek bir ölçüye ihtiyaç duyulmaktadır. Sapma, geliştirilmiş modeller için kalıntı kareleri toplamı rolü üstlenir ve modellerin uyum iyiliği ölçüsünün değerlendirilmesi için kullanılır.

Bunun yanı sıra semiparametrik logit modelin parametrik logit modelden daha iyi performans gösterip göstermediğini değerlendirmek için çeşitli istatistiksel göstergelerin hesaplanmasında önemli bir işleve sahiptir.

Model karşılaştırmalarında kullanılacak bu göstergeler belirlilik katsayısı ve hipotez testlerinin gerçekleştirilmesi için kullanılacak yaklaşık F-test istatistiğidir. D sapmayı ifade etmek suretiyle semiparametrik logit model için model karşılaştırılmasında kullanılacak yaklaşık F test istatistiği,

$$F = \frac{D(Kısıtlı) - D(Kısıtsız)}{[df(Kısıtsız) - df(Kısıtlı)]\emptyset}$$

olmaktadır. \emptyset dağılım parametresi olup binom dağılımı için 1'e eşittir. Bu test istatistiği pay ve paydada yer alan serbestlik dereceli F dağılımı ile karşılaştırılarak sıfır hipotezi altında karar verilir.

Semiparametrik logit model için bir diğer model uyum iyiliği ölçüsü olan belirlilik katsayısı yine sapmalar aracılığıyla hesaplanmaktadır. Burada $D(Null)$ yalnızca sabit terim içeren bir modelin sapması iken $D(Kalıntı)$ kalıntı sapma yani tahmin edilen modelin sapması olarak tanımlanmaktadır. Semiparametrik logit model için bu iki sapma değerinin birbirine oranlanması ile belirlilik katsayısı,

$$R^2 = 1 - \frac{D(Kalıntı)}{D(Null)}$$

olacak şekilde hesaplanmaktadır. Tahmin edilen modeller karşılaştırılırken R^2 değeri yüksek olan modelin veriyi daha iyi açıkladığı düşünülmektedir. Fakat daha önce bahsettiğimiz gibi parametrik logit model için hesaplanan R^2 değeri model uyum iyiliği için zayıf bir ölçüdür.

Semiparametrik logit model için de bağımlı değişkenin ikili yapısı gereği bu ifadenin geçerli olduğu düşünülebilir ve parametrik logit modelde olduğu gibi semiparametrik logit modele ait $Pseudo - R^2$ değeri hesaplanmak istenebilir. Parametrik logit model için bu değer logaritmik benzerlik oranları kullanılarak,

$$Pseudo - R^2 = 1 - \frac{1}{1 + 2(\ln L1 - \ln L0)/N}$$

şeklinde hesaplanmaktadır. Esasen bir modelin sapması logaritmik benzerlik değerlerinin farkının eksi iki katı ile çarpılmış halidir⁵⁴. Bu nedenle benzerlik oranı ile sapma arasındaki ilişki,

$$D(Null) - D(Kalıntı) = 2(\ln L1 - \ln L0)$$

olarak ifade edilebilir. Bu ilişkidен hareketle $Pseudo - R^2$ eşitliğinin paydasında yer alan $2(\ln L1 - \ln L0)$ teriminin yerine sapmalar arasındaki farkı koyarsak semiparametrik logit model için $Pseudo - R^2$ değeri,

$$Pseudo - R^2 = 1 - \frac{1}{1 + [D(Null) - D(Kalıntı)]/N}$$

olarak elde edilir. Elde edilen bu değerin model karşılaştırmasında kullanılması yerinde olacaktır.

Semiparametrik logit modelin genel olarak model uyum iyiliğinin yanı sıra parametrik olmayan her bir toplamsal bileşenin uyumunu incelemek mümkündür. Bunu tahmini değerlere karşı açıklayıcı değişkenin gözlenen değerleri için çizilecek grafik çizimi ile gerçekleştirebiliriz. Grafik çizimi ile elde edilen nonparametrik eğrilerin gözlemlenen değerlere ne kadar iyi uyum sağladığını incelemek mümkündür.

Kısmi kalıntılar açıklayıcı değişkenlerin araştırılması için yararlıdır. Kısmi kalıntılar bütün model için kalıntıların toplamını ve ilgili açıklayıcı değişken için tahmin edilmiş düzgünleştirme fonksiyonunu ifade eder. Bu nedenle grafik çizimlerinde kısmi kalıntılar açıklayıcı değişken değerlerine karşı çizilir. Grafik çizimi için varyans değerlerinin ve güven bantlarının hesaplanması gerekmektedir. Parametrik olmayan regresyon için tahmini değerleri,

$$\hat{m} = Sy$$

olarak ifade edebiliriz. Burada \hat{m} tahmin edilmiş değerler vektörünü, S açıklayıcı değişkenlerin değerlendirme noktasında tahmin için uygun ağırlıklardan oluşan sıraları içeren bir düzgünleştirme matrisini, y ise açıklanan değişken için gözlem vektörünü göstermektedir. Parametrik olmayan kalıntı kareleri toplamı bu eşitlikten,

⁵⁴ Ruppert ve Diğerleri, s.209.

$$RSS = \sum (y_i - \hat{m}(x_i))^2$$

olarak elde edilir. Kalıntı karelerini kullanılarak hata varyansı,

$$\hat{\sigma}^2 = \frac{RSS}{df}$$

şeklinde hesaplanır. Eşitliğin paydasında yer alan df serbestlik derecesi olup doğrusal model için projeksiyon matrisinin izine eşittir. Bunun yaklaşık bir versiyonu parametrik olmayan regresyon için,

$$df = tr(S)$$

şeklinde oluşturulabilir. Hata varyansını kullanarak x_i 'nin herhangi bir değerindeki değişkenlik miktarını ölçen standart hataları hesaplayabiliriz. $m(x)$ bir değerlendirme noktasındaki $m(\cdot)$ 'in tahmini değerlerini simgeliyorsa varyans değerleri,

$$var(\hat{m}) = SS^T \sigma^2$$

olarak hesaplanmaktadır. Bu matrisin diyagonal elemanlarının karekökü x 'e ait her bir değerlendirme noktası için standart hataları üretir.

Parametrik olmayan değişken için elde edilen nonparametrik eğri üzerindeki her noktaya iki standart hata eklenip çıkarılarak eğrideki tahmini değişimi ifade eden güven bantlarını elde etmiş oluruz. Güven bantları ile birlikte çizilen grafik, bağımlı değişken ile açıklayıcı değişken arasındaki ilişkiyi gösteren bir dağılım grafiğine benzemektedir. Bu grafiği inceleyerek uyumun iyi olup olmadığına karar verebiliriz. Güven bantlarının geniş olmayan bir aralıkta ve nonparametrik eğri ile aynı doğrultuda olması tahminlerin istatistiksel olarak anlamlı ve güvenilir olduğu anlamına gelecektir⁵⁵. Bunun yanı sıra uyum iyi olduğunda kısmi kalıntılar tahmin edilen eğri çevresinde rastgele dağılacaktır.

⁵⁵ Münevver Turanlı ve Seda Bağdatlı Kalan, "Site İçerisindeki Daire Fiyatlarını Etkileyen Unsurların Semiparametrik Regresyon Analizi İle Belirlenmesi", **İstanbul Ticaret Üniversitesi Sosyal Bilimler Dergisi**, Cilt.11, Sayı.21, 2012, s.400.

4. UYGULAMA: İÇ GÖÇ SONRASI BİREYLERİN İŞGÜCÜNE KATILIMINI ETKİLEYEN FAKTÖRLER

4.1 ARAŞTIRMANIN KONUSU ve AMACI

Türkiye'nin Cumhuriyet tarihine kadar uzanan iç göç hareketleri geçmişte olduğu gibi günümüzde de önemli bir konu olarak karşımıza çıkmaktadır. Cumhuriyetin kuruluşundan başlayarak çeşitli karakteristik özellikler gösteren göç hareketleri toplumun ekonomik, siyasi ve sosyal yapısında değişime neden olmaktadır. Bu değişimden işgücü ve istihdam piyasaları da etkilenmektedir. İşgücü piyasası soyut bir kavramdır ve anlaşılması insan/işçi unsurunun bu piyasadaki konumunun tespit edilmesi ile mümkündür. İşgücü piyasasının merkezinde yer alan işgücü faktörü genel yapısı itibarıyla değişkendir. Değişkenliğin temel dinamiklerinden birisi de göçlerdir. Bu bağlamda göçün işgücü piyasalarında meydana getirdiği değişimin incelenmesi bu çalışmanın konusunu oluşturmaktadır. Göç eden bireylerin işgücüne katılımı üzerinde etkili olan bir dizi faktör bulunmaktadır. Bu faktörlerin tespit edilmesi işgücü piyasası sorunları üzerinde göçün etkili olup olmadığının anlaşılmasına olanak sağlayacaktır.

4.2 ARAŞTIRMANIN KAPSAMI VE KISITLARI

Bu çalışma göç eden bireyler üzerine odaklanmaktadır. Bu nedenle göç eden nüfusun belirlenmesi gerekmektedir. TÜİK'in Nüfus ve Konut Araştırması'nda araştırma tarihindeki daimi ikametgâhın bulunduğu yerleşim yeri ile araştırmadan bir yıl önceki daimi ikametgâhın bulunduğu yerleşim yeri farklı olan kişiler, göç eden nüfus olarak tanımlanmıştır⁵⁶. Burada daimi ikametgâh; 12 ay ve daha uzun süredir oturlan veya toplam 12 ay ve daha uzun süreyle oturulma niyetinde olunan adres olarak ifade edilmektedir. Bu tanıma ek olarak süresi ne olursa olsun askerlik ve hükümlülük nedeniyle yerleşim yerini değiştirmiş kişiler ile yatılı bölge okulunda ve benzeri ortaöğretim yurtlarında okumak üzere başka bir yerleşim yerinde bulunmuş kişiler göç kapsamına girmediğinden bu çalışmanın da kapsamı dışında kalmaktadır. Göçün ilgili

⁵⁶ Türkiye İstatistik Kurumu, **Nüfus ve Konut Araştırması 2011**, No:4030, Ankara: Türkiye İstatistik Kurumu Yayını, 2013, s. 20.

tınımı gereği, göçe ilişkin sorular, kurumsal yerler uygulamasında kışla, cezaevi ve islahevinde kalan kişilere sorulmamış, bu kişiler göç etmemiş kabul edilmiştir.

