SINIRLI BAĞIMLI
DEĞİŞKENLİ MODELLER
VE TAHMİNLEME YÖNTEMLERİ

Yüksek Lisans Tezi

İsmail YENİLMEZ

Eskişehir, 2017

SINIRLI BAĞIMLI DEĞİŞKENLİ MODELLER VE TAHMİNLEME YÖNTEMLERİ

İsmail YENİLMEZ

YÜKSEK LİSANS TEZİ

İstatistik Anabilim Dalı
Danışman: Doç. Dr. Yeliz MERT KANTAR

Eskişehir
Anadolu Üniversitesi
Fen Bilimleri Enstitüsü
Ocak, 2017

Bu Tez Çalışması BAP Komisyonunca kabul edilen 1606F563 no.lu proje kapsamında desteklenmiştir.

JÜRİ VE ENSTİTÜ ONAYI

İsmail YENİLMEZ'in "Sınırlı Bağımlı Değişkenli Modeller ve Tahminleme Yöntemleri" başlıklı tezi 17/01/2017 tarihinde aşağıdaki jüri tarafından değerlendirilerek "Anadolu Üniversitesi Lisansüstü Eğitim-Öğretim ve Sınav Yönetmeliği"nin ilgili maddeleri uyarınca, İstatistik Anabilim dalında Yüksek Lisans tezi olarak kabul edilmiştir.

	Unvanı-Adı Soyadı	İmza
Üye (Tez Danışmanı)	: Doç. Dr. Yeliz MERT KANTAR	.....
Üye	: Prof. Dr. Birdal ŞENOĞLU	.....
Üye	: Doç. Dr. Kadir Özgür PEKER	.....

……...

Enstitü Müdürü

ÖZET

SINIRLI BAĞIMLI DEĞİŞKENLİ MODELLER VE TAHMİNLEME YÖNTEMLERİ

İsmail YENİLMEZ
İstatistik Anabilim Dalı
Anadolu Üniversitesi, Fen Bilimleri Enstitüsü, Ocak, 2017
Danışman: Doç. Dr. Yeliz MERT KANTAR

Regresyon modelinde, yalnızca iki değer alabilen veya negatif değer alamayan veya belli aralıklarda değer alan bağımlı değişkenler sınırlı bağımlı değişken olarak tanımlanmaktadır. Sürekli bağımlı değişken belli aralıkta değer aldığında ise bu durum sansürlü veri durumu olarak ifade edilir ve bu durumda yaygın olarak kullanılan klasik sıradan en küçük kareler tahmin edicileri yanlı ve tutarsız sonuçlar verir. Tobit model veya sansürlenmiş normal regresyon modeli bu gibi sorunlara çözüm getirmek için Tobin (1958) tarafından geliştirilmiştir. Ancak tobit modelin en çok olabilirlik tahmini normallik varsayımına dayalıdır ve hata terimlerinin normal dağılmaması halinde etkin olmayan sonuçlar vermektedir. Normallik varsayımının esnetilebilmesi için literatürde çeşitli tahmin ediciler önerilmiştir. Kısmı uyarlamalı tahmin ediciler, normallik varsayımının ihlali halinde önerilen tahmin ediciler arasındadır.

Bu tez kapsamında, iyi bilinen sınırlı bağımlı değişkenli modeller ve bu modellere ilişkin tahminleme metotları incelenmiştir. Söz konusu tahmin ediciler, en çok olabilirlik, iki aşamalı heckit, iki aşamalı en küçük kareler, probit ve logittir. Bunun yanında normallik varsayımının bozulmasına karşı genelleştirilmiş normal dağılıma dayalı kısmı uyarlamalı bir tahmin edici ele alınmıştır. Ayrıca farklı hata dağılımları için ele alınan tahmin edicilerin göreli performansları simülasyon çalışması yardımıyla karşılaştırılmıştır. Elde edilen sonuçlar göstermiştir ki kısmı uyarlamalı tahmin edicinin performansı hata normal olmadığında en çok olabilirlik tahmin edicisine göre daha iyidir. Ayrıca elde edilen diğer tüm sonuçlar sunulmuş ve ilgili litetatür dikkate alınarak tartışılmıştır.

Anahtar Sözcükler: Sınırlı bağımlı değişkenli modeller, Tobit, İki aşamalı tahmin yöntemleri, Kısmı uyarlamalı tahmin edici, Sansürlü regresyon.

ABSTRACT

LIMITED DEPENDENT VARIABLE MODELS AND ESTIMATION METHODS

İsmail YENİLMEZ

Department of Statistics

Anadolu University, Graduate School of Science, January, 2017

Supervisor: Assoc. Prof. Dr. Yeliz MERT KANTAR

In the regression model, the dependent variable which takes only two values or do not take a negative value or takes values at certain intervals are defined as the limited dependent variable. If the continuous dependent variable takes values at a certain range, this situation is expressed as the censored variable. In this case, ordinary least squares estimates give biased and inconsistent results. To solve a part of this problem, the censored normal regression model or tobit model, was proposed by Tobin (1958). However, the maximum likelihood estimation of the tobit model depends on the assumption of normality and if the errors are non-normally distributed, the tobit model yields inefficient results. Various estimators have been proposed to relax the normality assumption. To cope with non-normality, one of the proposed estimator is partially adaptive estimators.

In this thesis, well-known limited dependent variable models and estimation methods for these models are examined. The considered estimators are maximum likelihood, two-stage heckit, two-stage least squares, probit and logit. Besides that, a partially adaptive estimator based on the generalized normal distribution is considered in the case of non-normallity problem. Furthermore, a simulation study is used to analyze the estimators' relative performance in the case of different error distributions. Different estimators are compared in this way. The results obtained show that the performance of the partially adaptive estimator is better than the maximum likelihood estimator when the errors are non-normally distributed. All other results obtained are presented and discussed regarding the relevant literature.

Keywords: Limited dependent variable models, Tobit, Two-stage estimation methods, Partially adaptive estimator, Censored regression.

ÖNSÖZ

Bu tezin ortaya çıkmasına bilgisi, deneyimi ve çalışkanlığı ile en önemli katkıyı sağlayan danışmanım Doç. Dr. Yeliz MERT KANTAR'a,

Bu süreçte zaman ayırıp yardımlarını esirgemeyen Araş. Gör. İbrahim ARIK'a,

Hayatımın her anında yanımda olan geniş aileme,

On yıldır en büyük destekçim olduğu gibi bu süreçte her an beni destekleyen sevgili eşim Begüm YENİLMEZ'e teşekkür ederim.

İsmail YENİLMEZ

Ocak 2017

17/01/2017

ETİK İLKE VE KURALLARA UYGUNLUK BEYANNAMESİ

Bu tezin bana ait, özgün bir çalışma olduğunu; çalışmamın hazırlık, veri toplama, analiz ve bilgilerin sunumu olmak üzere tüm aşamalarında bilimsel etik ilke ve kurallara uygun davrandığımı; bu çalışma kapsamında elde edilemeyen tüm veri ve bilgiler için kaynak gösterdiğimi ve bu kaynaklara kaynakçada yer verdiğimi; bu çalışmanın Anadolu Üniversitesi tarafından kullanılan "bilimsel intihal tespit programı"yla tarandığını ve hiçbir şekilde "intihal içermediğini" beyan ederim. Herhangi bir zamanda, çalışmamla ilgili yaptığım bu beyana aykırı bir durumun saptanması durumunda, ortaya çıkacak tüm ahlaki ve hukuki sonuçlara razı olduğumu bildiririm.

İsmail YENİLMEZ

İÇİNDEKİLER

Sayfa

TABLOLAR/ÇİZELGELER DİZİNİ

Sayfa

ŞEKİLLER DİZİNİ

Sayfa

SİMGELER VE KISALTMALAR DİZİNİ

OYF **:**Olasılık Yoğunluk Fonksiyonu (Probability Density Function)

BDF **:**Birikimli Dağılım Fonksiyonu (Cumulative Distribution Function)

EKK **:**En Küçük Kareler (Least Squares)

2AEKK **:**İki Aşamalı En Küçük Kareler (Two-Stage Least Squares)

EÇO **:**En Çok Olabilirlik (Maximum Likelihood)

UEÇO **:**Genelleştirilmiş Normal Dağılıma Dayalı Kısmi Uyarlamalı En Çok Olabilirlik Tahmin Edicisi (Partially Adaptive Estimator Based on Generalized Normal Distribution)

NES **:**Normal Eşdeğer Sapma (Normal Equivalent Deviation)

MÜF **:**Moment Üreten Fonksiyon (Moment Generating Function)

BM **:**Beklenti Maksimizasyonu (Expectation Maximization)

PEKK **:**Yalnızca Pozitif Değerleri Hesaba Katan Sıradan En Küçük Kareler Tahmin Edicisi (Positively Ordinary Least Squares)

SEKMS **:**Sansürlü En Küçük Mutlak Sapma (Censored Least Absolute Deviation)

GED-1 **:**Genelleştirilmiş Hata Dağılımı Tip-I (Generalized Error Distribution-I)

GED-2 **:**Genelleştirilmiş Hata Dağılımı Tip-II (Generalized Error Distribution-II)

GND **:**Genelleştirilmiş Normal Dağılım (Generalized Normal Distribution)

OO **:**Olabilirlik Oranı (Likelihood Ratio)

LÇ **:**Lagrange Çarpanı (Lagrange Multiplier)

HKO **:**Hata Kareler Ortalaması (Mean Squared Error)

$\phi(.)$:Standart Normal Dağılımın OYF

$\Phi(.)$:Standart Normal Dağılımın BDF

GİRİŞ

Sınırlı bağımlı değişkenler, olası değerler aralığı belli yollarla sınırlandırılmış değişkenlerdir (Wooldridge, 2002, s.451). Bu değişkenler sansürlenmiş, kırpılmış ve ayrık sonuçlar içeren değişkenleri içerir.

Değişkenlerin belli stokastik seçim mekanizmalarından dolayı tanım aralıklarında sınırlandırılmaları ve ekonometrik modellerde nitel değişkenlerin kukla değişkenlerle sıklıkla kullanımı, sınırlı bağımlı değişkenli modellerin kullanım alanını genişletmektedir (Maddala, 1983).

Bu modeller anket verilerinin analizinin yapıldığı deneysel (ampirik) çalışmalarda kullanıldığı gibi bazı zaman serisi modellerinde de uygulanabilirliğe sahiptir. Maddala (1983) tarafından sınırlı bağımlı değişkenli modeller örneklerle ifade edilerek üç farklı kategoride sınıflandırılmıştır.

- Kırpılmış regresyon modelleri,
- Sansürlenmiş regresyon modelleri,
- Kukla endojen modelleri.

Bu modellerin literatürde yaygın olarak uygulandığı örneklerden kısaca bahsedilirse,

1. Kesikli regresyon modelleri için negatif gelir vergisi deneyi örneği;

Hausman ve Wise (1976, 1977) tarafından ele alınan problemde $y = f(\text{eğitim}, \text{yaş}, \text{deneyim}, \text{v.b.})$ şeklinde bir gelir denklemi elde edebilmek için tanımlanan eşik noktasının üzerindeki değerler için, gözlem değerlerin belli bir noktadan kırpıldığı söylenebilir. Bu durumda, en küçük kareler (EKK) yönteminin yanlı tahminler verdiği görülmüştür. Kazanç olarak tanımlanan y değişkeni üzerinde eğitim süresinin etkisinin incelendiği regresyon denklemi $u_i \sim N\left(0, \sigma^2\right)$ olmak üzere;

$$y_i = x_i^{'}\beta + u_i$$

şeklinde yazılır. y_i, t herhangi bir kırpılma sabiti olmak üzere, $y_i \le t$ iken gözlemlenebilmektedir. Bu durumda;

$$u_i \le t - x_i^{'}\beta$$

olarak elde edilen eşitsizlik üzerinden yapılacak beklenen değer hesabı ile $E\left(u_i \left| u_i \le t - x_i^{'}\beta \right.\right)$'nın 0 olmadığı görülür. Aksine, artıklar x_i'nin bir fonksiyonu olup açıklayıcı değişken olan x_i ile ilişkilidir. Bu durumda elde edilecek β tahminleri tutarlı olmayacaktır. EKK tahmin edicisinin kullanılması durumunda β'nın pozitif olduğu ve x_i'nin artan değerleri için $E\left(u_i \left| u_i \le t - x_i^{'}\beta \right.\right)$ beklenen değerinin azaldığı görülecektir. Bu durumda β'nın EKK tahmin edicisinin aşağı yönlü olarak yanlı olduğu sonucuna varılmış olur. Örneğe dönülecek olursa, eğitimin gelir üzerindeki etkisinin araştırılması halinde negatif gelir vergisi deneyinden elde edilen veri için, gerçek etkinin altında tahminler elde edilecektir (Maddala, 1983, s.1-2).

2. Sansürlü regresyon modelleri için dayanaklı tüketim malları örneği;

Birçok kişinin araba ya da temel ev eşyaları (buzdolabı, çamaşır makinası vb.) için harcama yapmadığı bilinmektedir. Bunun yanında harcama miktarlarında anlık ciddi değişimlerin varlığından söz edilebilir. Bu durumda yapılacak modellemelerdeki eksiklik Tobin (1958) tarafından fark edilmiştir. Bu gibi durumlarda kullanılacak model Tobin tarafından y harcama ve x açıklayıcı değişkenler olacak şekilde;

$$y = x\beta + u \qquad y > 0$$

$$y = 0 \qquad d.d.$$

tanımlanmıştır. Diğer ifadeyle y bireysel harcama değeri iken y^* harcama eşiğidir. Bu eşik öznel olarak kabul edilebilecek en ucuz araba fiyatı olabilir. Bu durumda model aşağıdaki gibi yazılabilir;

$$y = x\beta + u \qquad y > 0$$

$$y^* = z\gamma + v \qquad d.d.$$

İlk modelde harcama bedeli eşik değerinden büyükse doğrudan alınırken eşik değerinden küçükse 0 alınır. İkinci modelde ise harcama eşik değerinden küçükse 0 yerine kişiden kişiye değişen bir değer kullanılmaktadır (Maddala, 1983, s.3-4).

3. Kukla endojen modelleri için konut talebi örneği;

Konut talebi analizinde alışılagelmiş olan metot, sahibi tarafından kullanılan konut talebi ile kiralanarak kullanılan konut taleplerinin ayrı ayrı incelenmesidir. Bir diğer

yöntem olarak ise konut harcamaları için ve sahibi tarafından kullanılan konut için izafi kiranın, kukla (gölge) değişken ile tanımlanmasıdır. Bu durumda;

$D = 1$ sahibi tarafından kullanılan konut

$D = 0$ *d.d.*

Burada öz seçim problemi mevcuttur. Bazı bireyler kendi ev sahibi olmayı bazı bireyler ise ev kiralamayı tercih edebilir. Talep fonksiyonunun tahmininde bu durum hesaba katılmalıdır. Bu problem ayrıntılı olarak Trost (1977), Lee ve Trost (1978) ile Rosen (1979) tarafından çalışılmıştır (Maddala, 1983, s.7).

Ayrıca kukla endojen modelleri için zorunlu öğrenim yasası örneğinde, rassal değişkenlerin normal dağıldığı varsayımı incelenmiştir. Buna bağlı olarak normal dağılımdan ayrılmanın, sonuçların etkinliğini düşüreceği açıktır. Hata teriminin dağılımı için alternatif olarak,

- Logistik dağılım
- Cauchy dağılımı
- Burr dağılımı

önerilmiştir.

Bu dağılımların alternatif kabul edilmelerindeki ortak özellik birikimli dağılım fonksiyonlarının (BDF) kapalı forma sahip olmasıdır. Ancak bunun gelişen bilgisayar teknolojileri ile çok büyük bir önemi kalmamıştır. Bu durumun yanında farklı bir alternatif olarak Burr dağılımının yalnızca pozitif değerler alan rassal değişkenlerin işlenmesinde kullanılabileceği ifade edilmiştir. Gamma ve beta dağılımları da kullanılan farklı alternatiflerdir.

Bu tez kapsamında, bağımlı değişkenlerin belli aralıklarda sınırlı olma durumlarında kullanılan ekonometrik modeller yani sınırlı bağımlı değişkenli regresyon modelleri çalışılmıştır. Sınırlı bağımlı değişkenler daha önce de ifade edildiği gibi sansürlenmiş, kırpılmış ve ayrık sonuçlar içeren değişkenleri içerir.

Genel anlamda iki durumlu tercih modelleri doğrusal olasılık modeli, logit model ve probit modeldir. Ancak eldeki verilerin sürekli olması ve belli kısıtlara (sansürlenme, kırpılma) sahip olması, farklı modellerin kullanımını gerekli kılmaktadır. Bu durumda devreye sansürlü regresyon modeli ile kırpılmış regresyon modeli girer.

Araştırmanın genel çerçevesi nitel tercih modelleri ve sınırlı bağımlı değişkenli modellerdir. Nitel tercih modelleri denilince literatürde sıklıkla kullanılan iki durumlu modellerin dışında, durum sayısının arttığı çoklu tercih modellerinin varlığı göz ardı edilemez. Ancak araştırma motivasyonunu sansürlenmiş ve kırpılmış veri grupları ile bu veri gruplarının modellenmesinden almaktadır. Bu aşamada kısmen probit modelle ilişkili olan ve kukla değişken içeren tobit model ve tahmin edicilerinin incelenmesinden önce tahmin metodu olarak kullanılan nitel tercih modellerinin iki durumlu olanlarının araştırmaya konu edilmesi gerekli görülmüştür. Ancak ikiden fazla tercihi barındıran nitel tercih modelleri bu araştırma kapsamına alınmamıştır.

- Kesikli regresyon modelleri başlığında iki durumlu modeller olan doğrusal olasılık modeli, logit model ve probit model ele alınmıştır.
- Sansürlenmiş ve kırpılmış regresyon modelleri başlığında ise sansürlenme durumları ile kırpılma durumları derinlemesine incelenmiştir.
- Sansürlü regresyon modeli olarak tobit model ve tobit model tipleri başlığı altında; sansürlenme durumunda modellemede sıklıkla kullanılan tobit model, tobit model tipleri, tobit model tahmin edicileri ve uygunluk ölçüleri ile testler paylaşılmıştır. Ayrıca tobit tip 1 model ayrıntılı olarak incelenerek; EKK, en çok olabilirlik (EÇO), iki aşamalı en küçük kareler (2AEKK), iki aşamalı heckit (2AHeckit), genelleştirilmiş normal dağılıma dayalı kısmi uyarlamalı en çok olabilirlik (UEÇO) tahmin edicileri ele alınmıştır.
- Uygulamalar başlığında ise ilgili tahmin ediciler MatLab programında yazılan bir algoritmayla karşılaştırılmıştır. Yine MatLab kullanılarak söz konusu model ve metotlar örneklendirilmiştir.
- Sonuç bölümünde ise araştırma kapsamındaki teorik bilgilerden elde edilen ve uygulamalarla paralellik gösteren sonuçlar ile geleceğe ilişkin öneriler sunularak tez çalışması neticelendirilmiştir.

1. KESİKLİ REGRESYON MODELLERİ

Bağımlı değişkenin ayrık sonuçlu olduğu modeller kesikli regresyon modelleridir. En ilkel hali ile ikili seçimin var olduğu modeller örnek gösterilebilir. Değişkenin iki özelliğe sahip olduğu durum, ikili (binary) ya da dikotom (dichotomous) olarak ifade edilebilir. Bu gibi durumlarda veri kategorik olup nominaldir. Daha açık bir ifade ile bağımlı değişkenin alabileceği bu iki seçenek iki farklı durumu niteler. Sonuç olarak evet-hayır, yaşıyor-yaşamıyor, kadın-erkek gibi dikotom durumların 0 ve 1 ile kodlanması ile bu değişkenler regresyona alınabilir. Bağımlı değişkenin ikiden fazla değer aldığı durumlarda söz konusudur. Bu durumda kategorik ve kategorik olmayan değişkenler devreye girer. Bu aşamada değişken türlerini açıklamakta fayda vardır. Literatürdeki ve özellikle Türkçe kaynaklarda çeviri kaynaklı ve kavram karmaşası sebepli birçok tanım birbirine girmiştir. Bu aşamada değişkenler için ayrımları netleştirmekte yarar vardır.

Veriler toplanma yöntemine göre kategorik (nitel) ve numerik (nicel) olarak ikiye ayrılır. Verilerin toplanma yöntemleri sayma ya da ölçme şeklinde olabilir. Kategorik veriler nominal ve ordinal olmak üzere ikiye ayrılırken numerik veriler ise kesikli ve sürekli olmak üzere iki farklı şekilde ifade edilebilir. Ayrıca kategorik değişkenlerin bir diğer ayrışımı şu şekilde yapılmıştır; sıralı olmayan (unordered), sıralı (ordered) ve dizisel (sequential) değişkenler (Amemiya, 1975; Cox, 1970). Burada yapılan ayrım ilk yapılan ayrımı karşılar ancak dizisel değişken dışarda kalır. Sıralı olmayan (unordered) değişken nominal değişkeni karşılarken sıralı (ordered) değişken ordinal değişkeni karşılamaktadır.

Tablo 1.1. *Toplanma türlerine göre veri türleri*

Kategorik (Nitel, Kalitatif)			Kategorik Olmayan (Nicel, Numerik, Kantitatif)	
Nominal-unordered	Ordinal	Dizisel	Kesikli	Sürekli

Kategorik değişkenlerin analizlerinde doğrudan sayı değerleri kullanıldığı gibi yüzdeleri de kullanılabilmektedir. Kategorik olmayan değişkenlerde ise ortalama ve standart sapma gibi değerler kullanılabilir.

İstatistiksel analiz programlarında da değişkenler farklı şekillerde ele alınmıştır. Örneğin SPSS, kategorik değişkenleri nominal ve ordinal diye ayırırken kategorik olmayan yani nümerik (nicel) değişkenleri tek tip olarak ele almaktadır. Hâlbuki kategorik olmayan (nümerik-nicel) değişkenlerde sürekli ve kesikli olmak üzere ikiye ayrılmaktadır.

Kategorik (nitel-kalitatif) değişkenler ve kategorik olmayan (nicel-nümerik-kantitatif) değişkenlerin yanında türetilmiş değişkenler mevcuttur. Bu değişkenler ilgili istatistiğin amacına göre çeşitlilik göstermektedir. Türetilmiş değişken olarak genellikle oran, orantı, yüzde ve hız kullanılmaktadır (Aktürk ve Acemoğlu, 2011 s.35).

Bu araştırmaya konu edilen bir önemli başlık kesikli regresyon modelleridir. Kesikli regresyon modelleri için kullanılabilecek değişkenler kategorik ve kategorik olmayan değişkenler olarak ele alınabilir. Tablo 1.1'de verilen değişkenler incelendiği zaman kesikli regresyon modellerinin konusuna kategorik değişkenlerin tamamı yani nominal, ordinal ve dizisel değişkenler girerken kategorik olmayan değişkenlerden kesikli değişken girer. Sırasıyla;

Nominal değişken örneği için;

$y = 1$ kişinin mesleği işçi ise

$y = 2$ kişinin mesleği akademisyen ise

$y = 3$ kişinin mesleği terzi ise

şeklinde ifade edilebilir. Burada gruplar arasında herhangi bir sıralama mevcut değildir. Yalnızca kategorize edilen her bir meslek sayısal bir değerde tutulmuştur.

Ordinal değişken örneği için;

$y = 1 \qquad x < 1000$

$y = 2 \quad 1000 \le x < 2000$

$y = 3 \quad 2000 \le x < 3000$

$y = 4 \quad 3000 \le x < 4000$

$y = 5 \qquad x \ge 4000$

biçiminde tanımlanan y değeri ordinal değişkendir ve burada bir tür sıralama söz konusudur. x rassal değişkeniyle ifade edilen gelir değerleri belli aralıklarla sınırlandırılmıştır. Bu aralıklara düşen her bir değer içinde belli bir atama yapılmıştır.

Dizisel değişken örneği için;

$y = 1$ kişi liseyi tamamlayamamış ise

$y = 2$ kişi liseyi tamamlamış ancak yüksek öğrenimi yok ise

$y = 3$ kişi yüksek öğrenimini tamamlamış ancak yüksek bir ortalamaya sahip değilse

$y = 4$ kişi yüksek öğrenimini yüksek bir ortalama ile tamamlamış ise

atanan değerlerde sarmallık mevcuttur ve yanıt değişkeninin adım adım tamamlanarak gittiği açıkça görülebilir.

Genel olarak, bağımlı değişkenin kesikli olduğu durumlardan en ilkel olanı olarak görülen ikili seçimin söz konusu olduğu durumların modellenmesinde doğrusal olasılık modeli, logit ve probit model kullanılır. Kullanılacak olan bu kesikli regresyon modelleri sırası ile sunulmuştur.

1.1. Doğrusal Olasılık Modeli

Bağımlı değişkenin ikili seçime sahip olduğu regresyon modellerinde kullanılan doğrusal olasılık modeli $E(u_i) = 0$ olmak üzere şu şekilde ifade edilir;

$$y_i = x_i'\beta + u_i \tag{1.1}$$

$E(y_i | x_i) = x_i'\beta$ olarak elde edilir. y 'nin regresyon eşitliğinden hesaplanan değer $\hat{y}_i = x_i'\hat{\beta}$, x 'in belli bir değeri için olayın oluşma olasılığını verir (Maddala, 1983, s.15).

Tablo 1.2. *Doğrusal olasılık modeli tablosu*

u_i	$f(u_i)$
$1 - x_i'\beta$	$x_i'\beta$
$-x_i'\beta$	$1 - x_i'\beta$

Kaynak: *Maddala, 1983, s.15.*

Özetle doğrusal olasılık modeli ikili seçim durumlarında tercihin olasılığını veren, bağımsız değişkenlerin doğrusal bir fonksiyonudur. Bu durum şu şekilde ifade edebilir.

$y_i = 1$ olayın gerçekleşme olasılığı : P_i

$y_i = 0$ olayın gerçekleşmeme olasılığı : $1 - P_i$

Bu durumda;

$$P_i = \begin{cases} 0 & E\left(y_i \middle| x_i\right) \le 0 \\ 1 & E\left(y_i \middle| x_i\right) \ge 1 \\ x_i'\beta & 0 < E\left(y_i \middle| x_i\right) < 1 \end{cases}$$

olur. Beklenen değer yardımı ile;

$$Y = y \qquad y_1, y_2, ..., y_n \tag{1.2}$$

$$f(y) = P(Y = y) \qquad f(y_1), f(y_2), ..., f(y_n)$$

$$E(Y_i) = \sum_{i=1}^{n} y_i f(y_i) = 1.P_i + 0.(1 - P_i) = P_i$$

elde edilir. Koşullu beklenen değer,

$$E\left(y_i \middle| x_i\right) = x_i'\beta = P_i \tag{1.3}$$

şeklindedir. $0 \le P_i \le 1$ ve (1.3) dikkate alınacak olursa;

$$0 \le E\left(y_i \middle| x_i\right) \le 1 \tag{1.4}$$

elde edilir. (Eren, 2012, s.3). Ancak uygulamalarda olasılık değeri veren bu işlem [0,1] aralığının dışına çıkabilmektedir. Bu durumun yanında doğrusal olasılık modelinin sahip olduğu farklı problemlerde mevcuttur.

Örneğin;

$$\begin{aligned} Var(u_i) &= x_i'\beta\left(1 - x_i'\beta\right)^2 + \left(1 - x_i'\beta\right)\left(x_i'\beta\right)^2 \\ &= x_i'\beta\left(1 - x_i'\beta\right) \\ &= E\left(y_i\right)\left[1 - E\left(y_i\right)\right] \end{aligned} \tag{1.5}$$

elde edilir.

(1.5) değişen varyans (heteroscedasticity) problemini açıkça ortaya koymaktadır. Bu durumda (1.1) modeli için EKK tahmin yöntemi ile elde edilecek olan β değerleri etkin olmayacaktır.

Bu durumda önerilen;

- Öncelikle (1.1) yardımı ile EKK tahmin edicisini elde etmek,
- Sonrasında ise $\hat{y}_i\left(1-\hat{y}_i\right)$ değerini hesaplayarak ağırlıklandırılmış EKK metodunu (1.6) kullanmaktır (Goldberger, 1964, s.250)

$$\mathrm{w}_i = \left[\hat{y}_i\left(1-\hat{y}_i\right)\right]^{1/2} \tag{1.6}$$

Bu süreçte meydana gelen problemler sırası ile;

1. Büyük örneklemlerde küçük bir olasılık olsa da zaman zaman $\hat{y}_i\left(1-\hat{y}_i\right)$ değeri negatif gelebilmektedir (Maddala, 1983, s.16). Ancak $\hat{y}_i\left(1-\hat{y}_i\right)$ değerinin $E\left(y_i\right)\left[1-E\left(y_i\right)\right]$ için sabit bir tahmin edici olduğu gösterilmiştir (McGillavray, 1970).
2. Hata terimlerinin $\left(u_i\right)$ normal dağılmayışı, EKK tahmin edicisinin etkinliğini azaltmaktadır. Bu bağlamda EKK tahmin edicisinden daha etkin olan doğrusal olmayan yöntemler vardır.
3. Olayın oluşma olasılığı olarak yorumlanabilen koşullu beklenen değer $E\left(y_i \middle| x_i\right)$ zaman zaman [0,1] aralığının dışında yer alabilmektedir (Maddala, 1983, s.16; Gujarati, 2004, s.584).

Ayrıca bu problemlerin yanında doğrusal olasılık modeli için şu sorunlardan bahsedilebilir.

- Hata terimlerinin $\left(u_i\right)$ değişen varyansa sahip olması

Yukarıda bahsedilen durumu ayrı olarak ele almak gerekecektir. (1.5) eşitliğinden ifade edilen sonuç, değişen varyans sorununu açıkça ifade etmektedir. Değişen varyans durumunda EKK tahmin edicisi sapmasızdır ancak en küçük varyanslı değildir. Bu aşamada da (1.6) eşitliğindeki varyanslar kullanılarak ağırlıklandırılmış EKK tahmin edicisi devreye sokulur ve problemin üstesinden gelinir.

- Uyum iyiliği ölçüsü olan R^2 ' nin kuşkulu değeri

Özelde doğrusal olasılık modeli için genelde ise kukla değişken ile tutulan bağımlı değişkenli modellerde R^2 uyum iyiliği için iyi bir ölçü olarak kabul edilmemektedir. Bu durumun başlıca sebebi konu kapsamında doğrusal olasılık modeli için değerlendirilecek olursa, katsayı değerinin 0,2 ile 0,6 sınırları arasında değer almasıdır ki bu kuşkulu bir sonuçtur (Aldrich ve Nelson, 1984, s.15).

- P_i değerinin x_i rassal değişkeninin bir fonksiyonu olarak kabul edilmesi

P_i değerinin x_i rassal değişkeninin bir fonksiyonu olarak kabul edilmesi sonucunda x_i rassal değişkeninin marjinal etkisi sabitlenmektedir. Bu durumda x_i'deki her bir birimlik değişim P_i üzerinde eşit değerde etki oluşturmaktadır. O halde x_i'deki her bir birimlik değişim, gerçekleşme olasılığını β_i kadar değiştirmektedir. Bu durum (1.7) eşitliği ile ifade edilebilir (Long, 1997, s.39).

$$\frac{\partial P_i}{\partial x_i} = \beta_i \tag{1.7}$$

Doğrusal modellerde karşılaşılan bu sorunun bazı modellerde çözülmesine rağmen doğrusal olasılık modeliyle çözülemiyor olması, doğrusal olasılık modelinin değişkenleri ile çalışabilen ve doğrusal olasılık modelinin sorunlarını içermeyen logit ve probit modeli alternatif haline getirir. $0 \le E\left(y_i \left| x_i \right.\right) \le 1$ varsayımının kontrolü için öncelikle doğrusal olasılık modeli için EKK tahmin edicisi ile $\hat{y}$ bulunarak varsayıma uygunluğu araştırılır. Doğrusal olasılık modelinde, bulunan değer 1'den büyükse 1, 0'dan küçükse 0 değeri verilir. Ancak bu durum tahmin edicinin yanlı tahminler verdiğini gösterir. Doğrusal olasılık modelinin en önemli problemlerinden olan ve en güçlü eleştirileri aldığı $0 \le E\left(y_i \left| x_i \right.\right) \le 1$ varsayımının sağlanamayışının çözüm yöntemlerinden biri yukarıda ifade edildiği gibi logit ve probit modelin kullanımıdır. Doğrusal yapıya sahip doğrusal olasılık modeline alternatif olarak doğrusal olmayan logit ve probit model sırasıyla ele alınmıştır.

1.2. Logit ve Probit Model

Goldberger (1964) tarafından isimlendirilen probit model için yazılacak regresyon eşitliği;

$$y_i^* = x_i'\beta + u_i \tag{1.8}$$

şeklindedir. y_i^* gözlenemez ve eldeki gölge değişken y şu şekilde tanımlanır;

$$\begin{array}{ll} y = 1 & y^* > 0 \\ y = 0 & d.d. \end{array} \tag{1.9}$$

Burada doğrusal olasılık modelinde olduğu gibi $x_i'\beta$, $E(y_i|x_i)$ değerine eşit değildir. (1.8) ile (1.9)'dan,

$$\begin{aligned} P(y_i = 1) &= P(u_i > -x_i'\beta) \\ &= 1 - F(-x_i'\beta) \end{aligned} \tag{1.10}$$

elde edilir.

Olabilirlik fonksiyonu ise;

$$L = \prod_{y_i=0}\left[F(-x_i'\beta)\right]\prod_{y_i=1}\left[1 - F(-x_i'\beta)\right] \tag{1.11}$$

şeklindedir. (1.11)'de kullanılan BDF (1.8)'deki u_i varsayımına bağlıdır. Buna bağlı olarak u_i 'nin BDF'sinin logistik olduğu kabul edilirse;

$$F(-x_i'\beta) = \frac{\exp(-x_i'\beta)}{1+\exp(-x_i'\beta)} = \frac{1}{1+\exp(x_i'\beta)} \text{ olacaktır.}$$

Bu durumda;

$$1 - F(-x_i'\beta) = \frac{\exp(x_i'\beta)}{1+\exp(x_i'\beta)} \tag{1.12}$$

olarak yazılabilir. Burada kullanılan BDF'de olduğu gibi, BDF her zaman kapalı bir forma sahip olmayabilir. Örneğin probit modelde u_i 'nin dağılımının $IN(0,\sigma^2)$ olduğu biliniyor ise;

$$F(-x_i'\beta) = \int_{-\infty}^{-x_i'\beta/\sigma} \frac{1}{(2\pi)^{1/2}} \exp\left(-\frac{t^2}{2}\right) dt \tag{1.13}$$

olacaktır. (1.11) ve (1.13)'dan yalnızca β/σ tahmin edilebilir. Bu oranın parametleri ayrı ayrı tahminlenmez. Ancak $\sigma = 1$ olarak varsayılarak başlanabilir (Maddala, 1983, s.22).

