



T.C.
NECMETTİN ERBAKAN ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

PROBİT REGRESYON MODELDE ÇOKLUBAĞLANTI PROBLEMİ
ÜZERİNE

Kadriye KILINÇ

YÜKSEK LİSANS TEZİ

Matematik Anabilim Dalı

HAZİRAN - 2020

KONYA

Her Hakkı Saklıdır

TEZ KABUL VE ONAYI

Kadriye KILINÇ tarafından hazırlanan "*PROBİT REGRESYON MODELDE ÇOK-LUBAĞLANTI PROBLEMİ ÜZERİNE*" adlı tez çalışması 02/06/2020 tarihinde aşağıdaki jüri tarafından oy birliği / oy çokluğu ile Necmettin Erbakan Üniversitesi Fen Bilimleri Enstitüsü Matematik Anabilim Dalı'nda YÜKSEK LİSANS TEZİ olarak kabul edilmiştir.

Jüri Üyeleri

İmza

Başkan

Prof. Dr. Murat ERİŞOĞLU

Danışman

Doç. Dr. Yasin ASAR

Üye

Doç. Dr. Neslihan İYİT

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

Prof. Dr. Süleyman 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.

Kadriye KILINÇ

Tarih: 02/06/2020

ÖZET

YÜKSEK LİSANS TEZİ

PROBİT REGRESYON MODELDE ÇOKLUBAĞLANTI PROBLEMİ ÜZERİNE

Kadriye KILINÇ

Necmettin Erbakan Üniversitesi Fen Bilimleri Enstitüsü

Matematik Anabilim Dalı

Danışman: Doç. Dr. Yasin ASAR

2020, 67 Sayfa

Jüri

Doç. Dr. Yasin ASAR

Prof. Dr. Murat ERİŞOĞLU

Doç. Dr. Neslihan İYİT

Bu çalışmada, çoklu doğrusal bağlantı probleminin en çok olabilirlik tahmincisi üzerindeki etkileri probit regresyon modelde analiz edilmiştir. Tasarım matrisindeki doğrusal çizgiye yakın bağımlılıkların, en çok olabilirlik tahminini olumsuz etkilediği bilinir yani; standart hatalar o kadar büyük hale gelir ki tahminlerin tutarsız olduğu gözlenir. Bu durumda en çok olabilirlik tahmincisine ve ridge tahmincisine alternatif olarak yeni bir jackknife ridge tahmincisi tanıtılmıştır. Tahmin edicilerin hata kareler ortalamasının özellikleri teorik olarak incelenmiştir. Tahmin edicilerin performansını değerlendirmek için bir Monte Carlo simülasyon çalışması tasarlanmış ve performans kriterleri olarak hata kareler ortalaması (MSE) ve karesel yanlılık (bias) kullanılmıştır. Elde edilen yeni tahmin edicinin faydaları gerçek bir veri uygulaması ile gözlenmiştir.

Anahtar Kelimeler: Çoklu doğrusallık, jackknife ridge tahmincisi, Monte Carlo simülasyon, probit regresyon model, ridge tahmincisi.

ABSTRACT

MS THESIS

ON THE PROBLEM OF MULTICOLLINEARITY IN PROBIT REGRESSION MODELS

Kadriye KILINÇ

**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, 67 Pages

Jury

Assoc. Prof. Dr. Yasin ASAR

Prof. Dr. Murat ERİŞOĞLU

Assoc. Prof. Dr. Neslihat İYİT

In this study, the effects of multicollinearity on the maximum likelihood estimator are analyzed in the probit regression model. It is known that the near-linear dependencies in the design matrix affect the maximum likelihood estimation negatively, namely, the standard errors become so large so that the estimations are said to be inconsistent. Therefore, a new jackknifed ridge estimator is introduced as an alternative to the maximum likelihood technique and the well-known ridge estimator. Mean squared error properties of the listed estimators are investigated theoretically. In order to evaluate the performance of the estimators, a Monte Carlo simulation study is designed and simulated mean squared error and squared bias are used as performance criteria. Finally, the benefits of the new estimator is illustrated via a real data application.

Anahtar Kelimeler: Jackknifed Ridge Estimator, Mean squared error, Monte Carlo simulation, Multicollinearity, Probit Model, Ridge Estimator, Squared bias.

ÖNSÖZ

Bu çalışma, Necmettin Erbakan Üniversitesi, Fen Fakültesi, Matematik ve Bilgisayar Bilimleri Bölümü, Uygulamalı Matematik Anabilim Dalından Doç. Dr. Yasin ASAR yönetiminde hazırlanarak Necmettin Erbakan Üniversitesi, Fen Bilimleri Enstitüsü'ne Yüksek Lisans Tezi olarak sunulmuştur. Çalışmalarım boyunca bilgilerimi benimle paylaşan, fikirleriyle bakış açımı geliştirip zenginleştiren ve çalışmam süresince sabırla desteğini sürdüren Doç. Dr. Yasin ASAR hocama teşekkürlerimi sunarım. Ayrıca çalışmalarım süresince sabır göstererek beni daima destekleyen aileme en içten teşekkürlerimi sunarım.

Kadriye KILINÇ

KONYA-2020

İÇİNDEKİLER

ÖZET	iv
ABSTRACT	v
ÖNSÖZ	vi
ŞEKİLLER LİSTESİ	x
TABLolar LİSTESİ	xi
SİMGELER VE KISALTMALAR	xii
1. GİRİŞ	1
2. KAYNAK ARAŞTIRMASI	3
3. MATERYAL VE METOT	5
3.1. Doğrusal Regresyon Model ve Çoklu Bağlantı	5
3.2. En Küçük Kareler Yöntemi	7
3.3. Çoklu Doğrusal Regresyon Model.....	8
3.4. Çoklu Doğrusal Regresyon Modelin Varsayımları.....	10
3.5. Tahmin Ediciler ve Özellikleri	10
3.6. β 'nın En Küçük Kareler Tahmini.....	12
3.7. Çoklu Bağlantı Problemi	13
3.7.1. Çoklu Bağlantı Probleminin Nedenleri	15
3.7.2. Çoklu Doğrusal Bağlantı Probleminin Etkileri	15
3.7.3. Çoklu Doğrusal Bağlantı Probleminin Belirlenmesi	17
3.7.3.1. Çoklu Doğrusal Bağlantı Probleminin $X^T X$ Korelasyon Matrisi ile Belirlenmesi	17
3.7.3.2. Çoklu Doğrusal Bağlantı Probleminin Varyans Şişirme Faktörü(VIF) ile Belirlenmesi.....	17
3.7.3.3. Çoklu Doğrusal Bağlantı Probleminin Regresyon Katsayılarının İşaret- leri ve Büyüklükleri İncelenerek Belirlenmesi.....	18
3.7.3.4. Çoklu Doğrusal Bağlantı Probleminin $X^T X$ Matrisinin Özdeğerleri ile Belirlenmesi	18
3.7.4. Çoklu Doğrusal Bağlantı Probleminin Giderilmesi	19

3.7.5. Ridge Regresyon	20
3.7.6. Ridge Tahmincisinin EKK ile İlişkisi	22
3.7.7. Ridge Tahmincisinin Beklenen Deęeri	22
3.7.8. Ridge Tahmincisinin Yanlılıęı.....	23
3.7.9. Ridge Tahmincisinin Varyansı	23
3.7.10. Ridge Tahmincisinin Hata Kareler Ortalaması	24
3.7.11. Jackknife Ridge Regresyon	24
3.7.12. Jackknife Ridge Tahmincisinin Beklenen Deęeri	25
3.7.13. Jackknife Ridge Tahmincisinin Yanlılıęı	26
3.7.14. Jackknife Ridge Tahmincisinin Varyansı	26
3.7.15. Jackknife Ridge Tahmincisinin Hata Kareler Ortalaması.....	27
4. PROBİT REGRESYON MODELİ	28
4.1. Probit Regresyon Modelin Matematiksel Yapısı	28
4.2. Probit Regresyon Modelinde Parametre Kestirimi	31
4.2.1. En Çok Olabilirlik Tahmini	31
4.3. Probit Ridge Tahmin Edicisi	34
4.3.1. Probit Ridge Tahmincisinin Beklenen Deęeri	34
4.3.2. Probit Ridge Tahmincisinin Yanlılıęı	35
4.3.3. Probit Ridge Tahmincisinin Varyans–kovaryans matrisi ve Varyansı	35
4.3.4. Probit Ridge Tahmincisinin Hata Kareler Ortalaması.....	36
4.3.5. Probit Ridge Tahmincisinin Parametre Seçimi	37
5. PROBİT JACKKNİFE RİDGE TAHMİN EDİCİSİ	41
5.1. Probit Jackknife Ridge Tahmincisinin Beklenen Deęeri	42
5.2. Jackknife Ridge Tahmincisinin Yanlılıęı	42
5.3. Probit Jackknife Ridge Tahmincisinin Varyansı	43
5.4. Probit Jackknife Ridge Tahmincisinin Hata Kareler Ortalaması	43
5.5. MSE Özellikleri ve Tahminciler Arasındaki Teorik Karşılaştırmalar	45
5.5.1. Tahmincilerin Yanlılık(Bias) Karşılaştırmaları	46
5.5.2. Tahmincilerin Varyans Karşılaştırmaları	47
5.5.3. Tahmincilerin MSE ve MMSE Karşılaştırmaları.....	49
5.6. Yanlılık Parametresinin Seçimi	53
5.7. Monte Carlo Simülasyonu	54

5.8. Gerçek Veri Uygulaması	57
6. SONUÇ VE ÖNERİLER	62
6.1. Sonuçlar	62
6.2. Öneriler	62
KAYNAKLAR	63
ÖZGEÇMİŞ	68



ŞEKİLLER LİSTESİ

<u>Şekil</u>	<u>Sayfa</u>
3.1 $\hat{\beta}$ ' nın yanlı ve yansız tahmin edicilerinin örnekleme dağılımı.[Montgomery ve Peck (2012)'den uyarlanmıştır.]	21
4.1 Probit modelin birikimli dağılım olarak gösterimi.	30
4.2 Ridge İzi.	38
5.1 Tahmincilerin MSE'lerinin k 'ya göre grafikleri (Banka verisi).	61
5.2 Tahmincilerin karesel yan değerlerinin k 'ya göre grafikleri (Banka verisi).	61

TABLolar LİSTESİ

<u>Tablo</u>	<u>Sayfa</u>
5.1 $p = 4$ 'iken tahmincilerin simülasyon sonucu elde edilen MSE değerleri.	57
5.2 $p = 8$ 'iken tahmincilerin simülasyon sonucu elde edilen MSE değerleri.	57
5.3 $p = 4$ 'iken tahmincilerin simülasyon sonucu elde edilen karesel yan değerleri.	58
5.4 $p = 8$ 'iken tahmincilerin simülasyon sonucu elde edilen karesel yan değerleri.	58
5.5 Banka Verileri için tahmin edicilere ait katsayılar, standart hatalar, MSE değerleri ve karesel yan değerleri	60

SİMGELER VE KISALTMALAR

Simgeler

y	Bağımlı değişken
y^*	Gizli değişken
X	Veri matrisi
k	Ridge regresyonu yanlılık parametresi
n	Örnekleme gözlem sayısı
N	Ana kütle hacmi
β	Regresyon katsayısı
$\hat{\beta}$	β katsayısının tahmini
α	β parametresinin kanonik formu
λ_i	i . özdeğer
ϕ	Standart normal dağılımın olasılık yoğunluk fonksiyonu
Φ	birikimli dağılım fonksiyonu
r	korelasyon katsayısı
R^2	Belirtme katsayısı
$\hat{\beta}$	En küçük kareler tahmincisi
$\hat{\beta}_{rr}$	Ridge regresyon tahmincisi
$\hat{\beta}_{jrr}$	Jackknife ridge regresyon tahmincisi
$\hat{\beta}_{MLE}$	En çok olabirlik tahmincisi
$\hat{\beta}_{PRR}$	Probit modelde ridge regresyon tahmincisi
$\hat{\beta}_{JRR}$	Probit modelde jackknife ridge regresyon tahmincisi
$\hat{\gamma}_{rr}$	Ridge regresyon tahmincisinin kanonik formdaki hali
$\hat{\gamma}_{jrr}$	Jackknife ridge regresyon tahmincisinin kanonik formdaki hali
$\hat{\alpha}_{MLE}$	En çok olabirlik tahmincisinin kanonik formdaki hali
$\hat{\alpha}_{PRR}$	Probit modelde ridge regresyon tahmincisinin kanonik formdaki hali
$\hat{\alpha}_{JRR}$	Probit modelde jackknife ridge regresyon tahmincisinin kanonik formdaki hali

$tr(\mathbf{X})$	\mathbf{X} matrisinin izi
$E(\hat{\beta})$	$\hat{\beta}$ tahmincisinin beklenen değeri
$Bias(\hat{\beta})$	$\hat{\beta}$ tahmincisinin yanlılığı
$Cov(\hat{\beta})$	$\hat{\beta}$ tahmincisinin varyans–kovaryans matrisi
$Var(\hat{\beta})$	$\hat{\beta}$ tahmincisinin varyansı
σ^2	Lineer regresyon modelde hata teriminin varyansı
\mathbf{I}_n	$n \times n$ 'lik birim matris
$\epsilon \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_n)$	$\mathbf{0}$ ortalamalı ve $\sigma^2 \mathbf{I}_n$ varyans–kovaryans matrisli çok değişkenli normal dağılıma sahip ϵ hata vektörü
Σ	Toplam sembolü
Π	Çarpım sembolü

Kısaltmalar

MSE	Hata Kareler Ortalaması (Mean Squared Error)
MMSE	Matris Hata Kareler Ortalaması
EKK	En Küçük Kareler Tahmincisi
MLE	En Çok Olabilirlik Tahmincisi (Maksimum Likelihood Estimator)
JRR	Jackknife Ridge Tahmincisi
PRR	Probit Ridge Tahmincisi
PJRR	Probit Jackknife Ridge Tahmincisi
VIF	Varyans Şişirme Faktörü (Variance Inflation Factor)
CN	Koşul Sayısı (Condition Number)
GRR	Genelleştirilmiş Ridge Tahmincisi
GJR	Genelleştirilmiş Jackknife Ridge Tahmincisi
IRLS	Yinelemeli Olarak Yeniden Ağırlıklandırılmış En Küçük Kareler (Iteratively Reweighted Least Squares)

1. GİRİŞ

Regresyon analizi, bağımlı değişken (y) ile bir veya daha fazla bağımsız değişken (x_i) arasındaki ilişkiyi açıklamak ve bu ilişkiye uygun model uydurmak için kullanılan bir yöntemdir. Bağımsız değişken sayısının bir olması durumu basit regresyon, birden fazla olması durumu ise çoklu (multiple) regresyon olarak isimlendirilir.