Ayrıca iç göçü tanımlarken zaman bakımından bir yıllık daimi ikâmet değişikliği göç olarak kabul edilmiştir. Bu nedenle araştırmada kullanılan veriler ile bir dönem analizi gerçekleştirilmesi mümkün olmadığından araştırma sonuçları yalnızca 2011 yılı özelinde değerlendirilmektedir.

4.3 ARAŞTIRMANIN YÖNTEMİ

Bu çalışmada Türkiye’de 2011 yılında iç göç sürecine katılan bireylerin işgücüne katılmasını etkileyen faktörleri belirleyebilmek amacıyla nitel tercih modellerinden parametrik ve semiparametrik logit modeli kullanılmıştır. Bu yöntemde sayısal olarak gözlenmiş olan değişkenler öncelikle parametrik biçimde daha sonra parametrik olmayan fonksiyon halinde modele eklenerek model tahminleri gerçekleştirilmiştir. Parametrik ve semiparametrik logit modelden elde edilen sonuçlar karşılaştırılarak göç eden bireylerin işgücüne katılmasını etkileyen faktörlerin hangi model tarafından daha iyi açıklandığı belirlenmeye çalışılmıştır. Model tahminleri için R programı kullanılmıştır. Semiparametrik logit model tahmini ve parametrik olmayan değişkenler için kısmi regresyon fonksiyonlarına ait grafikler “mgcv” paketi yardımıyla elde edilmiştir. Semiparametrik model tahmininde parametrik olmayan düzgünleştirme teriminin seçimi için otomatik düzeltme tekniğinden faydalanılmış, parametrik olmayan değişkenlere ait tahminler için spline düzeltme yöntemi kullanılmış ve model tahminleri backfitting algoritması ile gerçekleştirilmiştir.

4.4 ARAŞTIRMADA KULLANILAN VERİNİN YAPISI

Araştırmada kullanılan veriler Türkiye İstatistik Kurumu’nun 2011 yılı Nüfus ve Konut Araştırması mikro verisinden elde edilmiştir. NKA bir örnekleme çalışması olup araştırma kapsamında fert bazında toplam 1.048.575 gözlem mevcuttur. Bu gözlemlerin 45.871’i göç eden bireylere ait verileri oluşturmaktadır. Araştırmada göç edenlerin 1.481’i yurtdışından Türkiye’ye göç ederken 44.390’ı ise ülke içinde göç

etmiştir. Araştırmanın kapsamı gereği ülke içinde göç eden bu kesime ait veri kümesi araştırmanın örneklemini oluşturmaktadır. Bu veriye çeşitli filtreler uygulanmıştır. Göç eden 44.390 bireyin bir kısmına ait işgücü bilgisinin gözlenmediği görülmüştür. Bunun dışında araştırmaya katılanların bir kısmı sağlık sorunları nedeniyle göç etmek zorunda kalmıştır. Sağlık sorunları nedeniyle göçe katılan bu kesim istese bile çalışma yaşamı içerisinde bulunamayacak olması nedeniyle araştırma kapsamı dışında bırakılmıştır. Bu iki durum göz önünde bulundurularak uygulanan filtre sonucunda 29.772 gözlem elde edilmiş ve analizler bu gözlemler üzerinden gerçekleştirilmiştir.

4.5 VERİLERİN ANALİZİ

Veri analizi bölümünde NKA'dan elde edilen veriler ile iç göç sürecine katılan bireylerin işgücüne katılma durumunu etkileyen faktörler araştırılacaktır. Nitel tercih modelleri ile göç eden bireylerin işgücüne katılma durumları incelenmeden önce bireylerin demografik özelliklerine ve göç etme nedenlerine ilişkin bilgiler özetlenecektir.

4.5.1 Demografik Özelliklere İlişkin Frekanslar

Çalışma 29.772 göçmene ait gözlem verisi üzerinden gerçekleştirilmiştir. Göçmenlere ait demografik özellikleri incelediğimizde %47'si erkek %53'ü kadınlardan oluşmaktadır. Medeni durum açısından bakıldığında hiç evlenmemiş olanların oranı %33 iken evli olanların oranı %67'dir. Eğitim durumuna göre bir okul bitirmeyenlerin oranı %4, ilkokul mezunlarının oranı %16, ilköğretim-orta okul mezunlarının oranı ise %23'dür. Göç edenlerin %26'sı lise mezunu, %29'u yüksek okul-fakülte ve %2'si yüksek lisans-doktora mezunudur. Yaş dağılımı açısından baktığımızda göç edenlerin %37'si 15-24 yaş, %37'si 25-34 yaş, %15'i 35-44 yaş %8'i 45-54 yaş ve %4'ü ise 55-64 yaş aralığındadır.

4.5.2 Göç Etme Nedenine İlişkin Frekanslar

29.772 göçmene ait veriyi incelediğimizde göç etme nedenine bağlı oranlar şöyledir; göçmenlerin %17,4'ü iş aramak/bulmak amacıyla, %19,8'i tayin / iş değişikliği %14,1'i eğitim, %10,8'i evlilik/boşanma nedeniyle ve %35,9'u ise hanedeki fertlerden birine bağlı olarak göç etmiştir.

Göç etme nedenine kadın ve erkek açısından baktığımızda iş aramak bulmak amacıyla göç edenlerin %92'si erkek %8'i ise kadındır. Evlilik nedeniyle göçe katılanların %91'i, hanedeki fertlerden birine bağlı olarak göç edenlerin %80'i, tayin / iş değişikliği nedeniyle göç edenlerin ise %26'sı kadındır. Eğitim amacıyla göç edenlerin cinsiyet dağılımları birbirine çok yakın olup erkeklerin oranı %50,8 kadınların oranı ise %49,2'dir.

4.6 GÖÇMENLERİN İŞGÜCÜNE KATILIMINI ETKİLEYEN FAKTÖRLERİN LOGİT MODELİ İLE ANALİZİ: PARAMETRİK VE SEMİPARAMETRİK LOGİT MODELLERİ İÇİN UYGULAMA

Göçmenlerin işgücüne katılmasını etkileyen faktörlerin parametrik ve semiparametrik logit model ile incelenmesi bu araştırmanın konusunu oluşturmaktadır. Çalışmada tahmin edilen modeller için bağımlı değişken;

$$y_i = \begin{cases} 1 = \text{işgücünde} \\ 0 = \text{işgücünde değil} \end{cases}$$

olarak tanımlanmıştır. Birey işgücünde ise "1", değilse "0" değerini almaktadır. Modelde yer alan bağımsız değişkenler ise sırasıyla; eğitim, yaş, cinsiyet, medeni durum ve göçün akım yönü değişkenlerinden oluşmaktadır. 2011 yılı NKA mikro verisinde eğitim değişkeni kategorik olarak gözlenmiştir. Bu değişken için mezun olunan eğitim seviyesine bakılarak sayısallaştırma işlemi yapılmıştır. Eğitim değişkeni için eğitim düzeyine denk düşen eğitim yılı ilgili düzeye atanarak sayısal hale getirilmiştir. Yaş değişkeni sayısal olarak gözlendiğinden bu değişkenle ilgili herhangi bir işlem yapılmamıştır.

Cinsiyet deęişkeni için kadın ve erkek, medeni durum deęişkeni içinse hiç evlenmemiş ve evli olmak üzere iki kategori bulunmaktadır. Araştırmada cinsiyet deęişkeni için erkek medeni durum deęişkeni için hiç evlenmedi kategorisi referans kategori olarak kabul edilmiştir. İç göç çalışmalarında önemli bir dięer deęişken göçün akım yönü olup bu deęişken göçmenlerin göç etmeden önce yaşadığı yerleşim birimi ve göç ettiği yerleşim biriminin özelliğine baęlı olarak belirlenmektedir. Türkiye’de ikâmet edilen yer açısından yerleşim birimlerinin özellikleri; il merkezi, ilçe merkezi, belde veya köy olarak tanımlanmaktadır. Bireyin göç etmeden önceki yerleşim birimi özelliği ile göç ettiği yerleşim birimine ait özellik ilişkilendirilerek göçün akım yönü dört ayrı şekilde incelenmektedir. Burada TUİK’in yerleşim yerleri arası göç olarak ifade ettiği tanımdan yararlanarak şehirden-şehire, şehirden-köye, köyden-şehire ve köyden-köye olmak üzere dört akım yönü şu şekilde tanımlanmaktadır;

- Şehirden-şehire göç: Bir il merkezinden ilçe merkezine, ilçe merkezinden il merkezine, ilçe merkezinden ilçe merkezine, il merkezinden il merkezine (bir ilin merkezinden dięer ilin merkezine),
- Köyden-şehire göç: Köy veya beldeden il merkezine, köy veya beldeden ilçe merkezine,
- Şehirden-köye göç: İl merkezinden köy veya beldeye, ilçe merkezinden köy veya beldeye,
- Köyden-köye göç: Köy veya beldeden bir başka köy veya beldeye göç olarak tanımlanmaktadır.

Bu göç tipi dışında araştırmada il içi ve iller arası göç ayrımı olarak tanımlanan iki kategorili bir başka göç yönü deęişkeni de yer almaktadır. İl içi göç; bir il içindeki (büyükşehir ilçe merkezleri hariç) yerleşim yerleri arasındaki nüfus hareketliliği olarak ifade edilmektedir. Bir ilden başka bir ile yaşanan nüfus hareketliliği ise iller arası göç olarak tanımlanmaktadır. Bu deęişken için referans olarak iller arası göç kategorisi referans kategori olarak kabul edilmiştir. Eğitim ve yaş deęişkeni dışında modelde yer alan dięer baęımsız deęişkenler kukla deęişken olarak tanımlanmıştır. Kukla deęişken

oluşturulurken değişkenin istenen özelliği (referans kategori) için “1” bunun dışındaki tüm durumlar için değişkene “0” değeri atanmıştır⁵⁷.