Logistik dağılımla ilişkili olan logit model ile normal dağılım ile ilişkili olan probit modelin büyük örnekleme sahip olunmaması ve kuyruktaki gözlemlerin ciddi etkiye sahip olmaması durumlarında benzer sonuçlar vereceği vurgulanmıştır (Maddala, 1983, s.10-11).

Logistik dağılım ile normal dağılımın karşılaştırılması içinse bir dönüşümün yapılması gerekmektedir. Standart normal dağılımın varyansı 1 iken standart $\sec h^2$ dağılımının varyansı $\pi^2/3$'tür. Bu durumda logit değeri $(3/\pi)^{1/2}$ değeri ile çarpılmalıdır. Alternatif olarak logit değerin $1/1,6$ ya da $0,625$ değeri ile çarpılabileceği ifade edilmiştir (Amemiya, 1981).

1.2.1. Logit model

Doğrusal olmayan modellerin hata terimlerine dair dağılım bilgisi mevcuttur. Logit model kümülatif logistik dağılım fonksiyonuna dayanır ve herhangi *i.* bireyin bağımlı değişken olan tercihlerden birini seçme olasılığıdır (Pindyck ve Rubinfeld, 2009, s.736).

David Cox tarafından 1958 yılında geliştirilen logistik regresyon, ikili seçim durumlarında, ikili yanıtın olasılığının tahmininde kullanılır (Cox, 1958). Bir sınıflama metodu olmayıp ikili tercih modelidir. Logit model, bağımsız değişkenlere göre bağımlı değişkenin beklenen değerinin olasılık olarak bulunduğu bir yöntemdir (Özdamar, 2011).

İki tercih durumlu bağımlı değişken için yazılacak olan logit model;

$$P_i = E\left(Y=1|x_i\right) = \frac{1}{1+\exp\left(-x_i'\beta\right)} \tag{1.14}$$

şeklindedir ve bu eşitlikte $z_i = x_i'\beta$ dönüşümü yapılacak olursa,

$$P_i = E\left(Y=1|x_i\right) = \frac{1}{1+\exp\left(-z_i\right)} \tag{1.15}$$

elde edilecektir. Bu denklem lojistik dağılımın BDF'sidir. Bu denklemde $-z_i \xrightarrow{-\infty} P_i = 0$ ve $-z_i \xrightarrow{+\infty} P_i = 1$ olacaktır. Burada anakütle katsayılarının tahmininde EKK, z_i ile P_i arasında doğrusal olmayan ilişkiden dolayı kullanılamaz. Bu durumda lojistik dağılımın BDF'sinin tersi alınarak logit model doğrusallaştırılmış olur.

$P_i = \frac{1}{1+\exp\left(-z_i\right)}$ ve $1-P_i = \frac{1}{1+\exp\left(z_i\right)}$ kullanılarak,

$$\frac{P_i}{1-P_i} = \frac{\frac{1}{1+\exp(-z_i)}}{\frac{1}{1+\exp(z_i)}} = \frac{1+\exp(z_i)}{1+\exp(-z_i)} = \exp(z_i) \tag{1.16}$$

elde edilir. Burada kullanılan $P_I/1-P_I$ değeri bahis-fark-şans oranı (odds ratio) olarak adlandırılır. Bu aşamada oranın logaritmasının alınması ile logit model elde edilir.

$$L_i = \ln\frac{P_I}{1-P_I} = \ln\left(\exp(z_i)\right) = z_i = x_i'\beta \tag{1.17}$$

Burada L_i değeri de logit olarak adlandırılır. Dikkat çekici nokta logitin sadece x değerine göre değil anakütle katsayılarına göre de doğrusal olmasıdır (Gujarati ve Porter, 2012, s.554).

Bahis-fark-şans oranı (odds ratio) olarak yukarıda ifade edilen değer göreceli olasılıklar oranı ya da tahmini rölatif risk olarak da anılmaktadır ve değer $P_I/1-P_I$ ifadesinden de anlaşılacağı üzere ortaya çıkma olasılığının ortaya çıkmama yani gerçekleşmeme olasılığına oranıdır. Bu oran kat olarak yorumlanabildiği gibi olasılık oranı olarak da yorumlanabilir. İki örnek durum incelenir ve yorumlanırsa;

- Beraberlik olasılığının mümkün olmadığı bir müsabakada A takımının kazanma olasılığı $P(A) = 0,6$ olsun. Bu durumda A takımının kaybetme, dolayısı ile B takımının kazanma olasılığı $P(A') = P(B) = 0,4$ olacaktır. Bu durumda,

  $OR_A = odds(A) = P(A)/P(A') = 0,6/0,4 = 1,5$ olarak elde edilir. Diğer taraftan,

  $OR_B = odds(B) = P(B)/P(B') = 0,4/0,6 = 0,\overline{6} \approx 0,7$ bulunur.

  Bu durumda A takımının B takımını yenme olasılığı 1,5 kat daha fazladır denebilir.

- Analiz dersinin ilk sınavında 50 kız öğrenciden başarılı sayılacak kız öğrencilerin sayısı 40, 50 erkek öğrenciden başarılı sayılacakların sayısı da 10'dir. Bu durumda göreceli olasılık oranı,

  $$\frac{p}{1-p}\Big/\frac{q}{1-q} = \frac{p}{p'}\Big/\frac{q}{q'} \tag{1.18}$$

  yardımı ile hesaplanır.

Bu durumda bir kız öğrencisinin geçme olasılık oranı $0,8/0,2=4$; bir erkek öğrencisinin geçme olasılık oranı $0,2/0,8=0,25$ olarak elde edilir. Göreceli olasılıklar oranı ise $4/0,25=16$ olarak bulunur.

Kız öğrencilerin erkek öğrencilere göre $40/10=4$ kat daha fazla başarılı oldukları ancak başarılı olma olasılıklarının 16 kat daha fazla olduğu ifade edilebilir. Bu durumda göreceli olasılık oranının logaritması (olasılıkların logit değerleri farkı) alınarak değer yumuşatılır [$\log(16)=2,7726$; oran ters kullanılırsa $\log(1/16)=-2,7726$].

Yukarıda iki tercih durumlu bağımlı değişken için yürütülen süreç $P_i=E(Y|X_i)=\beta_0+\beta_1X_1+...+\beta_kX_k$ için işletilebilir. Ayrıca belirtilmelidir ki yukarıda bahsedilen doğrusal olmayan ilişki problemi yapılan logit işlemleri ile doğrusallaştırılmış olur.

Logit modelin tahmininde ise;

- Modeldeki bütün değişkenler kategorik ise EÇO yöntemi ile ağırlıklandırılmış EKK yöntemi kullanılabilir (Maddala, 1983, s.328).
- Kolay hesaplanabilir ancak amaç fonksiyonu tam maksimize edilemeyen minimum Ki-Kare yöntemi de bir diğer tahmin yöntemidir (Pampel, 2000, s.40).

Yukarıda verilen (1.17) eşitliğinde $x_i'\beta$ yerine $\beta_1+\beta_2x_i$ yazılacak olursa,

$$L_i=\ln\frac{P_i}{1-P_i}=\ln\left(\exp\left(z_i\right)\right)=z_i=\beta_1+\beta_2x_i \tag{1.19}$$

elde edilecektir.

Yukarıda verilen değere de hata terimi eklenerek tekrar yazılacak olursa,

$$L_i=\ln\frac{P_i}{1-P_i}=\ln\left(\exp\left(z_i\right)\right)=z_i=\beta_1+\beta_2x_i+u_i \tag{1.20}$$

elde edilir ki bunu tahmin etmek için x_i dışında bağımlı değişkene ya da L_i değerine ihtiyaç vardır. Bu durum eldeki verinin doğası ile ilişkilidir. Bu durumda;

i. Tekil düzeydeki veri

ii. Gruplanmış ya da yinelenmiş veri dikkate alınmalıdır (Gujarati ve Porter, 2012, s.554).

- Tekil düzeydeki veri

Frekans değeri mevcut olmayan verilerdir. Bu tipteki verilerde EKK yöntemi kullanılamaz. Sebebi ise EKK yönteminde kullanılacak olan ancak tekil düzey (mikro ya da kişisel düzey) sebebi ile 0 ya da 1 değerlerinin yerine yazılması ile anlamsız değerlerin ortaya çıkmasıdır.

$L_i = \ln \frac{1}{0}$ olayın gerçekleşmesi durumu

$L_i = \ln \frac{0}{1}$ olayın gerçekleşmemesi durumu

Tablo 1.3. *Tekil düzey veri örneği*

Aile Sıra	**Ev Sahipliği (Y)**	**Gelir (X)**
1	1	5.000 TL
2	1	4.500 TL
3	0	2 000 TL
4	0	3.000 TL
5	1	2.500 TL

- Gruplanmış veri ya da yinelenmiş veri

Gruplanma ya da yinelenme (tekrarlanma) sayıları bilinen verilerdir. $n_i \leq N_i$ olacak şekilde N_i karşı geldiği gelir seviyesindeki aile sayısını ifade ederken n_i bu ailelerden ev sahibi olanlarının sayısını ifade etmektedir.

Tablo 1.4. *Gruplanmış (Yinelenmiş) veri örneği*

Gelir (X)	**Ev Sahibi Aile Sayısı (n_i)**	**Gelir Düzeyindeki Aile Sayısı (N_i)**
2.000	3	10
3.000	2	6
4.000	4	5
5.000	2	3
6.000	3	3

O halde her x_i gelir seviyesi için göreli sıklık hesaplanırsa bu gelir seviyeleri için gerçek P_i tahmininde bu göreli sıklık değerleri kullanılabilir.

$$\hat{P}_i = \frac{n_i}{N_i} \tag{1.21}$$

Bu aşamada tekil düzeydeki veriler için kullanılamayan ancak gruplanmış ya da yinelenme (tekrarlanma) sayıları bilinen veriler için kullanılabilen EKK yönteminden bahsedilebilir.

1.2.1.1. En küçük kareler yöntemi (EKK)

Her x_i gelir seviyesi için göreli sıklığın hesaplanarak bu gelir seviyeleri için gerçek P_i tahmininde kullanılacağı ifade edilmişti $\left[\hat{P}_i = \frac{n_i}{N_i}\right]$. Burada N_i büyükse, $\hat{P}_i$ değeri P_i değerinin iyi bir tahminidir. Çünkü bir olayın olasılığı, göreli sıklıktaki örneklem büyüklüğünün sonsuza giderken hesaplanacak limit değeridir $\left(\hat{P}_i \xrightarrow{N_i \longrightarrow \infty} P_i\right)$. Tahmin edilen P_i değeriyle logit değeri şu şekilde tahmin edilebilir,

$$\hat{L}_i = \ln \frac{\hat{P}_i}{1-\hat{P}_i} = \hat{\beta}_1 + \hat{\beta}_2 x_i \tag{1.22}$$

burada da iyi bir P_i tahmini iyi bir L_i tahmini anlamına gelecektir.

Gruplanmış ya da yinelenme (tekrarlanma) sayıları bilinen verilerle, (1.20) eşitliğindeki modelin tahmini için bağımlı değişken ve logitlere dair veri elde edilebilir. Ancak EKK yöntemiyle katsayıların tahmininde hata terimlerinin özelliklerinin bilinmemesi sorun oluşturacaktır. Doğrudan EKK yöntemi ile katsayılar tahminlenebilir demek doğru olmayacaktır. N değerinin yeteri kadar büyük olduğu biliniyor ve verinin her x_i gelir seviyesi için binom değişkeni olarak bağımsız dağıldığı biliniyorsa u_i sıfır ortalamalı, $\frac{1}{N_i P_i (1-P_i)}$ varyanslı normal dağılıma sahiptir.

$$u_i \sim N\left[0, \frac{1}{N_i P_i (1-P_i)}\right] \tag{1.23}$$

Başarı oranı $\hat{P}_i$ ortalaması P_i ve varyansı $\frac{P_i(1-P_i)}{N_i}$'ye eşit olan binom dağılımına uyar ve $N_i \longrightarrow \infty$ iken dağılım normal dağılıma yakınsar (Theil, 1970).

Burada önemli nokta, doğrusal olasılık modelinde olduğu gibi logit modelin hata teriminin de değişen varyansa sahip olmasıdır. Bu durumda EKK yöntemi yerine ağırlıklandırılmış EKK yöntemi kullanılır. Bu aşamada bilinmeyen P_i yerine $\hat{P}_i$ kullanılarak σ^2'nin bir tahmin edicisi;

$$\hat{\sigma}^2 = \frac{1}{N_i \hat{P}_i \left(1-\hat{P}_i\right)} \tag{1.24}$$

elde edilir.

(1.20) eşitliğindeki regresyon modelinin tahmin aşamaları;

1. Her x gelir seviyesi için tahmin edilen olasılık $\hat{P}_i = \frac{n_i}{N_i}$ ile ayrı ayrı hesaplanır.

2. Her x_i için logit $\hat{L}_i = \ln \frac{\hat{P}_i}{1-\hat{P}_i}$ yardımı ile hesaplanır.

   Burada $\hat{P}_i = \frac{n_i}{N_i}$ olduğundan, $\hat{L}_i = \ln \frac{n_i}{N_i - n_i}$ yazılabilir. Uygulamalarda $\hat{P}_i$ değerinin 0 ve 1 değerini almaması için $\hat{L}_i = \ln\left(\mathrm{n}_i + 1/2\right)/\left(N_i - \mathrm{n}_i + 1/2\right) = \ln\left(\hat{P}_i + 1/2N_i\right)/\left(1-\hat{P}_i + 1/2N_i\right)$ olarak hesaplanır. Ayrıca her x_i değeri için N_i'nin en az 5 olması tavsiye edilir (Cox, 1970).

3. Değişen varyans problemi için eşitlik,

   $$\sqrt{w_i} L_i = \beta_1 \sqrt{w_i} + \beta_2 \sqrt{w_i} x_i + \sqrt{w_i} u_i \tag{1.25}$$

   ifadesine dönüştürülür. Değişen varyans durumunun hesaba katılmaması durumunda etkin olmayan tahmin ediciler elde edilir.

   Ağırlıklar sırasıyla; $w_i = N_i \hat{P}_i \left(1-\hat{P}_i\right)$, L^*_i ağırlıklandırılmış L_i, x^*_i ağırlıklandırılmış x_i, v_i dönüştürülmüş hata terimi olmak üzere,

   $$L^*_i = \beta_1 \sqrt{w_i} + \beta_2 x^*_i + v_i \tag{1.26}$$

   şeklinde yazılır.

Burada özgün hata varyansı $\left(\hat{\sigma}_u^{\,2} = \frac{1}{N_i \hat{P}_i (1-\hat{P}_i)} \right)$ eşitliğiyle v_i 'nin sabit varyanslı olduğunu ortaya koyar.

4. Ağırlıklandırılmış verilerle EKK yöntemi kullanılırken, eşitlik (1.25) EKK yöntemi ile tahmin edilir. Ayrıca (1.26) modelinde belli bir sabit terim yoktur. Bundan dolayı bu eşitlik tahmin edilirken 0'dan geçen regresyon kullanılmalıdır.
5. Güven aralıklarının klasik EKK yöntemi ile belirlenmesiyle ön savlar sınanır. Büyük örneklemlerde sonuçlar geçerlidir ancak küçük örneklemlerde bulgular daha iyi incelenerek yorumlanmalıdır (Gujarati ve Porter, 2012, s.558).

1.2.1.2. En çok olabilirlik yöntemi (EÇO)

Frekans değeri mevcut olmayan tekil düzeydeki verilerde EKK yönteminin kullanılamayacağı ifade edilmişti. Sebebinin ise EKK yönteminde kullanılacak olan ancak tekil düzey olan 0 ya da 1 değerlerinin P_i yerine yazılması ile anlamsız değerlerin ortaya çıkmasıydı.

$L_i = \ln \frac{1}{0}$ olayın gerçekleşmesi durumu

$L_i = \ln \frac{0}{1}$ olayın gerçekleşmemesi durumu

Bu durumda EÇO yöntemine başvurulur. EÇO yöntemi ile parametre tahmini, anakütle ile bu anakütleden çekilen örneklem arasındaki benzerlik ilişkisini kullanır. Söz konusu örneklemin elde edilme olasılığını maksimum yapan parametre değeri tahmin edilir (Piegorsch, 1992). Özetle EÇO yöntemi, elde edilecek olabilirlik fonksiyonunu maksimum yapan parametrelerin bulunması sürecidir.

Model için regresyon eşitliği;

$$y_i^* = x_i' \beta + u_i \tag{1.27}$$

şeklindedir. y_i^* gözlenemeyen değişkendir. Eldeki gölge değişken y şu şekilde tanımlanır;

$$\begin{aligned} y &= 1 \quad y^* > 0 \\ y &= 0 \quad d.d. \end{aligned} \tag{1.28}$$

Burada doğrusal olasılık modelinde olduğu gibi $x_i'\beta$, $E\left(y_i \middle| x_i\right)$ değerine eşit değildir. (1.27) ile (1.28)'dan,

$$P\left(y_i = 1\right) = P\left(u_i > -x_i'\beta\right) \tag{1.29}$$

$$= 1 - F\left(-x_i'\beta\right)$$

şeklinde elde edilir.

Olabilirlik fonksiyonu ise;

$$L = \prod_{y_i=0}\left[F\left(-x_i'\beta\right)\right]\prod_{y_i=1}\left[1 - F\left(-x_i'\beta\right)\right] \tag{1.30}$$

şeklindedir.

(1.30)'de kullanılan BDF (1.27)'deki u_i'nun dağılımının varsayımına bağlıdır. Buna bağlı olarak u_i 'nin BDF'sinin logistik olduğu kabul edilirse;

$$F\left(-x_i'\beta\right) = \frac{\exp\left(-x_i'\beta\right)}{1+\exp\left(-x_i'\beta\right)} = \frac{1}{1+\exp\left(-x_i'\beta\right)} \tag{1.31}$$

olur. Bu durumda;

$$1 - F\left(-x_i'\beta\right) = \frac{\exp\left(x_i'\beta\right)}{1+\exp\left(x_i'\beta\right)} \tag{1.32}$$

elde edilir. (1.30)'den yalnızca β/σ tahmin edilebilir. Bu oranın parametleri ayrı ayrı tahminlenmez. Ancak $\sigma = 1$ olarak varsayılarak başlanabilir.

(1.30), BDF'nin logistik olduğu bilgisi ile düzenlenirse;

$$L = \prod_{i=1}^{n}\left[\frac{1}{1+\exp\left(x_i'\beta\right)}\right]^{1-y_i}\left[\frac{\exp\left(x_i'\beta\right)}{1+\exp\left(x_i'\beta\right)}\right]^{y_i} \tag{1.33}$$

$$L = \frac{\exp(\beta') \sum_{i=1}^{n} x_i y_i}{\prod_{i=1}^{n} \left[1 + \exp\left(x_i' \beta\right)\right]} \tag{1.34}$$

elde edilir.

$t^* = \sum_{i=1}^{n} x_i y_i$ olarak tanımlansın. β 'nın EÇO tahmin edicisinin elde edilmesi için

$$\log L = \beta' t^* - \sum_{i=1}^{n} \log\left[1 + \exp\left(x_i' \beta\right)\right] \tag{1.35}$$

$\frac{\partial \log L}{\partial \beta} = 0$ olmak üzere

$$s(\beta) = -\sum_{i=1}^{n} \frac{\exp\left(x_i' \beta\right)}{1 + \exp\left(x_i' \beta\right)} x_i + t^* = 0 \tag{1.36}$$

elde edilir.

Buradan elde edilecek β değeri doğrusal olmayacaktır. Bu aşamada Newton-Raphson ya da hesaplama (scoring) yöntemleri eşitliğin çözümü için kullanılabilir (Maddala, 1983, s.25).

- Newton-Raphson yöntemi:

  Genel anlamda doğrusal olmayan bir denklemi doğrusallaştırmada kullanılan temel yaklaşım Taylor seri açılımıdır. Taylor seri açılımından yararlanarak doğrusal olmayan regresyon modellerinin tahmininde Gauss-Newton ve Newton-Raphson yineleme (iteration) yöntemleri kullanılır. Ayrıca Marquard yöntemi adında farklı bir yöntem vardır ki, bu yöntem de en hızlı düşüş yöntemi ile doğrusallaştırma yöntemlerinin birleşimidir (Gujarati ve Porter, 2012).

  Başlık gereği Newton metodu olarak da bilinen Newton-Raphson yineleme yönteminden kısa ve basit şekilde aşağıdaki gibi bahsedilebilir,

  r ; $f(x) = 0$ denkleminin köklerinden biri olmak üzere aynı zamanda $f(x) = y$ eğrisinin $Ox-$ ekseninde kestiği apsis noktalarından biridir.

1. Öncelikle x_0 başlangıç noktası tahmini olarak belirlenir. Bu tahmin fonksiyondan yararlanılarak yapılabilir.
2. x_0'dan yararlanarak r'ye daha yakın olan x_1; x_1'den yararlanarak r'ye daha yakın olan x_2 elde edilerek sırası ile $x_1, x_2, ..., x_n, x_{n+1}$ bulunur.
3. $x_1, x_2, ..., x_n, x_{n+1}$ şeklinde oluşturulacak olan (x_n) dizisi $f(x) = y$ eğrisinin $(x_n, f(x_n))$ noktasındaki teğet eğimi olan $m = f'(x_n)$ olmak üzere,

$$f'(x_n) = \frac{f(x_n) - 0}{x_n - x_{n+1}}$$

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$$

indirgeme bağıntısı elde edilmiş olur.

4. $x_1, x_2, ..., x_n, x_{n+1}$ şeklinde r'ye yakınsayan bir (x_n) dizisi oluşturulmuştur ve n ne kadar büyük alınırsa r de o kadar x_n'e yakın olmuş olur $\left(x_n \xrightarrow{n \longrightarrow \infty} r\right)$. (Balcı, 2008, s.383)

Bu durumda $\theta = (\alpha, \delta)'$ tahmin edilecek parametre vektörü olmak üzere $\log L$ log-olabilirlik fonksiyonunun eğim değişim oranı türev ile elde edilecektir. Bu durumda Hessian matrisi,

$$\frac{\partial^2 \log L}{\partial\theta\partial\theta'} = \begin{pmatrix} \frac{\partial^2 \log L}{\partial\alpha\partial\alpha} & \frac{\partial^2 \log L}{\partial\alpha\partial\delta} \\ \frac{\partial^2 \log L}{\partial\delta\partial\alpha} & \frac{\partial^2 \log L}{\partial\delta\partial\delta} \end{pmatrix} \tag{1.37}$$

olarak elde edilir.

$\frac{\partial^2 \log L}{\partial\alpha\partial\alpha}$ değeri $\frac{\partial^2 \log L}{\partial\delta\partial\delta}$ değerinden büyükse α'daki değişim, δ'daki değişime göre eğimi daha hızlı değiştirir. α değerinin tahmini için gerekli düzeltme,

$$\theta_{n+1} = \theta_n - \left(\frac{\partial^2 \log L}{\partial\theta_n\partial\theta_{n+1}}\right)^{-1} \frac{\partial \log L}{\partial\theta_n} \tag{1.38}$$

şeklinde olur (Cafri, 2009, s.33).

- Hesaplama (scoring) yöntemi: Hessian matrisinin beklenen değeri bilgi matrisi olarak adlandırılabilir. Bilgi matrisinin kontrol matrisi olarak kullanıldığı yöntem hesaplama metodudur (Long, 1997).

$$\theta_{n+1} = \theta_n + \left(E\left[\frac{\partial^2 \log L}{\partial \theta_n \partial \theta_{n+1}} \right] \right)^{-1} \frac{\partial \log L}{\partial \theta_n} \tag{1.39}$$

olmak üzere bilgi matrisi,

$$I(\beta) = E\left[-\frac{\partial^2 \log L}{\partial \beta \partial \beta'} \right] \tag{1.40}$$

şeklindedir ve (1.40) yardımıyla logit model için

$$I(\beta) = \sum_{i=1}^{n} \frac{\exp\left(x_i'\beta\right)}{1+\exp\left(x_i'\beta\right)^2} x_i x'_i \tag{1.41}$$

bulunur. İterasyon sürecinde yukarıda da ifade edildiği gibi β_0 başlangıç değeri ile $S(\beta_0)$ ve $I(\beta_0)$ elde edilir. Bu durumda β için tahmin şu şekilde elde edilir;

$$\beta_1 = \beta_0 + \left[I(\beta) \right]^{-1} S(\beta_0) \tag{1.42}$$

$S(\beta_0)$ ve $I(\beta_0)$ örneklem büyüklüğüne bölünür. İterasyon süreci yakınsama için devam ettirilecektir. Böylelikle başlangıç değerinin ne olduğu önemini yitirir ve köke yakınsama sağlanır. β tahmininden sonra *i*. gözlem için 1 olasılıklı gözlem değerleri tahmin edilir.

$$\hat{p}_i = \frac{\exp\left(x_i'\hat{\beta}\right)}{1+\exp\left(x_i'\hat{\beta}\right)} \tag{1.43}$$

$$\sum \hat{p}_i x_i = \sum y_i x_i \tag{1.44}$$

(1.35) ve (1.42) eşitlikleri kullanılarak (1.44) yazılır ve $\hat{\beta}$ ile $\hat{p}_i$ tahminlerinin (1.43) ile bir tür sağlaması yapılmış olur (Cafri, 2009).

Logit model için devam edilecek olursa hesaplama yöntemiyle yukarıda da ifade edildiği gibi, β_0 başlangıç değeri ile $S(\beta_0)$ ve $I(\beta_0)$ elde edilir. Bu durumda β için tahminleme yapılır. Burada $I(\beta)$ sürekli olarak pozitif olacaktır ve başlangıç değerinin ne olduğu önem ifade etmeksizin yakınsama sağlanacaktır (Maddala, 1983, s.27).

1.2.2. Probit model

Doğrusal olasılık modelinin en önemli problemlerinden olan ve en güçlü eleştirileri aldığı $0 \le E\left(y_i \left| x_i \right.\right) \le 1$ varsayımının sağlanamayışının çözüm yöntemlerinden biri de probit modeldir.

İki değerli bağımlı değişkenin açıklanabilmesi için uygun bir BDF kullanılmalıdır. Bu aşamada lojistik dağılımdan yararlanılarak logit model elde edilir. Normal BDF'den yararlanılarak elde edilen model ise probit (normit) model olarak isimlendirilir.

$$P_i = E\left(Y = 1 \left| x_i \right.\right) = \frac{1}{1 + \exp\left(-x_i^{'}\beta\right)} \tag{1.45}$$

Burada normal BDF'nin kullanımı ile işlemlere devam ettirilmesi probit modeli ortaya çıkaracaktır. McFadden tarafından geliştirilen fayda kuramına dayanan probit model üzerinde durulmaktadır (Gujarati, 2004, s.608).

McFadden tarafından geliştirilen fayda kuramında, *i.* bireyin tercih kararı gözlenemeyen bir fayda endeksine $\left(I_i\right)$ bağlıdır.

$$I_i = \beta_1 + \beta_2 x_i \tag{1.46}$$

I^*_i şeklinde ifade edilecek olan bir eşik değeri söz konusu olsun. I_i değeri, eşik değeri olan I^*_i aşarsa başarılı olay $Y = 1$, aşamazsa $Y = 0$ başarısız olay söz konusu olacaktır. I^*_i eşik değeri de I_i gibi gözlenemez ancak aynı ortalama ve varyans ile normal dağıldığı varsayılır. Bu durumda yalnızca I_i indeksinin anakütle katsayıları tahmin edilmiş olmaz aynı zamanda indekse dair bilgi de edinilir.

Normallik varsayımı altında ve standartlaştırılmış normal BDF'den yararlanarak, I^*_i'nin I_i'den küçük ya da I_i'ye eşit olması olasılığı;

$$P_i = P\left(Y = 1 \left| x_i \right.\right) = P\left(I^*_i \le I_i\right) = P\left(Z_i \le \beta_1 + \beta_2 x_i\right) = F\left(\beta_1 + \beta_2 x_i\right) \tag{1.47}$$

şeklindedir. Bu olasılık x_i değeri için bulunacak olasılığı ifade eder. Z_i standartlaştırılmış normal değişkendir $\left(Z_i \sim N\left(0, \sigma^2\right)\right)$. X değişkeninin μ ortalama ve σ^2 varyansla normal dağıldığı biliniyorsa sırasıyla OYF ve BDF;

$$f(x) = \frac{1}{\sqrt{2\sigma^2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \tag{1.48}$$

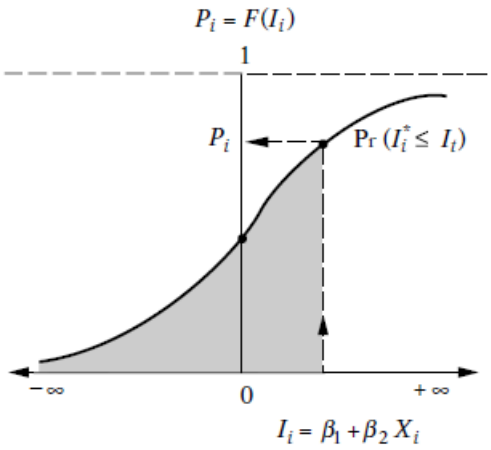
$$F(x) = \int_{-\infty}^{x_0} \frac{1}{\sqrt{2\sigma^2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \tag{1.49}$$

şeklinde olur. Bu durumda yukarıda bahsedilen eşitlik;

$$F(I_i) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{I_i} e^{-\frac{1}{2}(z)^2} dz \tag{1.50}$$

$$F(I_i) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\beta_1+\beta_2 x_i} e^{-\frac{1}{2}(z)^2} dz$$

şeklinde olacaktır ve burada ifade edilen alan Şekil 1.1'de gösterilmiştir.



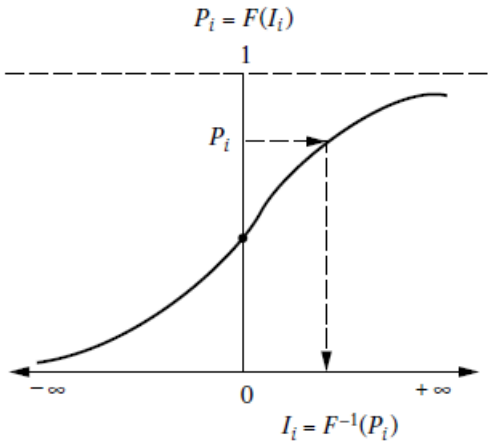
Şekil 1.1. *Fayda veri iken olasılığın değeri*

Kaynak: *Gujarati, 2004, s.609*

Bu durumun yanı sıra (1.47)'nin tersi alınarak β_1 ve β_2 'ye ilişkin bilgi edinilebilir.

$$I_i = F^{-1}(I_i) = F^{-1}(P_i) = \beta_1 + \beta_2 x_i$$

Şekil 1.1'de $I^*_i \leq I_i$ iken başarı olasılığı dikey eksende bulunurken, aşağıdaki Şekil 1.2'de, F^{-1} normal BDF'nin tersi olduğundan P_i veriyken I_i 'nin değeri yatay eksenden elde edilir.



Şekil 1.2. *Olasılık veri iken faydanın değeri*

Kaynak: *Gujarati, 2004, s.609*

Ancak burada önemli konu I_i, β_1 ve β_2'nin nasıl elde edileceğidir (Gujarati, 2004).

Bu aşamada tahminleme yapılacaktır. Logit modelde de olduğu gibi verilerin doğası bu basamakta önemlidir.

Logit modelde başlıklar altında incelenen tekil düzeydeki veri (frekans değeri mevcut olmayan veri) ile gruplanmış veri ya da yinelenmiş veri (yinelenme (tekrarlanma) sayıları bilinen veri) için kullanılacak uygun tahminle yöntemlerinin sırasıyla EÇO ve EKK tahmin edicileri olduğu ifade edilmelidir.

Frekans değeri mevcut olmayan verilerde tekil düzey (mikro ya da kişisel düzey) sebebi ile 0 ya da 1 değerlerinin yerine yazılması ile anlamsız değerlerin ortaya çıkması, EÇO yöntemini gerekli kılar. Her x_i gelir seviyesi için göreli sıklığın hesaplanıp, bu gelir seviyeleri için gerçek P_i tahmininde söz konusu değerlerin $\left(\hat{P}_i = \frac{n_i}{N_i}\right)$ kullanılması ile EKK yönteminden bahsedilebilir.

1.2.2.1. En küçük kareler yöntemi

Yukarıda da ifade edildiği üzere frekanslı seriler için kullanılacak olan EKK yöntemi için,

1. Öncelikle göreli sıklık değeri hesaplanır $\left(\hat{P}_i = \frac{n_i}{N_i}\right)$.
2. Şekil 1.2'de verilen F^{-1}'nin (normal BDF'nin tersi) P_i veriyken I_i'nin değerini verdiği bilinmektedir. Elde edilmek istenen $\hat{P}_i$ değerleri için tahmin edilen $I_i = \hat{I}_i$ değerleri ters fonksiyonun yardımı ile bulunur.
3. Probit model analizinde gözlenemeyen fayda endeksine (I_i), normit ya da normal eşdeğer sapma denmektedir. $P_i < 0$ için I_i negatif olur. Bu durumda $N.E.S.$ değerine 5 eklenir ve probit değeri elde edilir.

$$Probit = N.E.S. + 5 = I_i + 5 \tag{1.51}$$

4. β'nın tahmin edilebilmesi için,

$$I_i = \beta_1 + \beta_2 x_i + u_i \tag{1.52}$$

eşitliğinden faydalanılacağı gibi $I_i = \hat{I}_i$ değeri ya da (1.51) eşitliğindeki probit değer bağımlı değişken olarak kullanılabilir. Burada eğim ve belirginlik katsayısı sabit kalmak üzere yalnızca sabit terimde farklılaşma söz konusu olur.