Örnek olarak bir araştırmacının tutum puanlarının başarıyı yordama derecesini araştırdığı varsayalım. Tutum puanları açıklayıcı değişken olsun. Yalnızca başarı ile tutum puanları arasındaki ilişki açıklanıyorsa yani; bir bağımsız ve bir bağımlı değişken varsa bu model basit doğrusal regresyon modelidir. Bağımsız değişkenler artırılırsa bu analize çoklu doğrusal regresyon analizi denir.

Bir olayın sebep-sonuç ilişkisini belirleyebilmek için regresyon modeli sıkça kullanılan modellerden birisidir. Bu modelde yukarıdaki örnekte verilen olay ile bu olayın meydana gelmesinde etkili olan parametreler arasında veya olayı etkileyen bağımsız parametreler arasında bir bağıntının olabileceği öngörülür. Ancak bu bağıntının regresyon analizinde sıkça sorun yarattığı da bilinir. Regresyon sürecinde ortaya çıkan bu sorunun nedeni, bağımsız değişkenlerin bağımsızlık varsayımlarının bozulması ve sonuçta bu değişkenler arasında doğrusal bir bağıntının ortaya çıkmasıdır. "Çoklu doğrusal bağlantı veya çoklu bağlantı" olarak alandırılan bu soruna önerilen çözüm, yanlı regresyon yöntemlerinin kullanılmasıdır (Büyüküysal, 2010). Söz konusu yöntemler, değişken seçimi yaparak veya değişkenlerin hepsini modelde bırakarak en çok olabilirlik yöntemine (MLE) göre daha küçük varyansla kestirim yapan yöntemlerdir.

Çoklu doğrusal bağlantı problemi ile karşılaşıldığında yanlı tahmin yöntemlerinden biri olan ridge regresyon analizi kullanılabilir. Ridge regresyon analizi, hatayı en aza indirmeyi amaçlayan regresyon yöntemi olup genellikle modeldeki iki ya da daha fazla bağımsız değişken arasında yüksek dereceden ilişki olması durumunda kullanılır.

Regresyon analizinde temel bağımlı değişkenin normal dağıldığı varsayılarak, bağımlı değişkenin özellik belirten değişken (ister iki düzeyli ister çok düzeyli) olması durumunda probit regresyon analizine ihtiyaç duyulur (Özarıcı, 1996). Hata terimlerinin dağılımına i-

lişkin eğriyi normal eğri olarak kabul eden probit modellerin tahmini için en küçük kareler (EKK) yönteminin uygulanması uygun değildir. Bu yöntemle bulunan parametreler etkin olmayacağından parametrelerin tahmini için en çok olabilirlik tahmin yöntemi kullanılır (Cafrı, 2009).

Doğrusal olasılık modelindeki en önemli problem olasılık değerlerinin 0 ile 1 aralığı dışında bulunmasıdır. Bir olasılık değeri sıfırdan küçük veya birden büyük olamayacağına göre bu durum mantıklı değildir. Bu durumun üstesinden gelmek için özellikle iki kategorili bağımlı değişkenler için tasarlanmış regresyon modellerinden biri olan probit regresyon model geliştirilmiştir. (Gençer, 2016).

Bu çalışmada öncelikle çoklu doğrusal regresyon modelde en küçük kareler tahmin yönteminden, çoklu bağlantı probleminde ve problemin sebepleri ile sonuçlarından bahsedilmiştir. Çoklu bağlantı probleminin parametre tahminleri üzerine olan etkilerini azaltmak için önerilen yanlı tahmin edicilerden ridge regresyon ile jackknife ridge yöntemine ve bu yöntemlerin kullanım amaçlarına değinilmiştir. Sonraki bölümde ise probit regresyon modelde ridge tahmincisinin özellikleri incelenerek parametre tahminlerinin yapılabilmesi için gerekli olan k yanlılık parametresini belirleme yöntemleri verilerek ridge regresyon yöntemine ait parametre tahminleri elde edilmiştir. Ayrıca probit regresyonda jackknife ridge regresyon yöntemi önerilerek bu yöntemin özellikleri ve çalışma prensibi anlatılarak bu yöntemle ait parametre tahminleri yapılmıştır.

Bu tezin amacı; probit regresyon modelde çoklu bağlantı probleminin olduğu durumlarda problemin parametre tahminleri üzerine olan etkilerini azaltmak için önerilen yanlı tahmin edicilerden probit ridge ve probit jackknife ridge regresyon yöntemlerinden elde edilen analiz sonuçlarının karşılaştırılmalı olarak incelenmesidir. Literatür taramasından anlaşıldığı kadarıyla jackknife metodu henüz probit regresyon modelde ridge tahmincisinin karesel yanını düşürmek için kullanılmamıştır. Bu nedenle tezde önerilen probit jackknife ridge yöntemi literatüre kazandırılan yeni bir yöntem olacaktır.

2. KAYNAK ARAŞTIRMASI

Hoerl ve Kennard (1970a) çoklu doğrusal regresyon modelde tahmin vektörleri ortogonal değilse, en küçük kareler (EKK) toplamına dayalı olarak yapılan tahminlerin yanlış olmasa bile tatmin edici olma olasılığının düşük olduğunu göstermiştir. $\mathbf{X}^T \mathbf{X}$ matrisinin köşegen elemanlarına küçük pozitif bir değer ekleyerek ridge tahmincisini geliştirmiştir. Vektörlerin birbiriyle ortogonal olmadığı durumda ortaya çıkan etkiyi iki boyutlu uzayda göstermek için ridge izini tanıtmıştır. Daha küçük hata kareler ortalaması ile yanlış tahmin ediciler elde edebilmek için uygun yanlılık parametrelerinin nasıl seçilebileceğine değinmiştir.

Hoerl ve Kennard (1970b) çok faktörlü verilerdeki karmaşık ilişkileri gösterebilmek için iki boyutlu grafiksel bir prosedür olan ridge izini kullanarak en küçük kareler tahmininden daha iyi bir regresyon denklemi elde etmek için önerilerde bulunmuştur.

Kibria (2003) ridge regresyon analizinde ridge parametresini tahmin etmek için geliştirilmiş ridge regresyon yaklaşımına dayalı bazı yeni yöntemler önermiştir.

Kibria ve Saleh (2012) probit regresyon modelde regresyon parametrelerinin tahminini incelemiştir. Bu doğrultuda ağırlıklı tasarım matrisini kötü koşullu olduğunda ve β parametresinin $H\beta = h$ ile tanımlanan doğrusal bir alt uzaya ait olabileceğinden şüphelenildiğinde, β parametrelerini tahmin etmek amacıyla probit regresyon model için beş ridge regresyon model önermiştir. Tahmin edicilerin asimtotik özelliklerini, karesel yanlılıklarını, MSE matrislerini ve karesel risklerini inceleyerek önerilen tahmin edicilerin optimum bölgelerini ikinci dereceden risklere göre belirlemiştir.

Locking ve ark. (2013), Kibria ve ark. (2012)'nin çalışmasına dayanan probit ridge regresyon (PRR) modelin ridge parametresini tahmin etmek için bazı yöntemleri geliştirmiştir. Elde edilen yeni tahmincilerin performansını, Monte Carlo simülasyonları kullanarak hata kareler ortalaması kriterine göre değerlendirmiştir.

Quenouille (1956) ve Tukey (1958) lineer regresyon modelde çoklu doğrusal problemden kaynaklanan yanlılığı azaltmak için jackknife prosedürünü geliştirmiştir.

Singh ve ark. (1986), geliştirilmiş ridge regresyon tahmincisinin yanlılık miktarını

azaltmak için jackknife tekniğini uygulayarak hemen hemen yansız bir genelleştirilmiş ridge regresyon tahmincisi elde etmiştir. Khurana ve ark. (2014), herhangi bir dönüşüm işlemi yapılmaksızın orijinal (dönüştürülmemiş) parametrenin jackknife ridge tahmincisini üretmek için aynı yöntemin kullanılabilirliğini göstermiştir.

Nomura ve Ohkubo (1985), Singh ve ark. tarafından önerilen hemen hemen yansız genelleştirilmiş ridge regresyon tahmincisini MSE kriterine göre en küçük karelerle karşılaştırmış ve işlemsel olarak bir ridge tahmincisi önermiştir.

Özkale (2008), Trenkler ve Toutenburg (1990)'ın önerdiği ridge tahmincisine jackknife tekniğini uygulayarak yeni bir tahminci önermiş ve özelliklerini tartışmıştır.

Türkan ve Özel (2018) çoklu doğrusallığın neden olduğu genel problemi ortadan kaldırmak ve yanlılığı azaltmak için negatif binom regresyon modelde jackknife ridge tahmincilerinin MSE özelliklerini incelemiştir. Elde edilen sonuçlarda, negatif binom regresyon modelde kullanılan jackknife ridge tahmincisinin en çok olabilirlik ve ridge tahmincilerine tercih edilmesi gerektiğini gözlemlemiştir.

3. MATERYAL VE METOT

3.1. Doğrusal Regresyon Model ve Çoklu Bağlantı

Basit doğrusal regresyon modelde bir açıklayıcı değişken ve bağımlı değişken vardır. Açıklayıcı değişkenin birden fazla olması halinde model çoklu doğrusal regresyon olarak adlandırılır. Bu bölümde, bir bağımlı değişken ile açıklayıcı değişkenin olması halinde bu iki değişken arasındaki ilişkiyi açıklayacak modelin nasıl oluşturulduğu incelenecektir. N tane birimin her birinden bağımlı değişken $y = (y_1, y_2, \dots, y_N)$ ve açıklayıcı değişken $x = (x_1, x_2, \dots, x_N)$ değerleri belirlenmiş olsun. Bu durumda $(y_1, x_1), (y_2, x_2), \dots, (y_N, x_N)$ olmak üzere N tane gözlem değeri var olacaktır. Acaba bağımlı değişken ve açıklayıcı değişkenler arasındaki ilişki nasıldır? Bu ilişki matematiksel olarak anlatılabilir mi? Bu soruların cevaplarını verebilmek için $(y_i, x_i), i = 1, 2, \dots, N$ gözlem değerlerini koordinat sistemine yerleştirmek gerekir. Bu yapılan işleme regresyonda serpm diyagramı adı verilir. N tane gözlem değerinin kesişim noktaları serpm diyagramında bulunduğu N tane nokta oluşacaktır. Bu gözlem değerlerinin konumuna bakılarak modelin nasıl olduğu hakkında bilgi sahibi olunur. Eğer gözlem değerleri bir doğru etrafında toplanırsa doğrusal modelin kullanılması gerekir. x_i ve y_i değerleri verilmiş olsun. Bağımlı değişken ve açıklayıcı değişken arasındaki ilişkinin doğrusal biçimde olduğu varsayıldığında, basit regresyon model ortaya çıkar. β_0 ve β_1 sabit sayılar olmak üzere regresyon modelindeki fonksiyonel ilişki aşağıdaki şekilde yazılabilir

$$y_i = \beta_0 + \beta_1 x_i \quad (3.1)$$

öyle ki $i = 1, 2, \dots, N$. Yukarıdaki ilişkide iki değişken arasında determinist bir ilişki vardır. x açıklayıcı değişkeni, y bağımlı değişkenini kesin olarak belirler. Fakat iki değişken arasındaki böyle bir ilişkiyle nadiren karşılaşılır. Yukarıda verilen ilişkiye örnek olarak tutum puanlarının başarıyı yordama derecesi incelenebilir. Bu ilişkide y başarıdaki artışı, x tutum puanlarındaki artışı gösteriyorsa, yukarıdaki ilişki bu durumda eksik ifade edilmiş olur.