Tablo 5
Parametrik Logit Modeli’ne İlişkin Tahmin Sonuçları

Değişken adı	Katsayılar	Standart Hata	t-değeri	Olasılık
Sabit	-2.8216	0.0750	-37.591	0.000
Eğitim	0.1876	0.0039	47.776	0.000
Yaş	0.0063	0.0015	3.974	0.000
Cinsiyet	2.0496	0.0294	69.637	0.000
Medeni	-0.7987	0.0362	-22.061	0.000
Göçyönü1	0.1248	0.0340	3.668	0.000
Göçyönü22	0.4100	0.0445	9.211	0.000
Göçyönü23	0.1758	0.0389	4.518	0.000
Göçyönü24	1.0211	0.0724	14.103	0.000

Araştırmada öncelikle bütün değişkenlerin yer aldığı parametrik logit model tahmini yapılmıştır. Tahmin sonuçları yukarıda Tablo 5’de görüldüğü gibidir. Parametrik logit model için elde edilen tahmin sonuçları tüm bağımsız değişkenlerin bağımlı değişken üzerinde anlamlı etkisinin olduğunu göstermektedir. Daha sonra modelde sayısal olarak yer alan eğitim ve yaş değişkeninin modele parametrik olmayan bileşen olarak dahil edildiği semiparametrik logit model tahmini gerçekleştirilmiş ve bu modelin de tüm değişkenlerinin anlamlı olduğu görülmüştür. Dolayısıyla hangi modelin en iyi olduğunun belirlenmesi için modellerden elde edilen uyum kriterlerinin değerlendirilmesi gerekmektedir. Tahmin sonrasında parametrik ve semiparametrik logit modeli karşılaştırmak için her iki modele ait uyumun iyiliğini gösteren değerler hesaplanmış ve sonuçlar aşağıda Tablo 6’da sunulmuştur.

⁵⁷ Değişken tanımları için bkz. Ek 1.

Tablo 6
Genel Model İçin Uyum Kriterleri

Modeller	Kalıntı Sapma	R^2	Pseudo- R^2	LR test istatistiği	Olasılık Değeri
Parametrik	31654	0.23	0.24	3997.31	0.000
Semiparametrik	27654	0.33	0.31		

Parametrik ve semiparametrik logit modelin karşılaştırılmasında sapma değerleri kalıntı kareleri rolü üstlenir ve modellerin karşılaştırılmasında uyum iyiliği ölçüsü olarak kullanılabilir. Sapma değerinin düşük olması ilgili modelin daha iyi olduğunu ifade etmektedir. Tablo 6'yı incelediğimizde semiparametrik logit modelin sapmasının (27654) parametrik logit model sapmasından (31654) daha düşük olduğu görülmektedir. Dolayısıyla ilk bakışta semiparametrik logit modelin göçmenlerin işgücüne katılma olasılığını daha iyi açıkladığı görülmektedir. Fakat regresyon analizlerinde iki ayrı model karşılaştırılırken çoğunlukla kalıntılar arasındaki farkın hipotez testleri ile sınanması sonuçların güvenilirliği açısından oldukça önemlidir. Benzer şekilde burada parametrik ve semiparametrik logit modelin karşılaştırılmasında aynı tür sınama; modellerden elde edilen sapmalar vasıtasıyla, LR testi kullanılarak gerçekleştirilmiştir. Teste ait sıfır hipotezi açıklayıcı değişkenlerin modele doğrusal formda dahil edilmesi gerektiğini ifade ederken alternatif hipotez ise düzgünleştirme fonksiyonunun daha iyi uyum geliştirdiği yönündedir. Tablo 6'da LR test sonuçlarını incelediğimizde olasılık değerinin ($P=0.000<0.05$) istenen düzeyde olması sapmalar arasındaki farkın anlamlı olduğunu göstermektedir. Bu nedenle sıfır hipotezinin reddedildiği yönünde karar alınmış ve LR testi sonucuna göre semiparametrik logit modelin parametrik logit modelden daha iyi uyum geliştirdiği sonucuna ulaşılmıştır.

Yine tablodan görüleceği gibi semiparametrik logit modelin R^2 ve Pseudo- R^2 değerinin parametrik logit modelden daha yüksek olduğu görülmektedir. Uyum iyiliği ölçüleri ve LR testi sonucunu bir bütün olarak değerlendirdiğimizde semiparametrik logit modelin göç eden bireylerin işgücüne katılma olasılığını daha iyi açıkladığı sonucuna ulaşılmıştır. Dolayısıyla konuya ilişkin yorumlar bu model üzerinden yapılacaktır.

Tablo 7**Semiparametrik Logit Modeli'ne İlişkin Tahmin Sonuçları**

Değişken adı	Katsayılar	Standart Hata	t-değeri	Olasılık
Sabit	-1.0280	0.0286	-35.938	0.000
Cinsiyet	2.4018	0.0348	69.012	0.000
Medeni	-0.1312	0.0455	-2.880	0.003
Göçyönü1	0.1912	0.0366	5.223	0.000
Göçyönü22	0.4992	0.0481	10.378	0.000
Göçyönü23	0.1754	0.0419	4.182	0.000
Göçyönü24	1.0977	0.0762	14.397	0.000
Değişken adı	Düzenleme Parametresi		F test istatistiği	Olasılık
Eğitim	4.99		2096.0	0.000
Yaş	4.95		2010.0	0.000

Göç eden bireylerin işgücüne katılma olasılığına ilişkin semiparametrik logit model tahmin sonuçları yukarıda Tablo 7’de görüldüğü gibidir. Tablo 7’de yer alan modele ait sonuçları incelediğimizde modelde parametrik olarak yer alan değişkenlerin anlamlı olduğu görülmektedir. Bu değişkenlere ait t istatistiklerine baktığımızda 0.01 hata payına göre tablo kritik değeri olan 2.57 değerinden yüksek olduğu dolayısıyla değişkenlerin istatistiksel açıdan anlamlı olduğu görülmektedir.

Modele düzenleme fonksiyonu olarak dahil edilen eğitim ve yaş değişkenlerinin anlamlılık sınaması için Kısıtlı F testi sonuçları her iki değişkene ait düzenleme parametresinin anlamlı olduğunu göstermektedir.

Parametrik değişkenleri incelediğimizde sabit terim ve cinsiyet değişkeni için sıfır hipotezinin yüksek bir t değeriyle reddedildiği görülmektedir. Bu durum modelde yer alan bu iki değişkenin işgücüne katılma olasılığı üzerinde diğer değişkenlere kıyasla önemli derecede etkiye sahip olduğunu göstermektedir.

Sabit terime ait tahmin sonuçlarını yorumlayacak olursak 2011 yılında göç eden bireylerin genel anlamda işgücüne düşük katılım gösterdiklerini söylemek mümkündür. Bu durumun Türkiye işgücü yapısı görünümü ile yakından ilişkili olduğu

söylenebilir. Nitekim Türkiye işgücü yapısına baktığımızda kurumsal olmayan nüfusun işgücüne düşük oranda katılım gösterdiği görülmektedir. Öyle ki Türkiye’de işgücüne katılma oranları OECD ülkeleri ortalamasının oldukça altındadır⁵⁸. Buna göre göç eden bireylerin çeşitli nedenlerle işgücüne katılmadığı görülmektedir.

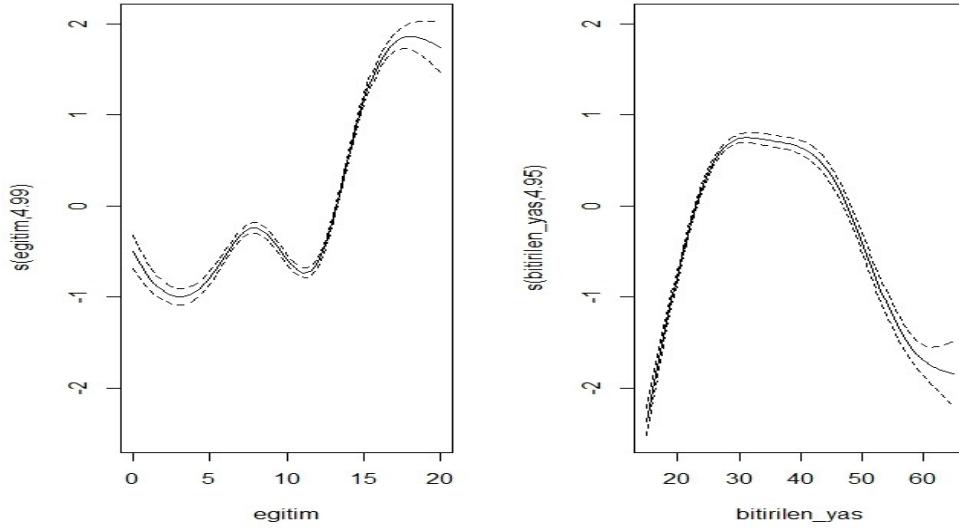
Cinsiyet açısından değerlendirildiğinde erkek referans kategori olarak modele dahil edilmiştir. Cinsiyet değişkeni için tahmin edilen katsayının yorumlanmasında bu durumun göz önünde bulundurulması gerekmektedir. Nitekim bu değişken için sonuçları yorumlayacak olursak erkeklerin kadınlara kıyasla işgücüne çok yüksek oranda katıldığı görülmektedir. Modelden elde edilen sonuçlara göre erkeklerin işgücüne katılma olasılığı kadınlarla kıyaslandığında oldukça yüksek (P=0.91) tahmin edilmiştir.

Medeni durum açısından bakıldığında hiç evlenmemiş olanların evlilere göre işgücüne daha az katıldığı görülmektedir. Fakat hiç evlenmemiş olanlar ve evliler arasındaki bu fark çok küçük bir fark olmakla beraber kadın ve erkek için sonuçların öneminin farklılaşması beklenmektedir.

Göçyönü1 değişkeni iller arası ve il içi göçlerin tanımlandığı kukla değişken olarak modelde yer almaktadır. Değişken için il içi göç bilgisi karşılaştırma grubu olarak kabul edilmiştir. Bu değişkene ait tahmin edilen parametre değerine baktığımızda iller arası göçe katılanların il içinde göç edenlere kıyasla daha fazla işgücüne katıldığı görülmektedir.

Göçyönü2 değişkeni için sonuçları şehirden-şehire (Göçyönü21) göç edenlerle kıyaslayarak yorumlarsak şehirden-köye (Göçyönü22), köyden-şehire (Göçyönü23) ve köyden-köye (Göçyönü24) göç edenlerin işgücüne daha yüksek katılım gösterdiği görülmektedir. Bu göç tipleri arasında şehirden-şehire göç edenlere kıyasla en yüksek işgücüne katılım köyden-köye göç edenlerde görülmektedir.

⁵⁸ Cahit Aydemir, “Türkiye’de İşgücü Yapısı, İşsizlik ve Kırsal Alan”, **Ankara Üniversitesi İktisadi ve İdari Bilimler Dergisi**, Cilt.27, Sayı.1, 2013, s.119.



