



T.C.
NECMETTİN ERBAKAN ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ



**TOBİT MODELDE BAZI YANLI TAHMİN EDİCİLERİN
PERFORMANSLARININ İNCELENMESİ**

Esra ÖĞÜTCÜOĞLU

YÜKSEK LİSANS TEZİ
Matematik Anabilim Dalı

Haziran - 2020
KONYA
Her Hakkı Saklıdır

TEZ KABUL VE ONAYI

Esra ÖĞÜTCÜOĞLU tarafından hazırlanan "*TOBIT MODELDE BAZI YANLI TAHMİN EDİCİLERİN PERFORMANSLARININ İNCELENMESİ*" adlı tez çalışması 22/06/2020 tarihinde aşağıdaki jüri tarafından oy birliği ile Necmettin Erbakan Üniversitesi Fen Bilimleri Enstitüsü Matematik Anabilim Dalı'nda YÜKSEK LİSANS TEZİ olarak kabul edilmiştir.

Jüri Üyeleri

Başkan

Prof. Dr. Murat ERİŞOĞLU

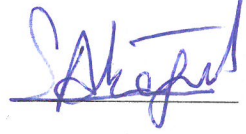
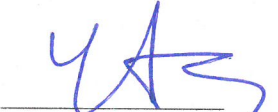
Danışman

Doç. Dr. Yasin ASAR

Üye

Dr. Öğr. Üyesi Serkan AKOĞUL

İmza



Fen Bilimleri Enstitüsü Yönetim Kurulu'nun .../.../2020 gün ve sayılı kararıyla onaylanmıştır.

Prof. Dr. S. Savaş DURDURAN
Enstitü Müdürü

TEZ BİLDİRİMİ

Bu tezdeki bütün bilgilerin etik davranış ve akademik kurallar çerçevesinde elde edildiğini ve tez yazım kurallarına uygun olarak hazırlanan bu çalışmada bana ait olmayan her türlü ifade ve bilginin kaynağına eksiksiz atıf yapıldığını bildiririm.

DECLARATION PAGE

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Esra ÖĞÜTCÜOĞLU
Tarih: 22/06/2020

ÖZET

YÜKSEK LİSANS TEZİ

TOBİT MODELDE BAZI YANLI TAHMİN EDİCİLERİN PERFORMANSLARININ İNCELENMESİ

Esra ÖĞÜTCÜOĞLU

Necmettin Erbakan Üniversitesi Fen Bilimleri Enstitüsü
Matematik Anabilim Dalı

Danışman: Doç. Dr. Yasin ASAR

2020, 53 Sayfa

Jüri

Doç. Dr. Yasin ASAR

Prof. Dr. Murat ERİŞOĞLU

Dr. Öğr. Üyesi Serkan AKOĞUL

Çoklu bağlantı probleminin en çok olabilirlik tahmin edicisi üzerindeki etkileri, tobit regresyon modelinde analiz edilmiştir. Çoklu bağlantı problemi yansız olan en çok olabilirlik tahmin edicisinin varyansını şişirir. Böylece tahminler tutarsız hale gelir. Khalaf ve ark. (2014) tarafından ridge regresyonu tahmin edicisinin kullanılması çoklu bağlantı probleminin çözümü olarak önerilmiştir. Bu çalışmada en çok olabilirlik tahmin edicisi ile ridge tahmin edicisinin hata kareler ortalaması özellikleri teorik olarak incelenmiş ve birbiriyle karşılaştırılmıştır. Tahmin edicilerin performanslarını değerlendirmek için bir Monte Carlo simülasyon çalışması tasarlanmıştır.

Çoklu bağlantı problemine başka bir alternatif olarak da Liu tahmin edicisinin geliştirilmesi olan yeni bir yanlı tahmin edici tanıtılmıştır. Tahmin edicilerin hata kareler ortalaması özellikleri teorik olarak incelenmiştir. Tahmin edicilerin performanslarını değerlendirmek için bir Monte Carlo simülasyon çalışması tasarlanmış ve bir performans kriteri olarak simüle edilmiş hata kareler ortalaması kullanılmıştır. Son olarak, yeni tahmin edicinin faydaları gerçek veri uygulamaları ile gösterilmiştir.

Anahtar Kelimeler: Çoklu bağlantı problemi, Liu regresyon, Ridge regresyon, Tobit model

ABSTRACT

MS THESIS

INVESTIGATING PERFORMANCES OF SOME BIASED ESTIMATORS IN THE TOBIT MODEL

Esra ÖĞÜTCÜOĞLU

**THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCE OF
NECMETTİN ERBAKAN UNIVERSITY
THE DEGREE OF MASTER OF SCIENCE IN MATHEMATICS**

Advisor: Assoc. Prof. Dr. Yasin ASAR

2020, 53 Pages

Jury

Assoc. Prof. Dr. Yasin ASAR

Prof. Dr. Murat ERİŞOĞLU

Assist. Prof. Dr. Serkan AKOĞUL

The effects of the multicollinearity problem on the maximum likelihood estimator are analyzed in the tobit regression model. The multicollinearity problem inflates the variance of the maximum likelihood estimator that is asymptotically unbiased. Thus the maximum likelihood estimates become inconsistent. The use of ridge regression estimator by Khalaf ve ark. (2014) has been proposed as a solution to the multicollinearity problem. In this thesis, the mean squared error properties of ridge estimators and likelihood estimators are analyzed and compared theoretically. A Monte Carlo simulation study is designed to evaluate their performance.

As another alternative to the multicollinearity problem, a new biased estimator is introduced, which is the generalization of the well-known Liu estimator. Mean squared error properties of the estimators are investigated theoretically. In order to evaluate the performances of the estimators, a Monte Carlo simulation study is designed and simulated mean squared error is used as a performance criterion. Finally, the benefits of the new estimator is illustrated via real data applications.

Anahtar Kelimeler: Multicollinearity problem, Liu regression, Ridge regression, Tobit model

ÖNSÖZ

Çalışmanın başından itibaren değerli yardım ve desteklerini esirgemeyen, her zaman bilgi birikimini paylaşmaktan sakınmayan danışman hocam Doç. Dr. Yasin ASAR'a teşekkür ederim.

Yüksek lisans eğitim süresinde maddi ve manevi destek sağlayan sevgili karedeşim Ahmet ÖĞÜTCÜOĞLU'na, aldığım her kararın arkasında duran ve kendileri ile gurur duyduğum aileme teşekkürlerimi sunarım.

Esra ÖĞÜTCÜOĞLU
KONYA-2020

İÇİNDEKİLER

ÖZET	iv
ABSTRACT	v
ÖNSÖZ	vi
İÇİNDEKİLER	vii
ŞEKİLLER LİSTESİ	ix
TABLOLAR LİSTESİ	x
SİMGELER VE KISALTMALAR	xii
1. GİRİŞ	1
2. KAYNAK ARAŞTIRMASI	5
3. TOBİT REGRESYON MODELİ	8
3.1. Sansürlü ve Kesilmiş Dağılımlar	9
3.1.1. Sansürlü Normal Dağılım	9
3.1.2. Kesilmiş Normal Dağılım	10
3.2. Tobit Regresyon Modeli	11
3.3. Çoklu Bağlantı Problemi	13
3.4. Tahmin Edicilerin Teorik Özellikleri	15
4. TOBİT RIDGE REGRESYON TAHMİN EDİCİSİ	17
4.1. Tobit Ridge Tahmin Edicisinin Teorik Özellikleri	17
4.2. k Yanlılık Parametresinin Tahmini	23
4.2.1. k Parametresi için Yeni Önerilen Tahmin Ediciler	23
4.2.2. Genelleştirilmiş Tobit Ridge Tahmin Edicisi için k_j 'lerin Seçimi	25
4.3. Monte Carlo Simülasyon Çalışması	25
4.3.1. Simülasyon Tasarımı	26
4.3.2. Simülasyon Sonuçları	27
5. TOBİT LIU REGRESYON TAHMİN EDİCİSİ	32
5.1. Tobit Liu Tahmin Edicisinin Teorik Özellikleri	32
5.2. d Yanlılık Parametresinin Tahmini	34
5.3. Monte Carlo Simülasyon Çalışması 1	36

5.3.1. Simülasyon Tasarımı	36
5.3.2. Simülasyon Sonuçları	37
5.4. Monte Carlo Simülasyon Çalışması 2	38
5.4.1. Simülasyon Tasarımı	39
5.4.2. Simülasyon Sonuçları	40
5.5. Gerçek Veri Uygulaması	40
5.5.1. Tobin Verisi	41
5.5.2. Çimento Verileri	43
6. SONUÇ VE ÖNERİLER	48
6.1. Sonuçlar	48
6.2. Öneriler	48
KAYNAKLAR	50
ÖZGEÇMİŞ	53

ŞEKİLLER LİSTESİ

<u>Şekil</u>	<u>Sayfa</u>
3.1 Standart Normal, Sansürlü ve Kesikli Normal Dağılımlar	10
5.1 Tobin veri setinde d' ye karşı MLE ve LE tahmin edicilerin MSE değerleri	43
5.2 Çimento veri setinde d' e karşı MLE ve LE tahmin edicilerin MSE değerleri – (sansürleme-1)	45
5.3 Çimento veri setinde d' ye karşı MLE ve LE tahmin edicilerin MSE değerleri – (sansürleme-2)	46

TABLolar LİSTESİ

<u>Tablo</u>	<u>Sayfa</u>
4.1 $\sigma = 1, p = 4$ olduğunda farklı ρ değerleri için MLE, GRR ve RR tahmin edicilerin MSE değerleri	27
4.2 $\sigma = 1, p = 8$ olduğunda farklı ρ değerleri için MLE, GRR ve RR tahmin edicilerin MSE değerleri	28
4.3 $\sigma = 1, p = 16$ olduğunda farklı ρ değerleri için MLE, GRR ve RR tahmin edicilerin MSE değerleri	28
4.4 $\sigma = 1, p = 32$ olduğunda farklı ρ değerleri için MLE, GRR ve RR tahmin edicilerin MSE değerleri	29
4.5 $\sigma = 5, p = 4$ olduğunda farklı ρ değerleri için MLE, GRR ve RR tahmin edicilerin MSE değerleri	30
4.6 $\sigma = 5, p = 8$ olduğunda farklı ρ değerleri için MLE, GRR ve RR tahmin edicilerin MSE değerleri	31
4.7 $\sigma = 5, p = 16$ olduğunda farklı ρ değerleri için MLE, GRR ve RR tahmin edicilerin MSE değerleri	31
4.8 $\sigma = 5, p = 32$ olduğunda farklı ρ değerleri için MLE, GRR ve RR tahmin edicilerin MSE değerleri	31
5.1 $p = 4$ için farklı ρ, n değerleri için MLE ve önerilen d parametrelerine sahip LE tahmin edicilerin MSE değerleri	38
5.2 $p = 8$ için farklı ρ, n değerleri için MLE ve önerilen d parametrelerine sahip LE tahmin edicilerin MSE değerleri	38
5.3 $p = 16$ için farklı ρ, n değerleri için MLE ve önerilen d parametrelerine sahip LE tahmin edicilerin MSE değerleri	39
5.4 $p = 32$ için farklı ρ, n değerleri için MLE ve önerilen d parametrelerine sahip LE tahmin edicilerin MSE değerleri	39
5.5 $\sigma = 1$ olduğunda farklı ρ, n, p değerleri için MLE ve LE tahmin edicilerin MSE değerleri	41
5.6 $\sigma = 5$ olduğunda farklı ρ, n, p değerleri için MLE ve LE tahmin edicilerin MSE değerleri	42

5.7	MLE ve önerilen d parametrelerine sahip LE tahmin edicileri için: Tobin verisinde tahmini kat sayılar ve MSE değerleri	42
5.8	MLE ve LE tahmin edicileri için: Tobin verilerinin tahmini katsayıları ve karşılık gelen MSE değerleri	43
5.9	MLE ve önerilen d parametrelerine sahip LE tahmin edicileri için: Çimento verisinde tahmini kat sayılar ve MSE değerleri – (sansürleme-1)	44
5.10	MLE ve LE tahmin edicileri için: Çimento verilerinin tahmini katsayıları ve karşılık gelen MSE değerleri – (sansürleme-1)	45
5.11	MLE ve önerilen d parametrelerine sahip LE tahmin edicileri için: Çimento verisinde tahmini kat sayılar ve MSE değerleri – (sansürleme-2)	46
5.12	MLE ve LE tahmin edicileri için: Çimento verilerinin tahmini katsayıları ve karşılık gelen MSE değerleri – (sansürleme-2)	46

SİMGELER VE KISALTMALAR

Simgeler

y	Bağımlı değişken
y^*	Gizil değişken
\mathbf{X}	Veri matrisi
β	Regresyon katsayı vektörü
$\hat{\beta}$	β katsayı vektörünün tahmin edicisi
λ_i	i . öz değer
n	Gözlem sayısı
p	Bağımsız değişken sayısı
σ^2	Modellerdeki hata teriminin varyansı
\mathbf{I}_n	$n \times n$ tipinde birim matris
$\epsilon_i \sim N(0, \sigma^2)$	0 ortalamalı ve σ^2 varyanslı normal dağılıma sahip rastgele hatalar
\mathbf{u}	$n \times 1$ tipinde rastgele hata vektörü
ϕ	Standart normal dağılımın olasılık yoğunluk fonksiyonu
Φ	Standart normal dağılımın birikimli dağılım fonksiyonu
$diag(\mathbf{X})$	\mathbf{X} matrisini köşegen elemanlarından oluşan vektör
$tr(\mathbf{X})$	\mathbf{X} matrisinin izi
\mathbf{X}^\top	\mathbf{X} matrisinin transpozu
$\ln(a)$	Logaritma e tabanında a 'nın değeri
$exp(a)$	e 'nin a . kuvveti
$E(\hat{\beta})$	$\hat{\beta}$ tahmin edicisinin beklenen değeri
$bias(\hat{\beta})$	$\hat{\beta}$ tahmin edicisinin yanlılığı
$Cov(\hat{\beta})$	$\hat{\beta}$ tahmin edicisinin varyans–kovaryans matrisi
$Var(\hat{\beta})$	$\hat{\beta}$ tahmin edicisinin varyansı
\sum	Toplam

Kısaltmalar

OLS	En Küçük Kareler (Ordinary Least Squares)
ML	En Çok Olabilirlik (Maximum Likelihood)
MLE	En Çok Olabilirlik Tahmin Edicisi (ML Estimator)
MSE	Hata Kareler Ortalaması (Mean Squared Error)
MMSE	Matris Hata Kareler Ortalaması
RR	Tobit Ridge Tahmin Edicisi
GRR	Genelleştirilmiş Tobit Ridge Tahmin Edicisi
LE	Tobit Liu Tahmin Edicisi
CN	Koşul Sayısı

1. GİRİŞ

Regresyon analizi deęişkenler arasındaki ilişkiyi modellemek ve incelemek için kullanılan istatistiksel bir tekniktir (Montgomery ve ark., 2012). Regresyon analizi hemen hemen tüm alanlarda kullanılmaktadır. Herhangi bir alandaki bir problemde deęişkenler arasında bir ilişkinin olduęu hissedilebilmektedir. Deęişkenler arasındaki ilişkinin şeklini, kuvvetini ve yönünü belirleyebilmemiz için regresyon analizine gerek vardır. Deęişkenler arasında ilişki olduęu hissediliyorsa, bu ilişki en iyi şekilde matematiksel bir denklem yardımıyla ifade edilebilir. Regresyon analizinin önemli bir amacı regresyon denklemindeki bilinmeyen parametreleri bulmaktır. Bulunan parametreler ile elde edilen regresyon denklemi, ilgili deęişkenler arasındaki gerçek fonksiyonel ilişkiye yalnızca bir yaklaşımdır. Regresyon analizinin sonraki evresinde modelin uygunluęu ve uyumun kalitesi incelenir. İnceleme sonucunda süreç ya modelin ya da uyumun deęiştirilmesi veya modelin kabulüyle sonuçlanır.

Doęrusal regresyon modelinde, tüm deęişkenlerin deęerleri tüm örneklem için bilinir. Ancak, gerçek hayatta her zaman durum böyle deęildir. Bazen bazı sebeplerden dolayı örneklem sınırlanabilir. Sansürleme (censoring) ve kesme (truncating) baęımlı deęişkenin sınırlı olduęu durumu ele almaktadır. Sansürleme ve kesme; tüm örneklem için baęımsız deęişkenleri gözlemlediğimizde bazı gözlemler için baęımlı deęişken hakkında bilginin sınırlı olmasıdır. Sansürlemede, baęımlı deęişkene ait bilgi sınırlı olmasına raęmen baęımsız deęişkenler veride gözlenebilir. Yani sınırlı baęımlı deęişkene ait gözlemler veriden çıkarılmaz. Kesme de ise sınırlı gözlemler hariç tutularak veriyi daha fazla sınırlandırılır. Bu yüzden kesme örneklemini deęiştirirken, sansürleme deęiştirmez (Long, 1997).

Sansürleme yapılmıř modelde en küçük kareler (OLS: ordinary least squares) tahmini üzerinde yanlı ve tutarsız sonuçlar vermektedir (Long, 1997). OLS'ye bir alternatif, Tobin (1958) tarafından önerilen sansürlü regresyon modelidir. Sansürlü regresyon modeline tobit model de denmektedir. Baęımlı deęişkenin alt veya üst limite sahip olduęu modeller olarak bilinen tobit model, ilk olarak 1958 yılında Tobin tarafından uygulanmıřtır (Tobin, 1958). Tobin (1958) hane halkı harcamalarını analiz etmiřtir. Harcamanın negatif olamayacaęı gerçeęine dayanarak geliri belirli düzeyi ařana kadar harcama sıfır kabul etmiřtir. Bu durum baęımlı deęişkenin sansürlenmesine klasik bir örnektir. Farklı veri gruplarında baęımlı

değişken farklı şekillerde sansürlenmiş olabilir. Farklı sansürlemeler arasında en sık kullanılan hane halkı harcamaları verisindeki gibi sansürlenmiş standart tobit modeldir (Emir, 2016). Amemiya (1985) beş farklı sansürleme ile tobit modelini tanımlamıştır. Bir başka tobit model tipi ise Heckman'ın iki aşamalı modeli ile genelleştirilmiştir (Heckman, 1979). Bu tez çalışmasında en sık karşılaşılan standart tobit model dikkate alınmıştır.

Bağımsız değişkenlerin doğrusal olarak ilişkili olduğu durumlarda regresyon modelinde OLS'ye dayanan çıkarımlar yanıltıcı ve hatalı olabilir. Bağımsız değişkenler arasında yakın doğrusal bağımlılık olduğunda çoklu bağlantı probleminin var olduğu söylenir. Çoklu bağlantı, hem OLS tarafından tahmin edilen doğrusal model hem de en çok olabilirlik (ML: maximum likelihood) yöntemi ile tahmin edilen doğrusal olmayan tobit model için iyi bilinen bir prob-lemidir. Bu problemde, tahmincilerin hata kareler ortalaması (MSE: mean squared error) çok büyük olduğu için bireysel katsayıların tahminlerini yorumlamak zorlaşır. Dolayısıyla, tahmin edilen parametre vektörü gerçek değerlerinden uzak olabilir. Bu sorunun bir sonucu olarak, en çok olabilirlik tahmin edicisi (MLE: ML estimator) dengesizleşir ve değişkenlik ortaya çıkar (Khalaf ve ark., 2014). Bu durumda, tahmin edilen parametreleri yorumlamak çok zordur. Bu nedenle, çoklu bağlantı probleminin varlığında, yansız tahmin edici MLE yerine yanlı tahmin ediciler gibi alternatif yöntemler kullanılabilir.

Literatürde tobit modellerinde çoklu bağlantı sorunu ile ilgili çok fazla çalışma yoktur. İstisnalar olarak, Khalaf ve ark. (2014), ridge regresyonu tahmin edicisinin kullanılmasını önermiştir ve Alhusseini ve Odah (2016) tobit modellerinde bu sorunun çözümü olarak temel bileşenler regresyonunu tanımlamıştır.

Bu tez çalışmasında, en çok olabilirlik tahmin edicisi ve çoklu bağlantı sorunun üstesinden gelmek için önerilen bazı yanlı tahmin yöntemleri göz önünde bulundurulmuştur. Ridge tahmin edicisi (Hoerl ve Kennard, 1970) ve Liu tahmin edicisi (Liu, 1993) çalışılmıştır.

Bu tahmin ediciler doğrusal regresyon modelinde başarıyla kullanılmaktadır. Khalaf ve ark. (2014) çalışmasında ridge tahmin edicisini tobit regresyon modeline uyarlamıştır ve doğrusal modelde yanlılık parametresi k için önerilen bazı tahmin ediciler tobit modele uyarlanarak Monte Carlo simulasyon çalışması ile karşılaştırma yapılmıştır. Bu tez çalışmasında farklı olarak matris hata kareler ortalaması (MMSE) ve hata kareler ortalaması (MSE) kriterlerine göre teorik karşılaştırmalar elde edilmiştir. Ayrıca doğrusal regresyon modelinde ridge tahmin edicisinin k yanlılık parametresi için tanımlanan birçok farklı tahmin edici tobit ridge tahmin edicisinin yanlılık parametresinin tahmini olarak önerilmiştir.

Bu tahmin edicilerden hangisinin MSE kriterine göre daha iyi performans sergilediğini bulmak için Monte Carlo simülasyon çalışması yapılmıştır.

Ayrıca, farklı bir yanlı tahmin edici olan Liu tahmin edicisi (Liu, 1993) tobit model için tanımlanmış ve bu tahmin edici tobit regresyon modelinde MMSE ve MSE kriterleri kullanılarak en çok olabilirlik tahmin edicisi ile karşılaştırılmıştır. Monte Carlo simülasyon çalışmaları tasarlanarak tahmin ediciler sayısal olarak karşılaştırılmıştır. Simülasyonlarda, tahmin edicilerin performanslarını karşılaştırmak için MSE kriteri kullanılmıştır.