5. Veriler, (1.52) modelinin EKK ile tahmin edilmesi ile ortaya çıkacak olan değişen varyans probleminin çözümü için dönüştürülmelidir. Logit modelde EKK yöntemi kullanılırken gerekli olan ağırlıklandırma işlemi burada da mevcuttur. Hata teriminin varyansının karekökü ağırlık olarak kullanılır. f_i, $F^{-1}(P_i)$'deki standart normal olasılık yoğunluk foksiyonu (OYF) olmak üzere,

$$\sigma_u^2 = \frac{P_i(1-P_i)}{N_i f_i} \tag{1.53}$$

dönüşümü yapılarak ağırlıklı EKK yöntemi kullanılır (Eren, 2012, s.14).

1.2.2.2. En çok olabilirlik yöntemi

Probit model için yazılan regresyon eşitliği;

$$y_i^* = x_i'\beta + u_i \tag{1.54}$$

olmak üzere y_i^* gözlenemez ve eldeki gölge değişken y şu şekilde tanımlanırsa;

$$y = 1 \qquad y^* > 0 \tag{1.55}$$

$$y = 0 \qquad d.d.$$

burada doğrusal olasılık modelinde olduğu gibi $x_i'\beta$, $E\left(y_i \middle| x_i\right)$ değerine eşit değildir ve (1.27) ile (1.28)'dan,

$$P\left(y_i = 1\right) = P\left(u_i > -x_i'\beta\right) \tag{1.56}$$

$$= 1 - F\left(-x_i'\beta\right)$$

olarak bulunur. Olabilirlik fonksiyonu ise;

$$L = \prod_{y_i=0}\left[F\left(-x_i'\beta\right)\right]\prod_{y_i=1}\left[1 - F\left(-x_i'\beta\right)\right] \tag{1.57}$$

olur. Burada önemli nokta (1.30)'da kullanılan BDF'dir. Bu BDF de (1.27)'deki u_i varsayımına bağlıdır. u_i'nin dağılımının $N\left(0, \sigma^2\right)$ olduğu biliniyorsa,

$$F\left(-x_i'\beta\right) = \int_{-\infty}^{-x_i'\beta/\sigma} \frac{1}{\left(2\pi\right)^{1/2}} \exp\left(-\frac{t^2}{2}\right) dt \tag{1.58}$$

olacaktır. (1.32)'dan yalnızca β/σ tahmin edilebilir. Bu oranın parametleri ayrı ayrı tahminlenmez.

(1.57) ve (1.58) kullanılarak yazılacak olabilirlik fonksiyonu, $\Phi(.)$ standart normal BDF ve $\phi(.)$ standart normal OYF'yi göstermek üzere,

$$L = \prod_{i=1}^{n}\left[\Phi\left(x_i'\beta\right)\right]^{y_i}\left[1 - \Phi\left(x_i'\beta\right)\right]^{1-y_i} \tag{1.59}$$

şeklindedir.

Log-olabilirlik fonksiyonu ise;

$$\log L = \sum_{i=1}^{n} y_i \log\Phi\left(x_i'\beta\right) + \sum_{i=1}^{n}\left(1 - y_i\right)\log\left[1 - \Phi\left(x_i'\beta\right)\right] \tag{1.60}$$

olur ve Log-olabilirlik fonksiyonunun β'ya görevi türevi alınır.

$$S(\beta) = \frac{\partial \log L}{\partial \beta} = \sum_{i=1}^{n} \frac{\left[y_i - \Phi\left(x_i^{'} \beta \right) \right]}{\Phi\left(x_i^{'} \beta \right) \left[1 - \Phi\left(x_i^{'} \beta \right) \right]} \phi\left(x_i^{'} \beta \right) x_i \tag{1.61}$$

(1.61) eşitliğinin 0'a eşitlenmesi ile EÇO tahmincisi $\left(\hat{\beta}_{ML} \right)$ elde edilir. Logit modelde olduğu gibi bu aşamada probit modelde de β parametresinin doğrusal olmamasından dolayı nümerik bir metoda başvurulur. Logit modelde ifade edilen farklı iterasyon süreçleri ile sonuca ulaşılabilir. Bu aşamalarda ihtiyaç duyulacak olan bilgi matrisi ise,

$$I(\beta) = E\left[-\frac{\partial^2 \log L}{\partial \beta \partial \beta'} \right] \tag{1.62}$$

şeklindedir ve (1.62) yardımıyla probit model için;

$$I(\beta) = \sum_{i=1}^{n} \frac{\left[\phi\left(x_i^{'} \beta \right) \right]^2}{\Phi\left(x_i^{'} \beta \right) \left[1 - \Phi\left(x_i^{'} \beta \right) \right]} x_i x_i^{'} \tag{1.63}$$

olarak elde edilir.

Logit model için EÇO yönteminin kullanımındaki gibi iterasyon sürecinde β_0 başlangıç değeri ile $S(\beta_0)$ ve $I(\beta_0)$ elde edilirken β için tahmin;

$$\beta_1 = \beta_0 + \left[I(\beta_0) \right]^{-1} S(\beta_0) \tag{1.64}$$

kullanılarak yapılır. Ayrıca burada da $S(\beta_0)$ ve $I(\beta_0)$ örneklem büyüklüğüne bölünür. Böylelikle iterasyon süreci yakınsama için devam ettirilir ve başlangıç değerinin ne olduğu önemli olmaksızın olabilirlik fonksiyonuna yakınsama sağlanır.

2. SANSÜRLENMİŞ VE KIRPILMIŞ REGRESYON MODELLERİ

Sınırlı bağımlı değişkenlerin kesikli olma durumunun yanında sürekli olduğu ve belli bir aralık değeriyle sınırlandırıldığı durumlar mevcuttur. Sınırlı değişkenli modellerin kesikli olma durumu ikili (binary) ya da dikotom (dichotomous) olarak en ilkel haliyle önceki başlıklarda ele alınmıştır. Bu başlık altında da sürekli bağımlı değişkenin alabileceği değerlerin belli bir aralık değeri ile sınırlandırılması ele alınacaktır. Bu durum sansürlenme ve kırpılma olmak üzere iki farklı şekilde ifade edilmektedir.

Sansürlenmiş veri yapısı ile kırpılmış veri yapılarının karıştırıldığı durumlar söz konusudur. Bu aşamada bu iki farklı veri yapısını ayırt etmek öncelikli ve önemlidir. Belli bir aralığın dışında kalan veriler tamamen kayıpsa kırpılmış (truncated) regresyon modeli söz konusudur. Ancak kayıp veriler için bağımsız değişkenlerin bilgisi mevcut ise bu durumda sansürlenmiş regresyon modeli söz konusu olacaktır. Diğer bir ifade ile bağımlı değişkende bilgi kaybı varken bağımsız değişken bilgileri mevcutsa sansürlü veri; hem bağımlı hem de bağımsız değişken için bilgi kaybı varsa kırpılmış veri söz konusudur (Davidson ve MacKinnon, 1999 s.473).

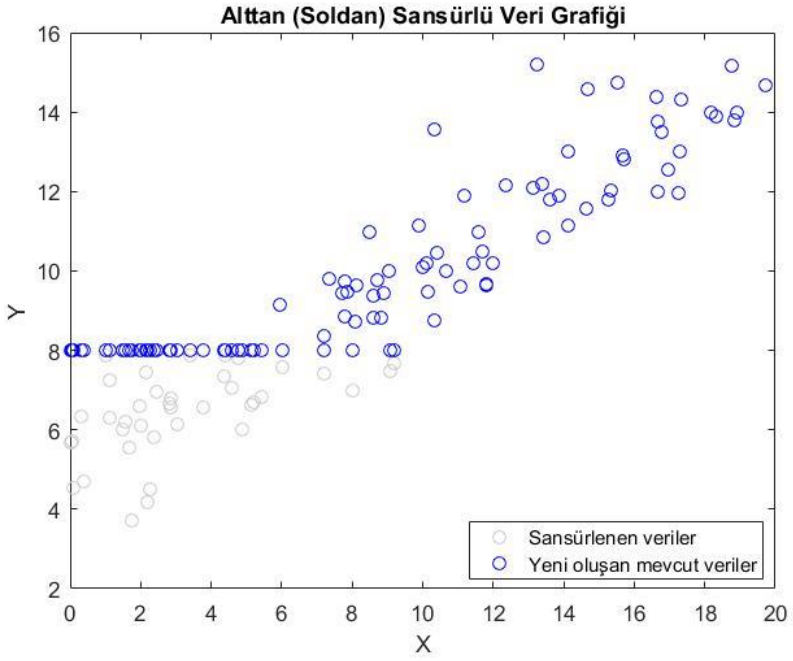
Sansürlenmiş ve kırpılmış veri yapısı sırası ile aşağıdaki gibi ayrıntılandırılmıştır.

- Sansürlenmiş Veri:
    - Bağımsız değişken gözleniyorken bağımlı değişken bilgisi yoktur.
    - Sistematik bir sınırlama veri seti için söz konusu değildir. Verinin doğası gereği sansürleme ortaya çıkar.
    - İki şekilde sansürlenmiş veri söz konusu olabilir:
        - Ölçüm değerinin belli bir eşik değeri için mevcut olmaması. Bir araştırmada belli bir gelir düzeyi üstündeki değerlerin, araştırma kapsamında belirlenmiş bir değere sabitlenmesi örnek olarak gösterilebilir. Benzer durum, alt bir eşik ile belli bir değer altındaki gelir seviyelerinin belli bir değere sabitlenmesidir.
        - Yaşam analizinde bağımlı değişken olarak ele alınan yaşam süresidir. Yaşam süresi ise başarısızlık zamanına kadar geçen süredir. Araştırma süresinde başarısızlığın gerçekleşmediği veriler sansürlenmiş veri olarak kabul edilir (Yılmaz vd., 2013). Örneğin

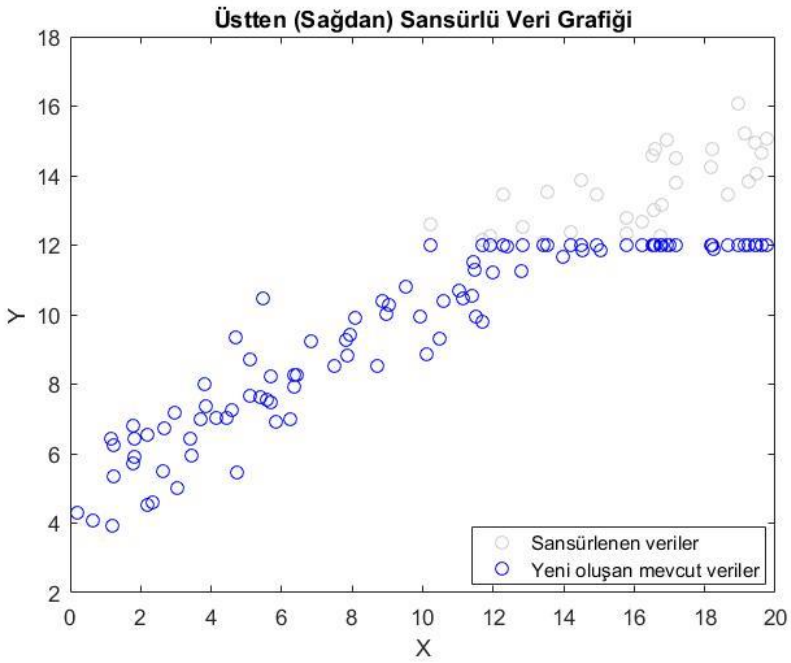
bir araştırma verilerinin (hastaların yaşam süresi) bir kısmı, henüz araştırma tamamlanmadan çeşitli nedenlerle (ölüm, yer değiştirme, araştırmaya katılmaktan vazgeçme v.b.) araştırmadan ayrılırsa sansürlenmiş veri olarak ele alınır. Bu veri grupları içinde sansürlü regresyon uygulamaları mevcuttur (Yenilmez, Kantar, 2016, s.107; Zorlutuna vd., 2016, s.13).

Şekil 2.1 ve Şekil 2.2'de sırasıyla soldan (alttan) ve sağdan (üstten) sansürlenmiş verilerin görselleri sunulmuştur. Dikkat edilmesi gereken noktalar,

1. Soldan sansürleme için sansür noktasından önceki; sağdan sansürleme için sansür noktasından sonraki değerlerin sansür noktasına eşit olmasıdır.
2. Sansür öncesi ve sansür sonrası örneklem sayısının değişmemesidir.



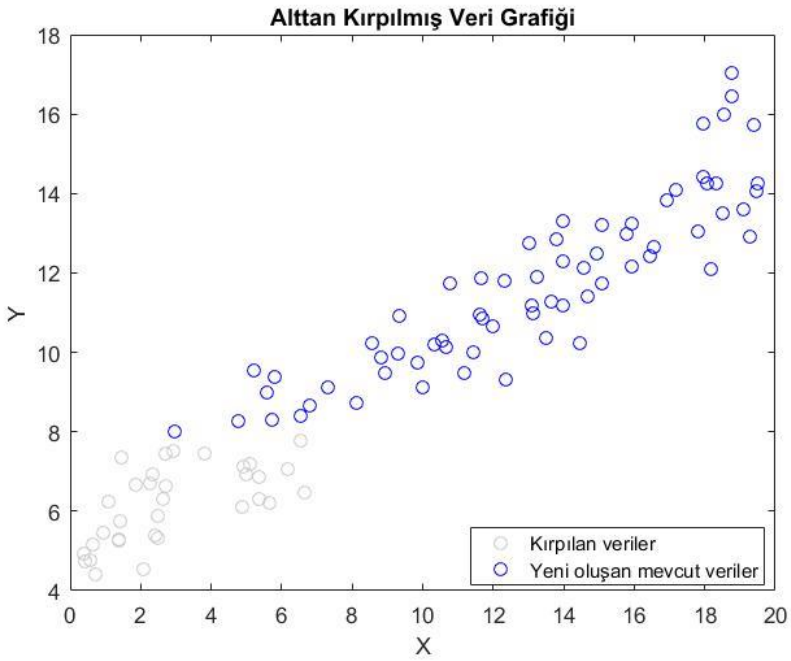
Şekil 2.1. *Soldan (alttan) sansürlü veri grubu*



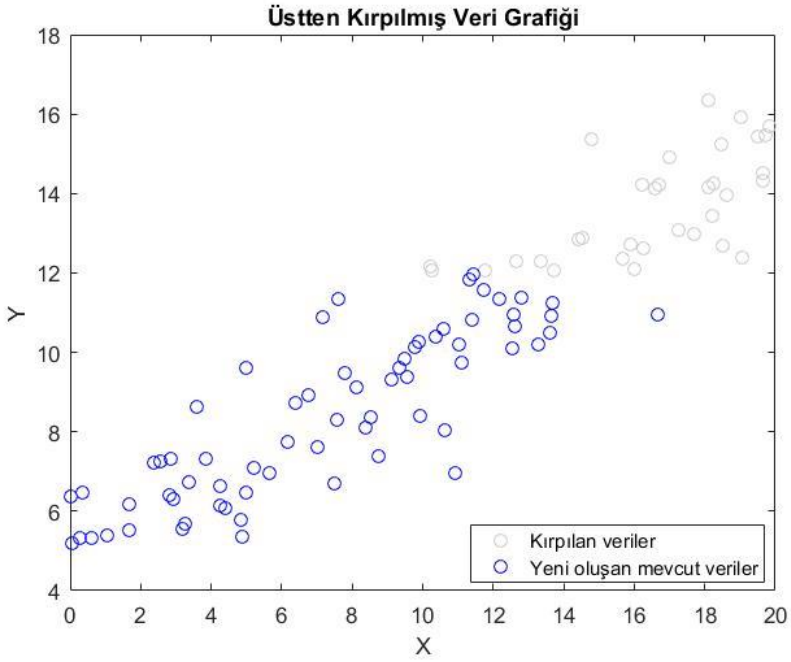
Şekil 2.2. *Sağdan (üstten) sansürlü veri grubu*

- Kırpılmış Veri:
    - Bağımlı-bağımsız değişkenlerin belli değerlerle sınırlandırıldığı kayıp değerleri içeren veri gruplarıdır.
    - Veri seti için sistematik bir sınırlama söz konusudur.
    - Örneğin 60 yaş sınırının zorunlu emeklilik yaşı olduğu ve kişilerin emeklilikten sonra herhangi bir şekilde çalışma hakkının bulunmadığı ülkede çalışma tercihi 60 yaş üstü için kırpılmıştır.

Şekil 2.3 ve Şekil 2.4'de de sırasıyla soldan ve sağdan kırpılmış verilerin görselleri sunulmuştur. Açıktır ki sansürlemeden farklı olarak kırpılma noktasından sonraki veriler örneklemden tamamen çıkarılmaktadır. Bu durum çalışılan örneklemin değiştiğinin açık ve net bir göstergesidir.



Şekil 2.3. *Soldan (alttan) kırpılmış veri grubu*



Şekil 2.4. *Sağdan (üstten) kırpılmış veri grubu*

2.1. Sansürlenmiş ve Kırpılmış Değişkenler

Sansürleme ve kırpılma farklı bölgelerden (sağdan-üstten, soldan-alttan) farklı sınırlandırmalarla (sansürleme, kırpma) ifade edilebilir.

- Soldan sansürlenmiş örneklem: $Y*$ rassal değişken olmak üzere n tane değeri içeren örneklem $\left(y_1^*, y_2^*, ..., y_n^*\right)$ için c sansür noktası olmak üzere,

$$\begin{aligned} y_i &= y_i^* \qquad y_i^* > c \\ y_i &= c \qquad d.d. \end{aligned} \tag{2.1}$$

şeklinde ifade edilebilir.

Buradan elde edilecek örneklem $\left(y_1, y_2, ..., y_n\right)$ sansürlenmiş örneklemdir. (2.1)'den anlaşılacağı üzere $y_i^* \le c$ durumunda $y_i = c$ gözlem değeri olarak alınmaktadır.

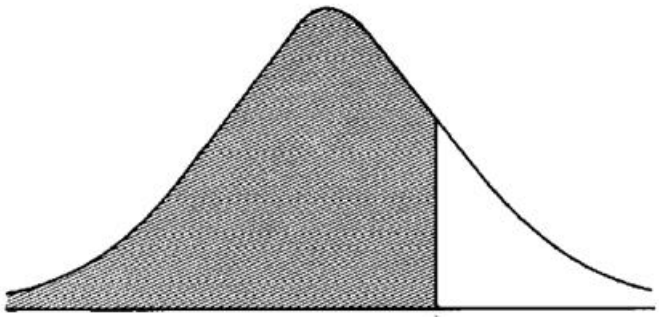
$$P\left(y_i = c\right) = P\left(y_i^* \le c\right) \tag{2.2}$$

$F\left(.\right)$ ve $f\left(.\right)$ sırasıyla $Y*$ rassal değişkeninin BDF ve OYF'si olmak üzere parametrelerin tahmini için yazılacak olabilirlik fonksiyonu,

$$L = K_1\left[F\left(c\right)\right]^{n_c} \prod_{i=1}^{n-n_c} f\left(x_i\right) \tag{2.3}$$

şeklinde olacaktır. Burada soldan sansürleme söz konusudur. Sansürlenen veri sayısı n_c olarak kabul edilmiştir. K_1 ise parametrelere bağlı olmayan düzenleme katsayısıdır (Cohen, 1991, s.6).

- Sağdan kırpılmış örneklem: Farklı örnek teşkil etmesi için burada da sağdan kırpılmış değişkenlerin elde edilmesi için $y_i^* = t$ kırpılma noktası olarak alınıp $y_i^* > t$ durumunda gözlem değerleri sınırlandırılırmıştır. Kolaylığı nedeni ile normal dağılım ile çalışılırsa tüm gözlemler Şekil 2.5'de gösterilen taralı bölgeden gelecektir.



Şekil 2.5. *Kırpılmış Normal Dağılım*

Kaynak: *Maddala, 1983 s.150*

Sağdan kırpılmış OYF, $\Phi\left(\frac{t-\mu}{\sigma}\right)$ normalleştirme katsayısı ve $-\infty < y^* \leq t$ olmak üzere;

$$f\left(y^* \middle| y^* < t\right) = \frac{1}{\sigma}\phi\left(\frac{y_i - \mu}{\sigma}\right) \Big/ \Phi\left(\frac{t-\mu}{\sigma}\right) \tag{2.4}$$

şeklinde olacaktır.

t noktasından sağdan kırpılmış normal dağılımın olabilirlik fonksiyonu yazılacak olursa,

$$L = \frac{1}{\Phi(t)} \prod_{i=1}^{n} \phi(x_i) \tag{2.5}$$

şeklindedir.

- Hem soldan hem sağdan (iki katlı) kırpılmış ve sansürlenmiş örneklem: Aslında olabilirlik fonksiyonu sansürleme ve kırpılma tiplerine göre (yalnızca sağdan veya yalnızca soldan kırpılmış, yalnızca sağdan veya soldan sansürlenmiş, iki katlı kırpılmış veya sansürlenmiş, merkezi sansürlenmiş veya kırpılmış, aşamalı (progressively) sansürlenmiş) ayrı ayrı yazılabilir (Cohen, 1991, s.3). Örneğin iki katlı (hem sağdan hem soldan) kırpılmış ve sansürlenmiş örneklemler için sırasıyla olabilirlik fonksiyonu, sansürlenme noktaları c_1 ve c_2; kırpılma noktaları t_1 ve t_2 olmak üzere şu şekilde yazılabilir;

$$L = \frac{1}{F(t_2) - F(t_1)} \prod_{i=1}^{n} f(x_i), \qquad t_1 \le x_i \le t_2 \tag{2.6}$$

$$L = K_2 \left[F(c_1) \right]_{(n_{c_1})} \left[1 - F(c_2) \right]_{(n_{c_2})} \prod_{i=1}^{n-n_{c_1}-n_{c_2}} f(x_i) \qquad c_1 \le x_i \le c_2 \tag{2.7}$$

(2.7)'de kullanılan K_2, (2.3) de kullanılan K_1 gibi parametrelere bağlı olmayan düzenleme katsayısıdır. Bu katsayı uygulanan sınırlamaya göre olabilirlik fonksiyonlarında farklılık gösterir. Ayrıca özel olarak (2.7)'de kullanılan n_{c_1} soldan sansürlenen örneklem sayısını, n_{c_2} ise sağdan sansürlenen örneklem sayısını ifade eder.

- Soldan sansürlü sağdan kırpılmış örneklem: Yukarıda verilen iki katlı kırpılmış ve sansürlenmiş örneklemlerin yanında, özel olarak aynı anda hem kırpılmış hem de sansürlenmiş örneklemlerle de çalışılabilir. Örneğin t kırpılma noktası, c sansürlenme noktası olmak üzere ($c < t$),

$$\begin{aligned} y_i &= y_i^* \qquad y_i^* > c \\ y_i &= c \qquad d.d. \end{aligned} \tag{2.8}$$

şeklinde kurulan modelde normal dağılım için olabilirlik fonksiyonu şu şekilde yazılacaktır;

$$L\left(\mu, \sigma^2 \middle| y_1, y_2, ..., y_n\right) = \left[\Phi\left(\frac{t-\mu}{\sigma} \right) \right]^{-n} \prod_{y_i^* > c} \frac{1}{\sigma} \phi\left(\frac{y_i - \mu}{\sigma} \right) \prod_{y_i^* \le c} \Phi\left(\frac{c-\mu}{\sigma} \right) \tag{2.9}$$

(Maddala, 1983 s.150)

2.2. Sansürlenmiş ve Kırpılmış Dağılımlar

2.2.1. Kırpılmış dağılımlar

Kırpılmış dağılım, kırpılma sonucu kalan kırpılmamış değerlerin dağılımıdır. Kırpılma, örneklem verisinin dağılımının karakteridir. Kırpılmanın anlaşılması ve modellerin uygulanması için özellikle kırpılmış dağılımların açık şekilde anlaşılması gerekmektedir. Ayrıca kırpılma durumunun teorik sonuçlarının sansürlü modellerin analizinde gerekli olduğu görülecektir (Greene, 2011, s.833).

2.2.1.1. Kırpılmış rassal değişkenin yoğunluğu

X sürekli rassal değişkeni için OYF $f(x)$ ve t soldan (alttan) kırpılma noktası olmak üzere;

$$f\left(x|x>t\right)=\left(\frac{f\left(x\right)}{P\left(x>t\right)}\right)=\left(\frac{f\left(x\right)}{1-P\left(x<t\right)}\right)=\left(\frac{f\left(x\right)}{1-F\left(t\right)}\right) \tag{2.10}$$

şeklinde elde edilir. Sağdan (üstten) kırpma işleminin uygulanmasıyla da $f\left(x|x<t\right)=\left(f\left(x\right)/P\left(x<t\right)\right)=\left(f\left(x\right)/F\left(t\right)\right)$ elde edilir. Kırpılmış dağılım bir tür koşullu dağılım olarak tanımlanır. Sürekli rassal değişken söz konusu olduğunda kırpılmış dağılım olarak kırpılmış normal dağılımın yaygın kullanımı göze çarpmaktadır.

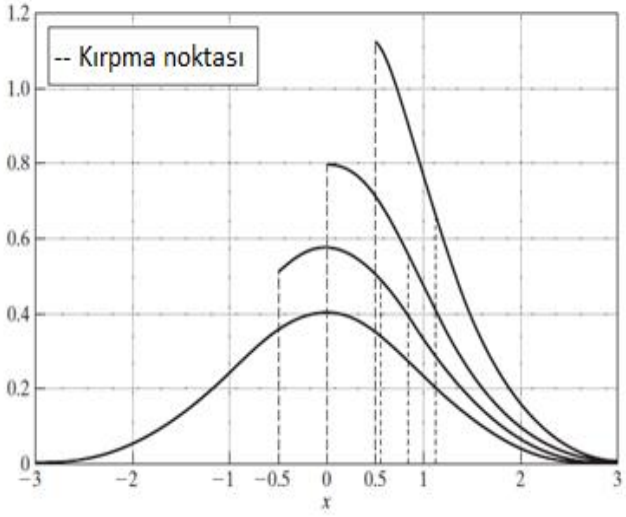
$X\sim N(\mu,\sigma^2)$, t eşik değeri ve $\alpha=\dfrac{t-\mu}{\sigma}$ olmak üzere;

$$P\left(x>t\right)=1-\Phi\left(\frac{t-\mu}{\sigma}\right)=1-\Phi\left(\alpha\right) \tag{2.11}$$

şeklinde olacaktır. O halde kırpılmış normal dağılımın yoğunluğu,

$$f\left(x|x>t\right)=\frac{f\left(x\right)}{1-\Phi\left(\alpha\right)}=\frac{(2\pi\sigma^2)^{-1/2}\exp\left(-\dfrac{(x-\mu)^2}{2\sigma^2}\right)}{1-\Phi\left(\alpha\right)}=\frac{\dfrac{1}{\sigma}\phi\left(\dfrac{x-\mu}{\sigma}\right)}{1-\Phi\left(\alpha\right)} \tag{2.12}$$

şeklinde olur. Ayrıca $\mu=0$ ve $\sigma=1$ için kırpılmış standart normal dağılımda elde edilebilir. $t=-0.5,0,0.5$ için tek bir şekilde üzerinde kırpılmış normal dağılım Şekil 2.6'de gösterilmiştir.



Şekil 2.6. *Kırpılmış Normal Dağılımlar*

Kaynak: *Greene, 2011, s. 835*

Burada dikkat edilmesi gereken noktalar örneklemin bozulduğu ve orijinal dağılımın bir alt kümesi ile çalışıldığıdır. Kırpılma sağdan (üstten) olursa koşullu beklenen değer orijinal beklenen değerden, koşullu varyansta orijinal varyanstan küçük olacaktır. Kırpılma soldan (alttan) olursa koşullu beklenen değer orijinal beklenen değerden bu defa büyük, koşullu varyans ise orijinal varyanstan yine küçük olacaktır. Soldan kırpma işlemi sonucunda oluşan durum eşitsizlikler yardımı ile gösterilmiştir:

$$E(X \mid X > t) > E(X) \tag{2.13}$$

$$Var(X \mid X > t) < Var(X)$$

Normal dağılımda olduğu gibi diğer sürekli ve kesikli dağılımlar için sağdan, soldan veya her iki taraftan kırpılmış olan dağılımlar ayrı olarak oluşturulabilir (Cohen, 1991).

2.2.1.2. Kırpılmış dağılımın momentleri

Kırpılmış rassal değişkenlerin ortalama ve varyans değerleri, sürekli rassal değişken için aşağıdaki gibi ele alınır. Bu aşamada ortalama ve varyans için kullanılacak olan genel formül,

$$\mathrm{E}[X \mid X > t] = \int_{t}^{\infty} x f\left(x \mid x > t\right) dx \tag{2.14}$$

şeklindedir.

Bu durumu örneklemek gerekirse X rassal değişkeni standart düzgün dağılıma sahip olmak üzere ($X \sim U(0,1)$), $x = 1/4$ noktasından dağılım kırpılırsa;

$$f\left(x \mid x > 1/4\right) = \frac{f\left(x\right)}{\mathrm{P}\left(x > 1/4\right)} = \frac{1}{3/4} = \frac{4}{3} \qquad \frac{1}{3} \le x \le 1 \tag{2.15}$$

şeklinde olacaktır. Bu durumda beklenen değer,

$$\mathrm{E}[X \mid X > 1/4] = \int_{1/4}^{1} x \frac{4}{3} dx = \frac{16}{27} \tag{2.16}$$

olarak bulunur. Bunun yanında herhangi bir T rassal değişkeninin L ve U arasında düzgün dağıldığı biliniyorsa varyans $\frac{(U-L)^2}{12}$ ifadesinden yararlanılarak bulunur. O halde;

$$\mathrm{Var}[X \mid X > 1/4] = \frac{3}{64} \tag{2.17}$$

olarak bulunur.

Normal şartlar altında kırpılmamış bir düzgün dağılımda ortalama ve varyans değerleri sırası ile $\frac{1}{2}$ ve $\frac{1}{12}$ 'dir.

Bu sonuçlar ışığında örnekten çıkarılacak iki sonuç,

1. Kırpılma soldan ise kırpılmış değişkenlerin ortalaması kırpılmamış olan orijinal değişkenlerin ortalamasından daha büyüktür. Kırpılma sağdan ise bu defa kırpılmış değişkenlerin ortalaması kırpılmamış olan orijinal değişkenlerin ortalamasından daha küçüktür.
2. Kırpma işlemi sonucu ortaya çıkan dağılımların varyansları ile kırpılmamış dağılımların varyansları karşılaştırılırsa kırpma işleminin varyans değerini küçülttüğü ifade edilebilir.

2.2.2. Kırpılmış normal dağılım

Öncelikle normal dağılımı özetlemek faydalı olacaktır. Aşağıda tanıtılan normal dağılım bilgileri bu başlık altında kullanıldığı gibi sansürlü normal dağılım başlığında da (sansürlü normal dağılım başlığında tekrarlanmaksızın) kullanılacaktır.

Y rassal değişkeninin μ ortalama ve σ^2 varyans ile normal dağılıma sahip olduğu bilinsin bu durumda OYF;

$$f(y) = \frac{1}{\sqrt{2\sigma^2\pi}} e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2} \qquad -\infty < x < +\infty \tag{2.18}$$

şeklinde olur. $Z = \frac{Y-\mu}{\sigma}$ olmak üzere Z rassal değişkeni 0 ortalama ve 1 varyans ile normal dağılıma sahiptir ve Z rassal değişkeninin standart normal dağılıma sahip olduğu ifade edilir ve OYF;

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(z)^2} \tag{2.19}$$

olarak bulunur.

$Y \sim N(\mu, \sigma^2)$, Z standart normal dağılmış rassal değişken olmak üzere;

$$f(y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2} = \frac{1}{\sigma}\phi\left(\frac{y-\mu}{\sigma}\right) \tag{2.20}$$

bilgisinden yararlanılır. Bu durumda $Z = \frac{Y-\mu}{\sigma}$ rassal değişkeninin BDF $\Phi(z)$;

$$\Phi\left(\frac{y-\mu}{\sigma}\right) = \Phi(z) = P(Y \le y) \tag{2.21}$$

$$1 - \Phi\left(\frac{y-\mu}{\sigma}\right) = 1 - \Phi(z) = P(Y > y)$$

olur. Bu durumda ayrı olarak belirtmek gerekirse tesadüfi hata terimi olan ε, μ ortalama ve σ^2 varyans ile normal dağılıma sahip ise OYF;

$$f(\varepsilon)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{\varepsilon-\mu}{\sigma}\right)^2}=\frac{1}{\sigma}\phi\left(\frac{\varepsilon-\mu}{\sigma}\right) \tag{2.22}$$

şeklinde yazılır (Gujarati, 1999).

X rassal değişkeni; μ ortalamalı, σ standart sapmalı normal dağılıma sahip ise yani $X \sim N\left(\mu,\sigma^2\right)$ olduğunda, t kırpılma katsayısı olmak üzere; $\alpha=\frac{t-\mu}{\sigma}$ ve $f(x)=\left(2\pi\sigma^2\right)^{-1/2}\exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right)=\frac{1}{\sigma}\phi\left(\frac{x-\mu}{\sigma}\right)$ şeklinde yazılabilir. Bu aşamada eşitlik (2.11) kullanılarak ve $\Phi(.)$ ve $\phi(.)$ yardımı ile kırpılmış normal dağılımın yoğunluk fonksiyonu;

$$\begin{aligned} f\left(x|x>t\right) &= \frac{f(x)}{1-\Phi(\alpha)} \\ &= \frac{\left(2\pi\sigma^2\right)^{-1/2}e^{-(x-\mu)^2/2\sigma^2}}{1-\Phi(\alpha)} \\ &= \frac{\frac{1}{\sigma}\phi\left(\frac{x-\mu}{\sigma}\right)}{1-\Phi(\alpha)} \end{aligned} \tag{2.23}$$

olarak bulunur (Greene, 2003 s.757).