Şekil 1: Genel Model İçin Eğitim ve Yaş'a Ait Düzgünleştirme Fonksiyonu

Şekil 1'de eğitim ve yaş değişkeni için kısmi regresyon fonksiyonları görülmektedir. Şekil 1 her iki değişken için tahmin edilen regresyon eğrilerine ait grafiği ifade etmektedir. Görüldüğü gibi açıklayıcı değişkenler ile bağımlı değişken arasında parametrik olmayan bir ilişki mevcuttur. Dikey eksen logit değerlerini yatay eksen ise modele nonparametrik biçimde dahil edilen değişkenlere ait değerleri göstermektedir. Grafikte görülen kesikli çizgiler güven bantları olup güven bantları ile birlikte çizilen grafik, bağımlı değişken ile bağımsız değişken arasındaki ilişkiyi gösteren dağılım grafiği olarak ifade edilebilir.

Nonparametrik değişkenler için elde edilen grafikler uyumun iyi olup olmadığına ilişkin karar alınmasına olanak sağlamaktadır. Güven bantlarının geniş olmayan bir aralıkta ve parametrik olmayan eğri ile aynı doğrultuda olması tahminlerin istatistiksel açıdan anlamlı ve güvenilir olduğunu göstermektedir. Şekil 1'i incelediğimizde tahmin edilen eğrilerin güven aralıklarının oldukça dar ve eğri ile aynı dağılımı izlediği görülmektedir. Dolayısıyla bu iki değişken için tahminler anlamlı ve güvenilir olarak elde edilmiştir.

Grafiklerin yorumlanmasında daha önceki bölümlerde anlatıldığı gibi logit ve olasılık değeri arasındaki ilişkiden yararlanılacaktır. Dikey ekseninde logit değerinin aldığı “0” değeri için göç eden bireylerin işgücüne katılma olasılığı 0.50 olarak hesaplanmaktadır. Logit değerinin sıfırın altına düşmesi olasılığın azalarak 0’a yaklaşması, 0’ın üstüne çıkması durumunda ise artarak 1’e yaklaşmasını ifade etmektedir.

Şekil 1’in sol tarafında yer alan eğriye baktığımız zaman eğitim ile işgücüne katılma olasılığı arasında pozitif yönlü bir ilişkinin olduğu görülmektedir. Eğitimdeki artış işgücüne katılma olasılığını artırmaktadır. Fakat eğitim düzeyi için bu etkiler sabit değildir. Eğitim yılının 12 olarak kabul edildiği lise mezuniyet düzeyine kadar işgücüne katılma olasılığının 0.50’nin altında olduğu görülmektedir. Lise seviyesinden yüksek okul düzeyine geçildiğinde işgücüne katılma olasılığının ciddi ölçüde arttığı ve bu düzeyden sonra artarak devam ettiği görülmektedir.

Sonuçları yaş açısından değerlendirecek olursak 15-24 yaş aralığındaki göçmenlerin işgücüne oldukça sınırlı katılım gösterdikleri görülmektedir. 30 yaşın en yüksek katılım seviyesine sahip olduğu ve 30-40 yaş aralığının işgücüne yüksek katılım gösteren yaş aralığı olduğu görülmektedir. 40 yaş sonrasında göçmenlerin işgücüne katılma olasılığının hızlı biçimde düşme eğilimine girdiği görülmektedir.

Genel modele ait sonuçlar kadın ve erkeğin işgücüne katılma durumunda önemli bir ayrım olduğunu gösterdiğinden bu iki grup için ayrı ayrı regresyon modelleri tahmin edilmiştir. Kadınlar için tahmin edilen parametrik logit modele ait sonuçlar aşağıda Tablo 8’de görüldüğü gibidir. Parametrik logit model için tahmin edilen katsayıların tamamının anlamlı olduğu görülmektedir ($P=0.000<0.05$). Dolayısıyla tahmin edilen tüm değişkenlerin işgücüne katılma olasılığı üzerinde anlamlı bir etkiye sahip olduğu söylenebilir.

Tablo 8
Kadın İçin Parametrik Logit Modeli'ne İlişkin Tahmin Sonuçları

Değişken adı	Katsayılar	Standart Hata	t-değeri	Olasılık
Sabit	-2.8484	0.1199	-23.748	0.000
Eğitim	0.2438	0.0057	42.782	0.000
Yaş	0.0235	0.0022	10.471	0.000
Medeni	0.3707	0.0464	7.980	0.000
Göçyönü1	0.2163	0.0474	4.560	0.000
Göçyönü21	-1.4908	0.0897	-16.604	0.000
Göçyönü22	-0.8863	0.0953	9.297	0.000
Göçyönü23	-1.4617	0.0899	-16.260	0.000

Kadınlar için tahmin edilen parametrik modele ait değişkenler ile semiparametrik logit model tahmini yapılmıştır. Semiparametrik logit modele ait sonuçlar incelenmiş ve sabit terimin anlamsız olduğu görülmüştür. Daha sonra model sabit terim olmadan yeniden tahmin edilmiştir. Her iki model için hesaplanan uyum iyiliği ölçüleri aşağıda Tablo 9'da sunulmuştur.

Tablo 9
Kadına Ait Model İçin Uyum Kriterleri

Modeller	Kalıntı Sapma	R^2	Pseudo- R^2	LR test istatistiği	Olasılık Değeri
Parametrik	16730	0.15	0.16	1310.84	0.000
Semiparametrik	15418	0.22	0.21		

Parametrik ve semiparametrik logit model için hesaplanan değerlere baktığımızda semiparametrik modelin sapması parametrik logit modelden daha düşük elde edilmiştir. Doğrusal fonksiyona karşı düzgünleştirme fonksiyonunun test edilmesinde kullanılan LR testi sonuçları da bu durumu destekler niteliktedir. LR test sonucuna bakılarak olasılık değerinin ($P=0.000<0.05$) anlamlı sonuç vermiş olması sıfır hipotezini reddederek alternatif hipotez olan düzgünleştirme fonksiyonunun tercih

edilmesi yönünde karar alınmıştır. Dolayısıyla semiparametrik logit modelin daha iyi uyum gösterdiği hipotez testi sonucunda ispatlanmış olmaktadır.

Ayrıca Tablo 9’da görüldüğü üzere parametrik logit modelden elde edilen R^2 değeri semiparametrik modelin R^2 değerinden daha küçük olarak tahmin edilmiş olması semiparametrik logit modelin daha iyi sonuç verdiğini göstermektedir. Genel model için olduğu gibi kadınlara ait model tahmin sonuçlarında da semiparametrik logit modelin daha iyi olduğu görülmektedir.

Tablo 10

Kadın İçin Semiparametrik Logit Modeli’ne İlişkin Tahmin Sonuçları

Değişken adı	Katsayılar	Standart Hata	t-değeri	Olasılık
Medeni	0.5916	0.0563	10.466	0.000
Göçyönü1	0.30245	0.0453	6.668	0.000
Göçyönü21	-1.3350	0.0353	37.723	0.000
Göçyönü22	-0.6829	0.0610	-11.179	0.000
Göçyönü23	-1.2880	0.0553	-23.276	0.000
Değişken adı	Düzenleme Parametresi	F test istatistiği	Olasılık	
Eğitim	4.98	2031.7	0.000	
Yaş	4.37	223.6	0.000	

Kadınlar için tahmin edilen semiparametrik logit modele ait sonuçları incelediğimizde modelde parametrik olarak yer alan değişkenlerin tamamının anlamlı olduğu görülmektedir. Bu değişkenlere ait t istatistiklerine baktığımızda tablo kritik değeri olan 2.52 değerinden yüksek olduğu dolayısıyla değişkenlerin 0.01 hata payına göre anlamlı olduğu görülmektedir.

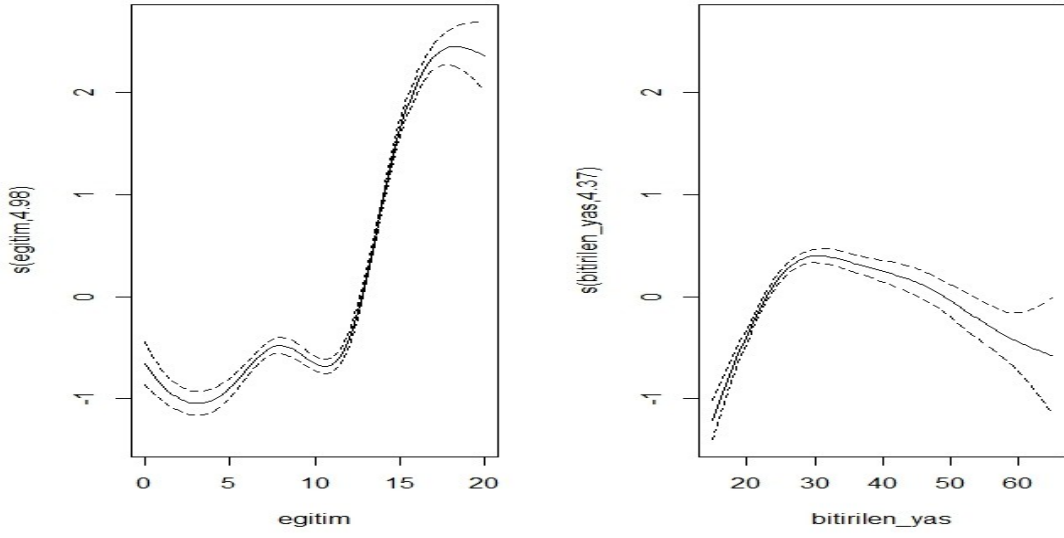
Modele nonparametrik olarak eklenen eğitim ve yaş değişkenlerinin anlamlılık sınaması için Kısıtlı F testi sonuçları her iki değişkene ait düzenleme parametresinin anlamlı olduğunu göstermektedir.

Modelde parametrik olarak yer alan medeni durum değişkenine ait sonuçları yorumlayacak olursak hiç evlenmemiş olan kadınların evli kadınlara kıyasla işgücüne

daha fazla katıldığı görülmektedir ($P= 0.64$). Bu değeri odds oranı ($e^{(0.5916)} = 1.80$) olarak ifade etmek gerekirse bekâr kadınların evlilere kıyasla işgücüne yaklaşık iki kat daha fazla katılım gösterdiği söylenebilir. Evli kadınların işgücüne katılmasını engelleyen temel unsur “ev kadınlığı” olarak ifade edilen durumdur. Gerçekten de göç eden kadınların yaklaşık %64’ünün işgücüne katılmama sebebi ev işleri ile meşgul olmasıdır (bkz. EK 8). Bu kesim aile içindeki rolü gereği ev içi işlerin (yemek yapmak, bulaşık yıkamak, temizlik yapmak vb.) yanı sıra çocuk, hasta ve yaşlı bakımıyla ilgilendikleri için işgücüne katılmamaktadır.