Tezin akışı şu şekildedir: İlk olarak Bölüm 2’de sınırlı bağımlı değişkenler analizinde kullanılan tobit model ve yanlı tahmin ediciler ile ilgili yapılmış çalışmalardan bazıları hakkında bilgi verilmiştir.

Bölüm 3’te doğrusal regresyon modeli ve modelin varsayımları hakkında kısaca bilgi verilmiştir. Sonra tobit regresyon modeli tanıtılmış ve parametre tahmini için en çok olabilirlik yöntemi gösterilmiştir. En çok olabilirlik tahmin edicisinin kovaryans matrisi ve MSE özellikleri elde edilmiştir. Ardından çoklu bağlantı problemine ve bu problemin etkilerine değinilmiştir. Son olarak alt bölüm 3.4’de tahmin edicilerin teorik özelliklerinin bulunma formülleri ve tobit modelde en çok olabilirlik yöntemi kullanılarak elde edilen tahmin edicinin kanonik formu verilmiştir.

Bölüm 4’te, ilk başta Khalaf ve ark. (2014) tarafından önerilen tobit ridge tahmin edicisi tanıtılmıştır. Ayrıca genelleştirilmiş tobit ridge tahmincisi bu çalışmada önerilmiştir. Tobit ridge ve genelleştirilmiş tobit ridge tahmin edicilerinin bazı teorik özellikleri alt bölüm 4.1’de verilmiştir. Bu alt bölümde tobit ridge tahmin edicisi ile tobit MLE tahmin edicisi için MMSE ve MSE kriterlerine göre teorik karşılaştırmalar da elde edilmiştir. Tobit ridge tahmin edicisinin k yanlılık parametresi için farklı öneriler Bölüm 4.2’de verilmiştir. Bölüm 4.3’te, Monte Carlo simülasyon çalışması ile farklı k parametrelerine sahip tobit ridge tahmin edicileri, genelleştirilmiş tobit ridge ile tobit MLE tahmin edicisi sayısal olarak karşılaştırılmıştır.

Bölüm 5’te, tobit model için karşılaşılan çoklu bağlantı problemine çözüm olarak tobit Liu tahmin edicisi önerilmiştir. Tobit Liu tahmin edicisinin bazı teorik özellikleri alt bölüm 5.1’de verilmiştir. Ayrıca bu alt bölümde tobit Liu tahmin edicisi ile tobit MLE tahmin edicisi için MMSE ve MSE kriterlerine göre teorik karşılaştırmalar elde edilmiştir. Tobit Liu tahmin edicisinin d yanlılık parametresi için farklı öneriler Bölüm 5.2 de verilmiştir. Bölüm 5.3 ve 5.4’te Monte Carlo simülasyon çalışması ile farklı d parametrelerine sahip tobit Liu tahmin edicileri ile tobit MLE tahmin edicisi sayısal olarak karşılaştırılmıştır. Bölüm 5.3’te

önerilen d parametreleri kullanılarak, Bölüm 5.4'te ise d parametresi için değerler verilerek simülasyon çalışması yapılmıştır. Bölüm 5.5'te iki gerçek veri uygulaması gösterilmiştir.

Son bölümde, çalışmanın sonuçları verilmiş ve araştırmacılar için bazı öneriler sunulmuştur.

2. KAYNAK ARAŞTIRMASI

Sınırlı bağımlı değişkenler analizinde kullanılmak üzere önerilen iki dağılım işlevi vardır: Sansür ve kesme. Bu iki dağılım yoluyla birçok model geliştirilmiştir. Sansürleme örneklemin yapısını değiştirmedeği için anlamlı yorumlar sağlar (Long, 1997).

Emeç (2001) çalışmasında, sansürlenmiş ve kesilmiş regresyon modellerinin kullanımını incelemiştir. Çalışmasında 1994 yılı Türkiye hanehalkı tüketim verilerini kullanarak en küçük kareler yönteminin problemlerine dikkat çekmiş ve tobit model yardımıyla verinin parametrelerini tahmin etmiştir. Ayrıca veride bireylerin harcama olasılıklarının tahmin edebilmek için sıralı logit modelini kullanmıştır.

Mujasi ve ark. (2016) çalışmalarının amacı, 2012/2013 yılında Uganda'daki sevk hastanelerinin teknik verimliliğini araştırmaktır. Birinci aşamada, veri zarflama analizini numunedeki her bir hastanenin verimlilik puanlarını hesaplamak için kullanmışlardır. İkinci aşamada, verimlilik puanlarını bağımlı bir değişken olarak kullanarak ve puanların sağ sansürlendiği göz önüne almışlardır. Her hastane için ayarlanan verimlilik puanlarını tahmin etmek için bir tobit regresyon modeli kullanmışlardır.

Yahia ve Essid (2019) çalışmalarının amacı, Tunus orta öğretiminin teknik verimliliğini değerlendirmek ve Uluslararası Öğrenci Değerlendirme Programı (PISA) 2015 anketini kullanarak performansının belirleyicilerini belirlemektir. İlk aşamada, veri zarflama analizi yaklaşımını kullanarak her okulun verimlilik puanını tahmin edimşlerdir. İkinci aşamada, okulun verimliliğini etkileyen faktörler bir tobit regresyon analizi ile incelemiştir.

Wilson ve ark. (2020) çalışmalarında, sıkça kullanılan tobit modeli ve kriminolojide kullanımını incelemiştir. Çalışmalarında, tobit model de dağılım varsayımlarına dikkat edilmesi gerektiğine aksi takdirde tutarsız tahminler elde edileceğine dikkat çekmişlerdir. "Kriminologlar, daha verimli önyargılı tahminlere geçmeden önce, sansürlü sonuç verileri analizlerine mevcut modellerin en az kısıtlayıcı olanıyla başlamalıdır." tavsiyesinde bulunmuşlardır.

Wulff ve Villadsen (2020) çalışmalarında, uluslararası işletmelerde oranların analizi için genellikle kullanılan tobit ve kesirli regresyona değinmişlerdir. Bu alanda, tobit regresyon analizlerinin genellikle eksik veya hatalı olarak raporlanan sonuçlardan kaynaklı yorumlar olduğunu görmüşlerdir. Buna göre, araştırmacıların tobit ve kesirli regresyon arasında

nasıl seçim yapması gerektiğini açıklıp ve sonuçlarını yorumlamışlardır.

Çoğu çalışmada, karayolu segmentlerinde belirli bir süre boyunca kaza sayısını etkileyen faktörleri araştırmak için çeşitli sayım verileri modellerini kullanılmıştır. Çarpışma sıklığı ile ilgili çalışmalardan farklı olarak, Hou ve ark. (2020) çarpışma oranlarını doğrudan sifirda sansürlenmiş sürekli bir değişken olarak görüp ve tobit regresyonuna dayalı alternatif bir yaklaşımın uygulanmasını araştırmışlardır.

Bununla birlikte, sansürlemede en sık kullanılan tobit model; endüstri alanında (Kıranoğlu, 2005), ziraat alanında (Aydın, 2014), veterinerlik alanında (Emir, 2015), biyoistatistik alanında (Emir, 2016), sivil havacılık alanında (Yazgan, 2012), işletme alanında (Keskin, 2017), ekonometri alanında (Çakır, 2019) çalışmaların da olduğu gibi benzer şekilde birçok alanda kullanılmaktadır.

Tobit modelde en çok olabilirlik tahmin yöntemi, bağımsız değişkenlerin lineer bağımsız olması varsayımına dayanır. Ancak, gerçek hayatta bu varsayım her zaman sağlamayabilir. Bağımsız değişkenler birbiriyle ilişkili olduğunda en çok olabilirlik tahmin edicisi kararsız hale gelir. Yüksek varyanstan dolayı, MLE'nin tahminleri artık güvenilir değildir. Bağımsız değişkenlerin ilişkili olduğu (çoklu bağlantı) sadece tobit modelde değil en küçük hata kareler yöntemiyle tahmin edilen lineer modelde ve diğer modellerde iyi bilinen problemdir. Bu sorunun bazı çözümleri vardır.

Bunlardan bir tanesi, ilk olarak Hoerl ve Kennard (1970) tarafından lineer model için tanımlanan ridge regresyonudur. Hoerl ve Kennard (1970) çalışmalarında, bağımsız değişkenlerin ilişkili olduğu durumda $X^T X$ matrisinin öz değerlerinin bir kısmı sıfıra yaklaştığı için OLS tahminlerinin yetersiz olma ihtimalinin yüksek olduğunu göstermişlerdir. Ridge regresyonunun, $X^T X$ matrisinin köşegen elemanlarına bir pozitif sayı (k) eklenmesine dayanan tahmin prosedürü olduğuna değinmişlerdir. Daha sonra, hata kareler ortalamasının (MSE) küçük olacak şekilde k değerinin nasıl seçilebileceğini göstermişlerdir.

Ridge tahmin edicisi başarıyla Khalaf ve ark. (2014) tarafından tobit regresyon modeline uygulanmış ve bazı özellikleri araştırılmıştır. Bir Monte Carlo simülasyonu yaparak, MLE ve tobit ridge tahmin edicilerinin farklı koşullar altında çoklu bağlantının varlığında performanslarını araştırmışlardır. Simülasyonun sonuçlarına göre, tobit ridge tahmin edicileri MLE'den daha iyi performansa sahiptir.

Alhusseini ve Odah (2016) çalışmasında bağımsız değişkenlerin ilişkili olduğu sorunu tedavi etmek için ana bileşen yöntemini ele almışlardır. Tobit modelde, bireylerin altın mik-

tarı alımına odaklanmakta ve bu veriyi incelemektedir.

Liu (1993) lineer regresyon modelinde çoklu bağlantı problemine bir başka çözüm olarak Liu tahmin edicisi adı verilen bir yanlı tahmin ediciyi tanıttı. Liu tahmin edicisinde kullanılan d küçülme parametresinin seçiminde, Liu tahmin edicinin MSE değeri dikkate alınmıştır. Yine MSE kriterine göre, Liu ile OLS tahmin edicilerinin performansları teorik olarak karşılaştırılmıştır. Bir veride elde ettiği teorik sonuçları uygulanmış ve Liu tahmin edicisinin OLS tahmin edicisine göre daha iyi performans sergilediği sonucuna ulaşılmıştır.

Lojistik regresyon modelinde, bağımsız değişkenlerin yüksek derecede ilişkili olduğunda yaygın olarak kullanılan maksimum olabilirlik yönteminin hata kareler ortalamasını şişirir. Bu durumun olumsuz sonuçlarından dolayı, Månsson ve ark. (2012a) tarafından logit modeli için Liu tahmin edicisini tanıtmışlardır. Hata kareler ortalaması kullanılarak, d küçülme parametresinin optimal değeri türetilip ve bunu tahmin etmek için bazı yöntemler önermişlerdir. Monte Carlo simülasyonları ile tahmini hata kareler ortalaması ve mutlak hata ortalaması kriterlerine göre Liu tahmin edicisinin daha iyi performans sergilediğini göstermişlerdir. Çalışmalarında son olarak, belediyelerin sakinlerde net artış olasılığını açıklamak için farklı ekonomik faktörlerin kullanıldığı ampirik bir uygulama gösterilmektedir.

Benzer şekilde Poisson regresyonda (Månsson ve ark., 2012b), negatif binom regresyonda (Månsson, 2013), gamma regresyonda (Qasim ve ark., 2018) ve genelleştirilmiş doğrusal modellerde (Kurtoğlu and Özkale, 2016) Liu tahmin edicisi çalışılmıştır.

3. TOBİT REGRESYON MODELİ

Çoklu doğrusal regresyon modeli düşünülürse

$$y = X\beta + \epsilon$$

burada

- p : bağımsız değişken sayısı
- n : gözlem sayısı
- y : $(y_1, y_2, \dots, y_n)^\top$ gözlenen bağımlı değişkenlerden oluşan sütun vektörü
- X : $n \times p$ tipinde gözlenen bağımsız değişkenlerin matrisi
- β : $p \times 1$ tipinde katsayıların vektörü
- ϵ : $n \times 1$ tipinde rastgele hatalar vektörüdür.

Rastgele hata teriminde bazı varsayımlar vardır (Montgomery ve ark., 2012):

1. $E(\epsilon_i) = 0, i = 1, 2, \dots, n.$

Bu ifadeye göre hata terimleri sıfır ortalamaya sahiptir, hata terimlerinin beklenen değeri X ile değişmediği ifade eder.

2. $Var(\epsilon_i) = \sigma^2, i = 1, 2, \dots, n.$

Buna göre hata terimlerin varyansının tüm gözlemler için aynıdır. Yani hata terimlerinin varyansı homojen ve sabittir.

3. $i \neq j, 1 \leq i < j \leq n$ için $Cov(\epsilon_i, \epsilon_j) = 0.$

Hata terimleri birbirinden bağımsızdır.

Bu modelde bağımsız değişkenler rassal değişken değildir. Bağımsız değişkenler arasında doğrusal ilişki olmadığı varsayımı üzerine model kurulur. Bu varsayımın ihmal edilmesi durumu çoklu bağlantı problemi olarak adlandırılır.

Doğrusal regresyon modelinde, tüm değişkenlerin değerleri tüm gözlemler için bilinir. Bazen veri gruplarında örneklemin tamamındaki bağımsız değişkenler gözlemlendiğinde, bazı gözlemler için bağımlı değişkene dair kısıtlı bilgi mevcuttur. Örneğin, bağımlı değişken

verisinin 50'den küçük olduğu bilinebilir ama ne kadar küçük olduğu bilinmeyebilir. Veri gruplarında bağımlı değişkene dair kısıtlı bilgi bazen değişkenin yapısından kaynaklı olabilir. Örneğin, hane halkı harcama verisinde harcamanın negatif değere sahip olamayacağı gibi. Bağımlı değişken hakkında sınırlı bilgiye sahip olduğumuz durumlarda sansürleme veya kesme modelleri kullanılır. Kesme, sınırlı bilgiye sahip bağımlı değişkendeki gözlemlerin tamamının çıkarılmasıdır. Bu durum örnekleme daha fazla kısıtlamaktadır. Hausman ve Wise (1977) çalışmaları, kesme modellerinin ilk uygulamalarından biridir. New Jersey negatif gelir vergi deneyinde elde edilen örnekleme kesme işlemi uygulamışlardır. Böylece değişikliğe uğrayan örneklem, artık toplumsal bir gösterge olmaktan çıkmıştır. Sansürlemede, hiçbir gözlem çıkarılmadan bağımlı değişken üzerinden işlemler yapılır. Kesme örnekleme değiştirirken, sansürleme değiştirmez (Long, 1997).

En küçük kareler modelin hata terimlerinden yola çıkarak hesaplanan tahmin yöntemidir. Yöntemde gerçek y_i gözlemleri ile regresyon doğrusundaki farkların karesi alınarak kareler toplamı bulunur. Elde edilen toplam minimize edilecek şekilde β katsayı vektörü bulunur tahmin edilir.

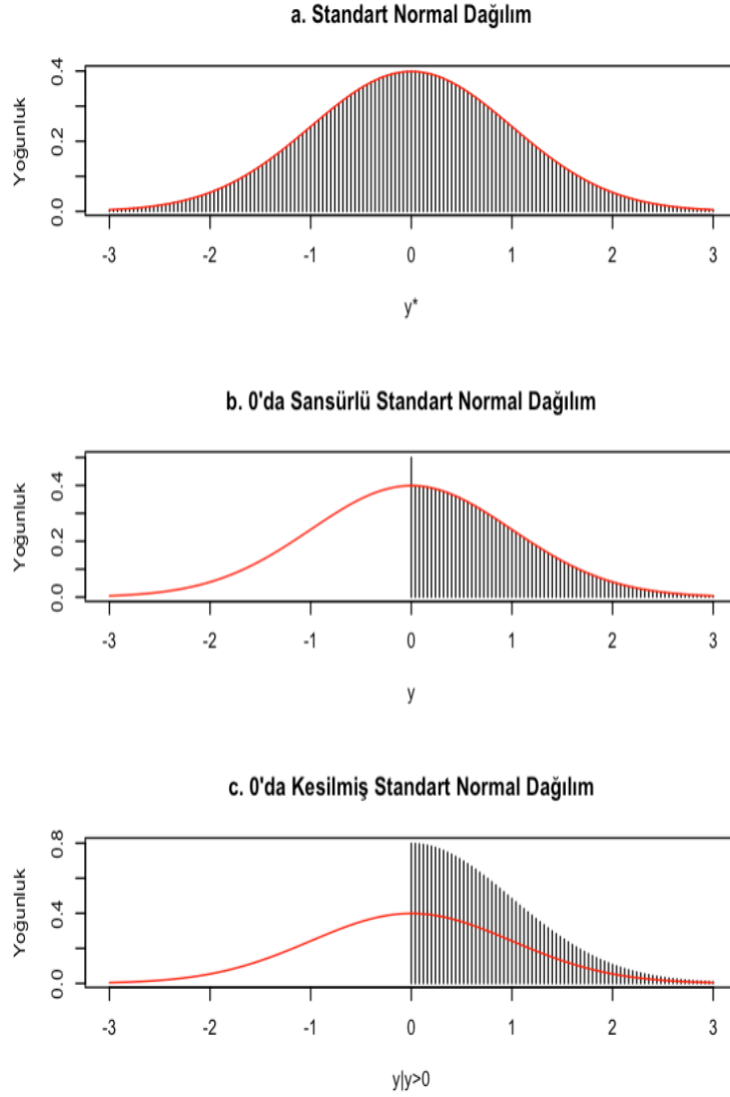
En çok olabilirlik yöntemi ise bilinmeyen parametreleri (β), verilen bağımlı değişkenlerinin gözlenme olasılığı maksimum olacak tarzda tahmini esasına dayanır. Bu yüzden tobit regresyon modeli için en çok olabilirlik tahmin yöntemi kullanılır.

3.1. Sansürlü ve Kesilmiş Dağılımlar

Tobit modelin için en çok olabilirlik tahmin yönteminden bahsedilmeden önce, kesilmiş ve sansürlenmiş normal dağılımlar hakkında bazı bilgilere değinilecektir. Tobit modelinde, bağımlı değişken (y^*) sıfırdan küçük değerler de sınırlıdır. Bu yüzden, kesilmiş ve sansürlenmiş dağılımların grafiğinde sıfırın solunda kesilme dikkate alınır. y^* standart normal dağılıma bir rasgele değişken olsun.

3.1.1. Sansürlü Normal Dağılım

y^* bağımlı değişkene ait sıfırın altındaki değerleri sıfır olarak alındığında y değişkeni ismi verilir. y değişkeni sansürlü normal dağılıma sahip olacaktır. Sansürlü normal dağılım,



Şekil 3.1. Standart Normal, Sansürlü ve Kesikli Normal Dağılımlar

y^* değerlerinin iki sete bölünmesini içerir. İlk set y^* değerleri ile aynı şekilde davrandığı sansürlü gözlemleri içerir. İkinci set ise sansürlü gözlemleri içerir. Sonuç olarak y dağılımı, kesikli ve sürekli parçaların bir karışımı haline gelir (Long (1997); sf 204).

Şekil (3.1)'de üç farklı grafik verilmiştir. Bu grafiklerde kırmızı çizgi ile gösterilen eğri standart normal dağılıma ait olasılık yoğunluk fonksiyonunun eğrisidir. Şekil (3.1)'in **a** panelinde standart normal dağılımın olasılık yoğunluk fonksiyonu gösterilmiştir. Sansürlü normal dağılım, Şekil (3.1)'in **b** panelinde siyah dikey çizgiler ile taralı olarak gösterilmiştir. Bu grafikten görülebileceği gibi standart normal dağılıma sahip y^* değişkeni sıfır değerinde soldan sansürlenmiştir.

3.1.2. Kesilmiş Normal Dağılım

y^* bağımlı değişkene ait sıfırın altındaki gözlemler silindiğinde $y|y > 0$ değişkeni kesilmiş normal dağılıma sahip olacaktır. y^* dağılımının sıfırdan küçük bölgedeki durumlar çıkarıldığından, sıfırdan büyük bölgedeki y^* dağılımı dikkate alınır. y^* dağılımlı olasılık yoğunluk fonksiyonunun, sıfırın sağındaki alanın dikkate alınması ile kesilmiş olasılık fonksiyonu ortaya çıkar. Bu durum, ortaya çıkan dağılımı belli bir alanda yer almaya zorlamıştır. Kesilmiş normal dağılım, Şekil (3.1)'in c panelinde siyah dikey çizgiler ile taralı olarak gösterilmiştir.

3.2. Tobit Regresyon Modeli

Tobit modeli için gözlemlenemeyen gizil değişkenin (latent variable) aşağıdaki biçimde olduğu kabul edilir:

$$y_i^* = \mathbf{x}_i^\top \boldsymbol{\beta} + u_i \quad (3.1)$$

Burada

- p : açıklayıcı değişkenler
- n : gözlem sayısı
- $\mathbf{y} = (y_1, y_2, \dots, y_n)^\top$ gözlenen bağımlı değişken
- \mathbf{X} : $n \times p$ tipinde data matris öyle ki i . satırı $\mathbf{x}_i^\top = (x_{i1}, x_{i2}, \dots, x_{ip})$
- $\boldsymbol{\beta}$: $p \times 1$ tipinde katsayıların vektörü
- \mathbf{u} : $n \times 1$ tipinde rastgele hata vektörü

Bağımsız değişkenler (\mathbf{x}_i) her durumda gözlemlenmektedir. Tobit modeller için hata terimlerinin normal dağıldığı $\mathbf{u} \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$ varsayılır. Doğrusal regresyon modelinde hata terimlerinin normal dağıldığı varsayımı ihlal edilse bile OLS tahmin edicileri yansız ve tutarlıdır. Tobit modelde bu durum farklıdır. ML yönteminde hata terimleri normal dağılmadığında sonuçlar tutarsızdır (Emeç (2001); sf 8).

y^* gizil değişkeni 0'dan büyük değerler için gözlenir ve 0'dan küçük veya ona eşit değerler için sansürlenir. Bu nedenle, gözlenen bağımlı değişken $\mathbf{y} = (y_1, y_2, \dots, y_n)^\top$ ile tanımlanır:

$$y_i = \begin{cases} y_i^* & \text{ise } y_i^* > 0 \\ 0 & \text{ise } y_i^* \leq 0 \end{cases} \quad (3.2)$$

x 'in her bir değerinde, x verilen y_i^* dağılımını gösteren normal bir eğridir. Sansürlü gözlemler için, gözlem ile normal eğri arasındaki mesafe, belirli bir β ve σ için bu gözlemin olasılığıdır. Şekil (3.1) de $y_i^* = 0$ satırı, sansürün nerede gerçekleştiğini gösterir. (x_i, y_i^*) gibi sansürlenmiş gözlemler için, y_i^* değerini bilinmez ve bu nedenle normal eğrinin yüksekliği bu noktada kullanılmaz. Sansürlü vakalar için bildiğimiz tek şey $y_i^* \leq 0$ olduğunda, sansür olasılığını olasılık olarak kullanılır (Long (1997); sf 204).