Çünkü en azından başarıdaki artışı belirleyen, tutum puanlarının yanında başka faktörler de etkili olabilir, örneğin özyeterlilik, özerklik, motivasyon v.b. Diğer yandan başarının (y) gözlenmesinde ya da ölçülmesinde hatalar yapılmış olabilir. Regresyonda bir model oluşturulurken tespit edilemediği ya da unutulduğu için modele katılmayan değişkenler bir rassal terim olarak katılmalıdır. Bu rassal terim ε_i ile gösterilirse x ve y arasındaki gerçek ilişki:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad i = 1, 2, \dots, N \quad (3.2)$$

eşitliğindeki gibi elde edilir. Gerçek regresyon doğrusu ise:

$$E(y_i) = \beta_0 + \beta_1 x_i, \quad i = 1, 2, \dots, N \quad (3.3)$$

olur. Bu şekilde bağımlı değişkende meydana gelen değişmeler, rassal ve deterministik değişikliklerin toplam etkisiyle ortaya çıkar. Bu gerçek ilişki ve gerçek regresyon doğrusu, x ve y 'ye ait tüm anakütle değerleri bilinmediği sürece bilinemezler. Örneklem üzerinden tahmin modeli şu şekilde ifade edilebilir:

$$y_i = \hat{\beta}_0 + \hat{\beta}_1 x_i + e_i, \quad i = 1, 2, \dots, n. \quad (3.4)$$

Tahmin edilen regresyon doğrusu ise:

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i, \quad i = 1, 2, \dots, n \quad (3.5)$$

olur. Burada \hat{y} tahmin edilen bağımlı değişkendir.

Basit doğrusal regresyon modelde yer alan $\hat{\beta}_0$ ve $\hat{\beta}_1$, sırasıyla yukarıda belirtilen gerçek ilişkideki β_0 ve β_1 parametrelerinin tahmin edilen değerleridir. Ayrıca e_i ' ler gerçek ilişkideki hataların tahmini değerleridir. Regresyonda e_i değerlerine regresyon artıkları adı verilir. (x_i, y_i) gözlem değerlerinin $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ ile ifade edilen regresyon doğrusu üzerinde ortaya çıkan sapmaların sebebi ise e_i artıklarıdır. Yani $y_i = \hat{y}_i + e_i$ ilişkisi vardır. Regresyon doğrusunda gözlem değerleri doğrunun üzerinde bulunuyorsa e_i 'ler pozitif, altında bulunuyorsa e_i 'ler negatif değer alır. Regresyonda hata terimlerinin cebirsel toplamı sıfırdır. Burada verilen $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ regresyon modeli sadece bir açıklayıcı değişken ihtiva ettiğinden basit regresyon adını alır. Modelde görüldüğü gibi, basit regresyonda tahmin edilmesi gereken $\hat{\beta}_0$ ve $\hat{\beta}_1$ parametreleri vardır. Açıklayıcı değişken 0 değerini aldığı zaman regresyon doğrusu bağımlı değişkeni $\hat{\beta}_0$ noktasında keser. Bu yüzden $\hat{\beta}_0$ parametresine kesme terimi denir. $\hat{\beta}_1$ parametresi ise modeldeki doğrunun eğimi olduğundan eğim katsayısı adını alır.

Regresyon parametrelerini tahmin etmek için kullanılan en yaygın yöntemlerden biri en küçük kareler (EKK) yöntemidir. Regresyon doğrusunun gözlem değerlerini ifade edebilmesi için gözlem noktalarını tam olarak ortalaması gerekir. Böylece hata terimleri minimize edilir. Bu yüzden, kullanılan EKK yönteminde modele bir hata terimi olarak eklenen ε değişkeni hakkında şu varsayımlar geçerlidir:

- ε_i hata terimi rasgele hata bileşenidir.
- ε_i hata teriminin beklenen değeri sıfırdır. $E(\varepsilon_i) = 0$
- ε_i hata teriminin varyansı sabittir. $Var(\varepsilon_i) = \sigma^2$
- ε_i hata terimleri ilişkisizdir. $Cov(\varepsilon_i, \varepsilon_j) = 0$.

Bu varsayımlar altında, artık kareler toplamını minimize ederek $\hat{\beta}_0$ ve $\hat{\beta}_1$ parametrelerinin belirlenmesine EKK yöntemi denir (Pamukçu, 2010).

3.2. En Küçük Kareler Yöntemi

Verileri temsil edecek bir regresyon modelinin tespit edilebilmesi için saçılım grafiğini incelediğimizde; doğrusal bir yönelme görülürse x 'in y değişkenine göre doğrusal olduğuna kesin olmasa da karar verilebilir. Ancak gözlem değerleri arasından farklı doğrular geçebilir. Bu doğrular arasında y bağımlı değişkene en yakın tahmini \hat{y} değerini minimum hata ile verecek olan doğru en uygun olanıdır.

$$e_i = y_i - \hat{y}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i, \quad i = 1, 2, \dots, n \quad (3.6)$$

olacak şekilde hataları minimize edecek bir fonksiyon seçilmelidir. Bütün gözlem değerleri için karesel hata;

$$S(\beta_0, \beta_1) = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 \quad (3.7)$$

şeklinde ifade edilir. Denklem (3.7) minimum olmalıdır. Hata kareler toplamını minimize edecek şekilde katsayıların tahmin edilmesi yöntemine en küçük kareler yöntemi (EKK) denir. Bu ifadenin minimize edilebilmesi için Denklem (3.7)'nin $\hat{\beta}_0$ ve $\hat{\beta}_1$ katsayılarına göre birinci mertebeden kısmi türevlerinin 0'a eşit olması gerekir.

$$\left. \frac{\partial S}{\partial \beta_0} \right|_{(\hat{\beta}_0, \hat{\beta}_1)} = -2 \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0 \quad (3.8)$$

$$\frac{\partial S}{\partial \beta_1} \Big|_{(\hat{\beta}_0, \hat{\beta}_1)} = -2 \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) x_i = 0 \quad (3.9)$$

Bu iki denklem düzenlendiğinde,

$$\sum_{i=1}^n y_i = n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^n x_i \quad (3.10)$$

$$\sum_{i=1}^n x_i y_i = \hat{\beta}_0 \sum_{i=1}^n x_i + \hat{\beta}_1 \sum_{i=1}^n x_i^2 \quad (3.11)$$

denklemleri elde edilir. Bu eşitliklere en küçük kareler normal denklemleri adı verilir. $\hat{\beta}_0$ parametresini elde edebilmek için (\bar{x}_i, \bar{y}_i) gözlem değeri denklemde yerine yazılırsa,

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \quad (3.12)$$

ve

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n x_i y_i - \frac{\sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n}}{\sum_{i=1}^n x_i^2 - \frac{(\sum_{i=1}^n x_i)^2}{n}} \quad (3.13)$$

olarak elde edilebilir.

3.3. Çoklu Doğrusal Regresyon Model

Çoklu doğrusal regresyon modelde asıl amaç; bağımlı değişkenin davranışını belirlemek, bu değişkenin sebep olduğu toplam değişimin neye bağlı olduğunu anlamaktır ve bağımlı değişken üzerinde değişime sebep olan açıklayıcı değişkenlerin, bağımlı değişken üzerinde nasıl bir etkide bulunduğunu matematiksel modellemeyle incelemektir. Elde edilen matematiksel modelin iki ana unsuru vardır. Bunlar ise

- Açıklayıcı değişkenler aracılığıyla bilinmeyen bağımlı değişkenin değerini tahmin etmek veya öngörmek
- Bağımlı değişkeni etkilediği düşünülen açıklayıcı değişkenlerden hangisi ya da hangilerinin bağımlı değişkeni daha fazla etkilediğini açıklamak

olarak verilebilir.

Bu modele örnek vermek gerekirse bir hastanedeki hasta memnuniyetini önceden tahmin edebilmek amacıyla memnuniyeti etkileyen faktörleri dikkate alarak bu iki değişken

arasında bir regresyon modeli oluşturulsun. Eğer hasta memnuniyetini etkileyen faktörler bu değişkeni açıklamakta zayıf kalıyorsa bağımlı değişkeni etkilediği düşünülen hastane çalışanları, hastanenin temel işleyişi, hastanenin fiziki koşulları, gelir,...vb. gibi açıklayıcı değişkenleri modele ilave ederek çoklu doğrusal regresyon model oluşturulabilir. Bu verdiğimiz örnekteki ilk amaç, belirlenecek çoklu regresyon denklemi yardımıyla bir sonraki yıl için hasta memnuniyetini ve hastane tercihini en çok hangi ögenin etkilediğini belirlemek, ikinci amaç ise bir sonraki yılın hasta memnuniyet oranını önceden belirleyerek faktörleri ona göre oluşturmaya çalışmaktır.

Genel olarak y bağımlı değişkenin k açıklayıcı değişkenli matematiksel modeli,

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik} + \varepsilon_i, \quad i = 1, 2, \dots, n \quad (3.14)$$

k açıklayıcı değişkenli çoklu doğrusal regresyon model olarak tanımlanır. Bu denklemin β_j parametreleri $j = 1, 2, 3, \dots, k$ regresyon katsayıları olarak isimlendirilir. Modelde β_j katsayısı, x_j haricindeki tüm açıklayıcı değişkenler sabit tutulduğunda x_j 'deki her bir birimlik değişime karşılık gelen y bağımlı değişkendeki ortalama değişimi ifade eder. Bu sebeple β_j katsayıları kısmi regresyon katsayıları ismini de alır. Burada ε_i hata terimi ortalaması 0, varyansı σ^2 olan normal dağılıma sahip bir rasgele bileşeni teşkil eder. Bağımlı değişkenin bir ortalama etrafında değişime sebep olan bilinmeyen ve gözlenemeyen tüm etkenlerin içinde tutulduğu bir değişkeni ifade eder. Bu amaçlara ulaşılabilmesi için β_j parametrelerinin tahmin edilmesi gerekir. β_j parametrelerinin tahmin değerleri $\hat{\beta}_j$ ile ifade edilirse çoklu doğrusal regresyon model;

$$y_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \dots + \hat{\beta}_j x_{ij} + \dots + \hat{\beta}_k x_{ik} + e_i, \quad (3.15)$$

$i = 1, 2, 3, \dots, n$, şeklinde elde edilebilir. Buna göre ilgilendiğimiz veri seti aşağıdaki gibi olur:

Gözlem	y	x_1	x_2	\dots	x_k
1	y_1	x_{11}	x_{12}	\dots	x_{1k}
2	y_2	x_{21}	x_{22}	\dots	x_{2k}
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
n	y_n	x_{n1}	x_{n2}	\dots	x_{nk}

N gözlemlenmiş anakütleden alınan n örneklemlenmiş veri seti yukarıdaki tabloda verilir. Bu modelin matris notasyonu ile gösterimi:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \quad (3.16)$$

şeklindedir. Burada \mathbf{y} bağımlı değişkene ilişkin gözlem değerleri $n \times 1$ boyutlu vektör; \mathbf{X} açıklayıcı değişkene ilişkin gözlem değerleri $n \times (k + 1)$ boyutlu matris; $\boldsymbol{\beta}$ bilinmeyen regresyon parametrelerini içeren $(k + 1) \times 1$ boyutlu vektör ve $\boldsymbol{\epsilon}$ hata değerleri $n \times 1$ boyutlu vektör olmak üzere gösterimleri aşağıdaki gibidir:

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1k} \\ 1 & x_{21} & x_{22} & \dots & x_{2k} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nk} \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}, \quad \boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}.$$

3.4. Çoklu Doğrusal Regresyon Modelin Varsayımları

Çoklu doğrusal regresyon modelin varsayımları aşağıda maddeler halinde belirtilir:

- ϵ_i hata teriminin ortalaması sıfırdır. $E(\epsilon_i) = 0$
- ϵ_i hata teriminin varyansı değişiklik göstermez, sabittir. $Var(\epsilon_i) = \sigma^2$
- ϵ_i hata terimleri ilişkisizdir ve birbirinden bağımsızdır. $i \neq j$ olmak üzere $Cov(\epsilon_i, \epsilon_j) = 0$ 'dır.

3.5. Tahmin Ediciler ve Özellikleri

Çoklu doğrusal regresyon modelde temel amaç modele ait parametre tahmini yapmaktır. Yani regresyon modelini analiz edebilmek için bağımsız değişkenlerin gözlem sonuçlarından faydalanarak bilinmeyen katsayıların tahmini yapılmak istenir. Bilinmeyen katsayıları tahmin etmek için kullanılan örneklem istatistiği tahminci olarak adlandırılır. Bu tahmincinin bazı özelliklere sahip olması gerekir (Demirci, 2014). Bu özellikler aşağıda verilmiştir.

- Yansızlık: İyi bir tahmincinin, modelde tahmin edilmeye çalışılan parametreye yakın olması beklenir. $\hat{\beta}$ tahmincisinin örneklem dağılımının beklenen değeri β katsayı vektörüne eşitse; bu $\hat{\beta}$ tahmincisi, β 'nın bir yansız tahmincisidir. Yani;

$$E(\hat{\beta}) = \beta$$

dır. Burada $\hat{\beta}$ tahmincisinin yanlılık miktarı,

$$Bias(\hat{\beta}) = E(\hat{\beta}) - \beta$$

şeklinde hesaplanabilir. Bir tahmincinin yanlı ya da yansız olması teorik olarak gösterilebilir ancak β parametresinin değeri hesaplanamayacağı için yanlılık miktarı bilinemez.