2.2.2.1. Kırpılmış normal dağılımın momentleri

Beklenen değer ve varyans için öncelikle moment üreten fonksiyonu (MÜF) ifade etmekte yarar vardır. Kırpılma noktası çalışma genelinde t olarak alınmıştır. MÜF içinde kullanılacak t değerinden dolayı notasyan karışıklığını önlemek adına bu başlık altında kırpılma noktası $(k=1,2)$ olmak üzere a_k alınmıştır. $X \sim N\left(\mu,\sigma^2\right)$ olmak üzere, $M(t)=E\left[e^{tX}\,\middle|\,X\in A\right]$ kullanılarak MÜF;

$$M(t)=\frac{\int_{a_1}^{a_2}e^{tx}f(x)dx}{\Phi\left(\frac{a_2-\mu}{\sigma}\right)-\Phi\left(\frac{a_1-\mu}{\sigma}\right)}=e^{\mu t+\sigma^2t^2/2}\,\frac{\Phi\left(\frac{a_2-\mu}{\sigma}-\sigma t\right)-\Phi\left(\frac{a_1-\mu}{\sigma}-\sigma t\right)}{\Phi\left(\frac{a_2-\mu}{\sigma}\right)-\Phi\left(\frac{a_1-\mu}{\sigma}\right)} \tag{2.24}$$

şeklindedir. Buradan

$$\frac{1}{\sigma\sqrt{2\pi}}\int_{a_1}^{a_2} e^{tx} e^{\frac{-1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = \frac{1}{\sigma\sqrt{2\pi}}\int_{a_1}^{a_2} e^{\frac{-1}{2\sigma^2}\left\{\left[x-\left(\sigma^2 t+\mu\right)\right]^2-\left(\sigma^2 t+\mu\right)^2+\mu^2\right\}} dx \tag{2.25}$$

$$= e^{\frac{-1}{2\sigma^2}\left[\mu^2-\left(\sigma^2 t+\mu\right)^2\right]} \frac{1}{\sigma\sqrt{2\pi}}\int_{a}^{a_2} e^{\frac{-1}{2}\left(\frac{x-\mu'}{\sigma}\right)^2} dx$$

$$= e^{\mu t+\sigma^2 t^2/2}\int_{a_1}^{a_2} \frac{1}{\sigma}\phi\left(\frac{x-\mu'}{\sigma}\right) dx$$

$$= e^{\mu t+\sigma^2 t^2/2}\left[\Phi\left(\frac{a_2-\mu'}{\sigma}\right)-\Phi\left(\frac{a_1-\mu'}{\sigma}\right)\right]$$

elde edilir. Burada $\mu' = \sigma t + \mu$ olarak alınmıştır.

MÜF'nun yerine yazılmasıyla,

$$E\left[X \middle| X \in A\right] = M'(t)\Big|_{t=0} = \mu - \sigma\frac{\phi(a_2)-\phi(a_1)}{\Phi(a_2)-\Phi(a_1)} \tag{2.26}$$

bulunur. Burada $\alpha_k = \frac{a_k-\mu}{\sigma}$'dir. Ayrıca a_2 sonsuza gittikçe,

$$E\left[X \middle| X > a_1\right] = \mu + \frac{\phi(a_1)}{1-\Phi(a_1)} = \mu + \sigma\lambda(a_1) \tag{2.27}$$

olacaktır. Burada $\lambda(\alpha)$ tehlike fonksiyonudur ve pozitiftir.

a_1 eksi sonsuza gittikçe ise,

$$E\left[X \middle| X < a_2\right] = \mu - \frac{\phi(a_2)}{\Phi(a_2)} = \mu - \sigma\lambda(-a_2) \tag{2.28}$$

elde edilir. Kırpılmanın olmadığı durumda a_2 sonsuza gittikçe $E[X] = \mu$ elde edilecektir. Kırpılmanın olmadığı duruma kıyasla kırpılmanın sağdan olması ortalamayı azaltır ($E\left[X \middle| X < a_2\right] < E[X]$). Bu eşitsizliğin soldan kırpılma durumu (2.13) de

varyansta dikkate alınarak verilmiştir. Soldan kırpılma durumunda ortalamanın artacağı açıktır[1].

Ayrıca kırpılmış normal dağılım için MÜF yardımı ile varyans da elde edilir.

$$Var\left[X \left| X > a_k \right. \right] = \sigma^2 \left[1 - \delta(\alpha_k) \right] \tag{2.29}$$

Buradan itibaren kırpılma noktası tekrar t olarak alınacak olursa, $\alpha = \dfrac{t-\mu}{\sigma}$ için,

$$\begin{aligned} \lambda(\alpha) &= \phi(\alpha)/[1-\Phi(\alpha)] \qquad x > t \\ \lambda(\alpha) &= -\phi(\alpha)/\Phi(\alpha) \qquad x < t \end{aligned} \tag{2.30}$$

olacaktır. Buradan,

$$\delta(\alpha) = \lambda(\alpha)\left[\lambda(\alpha) - \alpha \right] \tag{2.31}$$

eşitliği yazılır.

Burada önemli bir sonuç her α değeri için $0 < \delta(\alpha) < 1$ olmasıdır. Sonraki başlıklarda kullanılacak olan bir diğer sonuç;

$$d\,\phi(\alpha)/d\alpha = -\alpha\phi(\alpha) \tag{2.32}$$

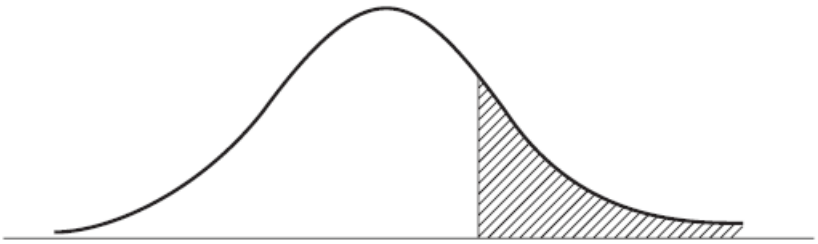
eşitliğidir. Ayrıca açıklanacak olursa $\lambda(\alpha)$ fonksiyonu ters Mills oranı (inverse Mills ratio) olarak adlandırılır ve bunun yanında $\lambda(\alpha) = -\phi(\alpha)/\Phi(\alpha)$ fonksiyonu standart normal dağılım için tehlike fonksiyonu (hazard function) olarak isimlendirilir (Greene, 2003, s.759). Burada Ters Mills Oranı olarak tanıtılan fonksiyon açıkça görülmektedir ki OYF'nin BDF'ye oranıdır $\left(\lambda(.) = \phi(.)/\Phi(.)\right)$.

2.2.3. Sansürlenmiş normal dağılım

Sansürlü değişken için ilgili dağılım teorisi kırpılmışa benzerdir. Burada da, kabul gören çalışmalarda olduğu gibi, normal dağılım ve normallik varsayımı ile işe başlanır. Yalnızca uygun bir standartlaştırma olmamasına rağmen sansür noktası genellikle sıfır varsayılmıştır. Kırpılmış dağılımda, kırpılma noktasının yine sıfır kabul edilmesi halinde

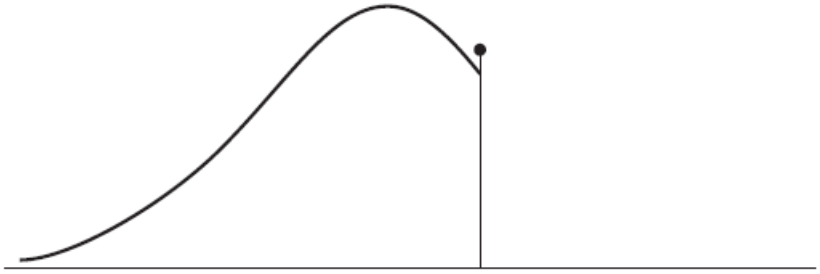
[1] https://disciplinas.stoa.usp.br/pluginfile.php/2028147/mod_resource/content/0/Normal%20truncada.pdf (Erişim Tarihi: 12.09.2016)

bu noktanın üzerindeki kısım hesaplamalar için ilgili alandır. Veri sansürlü ise örneklem verisi için geçerli dağılım kesikli ve sürekli dağılımların karması halini alır. Bu durum Şekil 2.7'de yatay eksen koltuk talebini göstermek üzere örneklendirilmiştir. Yukarıda izah edilen uygulama ise Şekil 2.8'da yatay eksen satışı yapılan bileti göstermek üzere verilmiştir. Bu her iki şekilde yatak eksene dik olarak verilen, ilgili alanın kapasitedir ve sansür noktasıdır.



Şekil 2.7. *Talep*

Kaynak: *Greene, 2011, s. 846*



Şekil 2.8. *Satış*

Kaynak: *Greene, 2011, s. 846*

Dağılımın analizi için yeni bir rassal değişken olarak Y, orijinal ancak gizil değişken olan Y^* den dönüştürülür (Greene, 2011, s.846).

Aslında kırpılmış değişkende olduğu gibi Y değişkeninin gözlenemediği durumlar için Y^* değişkeni kullanılır. Daha önce de tanımlandığı üzere, gözlenen değişken Y, herhangi bir sansür noktası c, $u \sim N\left(0,\sigma^2\right)$ ve dolayısıyla $Y^* \sim N\left(\mu,\sigma^2\right)$ olmak üzere, tanımlanan herhangi bir c katsayısına göre,

$$y = y^* \qquad y^* > c$$
$$y = c \qquad d.d. \tag{2.33}$$

şeklinde tanımlanır.

$c = 0$ olduğu varsayılırsa ve $Y^* \sim N\left(\mu, \sigma^2\right)$ olmak üzere,

$$P\left(Y = 0\right) = P\left(Y^* \le 0\right) = \Phi\left(-\mu / \sigma\right) = 1 - \Phi\left(\mu / \sigma\right) \tag{2.34}$$

şeklinde ifade edilebilir. Ayrıca $Y^* > 0$ için Y, Y^*'nin yoğunluğuna sahiptir (Greene, 2011, s.847).

2.2.3.1. Sansürlenmiş normal dağılımın momentleri

c yine sansürlenme noktası ve $Y^* \sim N\left(\mu, \sigma^2\right)$ için,

$$y = y^* \qquad y^* > c$$
$$y = c \qquad d.d. \tag{2.35}$$

olmak üzere, $P\left(Y^* \le c\right) = \Phi\left(\dfrac{c - \mu}{\sigma}\right) = \Phi\left(\alpha\right) = \Phi$ ifadesi kullanılarak beklenen değer ve varyans sırasıyla şu şekilde yazılabilir.

$$\begin{aligned} E\left[Y\right] &= P(Y = c)E\left[Y \middle| Y = c\right] + P(Y > c)E\left[Y \middle| Y > c\right] \\ &= P(Y^* \le c)c + P(Y^* > c)E\left[Y^* \middle| Y^* > c\right] = \Phi c + (1 - \Phi)(\mu + \sigma\lambda) \end{aligned} \tag{2.36}$$

$$Var\left[Y\right] = \sigma^2(1 - \Phi)\left[(1 - \lambda^2 - \lambda\alpha) + (\alpha - \lambda)^2\Phi\right] \tag{2.37}$$

Burada kırpılmış normal dağılımdan yararlanılmıştır. Kullanılan λ daha önce de ters Mills oranı olarak ifade edilen $\lambda = \dfrac{\phi}{1 - \Phi}$ eşitliğidir (Greene, 2011, s.847).

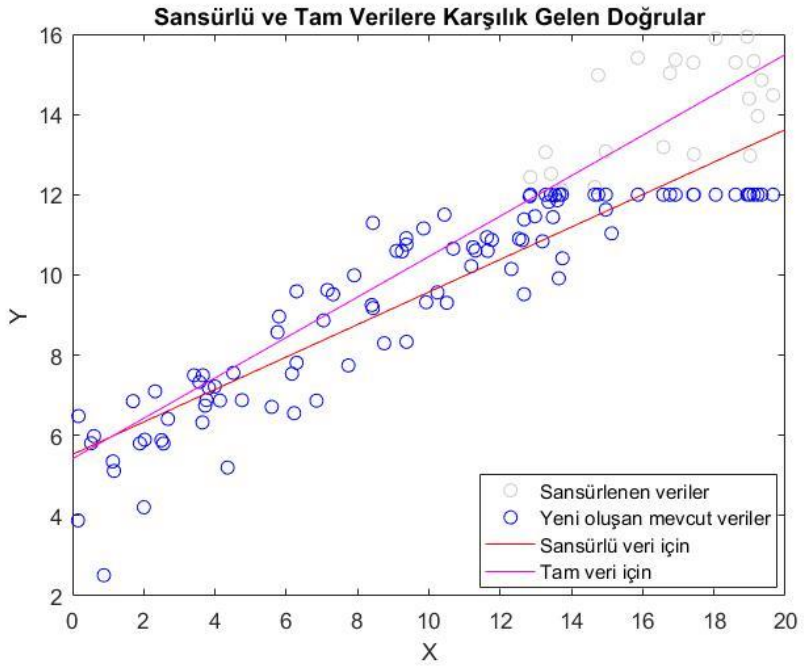
2.3. Sansürlenmiş ve Kırpılmış Regresyon Modelleri

Sürekli bağımlı değişkenler sansürleme ve kırpılma durumunda sadece belli aralıklar arasında değer alabilmektedir. Bu durumda Y^* gizil değişken ve $u_i \sim N\left(0, \sigma^2\right)$ olmak üzere model;

$$y_i^* = \beta_1 + \beta_2 x_i + u_i \tag{2.38}$$

şeklinde yazılır.

$y^* < 0$ olduğu durumlarda basitçe sansürleme veya kırpılmanın oluştuğu düşünüldüğünde, daha büyük bir hata teriminin daha büyük daha büyük bir y^* değerine işaret ettiği açık şekilde ifade edilebilir. O halde hata teriminin daha büyük olması $y^* \geq 0$ olma ihtimalini arttırır. Bu olasılık ayrıca x_i'ye bağlıdır. Böylesi bir durumun oluşması halinde u_i, 0 ortalamaya sahip olmayıp x_i ile korelasyona sahip olacaktır. Önemli varsayımların sağlanmadığı bu durumda klasik EKK yönteminin sapmalı ve tutarsız sonuçlar vereceği açıktır (Davidson ve MacKinnon, 1999 s.473). Klasik EKK yönteminin kullanıldığı örneklemlerde 0 gözlem çok ise parametre tahminleri yanlı; 0 gözlemler yok sayılırsa da parametre tahminleri için etkinlik kaybı söz konusudur. Özetle sansürlü örneklem durumunda, klasik EKK yönteminin sapmalı ve tutarsız oluşu farklı metotlara başvurulması gerektiğini işaret eder.



Şekil 2.9. *Sansürlü ve tam verilere karşılık gelen regresyon doğruları*

Sansürlü örneklem için çizilen doğrunun, sansürün olmadığını varsayan regresyonla elde edilen doğruya göre daha küçük bir eğime sahip olduğu, dolayısı ile tüm

örneklemi modelleyemediği Şekil 2.9'dan açıkça görülebilir. Eldeki veride sansürlü değerlerin olmasına rağmen iyi bir modellemenin istenmesi ise farklı bir modelleme sürecinin gerekliliğini ortaya koyar. Bu aşamada tobit model alternatif olarak önerilmiştir.

2.3.1. Kırpılmış regresyon modeli

$\varepsilon_i \sim N\left(0,\sigma^2\right)$ olmak üzere kırpılmış regresyon modeli $y_i = x_i^{'}\beta + \varepsilon_i$ şeklindedir. $\mu = x_i^{'}\beta$ ve x_i verilmişken model $Y_i \left| X_i \right. \sim N\left(x_i^{'}\beta,\sigma^2\right)$ olacaktır. t kesim noktası için,

$$E\left(Y\left|Y>t\right.\right) = x_i^{'}\beta + \sigma\frac{\phi\left(\frac{t-x_i^{'}\beta}{\sigma}\right)}{1-\Phi\left(\frac{t-x_i^{'}\beta}{\sigma}\right)} \tag{2.39}$$

olarak ifade edilir. (2.39) aslında kırpılmış normal dağılımın momentleri başlığında elde edilen $E\left(X\left|X>t\right.\right) = \mu + \sigma\phi\left(\frac{t-\mu}{\sigma}\right)\Big/1-\Phi\left(\frac{t-\mu}{\sigma}\right)$ eşitliğinde değerlerin yerine yazılmasıdır. Aynı şekilde soldan kırpılma için $E\left(X\left|X<t\right.\right) = \mu - \sigma\phi\left(\frac{t-\mu}{\sigma}\right)\Big/\Phi\left(\frac{t-\mu}{\sigma}\right)$ eşitliği kullanılır.

Ayrıca gizil değişken modelindeki hata terimlerinin dağılımının bilinmesi halinde EÇO yöntemi ile kırpılmış regresyon modeli kolaylıkla tahmin edilebilir. Özel bir durum oluşturacak olan örnek için yaygın kullanılan normal dağılıma sahip hata terimleri için regresyon fonksiyonunun $x_i^{'}\beta$ şeklinde olduğu kabul edilirse, y_i^{*}'nin örneklemde yer aldığı olasılık bulunabilir. Bu durum bir bakıma (2.11)'in özel halidir.

$$\begin{aligned} P\left(y_i^{*}\geq 0\right) &= P\left(x_i^{'}\beta + u_i \geq 0\right) \\ &= 1 - P\left(u_i < -x_i^{'}\beta\right) \\ &= 1 - P\left(\frac{u_i}{\sigma} < \frac{-x_i^{'}\beta}{\sigma}\right) \\ &= 1 - \Phi\left(\frac{-x_i^{'}\beta}{\sigma}\right) \\ &= \Phi\left(\frac{x_i^{'}\beta}{\sigma}\right) \end{aligned} \tag{2.40}$$

$y^* \geq 0$ olup y_i değeri gözlemlenebiliyorsa y_i ile y^*'nin yoğunlukları orantılıdır. Aksi durumda y_i'nin yoğunluğu 0'dır. y_i yoğunluğunun tekliğinin sağlanmasının zorunlu olduğu orantısallık faktörü, $y^* \geq 0$ olasılığının tersidir. Dolayısı ile y_i'nin yoğunluğu şu şekilde yazılabilir;

$$f(y_i) = \frac{\frac{1}{\sigma}\phi\left[\frac{\left(y_i - x_i'\beta\right)}{\sigma}\right]}{1 - \Phi\left[\frac{\left(t - x_i'\beta\right)}{\sigma}\right]} \tag{2.41}$$

Bu da tüm i değerleri için, y_i'nin yoğunluğunun logaritmasının toplamı $\left(\sum_{i=1}^{N} \log L_i\right)$ olan olabilirlik fonksiyonunu işaret eder;

$$\log L = \sum_{i=1}^{n} \ln\left[\sigma^{-1}\phi\left(\frac{y_i - x_i'\beta}{\sigma}\right)\right] - \sum_{i=1}^{n} \ln\left[1 - \Phi\left(\frac{\left(t - x_i'\beta\right)}{\sigma}\right)\right] \tag{2.42}$$

$$\log L(y, \beta, \sigma) = -\frac{n}{2}\log(2\pi) - n\log(\sigma) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(y_i - x_i'\beta\right)^2 - \sum_{i=1}^{n}\log\Phi\left(\frac{x_i'\beta}{\sigma}\right)$$

Burada $-\frac{n}{2}\log(2\pi) - n\log(\sigma) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - x_i'\beta)^2$ ifadesi y^*'nin yoğunluğunu, $\sum_{i=1}^{n}\log\Phi\left(\frac{x_i'\beta}{\sigma}\right)$ ise $y^* > 0$'in yoğunluğunu ifade eder.

İfadenin maksimum yapılarak parametrelerinin hesaplanması zor değildir. Olabilirlik fonksiyonu global olarak konkav (iç bükey) olmasa bile tek bir lokal EÇO tahmini mevcuttur (Orme ve Ruud, 2002).

(2.42)'deki ilk üç terim EKK regresyonuna karşılık gelen olabilirlik fonksiyonu iken son terim örnekleme ait olan regresyon fonksiyonu $x_i'\beta$ ile elde edilen gözlemlerin olasılıklarının logaritmasının toplamının negatif değeridir. Bu terim, olasılığın 1'den küçük olması gerekliliğinden her zaman pozitiftir. Olasılıklar küçültülerek bu terim büyültülebilir. Buradaki dördüncü terimin varlığı EÇO yöntemi ile tahmin edilen β ve

σ tahmin değerlerinin farklılığını sağlarken genellikle EKK tahminleri ile elde edilen emsallerine göre EÇO yöntemi ile elde edilen tahminlerin tutarlılığını sağlar.

Ayrıca bu modelin kırpılmanın diğer formlarına uyarlanması zor değildir. Örneklem soldan, sağdan ya da her iki şekilde kırpılabilir. Ancak kırpılma noktası, sabitlenmiş veya gözlem değerleri içinde değişen olsun fark etmeksizin, bilinmelidir (Davidson ve MacKinnon, 1999 s. 475).

2.3.1.1. Kırpılmış regresyonda parametre tahmini

Klasik istatistiksel tahminleme yöntemleri olan EKK yöntemi ile EÇO tahmin edicileri sırala ile incelendiğinde ve kırpılmış bir örneklemde gizil değişkenin varlığı dikkate alındığında belli problemler ortaya çıkmaktadır. Literatürde bu bağlamda tahmin ediciler geliştirilmiştir. Ancak bu iki tahmin edici kendi içinde kıyaslandığında, hata terimlerinin (u) koşullu beklenen değeri ($E(u|x) = E(u) = 0$) ile aynı hata terimlerinin x ile ilişki durumu ($Cov(x,u) = E(xu) = 0$) varsayımlarının sağlanamayışı, EKK tahmin edicisine göre EÇO tahmin edicisinin üstün geleceği öngörüsü verir (Koç, 2013).

Kırpılmış regresyon modellerinin EKK yöntemi ile tahmininde kırpılmış normal dağılımın momentleri başlığında verilen beklenen değerden $\left(E[y_i|y_i > t] = x_i'\beta + \sigma\lambda(\alpha_i)\right)$ yararlanılır. Bu durumda verilerin kırpıldığı alt popülasyon için yukarıda verilen beklenen değer aşağıdaki gibi yazılabilir.

$$\begin{aligned} E[y_i|x_i] &= x_i'\beta \\ E[y_i|x_i, y_i > t] &= x_i'\beta + \sigma\lambda\left(\frac{t - x_i'\beta}{\sigma}\right) \\ y_i = E[y_i|x_i, y_i > t] + u_i &= x_i'\beta + \sigma\lambda\left(\frac{t - x_i'\beta}{\sigma}\right) + u_i \end{aligned} \tag{2.43}$$

Soldan kırpılmayla oluşan model ve gözlenen verilere karşılık gelen regresyon fonksiyonu yukarıdaki gibidir. Ancak klasik regresyon modelinin $y_i = E[y_i|x_i] + u_i = x_i'\beta + u_i$ şeklinde yanlış tanımlanması $\lambda\left(\frac{t - x_i'\beta}{\sigma}\right)$ düzeltme değişkeninin ihmalinden kaynaklanmaktadır. Burada u_i, y_i değerinden y_i'nin koşullu

beklenen değerinin çıkarılması ile elde edilir. u_i, 0 ortalamaya sahiptir ancak değişen varyans söz konusudur;

$$Var[u_i] = \sigma^2(1-\lambda_i^2+\lambda_i\alpha_i) = \sigma^2(1-\delta_i) \tag{2.44}$$

(2.44)'den anlaşılacağı üzere $Var[u_i]$, x'in bir fonksiyonudur. (2.43) EKK yöntemi ile tahmin edilecek olursa doğrusal olmayan λ_i ihmal edilmelidir. İhmal edilecek bir değişkenden dolayı yanlılık muhtemel bir sonuçtur. Ayrıca X 'in dağılım bilgisine sahip olunmaması, ne derece bir yanlılığın olduğunun belirlenememesine neden olur. (Greene, 2003, s.761).

Sansürlenmiş ve kırpılmış değişkenler başlığında giriş amaçlı tanıtılan kesikli normal dağılımın olabilirlik fonksiyonu EÇO tahmin edicisinin elde edilişinde, $\Phi(.)$ ifadesi BDF'yi göstermek üzere $L = \prod_{i=1}^{n} \frac{f(x)}{1-\Phi(t)}$ yardımı ile bulunur. (2.41) ve (2.42)'de bu ifadenin açılımı yapılmıştır.

2.3.1.2. Kırpılmış regresyon modelinde marjinal etkiler

Modelde deterministik kısım $\mu_i = x_i'\beta$ olmak üzere klasik regresyon modeli;

$$y_i = x_i'\beta + \varepsilon_i \tag{2.45}$$

şeklindedir. Ayrıca $\varepsilon_i | X_i \sim N(0,\sigma^2)$ olduğunda,

$$Y_i | X_i \sim N\left(x_i'\beta, \sigma^2\right) \tag{2.46}$$

şeklinde olacaktır.

Kırpılma noktası olan t değerinden büyük olan Y_i değerlerinin dağılımı ile ilgilenilmektedir. Eşitlik (2.27)'de tanımlanan sonuca bağlı olarak,

$$E[Y_i | Y_i > t] = x_i'\beta + \sigma\frac{\phi[(t-x_i'\beta)/\sigma]}{1-\Phi[(t-x_i'\beta)/\sigma]} \tag{2.47}$$

eşitliği elde edilir. Koşullu ortalama t, σ, x ve β'nın doğrusal olmayan bir fonksiyonudur.

Alt popülasyonda bu modeldeki marjinal etki ise,

$$E[Y_i \,|\, Y_i > t] = x_i'\beta + \sigma\lambda(\alpha_i) \tag{2.48}$$

şeklinde elde edilir. Burada $\alpha_i = t - x_i'\beta/\sigma$, $\lambda_i = \lambda(\alpha_i)$ ve $\delta_i = \delta(\alpha_i)$ olmak üzere,

$$\begin{aligned} \frac{\partial E[Y_i \,|_i\, Y > t]}{\partial x_i} &= \beta + \sigma\left(\frac{d\lambda_i}{d\alpha_i}\right)\frac{\partial \alpha_i}{\partial x_i} \\ &= \beta + \sigma(\lambda_i^2 - \alpha_i\lambda_i)\left(\frac{-\beta}{\sigma}\right) \\ &= \beta(1 - \lambda_i^2 + \alpha_i\lambda_i) \\ &= \beta(1 - \delta_i) \end{aligned} \tag{2.49}$$

eşitliği elde edilmiş olur.

Kırpılmış varyanstan gelen ölçek faktörüne $(1-\delta_i)$ dikkat etmek gerekir. $(1-\delta_i)$, her x_i değeri için 0 ile 1 arasındadır ve marjinal etki ilgili katsayıdan daha küçüktür. Benzer bir küçülme varyans içinde söz konusudur. Alt popülasyonda $y_i > t$, varyans σ^2 değerine eşit olmayıp,

$$\text{Var}[Y_i \,|\, Y_i > t] = \sigma^2(1 - \delta_i) \tag{2.50}$$

şeklinde ifade edilir.

(2.50)'teki marjinal etki ya da β katsayısının kendisi olsun fark etmeksizin, çalışmanın amacına bağlı olarak katsayılarla ilgilenilir. Bu durumda ilgilenilen alt popülasyon ise marjinal etki; ilgilenilen tüm popülasyonsa β katsayısı dikkate alınır (Greene, 2011, s.838).

2.3.2. Sansürlü regresyon modeli

Sansürlenmiş veriler için en popüler model Tobin tarafından 1958'de önerilen tobit modeldir. Bu model tobit Tip I modeli olarak bilinmektedir. Daha önceki başlıklarda da ifade edildiği gibi en basit hali ile $u_i \sim N\left(0, \sigma^2\right)$ olmak üzere $y_i^* = x_i'\beta + u_i$ regresyon modeli;

$$\begin{aligned} y_i &= y_i^* \qquad y_i^* > 0 \\ y_i &= 0 \qquad d.d. \end{aligned} \tag{2.51}$$

şeklinde ifade edilir. Burada y_i^* gizil değişkendir ve pozitif olduğu zaman gözlemlenir; negatif olduğu zaman ise gözlem sansürlüdür ve bu durumda $y_i = 0$ olacaktır. Tobit model soldan, sağdan ya da hem soldan hem sağdan sansürlenme durumlarına göre farklı şekillerde ifade edilebilir.

Farklılık gösterse de üretilmesi çok güç olmayan tobit model için olabilirlik fonksiyonu elde edilirken öncelikle $P(y_i = 0)$ için;

$$\begin{aligned} P\left(y_i = 0\right) &= P\left(y_i^* \le 0\right) \\ &= P\left(x_i^{'}\beta + u_i \le 0\right) \\ &= P\left(u_i \le -x_i^{'}\beta\right) \\ &= P\left(\frac{u_i}{\sigma} \le \frac{-x_i^{'}\beta}{\sigma}\right) \\ &= \Phi\left(\frac{-x_i^{'}\beta}{\sigma}\right) \end{aligned} \tag{2.52}$$

şeklinde yazılır. $y_i = 0$ olması pozitif olasılık olduğundan $y_i = 0$ gözlemleri ile oluşturulan gözlemlerin olabilirlik fonksiyonuna katkısı, yoğunluğun logaritması olmayıp pozitif olasılığın logaritmasıdır.

$$\log\left[\Phi\left(\frac{-x_i^{'}\beta}{\sigma}\right)\right] \tag{2.53}$$

y_i pozitif ise y_i'nin yoğunluğu vardır ve olabilirlik fonksiyonuna katkısı;

$$\log\left[\frac{1}{\sigma}\phi\left(\frac{-x_i^{'}\beta_i}{\sigma}\right)\right] \tag{2.54}$$

olur. Olabilirlik fonksiyonuna bu katkı sansürlemenin olmadığı klasik normal doğrusal regresyon içindir.

Sansürlü gözlemler için olan olabilirlik fonksiyonu (2.53) ile sansürlemenin olmadığı gözlemler için olan olabilirlik fonksiyonu (2.54) birleştirilerek tobit model için yazılacak olan olabilirlik fonksiyonu;

$$\sum_{y_i=0} \log\left[\Phi\left(\frac{-x_i'\beta}{\sigma}\right)\right] + \sum_{y_i>0} \log\left[\frac{1}{\sigma}\phi\left(\frac{y_i - x_i'\beta}{\sigma}\right)\right] \tag{2.55}$$

şeklinde ifade edilir.

Bu olabilirlik fonksiyonunda ilk terim sansürlenmiş gözlemler için olasılıkların logaritmasının toplamı iken ikinci terim sansürlenmemiş veriler için yoğunlukların logaritmasının toplamıdır. Bu durum tobit modeldeki bağımlı değişkenlerin, sürekli rassal değişkenler ile kesikli rassal değişkenlerin karma dağılımına sahip olduğunu ifade eder. Ancak tobit model için EÇO tahmin edicisinin tutarlılık ve asimptotik normallik özelliklerinde bir problem söz konusu değildir. Bu durum Amemiya (1973) tarafından gösterilmiştir. (Davidson ve MacKinnon, 1999, s. 476).

2.3.2.1. Sağdan ve soldan sansürlü regresyonda marjinal etkiler

Burada iki katlı sansürleme söz konusudur. Daha önce c olarak verilen sansürleme noktası, bu başlık için iki sansürleme noktasının varlığı dikkate alınarak c_1 ve c_2 olarak alınmıştır. $y^* = x_i'\beta + \varepsilon$ ve c_1 ile c_2 $(c_1 < c_2)$ sabitler olmak üzere gözlenen değişken;

$$\begin{aligned} y &= c_1 \qquad y^* \le c_1 \\ y &= c_2 \qquad y^* \ge c_2 \\ y &= y^* \qquad d.d. \end{aligned} \tag{2.56}$$

şeklinde tanımlansın. $\varepsilon \sim N(0,\sigma^2)$ olmak üzere $f(\varepsilon)$ ile $F(\varepsilon)$ sırası ile ε'nın OYF ve BDF olsun. $f(\varepsilon|x) = f(\varepsilon)$ olmak üzere

$$\frac{\partial E[Y|X]}{\partial x} = \beta.P[c_1 < y^* < c_2] \tag{2.57}$$

şeklinde ifade edilir. İspatı ise tanımdan,

$$E\left[Y|X_i\right] = c_1 P[Y^* \le c_1|X_i] + c_2 P[Y^* \ge c_2|X_i] + P[c_1 < Y^* < c_2|X_i]E[Y^*|c_1 < Y^* < c_2|X_i]$$

şeklindedir.

Burada $j = 1,2$ için c_j sansür noktası ve $\alpha_j = \dfrac{(c_j - x_i^{'}\beta)}{\sigma}$, $F_j = F(\alpha_j)$, $f_j = f(\alpha_j)$ olacak şekilde,

$$E\left[Y\middle|X_i\right] = c_1 F_{c_1} + c_2(1-F_{c_1}) + (F_{c_2} - F_{c_1})E[Y^*\middle|c_1 < Y^* < c_2, \mathrm{X}_i] \tag{2.58}$$

yazılır. $y^* = x_i^*\beta + \sigma\left[\dfrac{y^* - x_i^*\beta}{\sigma}\right]$ olduğundan koşullu ortalama şu şekilde yazılabilir.

$$E\left[Y^*\middle|c_1 < Y^* < c_2, X_i\right] = x_i^{'}\beta \tag{2.59}$$

$$+\sigma E\left[\frac{Y^* - X_i^{*'}\beta}{\sigma}\middle|\frac{c_1 - X_i^{*'}\beta}{\sigma} < \frac{Y^* - X_i^{*'}\beta}{\sigma} < \frac{c_2 - X_i^{*'}\beta}{\sigma}\right]$$

$$= x_i^{'}\beta + \sigma\int_{\alpha_{c_1}}^{\alpha_{c_2}} \frac{(\varepsilon/\sigma)\, f\,(\varepsilon/\sigma)}{F_{c_2} - F_{c_1}} d(\frac{\varepsilon}{\sigma})$$

buradan hareketle terimler toplanarak,

$$E\left[Y\middle|X_i\right] = c_1 F_{c_1} + c_2(1-F_{c_1}) + (F_{c_2} - F_{c_1})x_i^{'}\beta + \sigma\int_{\alpha_{c_1}}^{\alpha_{c_2}} (\frac{\varepsilon}{\sigma})\, f\,(\frac{\varepsilon}{\sigma}) d(\frac{\varepsilon}{\sigma}) \tag{2.60}$$

elde edilir. x e göre türev alınacaktır. Bu aşamada tek zorluk, integralinin sınır değerleri açısından türev alımında sıkıntı çıkarabileceği son terimdir. $f(\varepsilon)$'nun x içermediği varsayılarak Leibnitz teoremi kullanılırsa,

$$\frac{\partial E\left[Y\middle|X_i\right]}{\partial x} = \left(\frac{-\beta}{\sigma}\right)\mathrm{c}_1\, f_{c_1} + \left(\frac{-\beta}{\sigma}\right)\mathrm{c}_2(1-F_{c_1}) + (F_{c_2} - F_{c_1})\beta \tag{2.61}$$

$$+x_i^{'}\beta(f_{c_2} - f_{c_1})\left(\frac{-\beta}{\sigma}\right) + \sigma\left[\alpha_{c_1} f_{c_2} - \alpha_{c_1} f_{c_1}\right]\left(\frac{-\beta}{\sigma}\right)$$

α_{c_1} ve α_{c_2}'nın tanımları eklenerek ve istenen sonuç için sıfır hariç tüm terimler toplanarak,

$$\frac{\partial E\left[Y\middle|X_i\right]}{\partial x} = (F_{c_2} - F_{c_1})\beta + \beta P[\mathrm{c}_1 < Y_i^* < c_2] \tag{2.62}$$

elde edilir.