Denklem (3.2)'de verilen sansürün mevcut olduğu varsayılarak, bağımlı değişkenlerin yoğunluk fonksiyonu aşağıdaki şekilde verilmiştir

$$f(y_i|\beta, \sigma^2) = \left[\frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left(-\frac{1}{2\sigma^2}(y_i - \mathbf{x}_i^\top \beta)^2\right) \right]^{d_i} \left[1 - \Phi\left(\frac{\mathbf{x}_i^\top \beta}{\sigma}\right) \right]^{1-d_i}$$

burada Φ normal dağılımın kümülatif dağılım fonksiyonu (cdf) ve d_i kukla değişken olarak

$$d_i = \begin{cases} 1 & \text{ise } y_i^* > 0 \\ 0 & \text{ise } y_i^* \leq 0 \end{cases} \quad (3.3)$$

şeklinde tanımlanır. Tobit regresyon model için log olabilirlik fonksiyonu

$$L(\beta, \sigma^2|\mathbf{x}) = \sum_{i=1}^n \left[-\frac{1}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} d_i (y_i - \mathbf{x}_i^\top \beta)^2 + (1 - d_i) \ln\left(1 - \Phi\left(\frac{\mathbf{x}_i^\top \beta}{\sigma}\right)\right) \right]$$

olarak elde edilir. β katsayı vektörünü en çok olabilirlik yöntemiyle tahmin etmek için, tobit log-olabilirlik fonksiyonunun β 'a göre kısmi türevi sifıra eşitlenir

$$S(\beta) = \frac{\partial L(\beta, \sigma^2|\mathbf{x})}{\partial \beta} = \sum_{i=1}^n \frac{1}{\sigma^2} \left[d_i (y_i - \mathbf{x}_i^\top \beta) - (1 - d_i) \frac{\sigma \phi\left(\frac{\mathbf{x}_i^\top \beta}{\sigma}\right)}{1 - \Phi\left(\frac{\mathbf{x}_i^\top \beta}{\sigma}\right)} \right] = 0 \quad (3.4)$$

ve Denklem (3.4) β 'ya göre çözülür. Burada ϕ standart normal dağılımın olasılık yoğunluk fonksiyonudur. Denklem (3.4), β 'nın lineer olmayan bir denklemi olduğundan bu denklemi çözebilmek için iteratif yöntemler kullanılır, örneğin Newton–Raphson metodu kullanılabilir. Newton–Raphson metodunun iterasyonları Denklem (3.5) ile gösterilmiştir:

$$\beta_r = \beta_{r-1} + \mathbf{I}^{-1}(\beta_{r-1}) S(\beta_{r-1}) \quad (3.5)$$

burada $\mathbf{I}(\boldsymbol{\beta}_{r-1})$ algoritmanın $(r-1)^{th}$ basamağında $\boldsymbol{\beta}_{r-1}$ kullanarak hesaplanan Fisher bilgi matrisidir, yani

$$\mathbf{I}(\boldsymbol{\beta}_{r-1}) = E \left(-\frac{\partial^2 L(\boldsymbol{\beta}, \sigma^2 | \mathbf{x})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top} \right) = \mathbf{X}^\top \widehat{\mathbf{W}} \mathbf{X} \quad (3.6)$$

öyle ki $\widehat{\mathbf{W}}(\boldsymbol{\beta}_{r-1}) = \text{diag} [\hat{a}_i(\boldsymbol{\beta}_{r-1})]$ şeklinde tanımlanır. $\delta = \frac{\mathbf{x}_i^\top \boldsymbol{\beta}}{\sigma}$ olsun, \hat{a}_i aşağıdaki gibi tanımlanır (bakınız Greene (2003))

$$\hat{a}_i = \sigma^2 \Phi(\delta) \left[1 - \left(\frac{\phi(\delta)}{\Phi(\delta)} \right)^2 - \delta \frac{\phi(\delta)}{\Phi(\delta)} + \left(\delta + \frac{\phi(\delta)}{\Phi(\delta)} \right)^2 (1 - \Phi(\delta)) \right].$$

Katsayı vektörü, bazı yakınsama kriterleri yerine getirilinceye kadar güncellenir. Algoritmanın son aşamasında, katsayılar vektörü, $\widehat{\boldsymbol{\beta}}_{MLE}$ ile gösterilen en çok olabilirlik tahmin edicisi MLE olur. Tobit MLE tahmin edicisi aşağıda verilen kovaryans matrisi ile asimptotik olarak normal dağılır

$$\text{Cov}(\widehat{\boldsymbol{\beta}}_{MLE}) = E \left(-\frac{\partial^2 L(\boldsymbol{\beta}, \sigma^2 | \mathbf{x})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top} \right)^{-1} = \left(\mathbf{X}^\top \widehat{\mathbf{W}} \mathbf{X} \right)^{-1} \quad (3.7)$$

ve MLE'nin skaler hata kareler ortalaması

$$MSE(\widehat{\boldsymbol{\beta}}_{MLE}) = \text{tr} \left(\mathbf{X}^\top \widehat{\mathbf{W}} \mathbf{X} \right)^{-1} = \sum_{j=1}^p \left(\frac{1}{\lambda_j} \right) \quad (3.8)$$

şeklinde elde edilir. $\mathbf{X}^\top \widehat{\mathbf{W}} \mathbf{X} = \mathbf{Q}^\top \boldsymbol{\Lambda} \mathbf{Q}$ tanımlandığı şekilde:

- \mathbf{Q} : $\mathbf{X}^\top \widehat{\mathbf{W}} \mathbf{X}$ matrisinin özvektörlerinden oluşan ortogonal matris
- $\boldsymbol{\Lambda} = \text{diag}(\lambda_j)$: $\mathbf{X}^\top \widehat{\mathbf{W}} \mathbf{X}$ matrisinin özdeğerlerinden oluşan köşegen matrisdir.

3.3. Çoklu Bağlantı Problemi

Regresyon analizinde, sütunları bağımsız değişkenlerden oluşan \mathbf{X} matrisinin j . sütunu \mathbf{x}_j olarak gösterilsin ve $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p]$ olsun. Her sıfırdan farklı t_1, t_2, \dots, t_p sabit reel sayıları için

$$\sum_{j=1}^p t_j \mathbf{x}_j = \mathbf{0}$$

ise $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p$ vektörleri lineer bağımlıdır. Bu toplamın değeri sıfıra yakın olması durumuna ise yakın doğrusal ilişki vardır denir. Bağımsız değişkenler arasında lineer bağımlılık

ve yakın doğrusal ilişki olduğunda çoklu bağlantı probleminin var olduğu söylenir (Montgomery ve ark. (2012): sf 285). Çoklu bağlantı probleminin varlığı, en küçük kareler ve en çok olabilirlik yöntemleri ile yapılan tahminlerde istenilmeyen durumdur.

Çoklu bağlantının sebeplerinin belirlenmesi bize çözüm arayışında bazı ipuçları verebilir. Çoklu bağlantı aşağıdaki belirtilen sebeplerden bir veya birkaç tanesinin ortak sonucu olarak ortaya çıkabilir. Bunlar (Montgomery ve ark. (2012) ,Gujarati (1988));

- Örneklem Yöntemleri: Araştırmacı sadece eğilimin yönünü belirleyen değişkenlerin bulunduğu bölgeden bir örneklem seçmesi durumunda çoklu bağlantı problemi oluşacaktır.
- Aşırı Tanımlanmış Model: Gözlem sayısının parametre sayısından az olması ($n < p$) durumu olarak da açıklanabilir. Daha çok tıbbi araştırmalarda karşılaşılan bu modellerde geçerli örnek birimi sayısı azdır. Bu noktadan kaynaklanan bir çoklu bağlantı probleminden kurtulabilmek için önemine göre bazı değişkenleri modelden çıkartmak veya gözlem sayısını artırmak gerekebilir.
- Model ve Kitle Üzerindeki Fiziksel Kısıtlar: Gerçekte kitlede var olan ilişkinin örneklemede de ortaya çıkması olarak açıklanabilir. Kitledeki zorunluluklar daha çok bağımsız değişkenlerin kimyasal veya üretim proseslerinden ortaya çıkar. Örneğin bir kimyasal reaksiyonun gerçekleşmesi için belli içeriklerin sabit oranlarda olması vb.
- Modelin Özellikleri: Veri setindeki değişkenlerin dönüşümlerinden yararlanarak ya da bu değişkenleri kullanarak hesaplanan yeni değişkenler üretmek, ilgili değişkenler arasında çoklu bağlantı ortaya çıkarabilir. Bu gibi durumlarda değişkenlerin oranları ve kuvvetleri, genellikle özgün değişkenlerle yaklaşık olarak çoklu bağlantılı olacaktır.

Çoklu bağlantı probleminin tobit MLE tahmin edicisi üzerinde ciddi etkileri vardır. Çoklu bağlantı probleminin derecesine göre $\mathbf{X}^T \widehat{\mathbf{W}} \mathbf{X}$ matrisinin özdeğerleri olan λ_j değerleri içinden en az bir tanesi küçük olacaktır. Bu nedenle, Denklem (3.7)'den açık olarak görüldüğü gibi çoklu bağlantı probleminin varlığı büyük varyansa yol açacaktır. Büyük varyansa bağlı olarak β katsayı parametresi tahminlerinin işaretlerini ve değerlerini etkileyecektir. Bu durum bağımlı ile bağımsız değişken arasındaki ilişkinin yönünü ve kuvvetini yanlış gösterecektir.

Aynı zamanda Denklem (3.8)'de görülebileceği gibi tobit MLE tahmin edicisinin MSE değerini yükseltecektir. Bu durum en çok olabilirlik tahmin edicisi $\widehat{\beta}_{MLE}$ 'nin gerçek

parametre vektörü olan β 'ya uzaklığının büyük olduğunu göstermektedir. Yani tobit MLE yöntemi ile yapılan çıkarımlar çoklu bağlantı probleminden dolayı yanıltıcı ya da hatalı olacaktır.

3.4. Tahmin Edicilerin Teorik Özellikleri

Tahmin edicilerin teorik özelliklerini incelemek için, tahmin ediciler ile ilgili önemli bilgileri içeren MMSE ve MSE fonksiyonları karşılaştırma kriteri olarak kullanılabilir. $\hat{\beta}^*$ tahmin edicisinin MMSE'si ve MMSE'nin izi olan MSE'si sırasıyla aşağıdaki gibi bulunur

$$\text{MMSE}(\hat{\beta}^*) = E \left[(\hat{\beta}^* - \beta) (\hat{\beta}^* - \beta)^\top \right] = \text{Cov}(\hat{\beta}^*) + \text{bias}(\hat{\beta}^*) \text{bias}(\hat{\beta}^*)^\top,$$

$$\text{MSE}(\hat{\beta}^*) = \text{tr}(\text{MMSE}(\hat{\beta}^*)) = \text{tr}[\text{Cov}(\hat{\beta}^*)] + \text{bias}(\hat{\beta}^*)^\top \text{bias}(\hat{\beta}^*).$$

Burada

- $\text{tr}(\cdot)$: iz ve $E(\cdot)$: beklenen değer operatörleri
- $\text{Cov}(\hat{\beta}^*)$: $\hat{\beta}^*$ tahmin edicisinin varyans–kovaryans matrisi
- $\text{bias}(\hat{\beta}^*) = E(\hat{\beta}^*) - \beta$: $\hat{\beta}^*$ tahmin edicisinin yan(bias) vektörüdür.

Bu çalışmada listelenen tahmin edicilerin MSE ve MMSE fonksiyonlarını karşılaştıran bazı teoremler elde edilmiştir. Bunu yapmak için, MSE ve MMSE farkları göz önünde bulundurularak bu farkların hangi koşullar altında pozitif oldukları araştırılmıştır. Eğer bir matris \mathbf{A} pozitif tanımlı ise, o zaman $\mathbf{A} > \mathbf{0}$ yazılır. Benzer şekilde, eğer \mathbf{A} negatif tanımlı değilse, o zaman $\mathbf{A} \geq \mathbf{0}$ yazılır. Tahmin edicilerin karşılaştırmalarını kolaylaştırmak için modelin kanonik formu dikkate alınmıştır.

$\mathbf{Q}^\top \mathbf{X}^\top \widehat{\mathbf{W}} \mathbf{X} \mathbf{Q} = \mathbf{\Lambda} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_p)$ olmak üzere burada $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p > 0$, $\mathbf{X}^\top \widehat{\mathbf{W}} \mathbf{X}$ matrisinin sıralı öz değerleri ve $\mathbf{Q} = [\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_p]$ matrisinin sütun vektörleri, $\mathbf{X}^\top \widehat{\mathbf{W}} \mathbf{X}$ matrisinin normalize edilmiş öz vektörlerinden oluşur. Bu nedenle, kanonik modelde $\mathbf{Z} = \mathbf{X} \mathbf{Q}$ ve $\alpha = \mathbf{Q}^\top \beta$ olarak ifade edilebilir. MLE'nin kanonik şekli $\hat{\alpha} = \mathbf{Q}^\top \hat{\beta}_{\text{MLE}}$ dir. $\hat{\alpha}$ nın asimptotik dağılımının $\hat{\alpha} \sim N(\mathbf{0}, \mathbf{\Lambda}^{-1})$ ile normal olduğu bilinmektedir.

MLE yansız tahmin edici olduğu için beklenen değeri:

$$\begin{aligned} E(\hat{\boldsymbol{\alpha}}) &= \boldsymbol{\alpha} \\ bias(\hat{\boldsymbol{\alpha}}) &= E(\hat{\boldsymbol{\alpha}}) - \boldsymbol{\alpha} = \mathbf{0} \end{aligned} \quad (3.9)$$

MLE tahmin edicisi yansız olmasından dolayı MMSE'si varyans–kovaryans matrisine eşittir

$$MMSE(\hat{\boldsymbol{\alpha}}) = Cov(\hat{\boldsymbol{\alpha}}) = \boldsymbol{\Lambda}^{-1}. \quad (3.10)$$

MLE tahmin edicisinin MSE fonksiyonu ise aşağıda verilmiştir

$$\begin{aligned} MSE(\hat{\boldsymbol{\alpha}}) &= tr(MMSE(\hat{\boldsymbol{\alpha}})) \\ &= \sum_{j=1}^p \frac{1}{\lambda_j}. \end{aligned} \quad (3.11)$$

Aşağıdaki lemma bu tezde verilen teoremlerin bazılarını ispatlamak için kullanılacaktır.

Lemma 3.1 (Farebrother (1976), Rao ve ark. (2008)) \mathbf{M} $n \times n$ boyutlu pozitif tanımlı bir matris ve $\boldsymbol{\alpha}$ $n \times 1$ boyutlu herhangi bir vektör olsun. $c\mathbf{M} - \boldsymbol{\alpha}\boldsymbol{\alpha}^\top \geq \mathbf{0}$ olması için gerek ve yeter şart $\boldsymbol{\alpha}^\top \mathbf{M}^{-1} \boldsymbol{\alpha} \leq c$.

4. TOBİT RİDGE REGRESYON TAHMİN EDİCİSİ

Doğrusal regresyon modelinde çoklu bağlantı probleminin varlığında, $\mathbf{X}^\top \mathbf{X}$ matrisinin öz değerlerinin bir kısmı sıfıra yaklaşır. Böylece, sıradan en küçük kareler tahmin edicisi, $\hat{\beta}_{OLS} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y}$ kararsız hale gelir. Hoerl ve Kennard (1970) $\mathbf{X}^\top \mathbf{X}$ matrisinin öz değerlerinin sıfıra yaklaşmasından dolayı köşegen elemanlarına sabit bir $k > 0$ ekleyerek tahminleri iyileştirmeyi amaçlamıştır. Bu amaçla, Hoerl ve Kennard (1970) tarafından, ridge tahmin edicisi $\hat{\beta}_k = (\mathbf{X}^\top \mathbf{X} + k\mathbf{I})^{-1} \mathbf{X}^\top \mathbf{X} \hat{\beta}_{OLS}$ 'i önerilmiştir.

Tobit model için de karşılaşılan çoklu bağlantı problemine çözüm önerisi olarak Khalaf ve ark. (2014) tarafından ridge tahmin edicisinin geliştirilmesi olarak tobit ridge regresyon tahmin edicisini aşağıdaki gibi tanımlanmıştır

$$\hat{\beta}_{RR} = (\mathbf{X}^\top \widehat{\mathbf{W}}\mathbf{X} + k\mathbf{I})^{-1} (\mathbf{X}^\top \widehat{\mathbf{W}}\mathbf{X}) \hat{\beta}_{MLE}. \quad (4.1)$$

$(\mathbf{X}^\top \widehat{\mathbf{W}}\mathbf{X})$ matrisinin her bir köşegen elemanına aynı pozitif k değeri yerine, her bir değişken için farklı pozitif k_j parametreleri eklenerek geliştirilmiş tobit ridge regresyon (GRR) tahmin edicisi aşağıdaki gibi önerilebilir

$$\hat{\beta}_{GRR} = (\mathbf{X}^\top \widehat{\mathbf{W}}\mathbf{X} + \mathbf{K})^{-1} (\mathbf{X}^\top \widehat{\mathbf{W}}\mathbf{X}) \hat{\beta}_{MLE} \quad (4.2)$$

öyle ki $\mathbf{K} = \text{diag}(k_1, k_2, \dots, k_p)$ matrisi köşegen elemanları $j = 1, 2, \dots, p$ için $k_j > 0$ olan $p \times p$ tipinde köşegen matristir.

4.1. Tobit Ridge Tahmin Edicisinin Teorik Özellikleri

Tahmin edicilerin karşılaştırmaları yapılırken kanonik formları dikkate alınmıştır. $\hat{\beta}_{RR}$ 'nin kanonik formu aşağıda verilen şekilde yazılabilir

$$\hat{\alpha}_{RR} = (\mathbf{\Lambda} + k\mathbf{I})^{-1} \mathbf{\Lambda} \hat{\alpha}. \quad (4.3)$$

Tahmin edicilerin MMSE ve MSE'sini hesaplamadan önce, yanı (bias) ve varyans–kovaryans matrisi elde edilmiştir. Tobit ridge tahmin edicisinin yan fonksiyonunu elde etmek için ilk başta tahmin edicinin beklenen değeri bulunmalıdır. Beklenen değer operatörünün

özellikleri ve (3.9) eşitliğinden yararlanarak tobit ridge tahmin edicisinin beklenen değeri

$$\begin{aligned}
E(\hat{\alpha}_{RR}) &= E[(\mathbf{\Lambda} + k\mathbf{I})^{-1} \mathbf{\Lambda} \hat{\alpha}] \\
&= (\mathbf{\Lambda} + k\mathbf{I})^{-1} \mathbf{\Lambda} E[\hat{\alpha}] \\
&= (\mathbf{\Lambda} + k\mathbf{I})^{-1} \mathbf{\Lambda} \alpha
\end{aligned} \tag{4.4}$$

olarak yazılabilir. Tobit ridge tahmin edicisinin yanı

$$\begin{aligned}
bias(\hat{\alpha}_{RR}) &= E(\hat{\alpha}_{RR}) - \alpha \\
&= (\mathbf{\Lambda} + k\mathbf{I})^{-1} \mathbf{\Lambda} \alpha - \alpha \\
&= (\mathbf{\Lambda} + k\mathbf{I})^{-1} [(\mathbf{\Lambda} + k\mathbf{I}) - k\mathbf{I}] \alpha - \alpha \\
&= [\mathbf{I} - k(\mathbf{\Lambda} + k\mathbf{I})^{-1}] \alpha - \alpha \\
&= -k(\mathbf{\Lambda} + k\mathbf{I})^{-1} \alpha.
\end{aligned} \tag{4.5}$$

şeklinde elde edilir. Kovaryans fonksiyonunun özellikleri ve (3.10) eşitliğinden yararlanarak RR tahmin edicisinin varyans–kovaryans matrisi

$$\begin{aligned}
Cov(\hat{\alpha}_{RR}) &= Cov[(\mathbf{\Lambda} + k\mathbf{I})^{-1} \mathbf{\Lambda} \hat{\alpha}] \\
&= (\mathbf{\Lambda} + k\mathbf{I})^{-1} \mathbf{\Lambda} Cov[\hat{\alpha}] \mathbf{\Lambda} (\mathbf{\Lambda} + k\mathbf{I})^{-1} \\
&= (\mathbf{\Lambda} + k\mathbf{I})^{-1} \mathbf{\Lambda} (\mathbf{\Lambda} + k\mathbf{I})^{-1}
\end{aligned} \tag{4.6}$$

olarak elde edilir. Varyans–kovaryans matrisi ve yanı bulunan RR tahmin edicisinin MMSE’si ise

$$\begin{aligned}
MMSE(\hat{\alpha}_{RR}) &= Cov(\hat{\alpha}_{RR}) + bias(\hat{\alpha}_{RR}) bias(\hat{\alpha}_{RR})^\top \\
&= Cov(\hat{\alpha}_{RR}) + k^2 (\mathbf{\Lambda} + k\mathbf{I})^{-1} \alpha \alpha^\top (\mathbf{\Lambda} + k\mathbf{I})^{-1} \\
&= (\mathbf{\Lambda} + k\mathbf{I})^{-1} \mathbf{\Lambda} (\mathbf{\Lambda} + k\mathbf{I})^{-1} + k^2 (\mathbf{\Lambda} + k\mathbf{I})^{-1} \alpha \alpha^\top (\mathbf{\Lambda} + k\mathbf{I})^{-1}
\end{aligned} \tag{4.7}$$

olarak elde edilir. Tobit ridge tahmin edicisinin MSE değeri aşağıdaki gibidir (Khalaf ve ark. (2014))

$$\begin{aligned}
MSE(\hat{\alpha}_{RR}) &= tr(MMSE(\hat{\alpha}_{RR})) \\
&= \sum_{j=1}^p \frac{\lambda_j}{(\lambda_j + k)^2} + k^2 \sum_{j=1}^p \frac{\alpha_j^2}{(\lambda_j + k)^2} \\
&= \kappa_1(k) + \kappa_2(k)
\end{aligned} \tag{4.8}$$