- Tutarlılık: Bir örneklemdeki n gözlem sayısı artarken $\hat{\beta}$ tahmincisinin değeri β katsayı vektörüne yaklaşıyorsa $\hat{\beta}$, β 'nin tutarlı bir tahmincisidir. Yani bu tahminci aracılığıyla hesaplanan tahminler tutarlı tahminlerdir. Tutarlılık ölçütü,

$$\lim_{n \rightarrow \infty} P(|\hat{\beta} - \beta| < \epsilon) = 1$$

denklemleri ile hesaplanır. ϵ değeri çok küçük pozitif bir sayıdır.

- Etkinlik: Bir örneklem dağılımında yansız tahmincinin etkinliği dağılımın varyansı ile tespit edilebilir. n gözlem değerli bir örneklem için söz konusu iki tahminci yansız tahminciyse, bu tahmincilerden varyansı küçük olan diğerine göre daha etkin bir durum sergiler.
- Yeterlilik: Tahmincinin parametre hakkında örneklemdeki tüm bilgiyi kullanması durumudur.

Regresyon modelinin varsayımları altında en küçük kareler veya bu gibi tahmin yöntemleri kullanılarak modelin katsayıları tahmin edilir. Fakat veride bu standart varsayımlar sağlanmaz ise bu tahmincilerin yukarıda sıralamış olduğumuz özelliklerinde bozulmalar meydana gelir. Herbir varsayımın bozulması farklı problemlere neden olur. Bu çalışmada, açıklayıcı değişkenlerin birbiriyle yüksek korelasyonlu olması durumunda ortaya çıkan problemlerden biri olan çoklu bağlantı problemi incelenir.

$y = X\beta + \epsilon$ regresyon modelinin örneklem tahmin denklemi;

$$\hat{y} = X\hat{\beta} \quad (3.17)$$

olarak elde edilir. Burada $\hat{\beta}$, $(p \times 1)$ boyutlu parametrelerin tahmin vektörüdür.

3.6. β 'nin En Küçük Kareler Tahmini

Bir regresyon modelinde en küçük kareler yöntemi, hata kareler toplamını minimize edecek şekilde β katsayı parametrelerinin bulunması üzerine kurulur. Regresyon modelinin matris notasyonu

$$\mathbf{y} = \mathbf{X}\beta + \epsilon$$

şeklinindedir. Bu modelin hata kareler toplamı

$$S(\beta) = \sum_{i=1}^n \epsilon^2 = \epsilon^\top \epsilon = (\mathbf{y} - \mathbf{X}\beta)^\top (\mathbf{y} - \mathbf{X}\beta) = \mathbf{y}^\top \mathbf{y} - 2\beta^\top \mathbf{X}^\top \mathbf{y} + \beta^\top \mathbf{X}^\top \mathbf{X}\beta \quad (3.18)$$

olarak yazılabilir. Bu fonksiyonda hata kareler toplamını minimize eden β parametreleri arandığına göre bu toplamın β 'ye göre türevi alınarak sifıra eşitlenir. O zaman,

$$\left. \frac{\partial S(\beta)}{\partial \beta} \right|_{\beta=\hat{\beta}} = -2\mathbf{X}^\top \mathbf{y} + 2\mathbf{X}^\top \mathbf{X}\hat{\beta} = 0 \quad \Rightarrow \quad \mathbf{X}^\top \mathbf{y} = \mathbf{X}^\top \mathbf{X}\hat{\beta} \quad (3.19)$$

normal denklemler elde edilir. Denklem (3.19)'den en küçük kareler tahmincisi,

$$\hat{\beta} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y} \quad (3.20)$$

olarak elde edilir. Belirtmiş olduğumuz varsayımlar eşliğinde $\hat{\beta}$ tahmin vektörünün; β katsayı parametresinin yansız, lineer ve minimum varyanslı tahmincisi olduğu gösterilebilir. Bu $\hat{\beta}$ yansız tahmincisinin beklenen değeri ile varyans–kovaryans matrisleri aşağıdaki gibidir: $\hat{\beta}$ 'nin Beklenen Değeri:

$$\begin{aligned} E(\hat{\beta}) &= E[(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y}] \\ &= E[(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top (\mathbf{X}\beta + \epsilon)] \\ &= (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{X}\beta + (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top E(\epsilon) \\ &= \beta \end{aligned}$$

$\hat{\beta}$ 'nin Kovaryansı:

$$\begin{aligned} Cov(\hat{\beta}) &= Cov[(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y}] \\ &= \sigma^2 (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{X} (\mathbf{X}^\top \mathbf{X})^{-1} \\ &= \sigma^2 (\mathbf{X}^\top \mathbf{X})^{-1} \end{aligned}$$

$\hat{\beta}$ 'nin toplam varyansı:

$$Var(\hat{\beta}) = \sigma^2 \sum_{i=1}^p \frac{1}{\lambda_i} \mathbf{v}_i \mathbf{v}_i^T \quad (3.21)$$

Denklem (3.21)'deki gibi bulunur. Burada $\mathbf{X}^T \mathbf{X}$ matrisinin özdeğerleri λ_i ve özvektörleri \mathbf{v}_i ile ifade edilir.

En küçük kareler yönteminde tahmin edilmek istenen katsayı parametresi β 'nin, tahmin vektörü $\hat{\beta}$ 'ya olan uzaklığı

$$L = (\hat{\beta} - \beta) \quad (3.22)$$

şeklinde tanımlanırsa bu uzaklığın karesinin beklenen değeri şu şekilde elde edilir:

$$E(L^2) = E\left[(\hat{\beta} - \beta)(\hat{\beta} - \beta)^T\right] \quad (3.23)$$

$$= \sigma^2 \text{iz}(\mathbf{X}^T \mathbf{X})^{-1}. \quad (3.24)$$

Bu denklem,

$$E(\hat{\beta}^T \hat{\beta}) = \beta^T \beta + \sigma^2 \text{iz}(\mathbf{X}^T \mathbf{X})^{-1} \quad (3.25)$$

şeklinde de ifade edilebilir. ϵ hata teriminin, normal dağıldığı varsayımına dayanılarak L^2 'nin varyansı,

$$Var(L^2) = 2\sigma^4 \text{iz}(\mathbf{X}^T \mathbf{X})^{-2} \quad (3.26)$$

olarak bulunur. $\mathbf{X}^T \mathbf{X}$ matrisinin özdeğerleri $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$ olacak şekilde L^2 uzaklığının beklenen değeri

$$E(L^2) = \sigma^2 \sum_{i=1}^p \left(\frac{1}{\lambda_i}\right) \quad (3.27)$$

olarak hesaplanır. Varyansın özdeğerler cinsinden eşiti ise

$$Var(L^2) = 2\sigma^4 \sum_{i=1}^p \left(\frac{1}{\lambda_i}\right)^2 \quad (3.28)$$

şeklinde hesaplanır.

Açıklayıcı değişkenler arasındaki korelasyonun yüksek olması $\mathbf{X}^T \mathbf{X}$ matrisinin determinantını 0'a yaklaştırır. Bu durum; özdeğerlerin bir ya da birkaçının 0'a yakınsamasına, modelde bulunan L^2 'nin beklenen değerinin artmasına ve varyansının şişmesine sebep olur (Demirci, 2014).

3.7. Çoklu Bağlantı Problemi

Çoklu doğrusal regresyon modelde açıklayıcı değişkenlerin birbiriyle ilişkisiz olması istenir ancak değişkenlerin korelasyonlu olması çoklu doğrusal bağlantı problemini ortaya çıkarır. Açıklayıcı değişkenler arasında doğrusal bir ilişki söz konusu değilse değişkenlerin ortogonal olduğu söylenir. Ancak bu çok karşılaşılan bir durum değildir (Montgomery ve Peck, 2012). Açıklayıcı değişkenler arasındaki ilişkinin az olması parametreleri çok etkilemeyeceğinden göz ardı edilebilir ancak bu ilişkinin güçlü olması tahminlerde yanıltıcı ve hatalı sonuçlara sebep olabilir (Kuvat, 2018).

\mathbf{X} ; $n \times k$ boyutlu veri matrisini gösterebilir, öyle ki, $\mathbf{X} = \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \dots & \mathbf{x}_k \end{bmatrix}$ şeklindedir. Çoklu bağlantı, \mathbf{X} 'in sütunlarının doğrusal bağımlılığı olarak isimlendirilebilir. \mathbf{X} veri matrisi, hepsi 0 olmayan t_1, t_2, \dots, t_k sabitleri ile çarpıldığında

$$t_1\mathbf{x}_1 + t_2\mathbf{x}_2 + \dots + t_k\mathbf{x}_k = \sum_{j=1}^k t_j\mathbf{x}_j = 0 \quad (3.29)$$

yukarıdaki eşitliği sağlıyorsa, bu duruma tam çoklu doğrusal bağlantı ismi verilir. Eğer açıklayıcı değişkenler doğrusal bağımlıysa $\mathbf{X}^T\mathbf{X}$ matrisinin rankı k 'dan küçük olur ve $\mathbf{X}^T\mathbf{X}$ matrisinin tersi elde edilemez. Ancak

$$t_1\mathbf{x}_1 + t_2\mathbf{x}_2 + \dots + t_k\mathbf{x}_k = \sum_{j=1}^k t_j\mathbf{x}_j \cong 0 \quad (3.30)$$

şeklinde bir durum söz konusu ise buna yakın çoklu doğrusal bağlantı denir. Bu durumun regresyon analizindeki olumsuz etkileri aşağıdaki gibidir:

- Herhangi bir açıklayıcı değişken veya veri setine ait değişkenler modele eklendiğinde ya da çıkarıldığında β katsayı parametrelerinde büyük değişiklikler meydana gelir.
- Bir veri setinde sadece bir veri değiştirildiğinde veya modelden çıkarıldığında β katsayı parametrelerinde yine büyük değişiklikler meydana gelir.
- Katsayı parametrelerinin negatif veya pozitif olması beklenenden farklı olabilir.
- Veri setinde önemli olan değişkenlerin regresyon katsayılarının ortalamadan uzaklıkları büyük ve bu katsayıların sonuçları sağlıklı olmaz.
- Açıklayıcı değişkenler arasındaki korelasyon katsayıları çok büyüktür ve bu istenilmeyen bir durumdur (Kuvat, 2018).

3.7.1. Çoklu Bağlantı Probleminin Nedenleri

Çoklu bağlantı probleminin ortaya çıkış nedenleri araştırılırsa şu sonuçlar elde edilir:

- Örneklem yöntemleri: Regresyon analizi yapılırken açıklayıcı değişkenlerin içinden bir örneklem alınması, çoklu doğrusallığa sebep olur. Aslında modelde çoklu doğrusal bağlantı problemi olmamasına karşın; açıklayıcı değişkenlerin eksik veya yetersiz seçilmesi bu problemin oluşmasına sebep olur.
- Aşırı tanımlanmış model: Modelde gözlem sayısının parametre sayısından az olması ($n < k$) durumunda, bu problemten kurtulabilmek için modeldeki gözlem sayısının artırılması veya önemine göre bazı değişkenlerin çıkarılması gerekir.
- Model ve anakütle üzerindeki fiziksel kısıtlar: Anakütleden kaynaklanan normalde de var olan ilişkinin örneklem üzerinde ortaya çıkması durumudur (Kırdemir, 2017).

Ayrıca araştırmacının açıklayıcı değişkenleri seçerken yaptığı hatalar çoklu doğrusal bağlantı problemine sebep olur. Örneğin bir öğrencinin boy uzunluğu, kol uzunluğu ve omuz genişliği ayrı birer değişken gibi görünse de aslında üçü de birbiriyle yakın ilişkilidir. Bu durumda araştırmacının her üçünü de ayrı birer değişken gibi modele koyması çoklu doğrusal bağlantı problemine sebep olur.

3.7.2. Çoklu Doğrusal Bağlantı Probleminin Etkileri

Çoklu doğrusal bağlantı modelinin EKK tahminleri üzerinde önemli etkileri vardır. Modelde en küçük kareler tahminlerinin varyans ve kovaryanslarını şişirir. Bu durum örneklemelerde farklılık gösterip beklenen tahminlerden farklı sonuçlar elde edilmesine sebep olabilir.