Bu genel sonuç dağılımın her iki kuyruğunun da sansürlendiği durumu içerir. Burada ε'nun normal dağılmama varsayımına dikkat edilmelidir. Soldan sıfırda sansürlemenin olduğu ve bozucu terimin normal dağıldığı standart durum için sonuç şu şekilde özelleştirilir,

$$\frac{\partial E\left[Y\middle|X_i\right]}{\partial x} = \beta \;\; \Phi\left(\frac{x_i'\beta}{\sigma}\right) \tag{2.63}$$

bu genel bir sonuç olmamakla beraber, EÇO tahminleri ile elde edilen tobit model katsayılarının EKK tahmin değerleri ile benzerliklerini açıklamada bir neden olarak önerilebilir.

McDonald ve Moffitt $\frac{\partial E\left[Y\middle|X_i\right]}{\partial x}$ için $\alpha_i = x_i'\beta$, $\Phi_i = \Phi_i(\alpha_i)$, $\lambda_i = \phi_i/\Phi_i$ olmak üzere kullanışlı bir ayrışma önermişlerdir (McDonald ve Moffitt, 1980).

$$\frac{\partial E\left[Y\middle|X_i\right]}{\partial x} = \beta.\left\{\Phi_i[1-\lambda_i(\alpha_i+\lambda_i)] + \phi_i(\alpha_i+\lambda_i)\right\} \tag{2.64}$$

Ayrı şekilde iki parça alınınca, bu sonuç eğim vektörünü ayrıştırır,

$$\frac{\partial E\left[Y\middle|X_i\right]}{\partial x} = P[Y_i>0]\frac{\partial E[Y_i\middle|X_i, Y_i>0]}{\partial x_i} + E[Y_i\middle|X_i, Y_i>0]\frac{\partial P[Y_i>0]}{\partial x_i} \tag{2.65}$$

böylece x' deki bir değişim iki etkiye sahiptir denir.

- Dağılımın pozitif kısmındaki y_i^*' nin koşullu ortalamasını etkiler.
- Gözlemin, dağılımın bu kısmına düşme olasılığını etkiler (Greene, 2011 s.849).

3. SANSÜRLÜ REGRESYON MODELİ OLARAK TOBİT MODEL VE TOBİT MODEL TİPLERİ

Regresyona dair sansürlü dağılım varsayımı varsa bu model sansürlü regresyon modeli ya da tobit model olarak anılmaktadır (Kmenta, 1990). Genelleştirilmiş sansürlü modelin sıfır noktasında sansürlenmesiyle "Standart Tobit Model" elde edilir (Carson, 2007). Aslında sansürlü regresyon modeli başlığı altında kısmen standart tobit model ifade edilmiştir. O halde $y_i^* = x_i'\beta + u_i$ regresyon modeli $u_i \sim N\left(0, \sigma^2\right)$ varsayımı altında ve c sansürleme noktası olmak üzere tobit model;

$$\begin{aligned} y_i &= y_i^* \qquad y_i^* > c \\ y_i &= c \qquad d.d. \end{aligned}$$

şeklinde ifade edilir. Burada c değerinden büyük değerler gerçek gözlenen değerlerken diğer durumlarda veri sansürlenmiş kabul edilmekte ve bu değerler için c değeri tanımlanmaktadır. Tobit modelin varyasyonları sansürlenmenin oluşturulduğu yerin ve zamanın değiştirilmesi ile elde edilir (Amemiya, 1985). Beş farklı tobit modeli tipi mevcuttur.

1. Tip I

y_i^* gizil değişkeni daima gözlemlenemezken bağımsız değişken x_i gözlemlenebilmektedir. Tobit modelin yaygın varyasyonları, 0'dan farklı olan c ile sansürlenmesidir. Ancak bu başlık altında sağdan, soldan ve hem sağdan hem soldan sansürlemenin yapılacağı dikkate alınarak soldan sansür noktası c_y, sağdan sansür noktası c_w olarak kabul edilmiştir.

c_y sansür noktası olmak üzere soldan sansürlenme;

$$\begin{aligned} y_i &= y_i^* \qquad y_i^* > c_y \\ y_i &= c_y \qquad d.d. \end{aligned} \tag{3.1}$$

Bir diğer varyasyon c_w gibi bir sansür noktası için sağdan sansürleme,

$$\begin{aligned} y_i &= y_i^* \qquad y_i^* < c_w \\ y_i &= c_w \qquad d.d. \end{aligned} \tag{3.2}$$

hem soldan hem de sağdan sansürleme ise,

$$
\begin{aligned}
y_i &= y_i^* \qquad & c_y < y_i^* < c_w \\
y_i &= c_y \qquad & y_i^* \leq c_y \\
y_i &= c_w \qquad & y_i^* \geq c_w
\end{aligned}
\tag{3.3}
$$

şeklinde olur.

Diğer modeller soldan 0 ile sınırlandırılarak sunulmuştur. Ancak tip I için yukarıda yapılan genelleştirmeler bu tipler için de yapılabilir. Bu çalışmada kullanılan ve tobit olarak ifade edilen model standart tobit ya da tobit tip 1'dir. Literatürde de yaygın olarak kullanılan bu tipin yanında 2AHeckit olarak ifade edilen tobit Tip 2 kullanılmıştır. Diğer tobit tipleri bilgi amaçlı tanıtılmıştır.

2. Tip II

Tobit modelin bir tür genelleştirilmiş hali olan model genelleştirilmiş tobit olarak anılmaktadır. İkinci bir gizil verinin kullanıldığı Tip II model tip I modelden farklı olarak bağımlı değişkenin hem sürece katılım kararını hem de ne ölçüde katıldığını hesaba katar. Heckman tarafından genelleştirilen bu tip, tip II model isminin yanında heckit model olarak da anılmaktadır. Heckman'ın iki aşamalı modeli olarak genelleştirilen model,

$$
\begin{aligned}
y_{2i} &= y_{2i}^* \qquad & y_{1i}^* > 0 \\
y_{2i} &= 0 \qquad & d.d.
\end{aligned}
\tag{3.4}
$$

şeklinde ifade edilir. Bu modele dayanan tahminleme tobit model alternatifleri başlığında ayrıntılandırılmıştır. İki aşamalı bir süreci işleyen bu model özetle ilk aşamada probit modeli, ikinci aşamada EKK sürecini işler.

3. Tip III

Tip III model ikinci bir gözlenmiş bağımlı değişken ile ikinci bir gölge değişkeni ortaya koymaktadır.

$$
\begin{aligned}
y_{1i} &= y_{1i}^* \qquad & y_{1i}^* > 0 \\
y_{1i} &= 0 \qquad & d.d.
\end{aligned}
\tag{3.5}
$$

$$
\begin{aligned}
y_{2i} &= y_{2i}^* \qquad & y_{1i}^* > 0 \\
y_{2i} &= 0 \qquad & d.d.
\end{aligned}
$$

4. Tip IV

Tip IV model üçüncü bir gözlenmiş bağımlı değişken ile üçüncü bir gölge değişkeni ortaya koymaktadır.

$$\begin{aligned} y_{1i} &= y_{1i}^{*} \qquad y_{1i}^{*} > 0 \\ y_{1i} &= 0 \qquad d.d. \end{aligned} \tag{3.6}$$

$$\begin{aligned} y_{2i} &= y_{2i}^{*} \qquad y_{1i}^{*} > 0 \\ y_{2i} &= 0 \qquad d.d. \end{aligned}$$

$$\begin{aligned} y_{3i} &= y_{3i}^{*} \qquad y_{1i}^{*} > 0 \\ y_{3i} &= 0 \qquad d.d. \end{aligned}$$

5. Tip V

Tip II'ye benzeyen bu tipte farklı olarak y_{1i}^{*}'nin yalnızca işareti gözlemlenmektedir.

$$\begin{aligned} y_{2i} &= y_{2i}^{*} \qquad y_{1i}^{*} > 0 \\ y_{2i} &= 0 \qquad d.d. \end{aligned} \tag{3.7}$$

$$\begin{aligned} y_{3i} &= y_{3i}^{*} \qquad y_{1i}^{*} > 0 \\ y_{3i} &= 0 \qquad d.d. \end{aligned}$$

3.1. Sansürlü Regresyon (Tobit) Modeli

Yalnızca pozitif değerler alan ve böylelikle sınırlandırılmış olan bağımlı değişkenli modeller standart tobit model olarak adlandırılır. Sınırlandırılma aşamasında kukla değişkene ihtiyaç duyulur. Daha önceki bölümlerde ifade edilen probit ve logit model için kukla değişken şu şekilde kullanılır:

$$\begin{aligned} y_i &= 1 \qquad y^{*} > 0 \\ y_i &= 0 \qquad y^{*} \le 0 \end{aligned} \tag{3.8}$$

$y_i^{*} = x_i^{'}\beta + u_i$ şeklinde oluşturulan tobit modelde ise $u_i \sim N(0, \sigma^2)$ olmak üzere,

$$\begin{aligned} y_i &= y_i^{*} \qquad x_i^{'}\beta + u_i > 0 \\ y_i &= 0 \qquad x_i^{'}\beta + u_i \le 0 \end{aligned} \tag{3.9}$$

şeklinde kukla değişkenler ile tobit model ifade edilir (Gujarati, 1999, s. 570).

Yukarıdaki model standart tobit model olarak adlandırılır. Daha önceki bölüm ve başlıklarda da belirtildiği üzere c sansürlenme sınır değeri olmak üzere $c = 0$ değeri için elde edilen tobit modelinin genelleştirilmiş hali,

$$\begin{aligned} y_i &= y_i^* \qquad c > 0 \\ y_i &= c \qquad c \le 0 \end{aligned} \tag{3.10}$$

şeklinde ifade edilir.

Tobit modelde negatif ve sıfır değerlerin tamamının ihmali halinde hata teriminin ortalaması sıfır olmaz. Ayrıca böylesi bir durumda hata teriminin yoğunluğu simetrikte olmayacaktır. Sınırlandırılmalarla beraber oluşan tüm verilerin, ki buna belirli aralıktaki değerlerin tamamının bir değere dönüştürülmesi de dahil, dağılımları sürekli ve süreksiz dağılımların karması olur. Hata terimleri normal dağılıyorsa EÇO ve diğer olabilirlik temelli süreçler tutarlı ve asimptotik normal dağılımlı tahmin edicileri verir. Ancak olabilirlik fonksiyonunun parametrik biçimi yanlış ise tahmin ediciler tutarsızdır (Baltagi, 2001, 212; Breen, 1996, 12–13'den aktaran Cafri, 2009, s.48).

3.2. Tobit Modelin Beklenen Değeri

$c = 0$ noktasında sansürlemenin olduğu standart tobit model için beklenen değer hesabı araştırmanın amacına ve veri setinin hangi kısmı ile ilgilenildiğine bağlı olarak üç farklı şekilde hesaplanır (Sigelman ve Zeng, 1999).Sadece gizil değişkenin temel alındığı araştırmalarda $E[Y]$' nin $E[Y_i^*]$' den daha kullanışlı olduğu ifade edilmiştir (Greene, 2003, s.764). Ancak bağımsız değişkenlerin bağımlı değişkenler üzerinde etkisinin incelendiği her durumda $E[Y]$' nin kullanılabileceği ve sansürsüz gözlemlerle ilgileniliyorsa da $E(Y|Y > c)$ beklenen değerinin kullanılabileceği ifade edilmiştir (Wooldridge, 2002, s.520). Bu noktada tam bir birlik yoktur.

Yukarıda bahsi edilen 3 farklı beklenen değer şu şekildedir:

1. Latent (gizli) değişken Y_i^*' nin beklenen değeri,

$$E[Y_i^*] = x_i'\beta \tag{3.11}$$

şeklinde ifade edilir. Ancak Y_i^* değişkenin kısmen gözlenebilir olmasından dolayı pratikte kullanışlı değildir.

2. $Y|Y>0$ için beklenen değer,

Öncelikle genel bir beklenen değer bilgisi için,

$E(Y|Y>c)=\dfrac{\phi(y)}{1-\Phi(y)}$ ve $y_i=x_i'\beta+u_i$ olmak üzere,

$$E\left(\frac{u_i}{\sigma}\middle|\frac{u_i}{\sigma}>\frac{c-x_i'\beta}{\sigma}\right)=\left[\frac{\phi\left(\dfrac{c-x_i'\beta}{\sigma}\right)}{1-\Phi\left(\dfrac{c-x_i'\beta}{\sigma}\right)}\right] \tag{3.12}$$

olarak elde edilir. Buradan,

$$E(Y|Y>c)=x_i'\beta+\sigma\left[\frac{\phi\left(\dfrac{c-x_i'\beta}{\sigma}\right)}{1-\Phi\left(\dfrac{c-x_i'\beta}{\sigma}\right)}\right] \tag{3.13}$$

bulunur. Sansür noktası $c=0$ için;

$$E(Y|Y>c) = x_i'\beta+\sigma\left[\frac{\phi\left[\dfrac{-x_i'\beta}{\sigma}\right]}{1-\Phi\left[\dfrac{-x_i'\beta}{\sigma}\right]}\right] \tag{3.14}$$

$$= x_i'\beta+\sigma\left[\frac{\phi\left[\dfrac{x_i'\beta}{\sigma}\right]}{\Phi\left[\dfrac{x_i'\beta}{\sigma}\right]}\right]$$

elde edilir. Bu aşamada $\lambda(\alpha)=\dfrac{\phi\left(\dfrac{x_i'\beta}{\sigma}\right)}{\Phi\left(\dfrac{x_i'\beta}{\sigma}\right)}$ ters Mills oranı olmak üzere,

$$E(Y|Y>c) = x_i'\beta+\sigma\lambda(\alpha) \tag{3.15}$$

şeklinde elde edilmiş olur.

3. Y bağımlı değişkeninin beklenen değeri,

Sansürlemenin sıfır noktasında olduğu sansürlü normal dağılımın beklenen değeri

$$E\left[Y\middle|X_i\right] = P(sansürsüz\middle|X_i).E(Y\middle|Y>c, X_i) + P(sansürlü\middle|X_i).c \tag{3.16}$$

olur ve eğer $c = 0$ ise,

$E\left[Y\right] = \Phi\left(\frac{\mu}{\sigma}\right)\left[\mu + \sigma\lambda\right]$ ve $\lambda = \frac{\phi\left(\frac{\mu}{\sigma}\right)}{\Phi\left(\frac{\mu}{\sigma}\right)}$, $\mu = x_i^{'}\beta$ olmak üzere,

$$E\left[Y\right] = \underbrace{\Phi\left(\frac{x_i^{'}\beta}{\sigma}\right)}_{P(y>0)}.\underbrace{\left[x_i^{'}\beta + \sigma\lambda\left(\alpha\right)\right]}_{E(y|y>0)} \tag{3.17}$$

şeklinde ifade edilir (Eren, 2012, s.30).

3.3. Tobit Modelin Marjinal Etkileri

Özetle denilebilir ki üç tane beklenen değer olduğu gibi üç tane de marjinal etki vardır (Eren, 2012 s.31).

1. Latent (gizli) bağımlı değişken olan Y^* için beklenen değer $E\left[Y^*\right] = x_i^{'}\beta$ olmak üzere marjinal etki,

$$\frac{\partial E\left[Y^*\middle|x_i\right]}{\partial x_i} = \beta_i \tag{3.18}$$

eşitliği ile hesaplanır (Eren, 2012, s.31). Bu durumda x_k bağımsız değişkeninin bir birimlik değişimi, Y^* latent bağımlı değişkeni üzerinde β_i kadar değişime tekabül eder.

2. $Y\middle|Y>0$ rassal değişkeni için beklenen değer $E\left[Y\middle|Y>0\right] = x_i^{'}\beta + \sigma\lambda(\alpha)$ olmak üzere marjinal etki:

$$\frac{\partial E\left[Y|Y>0\right]}{\partial x_k} = \beta_k \left\{1-\lambda(\alpha)\left[\frac{x_i'\beta}{\sigma}+\lambda(\alpha)\right]\right\} \tag{3.19}$$

şeklinde bulunur.

3. y 'nin beklenen değeri $E[Y] = \Phi\left(\frac{x_i'\beta}{\sigma}\right)[x_i'\beta + \sigma\lambda(\alpha)]$ olmak üzere marjinal etkisi:

$$\frac{\partial E\left[Y|x_i\right]}{\partial x_i} = \beta\Phi\left(\frac{x_i'\beta}{\sigma}\right) \tag{3.20}$$

şeklindedir. Daha önce (2.64)'de verildiği gibi burada McDonald ve Moffit $\frac{\partial E\left[Y|x_i\right]}{\partial x_i}$ tarafından önerilen ayrışımdan yararlanılır. Yukarıdaki eşitlik bu ayrışma neticesinde tekrar yazılır:

$$\frac{\partial E\left[Y|x_i\right]}{\partial x_i} = P(Y>0)\frac{\partial E\left[Y|x_i|Y>0\right]}{\partial x_i} + E\left[Y|x_i|Y>0\right]\frac{\partial P(Y>0)}{\partial x_i} \tag{3.21}$$

Daha önce sağdan ve soldan sansürlü regresyonda marjinal etkiler başlığı sonunda da verildiği gibi bu ifade x_i 'deki değişimin iki etkisini gösterir. Y^* 'nin koşullu beklenen değerinin, dağılımın pozitif kısmında olmasını etkilediği gibi gözlemin dağılım kısmına düşme olasılığını etkiler (Greene, 2011, s.850).

Burada hangi marjinal etkinin kullanılacağı, tahmin amacına göre değişebilir (Gezer, 2015, s.38).

Tobit modelde β katsayısının yorumu iki farklı şekilde yapılır.

- Bağımsız değişken sürekli ise, diğer tüm değişkenler sabitken x 'deki bir birimlik artış, bağımlı değişken y 'de β kadarlık bir değişim oluştururken,
- Bağımsız değişken kukla değişkenli ise, tüm değişkenler sabitken x değişkenine sahip olma olasılığı, y ' de β kadarlık bir değişim oluşturur (Koç, 2013, s.22).

3.4. Tobit Modelde Parametre Tahmini

Geçmiş yıllarda zor olsa da şuan gelişmiş bilgisayar paketleri ile tobit model üstesinden gelinebilir bir model halini almıştır. Şuan tahminlenmesi doğrusal bir regresyon modeli düzeyindedir.

d_i kukla değişken olmak üzere, Y_i sürekli değişkeninin OYF,

$$f(y_i) = \left[f(y_i^*) \right]^{d_i} \left[F(c) \right]^{1-d_i} \tag{3.22}$$

şeklindedir. Burada $y_i^* > c$ olduğunda kukla değişken 1 olup gözlem sansürsüzdür. Diğer durumlarda ise kukla değişken 0'a eşit olup gözlem sansürlüdür.

$d = 0$ durumunda y 'nin yoğunluğu $y^* \le c$ 'nin gözlenme olasılığına eşittir. $y^* \le c$ ve $y^* > c$ iken olasılıklarsa sırasıyla,

$$P\left(sansürlü\right) = P\left(y_i = c\right) = P\left(y^* \le c\right) \tag{3.23}$$

$$= P\left(x_i^{'}\beta + u_i \le c\right)$$

$$= P\left(u_i \le c - x_i^{'}\beta\right) = P\left(\frac{u_i}{\sigma} \le \frac{c - x_i^{'}\beta}{\sigma}\right)$$

$$= \Phi\left(\frac{c-\mu}{\sigma}\right) = \Phi(\alpha) = \Phi$$

$$P\left(sansürsüz\right) = \mathrm{P}\left(y^* > c\right) = 1 - \Phi\left(\frac{c-\mu}{\sigma}\right) = \Phi\left(\frac{\mu - c}{\sigma}\right) \tag{3.24}$$

şeklindedir. Bu eşitlikler dikkate alınarak olabilirlik fonksiyonu;

$$L = \prod_{i=1}^{N} \left[\frac{1}{\sigma} \phi\left(\frac{c-\mu}{\sigma}\right) \right]^{d_i} \left[1 - \Phi\left(\frac{\mu - c}{\sigma}\right) \right]^{1-d_i} \tag{3.25}$$

$$L = \prod_{y_{i=0}} \left[1 - \Phi\left(\frac{c - x_i^{'}\beta}{\sigma}\right) \right] \prod_{y_{i>0}} \frac{1}{\sigma} \phi\left(\frac{c - x_i^{'}\beta}{\sigma}\right)$$

$$\ln L = \sum_{y_i=0} \ln \left[1 - \Phi\left(\frac{c - x_i^{'}\beta}{\sigma}\right) \right] + \sum_{y_i>0} \ln \frac{1}{\sigma} \phi\left(\frac{y_i - x_i^{'}\beta}{\sigma}\right)$$

şeklinde elde edilir (Chay ve Powell, 2011; Park, 2003). $c = 0$ yani sıfırda sansürlenme varsayılırsa,

$$\sum_{y_i>0} \log\left(\frac{1}{\sigma}\Phi\left(\frac{y_i - x_i'\beta}{\sigma}\right)\right) + \sum_{y_i=0} \log\left(\Phi\left(\frac{-x_i'\beta}{\sigma}\right)\right) \tag{3.26}$$

şeklinde olacaktır. Bu aşamada kırpılmış regresyon ve probit model arasında ilginç bir ilişki vardır. Bu olabilirlik fonksiyonuna $\sum_{y_i>0} \log\left(\Phi\left(\frac{x_i'\beta}{\sigma}\right)\right)$ terimi eklenip çıkarılırsa ilk kısımda kırpılmış regresyon modeli, ikinci kısımda ise $\frac{x_i'\beta}{\sigma}$ indeks fonksiyonlu probit modelin olabilirlik fonksiyonu elde edilmiş olur. Bu kısımda β ve σ ayrı ayrı tanımlanamasa da ilk kısımın varlığından dolayı herhangi bir problem oluşmaz.

$$\sum_{y_i>0} \log\left(\frac{1}{\sigma}\Phi\left(\frac{y_i - x_i'\beta}{\sigma}\right)\right) - \sum_{y_i>0} \log\left(\Phi\left(\frac{x_i'\beta}{\sigma}\right)\right) \tag{3.27}$$

$$+\sum_{y_i=0} \log\left(\Phi\left(\frac{-x_i'\beta}{\sigma}\right)\right) + \sum_{y_i>0} \log\left(\Phi\left(\frac{x_i'\beta}{\sigma}\right)\right)$$

Açıkça ifade edilebilir ki tobit modelin olabilirlik fonksiyonu, kırpılmış regresyon modeli ile probit modelin olabilirlik fonksiyonlarının birleşimidir. Bu modellerdeki katsayı vektörleri orantılıdır. Kısıtlama k serbestlik dereceli olabilirlik oranı (OO) ile test edilir. Bu test neticesinde yokluk hipotezi reddedilirse tobit model kullanılmaz (Davidson ve MacKinnon, 1999, 477).

(3.25) farklı bir formda ifade edilecek olursa,

$$\ln L = \sum\nolimits_{y_i=0} \ln\left[1 - \Phi\left(\frac{x_i'\beta}{\sigma}\right)\right] + \sum\nolimits_{y_i>0} -\frac{1}{2}\left[\log(2\pi) + \ln\sigma^2 + \left(\frac{y_i - x_i'\beta}{\sigma}\right)^2\right] \tag{3.28}$$

şeklinde de yazılabilir. Buradaki iki kısım sırayla, sansürlü gözlemler için ilgili olasılıklar ve sansürsüz gözlemler için klasik regresyona karşılık gelir. Bu olabilirlik kesikli ve sürekli dağılımın karışımı olmasından dolayı standart bir tipe sahip değildir. Karmaşıklığına rağmen $\ln L$ 'nin maksimizasyonu için işletilen sürecin, tahmin edicide istenen ilgili özellikleri sağladığı Amemiya (1973) tarafından gösterilmiştir (Greene,

2011, s.850). Hessian'ın sürekli eksi tanımlanması ve buna bağlı olarak Newton metodu ile kolaylıkla yakınsaması sonrası $\gamma = \frac{\beta}{\sigma}$ ve $\theta = \frac{1}{\sigma}$ dönüşümleri ile;

$$\ln L = \sum_{y_i=0} \ln\left[1 - \Phi\left(\gamma' x_i\right)\right] + \sum_{y_i>0} -\frac{1}{2}\left[\ln(2\pi) - \ln\theta^2 + \left(\theta y_i - \gamma' x_i\right)^2\right] \tag{3.29}$$

elde edilir. Burada yeniden parametrelendirilerek yapılan düzenleme işlem kolaylığı sağlayacaktır (Olsen, 1978). Bu düzenlemeler neticesinde sansürlü regresyonun olabilirliği, kesikli regresyona oldukça benzer bir hal almış olur. Ancak yakınsama sonunda orijinal parametreler $\gamma = \frac{\beta}{\sigma}$ ve $\theta = \frac{1}{\sigma}$ kullanılarak elde edilir. Ayrıca bu tahminler için asimptotik kovaryans matrisi delta metodu kullanılarak $[\gamma, \theta]$ nin tahminleri için elde edilebilir (Greene, 2011, s.851).

Araştırmacılar tutarsız sonuçlarına rağmen genellikle EKK metodunu kullanmıştır. Neredeyse istisnasız olarak EKK tahminlerinin mutlak değer olarak EÇO tahmin edicisinden daha küçük olduğu bulunmuştur (Greene, 2011, s.850). Bu aşamada araştırmaların sağlıklı işleyişi sorgulanabilir. Belli koşullar ve varsayımlar altında kullanılan metotlar herhangi bir ihlal durumunda yanıltıcı sonuçlar doğurabilir.

Bu aşamada sansürlü veri için EKK metodu ile EÇO yöntemi (normal dağılım varsayımı altında EÇO tahmin edicisi) çalışma kapsamında incelenmiştir. Ayrıca EÇO tahmin edicisi için alternatif olarak, 2AEKK, UEÇO ve 2AHeckit ele alınmıştır. Bunun yanında daha sağlıklı çıkarsamalar için çalışmadaki referans model olan tobit model farklı bir algoritma ile hesaplanmıştır. Klasik olarak EÇO tahmin edicisi Newton-Rapson metodu olan ancak daha çok Newton metodu olarak literatürde anılan algoritmayı arka planda kullanmaktadır. Buna ek olarak BM (beklenti maksimizasyonu) algoritması kullanılmış ve her iki algoritma ile referans değer olarak kullanılacak tobit model çıkarsamaları sağlıklı bir tabana oturtulmuştur. Bu noktada BM algoritması hakkında kısa ve genel bir paylaşım gerekli görülmüştür.

- BM algoritması

BM algoritması, yaygın kullanılan algoritmaların EÇO tahmin edicisi için belli noktalarda yaşadıkları problemleri çözmek adına geliştirilmiş bir algoritmadır

(Dempster ve ark., 1977). Temelde gözlenen veri ile tam verinin olabilirlikleri arasındaki farka dayanan bir algoritmadır.

Verilerde sansürlemeden dolayı gözlenemeyen değerlerin varlığı gözlenen veri olabilirliğinde eksik bilgiye kaynaklık eder. Bu durum gözlenen verinin olabilirlik fonksiyonunun maksimizasyonunu güç hale getirir. Sansürlü regresyon için BM algoritması bu aşamada sunulmuştur (Amemiya, 1985). Bunun yanında kısmi uyarlamalı tahmin ediciler içinde BM algoritması, hata yapısının normalliğinin karması (Mixture-Normal) olması halinde geliştirilmiştir (Bartolucci ve Scaccia, 2004). Bu iki çalışma üzerine sansürlü regresyon için normal dağılımların karmasına dayanan tahminciler çalışılmıştır (Caudill, 2012). Bu çalışma kapsamında da genelleştirilmiş normal dağılıma dayanan kısmi uyarlamalı tahmin edici sansürlü veri için kullanılmıştır. Ancak BM algoritmasının adımları bu çalışmada referans model olan tobit modelin tahmini için gerçekleştirilmiştir.

Gizil değişken içeren veri setlerinde parametre tahmini için BM algoritmasının sıklıkla kullanıldığı görülmektedir (Yazıcı, 2005). İteratif bir süreç olan bu algoritmada EÇO ya da Bayesian istatistik için önsel bir olabilirlik bulmaya çalışılmaktadır. Sırasıyla,

- Log olabilirliğin beklentisi için fonksiyon oluşturulur (E step),

$$Q(\theta \left| \theta^{(t)} \right.) = E_{Z \left| X, \theta^{(t)} \right.} \left[\log L(\theta; X, Z) \right] \tag{3.30}$$

- Beklenti aşaması olan ilk aşamada bulunan fonksiyonun maksimizasyonu yapılır (M step).

$$\theta^{(t+1)} = \arg\max_{\theta} Q(\theta \left| \theta^{(t)} \right.) \tag{3.31}$$

Bu algoritma özetle, doğrudan çözülemeyen durumlarda lokal olabilirliklerden yola çıkarak çözüme ulaşır. Bu çalışma kapsamında BM algoritma süreci yukarıda da ifade edildiği gibi yalnızca referans değer alınan tobit model çıkarsamalarının sağlıklı bir tabana oturtulması amacıyla tobit model tahmini için uygulanmıştır. Alanda ciddi bir hacme sahip ve özel bir çalışma konusu olarak görülen bu algoritma için daha derin bilgiler ilgili literatürden elde edilebilir.

3.4.1. Parametre tahminde yaşanan sorunlar

Tahmin sürecindeki genel sorunlar: Verideki sansürlü yapının, klasik tahmin edicilerin kullanımı ile sorun oluşturacağı açıktır. Bu durumun başlıca nedeni verinin dağılımının, sürekli ve kesikli dağılımların karma yapısına sahip olmasıdır. Tobit modelin temel tipinde sıfır ya da diğer tiplerinde herhangi bir noktadan sansürlenmenin dikkate alınıyor olması, EKK tahmin edicilerinin yanlı sonuç vermesine neden olmaktadır. Sansürlü veri için tobit model ve normallik varsayımı altında bu modelin EÇO tahmin edicileri yani EÇO tahmin edicisi tutarlı ve asimptotik normaldir. Ancak burada da olabilirlik fonksiyonunun parametrik biçiminin yanlış belirlenmesi tahmin ediciyi yine tutarsız hala getirecektir. Alternatif bir diğer çözüm olarak sansürlü gözlemlerin genel olarak model dışı tutulması ise açık şekilde yanlılık doğuracaktır (Koç, 2013, s.25).

EKK tutarsızlığı: Sansürlemenin sıfır noktasında olduğu varsayımı altında, verideki sansür kaynaklı sıfırlar hata terimlerinin ortalamasını sıfırdan farklılaştırır. Sıfırların tamamen ihmal edilmesi ise etkinlik kaybına yol açmaktadır. Bu bağlamda çözüm, olabilirlik süreçlerini izleyen tobit model tahminleridir.

Sıfır gözlem sorunu ve tobit modelde etkinlik kaybı: Tobit modelde sıfır sansür değeri ile karşılaşmanın üç farklı nedeni gıda tüketim verisi örneği ile şu şekilde ortaya konmuştur:

1. Verilerin toplandığı dönemde tüketici ilgili gıda maddesine sahip olduğu için sıfır tüketim olarak kaydedilmiştir.
2. Tüketici, ilgili gıda maddesini ilgili dönemdeki gelir seviyesi ile satın alamamıştır (Sıfır tüketim). Ancak ilgili gıda maddesinin fiyatı ya da gelir seviyesinde yaşanabilecek değişimler ilgili ürünün tüketimini sıfırdan farklılaştırabilir.
3. İlgili gıda maddesi tüketicilerin tüketim sepetinde yer almayabilir. Dini inançlar, vejetaryenlik, sağlık koşulları vb. nedenler (Ekici, 1996; Şengül, 2004'den aktaran Koç, 2013, s.25).

Bu durumda aslında örneklem seçim (sample selection) problemi mevcuttur. Eğer kişi gerçekten ilgili ürünü hiç tüketmiyorsa tobit model analizi kullanmak etkin olmayan sonuçlar doğurabilir. Bu durumlarda kişinin kararını da içeren genelleştirilmiş denklemler devreye girer. Heckman'ın iki aşamalı tahmin edicisi heckit (1979) bu noktada önerilmiştir. Bu gibi modellerde de örneklem seçim sapması (sample selection

bias)'nın önlenmesi gerekmektedir. Burada örneklem seçim denklemi ayrı bir denklem olarak kabul edilmez. Seçim denklemi ile tahmin denklemi aynı denklemdir. Oysa bu durum birçok örneğin yapısı ile uyuşmamaktadır (Koç, 2013, s.26).

3.5. Tobit Model için Alternatif Tahmin Ediciler

Bir önceki başlıktan da açıkça anlaşıldığı üzere her bir tahmin edicinin kendi içinde belli kısıtı mevcuttur. Tobit model de bazı veri setlerinin analizinde, varsayımlarının sağlanmaması sebepli kullanılamayabilir. Tamda bu aşamada tobit modelin EÇO tahminine alternatif modellerden söz edilebilir. Tobit modelde alternatif olarak kullanılabilecek birçok tahmin edici mevcuttur. Ancak bu çalışma kapsamında belli tahmin ediciler, tobit modelin normal dağılım varsayımı altında EÇO tahmin edicisi ile birlikte incelenmiştir. Sırasıyla,

- EKK tahmin edicisi,
- Normal dağılım altında EÇO tahmin edicisi (EÇO tahmin edicisi),
- Heckman'ın iki aşamalı tahmin edicisi (2AHeckit),
- İki aşamalı EKK (2AEKK),
- Probit model ile sansürlü regresyonun tahmini,
- Logit model ile sansürlü regresyonun tahmini,
- Genelleştirilmiş normal dağılıma dayalı kısmi uyarlamalı tahmin edici (UEÇO)

EKK tahmin edicisi hata kareleri minimum yapmaya dayalı hesaplanan doğrusal regresyon modelleri için klasik yöntemdir ve diğer başlıklarda ilgili bilgiler paylaşılmıştır. Normal dağılım altında EÇO tahmin edicisi yine önceki bölümlerde normal dağılım için ve sansürlü regresyon için ele alınmıştır. Probit ve logit model ile sansürlü regresyonun tahmini içinse, pozitif değerler 1'e negatif değerler 0'a atanarak işlem yapılmıştır. Ayrıca bu iki tahmin edici tezin birinci bölümünde paylaşılmıştır. Bu bölümde ise, 2AHeckit, 2AEKK, UEÇO tahmin edicileri ele alınacaktır.