$\widehat{\beta}_{\text{GRR}}$ özelliklerini incelemek için $\widehat{\beta}_{\text{GRR}}$ tahmin edicisinin kanonik formu aşağıda verilen şekilde yazılabilir

$$\widehat{\alpha}_{\text{GRR}} = (\mathbf{\Lambda} + \mathbf{K})^{-1} \mathbf{\Lambda} \widehat{\alpha}. \quad (4.9)$$

Burada $\mathbf{K} = \text{diag}(k_1, k_2, \dots, k_p)$, elemanları $j = 1, 2, \dots, p$ için $k_j > 0$ olan $p \times p$ tipinde köşegen matristir. GRR tahmin edicisinin yan fonksiyonunu elde etmek için ilk adım da tahmin edicinin beklenen değeri bulunmalıdır. Beklenen değer fonksiyonunun özellikleri ve (3.9) eşitliğinden yararlanarak GRR tahmin edicisinin beklenen değeri

$$\begin{aligned} E(\widehat{\alpha}_{\text{GRR}}) &= E[(\mathbf{\Lambda} + \mathbf{K})^{-1} \mathbf{\Lambda} \widehat{\alpha}] \\ &= (\mathbf{\Lambda} + \mathbf{K})^{-1} \mathbf{\Lambda} E[\widehat{\alpha}] \\ &= (\mathbf{\Lambda} + \mathbf{K})^{-1} \mathbf{\Lambda} \alpha \end{aligned} \quad (4.10)$$

şeklinde bulunur. GRR tahmin edicisinin yanı aşağıdaki gibi bulunabilir

$$\begin{aligned} \text{bias}(\widehat{\alpha}_{\text{GRR}}) &= E(\widehat{\alpha}_{\text{GRR}}) - \alpha \\ &= (\mathbf{\Lambda} + \mathbf{K})^{-1} \mathbf{\Lambda} \alpha - \alpha \\ &= (\mathbf{\Lambda} + \mathbf{K})^{-1} [(\mathbf{\Lambda} + \mathbf{K}) - \mathbf{K}] \alpha - \alpha \\ &= [\mathbf{I} - (\mathbf{\Lambda} + \mathbf{K})^{-1} \mathbf{K}] \alpha - \alpha \\ &= -(\mathbf{\Lambda} + \mathbf{K})^{-1} \mathbf{K} \alpha. \end{aligned} \quad (4.11)$$

Kovaryans fonksiyonunun özellikleri ve (3.10) eşitliğinden yararlanarak GRR tahmin edicisinin varyans–kovaryans matrisi

$$\begin{aligned} \text{Cov}(\widehat{\alpha}_{\text{GRR}}) &= \text{Cov}[(\mathbf{\Lambda} + \mathbf{K})^{-1} \mathbf{\Lambda} \widehat{\alpha}] \\ &= (\mathbf{\Lambda} + \mathbf{K})^{-1} \mathbf{\Lambda} \text{Cov}[\widehat{\alpha}] \mathbf{\Lambda} (\mathbf{\Lambda} + \mathbf{K})^{-1} \\ &= (\mathbf{\Lambda} + \mathbf{K})^{-1} \mathbf{\Lambda} (\mathbf{\Lambda} + \mathbf{K})^{-1} \end{aligned} \quad (4.12)$$

olarak elde edilir. Kovaryans ve yan fonksiyonları bulunan GRR tahmin edicisinin MMSE'si ise

$$\begin{aligned} \text{MMSE}(\widehat{\alpha}_{\text{GRR}}) &= \text{Cov}(\widehat{\alpha}_{\text{GRR}}) + \text{bias}(\widehat{\alpha}_{\text{GRR}}) \text{bias}(\widehat{\alpha}_{\text{GRR}})^\top \\ &= (\mathbf{\Lambda} + \mathbf{K})^{-1} \mathbf{\Lambda} (\mathbf{\Lambda} + \mathbf{K})^{-1} \\ &\quad + (\mathbf{\Lambda} + \mathbf{K})^{-1} \mathbf{K} \alpha \alpha^\top \mathbf{K} (\mathbf{\Lambda} + \mathbf{K})^{-1} \end{aligned} \quad (4.13)$$

olarak bulunur. GRR tahmin edicisinin MSE değeri

$$\begin{aligned} \text{MSE}(\hat{\boldsymbol{\alpha}}_{\text{GRR}}) &= \text{tr}(\text{MMSE}(\hat{\boldsymbol{\alpha}}_{\text{GRR}})) \\ &= \sum_{j=1}^p \frac{\lambda_j}{(\lambda_j + k_j)^2} + \sum_{j=1}^p \frac{k_j^2 \alpha_j^2}{(\lambda_j + k_j)^2} \end{aligned} \quad (4.14)$$

olarak elde edilir.

Aşağıdaki teoremlerde RR ve MLE tahmin edicilerinin MMSE ve MSE kriterlerine göre karşılaştırmaları ele alınmıştır.

Teorem 4.1 *Tobit regresyon modelinde ridge tahmin edicisinin MMSE kriterine göre en çok olabilirlik tahmin edicisinden üstün olması için gerek ve yeter şart*

$$\boldsymbol{\alpha}^\top [k\boldsymbol{\Lambda}^{-1} + 2\mathbf{I}]^{-1} \boldsymbol{\alpha} \leq \frac{1}{k}$$

eşitsizliğinin sağlanmasıdır.

İspat MLE ve RR tahmin edicilerinin MMSE fonksiyonları arasındaki fark,

$$\Delta_1 = \text{MMSE}(\hat{\boldsymbol{\alpha}}) - \text{MMSE}(\hat{\boldsymbol{\alpha}}_{\text{RR}})$$

olsun. Δ_1 ise

$$\begin{aligned} \Delta_1 &= \boldsymbol{\Lambda}^{-1} - (\boldsymbol{\Lambda} + k\mathbf{I})^{-1} \boldsymbol{\Lambda} (\boldsymbol{\Lambda} + k\mathbf{I})^{-1} + k^2 (\boldsymbol{\Lambda} + k\mathbf{I})^{-1} \boldsymbol{\alpha} \boldsymbol{\alpha}^\top (\boldsymbol{\Lambda} + k\mathbf{I})^{-1} \\ &= (\boldsymbol{\Lambda} + k\mathbf{I})^{-1} [(\boldsymbol{\Lambda} + k\mathbf{I})\boldsymbol{\Lambda}^{-1}(\boldsymbol{\Lambda} + k\mathbf{I}) - \boldsymbol{\Lambda} - k^2 \boldsymbol{\alpha} \boldsymbol{\alpha}^\top] (\boldsymbol{\Lambda} + k\mathbf{I})^{-1} \\ &= (\boldsymbol{\Lambda} + k\mathbf{I})^{-1} [k^2 \boldsymbol{\Lambda}^{-1} + 2k\mathbf{I} - k^2 \boldsymbol{\alpha} \boldsymbol{\alpha}^\top] (\boldsymbol{\Lambda} + k\mathbf{I})^{-1} \end{aligned} \quad (4.15)$$

Burada $\mathbf{T} = k^2 \boldsymbol{\Lambda}^{-1} + 2k\mathbf{I}$ ve $\boldsymbol{\gamma}_1 = k\boldsymbol{\alpha}$ olsun. $\mathbf{T} = k(k\boldsymbol{\Lambda}^{-1} + 2\mathbf{I}) > \mathbf{0}$ dir. Denklem (4.15) aşağıdaki hale gelir

$$(\boldsymbol{\Lambda} + k\mathbf{I})^{-1} [\mathbf{T} - \boldsymbol{\gamma}_1 \boldsymbol{\gamma}_1^\top] (\boldsymbol{\Lambda} + k\mathbf{I})^{-1}$$

ve Lemma 3.1 kullanılarak $\Delta_1 \geq \mathbf{0}$ olması için gerek ve yeter şart

$$\begin{aligned} \boldsymbol{\gamma}_1^\top \mathbf{T}^{-1} \boldsymbol{\gamma}_1 &= k\boldsymbol{\alpha}^\top [k\boldsymbol{\Lambda}^{-1} + 2\mathbf{I}]^{-1} \boldsymbol{\alpha} \leq 1 \\ \boldsymbol{\alpha}^\top [k\boldsymbol{\Lambda}^{-1} + 2\mathbf{I}]^{-1} \boldsymbol{\alpha} &\leq \frac{1}{k}. \end{aligned}$$

olarak elde edilir. ■

Teorem 4.2 *Tobit regresyon modelinde $\forall j = 1, 2, \dots, p$ için*

Eğer $1 - \lambda_j \alpha_j^2 < 0$ ise MSE kriterine göre ridge tahmin edicisi en çok olabilirlik tahmin edicisinden üstündür.

Eğer $1 - \lambda_j \alpha_j^2 > 0$ ise MSE kriterine göre $\min\left(\frac{2\lambda_j}{\lambda_j \alpha_j^2 - 1}\right) > k > 0$ olacak şekilde k yanlılık parametresine sahip ridge tahmin edicisi en çok olabilirlik tahmin edicisinden üstündür.

İspat MLE ve RR tahmin edicilerinin MSE fonksiyonları arasındaki fark

$$\begin{aligned} \text{MSE}(\hat{\alpha}) - \text{MSE}(\hat{\alpha}_{\text{RR}}) &= \sum_{j=1}^p \frac{1}{\lambda_j} - \sum_{j=1}^p \frac{\lambda_j}{(\lambda_j + k)^2} - k^2 \sum_{j=1}^p \frac{\alpha_j^2}{(\lambda_j + k)^2} \\ &= \sum_{j=1}^p \frac{(\lambda_j + k)^2 - \lambda_j (\lambda_j + k^2 \alpha_j^2)}{\lambda_j (\lambda_j + k)^2} \\ &= k \sum_{j=1}^p \frac{2\lambda_j + k (1 - \lambda_j \alpha_j^2)}{\lambda_j (\lambda_j + k)^2} > 0 \end{aligned}$$

Burada k 'yı alnız bırakmak zordur. Toplam sembolünün içindeki her bir j . eleman için

$$\frac{2\lambda_j + k (1 - \lambda_j \alpha_j^2)}{\lambda_j (\lambda_j + k)^2} > 0 \iff 2\lambda_j > k (-1 + \lambda_j \alpha_j^2) \quad (4.16)$$

olduğundan, eğer $1 - \lambda_j \alpha_j^2 < 0$ ise her $k > 0$ değeri için MSE fonksiyonları arasındaki fark pozitifdir. Eğer $1 - \lambda_j \alpha_j^2 > 0$ ise son eşitsizlikteki k yalnız bırakılarak $\min\left(\frac{2\lambda_j}{\lambda_j \alpha_j^2 - 1}\right) > k$ elde edilir ve böylece ispat bitmiş olur. ■

Teorem 4.3 Tobit regresyon modelinde ridge tahmincisinin toplam varyans fonksiyonu olan $\kappa_1(k)$, k 'nin sürekli ve monoton azalan bir fonksiyonudur.

İspat $\kappa_1(k) = \sum_{j=1}^p \frac{\lambda_j}{(\lambda_j + k)^2}$ fonksiyonunun k göre birinci türevi

$$\frac{d\kappa_1(k)}{dk} = -2 \sum_{j=1}^p \frac{\lambda_j}{(\lambda_j + k)^3} < 0$$

bulunur. Dolayısıyla $\kappa_1(k)$ azalan bir fonksiyondur. Ayrıca her $k_1 > k_2$ için

$$\begin{aligned} \kappa_1(k_1) - \kappa_1(k_2) &= \sum_{j=1}^p \frac{\lambda_j}{(\lambda_j + k_1)^2} - \sum_{j=1}^p \frac{\lambda_j}{(\lambda_j + k_2)^2} \\ &= \sum_{j=1}^p \frac{\lambda_j ((\lambda_j + k_2)^2 - (\lambda_j + k_1)^2)}{(\lambda_j + k_1)^2 (\lambda_j + k_2)^2} \\ &= (k_2 - k_1) \sum_{j=1}^p \frac{\lambda_j (2\lambda_j + k_1 + k_2)}{(\lambda_j + k_1)^2 (\lambda_j + k_2)^2} < 0 \end{aligned}$$

olur. $\kappa_1(k_1) < \kappa_1(k_2)$ olduğundan $\kappa_1(k)$ fonksiyonu monoton azalan bir fonksiyondur. ■

Teorem 4.4 *Tobit regresyon modelinde ridge tahmincisinin karesel yan fonksiyonu olan $\kappa_2(k)$, k 'nin sürekli ve monoton artan bir fonksiyonudur.*

İspat $\kappa_2(k) = k^2 \sum_{j=1}^p \frac{\alpha_j^2}{(\lambda_j+k)^2}$ fonksiyonunun k göre birinci türevi

$$\frac{d\kappa_2(k)}{dk} = 2k \sum_{j=1}^p \frac{\lambda_j \alpha_j^2}{(\lambda_j + k)^3} > 0$$

bulunur. Dolayısıyla $\kappa_2(k)$ artan bir fonksiyondur. Ayrıca her $k_1 > k_2$ için

$$\begin{aligned} \kappa_2(k_1) - \kappa_2(k_2) &= k_1^2 \sum_{j=1}^p \frac{\alpha_j^2}{(\lambda_j + k_1)^2} - k_2^2 \sum_{j=1}^p \frac{\alpha_j^2}{(\lambda_j + k_2)^2} \\ &= \sum_{j=1}^p \frac{\alpha_j^2 (k_1^2 (\lambda_j + k_2)^2 - k_2^2 (\lambda_j + k_1)^2)}{(\lambda_j + k_1)^2 (\lambda_j + k_2)^2} \\ &= \sum_{j=1}^p \frac{\alpha_j^2 ((k_1 \lambda_j + k_1 k_2)^2 - (k_2 \lambda_j + k_1 k_2)^2)}{(\lambda_j + k_1)^2 (\lambda_j + k_2)^2} \\ &= (k_1 - k_2) \sum_{j=1}^p \frac{\alpha_j^2 \lambda_j (\lambda_j (k_1 + k_2) + 2k_1 k_2)}{(\lambda_j + k_1)^2 (\lambda_j + k_2)^2} > 0 \end{aligned}$$

olur. $\kappa_2(k_1) > \kappa_2(k_2)$ olduğundan $\kappa_2(k)$ fonksiyonu monoton artan bir fonksiyondur. ■

Teorem 4.5 *Tobit regresyon modelinde ridge tahmin edicisinin MSE kriterine göre en çok olabilirlik tahmin edicisinden üstün olduğu bir $k > 0$ değeri her zaman vardır.*

İspat Teorem (4.3) ve Teorem (4.4)'te, κ_1 ve κ_2 fonksiyonlarının sırasıyla monoton olarak azaldığı ve arttığı tespit edilmiştir. Bu iki fonksiyonun birinci dereceden türevine bakarak MSE fonksiyonunun birinci dereceden türevinin pozitif ya da negatif olması hakkında kesin bir bilgi vermek mümkün değildir. Bu nedenle teoremi kanıtlamak için, $\frac{d\text{MSE}(\hat{\alpha}_{\text{RR}})}{dk} < 0$ olacak şekilde her zaman bir $k > 0$ mevcut olduğunu göstermek gerekir, öyle ki

$$\begin{aligned} \frac{d\text{MSE}(\hat{\alpha}_{\text{RR}})}{dk} &= \frac{d\kappa_1(k)}{dk} + \frac{d\kappa_2(k)}{dk} \\ &= -2 \sum_{j=1}^p \frac{\lambda_j}{(\lambda_j + k)^3} + 2k \sum_{j=1}^p \frac{\lambda_j \alpha_j^2}{(\lambda_j + k)^3} \\ &= 2 \sum_{j=1}^p \frac{\lambda_j (-1 + k \alpha_j^2)}{(\lambda_j + k)^3} < 0 \\ \implies k &< \frac{1}{\alpha_{\max}^2} \end{aligned} \tag{4.17}$$

$k < \frac{1}{\alpha_{\max}^2}$ için türev negatif olduğundan $\hat{\alpha}_{\text{RR}}$ tahmin edicisinin MSE fonksiyonu azalandır. $k = 0$ için $\hat{\alpha}_{\text{RR}}$ tahmin edicisinin MSE değeri $\hat{\alpha}$ tahmin edicisine eşit olduğundan $k < \frac{1}{\alpha_{\max}^2}$ olacak şekilde bir k değeri vardır ki $\text{MSE}(\hat{\alpha}_{\text{RR}}) < \text{MSE}(\hat{\alpha})$ olur. ■

4.2. k Yanlılık Parametresinin Tahmini

Yanlı tahmin edicilerde yanlılık parametrelerinin seçimi her zaman önemli konudur. Tobit ridge regresyon yöntemi ile yapılacak çıkarımların kararlılığı, k yanlılık parametresinin optimum değerinin belirlenmesine bağlıdır. Uygun k parametresinin belirlenmesi durumunda RR tahmin edicisinin MSE'si MLE tahmin edicisinin MSE değerinden daha küçük olacaktır. Kısaca çoklu bağlantı probleminin varlığında RR tahmin edicisi ile yapılacak çıkarımların kararlılığı, yanlılık parametresi k için optimum değerinin belirlenmesine bağlıdır. Literatürde ridge regresyon tahmin edicisinin yanlılık parametresi k 'nın belirlenmesi için pek çok metot geliştirilmiştir. Bu metotlardan faydalanarak Khalaf ve ark. (2014) tobit model için yedi farklı k parametresi önermişlerdir.

İlk olarak, Hoerl ve Kennard (1970) çalışmasında bulunan klasik k tahmin edicisi tobit model için

$$\hat{k}_{HKL} = \frac{s^2}{\hat{\alpha}_{max}^2} \quad (4.18)$$

şeklinde önerilmiştir. Kibria (2003) çalışmasından yararlanarak \hat{k}_{GM} ve \hat{k}_{MED} tahmin edicilerini

$$\hat{k}_{GM} = \frac{s^2}{\left(\prod_{j=1}^p \hat{\alpha}_j^2\right)^{1/p}} \quad (4.19)$$

$$\hat{k}_{MED} = median\left(\frac{s^2}{\hat{\alpha}_j^2}\right) \quad (4.20)$$

olarak önermişlerdir. Muniz ve Kibria (2009) çalışmasından yararlanarak \hat{k}_{KM2} , \hat{k}_{KM4} , \hat{k}_{KM6} tahmin edicilerini

$$\hat{k}_{KM2} = max\left(\sqrt{\frac{\hat{\alpha}_j^2}{s^2}}\right) \quad (4.21)$$

$$\hat{k}_{KM4} = \left(\prod_{j=1}^p \sqrt{\frac{\hat{\alpha}_j^2}{s^2}}\right)^{1/p} \quad (4.22)$$

$$\hat{k}_{KM6} = median\left(\sqrt{\frac{\hat{\alpha}_j^2}{s^2}}\right) \quad (4.23)$$

önermişlerdir. Alkhamisi ve ark. (2006) çalışmasından esinlenerek \hat{k}_{max}^{KS}

$$\hat{k}_{max}^{KS} = max\left(\frac{\lambda_j s^2}{(n-p)s^2 + \lambda_j \hat{\alpha}_j^2}\right) \quad (4.24)$$

tahmin edicisini sunmuşlardır.

4.2.1. k Parametresi için Yeni Önerilen Tahmin Ediciler

Bu tez çalışmasında yukarıda anlatılan yöntemlere benzer şekilde tobit modelde k parametresinin optimal değerini tahmin etmek için Hoerl ve ark. (1975), Lawless ve Wang (1976), Hocking ve ark. (1976), Kibria (2003), Muniz ve Kibria (2009), Khalaf ve Shukur (2005) çalışmalarını takip ederek aşağıdaki tahmin ediciler önerilmiştir:

Hoerl ve ark. (1975)'in ardından tobit modelindeki ridge tahmin edicisi için \hat{k}_1 tahmin edicisi

$$\hat{k}_1 = \frac{ps^2}{\sum_{j=1}^p \hat{\alpha}_j^2}. \quad (4.25)$$

olarak önerilmiştir.

Bayeşçi yaklaşımla Lawless ve Wang (1976) doğrusal modelde farklı bir k tahmin edicisi sunmuşlardır. Benzer şekilde tobit regresyon modeli için bu tahmin edici

$$\hat{k}_2 = \frac{ps^2}{\sum_{j=1}^p \lambda_j \hat{\alpha}_j^2} \quad (4.26)$$

olarak geliştirilmiştir.

Hocking ve ark. (1976) çalışmasının geliştirilmesi olarak \hat{k}_3 tahmin edicisi ise

$$\hat{k}_3 = s^2 \frac{\sum_{j=1}^p (\lambda_j \hat{\alpha}_j)^2}{\left(\sum_{j=1}^p \lambda_j \hat{\alpha}_j^2\right)^2} \quad (4.27)$$

şeklinde önerilmiştir.

Kibria (2003) tarafından aritmetik ortalama kullanılarak önerilen k tahmin edicisi tobit regresyon modeli için

$$\hat{k}_4 = \frac{1}{p} \sum_{j=1}^p \frac{s^2}{\hat{\alpha}_j^2} \quad (4.28)$$

şeklinde uyarlanmıştır.