Öte yandan iller arası göç eden kadınların il içi göçe katılmış kadınlara kıyasla işgücüne daha yüksek oranda katıldığı görülmektedir. Elde edilen bu sonucu genel modele ait tahminler ile kıyaslayacak olursak yalnızca kadınlar için tahmin edilen katsayının genel modele kıyasla daha yüksek olduğu görülmektedir. Dolayısıyla iller arası göçe katılan kadınların il içi göçe katılan kadınlara oranla daha fazla işgücüne katıldığını belirtmek gerekir.

Göçyönü² değişkeni için sonuçlar incelendiğinde tüm akım yönlerinin kadınların işgücüne katılmasını azalttığı görülmektedir. Şehirden-şehire ve köyden-şehire göç edenlerin işgücüne çok az katıldığı bunun yanı sıra şehirden-köye göç edenlerin görece işgücüne daha yüksek katılım gösterdikleri görülmektedir. Bununla birlikte kadınlar için köyden-köye (Göçyönü²⁴) göç etme durumuna ilişkin tahmin edilen parametrenin anlamsız oluşu bu akım yönünün kadınların işgücüne katılma olasılığı üzerinde etkili olmadığını göstermektedir.



Şekil 2: Kadın İçin Eğitim ve Yaş'a Ait Düzgünleştirme Fonksiyonu

Kadınlar için tahmin edilen semiparametrik logit modele ilişkin parametrik olmayan değişkenlere ait kısmi regresyon fonksiyonlarını (Şekil 2) incelediğimizde eğitim ve yaş değişkeninin kadınların işgücüne katılma olasılığı üzerinde nonparametrik etkiye sahip olduğu görülmektedir.

Eğitim yılı ve kadınların işgücüne katılma olasılığı arasında pozitif yönlü bir ilişki mevcuttur. Eğitim yılı arttıkça kadınların işgücüne katılma olasılığı artmaktadır. Sonuçlar genel model ile benzerlik göstermektedir. Genel modelde olduğu gibi kadınlar için tahmin edilen model sonuçlarında da eğitim düzeyindeki artış işgücüne katılma olasılığı üzerinde sabit olmayan etkiye sahiptir. Fakat burada lise mezuniyet yılından sonra regresyon fonksiyonunun genel modelden elde edilen fonksiyona göre daha dik olduğu görülmektedir. Bu durum özellikle yüksek okul ve lisans mezunu olanlarda işgücüne katılma olasılığının çok yüksek olduğunu ifade etmektedir.

Yaş açısından bakıldığında genç kadınların işgücüne düşük oranda katılım gösterdiği görülmektedir. Bunun iki ayrı nedeni bulunmaktadır. İlki bu kesiminin çoğunluğunun (%43) eğitim amacıyla meslek edinmek için göç sürecine katılmış olmasıdır (bkz. EK 9). İkincisi ise genç kadınların evlilik nedeniyle göç etmeleridir. Evlilik nedeniyle göç eden kadınlar işgücüne katılım kararını çoğunlukla kendi başına

verememektedir. Çünkü evlenmemiş genç kadınlar ev dışında çalışma kararını daha özgürce alabilmekte iken evlenmeleri durumunda eşlerinin onayı olmadığında iş hayatını terk ettikleri gözlemlenmektedir⁵⁹. Fakat kısmi regresyon fonksiyonuna baktığımızda yaştaki artışın belirli bir noktaya kadar kadınların işgücüne katılımını olumlu yönde etkilediği ve erkeklere kıyasla 30 yaş sonrasında çok daha yavaş bir düşüş yaşandığı da görülmektedir.

Tablo 11
Erkek İçin Parametrik Logit Modeli'ne İlişkin Tahmin Sonuçları

Değişken adı	Katsayılar	Standart Hata	t-değeri	Olasılık
Sabit	3.8117	0.1914	19.914	0.000
Eğitim	0.1178	0.0065	17.893	0.000
Yaş	-0.0549	0.0029	-18.924	0.000
Medeni	-3.0216	0.0765	-39.464	0.000
Göçyönü21	-0.7574	0.1420	-5.333	0.000
Göçyönü22	-0.4904	0.1503	-3.261	0.001
Göçyönü23	-0.3622	0.1469	-2.465	0.014

Kadınlar için tahmin edilen model sonrasında erkekler için de parametrik ve semiparametrik logit model tahmini gerçekleştirilmiştir. Erkekler için tahmin edilen parametrik logit modele ait sonuçlar yukarıda Tablo 11'de özetlenmiştir. Parametrik logit model için tahmin edilen katsayıların tamamının anlamlı olduğu görülmektedir ($P=0.000<0.05$). Dolayısıyla tahmin edilen tüm değişkenlerin işgücüne katılma olasılığı üzerinde anlamlı bir etkiye sahip olduğu söylenebilir.

İkinci olarak erkekler için semiparametrik logit model tahmini yapılmıştır. Semiparametrik logit model için de elde edilen katsayıların anlamlı olduğu görülmüştür. Bu iki model arasından iyi olan modelin tespit edilmesi için erkekler için tahmin edilen parametrik ve semiparametrik modellere ait uyum kriterlerini gösteren değerlerin incelenmesi gerekmektedir.

⁵⁹ Nurcan, Önder, "Türkiye'de Kadın İşgücünün Görünümü", *Çalışma Dünyası Dergisi*, Cilt1, Sayı1, 2013, s.54.

Tablo 12
Erkeğe Ait Model İçin Uyum Kriterleri

Modeller	Kalıntı Sapma	R^2	Pseudo- R^2	LR test istatistiği	Olasılık Değeri
Parametrik	12301	0.18	0.16	3346.89	0.000
Semiparametrik	8946	0.40	0.30		

Diğer modellerde olduğu gibi erkekler için tahmin edilen parametrik ve semiparametrik logit modele ait sapma değerinin ve farklı uyum iyiliği ölçütlerinin değerlendirilmesi model seçimini kolaylaştırmaktadır. Tablo 12'ye baktığımızda semiparametrik logit modelin sapma değerinin parametrik logit modelden daha küçük olduğu görülmektedir. LR testi sonucunda sıfır hipotezinin reddedilmesi ($P=0.000<0.05$) modele nonparametrik olarak dahil edilen değişkenlerin- eğitim ve yaş değişkenlerinin-model uyumunu geliştirdiğini ifade etmektedir.

Yine Tablo 12'de görüldüğü üzere parametrik logit modelin R^2 ve Pseudo- R^2 değeri semiparametrik modelin R^2 ve Pseudo- R^2 değerinden daha düşük tahmin edilmiştir. Dolayısıyla erkekler için tahmin edilen modellere baktığımızda semiparametrik logit modelin parametrik logit modelden daha iyi uyum geliştirdiğini söyleyebiliriz.

Tablo 13
Erkek İçin Semiparametrik Logit Modeli'ne İlişkin Tahmin Sonuçları

Değişken adı	Katsayılar	Standart Hata	t-değeri	Olasılık
Sabit	3.4428	0.1824	18.870	0.000
Medeni	-1.8313	0.1157	15.822	0.000
Göçyönü1	0.1473	0.0690	2.133	0.033
Göçyönü21	-0.6044	0.1680	-3.597	0.000
Göçyönü22	-0.3612	0.1747	-2.067	0.039
Göçyönü23	-0.3391	0.1712	-1.980	0.048
Değişken adı	Düzgünleştirme Parametresi		F test istatistiği	Olasılık

Eđitim	4.97	294.8	0.000
Yaş	4.77	2246.7	0.000

Erkekler için tahmin edilen semiparametrik logit modele ait sonuçları incelediğimizde modelde parametrik olarak yer alan deęişkenlerin anlamlı olduđu görölmektedir. Bu deęişkenlere ait t istatistiklerine baktığımızda sabit terim, medeni durum ve göçyönü²¹ için elde edilen parametrelere ait t deęerlerinin tablo kritik deęeri olan 2.52 deęerinden yüksek olduđu dolayısıyla deęişkenlerin 0.01 hata payına göre anlamlı olduđu görölmektedir. Bu deęişkenler dışında kalan göçyönü¹, göçyönü²² ve göçyönü²³ deęişkenlerine ait t istatistiklerinin tablo kritik deęeri olan 1.96 deęerinden yüksek olduđu ve bu deęişkenlerin 0.05 hata payına göre anlamlı olduđu görölmektedir.

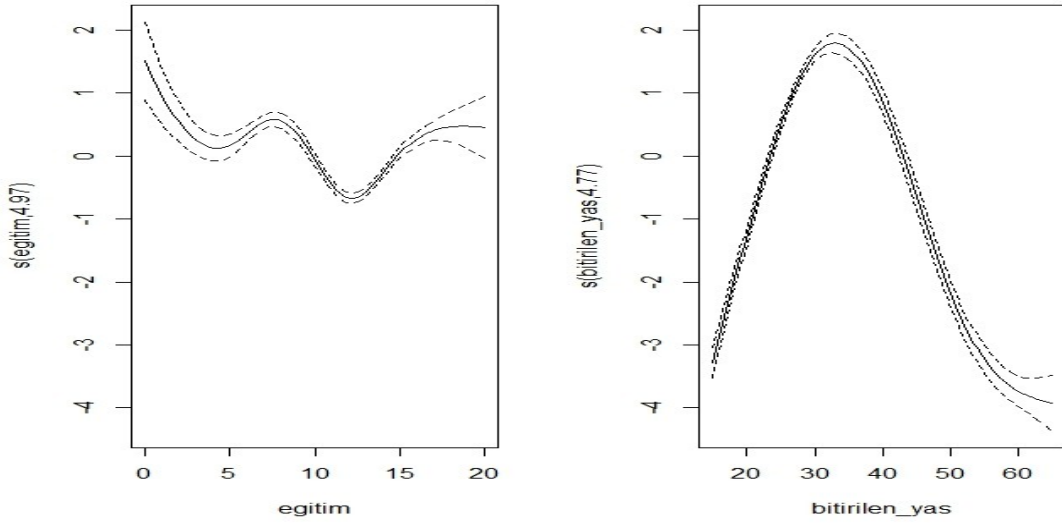
Modelde yer alan nonparametrik deęişken olan eğitim ve yaş deęişkenlerinin anlamlılık sınaması için F testi sonuçları her iki deęişkene ait düzgünleştirme parametresinin anlamlı olduđunu göstermektedir.