Çoklu doğrusal bağlantının etkisini daha iyi gözlemleyebilmek için iki açıklayıcı değişkeni olan aşağıdaki regresyon modeli incelenirse;

$$\mathbf{y} = \beta_0 + \beta_1 \mathbf{x}_1 + \beta_2 \mathbf{x}_2 + \epsilon \quad (3.31)$$

$i = 1, 2, \dots, N$ olmak üzere açıklayıcı değişkenlerin katsayı parametrelerinin tahmini,

$$\hat{\boldsymbol{\beta}} = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} \quad (3.32)$$

şeklinde elde edilir. Bağımlı ve açıklayıcı değişkenin verileri standartlaştırıldığında $\mathbf{X}^\top \mathbf{X}$ matrisi aşağıdaki gibi korelasyon matrisi halini alır.

$$\begin{bmatrix} 1 & r_{12} \\ r_{12} & 1 \end{bmatrix} \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = \begin{bmatrix} r_{1y} \\ r_{2y} \end{bmatrix} \quad (3.33)$$

Burada r_{12} , x_1 ve x_2 açıklayıcı değişkenleri arasındaki korelasyonu; r_{1y} ve r_{2y} ise açıklayıcı değişkenler ile bağımlı değişken arasındaki korelasyonu gösterir. $\mathbf{X}^\top \mathbf{X}$ matrisinin tersi \mathbf{C} ile ifade edilecek olursa,

$$\mathbf{C} = (\mathbf{X}^\top \mathbf{X})^{-1} = \begin{bmatrix} \frac{1}{1-r_{12}^2} & \frac{-r_{12}}{1-r_{12}^2} \\ \frac{-r_{12}}{1-r_{12}^2} & \frac{1}{1-r_{12}^2} \end{bmatrix} \quad (3.34)$$

şeklinde elde edilir. Regresyon modelinin katsayı tahminleri olan

$$\hat{\beta}_1 = \frac{r_{1y} - r_{12}r_{2y}}{1 - r_{12}^2}, \quad \hat{\beta}_2 = \frac{r_{2y} - r_{12}r_{1y}}{1 - r_{12}^2} \quad (3.35)$$

denklemleri ile ifade edilir. x_1 ve x_2 açıklayıcı değişkenleri arasında kuvvetli bir ilişki varsa ilişkinin yönüne bağlı olarak,

$$|r_{12}| \rightarrow 1$$

olur. Bu durum ise,

$$\text{Var}(\hat{\beta}_1) = \sigma^2 \mathbf{C}_{11} = \frac{\sigma^2}{1 - r_{12}^2} \rightarrow \infty$$

sebeptir. r_{12} 'nin $+1$ veya -1 'e yakın olmasına göre,

$$\text{Cov}(\hat{\beta}_1, \hat{\beta}_2) = \sigma^2 \mathbf{C}_{12} \rightarrow \pm \infty$$

elde edilir. Yani x_1 ve x_2 arasındaki korelasyon, en küçük kareler tahminlerinin varyans ve kovaryansının şişmesine neden olur. Varyansın şişmiş olması durumu, regresyon katsayı parametrelerinin herbir örnek verisi için hassasiyet göstergesi olmasından dolayı, farklı örnekler kullanıldığında oldukça farklı katsayı parametreleri ortaya çıkar. $\mathbf{C} = (\mathbf{X}^\top \mathbf{X})^{-1}$ matrisinin köşegen elemanları,

$$\mathbf{C}_{jj} = \frac{1}{1 - R_j^2} \quad j=1,2,\dots,p \quad (3.36)$$

denklemden gibi elde edilir. Burada R_j^2 , x_j açıklayıcı değişkenin bağımlı değişkenin yüzde kaçını açıklayabildiğini gösteren bir belirtme katsayısıdır. Eğer açıklayıcı değişkenlerin belirtme katsayısı 1 'e yakınsa çoklu bağlantının varlığından söz edilebilir. Açıklayıcı değişkenlerin katsayı parametreleri için verilen varyans ve kovaryans denklemlerde yerine yazılırsa

belirlilik katsayısının 1'e yakın olması varyans ve kovaryansın şiştiğini gösterir. Artan varyansa bağlı olarak regresyon katsayı tahminlerinin güven aralıkları genişler. Çoklu doğrusal bağlantı problemi regresyon katsayı parametrelerinin işaretlerini de etkiler. Katsayı parametrelerinin işaretleri açıklayıcı değişkenle bağımlı değişken arasındaki korelasyonu yanlış gösterir. Açıklayıcı değişkenler arasındaki çoklu doğrusal bağlantı β ile $\hat{\beta}$ arasındaki mesafeyi artıracığından tahminler arasında büyük sapmalara sebep olur. Bu durum katsayı parametrelerinin kararsızlığına neden olur.

Çoklu bağlantı regresyon katsayılarını işaretçe ve değerce etkilediğinden oldukça farklı tahminler ortaya çıkarabilir. Bu durum y bağımlı değişkeni etkilediğinden \hat{y} tahminlerinin standart hataları da büyür (Yıldırım, 2010). R^2 değeri yüksek olacağından regresyon modeli hatalı olsa bile modelin uygun olduğu yanlışlanmasına sebep olur. Bundan dolayı açıklayıcı değişkenler arasında çoklu doğrusal bağlantının varlığı iyi araştırılmalı, varsa bu problem giderilmeye çalışılmalıdır.

3.7.3. Çoklu Doğrusal Bağlantı Probleminin Belirlenmesi

Çoklu doğrusal bağlantı probleminin belirlenmesinde birçok yöntem kullanılır. Bu yöntemlerden en çok kullanılanlar aşağıda verilir.

3.7.3.1. Çoklu Doğrusal Bağlantı Probleminin $X^T X$ Korelasyon Matrisi ile Belirlenmesi

Açıklayıcı değişkenlerin standartlaştırılmasıyla elde edilen korelasyon matrisi formundaki $X^T X$ matrisinin köşegen elemanları kontrol edilir. Farrar ve Glauber (1967) geometrik olarak r_{ij} 'yi, x_i ve x_j değişkenleri arasındaki açının kosinüs değeri olarak tanımlamıştır. x_i ve x_j açıklayıcı değişkenleri lineer bağımlı olduğunda $|r_{ij}|$ 'nin 1 ya da 1'e yakın olması bu değişkenler arasında çoklu doğrusal bağlantı probleminin olabileceğini düşündürür. Ancak bu değişkenler arasındaki korelasyon katsayısının küçük bir değer alması durumu çoklu bağlantı probleminin olmadığı anlamına gelmez. Albayrak (2012)'a göre basit korelasyon katsayısı çoklu korelasyon katsayısından küçük olursa çoklu doğrusal bağlantı problemi ortaya çıkmayabilir.

3.7.3.2. Çoklu Doğrusal Bağlantı Probleminin Varyans Şişirme Faktörü(VIF) ile Belirlenmesi

$X^T X$ korelasyon matrisinin çoklu doğrusal bağlantı problemini belirlemede önemli bir etkisi vardır (Montgomery ve Peck, 2012). $X^T X$ matrisinin j . köşegen elemanına, j . varyans şişirme faktörü denir ve VIF_j ile gösterilir (Marquardt, 1970). Varyans şişirme faktörü,

$$VIF_j = \frac{1}{1 - R_j^2} \quad (3.37)$$

denklemleri ile hesaplanır. Varyans şişirme faktörünün hesaplanmasını göstermek amacıyla üç açıklayıcı değişkenli bir model ele alınırsa;

$$y = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_3 x_3 + e \quad (3.38)$$

Yukarıda bahsedilen varyans şişirme faktörünün formülünden çoklu korelasyon katsayısı hesaplanır. Daha sonra her bir açıklayıcı değişken için VIF değerleri hesaplanır.

$$VIF(x_1) = \frac{1}{1 - R_1^2}, \quad VIF(x_2) = \frac{1}{1 - R_2^2}, \quad VIF(x_3) = \frac{1}{1 - R_3^2} \quad (3.39)$$

Bağımlı değişken ile açıklayıcı değişkenler arasındaki korelasyon 0 ise ($R^2 = 0$) denklemden $VIF = 1$ olur. Bağımlı değişken ile açıklayıcı değişkenler arasındaki korelasyon ($R^2 = 1$) olursa denkleme göre $VIF \rightarrow \infty$ olur. Montgomery ve Peck (2012) varyans şişirme faktörü için şu genel durumu belirtmiştir. VIF değeri 10'dan küçükse tam çoklu doğrusallık yoktur, 10'dan büyükse çoklu doğrusallık probleminin varlığından söz edilebilir.

3.7.3.3. Çoklu Doğrusal Bağlantı Probleminin Regresyon Katsayılarının İşaretleri ve Büyüklükleri İncelenerek Belirlenmesi

Bir regresyon analizinde katsayı parametrelerinin işaretleri ve büyüklüklerine bakılarak çoklu doğrusal bağlantı problemi hakkında fikir sahibi olunabilir. Bir modele açıklayıcı değişkenin eklenmesi veya çıkarılması durumunda katsayı tahminlerinde gözle görülür değişiklikler meydana geliyorsa çoklu doğrusal bağlantı probleminden söz edilebilir. Ayrıca katsayıların işaretleri beklenilenin tersine çıkıyorsa yine çoklu doğrusal bağlantı probleminin varlığından söz edilebilir.

3.7.3.4. Çoklu Doğrusal Bağlantı Probleminin $X^T X$ Matrisinin Özdeğerleri ile Belirlenmesi

Frisch (1934), çoklu doğrusal bağlantıyı belirlemek için açıklayıcı değişkenlerin korelasyon matrisi $X^T X$ 'in özdeğerlerinin kullanılabileceğini ifade etmiştir. Fakat bilgisayar teknolojisindeki yetersizlik nedeniyle sayısal olarak analizi desteklenmemiştir (Özkale, 2007).

Açıklayıcı değişkenlere ait $X^T X$ matrisinin özdeğerleri $(\lambda_1, \lambda_2, \dots, \lambda_p)$ çoklu doğrusal bağlantı probleminin tespit edilmesinde kullanılabilir. Bu amaçla iki tane ölçüt geliştirilir:

- $X^T X$ matrisinin özdeğerlerinin terslerinin toplamı

$$\sum_{i=1}^p \frac{1}{\lambda_i} \cong p \quad (3.40)$$

şeklinde oluyorsa çoklu doğrusal bağlantıdan söz edilemez. Ancak çoklu doğrusal bağlantı probleminin olduğu durumlarda denklemdeki toplamın çok büyük değerler verdiği görülür.

- $X^T X$ korelasyon matrislerinden elde edilen en büyük özdeğerin en küçük özdeğere bölünmesiyle elde edilen değere koşul sayısı denir ve CN ile ifade edilir. Elde edilen koşul sayısı 100'den küçük bir değere sahipse ciddi anlamda çoklu doğrusal bağlantı problemi olmadığı, 100 ile 1000 arasında değere sahipse orta düzeyde çoklu doğrusal bağlantı problemi ve 1000'den büyük değerde ise çok güçlü çoklu doğrusal bağlantı probleminin var olduğu kabul edilebilir (Stewart, 1973).

3.7.4. Çoklu Doğrusal Bağlantı Probleminin Giderilmesi

Önceki bölümlerde bahsedilen yöntemler yardımıyla çoklu doğrusal bağlantı problemi tespit edilmişse regresyon tahminleri üzerindeki olumsuz etkisinden kurtulmak için bu problemin ortadan kaldırılması ya da etkisinin azaltılması gerekir. Çoklu doğrusal bağlantı probleminin çözüm önerileri aşağıda verilmiştir (Gujarati, 1988).

- Bir veya daha fazla açıklayıcı değişken regresyon modelinden çıkarılabilir. Ancak hangi değişkenin regresyon modelinden çıkarılması gerektiğine karar vermek önem-

lidir. Yanlış bir açıklayıcı değişkenin modelden çıkarılması yanlış yorumlara sebep olabilir.

- Bazen yeni gözlem değerleri elde edilerek çoklu doğrusal bağlantı problemi ortadan kaldırılabılır. Ancak her zaman örnekleme genişletmek mümkün olmaz.
- Birbiriyle korelasyonlu olan değişkenler yerine bu değişkenlerin toplamı (tek bir değişken olarak) alınabilir.
- Regresyon modelinde herhangi bir değişiklik yapmadan, değişken ekleyip çıkarmadan esas veri üzerinde bazı yanlış tahmin edici yöntemler kullanılabilir (Demirci, 2014). Bu yöntemlerden en çok bilinen ve tercih edilen ridge regresyon yöntemidir.

3.7.5. Ridge Regresyon

Önceki konularda bahsedildiği gibi çoklu doğrusal bağlantı olması durumunda birçok problemle karşılaşılır. Bu problemlerden en önemli olanı, regresyon parametrelerinin varyans ve kovaryanslarının sonsuza doğru büyümesidir. Bununla birlikte, önemli olan açıklayıcı değişkenlere ait regresyon parametrelerinin standart hataları artacağından parametrelerin t testi anlamsız sonuç verir. Çoklu doğrusal bağlantı probleminde modelden herhangi bir değişken çıkarmak veya eklemek sonuçlarda daha büyük farklılıklar gösterir. Ayrıca kısmi regresyon parametreleri de işaret olarak beklenen sonuçlardan farklı olabilir. Sonuç olarak çoklu doğrusal bağlantı problemi olan verilerle çalışıldığı zaman hesaplanan regresyon parametreleri durağanlığını ve kararlılığını kaybeder.

Çoklu doğrusal bağlantı problemini gidermek için kullanılan yanlış tahmincilerden biri de ridge regresyondur. Ridge regresyon ilk olarak Hoerl ve Kennard (1970a) tarafından 1970'de öne sürülmüştür. Hoerl ve Kennard bu makalede "Ridge Regresyon: Ortogonal Olmayan Problemler için Yanlış Tahmin" ismiyle tam ranklı genel hipotez modeline uygun çoklu doğrusal regresyonda sapmasız tahminleme sorununun detaylı bir araştırmasını ileri sürmüştür.