\hat{k}_5 tahmin edicisi, Khalaf ve Shukur (2005) çalışmasından faydalanarak tobit model için aşağıdaki gibi önerilmiştir

$$\hat{k}_5 = \frac{\lambda_{max} s^2}{(n-p)s^2 + \lambda_{max} \hat{\alpha}_{max}^2} \quad (4.29)$$

öyle ki, burada λ_{max} , $\mathbf{X}^T \widehat{\mathbf{W}} \mathbf{X}$ matrisinin öz değerleri içinde en büyük olanıdır.

Ayrıca \hat{k}_6 , \hat{k}_7 , \hat{k}_8 ve \hat{k}_9 tahmin edicileri Muniz ve Kibria (2009) çalışmasından esinlenerek sırasıyla aşağıdaki şekilde önerilmiştir

$$\hat{k}_6 = \frac{1}{p} \sum_{j=1}^p \left(\frac{\lambda_j s^2}{(n-p)s^2 + \lambda_j \hat{\alpha}_j^2} \right) \quad (4.30)$$

$$\hat{k}_7 = \max \left(\sqrt{\frac{s^2}{\hat{\alpha}_j^2}} \right) \quad (4.31)$$

$$\hat{k}_8 = \left(\prod_{j=1}^p \sqrt{\frac{s^2}{\hat{\alpha}_j^2}} \right)^{1/p} \quad (4.32)$$

$$\hat{k}_9 = \text{median} \left(\sqrt{\frac{s^2}{\hat{\alpha}_j^2}} \right) \quad (4.33)$$

4.2.2. Genelleştirilmiş Tobit Ridge Tahmin Edicisi için k_j 'lerin Seçimi

Genelleştirilmiş tobit ridge tahmin edicisinde kullanılan k_j parametrelerini tahmin etmek için GRR'nin MSE fonksiyonu

$$\begin{aligned} \text{MSE}(\hat{\alpha}_{\text{GRR}}) &= \sum_{j=1}^p \frac{\lambda_j}{(\lambda_j + k_j)^2} + \sum_{j=1}^p \frac{k_j^2 \alpha_j^2}{(\lambda_j + k_j)^2} \\ &= \sum_{j=1}^p \frac{\lambda_j + k_j^2 \alpha_j^2}{(\lambda_j + k_j)^2} \end{aligned} \quad (4.34)$$

$\text{MSE}(\hat{\alpha}_{\text{GRR}}) = g(k_j)$ olsun. Her bir k_j 'nin değerleri uygun bir şekilde seçilmelidir. Böylece $g(k_j)$ fonksiyonu minimize edilmeye çalışılır. Bu nedenle, $g(k_j)$ fonksiyonunun k_j 'ye göre kısmi türevleri sıfıra eşitlenerek k_j 'nin optimal değeri elde edilebilir

$$\frac{\partial g(k_j)}{\partial k_j} = 2 \sum_{j=1}^p \frac{\lambda_j (k_j \alpha_j^2 - 1)}{(\lambda_j + k_j)^3} = 0. \quad (4.35)$$

Denklem (4.35)'den k_j değerlerini direk bulmak mümkün değildir. Bu yüzden toplam sembolündeki her bir eleman sıfıra eşitlenir

$$\begin{aligned} \frac{\lambda_j (k_j \alpha_j^2 - 1)}{(\lambda_j + k_j)^3} &= 0 \\ k_j &= \frac{1}{\alpha_j^2} \end{aligned} \quad (4.36)$$

elde edilir. Uygulamada ise $\hat{\alpha}$ yansız tahmin edicisi kullanılarak aşağıdaki tahmin edici elde edilir

$$\hat{k}_j = \frac{1}{\hat{\alpha}_j^2}. \quad (4.37)$$

4.3. Monte Carlo Simülasyon Çalışması

Bu bölümün amacı, farklı derecelerde çoklu bağlantı probleminin varlığında tahmin edicilerin performanslarını değerlendirmektir. Simülasyon çalışmasının ayrıntıları bu bölümde verilmiştir. Simülasyon tasarlamada etkili faktörler bağımsız değişken sayısı p , örneklemin gözlem sayısı n ve değişkenler arasındaki ilişkinin derecesi ρ^2 'dir. Çoklu bağlantı derecesine karşılık gelen ilişki katsayısının değerleri $\rho = 0.80, 0.90, 0.99$ olarak seçilmiştir. Ayrıca, simülasyonda örneklemin gözlem sayısının $n = 100, 250, 500$ olacak şekilde üç farklı değeri göz önünde bulundurulmuştur. Son olarak simülasyon tasarımı bağımsız değişken sayısı $p = 4, 8, 16, 32$ 'den oluşan tobit modeli dikkate alınmıştır. Tahmin edicilerin performanslarını karşılaştırmak için MSE kriteri kullanılmıştır.

4.3.1. Simülasyon Tasarımı

McDonald ve Galarneau (1975)'in çalışması takip edilerek farklı derecelerde çoklu bağlantıya sahip veri matrisi üretebilmek için

$$x_{ij} = (1 - \rho^2)^{1/2} w_{ij} + \rho w_{ip+1}, \quad i = 1, 2, \dots, n, \quad j = 1, 2, \dots, p \quad (4.38)$$

eşitliğini kullanılmıştır. Öyle ki burada w_{ij} standart normal dağılımından üretilen rasgele sayılardır ve herhangi iki bağımsız değişken arasındaki korelasyonun derecesini ρ^2 belirtir.

Gerçek parametre değerleri, alanda yaygın olarak kullanılan bir kısıtlama olan

$$\sum_{j=1}^p \beta_j^2 = 1$$

olacak şekilde seçilmiştir (bkz. Kibria (2003)).

Bağımlı değişkene ait n tane gözlem

$$y_i^* = \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + u_i, \quad i = 1, \dots, n \quad (4.39)$$

eşitliği kullanılarak üretilmiştir. Burada u_i ' ler, $N(0, \sigma^2)$ dağılımından üretilen rasgele hatalardır. Hataların varyansının etkisini araştırmak için σ 'nın değeri 1 ve 5 olarak alınmıştır. Sansürlenmiş bağımlı değişkeni elde etmek için

$$y_i = \begin{cases} y_i^* & \text{ise } y_i^* > 0 \\ 0 & \text{ise } y_i^* \leq 0 \end{cases}$$

kullanılmıştır.

Herbir adımda yeni rasgele hata terimleri üretilerek simülasyon 5000 kez tekrarlanmıştır. $\hat{\beta}^*$ tahmincisinin simüle edilmiş MSE'si

$$MSE(\hat{\beta}^*) = \frac{1}{5000} \sum_{r=1}^{5000} (\hat{\beta}^* - \beta)_r^\top (\hat{\beta}^* - \beta)_r \quad (4.40)$$

eşitliğinden yararlanılarak hesaplanmıştır. Burada $(\hat{\beta}^* - \beta)_r$ simülasyonun r . yinelenmesinde elde edilen parametre tahmini ve gerçek parametre vektörleri arasındaki farkı göstermektedir. Tüm hesaplamalar R programlama dili (R Development Core Team, 2019) kullanılarak gerçekleştirilmiştir.

4.3.2. Simülasyon Sonuçları

Monte Carlo simülasyon çalışmasından elde edilen, tahmin edicilere ait MSE değerleri Tablo 4.1– 4.7'de verilmiştir.

Tablo 4.1. $\sigma = 1$, $p = 4$ olduğunda farklı ρ değerleri için MLE, GRR ve RR tahmin edicilerin MSE değerleri

n	MLE	GRR	RR						
			k_{GM}	k_{MED}	k_{KM2}	k_1	k_7	k_8	k_9
$\rho = 0.80$									
100	0.264	0.500	0.098	0.142	0.235	0.110	0.107	0.130	0.132
250	0.078	0.528	0.061	0.124	0.075	0.049	0.056	0.051	0.049
500	0.040	0.463	0.079	0.172	0.039	0.030	0.054	0.030	0.029
$\rho = 0.90$									
100	0.414	0.582	0.093	0.123	0.317	0.158	0.107	0.166	0.169
250	0.135	0.509	0.056	0.101	0.126	0.067	0.063	0.075	0.072
500	0.080	0.529	0.037	0.078	0.077	0.049	0.042	0.051	0.048
$\rho = 0.99$									
100	5.654	1.610	0.297	0.295	0.529	0.782	0.192	0.433	0.452
250	1.556	0.902	0.106	0.137	0.657	0.233	0.116	0.263	0.292
500	0.798	0.737	0.065	0.086	0.523	0.146	0.086	0.197	0.215

Buna göre Tablo 4.1'den aşağıdaki sonuçlara ulaşılabilir.

- $\rho = 0.99$ için GRR'nin MSE değeri, MLE tahmin edicisinin MSE değerinden daha küçüktür. $\rho = 0.80$ ve 0.90 değerleri için GRR tahmin edicisinin MSE değeri, MLE tahmin edicisinin MSE değerinden daha büyüktür. Yani çoklu bağlantı derecesinin yüksek olduğu durumda GRR tahmin edicisi, MLE tahmin edicisine göre daha iyi performans sergilemektedir.
- k_{GM} ve k_7 kullanılarak elde edilen RR tahmin edicisinin MSE değerleri birbirine yakın sonuçlara sahiptir.

- $\rho = 0.80$ için gözlem sayısı (n) arttıkça k_9 kullanılarak elde edilen RR tahmin edicisi diğer tahmin edicilere göre daha iyi performans sergilemektedir.
- $\rho = 0.90$ ve 0.99 değerleri için k_{GM} ve k_7 değerlerine sahip RR tahmin edicisi diğer tahmin edicilere göre iyi performans göstermektedir.

Tablo 4.2. $\sigma = 1$, $p = 8$ olduğunda farklı ρ değerleri için MLE, GRR ve RR tahmin edicilerin MSE değerleri

n	MLE	GRR	RR						
			k_{GM}	k_{MED}	k_{KM2}	k_1	k_7	k_8	k_9
$\rho = 0.80$									
100	0.449	0.566	0.099	0.105	0.408	0.132	0.128	0.179	0.186
250	0.180	0.619	0.144	0.158	0.173	0.066	0.111	0.085	0.085
500	0.090	0.594	0.106	0.129	0.089	0.043	0.067	0.050	0.049
$\rho = 0.90$									
100	1.248	0.903	0.095	0.097	0.948	0.116	0.130	0.225	0.246
250	0.380	0.756	0.046	0.051	0.350	0.073	0.065	0.122	0.126
500	0.174	0.663	0.050	0.060	0.168	0.054	0.055	0.075	0.075
$\rho = 0.99$									
100	11.086	3.124	0.154	0.229	1.577	0.744	0.113	0.556	0.681
250	4.378	1.821	0.063	0.108	1.563	0.290	0.079	0.373	0.464
500	2.128	1.324	0.028	0.043	1.247	0.129	0.056	0.257	0.300

Tablo 4.2'den aşağıda listelenen sonuçlara ulaşılmıştır:

- GRR tahmin edicisi Tablo 4.1'dekine benzer performans sergilemektedir.
- $\rho = 0.80$ için gözlem sayısı (n) arttıkça k_1 kullanılarak elde edilen RR tahmin edicisi diğer tahmin edicilerden daha üstün performans sergilemektedir.
- $\rho = 0.90$ ve 0.99 değerleri için k_{GM} kullanılarak elde edilen RR tahmin edicisi genel olarak daha üstündür.

Tablo 4.3. $\sigma = 1$, $p = 16$ olduğunda farklı ρ değerleri için MLE, GRR ve RR tahmin edicilerin MSE değerleri

n	MLE	GRR	RR						
			k_{GM}	k_{MED}	k_{KM2}	k_1	k_7	k_8	k_9
$\rho = 0.80$									
100	1.244	1.148	0.107	0.091	1.099	0.114	0.153	0.285	0.314
250	0.396	0.889	0.154	0.125	0.386	0.039	0.117	0.119	0.129
500	0.192	0.872	0.133	0.113	0.189	0.047	0.092	0.083	0.087
$\rho = 0.90$									
100	2.607	1.641	0.053	0.051	2.054	0.125	0.117	0.383	0.447
250	0.835	1.302	0.066	0.053	0.766	0.052	0.091	0.181	0.202
500	0.412	1.093	0.063	0.050	0.397	0.043	0.073	0.124	0.134
$\rho = 0.99$									
100	30.712	7.396	0.089	0.163	4.708	0.921	0.079	0.865	1.124
250	10.610	3.929	0.031	0.060	3.301	0.335	0.047	0.510	0.651
500	4.934	2.835	0.011	0.018	2.494	0.110	0.036	0.325	0.403

Tablo 4.3'ten elde edilen sonuçlar aşağıda listelenmiştir:

- Çoklu bağlantı derecesinin yüksek veya gözlem sayısı 100 olduğu durumlarda GRR tahmin edicisi, MLE tahmin edicisine göre daha iyi performans sergilemektedir.
- $n = 100$ için k_{MED} , çoklu bağlantı derecesi yüksek olduğunda k_7 değerlerine sahip RR tahmin edicisi üstündür.
- $n = 250$ ve 500 için k_1 , çoklu bağlantı derecesi yüksek olduğunda k_{GM} tahmini değerlerine sahip RR tahmin edicisinin performansı daha iyi çıkmaktadır.

Tablo 4.4. $\sigma = 1$, $p = 32$ olduğunda farklı ρ değerleri için MLE, GRR ve RR tahmin edicilerin MSE değerleri

n	MLE	GRR	RR						
			k_{GM}	k_{MED}	k_{KM2}	k_1	k_7	k_8	k_9
$\rho = 0.80$									
250	0.941	1.832	0.101	0.069	0.897	0.048	0.129	0.232	0.260
500	0.440	1.555	0.150	0.104	0.432	0.024	0.111	0.131	0.146
$\rho = 0.90$									
250	2.225	2.467	0.043	0.029	1.974	0.040	0.091	0.310	0.366
500	0.928	2.229	0.052	0.034	0.880	0.024	0.077	0.188	0.215
$\rho = 0.99$									
250	26.699	8.284	0.017	0.036	8.344	0.353	0.035	0.736	0.976
500	11.799	5.916	0.007	0.013	5.308	0.130	0.030	0.487	0.634

Tablo 4.4'ten elde edilen sonuçlar aşağıda sıralanmıştır:

- GRR tahmin edicisi Tablo 4.3'tekine benzer performans sergilemektedir.
- k_1 , k_{MED} ve k_{GM} kullanılarak elde edilen RR tahmin edicisi MSE kriterine göre daha iyi performans sergilemiştir.

Tablo 4.5'ten elde edilen sonuçlar Tablo 4.1'den elde edilen sonuçlara benzemektedir.

Tablo 4.5'ten elde edilen sonuçların Tablo 4.1'den elde edilen sonuçlardan farklı olarak:

- $n = 100$ için GRR tahmin edicisi, MLE tahmin edicisine göre daha iyi durumdadır.
- $\rho = 0.80$ ve 0.90 için k_1 parametresine sahip RR tahmin edicisi en iyi performans göstermektedir.

Tablo 4.6'dan elde edilen sonuçlar Tablo 4.2'den elde edilen sonuçlara benzemektedir. Tablo 4.6'dan elde edilen sonuçların Tablo 4.2'den elde edilen sonuçlardan farklı olarak:

- $n = 100$ ve 250 için GRR tahmin edicisinin performansı, MLE tahmin edicisine göre iyileşme artmıştır.

- k_1 parametresinin k_{GM} parametresinden performansı üstün olduğu durum sayısı artmıştır.

Tablo 4.7 ve Tablo 4.8 göre:

- $\rho = 0.80$, $n = 500$ durumu hariç GRR tahmin edicisi, MLE tahmin edicisine göre daha iyi performans sergilemektedir.
- k_1, k_{GM} ve k_{MED} parametrelerine sahip RR tahmin edicisinin MSE değerleri en küçük sonuçlara sahiptir.

Tablo 4.1- 4.8 göre:

- MLE'nin MSE değerleri gözlem sayısı artış olduğunda hızla azalır. Benzer şekilde k_1, k_7, k_8, k_9 parametrelerine sahip RR tahmin edicisinin MSE değerlerinde azalma görülür.
- Çoklu bağlantı derecesindeki artışın yansız tahmin edici olan MLE'nin MSE değerinde artmasına neden olur. Bu zaten beklenen durumdur. Derecedeki artış k_8 ve k_9 parametrelerine sahip RR tahmin edicisinin MSE değerini de artırmaktadır.
- Bağımsız değişken sayısındaki artışın MLE, GRR ve k_{KM2}, k_8, k_9 parametrelerine sahip RR tahmin edicilerin MSE değerlerinde artış sağladığı açıktır.
- Farklı k parametrelere sahip RR tahmin edicisinin MSE değeri, her zaman MLE'nin MSE değerinden küçüktür.

Tablo 4.5. $\sigma = 5$, $p = 4$ olduğunda farklı ρ değerleri için MLE, GRR ve RR tahmin edicilerin MSE değerleri

n	MLE	GRR	RR							
			k_{GM}	k_{MED}	k_{KM2}	k_1	k_7	k_8	k_9	
$\rho = 0.80$										
100	0.929	0.638	0.215	0.245	0.676	0.121	0.201	0.194	0.216	
250	0.315	0.592	0.186	0.246	0.285	0.070	0.142	0.115	0.121	
500	0.154	0.517	0.141	0.214	0.146	0.057	0.102	0.080	0.079	
$\rho = 0.90$										
100	1.681	0.885	0.158	0.185	0.891	0.141	0.177	0.222	0.252	
250	0.672	0.760	0.126	0.166	0.535	0.087	0.134	0.166	0.184	
500	0.303	0.615	0.105	0.153	0.273	0.063	0.097	0.108	0.111	
$\rho = 0.99$										
100	19.219	3.611	0.305	0.291	0.220	1.103	0.146	0.323	0.317	
250	7.010	1.653	0.152	0.146	0.624	0.405	0.141	0.281	0.286	
500	3.859	1.302	0.094	0.109	0.798	0.234	0.116	0.241	0.263	

Tablo 4.6. $\sigma = 5$, $p = 8$ olduğunda farklı ρ değerleri için MLE, GRR ve RR tahmin edicilerin MSE değerleri

n	MLE	GRR	RR						
			k_{GM}	k_{MED}	k_{KM2}	k_1	k_7	k_8	k_9
$\rho = 0.80$									
100	1.720	0.925	0.209	0.200	1.327	0.159	0.243	0.301	0.353
250	0.739	0.719	0.238	0.218	0.668	0.088	0.211	0.183	0.202
500	0.357	0.629	0.181	0.178	0.338	0.061	0.141	0.120	0.125
$\rho = 0.90$									
100	3.863	1.540	0.133	0.141	1.856	0.188	0.188	0.332	0.409
250	1.654	1.007	0.152	0.138	1.202	0.082	0.164	0.200	0.234
500	0.816	0.849	0.105	0.102	0.703	0.059	0.112	0.155	0.170
$\rho = 0.99$									
100	45.442	7.620	0.206	0.255	0.662	1.510	0.106	0.481	0.539
250	18.259	3.754	0.089	0.116	1.478	0.536	0.092	0.368	0.442
500	8.819	2.518	0.061	0.096	1.903	0.295	0.082	0.346	0.444

Tablo 4.7. $\sigma = 5$, $p = 16$ olduğunda farklı ρ değerleri için MLE, GRR ve RR tahmin edicilerin MSE değerleri

n	MLE	GRR	RR						
			k_{GM}	k_{MED}	k_{KM2}	k_1	k_7	k_8	k_9
$\rho = 0.80$									
100	4.237	1.687	0.237	0.187	3.113	0.134	0.300	0.383	0.470
250	1.725	1.209	0.165	0.128	1.513	0.072	0.194	0.263	0.309
500	0.859	1.026	0.189	0.147	0.799	0.052	0.176	0.183	0.207
$\rho = 0.90$									
100	10.378	3.054	0.133	0.137	4.079	0.335	0.256	0.582	0.744
250	3.672	1.991	0.127	0.096	2.395	0.093	0.192	0.306	0.374
500	1.811	1.430	0.087	0.066	1.495	0.060	0.133	0.245	0.289
$\rho = 0.99$									
100	116.849	21.158	0.243	0.403	1.166	2.708	0.103	0.858	1.080
250	49.036	9.053	0.062	0.101	2.984	0.835	0.081	0.523	0.681
500	19.608	5.129	0.031	0.061	4.290	0.303	0.063	0.437	0.608

Tablo 4.8. $\sigma = 5$, $p = 32$ olduğunda farklı ρ değerleri için MLE, GRR ve RR tahmin edicilerin MSE değerleri

n	MLE	GRR	RR						
			k_{GM}	k_{MED}	k_{KM2}	k_1	k_7	k_8	k_9
$\rho = 0.80$									
100	17.034	4.511	0.153	0.128	9.000	0.305	0.302	0.899	1.102
250	4.267	2.202	0.175	0.120	3.616	0.064	0.257	0.399	0.496
500	1.915	2.088	0.176	0.119	1.721	0.042	0.208	0.276	0.332
$\rho = 0.90$									
100	27.558	7.173	0.057	0.058	11.499	0.293	0.180	0.865	1.112
250	9.782	4.098	0.064	0.047	5.699	0.103	0.175	0.501	0.644
500	4.041	3.093	0.087	0.056	3.022	0.041	0.158	0.322	0.403
$\rho = 0.99$									
100	469.565	75.369	0.198	0.325	5.281	5.205	0.132	1.803	2.240
250	99.702	19.489	0.025	0.046	8.616	0.510	0.082	0.579	0.810
500	51.787	12.144	0.014	0.029	9.427	0.304	0.062	0.499	0.719

5. TOBİT LIU REGRESYON TAHMİN EDİCİSİ

Doğrusal regresyon modelinde çoklu bağlantı probleminin bir başka çözümü olarak Liu (1993) tarafından Liu tahmin edicisi

$$\hat{\beta}_d = (\mathbf{X}^\top \mathbf{X} + \mathbf{I})^{-1} (\mathbf{X}^\top \mathbf{X} + d\mathbf{I}) \hat{\beta}_{OLS}$$

olarak önerilmiştir. Burada d yanlılık veya küçültme (shrinkage) parameteresi $0 < d < 1$ kısıtını sağlamalıdır. Bu çalışmanın amacı, doğrusal regresyon modeli için önerilen Liu tahmin edicisinin tobit regresyon modeline bir genelleştirilmesi olan bir tobit Liu tahmin edicisini önermektir. Tobit Liu tahmin edicisi (LE)

$$\hat{\beta}_{LE} = (\mathbf{X}^\top \widehat{\mathbf{W}}\mathbf{X} + \mathbf{I})^{-1} (\mathbf{X}^\top \widehat{\mathbf{W}}\mathbf{X} + d\mathbf{I}) \hat{\beta}_{MLE} \quad (5.1)$$

şeklinde tanımlanabilir, öyle ki burada $0 < d < 1$.