Modelde parametrik olarak yer alan deęişkenleri incelediğimizde sabit terim için elde edilen deęerin yüksek olduđu görölmektedir. Bu durum göç eden erkeklerin işgücüne katılma olasılığının ($P=0.97$) çok yüksek olduđunu göstermektedir.

Kadınların aksine hiç evlenmemiş olan erkeklerin evli erkeklere oranla işgücüne daha az katıldığı görölmektedir. Öyle ki evli olan erkekler bekârlara kıyasla işgücüne yaklaşık ($e^{(1.8313)} = 6.23$) altı kat daha fazla katılmaktadır.

Öte yandan iller arası göç eden erkekler ile il içi göçe katılmış erkekler arasında işgücüne katılma durumu açısından benzerlik söz konusu olup büyük bir farkın olmadığı görölmektedir.

Göçyönü² deęişkeni için sonuçlar incelendiğinde referans kategori olan köyden-köye göçle karşılaştırıldığında tüm göç yönlerine göre erkeklerin işgücüne görece daha az katılım gösterdiğini söylemek mümkündür.



Şekil 3: Erkek İçin Eğitim ve Yaş'a Ait Düzgünleştirme Fonksiyonu

Şekil 3 erkekler için tahmin edilen parametrik olmayan değişkenlere ait regresyon fonksiyonlarını göstermektedir. Eğitime ait kısmi regresyon fonksiyonuna baktığımızda eğitim ve işgücüne katılma olasılığı arasında negatif yönlü bir ilişki olduğu görülmektedir. Eğitim yılındaki artış erkeklerin işgücüne katılma olasılığını azaltmaktadır. Genel modelden elde edilen sonuçların aksine eğitim düzeyi düşük olan bireylerin eğitim düzeyi yüksek olanlara göre işgücüne daha yoğun katıldıkları görülmektedir. Özellikle lise mezuniyet düzeyi en düşük katılıma sahip kesimi oluşturmaktadır. Yüksekokul ve sonrası mezunları için ise katılımın görece biraz daha yükseldiği görülmektedir. Yaşa ait kısmi regresyon fonksiyonuna baktığımızda genel modele benzemektedir.

Fakat burada regresyon fonksiyonunun daha dik ve tepe noktasının daha sivri olduğu görülmektedir. Bu da erkekler için 35 yaşın ciddi bir eşik olduğuna işaret etmektedir. 35 yaşa kadar erkeklerde işgücüne katılma olasılığı artarak devam ederken bu eşikten sonra işgücüne katılma olasılığında artış sona ermekte ve hızlı bir düşüş yaşanmaktadır. Gençlerle benzer etkiye sahip olan 45 ve üstü yaşta olanların ise emeklilik nedeniyle işgücünden çekildiğini söylemek mümkündür.

SONUÇ

Göç eden bireylerin işgücüne katılmasını etkileyen faktörlerin incelendiği bu çalışmada ekonometrik yöntem olarak parametrik ve semiparametrik logit model ile çalışılmıştır. Sonuçlar konunun açıklanmasında semiparametrik logit modelin parametrik modelden daha iyi olduğunu göstermektedir. Araştırma göç eden bireylerin işgücüne katılmasını etkileyen faktörleri tespit ederek göçün işgücü piyasasına etkisini ortaya koymayı amaçlamıştır.

Araştırma sonuçlarına baktığımızda genel olarak 2011 yılında Türkiye’de iç göç sürecine katılan bireylerde işgücüne katılımın düşük olduğu görülmektedir. Fakat kadın ve erkekler için sonuçlar farklılaşmaktadır.

Türkiye’nin işgücü piyasası sorunlarına baktığımızda işgücüne katılımın genel olarak düşük olduğu ve bunun da kadınların işgücüne oldukça sınırlı katılım göstermesinden kaynaklandığı bilinmektedir⁶⁰. Dolayısıyla göçe katılan bireylerin işgücüne katılım konusundaki genel düşüklüğün nedenin de kadınların katılımının düşük olmasından kaynaklandığı analiz sonuçlarıyla desteklenmektedir.

Toplumsal roller ve kadın-erkek arasındaki aile içi işbölümü kadınların işgücüne katılmasını engelleyen en önemli etmendir. Sonuçlar medeni durumun kadın ve erkeğin işgücüne katılma durumu üzerinde farklı etkiye sahip olduğunu göstermektedir. Evli kadınların işgücüne katılmasını engelleyen bir dizi faktör bulunmaktadır. Kadın-erkek arasındaki aile içi iş bölümü gereği evli kadınlar çoğunlukla ev içi işlerle (bulaşık yıkamak, yemek ve temizlik yapmak vb.) meşgul olduğundan işgücüne sınırlı katılım göstermektedir. Bunun dışında çocuk sayısı, çocuğun yaşı ve eşin gelir durumu da kadınların işgücüne katılmasını engelleyen unsurlar arasında sayılabilir. Öte yandan kadınlar için göç nedeni ile işgücüne katılım gösterme durumu arasında yakın bir ilişki bulunmaktadır. Gerçekten de kadınların göç serüveni erkeklerinkinden oldukça farklıdır⁶¹. Göç etme nedenine ait oranlara kadın-

⁶⁰ Seyfettin Gürsel ve Gökçe Uysal, “Türkiye’de İşgücü Piyasası Dinamikleri ve Yapısal Sorunlar”, **TİSK Akademi**, Cilt7, Sayı14, 2012, s.9.

⁶¹ Göç nedenleri açısından bağlantılı göç çoğunlukla kadınlara özgü bir göç nedeni. Bağlantılı göç ailenin herhangi bir nedenle (iş bulmak, iş tayini vb.) göç eden erkek üyelerini takip eden kadınların hareketini tanımlayan bir olgu. Burada kadının birey olarak bağımsız bir göç kararından ziyade, aile içindeki konumuna (eş, anne, kız çocuk) bağlı

erkek açısından baktığımızda iş aramak bulmak amacıyla göç edenlerin yalnızca %10'u kadındır. Oysa evlilik nedeniyle göç edenlerin yaklaşık %90'ı, hanedeki fertlerden birine bağlı olarak göç edenlerin ise %65'inin kadın olduğu görülmektedir (bkz. EK 10). Bu durum bize kadınların çoğunlukla göç kararını kendi başlarına almadıkları bir başkasına bağlı olarak göç ettiğini göstermektedir. Dolayısıyla evli olması nedeniyle bireysel göç kararını özgürce alamayan kadınlar ekonomik yaşama da özgürce katılım sağlayamamaktadır. Fakat eğitim düzeyindeki artışın kadınları bu genellemenin dışına çıkardığını söylemek mümkündür. Nitekim analiz sonuçları da eğitilmiş kadınların işgücüne katılma olasılığının çok yüksek olduğunu göstermektedir.

Göç edilen alan açısından değerlendirildiğinde kente göç eden kadınlara ait sonuçlar kadınların işgücüne çok az katılım gösterdiğini bunun yanında köye göç edenlerin ise görece daha yüksek katılımda bulduklarını ortaya koymaktadır. Türkiye'nin kırsal bölgelerinde yaşayan kadınların eğitim hayatına sınırlı şekilde katılması bu kişilerin kentlere göç ettiğinde kentteki çalışma yaşamının beklentilerini karşılayacak eğitime sahip olmamasından dolayı işgücüne katılmasını engellemektedir. Bununla birlikte kentsel alandan köye göç edenler ise aile ekonomisine katkıda bulunmak amacıyla tarla, bağ, bahçe vb. işlerde ücretsiz aile işçisi konumunda çalışarak işgücüne dahil olmaktadır.

İç göç sürecine katılan erkeklerin işgücüne katılımının genel olarak yüksek olduğu görülmektedir. Eğitilmiş bireylerin yanında eğitimsizler de göç sürecine katılmaktadır. Model tahmin sonuçları düşük eğitime sahip olan erkeklerin işgücüne daha yüksek oranda katıldığını göstermektedir. Bu durum kentsel alana göç eden erkeklerin formel piyasalara kıyasla daha kolay girilebilen kayıt dışı/enformel istihdam alanlarında rahatlıkla iş bulabildiklerini göstermektedir.

15-24 yaş arası gençlerin çoğunun eğitim yoluyla meslek edinme isteğinde olduğu ve bu nedenle kentlere göç ettiği görülmektedir. Bireysel olarak göç eden ve eğitimde olan bu kesimin işgücüne aktif katılım göstermediği görülmektedir. Fakat aile

olarak yeni bir mekâna taşınması söz konusu. Aynı bağlamda yine ağırlıklı olarak kadınlara özgü başka bir göç nedeni olan "evlilik göçü"nü de ele almak mümkün. Bkz. Pınar İlkaracan ve İpek İlkaracan, "1990'lar Türkiye'sinde Kadın ve Göç", Oya Köymen (Ed.), **Bilanço 98: 75 Yılda Köylerden Şehirlere** içinde (305-322), İstanbul: Tarih Vakfı Yayınları, 1998, s. 307.

ile birlikte göç eden genç bireylerde bu durum tersine sonuçlar verecek ve genç bireyi çalışma hayatının içine çekecektir.

Bütün bu sonuçların yanı sıra göçün bir de göç edenler üzerinde etkileri bulunmaktadır. Bunlardan ilki kadınlara yönelik olup kırsal alanda ücretsiz aile işçisi olarak üretim süreci içerisinde yer alan ve kentlere doğru göç eden kadınların işgücünden çekilmesiyle sonuçlanmaktadır. Bu durum çalışma hayatından çekilen kadının sosyal hayattan da kopmasına neden olmaktadır. Öte yandan göçün, kırdan kente göç eden erkeklere etkisine baktığımızda göç etmeden önceki alan ve göç ettikleri yerleşim yerinin niteliği arasındaki üretim tarzından kaynaklı bir işçileşme-enformel işlerde çalışma sonucu sınıfsal farklılaşmalar meydana gelmektedir⁶². Nitekim göç ile birlikte kırsal alanda tarımsal faaliyetle uğraşan ve kendi hesabına çalışan küçük üreticilerde kente göç ettikten sonra işçileşme sonucu yaratmaktadır (bkz EK 11). Ayrıca sonuçlar eğitim nedeniyle göç eden genç kesimin bir kısmının çalışma yaşamı içerisine olduğunu göstermektedir (bkz EK 12). Bu durumda kentlere eğitim amaçlı göç eden gençlerin gelir yetersizliği nedeniyle çalışmak zorunda olduğu anlaşılmaktadır. Dolayısıyla eğitim nedeniyle kentlere göç eden bu kesim mevcut niteliklerinin dışında düşük ücretli olarak part-time ve güvencesiz çalışma içerisinde istihdam edilecektir.