3.5.1. Heckman'ın iki aşamalı tahmin edicisi (2AHeckit)

X $Kx1$'lik açıklayıcı değişkenler, β $Kx1$'lik bilinmeyen katsayı vektörü ve ε_i, $f(\varepsilon)$ OYF'ye sahip rassal hata değişkeni olmak üzere, $Y_i = \max\{0, x_i'\beta + \varepsilon_i\}$ modelinin

tahmini normallik varsayımı altında Tobin (1958) tarafından çalışılmıştır. Ancak tobit model rassal değişkenin dağılımının doğasından çabuk etkilenir. Yani hatalar normal dağılmıyorsa EÇO tabanlı tahmin edici tutarlı sonuçlar vermeyebilir.

$Y_i = \max\left\{0, x_i^{'}\beta + \varepsilon_i\right\}$'de β'nın tahmininde EKK tahmin edici yalnızca pozitif değerleri kullanırsa (PEKK), EKK tahmin edicisinin yanlılığı ve tutarsızlığı görülecektir. PEKK tahmin edici için bu durum asimptotik yan olarak gösterilmiştir (Goldberger, 1980). Bu aşamada $Y_t > 0$ ya da $Y_t = 0$ bilgisi altında probit model tahmin edicisinden gelen ters mills oranını (Inverse Mills Ratio) dikkate alan iki aşamalı bir tahmin süreci, Heckman (1979) tarafından önerilmiştir (McDonald, Xu, 1996,154).

İlkel hali ile klasik tobit model (Tip I);

$$\begin{aligned} y_i &= y_i^* = x_i^{'}\beta + u \qquad & x_i^{'}\beta + u_i > 0 \\ y_i &= 0 & x_i^{'}\beta + u_i \le 0 \end{aligned} \tag{3.32}$$

şeklinde ifade edilebilir. Burada önemli nokta sansürsüz gözlemlerin koşullu ortalama üzerine etkisi ile X'in bir gözlemin sansürlenmesi olasılığı üzerindeki etkisinin aynı (β) olmasıdır.

Tip II tobit modelde ise iki aşamalı bir tahmin sürecini kullanır. Heckman seçim modeli (Heckman selection model) ya da probit seçim modeli (probit selection model) olarak da adlandırılan heckit model daha karmaşık bir yapıya sahiptir.

X katılım kararı üzerinde etkiye sahip olduğu gibi karar değeri üzerinde de farklı etkilere sahiptir. İlk etki probit kısmı (probit part) ikinci etki ise kırpılmış regresyon (truncated regression) aşaması olarak adlandırılır. Ayrıca Tip I bu durumda Tip II'nin özel bir hali olmuş olur.

Heckman'ın bu iki aşamalı tahmin edicisinin yararları ayrı olarak literatürde çalışılmıştır. Daha çok deneysel (ampirik) çalışmalarda karşılaşılan mikroekomometrik bir durum olduğu ifade edilmiştir. Bu modeli uygulamadan önce eşdoğrusallık probleminin incelenmesi gerektiği vurgulanmıştır. Eşdoğrusallık probleminin olmaması halinde tam olabilirliğin (full-information maximum likelihood-FIML), sınırlı olabilirliğe (limited-information maximum likelihood-LIML) nazaran tercih edilebilir olduğu ifade

edilmiştir. Ancak eşdoğrusallık probleminin varlığı halinde iki aşamalı modelin (sınırlı olabilirlik, LIML) güçlü sonuçlar verdiği ifade edilmiştir (Puhani, 2000, s.53).

Heckman'ın önerisini bir örnekle ayrıntılandırmak gerekirse,

$$y_{1i}^{*} = x_{1i}^{'}\beta_1 + u_{1i} \tag{3.33}$$

$$y_{2i}^{*} = x_{2i}^{'}\beta_2 + u_{2i} \tag{3.34}$$

$$\begin{aligned} y_{1i} &= y_{1i}^{*} \qquad y_{2i}^{*} > 0 \\ y_{1i} &= 0 \qquad\quad y_{2i}^{*} \le 0 \end{aligned} \tag{3.35}$$

şeklinde ifade edilen model yukarıda ifade edilen modelle paralel olup (3.34) eşitliğinde probit kısım olarak tanımlanan karar aşaması tanımlanmaktadır. Kişinin çalışma eğilimi ya da sahip olduğu gözlenen ücreti alınmış olur. x_i eğitim süresi ise araştırılan, okulda fazladan geçirilen sürenin iş piyasasında ücrete yansıyacak artı bir etkisinin olup olmadığıdır. (3.34) ve (3.35) çalışmayan dolayısıyla gözlenen ücreti bilinmeyen kişileri temsil eder. Ekonomik teori, eğitim süresi yüksek olan kişilerin onlara verilen nispeten düşük bir ücrete karşılık çalışmamayı tercih edeceklerini ön görür. O halde u_{1i} ve u_{2i} arasında pozitif bir ilişki vardır. Genel olarak da bu ikisinin iki değişkenli normal (bivariate normal) dağılıma sahip olduğu varsayılır (Puhani, 2000, s.54). Başka bir örnekle ifade etmek gerekirse kişilerin spor salonunda geçirdikleri sürenin tahmini için önce kişilerin spor salonuna gidip gitmeme kararları incelenmelidir. İşte bu aşamada probit kısım söz konusu olacaktır. İkinci kısımda ise gelmeyi tercih etmeyen kişilerin spor salonunda geçirecekleri sürenin bilinmemesi örneklem seçim sorunu olarak görülebilir ve eldeki verilerle (kırpık veri) kırpılmış regresyon kısmı işletilir.

$$\begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \sim BN\left[\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_1^2 \sigma_{12} \\ \sigma_{12} \sigma_2^2 \end{bmatrix}\right] \tag{3.36}$$

olmak üzere, modelin olabilirlik fonksiyonu aşağıdaki gibi yazılabilir:

$$L = \prod_{y_1=0} 1 - \Phi\left(\frac{x_2^{'}\beta_2}{\sigma_2}\right) \prod_{y_1>0} \Phi\left\{\left(x_2^{'}\beta_2 + \frac{\sigma_{12}}{\sigma_1^2}\left(y_1 - x_1^{'}\beta_1\right)\right)\sqrt{\sigma_2^2 - \frac{\sigma_{12}^2}{\sigma_1^2}}\right\} \times \frac{1}{\sigma_1}\phi\left(\frac{y_1 - x_1^{'}\beta_1}{\sigma_1}\right) \tag{3.37}$$

Bu aşamada tam olabilirliğin uzun hesaplamalar gerektirmesine bağlı olarak yukarıda verilen olabilirliğin sınırlı olabilirlik ile hesaplanacağı Heckman (1979) tarafından ileri sürülmüştür.

Bir pozitif y_1^* ile alt örneklem için y_1^*'nin koşullu beklentisi,

$$E\left(Y_{1i}^* \middle| X_{1i}, Y_{2i}^* > 0\right) = x_{1i}^{'}\beta_1 + E\left(u_{1i} \middle| u_{2i} > -X_{2i}^{'}\beta_2\right) \tag{3.38}$$

şeklindedir. (3.36) deki varsayım altında hata teriminin koşullu beklentisi ise;

$$E\left(u_{1i} \middle| u_{2i} > -x_{2i}^{'}\beta_2\right) = \frac{\sigma_{12}}{\sigma_2} \frac{\phi\left(-(x_{2i}^{'}\beta_2/\sigma_2)\right)}{1-\Phi\left(-\left(x_{2i}^{'}\beta_2/\sigma_2\right)\right)} \tag{3.39}$$

şeklinde ifade edilebilir. Bu aşamada y_1^*'nin koşullu beklentisi tekrar yazılacak olursa;

$$E\left(Y_{1i}^* \middle| X_{1i}^{'}, Y_{2i}^* > 0\right) = x_{1i}\beta_1 + \frac{\sigma_{12}}{\sigma_2} \frac{\phi\left(-(x_{2i}^{'}\beta_2/\sigma_2)\right)}{1-\Phi\left(-\left(x_{2i}^{'}\beta_2/\sigma_2\right)\right)} \tag{3.40}$$

şeklinde ifade edilebilir.

$$\lambda\left(x_{2i}^{'}\beta_2/\sigma_2\right) = \frac{\phi\left(-(x_{2i}^{'}\beta_2/\sigma_2)\right)}{1-\Phi\left(-\left(x_{2i}^{'}\beta_2/\sigma_2\right)\right)} \tag{3.41}$$

olmak üzere Heckman'ın iki aşamalı önerisi bu eşitlikte de yer alan ters mills oranının (ters mills ratio) tahminidir. Bu tahmin Probit model yoluyla yapılır. Bu aşamadan sonra aşağıdaki eşitlik tahmin edilir.

$$y_{1i} = x_{1i}^{'}\beta_1 + \frac{\sigma_{12}}{\sigma_2}\lambda\left(x_{2i}^{'}\beta_2/\sigma_2\right) + u_1 \tag{3.42}$$

(3.42) eşitliğinin tahmini ikinci aşama olarak ifade edilmektedir.

EKK tahmin edicisinin $y_1^* > 0$ için alt örneklemde kullanılması halinde ters mills oranıyla (λ) örneklem dışı bırakılan değişkenlerin özel bir durum olarak oluşturduğu problem, örneklem seçim problemi olarak Heckman tarafından ifade edilmiştir. Burada Heckman'ın iki aşamalı modelinin tutarlı olması için u_2'nin normal dağılıma sahip olması ve u_1'nin λ'dan bağımsız olması gerekmektedir. Ancak bu durum u_1'nin değişen

varyans sorununa (heteroscedastic) sahip olması durumunda etkinliğini yitirir. u_1'nin varyansı;

$$V(u_1) = \sigma_1^2 - \frac{\sigma_{12}^2}{\sigma_2^2}\left[\frac{x_{2i}^{'}\beta}{\sigma_2}\lambda\left(\frac{x_{2i}^{'}\beta}{\sigma_2}\right) + \lambda\left(\frac{x_{2i}^{'}\beta}{\sigma_2}\right)^2\right] \tag{3.43}$$

şeklinde ifade edilebilir. Buradan da anlaşılacağı üzere $V(u_1)$ sabit değildir (Puhani, 2000, s.55). Bu aşamada asimptotik varyans-kovaryans matrisinin basit ve sabit tahmin edicilerini elde etmek için White (1980) metodunun kullanılabileceği ifade edilmiştir (Lee, 1982). Seçim yanlılığının olmamasının H_0 hipotezi altında, seçim yanlılığını test etmek için önerilen test, ters mills oranının (λ) katsayısının t-testi yoluyla sınanmasına dayanır (Heckman, 1979, s.158). Bu t-istatistiğinin Langrange çarpan istatistiğine karşılık geldiği, böylece optimallik özelliklerine sahip olduğu ifade edilmiştir (Melino, 1982).

3.5.2. İki aşamalı EKK (2AEKK)

İki aşamalı en küçük kareler (2AEKK) doğrusal eşanlı denklem sisteminde tek bir yapısal denklemin parametrelerini tahmin etmenin bir yöntemi olarak önerilmiştir. Az sayıda araştırmacının bu içeriğe ait bilgi sahibi olması, bu tahmin edicinin belli notasyonlarının belirsiz bırakılmasıyla sonuçlanmıştır. İlk olarak 1950ler'de ortaya atılan bu model 1960-70ler'de sıklıkla kullanılmıştır.

İstatistik, ekonometri, epidemiyoloji ve ilgili disiplinlerde, kontrollü deneyler mümkün olmadığında veya bir tedavinin her birime rasgele bir deneyde başarıyla verilememesi durumunda nedensel ilişkileri tahmin etmek için enstrümantal (araç) değişkenler yöntemi (IV) kullanılır (Imbens, Angrist, 1994). 2AEKK tahmin edicisi daha çok enstrümantal değişkenlerin yardımı ile hesaplanır.

Regresyon modelinde açıklayıcı değişkenler ile hata teriminin korelasyona sahip olması ($Cor(X,\varepsilon) \neq 0$) EKK tahmincilerini hem yanlı hemde tutarsız yapmaktadır. Bu durumun çözümünde 2AEKK veya araç değişkenler metodu alternatif yaklaşımların başında gelmektedir. 2AEKK tahminlemesinde aşağıdaki koşulları sağlayan bir Z değişkenine ihtiyaç duyulur. Bu araç değişken adını alan Z değişkeni aşağıdaki koşulları sağlamalıdır:

$$Cor(Z,\varepsilon) = 0 \tag{3.44}$$

$$Cor(Z,X) \neq 0$$

i. Z ile ε korelasyona sahip olmamalıdır.

ii. Z ile X korelasyona sahip olmalıdır.

2AEKK tahminleme yönteminde EKK aşağıdaki gibi iki kez hesaplanır.

Adım 1. X bağımlı değişken kabul edilerek Z arasında EKK uygulanır ve $\hat{X}$ tahminleri bulunur.

$$X = \gamma Z + v \tag{3.45}$$

olmak üzere,

$$\hat{\gamma} = (Z^T Z)^{-1} Z^T X \tag{3.46}$$

alınarak tahmin değeri;

$$\hat{X} = Z\hat{\gamma} = Z(Z^T Z)^{-1} Z^T X = P_Z X \tag{3.47}$$

şeklinde bulunur.

Adım 2. Y ile $\hat{X}$ arasında EKK uygulanır.

$$Y = \hat{X}\beta + \varepsilon \tag{3.48}$$

buradan da,

$$\begin{aligned} \hat{\beta}_{2AEKK} &= (\hat{X}^T \hat{X})^{-1} \hat{X}^T Y \\ Y = X\hat{\beta}_{2AEKK} &= \hat{X}(\hat{X}^T \hat{X})^{-1} \hat{X}^T Y \\ \hat{\beta}_{2AEKK} &= \left((ZX)^T (ZX) \right)^T (ZX)^T Y \\ \hat{\beta}_{2AEKK} &= \left(X^T Z^T ZX \right) X^T Z^T Y \end{aligned} \tag{3.49}$$

olarak bulunur.

3.5.3. Kısmı uyarlamalı EÇO tahmin edici (UEÇO)

Kısmı uyarlamalı tahmin ediciler-UEÇO (partially adaptive estimators-PAE) ya da bir diğer ismiyle yarı olabilirlik tahmin edicileri (quasi-maximum likelihood estimators)

sansürlü regresyon modelinin log-olabilirlik fonksiyonunun uygun hale getirilmesi (optimize) ile elde edilir. Bu işlem regresyon parametresi (β) ile dağılım parametresinin eş zamanlı tahmin edilmesi ile yapılır.

Sansürlenmiş veride hatanın normal dağıldığı varsayımı altında kullanılan EÇO tahmin edicisinin, hata dağılımının normal olmaması durumunda EÇO tahmin edicisinin dolayısı ile tobit model analizi sonuçlarının etkin olmadığı daha önce ifade edilmiştir. Bu durum normalliğide içinde barındıran tahmin edici arayışına yol açmıştır (Caudill, 2012, 121). Sansürlü regresyon için yarı parametrik tahmin ediciler yoğunluk tabanlı (density based) ve yoğunluk tabanlı olmayan (non-density based) olmak üzere kategorize edilmiştir (Pagan ve Ullah, 1999). Yoğunluk tabanlı olmayan tahmin edicilerin sansürlü en küçük mutlak sapma (SEKMS) ve simetrik kırpılmış en küçük kareler (STLS) gibi tahmin ediciler olduğu ifade edilmiştir. Tamamen uyarlamalı tahmin edicileri (fully adaptive estimators) ve UEÇO tahmin edicileri ya da bir diğer ismiyle yarı olabilirlik tahmin edicileri ise yoğunluk tabanlı tahmin edicilerdir. Tamamen uyarlamalı tahmin edicilerin (fully adaptive estimators) bilinmeyen dağılımın parametrik olmayan tahminine dayandığı ifade edilirken, UEÇO tahmininin doğru bilinmeyen hata dağılımı için parametrik bir yaklaşım olduğu ifade edilmiştir (Caudill, 2012, 122).

Hata dağılımının sınırlı dağılımlara uygunluk göstermesinin yanında bu dağılımlara alternatif dağılımlarda (Laplace, Cauchy) Monte Carlo deneyi ile farklı örneklem genişliklerinde (50, 100, 200) ve farklı sansürleme seviyelerinde (25%, 50%) sınanmıştır (Paarsch, 1984). Paarsch'ın çalışması genişletilerek hata dağılımının simetrik olmadığı durumlar sınanmış ve kullanılan tahmin edicilere ek olarak UEÇO tahmin edicileri ile yarı parametrik tahmin ediciler (semi-parametric estimator) kullanılmıştır. McDonald ve Xu çalışmalarında örneklem medyanı, standart hata ve hata kareler ortalaması (HKO) kullanarak esnek dağılımlara dayalı kısmı uyarlamalı tahmin ediciler ile tobit, SEKMS, Heckman, PEKK, yarı parametrik en küçük kareler (SP-LS) ve yarı parametrik en çok olabilirlik (SP-MLE) tahmin edicilerini karşılaştırmıştır. Bu aşamalarda farklı hata dağılımları (Normal, Cauchy, Log-normal) kullanılmıştır (McDonald, Xu, 1996,156). Bir farklı çalışmada da yine Paarsch (1984) temel alınmış ve Cauchy, Laplace ve log-normal hata dağılımı kabul edilmek üzere farklı örneklem boyutlarında tobit, SEKMS, iki aşamalı en küçük kareler (TSLS/2SLS) ve normal dağılımların konum-ölçek karma tanımlı (location-scale mixture of normal distributions-

PAM) tahmin edicileri sansürlü veri için karşılaştırılmıştır. Ayrıca bu çalışmada hata terimlerinin ortalaması sıfırda, varyansları ise yüzde sabitlenmiştir. Cauchy dağılımının da ölçek parametresi ise 10 olarak alınmıştır. Eğim parametresinin gerçek değeri ise 1 olarak verilmiştir ($\beta = 1$). Her bir Monte Carlo deney kurgusu 1000 rassal deneme üzerine kurulmuştur. Simülasyon sonuçları ise ortalama, medyan ve RMSE değerleri ile sunulmuştur (Caudill, 2012, s.128). Normal, karma normal ve log-normal hata dağılımları ile normal, kalın kuyruklu ve çarpık hata dağılımlı durumlarda UEÇO tahmin edicisinin sansürlü regresyon modelindeki durumları diğer tahmin ediciler ile karşılaştırılmıştır (Lewis ve McDonald, 2014, 742).

Kısmi uyarlamalı tahmin edicilerin çözmeye çalıştığı üç ana durum mevcuttur. Bunlar sırasıyla,

- Farklı sınırlı bağımlı değişkenli modellere UEÇO tahmin edicisinin uygulamaları,
- Tahmin edici performansları üzerine farklı esnek hata yapılarının etki ve kullanımlarının incelenmesi,
- Sınırlılık için koşullar (Caudill, 2012, 122).

Bu çalışmada kullanılan UEÇO tahmin edicisi ikinci duruma uygunluk göstermektedir. Bu çalışma kapsamında genelleştirilmiş normal dağılımına dayanan kısmi uyarlamalı en çok olabilirlik tahmin edici kullanılmıştır.

- Genelleştirilmiş normal dağılım

  Genelleştirilmiş normal dağılım (generalized normal distribution) ya da diğer isimleri ile genelleştirilmiş Gaussian dağılım (generalized Gaussian distribution-GGD) iki versiyona sahiptir. Her ne kadar sabit bir isimlendirme olmasa da versiyon 1 üstel güç dağılımı (exponential power distribution) ya da genelleştirilmiş hata dağılımı (generalized error distribution) ismi ile anılmaktadır.

  X rassal değişkeni $(-\infty, +\infty)$ aralığında tanımlı ve μ, α ve β sırasıyla konum ölçek ve şekil parametreleri olmak üzere sırasıyla OYF ve BDF genelleştirilmiş normal dağılım (GND) için aşağıda sırasıyla sunulmuştur.

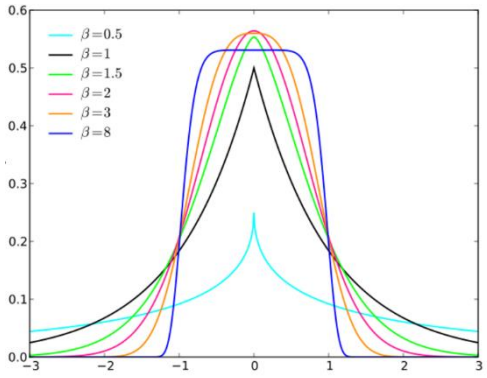
$$f(x) = \frac{\beta}{2\alpha\Gamma(1/\beta)} e^{-(|x-\mu|/\alpha)^{\beta}} \tag{3.50}$$

Bu eşitlikte $\Gamma(.)$ sembolü gamma fonksiyonunu temsil etmektedir.

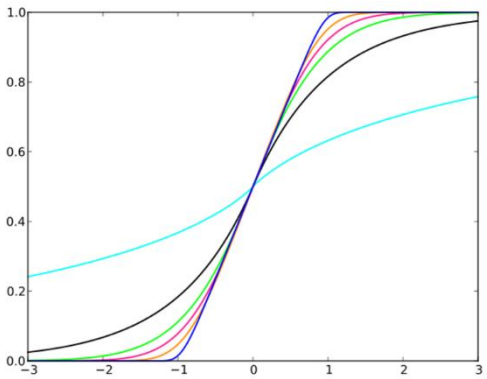
$$F(x) = \begin{cases} \dfrac{\Gamma\left(1/s, \left((\mu - x)/\sigma\right)^{s}\right)}{2\Gamma\left(1/s\right)} & , \quad x \le \mu \\ 1 - \dfrac{\Gamma\left(1/s, \left((x - \mu)/\sigma\right)^{s}\right)}{2\Gamma\left(1/s\right)} & , \quad x > \mu \end{cases} \tag{3.51}$$

Bu eşitlikte $\Gamma(.,.)$ sembolü tamamlanmamış gamma (incomplete gamma) fonksiyonunu temsil etmektedir.

Ayrıca OYF ve BDF sırasıyla aşağıdaki şekillerde sunulmuştur.



Şekil 3.1. *GND ailesinin belli değerler için OYF grafiği*



Şekil 3.2. *GND ailesinin belli değerler için BDF grafiği*

.

Yukarıda sunulan GND daha önce de ifade edildiği gibi versiyon 1 olarak anılıp genelleştirilmiş hata dağılımı ismiyle de kaynaklarda geçmektedir. Literatürle bütünlük arz etmesi açısından paylaşılacak olursa, genelleştirilmiş hata dağılımının ilk tipi GED-1, ikinci tipi GED-2 olarak geçmektedir. GED-1'in ağır kuyruklu (heavy tail), GED-2'nin ise yüksek eğik kuyruklu (highly skewed tail) olduğu ifade edilmiştir (Vasudeva ve Kumari, 2013).

Bu dağılımın momentlerinin elde edilmesi için GND için farklı bir gösterim yolu kolaylık amacı ile izlenebilir. Bilindiği üzere $X \sim N\left(\mu, \sigma^2\right)$ için OYF;

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-((x-\mu)^2/2\sigma^2)} \tag{3.52}$$

şeklindedir ve bu eşitlik şu şekilde genelleştirilebilir.

$$f(x) = Ke^{-(|x-\mu|/\alpha)^s} \tag{3.53}$$

açıktır ki (3.52) ile (3.53) $K = \frac{s}{2\alpha\Gamma(1/s)}$ için aynı OYF'yi ifade eder. Burada $s = 2$ için normal dağılım, $s = 1$ için Laplace dağılımları elde edilir.

X rassal değişkeninin GND'a sahip olduğu ve dolayısı ile X rassal değişkeninin OYF'sinin (3.53)'deki gibi olduğu bilinsin. Bu durumda $Z = (X - \mu)/\sigma$ için OYF;

$$f(\mathrm{z}) = \frac{se^{-|z|^s}}{2\Gamma(1/s)} \tag{3.54}$$

olur.

Bu aşamada Z için k. moment;

$$E(\mathrm{Z}^k) = \frac{1+(-1)^k}{2\Gamma(1/s)}\Gamma\left(\frac{k+1}{s}\right) \tag{3.55}$$

şeklinde elde edilir.

O halde X için n. moment;

$$E(\mathrm{X}^n) = E\left[(\mu + \sigma\,\mathrm{Z})^n\right] \tag{3.56}$$

$$= \sum_{k=0}^{n} \binom{n}{k} \mu^{n-k} \sigma^k E(\mathrm{Z})^k$$

$$= \frac{\mu^n \sum_{k=0}^{n} \binom{n}{k} (\sigma/\mu)\left[1 + (-1)^k\right] \Gamma\left((k+1)/s\right)}{2\Gamma\left(1/s\right)}$$

şeklinde elde edilir.

Bu durumda ilk dört moment yazılmak istenirse;

$$\begin{aligned} E(\mathrm{X}) &= \mu \\ E(\mathrm{X}^2) &= \mu^2 + \frac{\sigma^2 \Gamma\left(3/s\right)}{\Gamma\left(1/s\right)} \\ E(\mathrm{X}^3) &= \mu^3 + \frac{3\mu\sigma^2 \Gamma\left(3/s\right)}{\Gamma\left(1/s\right)} \\ E(\mathrm{X}^4) &= \mu^4 + \frac{6\mu^2\sigma^2 \Gamma\left(3/s\right)}{\Gamma\left(1/s\right)} + \frac{\sigma^4 \Gamma\left(5/s\right)}{\Gamma\left(1/s\right)} \end{aligned} \tag{3.57}$$

X için $n.$ merkezi moment ise;

$$E\left(\left(\mathrm{X} - \mu\right)^n\right) = \sigma^n \left[1 + (-1)^n\right] \int_{\mu}^{\infty} \left(\frac{x-\mu}{\sigma}\right)^n \frac{s \exp\left\{-\left(\left(x-\mu\right)/\sigma\right)^2\right\}}{2\sigma\Gamma\left(1/s\right)} dx \tag{3.58}$$

$$= \frac{s\sigma^n \left[1 + (-1)^n\right]}{2\Gamma\left(1/s\right)} \int_0^{\infty} z^n \exp\left(-z^s\right) dz$$

$$= \frac{\sigma^n \left[1 + (-1)^n\right]}{2\Gamma\left(1/s\right)} \int_0^{\infty} y^{(n+1)(s-1)} \exp\left(-y\right) dy$$

$$= \frac{\sigma^n \left[1 + (-1)^n\right] \Gamma\left(n + 1/s\right)}{2\Gamma\left(1/s\right)}$$

şeklinde elde edilir.

Bu durumda;

$$Var(X) = \frac{\sigma^2 \Gamma(3/s)}{\Gamma(1/s)} \tag{3.59}$$

$$E\left[\left\{X - E(X)\right\}\right]^3 = 0$$

$$E\left[\left\{X - E(X)\right\}\right]^4 = \frac{\sigma^4 \Gamma(5/s)}{\Gamma(1/s)}$$

$$skew(X) = 0$$

$$Kurt(X) = \frac{\Gamma(1/s)\Gamma(5/s)}{\Gamma^2(3/s)}$$

şeklinde elde edilir. Görüldüğü gibi ortalama μ, çarpıklık 0'dır. Ancak varyans ve basıklık şekil parametresine bağlıdır (Nadarajah, 2005, s.687).

Yukarıda kısa bir şekilde tanıtılan GND hakkında daha geniş bilgiye (Nadarajah, 2005, Do ve Vetterli, 2002, Varanasi ve Aazhang 1989) çalışmalarından ulaşılabilir.

Parametrik bir aile seçiminde esneklik oldukça önemlidir. Çarpıklık ve basıklık (skewness-kurtosis) esnekliğin bir ölçüsü olarak kullanılabilmektedir. Parametrik ailelerin esneklik bakımından karşılaştırılmasında şu formülden yararlanılmaktadır (Caudill, 2012, 123).

$$basıklık \geq çarpıklık^2 + 1 \tag{3.60}$$

Tablo 3.1. *GND'nin farklı parametreleri için basıklık (kurtosis) değerleri*

Şekil parametresi (s)	0.25	0.5	1	2	4	8
Basıklık Değerleri	458.0727	25.20	6.000	3.000	2.1884	1.9234

Tablo 4.1'deki değerlere bakılırsa, bu çalışma kapsamında kullanılan ve yukarıda kısaca tanıtılan GND'ı esnek bir dağılımdır. Ayrıca PAM hata yapılarına dayanan UEÇO tahmin edicisi sansürlü regresyon için tanıtılmıştır. Söz konusu tahmin edici, normal dağılımı ve Laplace dağılımını içermesinden dolayı, tobit modeli barındırmaktadır. Çalışma kapsamında kullanılan dağılımda, simetrik dağılımların parametrik aile üyesidir. Söz konusu dağılım, normal ve Laplace dağılımları ile sınırlı durumlarda düzgün dağılımı

barındırır. $s = 2$ iken μ ortalamalı, $\frac{\alpha^2}{2}$ varyanslı normal dağılım, $s = 1$ iken Laplace dağılımı, $s \to \infty$ iken yoğunluk noktasal olarak $(\mu - \alpha, \mu + \alpha)$ üzerinde düzgün dağılıma yakınsar.

GND'in tam veri için likelihood ve log likelihood fonksiyonu sırası ile verilmiştir:

$$f(x_1,....,x_n) = \left(\frac{s}{2\sigma}\right)^n \left(\Gamma\left(\frac{1}{s}\right)\right)^{-n} e^{-\sum\left|\frac{x_i-\mu}{\sigma}\right|^s} \tag{3.61}$$

$$L(\mu,\sigma,s) = \ln(f(x_1,....,x_n)) = n\ln(n) - n\log(2\sigma) - n\ln\left(\Gamma\left(\frac{1}{s}\right)\right) - \sum\left|\frac{x_i-\mu}{\sigma}\right|^s \tag{3.62}$$

Soldan Kırpılmış GND'in OYF'si aşağıda verilmiştir.

$$f(x) = \frac{\left(\frac{s}{2\sigma}\right)^n \left(\Gamma\left(\frac{1}{s}\right)\right)^{-n} e^{-\left|\frac{x_i-\mu}{\sigma}\right|}}{F(x)} \tag{3.63}$$

(3.63)'te kullanılan BDF (3.51)'de tanımlandığı gibidir.

Tobit Tip 1 model yani soldan sıfırda sansürlü regresyon için GND'nin likelihood fonksiyonu $(c = 0)$;

$$L(\mu,\sigma,s) = \prod_{y\le 0} F(y) \prod_{y>0} f(y) \tag{3.64}$$

$$L(\mu,\sigma,s) = \sum_{y\le 0} \ln((F(y)) + \sum_{y>0} \ln(f(y))$$

$$= (n-n_c)\ln(\mathrm{n}) - (n-n_c)\log(2\sigma) - (n-n_c)\ln\left(\Gamma\left(\frac{1}{s}\right)\right)$$

$$-\sum_{y_i>0}\left|\frac{y-x_1'\beta}{\sigma}\right|^s - n_c\ln(2) - n_c\ln\left(\Gamma\left(\frac{1}{s}\right)\right) + \sum_{y=0} 1 - \frac{\Gamma\left(\frac{1}{s}, \left(\frac{y-x_1'\beta}{\sigma}\right)^s\right)}{2\Gamma\left(\frac{1}{s}\right)}$$

şeklindedir. Burada $(n-n_c)$ sansürsüz gözlem sayısı, n_c ise sansürlü gözlem sayısıdır.

Olabilirlik fonksiyonunun içinde tamamlanmamış (incomplete) gamma bulunması sebebiyle, olabilirlik fonksiyonun maksimizasyonu çok zor olacaktır. Bu nedenden, tamamlanmamış (incomplete) gamma fonksiyonunun aşağıda verilen seri açılımı hesaplamalarda kullanılmıştır. Aşağıda, söz konusu tamamlanmamış (incomplete) gamma fonksiyonunun seri açılımı verilmiştir (Amore, 2005). $a > 0$ olmak üzere,

$$\Gamma(a,x) \approx x^{a-1}e^{-x}\left[1+\frac{a-1}{x}+\frac{(a-1)(a-2)}{x^2}+...\right] \tag{3.65}$$

Olabilirlik fonksiyonunda $\Gamma\left(\frac{1}{s},\left(\frac{y-x_1'\beta}{\sigma}\right)^s\right)$ için seri açılımı aşağıdaki şekildedir.

$$\Gamma\left(\frac{1}{s},\left(\frac{y-x_1'\beta}{\sigma}\right)^s\right)=\left(\frac{y-x_1'\beta}{\sigma}\right)^{\frac{1}{s}-1}\exp\left(-\left(\frac{y-x_1'\beta}{\sigma}\right)^s\right)\left[1+\frac{\frac{1}{s}-1}{\left(\frac{y-x_1'\beta}{\sigma}\right)^s}+...\right] \tag{3.66}$$

Bir diğer açıdan literatür incelendiğinde doğrusal regresyon modeli için genelleştirilmiş t-dağılımı (generalized t-distribution), normal hata yapılarının karmaları (mixture of normals error structure), maksimum entropi dağılımı (maximum entropy distribution) gibi dağılımlara ve yapılara dayanan UEÇO tahmin ediciler sınanmıştır (McDonald ve Newey, 1988, Phillips, 1994,Wu ve Stengos, 2005). Bu çalışmaların son dönemlerde sansürlü regresyona uyarlamaları da mevcuttur.

Hatanın normal olmama problemi, artan örneklem sayısı ile uygun hale gelirken bu problem bir noktada sürüncemede bırakılmıştır. Tobit modelde de normallik varsayımı vardır ve bu varsayımın esnetilmesi için çalışmalar yürütülmektedir. Ancak normallik varsayımının esnetilebilmesinin tek yolunun, esnek bir OYF ile hata dağılımının modellenmesi olduğu ifade edilmiştir. Kullanılacak esnek dağılımın OYF'sinin ve BDF'sinin normal dağılıma karşılık geldiği durumların varlığı, bu esnek dağılımın tobit modeli bir noktada içerdiğini göstermektedir. Bu durum olabilirlik oran testi ile kolaylıkla test edilebilir (Lewis ve McDonald, 2014, 733).