5.1. Tobit Liu Tahmin Edicisinin Teorik Özellikleri

$\hat{\beta}_{LE}$ 'nin kanonik formu aşağıda verilen şekilde yazılabilir

$$\hat{\alpha}_{LE} = (\mathbf{\Lambda} + \mathbf{I})^{-1} (\mathbf{\Lambda} + d\mathbf{I}) \hat{\alpha}. \quad (5.2)$$

Tahmin edicilerin MMSE ve MSE fonksiyonlarını hesaplamadan önce, yan ve kovaryans fonksiyonları elde edilmiştir. Tobit Liu tahmin edicisinin yan fonksiyonunu elde etmek için ilk başta tahmin edicinin beklenen değeri

$$\begin{aligned} E[\hat{\alpha}_{LE}] &= E[(\mathbf{\Lambda} + \mathbf{I})^{-1} (\mathbf{\Lambda} + d\mathbf{I}) \hat{\alpha}] \\ &= (\mathbf{\Lambda} + \mathbf{I})^{-1} (\mathbf{\Lambda} + d\mathbf{I}) E[\hat{\alpha}] \\ &= (\mathbf{\Lambda} + \mathbf{I})^{-1} (\mathbf{\Lambda} + d\mathbf{I}) \alpha \end{aligned} \quad (5.3)$$

olarak elde edilmiştir. Beklenen değeri (5.3)'te bulunan LE tahmin edicisinin yanı

$$\begin{aligned} bias(\hat{\alpha}_{LE}) &= E(\hat{\alpha}_{LE}) - \alpha \\ &= (\mathbf{\Lambda} + \mathbf{I})^{-1} (\mathbf{\Lambda} + d\mathbf{I}) \alpha - \alpha \\ &= (\mathbf{\Lambda} + \mathbf{I})^{-1} [(\mathbf{\Lambda} + \mathbf{I}) - (1 - d)\mathbf{I}] \alpha - \alpha \\ &= [\mathbf{I} - (1 - d)(\mathbf{\Lambda} + \mathbf{I})^{-1}] \alpha - \alpha \\ &= -(1 - d)(\mathbf{\Lambda} + \mathbf{I})^{-1} \alpha \end{aligned} \quad (5.4)$$

olarak elde edilir. Kovaryans fonksiyonunun özellikleri ve (3.10) eşitliğinden yararlanarak LE tahmin edicisinin varyans–kovaryans matrisi

$$\begin{aligned}
Cov(\hat{\boldsymbol{\alpha}}_{LE}) &= Cov[(\boldsymbol{\Lambda} + \mathbf{I})^{-1}(\boldsymbol{\Lambda} + d\mathbf{I})\hat{\boldsymbol{\alpha}}] \\
&= (\boldsymbol{\Lambda} + \mathbf{I})^{-1}(\boldsymbol{\Lambda} + d\mathbf{I})Cov[\hat{\boldsymbol{\alpha}}](\boldsymbol{\Lambda} + d\mathbf{I})(\boldsymbol{\Lambda} + \mathbf{I})^{-1} \\
&= (\boldsymbol{\Lambda} + \mathbf{I})^{-1}(\boldsymbol{\Lambda} + d\mathbf{I})\boldsymbol{\Lambda}^{-1}(\boldsymbol{\Lambda} + d\mathbf{I})(\boldsymbol{\Lambda} + \mathbf{I})^{-1} \quad (5.5)
\end{aligned}$$

şeklinde bulunur. LE tahmin edicisinin MMSE’si

$$\begin{aligned}
MMSE(\hat{\boldsymbol{\alpha}}_{LE}) &= Cov(\hat{\boldsymbol{\alpha}}_{LE}) + bias(\hat{\boldsymbol{\alpha}}_{LE})bias(\hat{\boldsymbol{\alpha}}_{LE})^T \\
&= Cov(\hat{\boldsymbol{\alpha}}_{LE}) + (1-d)^2(\boldsymbol{\Lambda} + \mathbf{I})^{-1}\boldsymbol{\alpha}\boldsymbol{\alpha}^T(\boldsymbol{\Lambda} + \mathbf{I})^{-1} \quad (5.6)
\end{aligned}$$

ve MSE fonksiyonu

$$\begin{aligned}
MSE(\hat{\boldsymbol{\alpha}}_{LE}) &= tr(MMSE(\hat{\boldsymbol{\alpha}}_{LE})) \\
&= \sum_{j=1}^{p+1} \frac{(\lambda_j + d)^2}{\lambda_j(\lambda_j + 1)^2} + (1-d)^2 \sum_{j=1}^{p+1} \frac{\alpha_j^2}{(\lambda_j + 1)^2} \quad (5.7)
\end{aligned}$$

olarak bulunur.

Aşağıdaki teoremlerde LE ve MLE tahmin edicilerinin MMSE ve MSE kriterlerine göre karşılaştırmaları ele alınmıştır.

Teorem 5.1 *Tobit regresyon modelinde Liu tahmin edicisinin MMSE kriterine göre en çok olabilirlik tahmin edicisinden üstün olması için gerek ve yeter şart*

$$\boldsymbol{\alpha}^T[(1+d)\boldsymbol{\Lambda}^{-1} + 2\mathbf{I}]^{-1}\boldsymbol{\alpha} \leq \frac{1}{(1-d)}$$

eşitsizliğin sağlanmasıdır.

İspat MLE ve LE tahmin edicilerinin MMSE fonksiyonları arasındaki fark,

$$\Delta_2 = MMSE(\hat{\boldsymbol{\alpha}}) - MMSE(\hat{\boldsymbol{\alpha}}_{LE})$$

olsun, Δ_2 ise

$$\begin{aligned}
\Delta_2 &= \boldsymbol{\Lambda}^{-1} - (\boldsymbol{\Lambda} + \mathbf{I})^{-1}(\boldsymbol{\Lambda} + d\mathbf{I})\boldsymbol{\Lambda}^{-1}(\boldsymbol{\Lambda} + d\mathbf{I})(\boldsymbol{\Lambda} + \mathbf{I})^{-1} \\
&\quad - (1-d)^2(\boldsymbol{\Lambda} + \mathbf{I})^{-1}\boldsymbol{\alpha}\boldsymbol{\alpha}^T(\boldsymbol{\Lambda} + \mathbf{I})^{-1} \\
&= (\boldsymbol{\Lambda} + \mathbf{I})^{-1} [(\boldsymbol{\Lambda} + \mathbf{I})\boldsymbol{\Lambda}^{-1}(\boldsymbol{\Lambda} + \mathbf{I}) - (\boldsymbol{\Lambda} + d\mathbf{I})\boldsymbol{\Lambda}^{-1}(\boldsymbol{\Lambda} + d\mathbf{I}) - (1-d)^2\boldsymbol{\alpha}\boldsymbol{\alpha}^T] \\
&\quad \times (\boldsymbol{\Lambda} + \mathbf{I})^{-1} \\
&= (\boldsymbol{\Lambda} + \mathbf{I})^{-1} [(1-d^2)\boldsymbol{\Lambda}^{-1} + 2(1-d)\mathbf{I} - (1-d)^2\boldsymbol{\alpha}\boldsymbol{\alpha}^T] (\boldsymbol{\Lambda} + \mathbf{I})^{-1} \quad (5.8)
\end{aligned}$$

olarak düzenlenebilir, burada $\mathbf{K} = (1 - d^2)\mathbf{\Lambda}^{-1} + 2(1 - d)\mathbf{I}$ ve $\boldsymbol{\gamma}_2 = (1 - d)\boldsymbol{\alpha}$ olursa (5.8) aşağıdaki hale gelir

$$(\mathbf{\Lambda} + \mathbf{I})^{-1} [\mathbf{K} - \boldsymbol{\gamma}_2 \boldsymbol{\gamma}_2^\top] (\mathbf{\Lambda} + \mathbf{I})^{-1}.$$

Lemma 3.1 kullanılırsa, $\mathbf{\Delta}_2 \geq \mathbf{0}$ olması için gerek ve yeter şart

$$\begin{aligned} \boldsymbol{\gamma}_2^\top \mathbf{K}^{-1} \boldsymbol{\gamma}_2 &= (1 - d) \boldsymbol{\alpha}^\top [(1 + d)\mathbf{\Lambda}^{-1} + 2\mathbf{I}]^{-1} \boldsymbol{\alpha} \leq 1 \\ \iff \boldsymbol{\alpha}^\top [(1 + d)\mathbf{\Lambda}^{-1} + 2\mathbf{I}]^{-1} \boldsymbol{\alpha} &\leq \frac{1}{(1 - d)}. \end{aligned}$$

olarak elde edilir. Böylece, ispat bitmiştir. ■

Teorem 5.2 *Tobit regresyon modelinde Liu tahmin edicisinin MSE kriterine göre en çok olabilirlik tahmin edicisinden üstün olması için yanlılık parametresi d 'nin*

$$d > \max \left(1 - 2 \frac{\lambda_j + 1}{1 + \lambda_j \alpha_j^2} \right)$$

eşitsizliğini sağlaması gerekir.

İspat MLE ve LE tahmin edicilerinin MSE fonksiyonları arasındaki fark

$$\begin{aligned} \text{MSE}(\hat{\boldsymbol{\alpha}}) - \text{MSE}(\hat{\boldsymbol{\alpha}}_{\text{LE}}) &= \sum_{j=1}^{p+1} \frac{1}{\lambda_j} - \sum_{j=1}^{p+1} \frac{(\lambda_j + d)^2}{(\lambda_j + 1)^2 \lambda_j} - (1 - d)^2 \sum_{j=1}^{p+1} \frac{\alpha_j^2}{(\lambda_j + 1)^2} \\ &= (1 - d) \sum_{j=1}^{p+1} \frac{2\lambda_j + 1 - \alpha_j^2 \lambda_j + d(1 + \alpha_j^2 \lambda_j)}{\lambda_j (\lambda_j + 1)^2} \end{aligned} \quad (5.9)$$

olarak yazılabilir. O zaman (5.9) ile gösterilen farkın, $2\lambda_j + 1 - \alpha_j^2 \lambda_j + d(1 + \alpha_j^2 \lambda_j) > 0$ için pozitif olduğu açıktır. Son eşitsizliği basitleştirerek $d > \max \left(1 - 2 \frac{\lambda_j + 1}{1 + \lambda_j \alpha_j^2} \right)$ elde edilir ve böylece ispat bitmiştir. ■

5.2. d Yanlılık Parametresinin Tahmini

Yanlı tahmin edicilerde küçülme parametrelerinin seçimi her zaman önemli bir konudur. Tobit Liu regresyon yöntemi ile yapılacak çıkarımların kararlılığı, d küçülme parametresinin optimum değerinin belirlenmesine bağlıdır. Uygun d parametresinin belirlenmesi durumunda LE tahmin edicisinin MSE'si daha küçük olacaktır. Kısaca çoklu bağlantı probleminin varlığında LE tahmin edicisi ile yapılacak çıkarımların kararlılığı, yanlılık parametresi d için optimum değerinin belirlenmesine bağlıdır. Optimum değerinin belirlenmesi için

MSE ($\widehat{\alpha}_{LE}$) fonksiyonunun d 'ye göre birinci türevi sıfıra eşitlenir

$$\frac{dMSE(\widehat{\alpha}_{LE})}{dd} = 2 \sum_{j=1}^{p+1} \frac{\lambda_j(1 - \alpha_j^2) + d(1 + \alpha_j^2\lambda_j)}{\lambda_j(\lambda_j + 1)^2} = 0. \quad (5.10)$$

elde edilir. (5.10) eşitliğinin d 'ye göre çözülmesi ile

$$\begin{aligned} d &= \frac{\sum_{j=1}^{p+1} \frac{\lambda_j(-1+\alpha_j^2)}{\lambda_j(\lambda_j+1)^2}}{\sum_{j=1}^{p+1} \frac{1+\alpha_j^2\lambda_j}{\lambda_j(\lambda_j+1)^2}} \\ &= \frac{\sum_{j=1}^{p+1} \frac{1+\alpha_j^2\lambda_j}{\lambda_j(\lambda_j+1)^2} - \sum_{j=1}^{p+1} \frac{\lambda_j+1}{\lambda_j(\lambda_j+1)^2}}{\sum_{j=1}^{p+1} \frac{1+\alpha_j^2\lambda_j}{\lambda_j(\lambda_j+1)^2}} \\ &= 1 - \frac{\sum_{j=1}^{p+1} \frac{\lambda_j+1}{\lambda_j(\lambda_j+1)^2}}{\sum_{j=1}^{p+1} \frac{1+\alpha_j^2\lambda_j}{\lambda_j(\lambda_j+1)^2}} \end{aligned} \quad (5.11)$$

elde edilir. Liu yanlılık parametresi, tanımından dolayı her zaman sıfır ile bir arasında olmalıdır. (5.11) eşitliğinden, d değerinin her zaman birden küçük olduğu açıktır. Ayrıca d 'nin her zaman sıfırdan büyük olduğu hakkında kesin bilgi elde edilemez. Bu yüzden (5.10)'da bulunan toplam sembolündeki her bir eleman sıfıra eşitlenir. $j = 1, 2, \dots, p + 1$ için

$$d_j = \frac{\alpha_j^2 - 1}{\frac{1}{\lambda_j} + \alpha_j^2} \quad (5.12)$$

elde edilir.

$$\begin{aligned} d_j &= \frac{\alpha_j^2 - 1}{\frac{1}{\lambda_j} + \alpha_j^2} \\ &= \frac{\alpha_j^2 + \frac{1}{\lambda_j} - \frac{1}{\lambda_j} - 1}{\frac{1}{\lambda_j} + \alpha_j^2} \\ &= 1 - \frac{1 + \frac{1}{\lambda_j}}{\frac{1}{\lambda_j} + \alpha_j^2} \end{aligned} \quad (5.13)$$

(5.13) eşitliğinden, her bir d_j elemanları her zaman birden küçüktür. d_j elemanlarının bu özelliğinden faydalanarak ve pozitif bir d parametresi tanımlanabilir. Tobit modelde d parametresinin tahmini için Hoerl ve Kennard (1970), Kibria (2003), Khalaf ve Shukur (2005) ve Månsson ve ark. (2015)'nin çalışmaları takip edilebilir. Bu çalışmalardan faydalanılarak aşağıdaki tahmin ediciler önerilmiştir.

Hoerl ve Kennard (1970) çalışmasından esinlenerek tobit Liu tahmin edicisinin d parametresini tahmin etmek için

$$d_1 = \max \left(0, \frac{\widehat{\alpha}_{max}^2 - 1}{\frac{1}{\lambda_{max}} + \widehat{\alpha}_{max}^2} \right) \quad (5.14)$$

önerilmiştir, öyle ki, burada λ_{max} , $\mathbf{X}^\top \widehat{\mathbf{W}} \mathbf{X}$ matrisinin en büyük özdeğeridir. Ayrıca, d yanlılık parametresinin değeri her zaman 0 ile 1 arasında olması gerektiğinden, bu tezde önerilen tüm tahmin edicilerde maksimum operatörü kullanılmıştır.

d_2 tahmin edicisi Kibria (2003)'ün çalışmasının genelleştirilmesidir ve

$$d_2 = \max \left(0, \max \left(\frac{\hat{\alpha}_j^2 - 1}{\frac{1}{\lambda_j} + \hat{\alpha}_j^2} \right) \right) \quad (5.15)$$

şeklinde tanımlanmıştır.

Ayrıca d_3 tahmin edicisi Khalaf ve Shukur (2005) ve Månsson ve ark. (2015) çalışmalarına dayanmaktadır ve

$$d_3 = \max \left(0, \max \left(\frac{\hat{\alpha}_j^2 - 1}{\max(\frac{1}{\lambda_j}) + \hat{\alpha}_j^2} \right) \right) \quad (5.16)$$

olarak önerilmiştir.

Diğer bir d_4 tahmin edicisi Qasim ve ark. (2018)'in çalışmalarına dayanmaktadır ve

$$d_4 = \max \left(0, \frac{1}{p} \max \left(\frac{\hat{\alpha}_j^2 - 1}{\max(\frac{1}{\lambda_j}) + \hat{\alpha}_{max}^2} \right) \right) \quad (5.17)$$

olarak önerilmiştir.

5.3. Monte Carlo Simülasyon Çalışması 1

Çoklu bağlantı probleminin etkileri ile ilgilendiğimiz için çoklu bağlantı derecesini belirleyen $\rho = \{0.8, 0.9, 0.99\}$ olacak şekilde dikkate alınmıştır. Örneklem büyüklüğünün tahmin edicilerin performansı üzerinde etkisi olduğundan, örneklem büyüklüğünün beş farklı değeri kullanılmıştır, $n = 50, 100, 200, 500$ ve 1000 . Ayrıca, değişken sayısının tahmin süreci üzerindeki etkisini anlayabilmek için bağımsız değişkenlerin sayısı $4, 8, 16$ ve 32 olarak alınmıştır. Tahmin edicilerin performanslarını karşılaştırmak için MSE kriteri kullanılmıştır.

5.3.1. Simülasyon Tasarımı

Farklı derecelerde çoklu bağlantıya sahip veri matrisi üretebilmek için

$$x_{ij} = (1 - \rho^2)^{1/2} w_{ij} + \rho w_{ip+1}, \quad i = 1, 2, \dots, n, \quad j = 1, 2, \dots, p \quad (5.18)$$

eşitliği kullanılmıştır, Öyle ki burada w_{ij} standart normal dağılımından üretilen rasgele sayılardır ve herhangi iki bağımsız değişken arasındaki korelasyonun derecesini ρ^2 belirtir.

Gerçek parametre değerleri, alanda yaygın olarak kullanılan bir kısıtlama olan

$$\sum_{j=1}^p \beta_j^2 = 1$$

olacak şekilde seçilmiştir (bkz. Kibria (2003)).

Bağımlı değişkene ait n tane gözlem

$$y_i^* = \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + u_i, \quad i = 1, \dots, n \quad (5.19)$$

eşitliği kullanılarak üretilmiştir. Burada u_i ' ler, $N(0, \sigma^2)$ dağılımından üretilen rasgele hatalardır. Hataların varyansının etkisini araştırmak için σ ' nın değeri 5 olarak alınmıştır. Sansürlenmiş bağımlı değişkeni elde etmek için

$$y_i = \begin{cases} y_i^* & \text{ise } y_i^* > 0 \\ 0 & \text{ise } y_i^* \leq 0 \end{cases}$$

kullanılmıştır.

Herbir adımda yeni rasgele hata terimleri üretilerek simülasyon 5000 kez tekrarlanmıştır. $\hat{\beta}^*$ tahmincisinin simüle edilmiş MSE'si

$$MSE(\hat{\beta}^*) = \frac{1}{5000} \sum_{r=1}^{5000} (\hat{\beta}^* - \beta)_r^\top (\hat{\beta}^* - \beta) \quad (5.20)$$

eşitliğinden yararlanılarak hesaplanmıştır. Burada $(\hat{\beta}^* - \beta)_r$ simülasyonun r . yinelemesinde elde edilen parametre tahmini ve gerçek parametre vektörleri arasındaki farkı göstermektedir. Tüm hesaplamalar R programlama dili (R Development Core Team, 2019) kullanılarak gerçekleştirilmiştir.

5.3.2. Simülasyon Sonuçları

Monte Carlo simülasyonundan elde edilen simüle edilmiş MSE değerleri Tablo 5.1–5.4'te sunulmuştur.

Bu tablolardan aşağıdaki sonuçlara ulaşılmıştır:

- Gözlem sayısı (n) artıkça tahmin edicilerin MSE değerleri küçülmektedir.

Tablo 5.1. $p = 4$ için farklı ρ, n değerleri için MLE ve önerilen d parametrelerine sahip LE tahmin edicilerin MSE değerleri

	n	MLE	LE			
			d_1	d_2	d_3	d_4
$\rho = 0.80$						
	50	2.049	1.382	1.243	1.219	1.036
	100	0.881	0.606	0.596	0.592	0.554
	200	0.440	0.342	0.341	0.341	0.336
	500	0.170	0.152	0.152	0.152	0.152
	1000	0.084	0.079	0.079	0.079	0.079
$\rho = 0.90$						
	50	3.450	2.407	1.961	1.884	1.415
	100	1.864	1.194	1.096	1.076	0.983
	200	0.776	0.533	0.526	0.524	0.500
	500	0.338	0.269	0.269	0.269	0.267
	1000	0.165	0.147	0.147	0.147	0.147
$\rho = 0.99$						
	50	47.186	44.629	25.679	21.346	20.680
	100	18.635	16.417	10.037	8.631	9.819
	200	9.382	7.557	4.974	4.439	3.603
	500	3.510	2.391	1.867	1.786	1.448
	1000	1.959	1.236	1.082	1.066	0.858

Tablo 5.2. $p = 8$ için farklı ρ, n değerleri için MLE ve önerilen d parametrelerine sahip LE tahmin edicilerin MSE değerleri

	n	MLE	LE			
			d_1	d_2	d_3	d_4
$\rho = 0.80$						
	50	6.000	4.565	3.905	3.795	2.273
	100	2.201	1.549	1.478	1.461	1.253
	200	0.915	0.713	0.711	0.709	0.693
	500	0.378	0.330	0.330	0.330	0.329
	1000	0.193	0.180	0.180	0.180	0.180
$\rho = 0.90$						
	50	10.250	8.280	6.462	5.924	2.701
	100	4.731	3.472	2.871	2.762	1.744
	200	1.980	1.372	1.310	1.300	1.120
	500	0.797	0.615	0.613	0.612	0.598
	1000	0.394	0.337	0.337	0.337	0.336
$\rho = 0.99$						
	50	115.411	111.946	73.619	54.662	13.795
	100	50.853	47.533	30.999	23.063	5.073
	200	24.786	21.997	14.530	11.417	2.905
	500	9.545	7.543	5.486	4.878	2.237
	1000	4.515	3.227	2.629	2.543	1.725

- Çoklu bağlantı derecesindeki artış tahmin edicilerin MSE değerlerinin artmasına neden olur.
- $\rho = 0.8$ ve 0.9 için bağımsız değişken sayısındaki artış tahmin edicilerin MSE değerlerinde artışı sağladığı açıktır.
- Her zaman, farklı d parametrelere sahip LE tahmin edicisi MLE tahmininden daha iyi performans sergilemektedir.
- d parametreleri içinde d_4 parametresine sahip LE tahmin edicisi MSE kriterine göre daha üstün çıkmıştır.