Araştırma sonuçlarına baktığımız zaman göçün işgücü piyasası yapısal sorunlarına kaynaklık ettiğini söyleyebiliriz. Türkiye'nin işgücü piyasasına ait en önemli sorunların başında kayıt dışı istihdam ve kadın katılımının düşük olması gelmektedir. Görüldüğü gibi göç, işgücü piyasasındaki bu olumsuzluklara pozitif yönde etki etmektedir.

Çözüm olarak kadınların işgücüne katılmasını destekleyici politikaların gerçekleştirilmesi gerekmektedir. Özellikle kadınlar için çocuk bakım ve eğitim merkezlerinin kurulması onların ekonomik ve sosyal yaşama katılması önündeki engeli kaldıracaktır. Öte yandan kırsal kesimde kalkınma politikaları çerçevesinde gerçekleştirilecek ekonomi politikaları kırsaldan kentsel alanlara göçü dolayısıyla kayıt

⁶² Kırsal göçün analizini yapan Aşkın vd (2013) Yeşilyurt ilçesindeki göç eğilimlerini incelemiş ve şu sonuçlara ulaşmıştır; ilçede yaşayan nüfusun %52,7 si yaşadığı yerden göç etme eğiliminde olup göç etmeyi düşünenlerin %42,3'ü göç ettikleri kentlerde fabrikada işçilik yapmayı, %10,4'ü ise pazarcılık, seyyar satıcılık, kapıcılık gibi işlerde çalışmayı istemektedir. Bkz. Elif Özlem Aşkın, Rüştü Yayar ve Zafer Oktay, "Kırsal Göçün Ekonometrik Analizi: Yeşilyurt İlçesi Örneği", C.Ü. İktisadi ve İdari Bilimler Dergisi, Cilt 14, Sayı 2, 2013, s.241.

dışı istihdamdaki artışı engelleyecektir. Yine kırsal alanda tarım dışı istihdam alanlarının geliştirilmesi ve eğitim olanaklarının yanı sıra sağlık vb. kamusal hizmetlerin de yaygınlaştırılması önerilmektedir.

KAYNAKÇA

Kitaplar

Baltagi, Badi H. **Econometrics**. Four Edition. Berlin: Springer, 2008.

Bertschek, Irene ve Marlene Müller. “Productivity Effects of IT-Outsourcing: Semiparametric Evidence for German Companies”, Stefan Sperlich, Wolfgang Hardle ve Gökhan Aydınli(Ed.), **The Art of Semiparametrics** içinde (130-154), Berlin: Springer Science & Business Media, 2006, s.141-142.

Cameron, A. Colin ve Pravin K. Trivedi. **Microeconometrics**. New York: Cambridge University Press, 2005.

Christian Gourieroux. **Econometrics of Qualitative Dependent Variables**. Paul B. Klassen (çev.). New York: Cambridge University Press, 1991.

Cramer, Jan Salomon. **Logit Models From Economics And Other Fields**. New York: Cambridge University Press, 2003.

Çağlayan, Ebru. **Nonparametrik Regresyon Modelleri**. İstanbul: Derin Yayınları, 2012.

Eubank, Randall L. **Nonparametric Regression And Spline Smoothing**. Second Edition, New York: Marcel Dekker, 1999.

Greene, William H. **Econometric Analysis**. Seventh Edition. California: Pearson Education, 2012.

Gujarati, Damodar N. **Basic Econometrics**. Fourth Edition. New York: McGraw-Hill, 2003.

Güriş, Selahattin ve Ebru Çağlayan. **Ekonometri: Temel Kavramlar**. 3. Basım. İstanbul: Der Yayınları, 2010.

Hardle, Wolfgang. **Applied Nonparametric Regression**. United Kingdom: Cambridge University Press, 1990.

Hardle, Wolfgang ve Diğerleri. **Nonparametric and Semiparametric Models**. Berlin: Springer, 2004.

Hastie, Trevor ve Robert Tibshirani. **Generalized Additive Models**. London: Chapman and Hall, 1990.

Hastie, Trevor, Robert Tibshirani ve Jerome Friedman. **The Elements of Statistical Learning: Data Mining, Inference and Prediction**, Second Edition, Berlin: Springer, 2008.

Horowitz, J. L. **Semiparametric and Nonparametric Methods in Econometric**, New York: Springer, 2009.

İlkkaracan, Pınar ve İpek İlkkaracan, “1990’lar Türkiye’inde Kadın ve Göç”, Oya Köymen (Ed.), **Bilanço 98: 75 Yılda Köylerden Şehirlere** içinde (305-322), İstanbul: Tarih Vakfı Yayınları, 1998, s. 305-322.

Keele, Luke. **Semiparametric Regression for the Social Sciences**. England: John Wiley & Sons, 2008.

Long, Scott. **Regression Models for Categorical and Limited Variables: Advanced Quantitative Techniques in the Social Sciences 7**. USA: Sage Publications, 1997.

Maddala, Gangadharrao Soundaryarao. **Introduction to Econometrics**. Second Edition. New York: Macmillan Publishing Company, 1992.

Maddala, Gangadharrao Soundaryarao. **Limited Dependent and Qualitative Variables in Econometrics**. New York: Cambridge University Press, 1993.

Pampel, Fred C. **Logistic Regression: A Primer**. USA: Sage, 2000.

Ruppert, D., M. P. Wand M. P. ve R. J. Carroll. **Semiparametric Regression**, New York: Cambridge University Press, 2003.

Takezawa, Kunio. **Introduction to Nonparametric Regression**. Canada: John Wiley & Sons Publication, 2006.

Verbeek, Marno **A Guide to Modern Econometrics**. Second Edition. England: John Wiley & Sons, 2004.

Wood, Simon. **Generalized Additive Models: An Introduction With R**. Boca Raton: Chapman & Hall/CRC, 2006.

Wooldridge, Jeffery M. **Introductory Econometrics a Modern Approach**. Fourth Edition. Mason, Ohio: South Western Cengage Learning, 2009.

Sürekli Yayınlar

Akkuş, Özge, Serdar Demir ve Durdu Karasoy, “İki Düzeyli Bağımlı Değişken Modelinin Yarı Parametrik Tahmini”, **İstatistikçiler Dergisi**, Sayı.1, s.135-143.

Aşkın, Elif Özlem, Rüştü Yayar ve Zafer Oktay. “Kırsal Göçün Ekonometrik Analizi: Yeşilyurt İlçesi Örneği”, **C.Ü. İktisadi ve İdari Bilimler Dergisi**, Cilt 14, Sayı 2, 2013, s.231-252.

Aydemir, Cahit. “Türkiye’de İşgücü Yapısı, İşsizlik ve Kırsal Alan”, **Ankara Üniversitesi İktisadi ve İdari Bilimler Dergisi**, Cilt.27, Sayı.1, 2013, s.115-138.

Aydın, Dursun, Mammadagha Mammadov, Ali Faut Yüzer ve Embiya Ağaoğlu. “Yarı Parametrik Modellerde Splayn Düzeltme ile Tahmin ve Çıkarsamalar”. **Anadolu Üniversitesi Bilim ve Teknoloji Dergisi**. Cilt.8, Sayı.1, (2007), s.41-52.

Çağlayan, Ebru ve Melek Astar. “Logit ve Probit Modellerinde Uyum İyiliği Ölçüleri”. **Trakya Üniversitesi Sosyal Bilimler Dergisi**. Cilt.12, Sayı.2, (Aralık 2010), s.1-12.

Davidson, Russell ve James G MacKinnon. "Convenient Specification Tests for Logit and Probit Models" **Journal of Econometrics**, Cilt.25, Sayı.3, 1984, s.241-262.

Gerfin, Michale. “Parametric and Semi-Parametric Estimation of the Binary Response Model of Labour Market Participation”, **Journal of Applied Econometrics**, Cilt.11, Sayı.3, (Mayıs-Haziran, 1996), s.321-339.

Gürsel, Seyfettin ve Gökçe Uysal. “Türkiye’de İşgücü Piyasası Dinamikleri ve Yapısal Sorunlar”, **TİSK Akademi**, Cilt7, Sayı14, 2012, s.6-33.

Klein, Roger W. ve Richard H. Spady. "An Efficient Semiparametric Estimator for Binary Response Models”, **Econometrica: Journal of the Econometric Society**, Cilt.61, Sayı.2, (Mart,1993), s.387-341.

Önder, Nurcan. “Türkiye’de Kadın İşgücünün Görünümü”, **Çalışma Dünyası Dergisi**, Cilt1, Sayı1, 2013, s.35-61.

Turanlı, Münevver ve Seda Bağdatlı. “Semiparametrik Regresyon”, **Öneri Dergisi**, Cilt.9, Sayı.35, (Ocak 2011), s.207-213.

Turanlı, Münevver ve Seda Bağdatlı Kalan. “Site İçerisindeki Daire Fiyatlarını Etkileyen Unsurların Semiparametrik Regresyon Analizi İle Belirlenmesi”, **İstanbul Ticaret Üniversitesi Sosyal Bilimler Dergisi**, Cilt.11, Sayı.21, 2012, s.383-402.

Diğer Yayınlar

Berthet J-L, Patilea V. “A review of some semiparametric regression models with applications to scoring”. **Proceedings of the XIth International Symposium in Applied Stochastic Models and Data Analysis**. 2005, Brest, France, 594-17–30 May 2005.

Bowman, Adrian ve Evers Ludger, “**Nonparametric Smoothing**”, 2015, School of Mathematics & Statistics The University of Glasgow, <https://warwick.ac.uk/fac/sci/statistics/apts/students/resources-1415/smoothing-preliminary-material.pdf> (15 Şubat 2018).

Dayıoğlu, Tuğba. “Parametrik ve Semiparametrik Logit Modelleri İle Hanehalkı Yoksulluğunun Analizi”, **Doktora Tezi**, Marmara Üniversitesi SBE, 2011.

Hansen, Bruce E., “**Econometrics**”, 2017, University of Wisconsin Department of Economics, <http://www.ssc.wisc.edu/~bhansen/econometrics/EconometricsKindle.pdf> (17 Şubat 2017).

Öztürk, Elif “Parametrik Olmayan ve Yarı Parametrik Panel Veri Modelleri: Çevresel Kuznets Eğrisinin Analizi”, **Doktora Tezi**, Marmara Üniversitesi SBE, 2011.

Türkiye İstatistik Kurumu, **Nüfus ve Konut Araştırması 2011**, No:4030, Ankara: Türkiye İstatistik Kurumu Yayını, 2013.