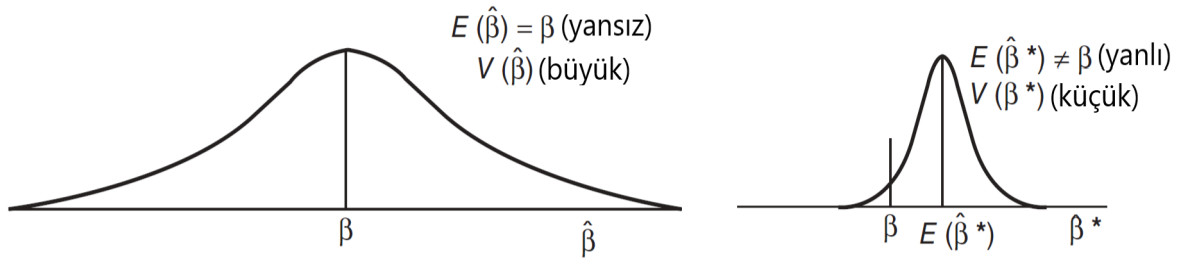
Hoerl ve Kennard (1970a,1970b) ridge regresyon modelini aşağıda verilen üç ana amaç için önermiştir:

- İleri derecede çoklu doğrusal bağlantı probleminin varlığında parametrelerdeki tutarsızlığın grafik üzerinde gösterilmesi,
- Çoklu doğrusal regresyon modelinde açıklayıcı değişkenlerin birbiriyle korelasyonlu olduğu durumlarda EKK tahmininden daha küçük varyanslı tahminlerin elde edilmesi,
- Regresyon modelinden çıkarılması gereken gereksiz açıklayıcı değişkenlere karar verilmesinde kullanılabilir.

Ridge regresyon EKK yöntemine göre yanlı bir tahminci olmasına rağmen EKK yöntemine göre avantajları aşağıdaki gibidir.

- Açıklayıcı değişkenlerdeki çoklu doğrusal bağlantıyı gidermede etkisi vardır.
- Modele bir miktar yanlılık ilave ederek tahmin edecinin varyansı ve hata kareler ortalaması azalır.

Ridge tahmincileri yanlı olmasına rağmen en küçük kareler tahminlerinden daha kararlı olabilmektedir. Varyanstaki azalıştan dolayı dikkate değer bir ilgi toplamış ve çok çeşitli alanlarda uygulanabilmektedir. Genellikle $X^T X$ matrisinin köşegen elemanlarına küçük bir yanlılık parametresi olan k sabitinin eklenmesinin dışında, ridge regresyon ve EKK yöntemlerinin uygulanışı aynıdır. Yani ridge regresyonla bir yandan tahmincilerin varyansı azalır, diğer yandan ise bu katsayı k oranında yanlı tahminler elde edilir. Böylece yansız tahmincilerle yüksek varyans, yanlı tahmincilerle düşük varyans elde edilir.



Şekil 3.1. $\hat{\beta}$ 'nin yanlı ve yansız tahmin edicilerinin örnekleme dağılımı.[Montgomery ve Peck (2012)'den uyarlanmıştır.]

Şekil 3.1'de verilen soldaki grafikte kestirim yansızdır ancak varyansı büyüktür. Sağdaki grafikte ise kestirim yanlıdır ancak varyansı küçüktür. Kestirimin yansız olmasıyla birlikte varyansın büyük olması, kestirimin kararsız olmasına neden olur. Bu yüzden $\hat{\beta}$ katsayı tahmininin yansız olma zorunluluğundan vazgeçilmelidir.

Ridge regresyon tahmin edicisi aşağıda verilir:

$$\hat{\beta}_{rr} = (\mathbf{X}^\top \mathbf{X} + k\mathbf{I})^{-1} \mathbf{X}^\top \mathbf{y} \quad (3.41)$$

burada k ridge parametresi, matrisin koşul sayısını önemli miktarda küçültmek amacıyla eklenir. k parametresi $0 < k < \infty$ aralığında değerler alır. k parametresi 0 değerini aldığı anda EKK'dan bir farkı olmayacaktır. Bu bakımdan ridge tahmincisinin EKK'nın doğrusal bir dönüşümü olduğu söylenebilir (Sakallıoğlu ve Kaçıranlar, 2008).

3.7.6. Ridge Tahmincisinin EKK ile İlişkisi

En küçük kareler tahmin edicisi $\hat{\beta} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y}$ şeklinde tanımlanmıştır. Buradan ridge tahmincisi ise $\hat{\beta}_{rr} = (\mathbf{X}^\top \mathbf{X} + k\mathbf{I})^{-1} \mathbf{X}^\top \mathbf{y}$ denkleminde $\mathbf{X}^\top \mathbf{y}$ yerine $\mathbf{X}^\top \mathbf{X} \hat{\beta}$ yazılırsa,

$$\hat{\beta}_{rr} = (\mathbf{X}^\top \mathbf{X} + k\mathbf{I})^{-1} \mathbf{X}^\top \mathbf{X} \hat{\beta} \quad (3.42)$$

elde edilir. Bu denklem ise

$$\hat{\beta}_{rr} = (\mathbf{X}^\top \mathbf{X} + k\mathbf{I})^{-1} [(\mathbf{X}^\top \mathbf{X})^{-1}]^{-1} \hat{\beta} \quad (3.43)$$

olarak yazılabilir. Her iki matris de tekil olmadığından dolayı,

$$\hat{\beta}_{rr} = \left[(\mathbf{X}^\top \mathbf{X})^{-1} (\mathbf{X}^\top \mathbf{X} + k\mathbf{I}) \right]^{-1} \hat{\beta} \quad (3.44)$$

olarak yazılabilir. Denklem,

$$\hat{\beta}_{rr} = \left[(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{X} + k (\mathbf{X}^\top \mathbf{X})^{-1} \right]^{-1} \hat{\beta} \quad (3.45)$$

şeklinde yazılır. Sonuç olarak

$$\hat{\beta}_{rr} = [\mathbf{I} + k (\mathbf{X}^\top \mathbf{X})^{-1}]^{-1} \hat{\beta} \quad (3.46)$$

olur. $\mathbf{T}_k = [\mathbf{I} + k (\mathbf{X}^\top \mathbf{X})^{-1}]^{-1}$ şeklinde tanımlanırsa,

$$\hat{\beta}_{rr} = \mathbf{T}_k \hat{\beta} \quad (3.47)$$

elde edilir, bu eşitlik ridge tahmincisinin EKK tahmincisinin bir lineer dönüşümü olduğunu gösterir.

3.7.7. Ridge Tahmincisinin Beklenen Değeri

Ridge tahmincisinin beklenen değeri:

$$\begin{aligned} E(\hat{\beta}_{rr}) &= E\left((\mathbf{X}^\top \mathbf{X} + k\mathbf{I})^{-1} \mathbf{X}^\top \mathbf{y}\right) \\ &= E\left([\mathbf{I} + k(\mathbf{X}^\top \mathbf{X})^{-1}]^{-1} \hat{\beta}\right) \\ &= [\mathbf{I} + k(\mathbf{X}^\top \mathbf{X})^{-1}]^{-1} E(\hat{\beta}) \\ &= [\mathbf{I} - k(\mathbf{X}^\top \mathbf{X} + k\mathbf{I})^{-1}] \beta \end{aligned}$$

olarak elde edilir.

3.7.8. Ridge Tahmincisinin Yanlılığı

Ridge tahmincisinin yanlılığı:

$$\begin{aligned} Bias(\hat{\beta}_{rr}) &= E(\hat{\beta}_{rr}) - \beta \\ &= [\mathbf{I} - k(\mathbf{X}^\top \mathbf{X} + k\mathbf{I})^{-1}] \beta - \beta \\ &= \beta - k(\mathbf{X}^\top \mathbf{X} + k\mathbf{I})^{-1} \beta - \beta \\ &= -k(\mathbf{X}^\top \mathbf{X} + k\mathbf{I})^{-1} \beta \end{aligned}$$

eşitlikteki gibi ifade edilir.

3.7.9. Ridge Tahmincisinin Varyansı

Ridge tahmin edicisinin varyans-kovaryans matrisi:

$$\begin{aligned} Cov(\hat{\beta}_{rr}) &= Cov\left[(\mathbf{X}^\top \mathbf{X} + k\mathbf{I})^{-1} \mathbf{X}^\top \mathbf{X} \hat{\beta}\right] \\ &= (\mathbf{X}^\top \mathbf{X} + k\mathbf{I})^{-1} \mathbf{X}^\top \mathbf{X} Cov(\hat{\beta}) \mathbf{X}^\top \mathbf{X} (\mathbf{X}^\top \mathbf{X} + k\mathbf{I})^{-1} \\ &= (\mathbf{X}^\top \mathbf{X} + k\mathbf{I})^{-1} \mathbf{X}^\top \mathbf{X} \sigma^2 (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{X} (\mathbf{X}^\top \mathbf{X} + k\mathbf{I})^{-1} \\ &= \sigma^2 (\mathbf{X}^\top \mathbf{X} + k\mathbf{I})^{-1} \mathbf{X}^\top \mathbf{X} (\mathbf{X}^\top \mathbf{X} + k\mathbf{I})^{-1} \end{aligned}$$

olarak elde edilir. Ridge tahmin edicisinin varyansı yukarıdaki verilen varyans-kovaryans matrisinin köşegen elemanlarının toplamı olarak yazılabileceğinden

$$Var(\hat{\beta}_{rr}) = tr\left(Cov(\hat{\beta}_{rr})\right) = \sigma^2 \sum_{j=1}^p \frac{\lambda_j}{(\lambda_j + k)^2}$$

olarak elde edilir.

3.7.10. Ridge Tahmincisinin Hata Kareler Ortalaması

Ridge tahmincisinin matris MSE'si:

$$\begin{aligned} MMSE \left(\hat{\beta}_{rr} \right) &= Cov \left(\hat{\beta}_{rr} \right) + Bias \left(\hat{\beta}_{rr} \right) Bias \left(\hat{\beta}_{rr} \right)^{\top} \\ &= \sigma^2 \left(\mathbf{X}^{\top} \mathbf{X} + k\mathbf{I} \right)^{-1} \mathbf{X}^{\top} \mathbf{X} \left(\mathbf{X}^{\top} \mathbf{X} + k\mathbf{I} \right)^{-1} \\ &\quad + k^2 \left(\mathbf{X}^{\top} \mathbf{X} + k\mathbf{I} \right)^{-1} \beta \beta^{\top} \left(\mathbf{X}^{\top} \mathbf{X} + k\mathbf{I} \right)^{-1} \end{aligned}$$

olarak elde edilir. Ridge tahmincisinin skaler MSE'si yukarıda verilen matris MSE'nin köşegen elemanlarının toplamı olarak yazılabileceğinden

$$MSE \left(\hat{\beta}_{rr} \right) = tr \left(MMSE \left(\hat{\beta}_{rr} \right) \right) = \sigma^2 \sum_{j=1}^p \frac{\lambda_j}{(\lambda_j + k)^2} + k^2 \sum_{j=1}^p \frac{\beta_j^2}{(\lambda_j + k)^2}$$

olarak elde edilir. Burada birinci ifade ridge tahmincisinin toplam varyansını ve ikinci ifade toplam yanlılığın karesini gösterir. Burada k yanlılık parametresine bağlı olarak seçilen varyans terimi k 'nin azalan bir fonksiyonu, yanlılık terimi k 'nin artan bir fonksiyonudur. Yani k yanlılık parametresi artarken tahminci gerçek parametre değerlerinden uzaklaşmasına rağmen varyans değeri azalacaktır.

3.7.11. Jackknife Ridge Regresyon

Bir regresyon modelinde jackknife prosedürü; örneklemden her bir gözlem değerinin sistematik olarak dışarı atılması, tahmin sonuçlarının hesaplanması ve elde edilen sonuçların ortalamasının hesaplanması ile uygulanır.

Araştırmacılar regresyon modelinde çoklu doğrusal regresyondan kaynaklanan yanlılığı azaltabilmek için Quenouille (1956) ve Tukey (1958)'in geliştirdiği yanlı bir tahmin yöntemi olan jackknife prosedürünü kullanır. Singh ve ark. (1986) çoklu doğrusal bağlantı varlığında kullanılan genelleştirilmiş ridge tahmincisinin yanlılığını azaltabilmek için jackknife prosedürünü uygulamışlardır (Arıcan, 2016). Ohtani (1986) ve Singh ve ark. (1987) jackknife ridge tahmincisinin özelliklerini incelemişlerdir. Gruber (1991) sıradan ve jackknife ridge tahmincilerini farklı iyilik ölçümleri ile karşılaştırmıştır. Jackknife ridge tahmin

edici, yanlılığı önemli ölçüde azaltsa da ridge tahmin edicisine göre daha büyük varyansa sahiptir. Nyquist (1988) jackknife tekniğinin ridge tahmincisine uygulanmasını, yanlılık parametresinin tahminini incelemiştir.

$\mathbf{X}^T \mathbf{X}$ matrisinin özdeğerlerin köşegen matrisi $\mathbf{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_p)$ ve \mathbf{T} özvektör matrisi olmak üzere:

$$\mathbf{T}^T \mathbf{X}^T \mathbf{X} \mathbf{T} = \mathbf{\Lambda}, \quad \mathbf{T}^T \mathbf{T} = \mathbf{T} \mathbf{T}^T = \mathbf{I}$$

eşitliği kullanılarak modelin kanonik formu

$$\mathbf{y} = \mathbf{Z} \boldsymbol{\gamma} + \boldsymbol{\epsilon}$$

olarak yazılabilir. Burada $\mathbf{Z} = \mathbf{X} \mathbf{T}$, $\boldsymbol{\gamma} = \mathbf{T}^T \boldsymbol{\beta}$ dır.