Tablo 5.3. $p = 16$ için farklı ρ, n değerleri için MLE ve önerilen d parametrelerine sahip LE tahmin edicilerin MSE değerleri

	n	MLE	LE			
			d_1	d_2	d_3	d_4
$\rho = 0.8$	100	4.891	3.720	3.435	3.362	2.573
	200	2.337	1.783	1.745	1.739	1.606
	500	0.852	0.734	0.733	0.733	0.729
	1000	0.427	0.390	0.390	0.390	0.389
$\rho = 0.9$	100	12.371	10.327	8.427	7.687	3.949
	200	4.891	3.647	3.347	3.276	2.533
	500	1.878	1.422	1.382	1.380	1.252
	1000	0.875	0.726	0.724	0.724	0.713
$\rho = 0.99$	100	18.635	16.417	10.037	8.631	9.819
	200	9.382	7.557	4.974	4.439	3.603
	500	3.510	2.391	1.867	1.786	1.448
	1000	1.959	1.236	1.082	1.066	0.858

Tablo 5.4. $p = 32$ için farklı ρ, n değerleri için MLE ve önerilen d parametrelerine sahip LE tahmin edicilerin MSE değerleri

	n	MLE	LE			
			d_1	d_2	d_3	d_4
$\rho = 0.80$	200	6.334	5.115	4.697	4.629	3.554
	500	2.010	1.666	1.651	1.651	1.588
	1000	0.898	0.793	0.793	0.793	0.787
$\rho = 0.90$	200	12.955	10.869	9.359	8.818	5.613
	500	4.304	3.400	3.176	3.170	2.624
	1000	1.967	1.574	1.537	1.537	1.410
$\rho = 0.99$	200	148.769	143.729	108.897	70.419	10.036
	500	49.887	46.079	33.222	25.609	8.443
	1000	23.319	20.442	15.001	13.417	6.589

5.4. Monte Carlo Simülasyon Çalışması 2

Bu bölümün amacı, farklı derecelerde çoklu bağlantı probleminin varlığında tahmin edicilerin performanslarını değerlendirirken d parametresinin tahmini yerine sıfır ile bir arası değerler verildiğinde LE tahmin edicisinin performansını incelemektir. Simülasyon tasarımında etkili faktörler bir önceki bölümdeki faktörler ile aynı olduğundan burada tekrar belirtilmemiştir. Çoklu bağlantı derecesine karşılık gelen $\rho = 0.90, 0.99$ iki farklı değeri seçilmiştir. Ayrıca, örneklemin gözlem sayısı $n = 100, 250, 500$ olacak şekilde seçilmiştir. Son olarak simülasyon tasarımda bağımsız değişken sayısı $p = 4, 8, 16, 32$ den oluşan tobit modeli alınmıştır. Tahmin edicilerin performanslarını karşılaştırmak için MSE kriteri kullanılmıştır.

5.4.1. Simülasyon Tasarımı

Veri üretim süreci bir önceki bölüm ile tamamen aynı olduğundan burada tekrar anlatılmamıştır. Ancak, d küçülme parametresi için herhangi bir tahmin yerine, Varathan ve Wijekoon (2019)'ı takiben, d' nin değerlerini $\{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$ olarak seçilerek LE tahmin edicisi hesaplanmıştır.

5.4.2. Simülasyon Sonuçları

Monte Carlo simülasyonundan elde edilen simüle edilmiş MSE değerleri Tablo 5.5 ve Tablo 5.6'da sunulmuştur. Buna göre Tablo 5.5 ve Tablo 5.6'dan aşağıdaki sonuçlara ulaşılabılır.

- Tüm durumlarda, tobit Liu tahmin edicisi, simüle edilmiş MSE değerleri açısından tobit MLE tahmin edicisinden daha iyidir.
- Çoklu bağlantının derecesindeki artış, hem MLE tahmin edicisinin hem de LE tahmin edicisinin MSE değerlerinde artışa sebep olmaktadır.
- Ayrıca, tahmin edicilerin simüle edilmiş MSE değerleri, bağımsız değişkenlerin sayısındaki artıştan olumsuz etkilenmektedir.
- Örneklem büyüklüğü arttığında, tahmin edicilerin MSE'lerinin tüm durumlarda azaldığı gözlenmiştir.
- d küçülme parametresinin değeri bire yaklaştıkça, tobit Liu tahmin edicisinin MSE'si MLE'nin MSE'sine yaklaşmaktadır.
- Tobit Liu tahmin edicisinin küçülme parametresi d' nin küçük değerleri için yüksek değerlere kıyasla daha iyi performans sağladığı gözlemlenmektedir.

5.5. Gerçek Veri Uygulaması

Bu bölümde, tezin 5. bölümünde tartışılan yöntemleri göstermek için iki veri seti analiz edilmiştir. Her iki veri kümesi için de, Eşitlik (3.8) ve (5.7)'de verilen MLE ve

Tablo 5.5. $\sigma = 1$ olduğunda farklı ρ, n, p değerleri için MLE ve LE tahmin edicilerin MSE değerleri

$p = 4$	MLE	LE								
		$d = 0.1$	$d = 0.2$	$d = 0.3$	$d = 0.4$	$d = 0.5$	$d = 0.6$	$d = 0.7$	$d = 0.8$	$d = 0.9$
$\rho = 0.90$										
100	0.2815	0.2450	0.2489	0.2529	0.2568	0.2609	0.2649	0.2690	0.2732	0.2773
250	0.1096	0.1035	0.1042	0.1049	0.1055	0.1062	0.1069	0.1076	0.1082	0.1089
500	0.0567	0.0550	0.0552	0.0554	0.0556	0.0558	0.0560	0.0561	0.0563	0.0565
$\rho = 0.99$										
100	2.7226	0.9610	1.1091	1.2692	1.4411	1.6249	1.8207	2.0283	2.2478	2.4793
250	1.2447	0.6841	0.7377	0.7935	0.8515	0.9116	0.9739	1.0384	1.1050	1.1738
500	0.5321	0.4015	0.4150	0.4289	0.4429	0.4572	0.4717	0.4864	0.5014	0.5166
$p = 8$										
$\rho = 0.90$										
100	0.7658	0.6439	0.6568	0.6699	0.6831	0.6965	0.7101	0.7238	0.7376	0.7517
250	0.2685	0.2518	0.2536	0.2554	0.2573	0.2591	0.2610	0.2629	0.2647	0.2666
500	0.1339	0.1296	0.1301	0.1305	0.1310	0.1315	0.1320	0.1325	0.1329	0.1334
$\rho = 0.99$										
100	6.8062	2.2950	2.6674	3.0721	3.5089	3.9780	4.4792	5.0127	5.5783	6.1762
250	2.6271	1.5157	1.6235	1.7351	1.8508	1.9703	2.0938	2.2212	2.3526	2.4879
500	1.1467	0.8795	0.9074	0.9358	0.9646	0.9938	1.0235	1.0536	1.0842	1.1153
$p = 16$										
$\rho = 0.90$										
100	1.6834	1.3802	1.4119	1.4441	1.4768	1.5100	1.5437	1.5779	1.6125	1.6477
250	0.6188	0.5771	0.5817	0.5862	0.5908	0.5955	0.6001	0.6047	0.6094	0.6141
500	0.3090	0.2980	0.2992	0.3004	0.3016	0.3029	0.3041	0.3053	0.3065	0.3078
$\rho = 0.99$										
100	21.4201	5.9234	7.0960	8.4058	9.8530	11.4375	13.1594	15.0186	17.0151	19.1489
250	6.4692	3.5115	3.7914	4.0834	4.3877	4.7041	5.0327	5.3736	5.7266	6.0918
500	2.6524	1.9845	2.0537	2.1242	2.1959	2.2688	2.3430	2.4185	2.4952	2.5731
$p = 32$										
$\rho = 0.90$										
100	8.7579	5.9859	6.2502	6.5254	6.8115	7.1086	7.4166	7.7355	8.0654	8.4062
250	1.4849	1.3625	1.3758	1.3891	1.4025	1.4161	1.4297	1.4433	1.4571	1.4710
500	0.6808	0.6550	0.6578	0.6607	0.6635	0.6664	0.6693	0.6721	0.6750	0.6779
$\rho = 0.99$										
100	80.4358	19.5783	23.6111	28.3262	33.7235	39.8032	46.5651	54.0093	62.1359	70.9447
250	14.7318	7.8525	8.4918	9.1623	9.8641	10.5972	11.3616	12.1572	12.9841	13.8424
500	6.1368	4.5004	4.6686	4.8403	5.0153	5.1938	5.3756	5.5608	5.7494	5.9414

LE tahmin edicilerinin teorik MSE değerleri hesaplanmıştır. Gerçek parametre değerlerini bilmediğimiz için, Eşitlik (3.8)'de α yerine $\hat{\alpha}$ kullanılmıştır.

5.5.1. Tobin Verisi

İlk olarak Tobin (1958) tarafından analiz edilen Tobin verileri dikkate alınmıştır. Bağımlı değişken, dayanıklı tüketim malları alımıdır ve bağımsız değişkenler yıl olarak x_1 : yaş ve x_2 : likidite oranıdır (x 1000). Tobit regresyon modeline uygun olan bu veride 20 gözlem vardır. Bu veri seti X matrisi ölçeklendirilmeden analiz edilmiştir. İlk olarak, $X = [\mathbf{1}, x_1, x_2]$ veri matrisi oluşturulmuştur. $X^T \widehat{W} X$ matrisinin özdeğerleri: $\lambda_1 = 1.6539 \times 10^4$, $\lambda_2 = 2.6730 \times 10^1$ ve $\lambda_3 = 3.8675 \times 10^{-3}$ olarak hesaplanmıştır. Bu nedenle koşul sayısı $CN = \sqrt{\lambda_{max}/\lambda_{min}} = 2067.931$, çoklu bağlantı probleminin var olduğunu gösterir (Mackinnon ve Puterman, 1989).

Tablo 5.6. $\sigma = 5$ olduğunda farklı ρ, n, p değerleri için MLE ve LE tahmin edicilerin MSE değerleri

$p = 4$	MLE	LE								
		$d = 0.1$	$d = 0.2$	$d = 0.3$	$d = 0.4$	$d = 0.5$	$d = 0.6$	$d = 0.7$	$d = 0.8$	$d = 0.9$
$\rho = 0.90$										
100	1.1827	0.7119	0.7573	0.8044	0.8533	0.9038	0.9561	1.0102	1.0660	1.1235
250	0.4581	0.3667	0.3763	0.3860	0.3959	0.4059	0.4161	0.4264	0.4368	0.4474
500	0.2379	0.2118	0.2146	0.2175	0.2203	0.2232	0.2261	0.2290	0.2320	0.2349
$\rho = 0.99$										
100	12.3308	1.0780	1.7011	2.4809	3.4176	4.5111	5.7614	7.1686	8.7325	10.4532
250	4.3417	1.0034	1.2547	1.5359	1.8471	2.1881	2.5590	2.9598	3.3906	3.8512
500	2.0318	0.8370	0.9433	1.0562	1.1758	1.3019	1.4347	1.5740	1.7200	1.8726
$p = 8$										
$\rho = 0.90$										
100	3.2556	1.6889	1.8336	1.9856	2.1449	2.3116	2.4857	2.6671	2.8559	3.0521
250	1.1405	0.8788	0.9060	0.9337	0.9619	0.9905	1.0196	1.0491	1.0791	1.1096
500	0.5215	0.4616	0.4680	0.4746	0.4811	0.4877	0.4944	0.5011	0.5079	0.5147
$\rho = 0.99$										
100	27.8454	2.5251	3.9481	5.7187	7.8369	10.3026	13.1160	16.2770	19.7855	23.6417
250	10.0472	2.3065	2.8901	3.5428	4.2647	5.0557	5.9157	6.8449	7.8432	8.9107
500	4.9459	1.9752	2.2378	2.5173	2.8137	3.1269	3.4569	3.8039	4.1677	4.5483
$p = 16$										
$\rho = 0.90$										
100	8.7650	4.0772	4.4884	4.9270	5.3930	5.8865	6.4074	6.9557	7.5314	8.1345
250	2.5263	1.9226	1.9852	2.0489	2.1138	2.1797	2.2468	2.3150	2.3843	2.4547
500	1.2491	1.0824	1.1003	1.1183	1.1365	1.1549	1.1734	1.1921	1.2109	1.2299
$\rho = 0.99$										
100	62.1263	6.0276	9.2196	13.1719	17.8845	23.3574	29.5906	36.5841	44.3378	52.8519
250	24.7748	5.1607	6.5912	8.2090	10.0139	12.0061	14.1854	16.5520	19.1057	21.8466
500	11.8893	4.3865	5.0338	5.7276	6.4681	7.2551	8.0888	8.9690	9.8958	10.8692
$p = 32$										
$\rho = 0.90$										
100	36.2990	12.1969	13.9545	15.9423	18.1601	20.6080	23.2860	26.1941	29.3323	32.7006
250	7.1548	4.9517	5.1726	5.3994	5.6322	5.8710	6.1158	6.3666	6.6234	6.8861
500	2.6530	2.2901	2.3290	2.3682	2.4078	2.4478	2.4881	2.5288	2.5698	2.6112
$\rho = 0.99$										
100	289.1248	19.4557	32.1230	49.1143	70.4296	96.0689	126.0321	160.3194	198.9305	241.8657
250	66.5458	11.6919	15.4003	19.7053	24.6070	30.1052	36.2001	42.8916	50.1797	58.0645
500	25.3014	9.3011	10.6698	12.1408	13.7141	15.3896	17.1674	19.0475	21.0298	23.1145

LE tahmin edicisi için Bölüm 5.2 önerilen d küçülme parametreleri $d_1 = 0.9956$, $d_2 = 0.4680$, $d_3 = 0.4680$, $d_4 = 0.1560$ olarak hesaplanmıştır. Bu küçülme parametrelerine sahip LE tahmin edicisi ve MLE tahmin edicisinin değerleri ve karşılık gelen tahmini MSE değerleri Tablo 5.7’de verilmiştir. d ’nin değerini 0 ile 1 arasında farklı değerler olarak elde edilen LE tahmin edicisinin değerleri ve karşılık gelen tahmini MSE değerleri ise Tablo 5.8’de sunulmuştur.

Tablo 5.7. MLE ve önerilen d parametrelerine sahip LE tahmin edicileri için: Tobin verisinde tahmini kat sayılar ve MSE değerleri

	MLE	d_1	d_2	d_3	d_4
$\hat{\beta}_0$	15.1449	15.0791	7.1195	7.1195	2.4120
$\hat{\beta}_1$	-0.1291	-0.1286	-0.0747	-0.0747	-0.0428
$\hat{\beta}_2$	-0.0455	-0.0454	-0.0230	-0.0230	-0.0098
MSE	258.6000	256.3633	121.5867	121.5867	168.7299

Buna göre Tablo 5.7 ve Tablo 5.8’den aşağıdaki sonuçlara ulaşılabilir.

- d_1 parametresinin değeri bire yakın olduğu için LE ve MLE tahmin edicilerinin MSE

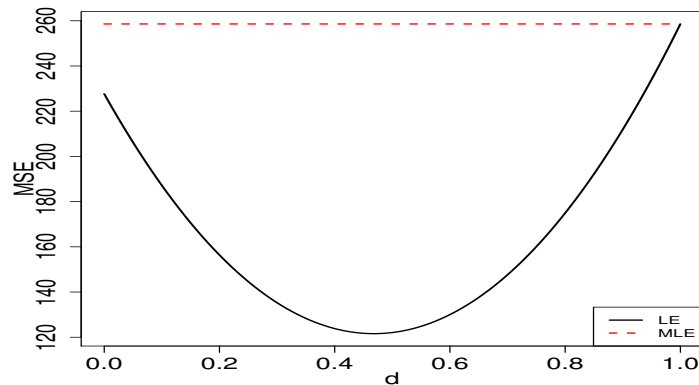
Tablo 5.8. MLE ve LE tahmin edicileri için: Tobin verilerinin tahmini katsayıları ve karşılık gelen MSE değerleri

	MLE	LE								
		$d = 0.1$	$d = 0.2$	$d = 0.3$	$d = 0.4$	$d = 0.5$	$d = 0.6$	$d = 0.7$	$d = 0.8$	$d = 0.9$
$\hat{\beta}_0$	15.1449	1.5668	3.0755	4.5842	6.0928	7.6015	9.1102	10.6189	12.1275	13.6362
$\hat{\beta}_1$	-0.1291	-0.0371	-0.0473	-0.0575	-0.0677	-0.0780	-0.0882	-0.0984	-0.1086	-0.1188
$\hat{\beta}_2$	-0.0455	-0.0075	-0.0117	-0.0159	-0.0202	-0.0244	-0.0286	-0.0329	-0.0371	-0.0413
MSE	258.6000	187.1760	156.3764	135.2607	123.8289	122.0810	130.0170	147.6369	174.9407	211.9284

değerleri birbirine yakındır.

- d_2 ve d_3 parametrelerinin değerleri aynıdır. Bu parametrelere sahip LE tahmin edicisi diğerlerine göre daha iyi performans sergiler.
- MLE'nin MSE değeri 258.6 olarak hesaplanmıştır ve d parametresinin 0.1'den 0.9'a kadar olan artan değerlerine karşılık LE'nin MSE değerlerinin önce azalan sonra artan bir yapıda olduğu gözlenmiştir.
- $d = 0.5$ 'e yaklaştığında LE'nin MSE'si bir noktaya kadar azalır ve daha sonra artmaya başlar.
- $d = 0.5$ 'e yakın olduğunda LE tahmin edicisinin MSE değeri minimum değerine ulaşır.

Tobin verisinde tobit Liu tahmin edicisinin MLE'ye karşı farklı d değerlerine göre üstünlüğünü değerlendirmek için Şekil 5.1'de gösterilen d ' e karşı tahmin edicilerin MSE değerlerinin grafiği çizilmiştir. Yukarıda özetlenen bulgular bu grafikte gösterilmiştir.



Şekil 5.1. Tobin veri setinde d 'ye karşı MLE ve LE tahmin edicilerin MSE değerleri

5.5.2. Çimento Verileri

Bu bölümde, ilk olarak Woods ve ark. (1932) tarafından incelenen ve daha sonra pek çok çalışmalarda analiz edilen çimento verisini incelenecektir. Bu veri setinde çimentonun sertleştirilmesi ve oluşturulmasında açığa çıkan ısı miktarı gösterilmektedir. Veri, çimentonun bileşenleri olan dört tane bağımsız değişkenden ve çimento gramı başına kalori cinsinden üretilen ısı miktarını gösteren bağımlı değişken (y_i^*)'den oluşmaktadır. $\mathbf{X} = [\mathbf{1}, \mathbf{x}_1, \dots, \mathbf{x}_4]$ veri matrisi oluşturulur. Veride 13 gözlem vardır. Bu veriye tobit regresyon modeli uygulamak için bağımlı değişken sansürlenmiştir. Sansürün verilerdeki etkilerini anlamak için iki farklı sansür uygulanmıştır. İlk olarak, sansürlemede keyfi olarak kesme değerini 80 olarak seçerek, Eşitlik (5.21) kullanılarak bağımlı değişken y_i elde edilir.

$$y_i = \begin{cases} y_i^* & \text{ise } y_i^* > 80 \\ 80 & \text{ise } y_i^* \leq 80 \end{cases} \quad (5.21)$$

Sansürlemeden sonra $\mathbf{X}^\top \widehat{\mathbf{W}} \mathbf{X}$ matrisinin özdeğerleri hesaplanmıştır ve bu öz değerler: $\lambda_1 = 1.0542 \times 10^4$, $\lambda_2 = 1.4076 \times 10^3$, $\lambda_3 = 1.9111 \times 10^2$, $\lambda_4 = 2.4874 \times 10^1$ ve $\lambda_5 = 2.8740 \times 10^{-4}$ olarak bulunmuştur. Koşul sayısı $CN = \sqrt{\lambda_{max}/\lambda_{min}} = 479938054$, çoklu bağlantı probleminin var olduğunu göstermektedir. LE tahmin edicisi için Bölüm 5.2'de önerilen d küçülme parametreleri $d_1 = 0.9996$, $d_2 = 0.7153$, $d_3 = 0.4383$, $d_4 = 0.1561$ olarak hesaplanmıştır. Bu küçülme parametrelerine sahip LE tahmin edicisi ve MLE tahmin edicisinin değerleri ve karşılık gelen tahmini MSE değerleri Tablo 5.9'da verilmiştir. d 'nin değeri 0 ile 1 arasından seçilen değerler olarak alındığında LE tahmin edicisinin performansı ise Tablo 5.10'da sunulmuştur.