Wand, Matt P. ve M. Chris Jones, “**Kernel Smoothing**”, 1995, Monographs on Statistics and Applied Probability Chapman and Hall, http://compdiag.molgen.mpg.de/docs/talk_05_01_04_stefanie.pdf (12 Ocak 2017), s15.

EKLER

EK 1-Değişken Açıklamaları

Değişken Adı	Etiket	Açıklama
İşgücü	İşgücü	İşgücünde=1 İşgücünde değil=0
Eğitim	Eğitim	Kişinin en son mezun olduğu eğitim düzeyine karşılık gelen yıl
Yaş	Yaş	Kişinin bitirmiş olduğu yaş
Cinsiyet	Cinsiyet	Erkek=1 Kadın=0
Medeni durum	Medeni	Hiç evlenmedi=1 Evli=0
Göç yönü-1	Göçyönü1	İller arası göç=1 İl içi göç=0
Göç yönü-2	Göçyönü21 Göçyönü22 Göçyönü23 Göçyönü24	Şehirden-Şehire=1 Diğer durumlar=0 Şehirden-Köye =1 Diğer durumlar=0 Köyden-Şehire =1 Diğer durumlar=0 Köyden-Köye =1 Diğer durumlar=0

EK 2-Parametrik Logit Model İçin R Program Çıktısı

```
Deviance Residuals:
  Min       1Q   Median       3Q      Max
-2.7109  -0.8237   0.4191   0.8209   2.6548

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -2.821606   0.075061 -37.591 < 2e-16 ***
egitim       0.187604   0.003927  47.776 < 2e-16 ***
bitirilen_yas 0.006311   0.001588   3.974 7.07e-05 ***
cinsiyet     2.049461   0.029431  69.637 < 2e-16 ***
gcyn2        0.124801   0.034025   3.668 0.000245 ***
gocyonu2     0.410080   0.044522   9.211 < 2e-16 ***
gocyonu3     0.175876   0.038926   4.518 6.24e-06 ***
gocyonu4     1.021105   0.072402  14.103 < 2e-16 ***
medeni1     -0.798766   0.036207 -22.061 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 41092  on 29771  degrees of freedom
Residual deviance: 31654  on 29763  degrees of freedom
AIC: 31672

Number of Fisher Scoring iterations: 4
```

EK 3-Semiparametrik Logit Model İçin R Program Çıktısı

```
Family: binomial
Link function: logit

Formula:
isgc ~ s(egitim, k = 6) + s(bitirilen_yas, k = 6) + cinsiyet +
      gcyn2 + gocyonu + medenil

Parametric coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.02809    0.02861 -35.938 < 2e-16 ***
cinsiyet     2.40185    0.03480  69.012 < 2e-16 ***
gcyn2        0.19125    0.03661   5.223 1.76e-07 ***
gocyonu2     0.49923    0.04811  10.378 < 2e-16 ***
gocyonu3     0.17541    0.04195   4.182 2.89e-05 ***
gocyonu4     1.09778    0.07625  14.397 < 2e-16 ***
medenil     -0.13112    0.04552  -2.880 0.00397 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
              edf Ref.df Chi.sq p-value
s(egitim)      4.994  5.000  2096 <2e-16 ***
s(bitirilen_yas) 4.945  4.998  2010 <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq. (adj) = 0.408   Deviance explained = 32.7%
UBRE = -0.06992   Scale est. = 1           n = 29772
```

EK 4-Kadınlar İçin Tahmin Edilen Parametrik Logit Model İçin R Program Çıktısı

```
      Min       1Q   Median       3Q      Max
-2.1289 -0.8788 -0.5297  1.0519  2.8298

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -2.848475    0.119946 -23.748 < 2e-16 ***
egitim       0.243860    0.005700  42.782 < 2e-16 ***
bitirilen_yas 0.023586    0.002252  10.471 < 2e-16 ***
gcyn2        0.216352    0.047448   4.560 5.12e-06 ***
gcyn11      -1.490806    0.089786 -16.604 < 2e-16 ***
gcyn12      -0.886316    0.095329  -9.297 < 2e-16 ***
gcyn13      -1.461797    0.089900 -16.260 < 2e-16 ***
medenil      0.370774    0.046465   7.980 1.47e-15 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 19691  on 15639  degrees of freedom
Residual deviance: 16730  on 15632  degrees of freedom
AIC: 16746

Number of Fisher Scoring iterations: 4
```

EK 5-Kadınlar İçin Tahmin Edilen Semiparametrik Logit Model İçin R Program Çıktısı

```
Formula:
isgc ~ s(egitim, k = 6) + s(bitirilen_yas, k = 6) + gcyn2 + gcyn11 +
      gcyn12 + gcyn13 + medeni1 - 1

Parametric coefficients:
      Estimate Std. Error z value Pr(>|z|)
gcyn2      0.30245   0.04536   6.668 2.6e-11 ***
gcyn11     -1.33508   0.03539  -37.723 < 2e-16 ***
gcyn12     -0.68295   0.06109  -11.179 < 2e-16 ***
gcyn13     -1.28806   0.05534  -23.276 < 2e-16 ***
medeni1     0.59164   0.05653   10.466 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
      edf Ref.df Chi.sq p-value
s(egitim)      4.984  5.000 2031.7 <2e-16 ***
s(bitirilen_yas) 4.370  4.795  223.6 <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.271   Deviance explained = 21.7%
UBRE = -0.01227  Scale est. = 1           n = 15640
```

EK 6-Erkekler İçin Tahmin Edilen Parametrik Logit Model İçin R Program Çıktısı

```
Deviance Residuals:
      Min       1Q   Median       3Q      Max
-2.8103   0.2317   0.3629   0.7357   2.1762

Coefficients:
      Estimate Std. Error z value Pr(>|z|)
(Intercept)   3.811725   0.191410  19.914 < 2e-16 ***
egitim        0.117847   0.006586  17.893 < 2e-16 ***
bitirilen_yas -0.054997   0.002906  -18.924 < 2e-16 ***
gcyn11       -0.757491   0.142026  -5.333 9.64e-08 ***
gcyn12       -0.490408   0.150397  -3.261 0.00111 **
gcyn13       -0.362200   0.146916  -2.465 0.01369 *
medeni1      -3.021630   0.076566 -39.464 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

      Null deviance: 14986  on 14131  degrees of freedom
Residual deviance: 12301  on 14125  degrees of freedom
AIC: 12315

Number of Fisher Scoring iterations: 5
```

EK 7-Erkekler İçin Tahmin Edilen Semiparametrik Logit Model İçin R Program Çıktısı

```

Formula:
isgc ~ s(egitim, k = 6) + s(bitirilen_yas, k = 6) + gcyn2 + gcyn11 +
      gcyn12 + gcyn13 + medenil

Parametric coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  3.44288    0.18245  18.870 < 2e-16 ***
gcyn2        0.14733    0.06907   2.133 0.032917 *
gcyn11      -0.60441    0.16803  -3.597 0.000322 ***
gcyn12      -0.36120    0.17474  -2.067 0.038725 *
gcyn13      -0.33914    0.17126  -1.980 0.047680 *
medenil     -1.83136    0.11575 -15.822 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
              edf Ref.df Chi.sq p-value
s(egitim)    4.973  5.000  294.8 <2e-16 ***
s(bitirilen_yas) 4.773  4.974 2246.7 <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.431   Deviance explained = 40.3%
UBRE = -0.36419   Scale est. = 1           n = 14132

```

EK 8- Cinsiyete Göre Göç Edenlerin İşgücüne Katılmama Nedeni (Bin)

	Erkek		Kadın		Toplam	
	N	%	N	%	N	%
Ev işleri	1	0.26	586	63.63	587	44.84
Eğitim/öğretim	262	67.53	255	27.69	517	39.50
Emekli	75	19.33	20	2.17	94	7.18
Çalışamaz halde	28	7.22	42	4.56	71	5.42
Diğer	21	5.41	18	1.95	39	2.98
Toplam	388	100.00	921	100.00	1309	100.00

Kaynak:TÜİK, 2011 NKA

EK 9- Cinsiyete Göre 15-24 Yaş Gençlerin Göç Etme Nedeni (Bin)

	Erkek		Kadın		Toplam	
	N	%	N	%	N	%
İş aramak/bulmak	52	12.12	15	2.62	67	6.69
Tayin/ İş değişikliği	26	6.06	22	3.84	48	4.79
Eğitim	249	58.04	247	43.11	496	49.50
Evlilik/Boşanma	2	0.47	147	25.65	149	14.87
Hanedeki fertlerden birine bağlı göç	97	22.61	139	24.26	236	23.55
Diğer	3	0.70	3	0.52	6	0.60
Toplam	429	100.00	573	100.00	1002	100.00

Kaynak:TUIK, 2011 NKA

EK 10-Cinsiyete Göre Göç Etme Nedeni (Bin)

	Erkek		Kadın		Toplam
	N	%	N	%	N
İş aramak/bulmak	359	89.9	40	10.1	399
Tayin/ İş değişikliği	274	74.2	95	25.8	369
Eğitim	280	51.0	269	49.0	549
Evlilik/Boşanma	26	9.9	237	90.1	263
Sağlık	15	48.3	16	51.7	31
Hanedeki fertlerden birine bağlı göç	492	34.5	936	65.5	1428
Diğer	45	26.3	16	73.7	61
Bilinmeyen	1	100	0	-	1
Toplam	1492		1608		3100

Kaynak:TUIK, 2011 NKA

EK 11-Göç Edilen Yerin Niteliği ve Cinsiyete Göre İstihdamda Olanların İşteki Durumu (Bin)

	Şehirden şehire		Şehirden köye		Köyden şehire		Köyden köye		Toplam	
	Erkek	Kadın	Erkek	Kadın	Erkek	Kadın	Erkek	Kadın	Erkek	Kadın
Ücretli, Maaşlı	372	177	77	26	131	34	26	7	606	244
İşveren	11	2	3	-	5	-	1	-	19	3
Kendi hesabına	23	4	26	5	19	4	8	2	76	15
Ücretsiz aile işçisi	3	4	9	28	1	5	2	20	16	58
Toplam	409	187	115	59	156	43	37	30	717	320

Kaynak:TÜİK, 2011 NKA

EK 12- Eğitim Nedeniyle Göç Edenlerin Cinsiyete Göre İşgücü Durumu (Bin)

	Erkek	Kadın	Toplam
İstihdam	54	38	92
İşsiz	22	24	46
İşgücünde olmayan	204	206	410
Toplam	280	268	548

Kaynak:TÜİK, 2011 NKA