$\boldsymbol{\gamma}$ 'nin en küçük kareler tahmincisi

$$\hat{\boldsymbol{\gamma}} = (\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T \mathbf{y} = \mathbf{\Lambda}^{-1} \mathbf{Z}^T \mathbf{y} \quad (3.48)$$

olarak elde edilir ki bu eşitlik $\boldsymbol{\gamma} = \mathbf{T}^T \boldsymbol{\beta}$ ve $\mathbf{T}^T \mathbf{T} = \mathbf{I}$ olduğundan

$$\hat{\boldsymbol{\beta}} = \mathbf{T} \hat{\boldsymbol{\gamma}}$$

denklemden yazılabilir. Singh ve ark. (1986) jackknife ridge tahmincisi, ridge tahmincisinin bir lineer dönüşümü olarak;

$$\hat{\boldsymbol{\gamma}}_{\text{jrr}} = \left[\mathbf{I} + k (\mathbf{Z}^T \mathbf{Z} + k \mathbf{I})^{-1} \right] \hat{\boldsymbol{\gamma}} \quad (3.49)$$

eşitliğindeki gibi önermişlerdir. $\hat{\boldsymbol{\gamma}}_{\text{jrr}}$, $\boldsymbol{\gamma}$ 'nin kanonik formdaki ridge tahmincisi olmak üzere

$$\hat{\boldsymbol{\gamma}}_{\text{jrr}} = \left[\mathbf{I} - k (\mathbf{Z}^T \mathbf{Z} + k \mathbf{I})^{-1} \right] \hat{\boldsymbol{\gamma}}$$

olarak en küçük kareler tahmincisinin lineer dönüşümü şeklinde yazılırsa

$$\hat{\boldsymbol{\gamma}}_{\text{jrr}} = \left[\mathbf{I} - k^2 (\mathbf{Z}^T \mathbf{Z} + k \mathbf{I})^{-2} \right] \hat{\boldsymbol{\gamma}} \quad (3.50)$$

eşitliği elde edilir. $\hat{\boldsymbol{\gamma}} = \mathbf{T}^T \hat{\boldsymbol{\beta}}$ ve $\mathbf{Z}^T \mathbf{Z} = \mathbf{T}^T \mathbf{X}^T \mathbf{X} \mathbf{T}$ olduğundan dolayı en küçük kareler kestirimiyle $\boldsymbol{\beta}$ 'nin jackknife ridge tahmincisi:

$$\hat{\boldsymbol{\beta}}_{\text{jrr}} = \mathbf{T} \hat{\boldsymbol{\gamma}}_{\text{jrr}} = \left[\mathbf{I} - k^2 (\mathbf{X}^T \mathbf{X} + k \mathbf{I})^{-2} \right] \hat{\boldsymbol{\beta}} \quad (3.51)$$

olarak elde edilir.

3.7.12. Jackknife Ridge Tahmincisinin Beklenen Değeri

Jackknife ridge tahmincisinin beklenen değeri

$$\begin{aligned} E(\hat{\beta}_{jrr}) &= E\left[(\mathbf{I} + k(\mathbf{X}^\top \mathbf{X} + k\mathbf{I})^{-1}) \hat{\beta}_{rr}\right] \\ &= E\left[(\mathbf{I} + k(\mathbf{X}^\top \mathbf{X} + k\mathbf{I})^{-1}) (\mathbf{I} - k(\mathbf{X}^\top \mathbf{X} + k\mathbf{I})^{-1}) \hat{\beta}\right] \\ &= \left(\mathbf{I} - k^2 (\mathbf{X}^\top \mathbf{X} + k\mathbf{I})^{-2}\right) E(\hat{\beta}) \\ &= \left[\mathbf{I} - k^2 (\mathbf{X}^\top \mathbf{X} + k\mathbf{I})^{-2}\right] \beta \end{aligned}$$

olarak elde edilir.

3.7.13. Jackknife Ridge Tahmincisinin Yanlılığı

Jackknife ridge tahmincisinin yan vektörü

$$\begin{aligned} Bias(\hat{\beta}_{jrr}) &= E(\hat{\beta}_{jrr}) - \beta \\ &= \left[\mathbf{I} - k^2 (\mathbf{X}^\top \mathbf{X} + k\mathbf{I})^{-2}\right] \beta - \beta \\ &= \beta - k^2 (\mathbf{X}^\top \mathbf{X} + k\mathbf{I})^{-2} \beta - \beta \\ &= -k^2 (\mathbf{X}^\top \mathbf{X} + k\mathbf{I})^{-2} \beta \end{aligned}$$

eşitlikteki gibi ifade edilir.

3.7.14. Jackknife Ridge Tahmincisinin Varyansı

Jackknife ridge tahmincisinin varyans–kovaryans matrisi

$$\begin{aligned} Cov(\hat{\beta}_{jrr}) &= Cov\left[(\mathbf{I} - k^2(\mathbf{X}^\top \mathbf{X} + k\mathbf{I})^{-2}) \beta\right] \\ &= \left[\mathbf{I} - k^2 (\mathbf{X}^\top \mathbf{X} + k\mathbf{I})^{-2}\right] Cov(\hat{\beta}) \left[\mathbf{I} - k^2 (\mathbf{X}^\top \mathbf{X} + k\mathbf{I})^{-2}\right] \\ &= \left[\mathbf{I} - k^2 (\mathbf{X}^\top \mathbf{X} + k\mathbf{I})^{-2}\right] \sigma^2 (\mathbf{X}^\top \mathbf{X})^{-1} \left[\mathbf{I} - k^2 (\mathbf{X}^\top \mathbf{X} + k\mathbf{I})^{-2}\right] \\ &= \sigma^2 \left[\mathbf{I} - k^2 (\mathbf{X}^\top \mathbf{X} + k\mathbf{I})^{-2}\right] (\mathbf{X}^\top \mathbf{X})^{-1} \left[\mathbf{I} - k^2 (\mathbf{X}^\top \mathbf{X} + k\mathbf{I})^{-2}\right] \end{aligned}$$

olarak elde edilir. Jackknife ridge tahmincisinin varyansı yukarıda elde edilen varyans-kovaryans matrisinin köşegen elemanlarının toplamı olarak yazılabileceğinden;

$$\begin{aligned}
Var(\widehat{\beta}_{jrr}) &= tr(Cov(\widehat{\beta}_{jrr})) = \sigma^2 \sum_{j=1}^p \left(1 - \frac{k^2}{(\lambda_j + k)^2}\right) \frac{1}{\lambda_j} \left(1 - \frac{k^2}{(\lambda_j + k)^2}\right) \\
&= \sigma^2 \sum_{j=1}^p \left(\frac{(\lambda_j + k)^2 - k^2}{(\lambda_j + k)^2}\right) \frac{1}{\lambda_j} \left(\frac{(\lambda_j + k)^2 - k^2}{(\lambda_j + k)^2}\right) \\
&= \sigma^2 \sum_{j=1}^p \frac{(\lambda_j + 2k) \lambda_j}{(\lambda_j + k)^2} \frac{1}{\lambda_j} \frac{(\lambda_j + 2k) \lambda_j}{(\lambda_j + k)^2} \\
&= \sigma^2 \sum_{j=1}^p \frac{\lambda_j (\lambda_j + 2k)^2}{(\lambda_j + k)^4}
\end{aligned}$$

şeklinde ifade edilir.

3.7.15. Jackknife Ridge Tahmincisinin Hata Kareler Ortalaması

Jackknife ridge tahmincisinin matris MSE'si

$$\begin{aligned}
MMSE(\widehat{\beta}_{jrr}) &= Cov(\widehat{\beta}_{jrr}) + Bias(\widehat{\beta}_{jrr}) Bias(\widehat{\beta}_{jrr})^\top \\
&= \sigma^2 \left[\mathbf{I} - k^2 (\mathbf{X}^\top \mathbf{X} + k\mathbf{I})^{-2} \right] (\mathbf{X}^\top \mathbf{X})^{-1} \left[\mathbf{I} - k^2 (\mathbf{X}^\top \mathbf{X} + k\mathbf{I})^{-2} \right] \\
&\quad + k^4 (\mathbf{X}^\top \mathbf{X} + k\mathbf{I})^{-2} \beta \beta^\top (\mathbf{X}^\top \mathbf{X} + k\mathbf{I})^{-2}
\end{aligned}$$

olarak elde edilir. Jackknife ridge tahmincisinin skaler MSE'si matris MSE'nin köşegen elemanlarının toplamı olarak yazılabileceğinden

$$\begin{aligned}
MSE(\widehat{\beta}_{jrr}) &= \sigma^2 \sum_{j=1}^p \left(1 - \frac{k^2}{(\lambda_j + k)^2}\right) \frac{1}{\lambda_j} \left(1 - \frac{k^2}{(\lambda_j + k)^2}\right) + k^4 \sum_{j=1}^p \frac{\beta_j^2}{(\lambda_j + k)^4} \\
&= \sigma^2 \sum_{j=1}^p \left(\frac{(\lambda_j + k)^2 - k^2}{(\lambda_j + k)^2} \frac{1}{\lambda_j} \frac{(\lambda_j + k)^2 - k^2}{(\lambda_j + k)^2}\right) + k^4 \sum_{j=1}^p \frac{\beta_j^2}{(\lambda_j + k)^4} \\
&= \sigma^2 \sum_{j=1}^p \frac{\lambda_j (\lambda_j + 2k)^2}{(\lambda_j + k)^4} + k^4 \sum_{j=1}^p \frac{\beta_j^2}{(\lambda_j + k)^4}
\end{aligned}$$

olarak elde edilir. Burada birinci ifade jackknife ridge tahmincisinin toplam varyansını ve ikinci ifade toplam karesel yanlılığı gösterir. Burada k yanlılık parametresine bağlı olarak seçilen varyans terimi k 'nin azalan bir fonksiyonu, yanlılık terimi ise k 'nin artan bir fonksiyonudur. Yani k yanlılık parametresi artarken tahminci gerçek parametre değerlerinden uzaklaşmasına rağmen varyans değeri azalacaktır.

4. PROBİT REGRESYON MODELİ

Denklem (3.14) ile verilen çoklu doğrusal regresyon modelinde bağımlı değişken y iki kategorili bir değişken olduğunda model doğrusal olasılık modeli olarak isimlendirilir. Doğrusal olasılık modelinde

$$E(y | \mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p) = P(y = 1 | \mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p)$$

öyle ki

$$P(y_i = 1 | \mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik}$$

olur. Burada β_j , diğer k değişken sabit iken $y_i = 1$ olma olasılığındaki değişimi ifade eder. Parametre tahmini sıradan çoklu doğrusal regresyonda olduğu gibi EKK yöntemi ile yapılır. Ancak doğrusal olasılık modelindeki en önemli sorunlardan biri koşullu olasılık fonksiyonunun doğrusal olduğu varsayılır. Bu durum, P olasılık değerinin 0 ile 1 arasında olmasını kısıtlanmaz. Bazı durumlarda, P olasılık değeri negatif ya da 1'den büyük değerler aldığından doğrusal olasılık modelinin burada anlamlı bir yorumu yoktur. Bu durum, ikili bağımlı değişkenin koşullu olasılık fonksiyonunu modellemek için doğrusal olmayan bir fonksiyon kullanmayı gerektiren yaklaşıma sebep olur. Bu yaklaşımlardan birisi, özellikle ikili bağımlı değişkenler için geliştirilen probit regresyon modelidir.

4.1. Probit Regresyon Modelin Matematiksel Yapısı

Doğrusal olasılık modelinde P olasılık değeri 0 ile 1 aralığının dışında yer aldığı zaman modelin anlamlı bir yorumu yoktur. Doğrusal olasılık modeliyle yapılan kestirimlerin 0 ve 1 aralığının dışına çıkma olasılığına engel olabilmek amacıyla alternatif modeller üzerinde durulmaktadır. Doğrusal olasılık modeli tahminlerinin etkin ve yansız olmaması gibi olumsuz durumlara rağmen, modelin kısıtlanmalı formu, yani 0'dan küçük veya 1'den büyük önkestimlerin 0 veya 1'e eşitlenerek olasılıklarının $[0, 1]$ aralığı içinde kısıtlanmış hali, kullanılarak 0 ile 1 aralığında kalması sağlanır. Bağımlı değişkeni bir seçim yapmanın olasılığı olarak yorumlamak amaçlandığından, bir olasılık fonksiyonunun kullanılması daha uygun

olur. Bu nedenden dolayı, açıklayıcı değişkenin regresyon doğrusundaki tüm gerçek değerini 0'dan 1'e kadar uzanan bir olasılığa dönüştürmek için gerekli olan dönüşümler yapılmalıdır. Bu dönüşümler birikimli dağılım fonksiyonu kullanılarak yapılmaktadır.

Denklem (3.14) ile birlikte

$$P(y_i = 1 | \mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p) = \Phi(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik})$$

probit regresyon modeli olarak isimlendirilir, burada Φ standart normal dağılımın birikimli olasılık fonksiyonu olmak üzere $i = 1, 2, \dots, n$.