Tablo 5.9. MLE ve önerilen d parametrelerine sahip LE tahmin edicileri için: Çimento verisinde tahmini kat sayılar ve MSE değerleri – (sansürleme-1)

	MLE	LE			
		d_1	d_2	d_3	d_4
$\widehat{\beta}_0$	52.1634	52.1442	37.3306	22.8996	8.1967
$\widehat{\beta}_1$	1.6245	1.6247	1.7606	1.8930	2.0279
$\widehat{\beta}_2$	0.5818	0.5820	0.7385	0.8909	1.0463
$\widehat{\beta}_3$	0.3458	0.3460	0.4898	0.6298	0.7726
$\widehat{\beta}_4$	0.0062	0.0064	0.1585	0.3068	0.4578
MSE	3479.5374	3476.9775	2000.8712	1525.7484	2018.9529

Buna göre Tablo 5.9 ve 5.10'dan aşağıdaki sonuçlara ulaşılmıştır.

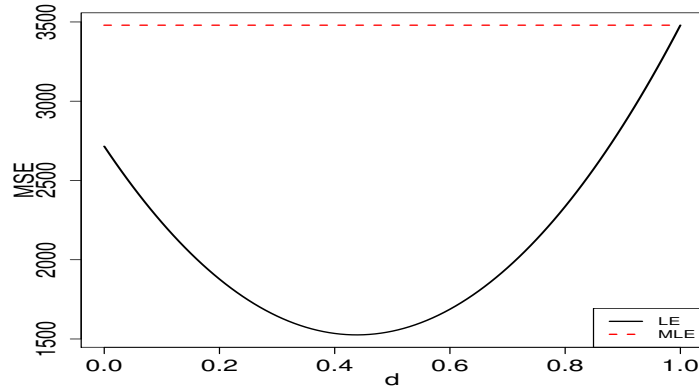
- d_1 parametresinin değeri bire yakın olduğu için LE ve MLE tahmin edicilerinin MSE değerleri birbirine yakındır.

Tablo 5.10. MLE ve LE tahmin edicileri için: Çimento verilerinin tahmini katsayıları ve karşılık gelen MSE değerleri – (sansürleme-1)

	MLE	LE								
		$d = 0.1$	$d = 0.2$	$d = 0.3$	$d = 0.4$	$d = 0.5$	$d = 0.6$	$d = 0.7$	$d = 0.8$	$d = 0.9$
$\hat{\beta}_0$	52.1634	5.2728	10.4828	15.6929	20.9030	26.1131	31.3231	36.5332	41.7433	46.9533
$\hat{\beta}_1$	1.6245	2.0547	2.0069	1.9591	1.9113	1.8635	1.8157	1.7679	1.7201	1.6723
$\hat{\beta}_2$	0.5818	1.0772	1.0221	0.9671	0.9120	0.8570	0.8019	0.7469	0.6919	0.6368
$\hat{\beta}_3$	0.3458	0.8010	0.7504	0.6998	0.6492	0.5987	0.5481	0.4975	0.4469	0.3964
$\hat{\beta}_4$	0.0062	0.4878	0.4343	0.3808	0.3273	0.2737	0.2202	0.1667	0.1132	0.0597
MSE	3479.5	2234.6	1877.5	1644.2	1534.8	1549.3	1687.6	1949.8	2335.8	2845.7

- d_3 parametresine sahip LE tahmin edicisi diğerlerine göre daha iyi performans sergiler.
- MLE'nin tahmini MSE değeri 3479.5374 olarak hesaplanmıştır.
- LE'nin MSE'sinde önemli bir azalma vardır. d 'nin değeri, $d = 0.4$ 'e yaklaştığında LE'nin MSE'si bir noktaya kadar azalır ve daha sonra artmaya başlar.
- $d = 0.4$ 'e yakın olduğunda LE tahmin edicisinin MSE değeri minimum değerine ulaşır.

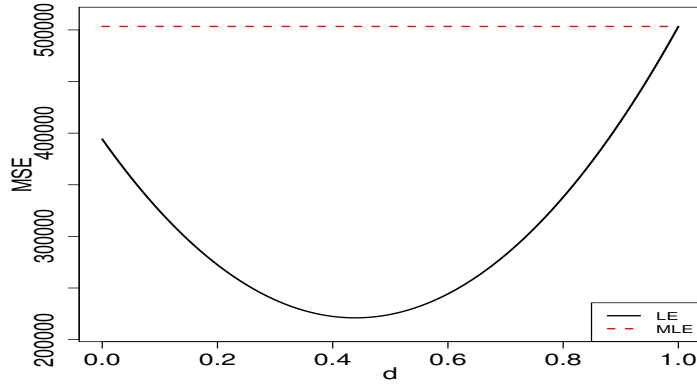
Tobit Liu tahmin edicisinin MLE'ye karşı farklı d değerlerine göre üstünlüğünü değerlendirmek için Şekil 5.2'de gösterilen d 'ye karşı tahmin edicilerin MSE değerlerinin grafiği çizilmiştir. Bu grafikte yukarıda özetlenen bulgular gözlenmektedir.



Şekil 5.2. Çimento veri setinde d ' e karşı MLE ve LE tahmin edicilerin MSE değerleri – (sansürleme-1)

Ayrıca, veri seti başka bir sansürleme senaryosu kullanılarak yeniden incelenmiştir. Sansürleme senaryosu

$$y_i = \begin{cases} y_i^* & \text{ise } y_i^* > 80 \\ 0 & \text{ise } y_i^* \leq 80 \end{cases} \quad (5.22)$$



Şekil 5.3. Çimento veri setinde d 'ye karşı MLE ve LE tahmin edicilerin MSE değerleri – (sansürleme-2)

şeklinde alınmıştır. Benzer şekilde, $\mathbf{X}^T \widehat{\mathbf{W}} \mathbf{X}$ matrisinin özdeğerleri $\lambda_1 = 71.3741$, $\lambda_2 = 7.9421$, $\lambda_3 = 1.3049$, $\lambda_4 = 0.1654$ ve $\lambda_5 = 1.9866 \times 10^{-6}$ olarak hesaplanmıştır. Koşul sayısı $CN = \sqrt{\lambda_{max}/\lambda_{min}} = 5994.003$, çoklu bağlantı probleminin var olduğunu göstermektedir. LE tahmin edicisi için Bölüm 5.2 önerilen d küçülme parametreleri $d_1 = 0.9999974$, $d_2 = 0.7063330$, $d_3 = 0.4391766$, $d_4 = 0.1566181$ olarak hesaplanmıştır. Bu küçülme parametrelerine sahip LE tahmin edicisi ve MLE tahmin edicisinin değerleri ve karşılık gelen tahmini MSE değerleri Tablo 5.11'de verilmiştir. d 'nin değeri 0 ile 1 arasından seçilen değerler olarak alındığında LE tahmin edicisinin performansı Tablo 5.12'de sunulmuştur. Ayrıca farklı d değerleri için tahmin edicilerin MSE değerlerinin grafiği Şekil 5.3'te verilmiştir.

Tablo 5.11. MLE ve önerilen d parametrelerine sahip LE tahmin edicileri için: Çimento verisinde tahmini kat sayılar ve MSE değerleri – (sansürleme-2)

	MLE	LE			
		d_1	d_2	d_3	d_4
$\widehat{\beta}_0$	627.7776	627.7759	443.4274	275.7192	98.3424
$\widehat{\beta}_1$	-1.8691	-1.8691	-0.9253	-0.0666	0.8415
$\widehat{\beta}_2$	-5.3012	-5.3012	-3.2369	-1.3589	0.6274
$\widehat{\beta}_3$	-5.1176	-5.1176	-3.6857	-2.3830	-1.0052
$\widehat{\beta}_4$	-7.4703	-7.4703	-5.4325	-3.5787	-1.6180
MSE	503384.1005	503381.5107	285138.9317	221078.5979	292743.8645

Tablo 5.12. MLE ve LE tahmin edicileri için: Çimento verilerinin tahmini katsayıları ve karşılık gelen MSE değerleri – (sansürleme-2)

	MLE	LE								
		$d = 0.1$	$d = 0.2$	$d = 0.3$	$d = 0.4$	$d = 0.5$	$d = 0.6$	$d = 0.7$	$d = 0.8$	$d = 0.9$
$\widehat{\beta}_0$	627.7776	62.8002	125.5755	188.3507	251.1260	313.9012	376.6765	439.4518	502.2270	565.0023
$\widehat{\beta}_1$	-1.8691	1.0234	0.7020	0.3806	0.0593	-0.2621	-0.5835	-0.9049	-1.2263	-1.5477
$\widehat{\beta}_2$	-5.3012	1.0254	0.3224	-0.3805	-1.0835	-1.7864	-2.4894	-3.1923	-3.8953	-4.5982
$\widehat{\beta}_3$	-5.1176	-0.7292	-1.2168	-1.7044	-2.1920	-2.6796	-3.1672	-3.6548	-4.1424	-4.6300
$\widehat{\beta}_4$	-7.4703	-1.2251	-1.9190	-2.6129	-3.3069	-4.0008	-4.6947	-5.3886	-6.0825	-6.7764
MSE	503384.1	324340.7	272427.7	238466.3	222456.6	224398.6	244292.3	282137.7	337934.8	411683.6

Buna göre Tablo 5.11 ve 5.12'den görülebileceđi gibi ilk sansürleme senaryosundaki sonuçlara benzer sonuçlar elde edilmiştir. Ancak iki sansürleme senaryosunun tahmin edicilerin hem katsayı işaretlerini hem büyüklüklerini hem de MSE değerlerini önemli derecede etkilediđi gözlenmiştir.

6. SONUÇ VE ÖNERİLER

6.1. Sonuçlar

Bu tezde tobit regresyon modelinde çoklu bağlantı problemi ele alınmıştır. Bölüm 3'te tobit modelin parametre tahmini, çoklu bağlantı probleminin tanımı, sorunun kaynakları ve sonuçları verilmiştir. Bölüm 4 ve 5'te, çoklu bağlantı probleminin üstesinden gelmek için kullanılacak bazı yanlı tahmin ediciler verilmiştir.

Bölüm 4'te, tobit ridge tahmin edicisi ile en çok olabilirlik tahmin edicisi hata kareler ortalaması ve matris hata kareler ortalaması kriterlerine göre teorik olarak karşılaştırılmıştır. Çoklu doğrusal regresyon modelinde ridge tahmin edicisinin k parametresi için literatürde bulunan tahmin ediciler tobit ridge regresyonu için tanımlanmış ve bu yöntemlerin performansları Monte Carlo simülasyonu ile değerlendirilmiştir. Bu değerlendirme sonucunda farklı k parametrelerine sahip tobit ridge tahmin edicisinin en çok olabilirlik tahmin edicisinden hata kareler ortalaması kriterine göre daha iyi performans sergilediği sonucuna ulaşılmıştır.

Son olarak Bölüm 5'te, yanlı tahmin edicilerden biri olan Liu tahmin edicisi tobit modelde tanımlanmıştır. Tobit Liu tahmin edicisi ile tobit en çok olabilirlik tahmin edicisinin matris hata kareler ortalaması ve hata kareler ortalaması kriterlerine göre teorik karşılaştırmaları elde edilmiştir. Tobit Liu tahmin edicisinde d parametresinin tahmini için yeni tahmin ediciler tanımlanmış ve bunların performansları Monte Carlo simülasyon çalışması ile değerlendirilmiştir. Tobit Liu tahmin edicisinin d parametresine bağlı performansını daha detaylı incelemek için d 'nin değeri sıfır ile bir aralığında seçilerek Monte Carlo simülasyon yapılmış ve değerlendirilmiştir. İki farklı simülasyon çalışmasında da d parametresini optimal olarak seçmenin önemi gözlenmiş ve tobit Liu tahmin edicisinin en çok olabilirlik tahmin edicisinden daha iyi performansa sahip olduğu gösterilmiştir. Ayrıca iki farklı gerçek veri uygulamasında hata kareler ortalaması kriterine göre tobit Liu tahmin edicisinin en çok olabilirlik tahmin edicisinden daha iyi performansa sahip olduğu sonucuna varılmıştır.

6.2. Öneriler

Tobit regresyon modelinde çoklu bağlantı probleminden kaynaklı sonuçlarının üstesinden gelmek için yanlı tahmin ediciler kullanılabilir. Şimdiye kadar ridge tahmin edicisi, tobit modele uyarlanmıştır. Bu tez çalışmasında Liu tahmin edicisi tobit model için incelenmiştir. Literatürde lineer modeller için tanımlanan bir çok yanlı tahmin edici vardır. Bu yanlı tahmin ediciler tobit regresyon modeline genelleştirilebilir.

KAYNAKLAR

- Alhousseini, F. H. H. and Odah, M. H., 2016, Principal component regression for tobit model and purchases of gold, *In Proceedings of the International Management Conference (Vol. 10, No. 1, pp. 491-500)*, Faculty of Management, Academy of Economic Studies, Bucharest, Romania.
- Alkhamisi, M., Khalaf, G. and Shukur, G., 2006, Some modifications for choosing ridge parameters, *Communications in Statistics–Theory and Methods*, 35(11), 2005–2020.
- Amemiya, T., 1985, *Advanced Econometrics*. Basil Blackwell. ISBN 0-631-13345-3, Oxford, 521s.
- Aydın, B., 2014, Trakya bölgesinde faaliyet gösteren tarım işletmelerinin yapısal özellikleri ve etkinliklerinin belirlenmesi, Doktora Tezi, *Namık Kemal Üniversitesi Fen Bilimleri Enstitüsü*, Tekirdağ.
- Çakır, N. Z., Sağlık harcamalarını etkileyen faktörlerin panel tobit modelleri ile analizi: Avrupa Birliği ülkeleri uygulaması, Yüksek Lisans Tezi, *Marmara Üniversitesi Sosyal Bilimleri Enstitüsü*, İstanbul.
- Emeç, H., 2001, Farklı harcama grupları için sıralı logit ve tobit modeller–bölgelerarası karşılaştırma, Doktora Tezi, *Dokuz Eylül Üniversitesi Sosyal Bilimleri Enstitüsü*, İzmir.
- Emir, B., 2016, Standart tobit regresyon modelinde kullanılan parametre tahmin yöntemlerinin karşılaştırılması, Yüksek Lisans Tezi, *Eskişehir Osmangazi Üniversitesi Sağlık Bilimleri Enstitüsü*, Eskişehir.
- Emir, M., 2015, Türkiye’de aracıların sosyo-ekonomik yapısı ve üretim etkinliği, Doktora Tezi, *Ondokuz Mayıs Üniversitesi Fen Bilimleri Enstitüsü*, Samsun.
- Farebrother, R. W., 1976, Further results on the mean square error of ridge regression, *Journal of the Royal Statistical Society B*, 38,248–250.
- Greene, W. H., 2003, *Econometric analysis*. Pearson Education India.
- Gujarati, D. N., 1988, *Basic Econometrics*, McGraw-Hill Book Company, U.S.A.
- Hausman, J. A., and Wise, D. A., 1977, Social experimentation, truncated distributions, and efficient estimation, *Econometrica: Journal of the Econometric Society*, 919–938.
- Heckman, J. J., 1979, Sample selection bias as a specification error, *Econometrica*, 47, 153–161.
- Hocking, R. R., Speed, F. M. and Lynn, M. J., 1976, A class of biased estimators in linear regression, *Technometrics*, 18:425–438.
- Hoerl, A. E. and Kennard, R. W., 1970, Ridge regression: biased estimation for nonorthogo-

- nal problems, *Technometrics*, 12(1), 55–67.
- Hoerl, A. E., Kennard, R. W. and Baldwin, K. F., 1975, Ridge regression: some simulation, *Communications in Statistics*, 4:105–123.
- Hou, Q., Huo, X. and Leng, J., 2020, A correlated random parameters tobit model to analyze the safety effects and temporal instability of factors affecting crash rates, *Accident Analysis and Prevention*, 134, 105326.
- Keskin, Ö. K., 2017, Kamu hastaneleri ile özel hastanelerin etkinliğini etkileyen faktörler (Ankara ili örneği), Doktora Tezi, *Gazi Üniversitesi Sosyal Bilimleri Enstitüsü*, Ankara.
- Khalaf, G., Månsson, K., Sjölander, P. and Shukur, G., 2014, A Tobit ridge regression estimator, *Communications in Statistics–Theory and Methods*, , 43(1), 131–140.
- Khalaf, G. and Shukur, G., 2005, Choosing ridge parameter for regression problems, *Communications in Statistics–Theory and Methods*, 34:5, 1177–1182.
- Kiranoğlu, M. E., 2005, Measuring secondary school education efficiencies of Turkish provinces: an application of data envelopment analysis, Yüksek Lisans Tezi, *Marmara Üniversitesi Fen Bilimleri Enstitüsü*, İstanbul.
- Kibria, B. G., 2003, Performance of some new ridge regression estimators, *Communications in Statistics-Simulation and Computation*, 32(2), 419–435.
- Kurtoğlu, F. and Özkale, M. R., 2016, Liu estimation in generalized linear models: application on gamma distributed response variable, *Statistical Papers*, 57(4), 911–928.
- Lawless, J. F. and Wang, P., 1976, A simulation study of ridge and other regression estimators, *Communications in Statistics–Theory and Methods*, 5(4), 307-323.
- Liu, K., 1993, A new class of biased estimate in linear regression, *Communications in Statistics–Theory and Methods*, 22(2), 393-402.
- Long, J. S., 1997, Regression models for categorical and limited dependent variables, Sage Publications Inc.
- Månsson, K., Kibria, B. G. and Shukur, G., 2012a, On Liu estimators for the logit regression model, *Economic Modelling*, 29(4), 1483–1488.
- Månsson, K., Kibria, B. G., Sjölander, P. and Shukur, G., 2012b, Improved Liu estimators for the Poisson regression model, *International Journal of Statistics and Probability*, 1(1), 2–6.
- Månsson, K., 2013, Developing a Liu estimator for the negative binomial regression model: method and application, *Journal of Statistical Computation and Simulation*, 83(9), 1773–1780.

- Månsson, K., Kibria, B. G. and Shukur, G., 2015, Performance of some weighted Liu estimators for logit regression model: an application to Swedish accident data, *Communications in Statistics–Theory and Methods*, 44(2), 363–375.
- Mackinnon, M. J. and Puterman, M. L., 1989, Collinearity in generalized linear models, *Communications in statistics–Theory and methods*, 18(9), 3463–3472.
- McDonald, G. C. and Galarneau, D. I., 1975, A Monte Carlo evaluation of some ridge-type estimators, *Journal of the American Statistical Association*, 70(350), 407–416.
- Montgomery, D. C., Peck, E. A. and Vining, G. G., 2012, Doğrusal Regresyon Analizine Giriş (Çev. Editörü: M. Aydın Erar), Nobel Yayınevi, Ankara.
- Muniz, G. and Kibria, B. G., 2009, On some ridge regression estimators: An empirical comparisons, *Communications in Statistics–Simulation and Computation*, 38(3), 621–630.
- Mujasi, P. N., Asbu, E. Z. and Puig-Junoy, J., 2016, How efficient are referral hospitals in Uganda? A data envelopment analysis and tobit regression approach, *BMC health services research* , 16(1), 230.
- Qasim, M., Amin, M. and Amanullah, M., 2018, On the performance of some new Liu parameters for the gamma regression model, *Journal of Statistical Computation and Simulation*, 88(16), 3065–3080.
- R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- R. Rao, C., Helge Toutenburg, S. and Heumann, C., 2008, Linear models and generalizations, least squares and alternatives. 3rd Edition Springer Berlin Heidelberg New York.
- Tobin, J., 1958, Estimation of relationships for limited dependent variables, *Econometrica*, 26(1), 24–36.
- Varathan, N. and Wijekoon, P., 2019, Logistic Liu Estimator under stochastic linear restrictions, *Statistical Papers*, 60(3), 595–612.
- Wilson, T., Loughran, T. and Brame, R. ,2020, Substantial bias in the Tobit estimator: Making a case for alternatives, *Justice Quarterly*, 37(2), 231-257.
- Woods, H., Steinour, H. H. and Starke, H. R., 1932, Effect of composition of Portland Cement on heat evolved during hardening. *Industr. Eng. Chem.* 24:1207–1214.
- Wulff, J. N. and Villadsen, A. R., 2020, Keeping it within bounds: Regression analysis of proportions in international business, *Journal of International Business Studies*, 51(2), 244-262.
- Yahia, F. B. and Essid, H., 2019, Determinants of Tunisian Schools’ efficiency: a DEA-Tobit approach, *Journal of Applied Management and Investments* , 8(1), 44-56.

Yazgan, A. E., 2012, Veri zarflama analizi ie etkinlik ölçümleri ve havacılık sektöründe bir uygulama, Doktora Tezi, *Gaziosmanpaşa Üniversitesi Sosyal Bilimleri Enstitüsü*, Tokat.

ÖZGEÇMİŞ

KİŞİSEL BİLGİLER

Adı Soyadı: Esra ÖĞÜTCÜOĞLU
Uyruğu: T.C.
Doğum Yeri ve Tarihi: KONYA-1995
Telefon: 0542 481 10 40
e-mail : esraogutcuoglu@gmail.com

EĞİTİM

Derece	Adı	Bitirme Yılı
Lise	Selçuklu Atatürk Anadolu Lisesi	2013
Lisans	Necmettin Erbakan Üniversitesi	2017
Yüksek Lisans	Necmettin Erbakan Üniversitesi	2020

İŞ DENEYİMLERİ

Yıl	Kurum	Görevi
2019-Devam ediyor	M.E.B.	Öğretmen

YABANCI DİLLER

İngilizce

YAYINLAR

Makale

Asar, Y., Öğütcüoğlu, E. (2020). A New Biased Estimation Method in Tobit Regression: Theory and Application, Hakemlik sürecince. (Yüksek Lisans tezinden yapılmıştır.)

Bildiri

Asar, Y., Öğütcüoğlu, E., 2018, An improved estimation method in tobit model, *International Conference on Mathematical Studies and Applications*, Karaman-Türkiye. (Yüksek Lisans tezinden yapılmıştır.)