3.6. Modelin Uygunluk Ölçüsü

Uyum iyiliği ölçüsü (R^2) doğrusal modellerde, tahmin edilen regresyon modeli ile gözlenen değerler arasındaki uyumu gösteren belirleme katsayısıdır. En ilkel haliyle

tahmin edilen model ile reel değerlerin uyumu (uyum iyiliği) temsil edilir. Çalışma kapsamında konu edilen sınırlı bağımlı değişkenli modellerde ise R^2 olduğundan daha küçük sonuçlar verir (Tatoğlu, 2005). Belirlilik katsayısının oldukça düşük değer vermesi, uyum iyili ölçüsü olarak belirlilik katsayısının sınırlı bağımlı değişkenli modellerde kabul görmemesine neden olmuştur (Vermek, 2004, s.182). Belirlilik katsayısının bu haliyle kullanılması çıkarsamalarda hataya sebep olabilir. Bu aşamada R^2 yerine düzeltimiş R^2 anlamına gelen $pseudoR^2$ kullanılır. $pseudoR^2$ de R^2 gibi 0-1 aralığında değer alır. Ayrıca $pseudoR^2$ de yine R^2 gibi sabit terimler dışında tüm değişkenlerin sıfır olduğu hipotezine dayanır (Eren, 2012, s.37).

$H_0 : \beta_1 = \beta_2 = ... = \beta_k = 0$ (kısıtlama durumunda) ise

L_u : Genel modelin olabilirliği

L_c : Sadece sabit terim içeren modelin olabilirliği

N : Gözlem sayısı

K : Tahmin edilen parametre sayısı olmak üzere,

$$\ln\left(\frac{P_i}{1-P_i}\right) = \beta_0 + u_i$$

biçimindeki olabilirlik fonksiyonunun maksimum değeri L_c ,

$$\ln\left(\frac{P_i}{1-P_i}\right) = \beta_0 + \beta_1 X_1 + ... + \beta_k X_k + u_i \tag{3.67}$$

şeklindeki olabilirlik fonksiyonun maksimum değeri L_u olur.

Bu bilgiler ışığında uyum iyiliğinin ölçüsü olarak kullanılan istatistikler aşağıdaki gibi sıralanabilir.

- Mc-Fadden tarafından önerilen $pseudoR^2$;

$$pseudoR_{MF}{}^2 = 1 - \frac{\ln L_u}{\ln L_c} \tag{3.68}$$

Teorik Aralığı: $0.2 \leq pseudoR_{MF}{}^2 \leq 0.4$

Bu değer aynı zamanda olabilirlik oran indeksi (likelihood ratio index) olarakta adlandırılmaktadır (Ramanathan, 2002: 284).

Ayrıca bir diğer düzeltilmiş (adjusted) $pseudoR_{MF}{}^2$ ise;

$$pseudoR_{MF}{}^2 = 1 - \frac{\ln L_u - K}{\ln L_c} \tag{3.69}$$

olarak ifade edilir.

- Estella tarafından önerilen $pseudoR^2$;

  İki farklı şekilde gösterilmek üzere;

$$pseudoR_{E1}{}^2 = 1 - \left(\frac{L_u}{L_c}\right)^{\frac{2}{N}\ln L_c} \tag{3.70}$$

$$pseudoR_{E2}{}^2 = 1 - \left(\frac{\ln L_u - K}{\ln L_c}\right)^{\frac{2}{N}\ln L_c} \tag{3.71}$$

  şeklindedir.

Bu iki $pseudoR^2$'nin yanında Cox-Snell tarafından önerilen $pseudoR^2$ (Teorik Aralığı: $0 \le pseudoR_{CS}{}^2 \le 1 - L_c^{2/N}$); Cragg-Uhler tarafından önerilen $pseudoR^2$ (Teorik Aralığı: $0 \le pseudoR_{CU}{}^2 \le 1$); Aldrich-Nelson tarafından önerilen $pseudoR^2$; Veall-Zimmermann tarafından önerilen $pseudoR^2$; McKelvey-Zavoina tarafından önerilen $pseudoR^2$; Nagelkerke/Cragg ve Uhler tarafından önerilen $pseudoR^2$ mevcuttur. Ancak önerilen $pseudoR^2$ değerleri birbirine yakın sonuçlar vermektedir. Bu durum ve hesaplama kolaylığı göz önünde bulundurulduğunda Mc-Fadden ve Estelle tarafından önerilen $pseudoR^2$ değerlerinin daha yaygın kullanıldığı ifade edilmiştir (Yerdelen Tatoğlu, 2005, s.89).

Yukarıdaki ifadelere rağmen sınırlı bağımlı değişkenli herhangi bir modelin dağılımının karma yapısı (kesikli-sürekli), R^2 değerleri gibi $pseudoR^2$ değerlerinin de [0,1] aralığının dışına çıkmasına neden olabilir. Bu durum karma yapı göz önünde bulundurularak farklı şekillerde yorumlanabilir. Kesikli, sürekli ve sürekli/kesikli modeller için;

I. Kesikli modellerde;

Log-olabilirlik değeri olasılığın logaritmasıdır ve her zaman negatif ya da 0 değer alır. Bu durumda $0 \geq L \geq L_0$ ve $1 \geq L / L_0 \geq 0$ yazılabilir. $pseudoR^2$, $1 - L / L_0$ formülünden $0 \leq pseudoR^2 \leq 1$ aralığında yer alır.

II. Sürekli modellerde;

Log-olabilirlik değeri yoğunluğun logaritması olup bu değer pozitif ve negatif değer alabilir.

III. Sürekli/kesikli modellerde;

a. Eğer $L < 0, L_0 > 0$ ise ;

$L / L_0 < 0$ olur ve dolayısıyla $1 - L / L_0 > 0$ bulunur.

b. Eğer $L > L_0 > 0$ ise;

$L / L_0 > 1$ olur ve dolayısıyla $1 - L / L_0 < 0$ bulunur.

Bunların yanında Pearson Ki-kare uygunluk testi de bir diğer uygunluk ölçüsüdür. Verilerin herhangi bir dağılıma uygun olup olmadığı, beklenen değerler ile gerçek değerlerin arasında farklılığın olup olmadığı bu test istatistiği ile incelenebilir.

3.7. Parametreler İçin Sınırlama Testleri

Kullanılan modelde hangi parametrenin kullanılacağına ya da model dışı bırakılacağına sınırlama testleri (tahmin sonuçlarının anlam testleri) ile karar verilir. Tahmin sonuçlarının anlamlılığının testi iki şekilde yapılabilir. Parametrelerin ayrı ayrı anlamlılıklarının test edilebileceği gibi, parametrelerin aynı anda ve birlikte anlamlılıkları test edilebilir. Kullanılacak bu testlerin her birinin sınama istatistiği ki-kare dağılımına uygunluk gösterir (Arıcan, 2010). Bu test istatistikleri hem doğrusal hem de doğrusal olmayan modellerde kullanılabilir. Ancak F ve t testleri, katsayıların anlamlılığını ölçmekle birlikte doğrusal olmayan modellerde kullanılmaz. Maksimum olabilirliğe dayalı analizlerde, büyük örneklemlerde eşit sonuçlar veren uygun test istatistikleri;

- OO,
- Wald,
- Lagrange Çarpanı (LÇ) testleridir.

❖ Parametrelerin ayrı ayrı anlamlılığı test edilecekse, T testi kullanılabilir;

$$t = \frac{\hat{\beta}_i}{\sigma_{\hat{\beta}_i}} \tag{3.72}$$

olmak üzere buradan elde edilecek değer ilgili anlam düzeyinde tablodan okunacak değerle karşılaştırılır. Neticede ise yokluk hipotezinin ($H_0 : \beta_i = 0$) kabulüne ya da reddine karar verilir.

❖ Parametrelerin birlikte anlamlılığı test edilecekse, OO, Wald, LÇ testleri kullanılabilir; her biri ayrı ayrı ele alınmadan önce bilinmelidir ki, LÇ testi sadece kısıtlandırılmış ($H_0 : \beta_1 = \beta_2 = ... = \beta_k = 0$) modelin tahmin sonuçlarına, Wald testi sadece kısıtlandırılmamış (H_1 : En az biri sıfırdan farklıdır.) modelin tahmin sonuçlarına, OO testi ise hem kısıtlandırılmış hemde kısıtlandırılmamış model sonuçlarına dayanır (B. Güriş, 2005: 19).

- Langrange çarpan testi (LM);

R_U^2 : Kısıtsız (unresicted) modelin belirlilik katsayısı

SSR : Sum square of regression (KKT : Kalıntı kareler toplamı)

SST : Sum square total (BKT : Bütün kareler toplamı) olmak üzere,

$$R^2 = \frac{\sum\left(\hat{Y}_i - \bar{Y}\right)^2}{\sum\left(\hat{Y}_i - \bar{Y}\right)^2} = \frac{SSR}{SST} = \frac{KKT}{BKT} \tag{3.73}$$

şeklinde ifade edilir ve Langrange çarpan testi (LM),

$$LM = nR_U^2 \tag{3.74}$$

olarak ifade edilir.

Bu şekilde tüm bağımsız değişkenlerin sıfırdan farklı olup olmadığı test edilmiş olur. Bu test istatistiği de χ^2 (ki-kare) dağılımına uyar ve $(k-1)$ serbestlik derecesi olmak üzere χ^2 değeri okunur. Bu aşamada ilgili önem derecesi ve belirtilen serbestlik derecesi değeriyle yukarıdaki eşitlikten elde edilen test istatistiği kıyaslanır. $LM > \chi^2_{k-1}$ ise yokluk hipotezi reddedilir.

- Wald testi;

SSE_R (resicted sumsquare of error): kısıtlı hata kareler toplamı

SSE_U (unresicteds umsquare of error): kısıtsız hata kareler toplamı ve

$$\hat{\sigma}_e^2 = \frac{SSE_U}{n} \tag{3.75}$$

olmak üzere kısıtsız tahmini gerektiren (Thomas, 1997: 258) Wald test istatistiği;

$$W = \frac{SSE_R - SSE_U}{\hat{\sigma}_e^2} \tag{3.76}$$

şeklinde ifade edilir.

Burada kısıt sayısı 1'dir. Bu anlamda 1 serbestlik dereceli χ^2 (ki-kare) dağılımına uygunluk vardır. Yukarıdaki eşitlik yardımı ile elde edilen test istatistiği, ilgili anlam düzeyinde ilgili tablo değeri kullanılarak karşılaştırılır. Bu bağlamda yokluk hipotezinin varlığı-yokluğu test edilirken aslında katsayıların birlikte anlamlılıkları test edilmiş olur.

- Olabilirlik oran (OO) testi (Likelihood Ratio Test),

l_R : Kısıtlı (resicted) modelin EÇO değerinin logaritması,

l_U : Kısıtsız (unresicted) modelin EÇO değerinin logaritması olmak üzere

$$LR = -2(l_R - l_U) = -2(\log L_c - \log L_u) \tag{3.77}$$

ifadesi olabilirlik oran test istatistiğini verir. Bu test logit ve probit model için F testi yerine kullanılabilir.

Kısıt sayısını serbestlik derecesi kabul eden χ^2 (ki-kare) değeri ile bu istatistik kıyaslanır. Bu duruma bağlı olarak ilgili anlam düzeyindeki olabilirlik oran testinin kısıt sayısını serbestlik derecesi kabul eden χ^2 değerinden büyük olması durumunda yokluk hipotezi reddedilir. Bu bağlamda kısıtların birlikte istatistiksel açıdan anlamlı olduğu ifade edilir (Eren, 2012, s.32-35).

4. SANSÜRLÜ REGRESYON İÇİN TAHMİN EDİCİLERİN PERFORMANSLARININ İNCELENMESİ

4.1. Simülasyon Çalışması

Literatürle paralel olarak simülasyon çalışmasında, hata dağılımları için Normal, Normal-Normal karışım, Student-t ve Laplace dağılımları kullanılmıştır. Örneklem hacimleri 50, 100, 200, 400, 500 ve 800 seçilmiştir. Simülasyonlar, iterasyon sayısı 100.000/n olacak şekilde gerçekleştirilmiştir.

Hata dağılımları aşağıdaki şekildedir:

- Normal N(0,1)
- 0.90 Normal N(0,9)+0.10 Normal N(0,1/9)
- 0.80 Normal N(0,9)+0.20 Normal N(0,1/9)
- Student (3)
- Laplace (0,1)

$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \;\; i = 1,...,n$ doğrusal regresyon modelinde $\beta_0 = 0$ ve $\beta_1 = 1$ alınmış ve $X_i \sim U(0,1)$ olmak üzere yukarıda bahsedilen dağılımlar hata teriminin dağılımı olacak şekilde veri üretilmiştir. Elde eldilen veriler için bahsedilen tahmin ediciler yardımıyla tahminler hesaplanarak HKO (MSE) ve Yan (Bias) değerleri hesaplanmıştır.

Tahmin edicilerin karşılaştırılmasında n örneklem hacmini göstermek üzere kullanılan Yan ve HKO sırasıyla,

$$Yan(\hat{\beta}) = \left(\frac{1}{100000/n} \sum \hat{\beta} \right) - \beta \tag{4.1}$$

$$HKO(\hat{\beta}) = \frac{1}{100000/n} \sum (\hat{\beta} - \beta)^2 \tag{4.2}$$

şeklindedir.

Ele alınan tahmin ediciler ve tablolardaki kısaltmaları aşağıda verildiği gibidir:

- Yalnızca gözlenen veriler için EKK – EKK_g
- EM algoritması kullanılarak elde edilen EÇO tahmin edicisi – EÇO_EM
- Heckit tahmin edicisi – 2AHeckit,

- İki aşamalı EKK tahmin edicisi – 2AEKK,
- Probit model – Probit,
- Logit model – Logit,
- Genelleştirilmiş normal dağılıma dayalı kısmı uyarlanabilir tahmin edici-UEÇO

Hatanın normal dağılımdan gelmesi halinde 50, 100, 200, 400, 500 ve 800 örneklem hacmi için elde edilen sonuçlar Tablo 4.1'de sunulmuştur.

Tablo 4.1. *Hatanın normal dağılımdan gelmesi halinde elde edilen Yan ve HKO değerleri*

$y = a + b * x$	Normal					
Eğim $b = 1$	n=50		n=100		n=200	
	Yan	HKO	Yan	HKO	Yan	HKO
EKK_g	0,31618	0,23880	0,32714	0,17081	0,31389	0,12975
EÇO_EM	-0,00079	0,25920	0,01872	0,12082	0,00217	0,05479
2AHeckit	-0,43650	12,7764	-0,31510	2,44711	0,21643	1,04927
2AEKK	-0,00078	0,26196	0,01805	0,12230	0,00209	0,05572
Probit	-0,03723	0,48039	0,00127	0,20939	-0,00367	0,09784
Logit	-0,68377	1,76747	-0,61156	0,93375	-0,61636	0,64110
UEÇO	-0,04180	0,26670	0,03637	0,16811	0,05251	0,09613
	n=400		n=500		n=800	
	Yan	HKO	Yan	HKO	Yan	HKO
EKK_g	0,31990	0,11902	0,31222	0,11018	0,30085	0,09980
EÇO_EM	0,01239	0,02862	-0,00721	0,02326	-0,01370	0,01490
2AHeckit	0,04531	0,07204	0,09189	0,09120	-0,02530	0,01420
2AEKK	0,00980	0,02965	-0,00511	0,02381	-0,01509	0,01558
Probit	0,00527	0,04608	-0,01290	0,03657	-0,01164	0,02572
Logit	-0,59973	0,48214	-0,62906	0,49240	-0,62674	0,46085
UEÇO	0,05033	0,05775	0,02613	0,04403	0,02911	0,03432

Tablo 4.1'e göre,

- Küçük örneklemde (n=50), sansürlü veri sayısının düşük olmasından dolayı EKK en küçük HKO değeri ile iyi performans ortaya koysa da örneklem sayısının arttırılmasına bağlı olarak performans kaybına uğramıştır. Teorik olarak daha önce verilen nedenlerden dolayı sansürlü veri sayısının artmasına bağlı olarak bu sonuç beklenendir.

- 2AEKK eşanlı denklem sistemlerini kullanmasına bağlı olarak EKK'dan daha iyi sonuçlar ortaya koymuştur. Genel anlamda ise sonuçları EÇO tahmin edicisi ile mukayese edilecek derecede iyidir.
- Genel anlamda normallik varsayımı altında EÇO ve 2AEKK tahmin edicilerinin iyi sonuçlar verdiği ortadadır. Normallik varsayımı altında bu sonucun alınması beklenendir.
- Genel olarak logit, büyük örneklemlerde hem HKO hem de Yan değerlerine göre diğer tahmin edicilere göre kötü sonuçlar vermiştir. Buna karşın normal dağılıma dayalı Probit model, iyi performans ortaya koymuştur.
- 2AHeckit küçük örneklemlerde HKO değerlerine göre kötü sonuçlar ortaya koymaktadır.
- UEÇO tahmin edicisi artan örneklem sayısına bağlı olarak EKK ve logit'e göre iyi performans sergilese de 2AEKK, EÇO ve 2AHeckit tahmin sonuçlarının ardında kalmıştır.
- BM algoritmasına dayalı EÇO tahmini büyük iterasyon sayısı altında Newton metodunu dikkate alan EÇO tahmini ile çakışmaktadır. İterasyon sayısının azaltılması farklılaştırmayı bir miktar arttırsa da yakınsama mevcuttur. Analizlerin bu anlamda sağlaması yapılmıştır. Ayrıca işlem süresinin programda saydırılması ile BM algoritmasının Newton yöntemine göre daha hızlı sonuçlar verdiği gözlemlenmiştir. İşlem süresi bakımından BM algoritmasının daha etkin olduğu ifade edilebilir.
- *n* örneklem boyutu arttıkça tüm tahmincilerin HKO değeri azalmaktadır.

Hatanın mixture-normal dağılımdan gelmesi ve birincil dağılımın etkisinin %90 olması halinde 50, 100, 200, 400, 500 ve 800 örneklem hacmi için elde edilen sonuçlar Tablo 4.2'de sunulmuştur.

Tablo 4.2. *Hatanın mixture-normal dağılımdan (birincil %90) gelmesi halinde elde edilen Yan ve HKO değerleri*

$y = a + b*x$	Mixture-normal %90					
Eğim $b = 1$	n=50		n=100		n=200	
	Yan	HKO	Yan	HKO	Yan	HKO
EKK_g	0,14413	0,14372	0,13312	0,08348	0,13268	0,04923
EÇO_EM	0,00166	0,24500	-0,00989	0,12307	-0,00682	0,06188
2AHeckit	0,05997	6,02597	0,16604	4,75637	0,05404	2,12433
2AEKK	0,00307	0,24763	-0,01198	0,12432	-0,00532	0,06280
Probit	-1,74182	3,81412	-1,68231	3,17447	-1,66715	2,94140
Logit	-3,62486	15,5930	-3,50273	13,38520	-3,45659	12,4666
UEÇO	-0,00137	0,09237	-0,01641	0,05098	-0,00329	0,02363
	n=400		n=500		n=800	
	Yan	HKO	Yan	HKO	Yan	HKO
EKK_g	0,14567	0,03675	0,12608	0,02910	0,14426	0,02860
EÇO_EM	0,00060	0,02996	-0,01752	0,02635	-0,00268	0,01585
2AHeckit	0,11533	2,51199	0,13139	0,75754	0,00018	0,07529
2AEKK	0,00056	0,02977	-0,01610	0,02707	-0,00160	0,01619
Probit	-1,60943	2,66465	-1,65571	2,81126	-1,63478	2,70292
Logit	-3,34789	11,4438	-3,43021	11,9898	-3,38799	11,5763
UEÇO	-0,00876	0,00856	-0,02881	0,01075	-0,01012	0,00389

Tablo 4.2'ye göre,

- Tüm örneklemlerde UEÇO tahmin edicisi, HKO değerlerine göre en iyi performansı ortaya koymuştur.
- Yan değerleri dikkate alındığında ise küçük örneklemlerde yine UEÇO tahmin edicisi en iyi performansı ortaya koyarken büyük örneklemlerde 2AEKK ile EÇO tahmin edicisi sonuçlarını takip etmektedir.
- Tüm örneklemlerde HKO ve Yan değerlerine göre logit model en kötü performansı ortaya koymaktadır.
- EKK küçük örneklemlerde UEÇO tahmin edicisini takip etse de artan örneklem sayısına bağlı olarak artan sansürlü veri sebepli etkinlik kaybı ve yan ortaya koymaktadır.
- Artan örneklem sayısına paralel olarak EÇO ve 2AEKK sonuçları, UEÇO tahmin edicisi sonuçlarını takip etmektedir.
- BM algoritmasına dayalı EÇO tahmini büyük iterasyon sayısı altında Newton metodunu dikkate alan EÇO tahminiyle bu hata dağılımı varsayımı altında da

çakışmaktadır. Zaman açısından bu dağlım varsayımı altında da daha hızlı sonuçlar alınmıştır.

- *n* örneklem boyutu arttıkça tüm tahmincilerin HKO değeri burada da azalmaktadır.

Hatanın mixture-normal dağılımdan gelmesi ve birincil dağılımın etkisinin %80 olması halinde 50, 100, 200, 400, 500 ve 800 örneklem hacmi için elde edilen sonuçlar Tablo 4.3'de sunulmuştur.

Tablo 4.3. *Hatanın mixture-normal dağılımdan (birincil %80) gelmesi halinde elde edilen Yan ve HKO değerleri*

$y = a + b*x$	Mixture-normal %80					
Eğim $b = 1$	n=50		n=100		n=200	
	Yan	HKO	Yan	HKO	Yan	HKO
EKK_g	0,16188	0,25854	0,18901	0,14751	0,16915	0,10859
EÇO_EM	-0,01611	0,49154	0,01795	0,22026	0,00543	0,12168
2AHeckit	0,18785	7,68814	-0,04797	2,89766	0,08264	1,19266
2AEKK	-0,01312	0,50053	0,01734	0,22386	0,00539	0,12491
Probit	-1,29013	2,36246	-1,26654	1,91047	-1,26947	1,75295
Logit	-2,84307	10,2060	-2,77380	8,63088	-2,76437	8,07939
UEÇO	-0,00029	0,11792	0,00795	0,07628	-0,00535	0,03784
	n=400		n=500		n=800	
	Yan	HKO	Yan	HKO	Yan	HKO
EKK_g	0,18331	0,06932	0,15730	0,04608	0,17516	0,04208
EÇO_EM	0,01917	0,06777	-0,00035	0,04390	0,00794	0,02639
2AHeckit	0,13339	0,58227	0,01087	0,12262	-0,00615	0,06812
2AEKK	0,01975	0,06733	-0,00052	0,04583	0,00602	0,02665
Probit	-1,30394	1,76640	-1,29487	1,73035	-1,28922	1,70093
Logit	-2,81323	8,11626	-2,79727	7,98810	-2,78172	7,85570
UEÇO	-0,00674	0,01665	-0,00487	0,00959	-0,00480	0,00505

Tablo 4.3'e göre,

- Tüm örneklemlerde UEÇO tahmin edicisi, HKO ve Yan değerlerine göre en iyi performansı ortaya koymuştur (Yan için n=500 hariç).
- Genellikle tüm örneklemlerde HKO ve Yan değerlerine göre ise logit model en kötü performansı ortaya koymaktadır.

- HKO değerlerine göre EKK, küçük örneklemlerde UEÇO tahmin edicisini takip etsede artan örneklem sayısına bağlı olarak artan sansürlü veri sebepli etkinlik kaybı ve yan ortaya koymaktadır.
- Burada da artan örneklem sayısına paralel olarak EÇO ve 2AEKK sonuçları, UEÇO tahmin edicisi sonuçlarını takip etmektedir.
- BM algoritmasına dayalı EÇO tahmini büyük iterasyon sayısı altında Newton metodunu dikkate alan EÇO tahmini ile bu hata dağılımı varsayımı altında da çakışmaktadır ve geçen süre bakımından daha etkindir.
- n örneklem boyutu arttıkça tüm tahmincilerin HKO değeri burada da azalmaktadır.
- Ara sonuç olarak ifade edilebilir ki, normal-normal karışımına dayalı hata dağılımlarının değişen oranlarına rağmen sonuçlar paraleldir. UEÇO tahmin edicisi, diğer tahmin edicilere nispeten normal-normal karışımına dayalı hata dağılımlarında üstün performansa sahiptir.

Hatanın Student-t dağılımdan gelmesi halinde 50, 100, 200, 400, 500 ve 800 örneklem hacmi için elde edilen sonuçlar Tablo 4.4'de sunulmuştur.

Tablo 4.4. *Hatanın Student-t dağılımdan gelmesi halinde elde edilen Yan ve HKO değerleri*

$y = a + b * x$	Student-t					
Eğim $b = 1$	n=50		n=100		n=200	
	Yan	HKO	Yan	HKO	Yan	HKO
EKK_g	0,30345	0,45763	0,35217	0,30306	0,32369	0,19714
EÇO_EM	-0,03881	1,43762	0,02536	0,35605	0,00122	0,17710
2AHeckit	0,25264	5,94914	-0,20823	5,88085	0,17849	5,01388
2AEKK	-0,03747	1,30779	0,02077	0,36168	-0,00271	0,18237
Probit	0,04106	0,47348	0,08409	0,21609	0,10260	0,11388
Logit	-0,55770	1,59481	-0,47792	0,78367	-0,44371	0,47024
UEÇO	-0,14577	0,39731	-0,03531	0,25104	-0,00942	0,14137
	n=400		n=500		n=800	
	Yan	HKO	Yan	HKO	Yan	HKO
EKK_g	0,32746	0,15250	0,34208	0,15517	0,31888	0,12668
EÇO_EM	0,01192	0,09390	0,02863	0,06974	-0,00999	0,04534
2AHeckit	-0,03351	3,23238	0,01054	0,13890	0,05604	0,27296
2AEKK	0,00890	0,09432	0,02741	0,06990	-0,00953	0,04650
Probit	0,09036	0,06875	0,08887	0,04682	0,08464	0,03291
Logit	-0,46199	0,37305	-0,46345	0,31720	-0,46988	0,28837
UEÇO	0,02928	0,12610	0,08556	0,10578	0,06060	0,10010

Tablo 4.4'e göre,

- Probit model artan örneklem sayısına bağlı olarak büyük örneklemlerde en küçük HKO değerlerini vermiştir. UEÇO tahmin edicisi küçük örneklemlerde HKO değerine göre diğer tahmin edicilere göre iyi sonuç getirse de artan örneklem sayısına bağlı olarak bu hata dağılımı varsayımı altında yerini Probit modele bırakmıştır.
- Genellikle tüm örneklemlerde Yan değerine göre 2AEKK en iyi performansı göstermiştir.
- HKO değerleri dikkate alındığında 2AHeckit genellikle küçük örneklem sayılarında en yüksek değeri getirmiştir. Büyük örneklem sayılarında ise logit model kötü performans ortaya koymuştur. Bunun yanında Yan değerine göre yine logit tüm örneklem sayılarında en kötü performansı sergilemiştir.
- Student-t dağılımı varsayımı altında da n örneklem boyutu arttıkça EÇO tahmin değerleri ve BM algoritmasına dayalı EÇO tahmin değerleri birbirine yakınsamaktadır. İşlem süresi dikkate alındığında bu hata dağılımı varsayımı altında da BM algoritması daha etkindir.
- n örneklem boyutu arttıkça genellikle bir çok tahmin edicisinin HKO değeri azalmaktadır.

Hatanın Laplace dağılımından gelmesi halinde 50, 100, 200, 400, 500 ve 800 örneklem hacmi için elde edilen sonuçlar Tablo 4.5'de sunulmuştur.

Tablo 4.5. *Hatanın Laplace dağılımından gelmesi halinde elde edilen Yan ve HKO değerleri*

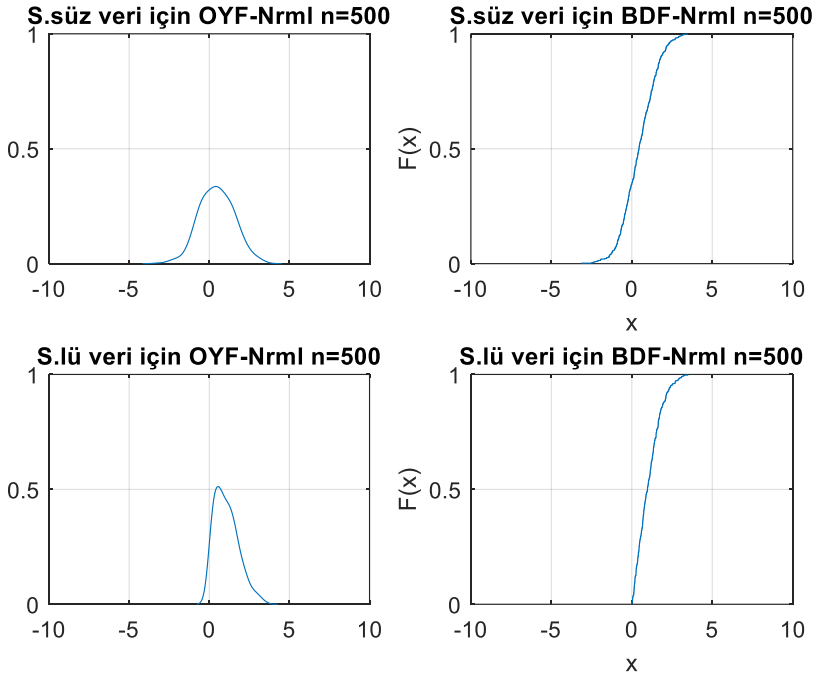
$y = a + b*x$	Laplace					
Eğim $b = 1$	n=50		n=100		n=200	
	Yan	HKO	Yan	HKO	Yan	HKO
EKK_g	0,20129	0,18012	0,06288	0,09163	0,24898	0,09727
EÇO_EM	-0,08025	0,23056	0,01201	0,10595	-0,01853	0,06920
2AHeckit	-0,05663	0,70509	-0,07501	0,50818	0,00402	0,02474
2AEKK	-0,07879	0,26136	0,01975	0,11689	-0,02473	0,06954
Probit	-0,58950	0,84549	-0,46925	0,44945	-0,49584	0,38111
Logit	-1,61139	4,06848	-1,39292	2,57993	-1,43012	2,42228
UEÇO	-0,14431	0,21221	-0,01745	0,07228	-0,00342	0,05224
	n=400		n=500		n=800	
	Yan	HKO	Yan	HKO	Yan	HKO
EKK_g	0,26469	0,08314	0,07379	0,02708	0,06838	0,01873
EÇO_EM	0,01019	0,02053	0,00844	0,02871	0,00791	0,01690
2AHeckit	-0,02488	0,03355	-0,01104	0,04438	-0,00309	0,06432
2AEKK	0,01054	0,02101	0,00576	0,03282	0,00997	0,01715
Probit	-0,41513	0,21053	-0,42294	0,23893	-0,46238	0,24678
Logit	-1,29263	1,77564	-1,30553	1,86827	-1,37009	1,96898
UEÇO	0,03325	0,02000	0,01700	0,01992	0,00311	0,00919

Tablo 4.5'e göre,

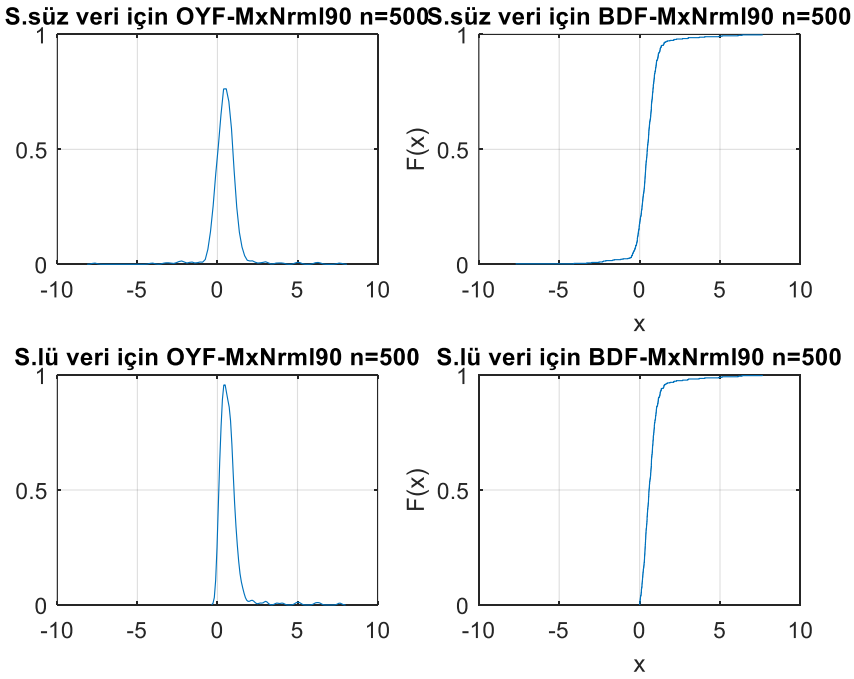
- Genellikle tüm örneklemlerde HKO ve Yan değerine göre UEÇO tahmin edicisi küçük değerler verirken bu tahmin ediciyi farklı örneklem sayıları için 2AEKK, EÇO ve 2AHeckit tahmin edicileri takip etmektedir.
- Genellikle tüm örneklemlerde Yan ve HKO değerleri incelendiğinde logit model en yüksek değeri getirmiştir. Bu anlamda logit modelin Laplace hata dağılımı varsayımı altında en düşük performanslı tahmin edici olduğu ifade edilebilir.
- BM algoritmasına dayalı EÇO tahmini büyük iterasyon sayısı altında Newton metodunu dikkate alan EÇO tahmini ile bu hata dağılımı varsayımı altında da çakışmaktadır. İşlem süresi dikkate alındığında bu hata dağılımı varsayımı altında da BM algoritması daha etkindir.
- n örneklem boyutu arttıkça tüm tahmincilerin HKO değeri genellikle azalmaktadır.

Yukarıda farklı dağılım varsayımları altında farklı örneklem sayılarına dair analizleri yapılan tablolar için, simülasyon sürecinde üretilen verilerin n=500 örneklem sayısı için grafikleri MatLab'tan elde edilerek Şekil 4.1-5'te sunulmuştur. Şekillerin

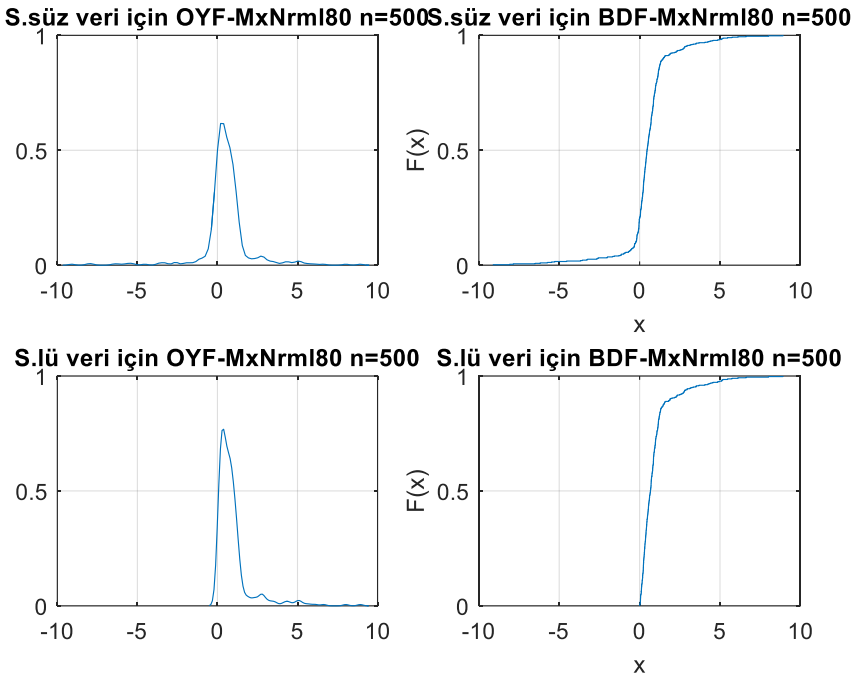
verilişinde yukarıda izlenen dağılım sırası gözetilmiştir. Söz konusu şekillerde S.lü ifadesi sansürlü; S.süz ifadesi sansürsüze karşı gelmektedir.



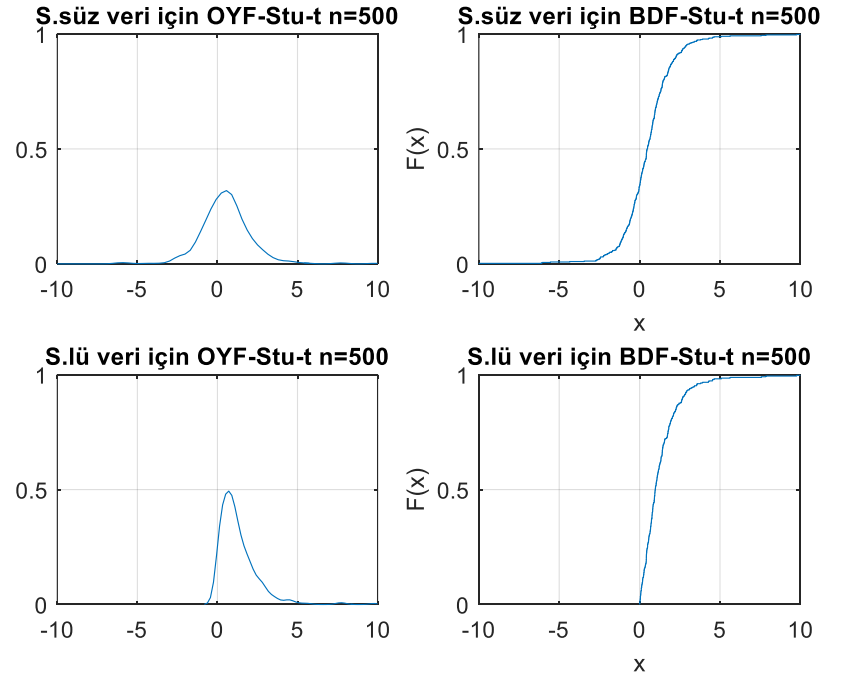
Şekil 4.1. *Hata dağılımının normal olması ve n=500 durumunda sansürlü ve sansürsüz veri için OYF ve BDF*



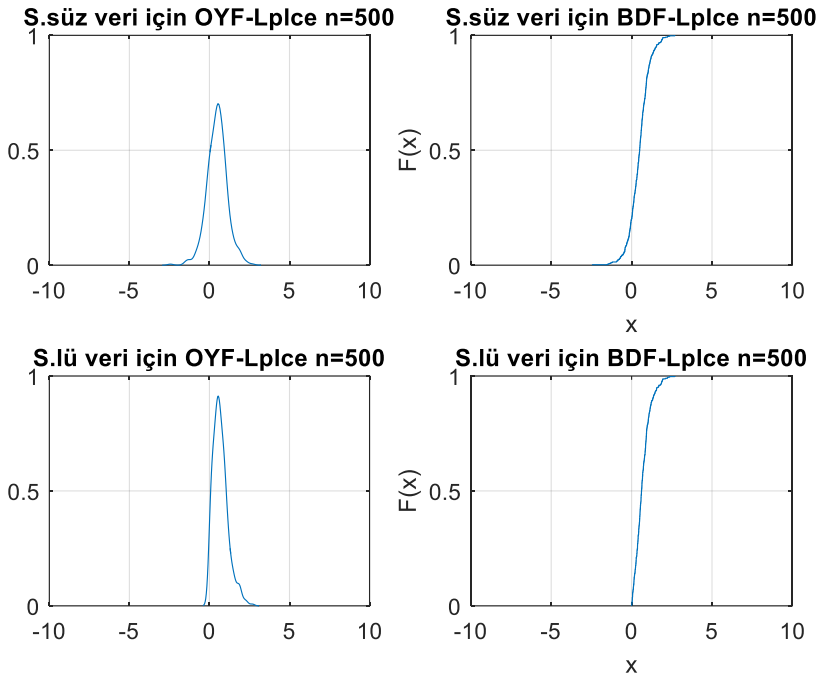
Şekil 4.2. *Hata dağılımının mixture normal (%90 birincil) olması ve n=500 durumunda sansürlü ve sansürsüz veri için OYF ve BDF*



Şekil 4.3. *Hata dağılımının mixture normal (%80 birincil) olması ve n=500 durumunda sansürlü ve sansürsüz veri için OYF ve BDF*



Şekil 4.4. *Hata dağılımının Student-t olması ve n=500 durumunda sansürlü ve sansürsüz veri için OYF ve BDF*



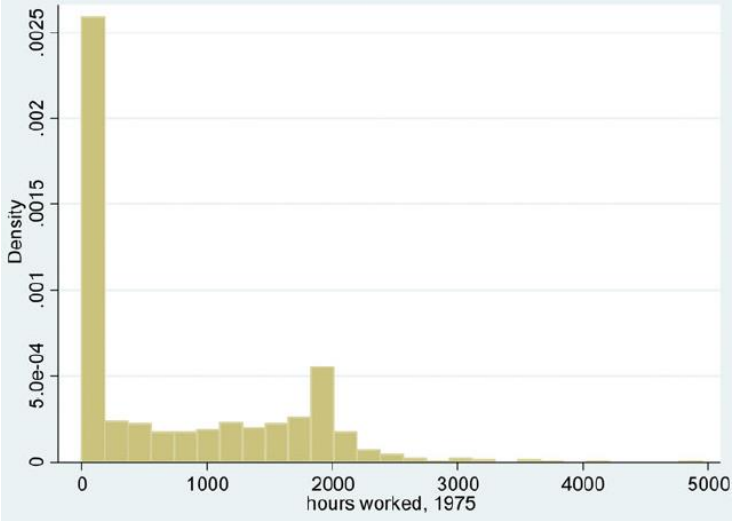
Şekil 4.5. *Hata dağılımının Laplace olması ve n=500 durumunda sansürlü ve sansürsüz veri için OYF ve BDF*

Şekiller yardımıyla simülasyon süresince farklı hata dağılımlarından üretilen verilerin sansür durumları ve buna bağlı olarak değişimler net şekilde görülmektedir. Özellikle hata dağılımının (artık dağılımının) sansürleme sonrası, beklenen değerinin sıfırdan pozitif tarafa doğru kaydığı görülmektedir.

4.2. Uygulama

Mroz (1987) tarafından evli kadınların çalışma saatleri üzerine bağımsız değişkenlerin etkileri literatürde tartışılmıştır. Araştırma kapsamında derlenen veri seti birçok araştırmacıya kaynaklık etmiştir. Bu çalışma kapsamında da Mroz verisi olarak da anılan bu veri setinden yararlanılmıştır. Simülasyonun yanında, çalışma kapsamında kullanılan tahmin edicilerin gerçek veri üzerinden incelenmesi uygulanabilirlik açısından gerekli görülmüştür. Kısaca Mroz verisini tanıtmak gerekirse; 753 evli kadına ait gözlem değerleri mevcuttur. Bu kadınların 428'i ev dışında bir ücret karşılığı çalışırken kalan 325'i sıfır saat çalışıyor olarak kaydedilmiştir. Verilerin yaklaşık olarak %43'ünün sansürlü olduğu söylenebilir. Evli kadınların çalışma saatlerini etkileyen pek çok değişken olmasına rağmen, tezin tüm kurgusu basit doğrusal regresyon modeli olduğu

için bu açıklayıcı değişkenlerden yalnızca biri ele alınmıştır. Uygulamada, bağımlı değişken ev hanımlarının yıllık çalışma saatleri, bağımsız değişken ise kadının yaşı olarak kabul edilmiştir. Bağımlı değişkenin değerlerinin histogramı aşağıda sunulmuştur.



Şekil 4.6. *Mroz verisi: Çalışılan süre*

Kaynak: *McDonald, Nguyen, 2015, s. 2155*

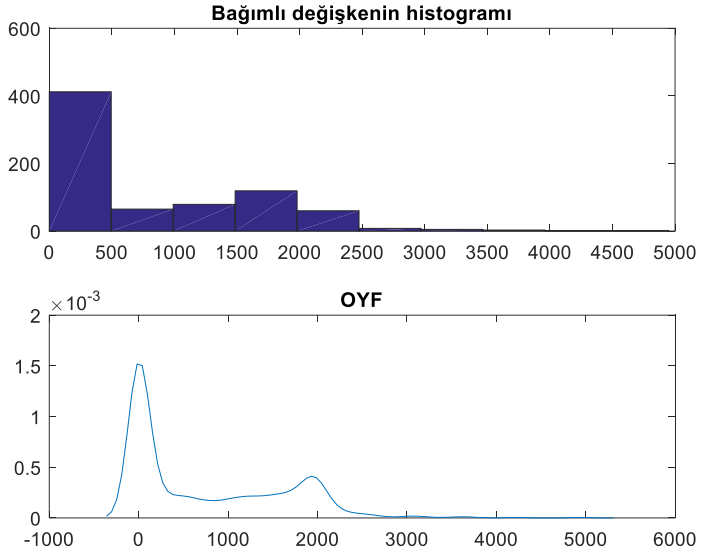
Bu çalışma kapsamında basit doğrusal regresyon için tahmin ediciler ve performanslarının ele alındığı düşünülürse, tahmin edilecek model aşağıdaki gibidir.

$\hat{y}$: Çalışma süresi,

x_1 : Yaş olmak üzere,

$$\hat{y} = \hat{\beta}_1 + \hat{\beta}_2 x_1$$

Bu bağlamda modelde kullanılan bağımlı değişkenin histogramı ve deneysel OYF Şekil 4.7'de verilmiştir.



Şekil 4.7. *Bağımlı değişkenin histogramı*

Simülasyon çalışmasında, EÇO ve UEÇO tahmin edicilerinin performanslarının genel olarak diğer tahmin edicilere göre üstün olduğu gözlemlenmiştir. Çalışmada EÇO tahminedicisi için UEÇO tahmin edicisinin alternatif olarak önerildiği de dikkate alınarak bu iki tahmin edicinin sonuçları Tablo 4.7 de sunulmuştur.

Tablo 4.6. *EÇO ve UEÇO tahmin sonuçları*

	Tahminler	
Tahminciler	$\hat{\beta}_1$	$\hat{\beta}_2$
EÇO_EM	734,2941	-9,87072
UEÇO	763,1398	-10,7933

Tabloda görüldüğü üzere UEÇO tahmin edicisi, tobit model tahmini için alternatif olarak kullanılabilir bir tahmin edicidir. Gerçek veri uygulamalarında en iyi tahmin ediciyi tüm veri setleri için genellemek mümkün olmayacaktır. Değişen veri doğasına göre tahmin edicilerin performansları değişim gösterebilir. Ancak bu çalışma kapsamında sansürlü veri durumunda sıklıkla kullanılan tobit model için önerilen alternatif tahmin edicinin farklı hata dağılımları ve farklı örneklem büyüklüklerindeki kayda değer performansı, literatürde sıklıkla kullanılan Mroz verisi uygulamasıyla pekişmiştir.

SONUÇ VE ÖNERİLER

Sınırlı bağımlı değişkenler; sansürlenmiş, kırpılmış ve ayrık sonuçlar içeren değişkenleri içerir. Bu değişkenlere bağlı olarak; kırpılmış regresyon modelleri, sansürlenmiş regresyon modelleri, kukla endojen modelleri tanımlanabilir. Bu çalışma kapsamında, söz konusu değişkenler ve modeller ele alınmıştır.

Sırasıyla, doğrusal olasılık modeli ve bu modele alternatif probit ve logit model; kırpılmış ve sansürlenmiş dağılım ile bu sınırlılıklar için regresyon bilgileri paylaşılmıştır. Birbirleriyle ilişkili ve birbirlerini tamamlayan bu modeller birikimli olarak, literatür sırası dikkate alınarak ayrıntılandırılmıştır. Sansürlü regresyonun odağa alınması ile bu genel çerçeveden özele geçilmiştir. Sürekli bağımlı değişkenin belli aralıklarla sınırlandırıldığı durum, yani sansürlü veri durumu ve bu durum için geliştirilmiş sansürlü regresyon modeli tobit tip I (normallik varsayımı altında sansürlü regresyon) ayrıntılı olarak incelenerek, modelin EÇO tahmini ile alternatif tahmin edicileri performansları bakımından incelenmiştir.

Bilinmektedir ki sansürlü regresyon modeli için sıradan EKK tahmin edicileri yanlı ve tutarsız sonuçlar vermektedir. Tobit model veya sansürlenmiş normal regresyon modeli bu aşamada çözüm olarak önerilen bir modeldir. Ancak tobit model normallik varsayımına dayalıdır ve hata terimlerinin normal dağılmaması halinde EÇO tahminleri tutarsız sonuçlar vermektedir. Normallik varsayımının esnetilebilmesi için literatürde çeşitli tahmin ediciler önerilmiştir. Kısmı uyarlamalı tahmin ediciler, normallik varsayımının ihlali halinde önerilen tahmin ediciler arasındadır. Ayrıca modeli yansızlaştırmaya yönelik önerilen 2AHeckit ve 2AEKK tahmin edicisi tobit modelin alternatif tahmin edicileri olarak incelenmiştir.

Hatanın normallikten ayrılışına karşı, UEÇO tahmin edicisi çalışma kapsamında önerilmiştir. GND şekil parametresine bağlı olarak normal dağılımı ve Laplace dağılımını içermektedir. Dolayısıyla özel durumda EÇO tahmin edicisini kapsadığı söylenebilir. Ayrıca farklı hata dağılımları için ele alınan tahmin edicilerin göreli performansları simülasyon çalışması yardımıyla değerlendirilmiştir. Tüm tez ve simülasyon sonuçları dikkate alındığında elde edilen sonuçlar aşağıdaki şekilde özetlenebilir:

- EÇO tahmin edicisi UEÇO tahmin edicisinin özel bir halidir.
- EÇO tahmin edicisine dayalı tahminlerin elde edilmesinde BM algoritması kullanılabilir.

- Tobit modelin parametrelerinin hesaplanması BM algoritması ile daha kolay ve hızlıdır.
- Hata dağılımının normal olduğu durumlarda EÇO tahmin edicisi beklendiği gibi diğer tahmin edicilerden üstün gelirken, UEÇO tahmin edicisi oldukça küçük bir etkinlik kaybı ile EÇO tahmin edicisine yakın sonuçlar vermiştir.
- Hata dağılımının normal olmadığı zamanlarda UEÇO tahmin edicisi, EÇO tahmin edicisi dahil olmak üzere diğer tahmin edicilere karşı ciddi bir üstünlüğe sahiptir. Bu durum GND'ın esnekliğine bağlı bir sonuçtur.
- Logit model genel anlamda farklı hata dağılımlarında en kötü performansı ortaya koyan tahmin edicidir.
- Probit model hataların Student-t dağılımdan gelmesi dışında kayda değer bir performans ortaya koyamamıştır.
- 2AEKK ve 2AHeckit, iki aşamalı bir tahmin sürecini işleterek EÇO tahmin edici ile mukayese edilebilir performanslar ortaya koymuşlardır. Ancak 2AHeckit, tahmin ediciler arasında zaman zaman en kötü sonuçları vermiştir. 2AEKK ve 2AHeckit kendi içinde karşılaştırılacak olursa, 2AEKK tahmin edicisi 2AHeckit'e göre daha iyi bir performans sergilemiştir.
- n örneklem boyutu arttıkça tüm tahmincilerin HKO değeri, incelenen farklı hata dağılımları için genellikle azalmaktadır. Bu anlamda yapılan analizlerin tutarlı olduğu ifade edilebilir.
- Araştırma kapsamında işletilen ve yararlılığı görülen durum farklı esnek dağılımlara dayalı tahmin edicilerin geliştirilmesi ile sansürlü regresyona uyarlanabilir.
- Bu yapı diğer sınırlı bağımlı değişkenli modellere de uyarlanabilir.

Bu tezdeki çalışmalar; sansürlü regresyonda sağdan ve aralıklı sansürleme durumları, dışarıdan belirlenebilir farklı sansürleme seviyeleri, farklı hata dağılımları, değişen varyans probleminin çözümüne yönelik incelemeler ve sansürlü regresyonsa robust tahmin edicileriyle devam edecektir.

KAYNAKÇA

Aktürk, Z. ve Acemoğlu, H. (2011). *Sağlık Çalışanları İçin Araştırma ve Pratik İstatistik (2)*. İstanbul: Anadolu Matbaası.

Aldrich and Nelson. (1984). Linear Probability, Logit, and Probit Models, 45, 15.

Amemiya, T. (1973). Regression Analysis when the Dependent Variable Is Truncated Normal. *Econometrica*, 41 (6), 997–1016.

Amemiya, T. (1975). Qualitative Models. *Annals of Economic and Social Measurament*, 4, 363-72.

Amemiya, T. (1981). Qualitative Response Models: A Survey. *Journal of Economics Literature*, 19 (4), 483-536.

Amemiya, T. (1985). *Advanced Econometrics*. Oxford: Basil Blackwell.

Amore, P. (2005). Asymptotic and Exact Series Representations for The Incomplete Gamma Function. *EPL (Europhysics Letters)* 71(1), 1-8.

Arıcan, E. (2010). *Hanehalkı Harcamaları Olasılıklarını Sıralı Regresyon Modeli İle Tahmin Etme*. TUİK Uzmanlık Tezi, Adana: T.C. Başbakanlık Türkiye İstatistik Kurumu.

Balcı, M. (2008). *Genel Matematik*. 5. baskı. Ankara: Balcı Yayınları.

Baltagi, Badi. (2001). *Econometric Analysis of Panel Data (2)*. NewYork: JohnWiley&Sons Ltd.

Bartolucci, F. and Scaccia, L. (2004). The Use of Mixtures for Dealing with Non-normal Regression Errors. Submitted to Computational Statistics and Data Analysis.

Breen, R. (1996). Regression Models: Censored, Sample Selected or Truncated Data. *Quantitative Applications in the Social Sciences*, 7-111.

Cafri, R. (2009). *Adana ilinde yoksulluğun analizi: Sınırlı bağımlı değişkenli modellerle bir inceleme.* Yüksek Lisans Tezi. Adana: Çukurova Üniversitesi Sosyal Bilimler Enstitüsü.

Caudill, S. B. (2012). A Partially Adaptive Estimator for The Censored Regression Model Based on A Mixture of Normal Distributions. *Stats Methods Appl*, 21:121-137

Carson, R.T. and Sun Y. (2007). The Tobit Model with A Non-Zero Threshold. *Econometrics Journal,*10: 488-502.

Chay, K.Y. and Powell, J. L. (2001). Semiparametric Censored Regression Models. *Journal of Economic Perspectives*, 15 (4).

Cohen, A.C. (1991). *Truncated and Censored Samples: Theory and Applications.* NY: Taylor & Francis Group

Cox, D.R. (1958). The Regression Analysis of Binary Sequences (with discussion). *J Roy Stat Soc B*, 20: 215–242.

Cox, D.R. (1970). *Analysis of Binary Data*. Londra: Methuen.

Davidson, R. and MacKinnon J.G. (1999). *Econometric Theory*. USA: Oxford University Press.

Dempster, A. P., Laird N. M. and Rubin D. B. (1977). Maximum Likelihood Estimation from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society, Series B*, 39, 1-38.

Do, M.N., Vetterli, M. (2002). Wavelet-based Texture Retrieval Using Generalised Gaussian Density and Kullback-Leibler Distance, Transaction on Image Processing. 11, 146–158.

Eren. (2012*). Sınırlı bağımlı değişkenli modeller ve ülkelerin gelişmişlik düzeyleri üzerine uygulama*. Yüksek Lisans Tezi. Erzurum: Atatürk Üniversitesi Sosyal Bilimler Enstitüsü.

Gezer, E. (2015). *Türkiye'de kamu kaynaklı sosyal yardımlar: Sansürlü regresyon analizi.* Yüksek Lisans Tezi. İzmir: Dokuz Eylül Üniversitesi Sosyal Bilimler Enstitüsü.

Goldberger. (1964). Econometric Theory. New York: Wiley.

Goldberger. (1980). *Abnormal Selection Bias*. Wisconsin: University of Wisconsin.

Greene, H.W. (2003). Econometric Analysis. *Prentise Hall*, (5), 1026.

Greene, H.W. (2011). Econometric Analysis. *Prentise Hall*, (7), 1231.

Gujarati, D.N. (1999). Essentials of Econometrics. Second Edition. *McGraw-Hill.*

Gujarati, D. N (2004). Basic Econometrics. *The McGraw-Hill Companies*.

Gujarati, D. N. and Porter, D. C. (2012), Temel Ekonometri, (Çev., Ü. Şenesen ve G.G. Şenesen), *Literatür Yayıncılık*, 554.

Güriş, B. (2005). *Ülkelerin kalkınmışlık düzeylerini etkileyen faktörlerin çok seçenekli tercih modelleri ile incelenmesi.* Yayımlanmış Yüksek Lisans Tezi. İstanbul: İstanbul Üniversitesi Sosyal Bilimler Enstitüsü.

Hausman, J. A. and Wise, D. A. (1976). The Evaluation of Result from Truncated Samples: The New Jersey Negative Income Tax Experiment. *Annals of Economic and Social Measurement*, 5, 421-45.

Hausman, J. A. and Wise, D. A. (1977). Social Experimentation, Truncated Distribution and Efficent Estimation. *Econometrica*, 45, 919-39.

Imbens, G. and Angrist, J. (1994). Identification and Estimation of Local Average Treatment Effects. *Econometrica*, 62 (2), 467–476.

Kmenta, J. (1990). Elements of Econometrics. *McMillan Publishing*, 2.

Koç Ş. (2013). *Tobit regresyon analizi ve bir uygulama*. Yüksek Lisans Tezi. Kahramanmaraş: Kahramanmaraş Üniversitesi, Fen Bilimleri Enstitüsü.

Lee, L. F. (1982). Some Approaches to the Correction of Selectivity Bias. *Review of Economic Studies*, 49, 355-372.

Lee, L. F. and Trost R. P. (1978). Estimation of Some Limited Dependent Variable Models with Applications to Housing Demand. *Journal of Econometrics*, 8, 357.

Lewis, R. A. and McDonald, J. B. (2014). Partially Adaptive Estimation of the Censored Regression Model. *Economic Reviews*, 33 (7), 732-750.

Long, J. S. (1997). *Regression Models for Categorical and Limited Dependent Variables*.Thousand Oaks, London: Sage Publications, International Educational and Professional Publisher .

Maddala, G.S. (1983), *Limited-Dependent and Qualitative Variables in Econometrics*. Cambridge, UK: Cambridge University Press.

McDonald and Moffitt. (1980). *The Review of Economics and Statistics*, 62 (2), 318-321.

McDonald, J. B. and Newey, W.K. (1988). Partially Adaptive Estimation of Regression Models via the Generalized T Distribution. *Econometric Theory*, 4, 428-457.

McDonald, J. B., Xu Y. J. (1996). A Comparison of Semi-parametric and Partially Adaptive Estimators of the Censored Regression Model with Possibly Skewed and Leptokurtic Error Distributions. *Economics Letter*, 51(2), 153-159.

McDonald, J. B. and Nguyen H. (2015). Heteroscedasticity and Distributional Assumptions in The Censored Regression Model. *Communications in Statistics-Simülation and Computions*, 44, 2151-2168.

McGillavray, (1970). Journal of Transport Economics and Policy.

Melino, A. (1982). Testing for Sample Selection Bias. *Review of Economic Studies*, 49, 151-153.

Mroz, T. (1987). The sensitivity of an empirical model ofmarried women's hours of work to economic and statistical assumptions. *Econometrica* 55:765–799.

Nadarajah, S. (2005). A Generalized Normal Distribution. *Journal of Applied Statistics*, 32 (7), 685–694.

Olsen, R.J. (1978). Note on the Uniqueness of the Maximum Likelihood Estimator for the Tobit Model. *Econometrica*, 46 (5), 1211-1215.

Orme, Chris D. and Ruud, Paul A., (2002). On The Uniqueness of The Maximum Likelihood Estimator. *Economics Letters, Elsevier*, vol. 75(2), p209-217.

Özdamar, (2011). *Paket Programlar ile İstatistiksel Veri Analizi*. Eskişehir: Kaan Yayınevi.

Paarsch, H. (1984). A Monte Carlo Comparison of Estimators for Censored Regression Models. *Journal of Econometrics*, 24, 197-213.

Pagan, A. and Ullah A. (1999). Nonparametric ecometrics, Cambridge, UK: Cambridge University Press.

Pampel, F.C. (2000). *Logistic Regression – A Primer*. California : Sage Publications Inc.

Park, S.Y. (2003). Unbiasedness or Statistical Efficiency: Comparison between One-stage Tobit of MLE and Two-step Tobit of OLS. *International Journal of Human Ecology*, 4 (2), 77-86.

Phillips, R. F. (1994). Partially Adaptive Estimation via a Normal Mixture. *Journal of Econometrics*, 64, 123-144.

Piegorsch, W. (1992). Complementary Log Regression for Generalized Linear Models. *American Statistician*, 46 (2), 94-99.

Pindyck R.S. and Rubinfield D.L. (2009). Microeconomics. *Pearson/Prentice Hall*, ISBN: 01 320 80230, 736.

Puhani, P. A. (2000). The Heckman Correction for Sample Selection and Its Critique, Journal of Econometric Surveys, 14 (1), 55.

Ramanathan, R. (2002). *Introductoy Econometrics with Applications*. Texas: Harcount Collage Publishers.

Rosen, H. S. (1979). Housing Decisions and the U.S. Income Tax: An Econometric Analysis. *Journal of Public Economics*, 11, 1-23.

Sigelman, L. And Zeng, L. (1999). Analyzing Censored and Sample-Selected Data with Tobit and Heckit Models. The George Washington University,December 16,WV002-05.

Tatoğlu Yerdelen, F. (2005). *Sermaye Piyasası'nda Riskin Sınırlı Bağımlı Değişkenli Panel Veri Modelleri ile Analizi*. Yayımlanmış Doktora Tezi. İstanbul: İstanbul Üniversitesi Sosyal Bilimler Enstitüsü.

Theil, H. (1970). On the Estimation of Relationships Involving Qualitative Variables. *American Journal of Sociology*, 76, 103-154.

Thomas, Richard L. (1997). *Modern Econometrics: An Introduction*. England: Addison Wesley.

Tobin, J. (1958). Estimation of Relationships for Limited Dependent Variables. *Econometrica*, 26, 24-36.

Trost, R. P. (1977). *Demand for Housing: A Model Based on Inter-related Choices Between Owning and Renting*. Unpublished Ph. D. Dissertation. University of Florida.

Varanasi, M.K. and Aazhang, B. (1989). Parametric Generalized Gaussian Density Estimation. *Journal of the Acoustical Society of America*, 86 (4), 1404–1415.

White, H. (1980). A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity. *Econometrica*, 48 (4), 817-828.

Wooldridge, J. (2002). *Econometric Analysis of Cross Section and Panel Data*, Cambridge: MIT Press.

Wu, X. and Stengos, T. (2005). Partially Adaptive Estimation via the Maximum Entropy Densities. *Econometrics Journal*, 9, 1-15.

Yazıcı, F. (2005). *EM Algortiması ve Uzantıları*, Yüksek Lisans Tezi. Ankara: Hacettepe Üniverisitesi FBE.

Yenilmez, I. ve Kantar, Y.M. (2016). Tobit Regression For Modeling Gastric Cancer. *International Congress on Fundamental and Applied Sciences ICFAS2016*, Istanbul, Yildiz Technical University, s.107

Yılmaz A., Karasoy D. ve Eroğlu A. (2013). Yaşam Çözümlemesinde Cox Orantılı Tehlikeler ve Orantılı Odds Modelleri. *İstatistikçiler Dergisi: İstatistik&Aktüerya*, 6, 70-78.

Zorlutuna Ş., Erilli N.A., Yücel B. (2016). Lung Cancer Study with Tobit Regression Analysis: Sivas Case. *Eurasian Eononometrics, Statistics and Emprical Economics Journal*, 3(3),13-22.

ÖZGEÇMİŞ

Adı-Soyadı : İsmail YENİLMEZ

Yabancı Dil : İngilizce

Doğum Yeri ve Yılı : Eskişehir/1989

E-Posta : ismailyenilmez@anadolu.edu.tr

Eğitim ve Mesleki Geçmişi:

2012, Marmara Üniversitesi, Ortaöğretim Fen ve Matematik Alanları Eğitimi, Matematik Öğretmenliği

2012, Marmara Üniversitesi, Ortaöğretim Fen ve Matematik Alanları Eğitimi, Matematik Öğretmenliği Tezsiz YL

2016, Marmara Üniversitesi, Ortaöğretim Fen ve Matematik Alanları Eğitimi, Matematik Öğretmenliği Tezli YL

Yayınları ve Bilimsel/Sanatsal Faaliyetleri:

Uluslararası hakemli dergilerde yayınlanan makaleler;

- Yeliz Mert KANTAR, Ilhan USTA, Ismail YENILMEZ, Ibrahim ARIK, A Study on Estimation of Wind Speed Distribution by Using the Modified Weibull Distribution. International Journal of Informatics Technologies. Vol 9, No 2 (2016). DOI: 10.17671/btd.49478. ISSN: 1307-9697
- Yeliz MERT KANTAR, Ibrahim ARIK, Ilhan USTA, Ismail YENILMEZ, Comparison of Some Estimation Methods of the two parameter Weibull Distribution for Unusual Wind Speed Data Cases. International Journal of Informatics Technologies. Vol 9, No 2 (2016). DOI: 10.17671/btd.61183. ISSN: 1307-9697
- Ibrahim ARIK, Yeliz Mert KANTAR, Ismail YENILMEZ, The Evaluation of Robust and Efficient Estimators for Log-Logistic Distribution for Censored Data with/without Outliers, Journal of Scientific Research and Development 2 (12): 24-32, 2015 Available online at www.jsrad.org ISSN 1115-7569 © 2015 JSRAD

Uluslararası bilimsel toplantılarda sunulan ve bildiri kitabında (Proceeding) basılan bildiriler;

- MERT KANTAR YELIZ, USTA ILHAN, ARIK IBRAHIM, YENILMEZ ISMAIL (2016). Distributions of Wind Speed at Different Heights. 2016 International Conference on Engineering & MIS (ICEMIS'16), Doi: 978-1-5090-5579-1/16 (Tam metin bildiri)
- USTA ILHAN, MERT KANTAR YELIZ, ARIK IBRAHIM, YENILMEZ ISMAIL (2016). A Statistical Investigation on Wind Energy Potential of Northwest of Turkey. 2016 International Conference on Engineering & MIS (ICEMIS'16), Doi: 978-1-5090-5579-1/16 (Tam metin bildiri)
- USTA ILHAN, MERT KANTAR YELIZ, ARIK IBRAHIM, YENILMEZ ISMAIL (2016). The Generalized Lindley Distribution to Model Wind Speed. 4. European Conference on Renewable Energy Systems 4. European Conference on Renewable Energy Systems (ECRES'16) (Tam metin bildiri)
- USTA ILHAN, MERT KANTAR YELIZ, ARIK IBRAHIM, YENILMEZ ISMAIL (2016). A New Estimator Based on Median and Mode for Weibull Distribution in Wind Energy. 4. European Conference on Renewable Energy Systems (ECRES'16) (Tam metin bildiri)
- YENILMEZ ISMAIL, MERT KANTAR YELIZ (2016). Tobit Regression For Modeling Gastric Cancer. International Congress on Fundamental and Applied Sciences (ICFAS'16) (Özet bildiri)
- YENILMEZ ISMAIL, YAVUZ ILYAS (2016). Usıng Technology in Statistical Education at High School Level: VUstat Software and a Case of Eskişehir Tepebaşı. 7th International Congress on New Trends in Education, 55-55. (Özet bildiri)
- ARIK IBRAHIM, MERT KANTAR YELIZ, USTA ILHAN, YENILMEZ ISMAIL (2015). Evaluation of Robust Estimation Methods in Estimating Weibull Parameters for Wind Energy Application. Proceedings of the The International Conference on Engineering & MIS 2015 - ICEMIS'15, Doi: 10.1145/2832987.2833041 (Tam metin bildiri)
- YENILMEZ ISMAIL, MERT KANTAR YELIZ, USTA ILHAN,ARIK IBRAHIM (2015). Analysis of the Modified Weibull Distribution for Estimation of Wind Speed Distribution. Proceedings of the International Conference on Engineering & MIS 2015 - ICEMIS'15, Doi: 10.1145/2832987. 2833059 (Tam metin bildiri)

Projeler;

- 1506F532 nolu BAP projesi, "Rüzgar Hızını Modellemede Önerilen İstatistiksel Dağılım Aileleri: Türkiye'nin Farklı Bölgelerinin Rüzgâr Gücü Potansiyelinin Önerilen Dağılımlarla Araştırılması", Araştırmacı, Eylül 2015- Eylül 2016.

Ödülleri:

- Hüsnü M. Özyeğin Vakfı Bursu, Türkiye İlk 1000 Bursu (5 Yıl Süreyle); 2007 ÖSS SAY-1 TR 742.'si SAY-2 TR 766.'sı