İki düzeyli özellik belirten bağımlı değişkenin, evet-hayır, başarılı-başarısız, var-yok gibi yanıtlardan oluştuğu düşünülürse, bu değerlerin ortaya çıkma olasılıkları,

$$P(y_i = 1) = \Phi(\mathbf{x}_i^\top \boldsymbol{\beta}) \quad (4.1)$$

$$P(y_i = 0) = 1 - \Phi(\mathbf{x}_i^\top \boldsymbol{\beta}) \quad (4.2)$$

şeklinde (Alp, 2007), $i = 1, 2, \dots, n$. Burada $\boldsymbol{\beta}$ katsayı parametre vektörü, açıklayıcı değişken ya da değişkenlerdeki değişimin etkisini olasılığa yansıtır. Bu yansıtılan etkiyi doğrusal regresyon modeliyle göstermek mümkündür

$$E(y_i) = \Phi(\mathbf{x}_i^\top \boldsymbol{\beta}) = \mathbf{x}_i^\top \boldsymbol{\beta} \quad (4.3)$$

ayrıca $E(y_i) = \Phi(\mathbf{x}_i^\top \boldsymbol{\beta})$ olduğundan dolayı regresyon modeli aşağıda verildiği gibi yazılabilir

$$\mathbf{y} = E(\mathbf{y}) + [\mathbf{y} - E(\mathbf{y})] = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}. \quad (4.4)$$

Böylece elde edilen olasılıklar

$$\pi_i = \Phi(\mathbf{x}_i^\top \boldsymbol{\beta}) = \Phi(I_i) \quad (4.5)$$

şeklinde (İşyar, 1994), öyle ki $I_i = \mathbf{x}_i^\top \boldsymbol{\beta}$. Burada I_i gerçekte ölçülmemiş normal ve sürekli rasgele değişkendir ve bu değerler için gözlemler mevcut değildir. Ancak bu I_i indeksinin küçük ve büyük değerlerine bakılarak bireysel gözlemlerin hangi kategoriye ait olduğu bilinebilmektedir.

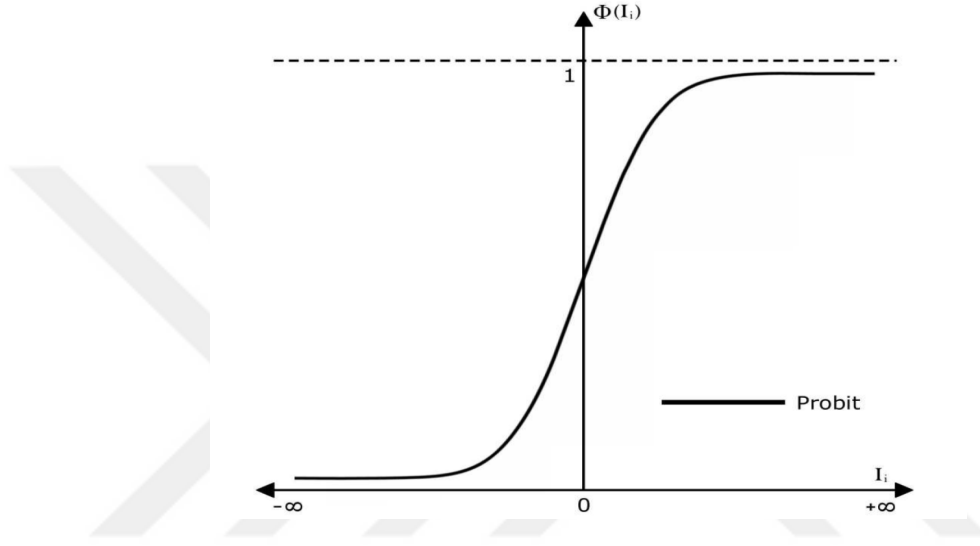
Probit analizi, gerçekte ölçülmemiş olan I_i indeksi hakkında bilgi sağlayarak $\boldsymbol{\beta}$ katsayı parametrelerinin tahmini için bir yaklaşım sağlamaktadır.

Her bir gözlem için I_i 'nin belli bir değerinden (kritik değer) itibaren olayın gerçekleşmesi söz konusudur. Bu değer I_i^* ile ifade edilecek olursa $I_i > I_i^*$ ise olay gerçekleşecek,

$I_i < I_i^*$ ise olay gerçekleşmeyecektir. I_i^* normal dağılımlı rasgele değişken olarak varsayıldığından $I_i < I_i^*$ olasılığı birikimli normal dağılım fonksiyonundan hesaplanabilecektir (Özarıcı, 1996). Probit regresyon modelinin birikimli normal dağılım fonksiyonu

$$\pi_i = \Phi(I_i) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\mathbf{x}_i^\top \boldsymbol{\beta}} \phi(t) dt \quad (4.6)$$

şeklinde gösterilebilir. Burada ϕ standart normal dağılımın olasılık yoğunluk fonksiyonudur (Greene, 2018). Bu fonksiyonun grafiği Şekil 4.1’de verilmiştir.



Şekil 4.1. Probit modelin birikimli dağılım olarak gösterimi.

Bir olayın gerçekleşme olasılığını ifade eden π_i , 0 ile 1 değerleri arasında yer alır. Bu olasılık standart normal eğrinin $-\infty$ ile I_i arasındaki bölgenin alanına eşit olup I_i değerinin büyük olması olayın gerçekleşme olasılığının yüksek olduğunu göstermektedir (İşyar, 1994).

Birikimli dağılım fonksiyonu monoton bir fonksiyondur ve bir fonksiyon monoton olduğu sürece tersi vardır. Probit regresyon model için bu fonksiyonun monoton artan olduğu aşıkardır ve probit regresyon modelin denklemini ifade edebilmek için (4.6)’daki denklemin tersi alınmalıdır. Probit regresyon model, matematiksel olarak ifade edilecek olursa

$$I_i = \Phi^{-1}(\pi_i) = \mathbf{x}_i^\top \boldsymbol{\beta} \quad (4.7)$$

şeklinde yazılabilir. Φ^{-1} birikimli normal dağılım fonksiyonunun tersini ifade eder ve (4.6)’da verilen birikimli normal dağılım fonksiyonunun tersi alınarak probit fonksiyonun doğrusal hale gelmesi sağlanır (Alp, 2007). (4.7)’de elde edilen ifade probit regresyon modeli olarak isimlendirilir.

Denklem (4.1)'den anlaşılacağı gibi, \mathbf{x}_i^\top gözlem vektörü için

$$\lim_{\mathbf{x}_i^\top \beta \rightarrow +\infty} P(y_i = 1 | \mathbf{x}_i^\top) = 1 \quad (4.8)$$

$$\lim_{\mathbf{x}_i^\top \beta \rightarrow -\infty} P(y_i = 1 | \mathbf{x}_i^\top) = 0 \quad (4.9)$$

olur, öyle ki $0 < P(y_i = 1 | \mathbf{x}_i^\top) < 1$ (Greene, 2018).

Probit regresyon modelin varsayımları şu şekildedir:

- $y_i \in \{0, 1\}, i = 1, 2, \dots, n$
- $P(y_i = 1 | \mathbf{x}_i^\top) = \Phi(\mathbf{x}_i^\top \beta)$
- y_1, y_2, \dots, y_n bağımlı değişkenler istatistiksel olarak birbirinden bağımsızdır.
- Bütün \mathbf{x}_i açıklayıcı değişkenleri arasında tam ya da yaklaşık doğrusal bağımlılık yoktur (Aldrich ve ark., 1984).

4.2. Probit Regresyon Modelinde Parametre Kestirimi

İki düzeyli bağımlı değişken içeren modellerden probit regresyonda katsayı parametrelerinin kestirimi yapılırken genellikle en çok olabilirlik yöntemi kullanılmaktadır (Greene, 2018).

4.2.1. En Çok Olabilirlik Tahmini

Regresyon modellerinde parametre tahmini yapmak için en küçük kareler tekniğine alternatif olarak en çok olabilirlik tahmin yöntemi kullanılabilir. Bu yöntem ile elde edilen parametre tahminleri yeterince büyük örneklem için tutarlıdır, asimtotik olarak etkindir ve elde edilen tahminler asimtotik olarak normal dağılıma sahiptir.

Birikimli normal dönüştürme yaklaşımı doğrusal olmadığından probit modelin parametre kestiriminde en küçük kareler tekniği uygun değildir. Probit modelin doğrusal olasılık modeline karşı daha üstün özelliklere sahip olmasına rağmen ek hesaplamalar ve kestirim külfetinden dolayı bu modelin uygulanma ihtimalini azaltmaktadır. Bu nedenle probit regresyon modelinin parametrelerinin kestiriminde en çok olabilirlik tekniğinin kullanılması daha uygun olur (İşyar, 1994).

Modelde bilinmeyen parametrelerin gözlemlenmiş veri setinden gözlenme ihtimalini en büyük yapacak biçimde kestirim esasına dayanan en çok olabilirlik tekniğinin uygulanmasında ilk olarak olabilirlik fonksiyonu oluşturulur (Özarıcı, 1996).

Probit regresyon modelinin gözlenemeyen bağımlı değişkeni

$$y_i^* = \mathbf{x}_i^\top \boldsymbol{\beta} + \varepsilon_i \quad (4.10)$$

şeklinindedir. Burada y_i^* gözlemlenmemiş veya gizli değişkeni, $\mathbf{x}_i^\top = [x_{i1}, x_{i2}, \dots, x_{ip}]$ ise $n \times p$ boyutlu \mathbf{X} veri matrisinin i . satırını, $\boldsymbol{\beta}$ $p \times 1$ boyutlu vektörü regresyon katsayılarını ve ε_i normal dağılıma sahip bir hata terimini temsil etmektedir. Gerçek hayatta gizli değişken gözlemlenemez olduğundan gizli değişkenin yerine Bernoulli(μ_i) dağılımına sahip kukla değişkeni;

$$y_i = \begin{cases} 1, & \text{if } y_i^* > 0, \\ 0, & \text{aksi takdirde,} \end{cases} \quad (4.11)$$

şeklinde elde edilir, öyle ki, $\mu_i = \Phi(\mathbf{x}_i^\top \boldsymbol{\beta})$ ve Φ standart normal dağılımın birikimli dağılım fonksiyonudur.

Probit regresyon modelinin olabilirlik fonksiyonu

$$L(y; \boldsymbol{\beta}) = \prod_{i=1}^n \Phi(\mathbf{x}_i^\top \boldsymbol{\beta})^{y_i} [1 - \Phi(\mathbf{x}_i^\top \boldsymbol{\beta})]^{1-y_i} = \prod_{i=1}^n \Phi_i^{y_i} (1 - \Phi_i)^{1-y_i}.$$

ile ifade edilir. En çok olabilirlik tekniği, probit regresyon modelinde olabilirlik değerini maksimize edecek model parametre değerlerini hesaplamaya çalışır. Yukarıda verilen olabilirlik fonksiyonunu maksimize etmek ile onun logaritmasını maksimize etmek birbirine denk problemler olduğu için olabilirlik fonksiyonunun logaritması alınarak maksimize edilir. Log-olabilirlik fonksiyonu ise

$$\ln L = \sum_{i=1}^n y_i \ln \Phi_i + (1 - y_i) \ln (1 - \Phi_i). \quad (4.12)$$

şeklinde elde edilir. Denklem (4.12)'de verilen fonksiyonu maksimize etmek için her bir model parametresine göre kısmi türevleri alınarak eşitlik 0'a eşitlenir (Aldrich ve ark., 1984) ve

$$\frac{\partial \ln L}{\partial \boldsymbol{\beta}} = \sum_{i=1}^n \mathbf{x}_i \left\{ \frac{y_i \phi_i}{\Phi_i} + \frac{(1 - y_i)(-\phi_i)}{1 - \Phi_i} \right\} = \sum_{i=1}^n x_i \left\{ \frac{y_i - \Phi_i}{\Phi_i(1 - \Phi_i)} \phi_i \right\} = 0 \quad (4.13)$$

elde edilir, burada Φ_i birikimli dağılım fonksiyonunu ve ϕ_i ise standart normal dağılım için olasılık yoğunluk fonksiyonunu temsil etmektedir (Greene, 2018). (4.13)'de verilen denklem

sistemi doğrusal yapıda olmadığından dolayı denklemin çözümünde iteratif olarak yeniden ağırlıklandırılmış en küçük kareler algoritması aşağıdaki gibi kullanılabilir (Hardin ve Hilbe, 2007).

İkili probit regresyon model için IRLS algoritması

- 1: Dev = 0
 - 2: $\mu_i = (y_i + 0.5)/2, i = 1, 2, \dots, n$ ▷ μ 'nün başlangıç değeri
 - 3: $\eta_i = \Phi(\mu_i), i = 1, 2, \dots, n$ ▷ η 'nın başlangıç değeri
 - 4: While abs(Δ Dev > TOL) do
 - 5: $\phi_i = \exp(-\eta_i^2/2) / \sqrt{2\pi}, i = 1, 2, \dots, n$ ▷ standart normal yoğunluk
 - 6: $\mathbf{W} = \text{diag}(\phi_i^2 / (\mu_i(1 - \mu_i)))$ ▷ ağırlık matrisi
 - 7: $\mathbf{z} = \boldsymbol{\eta} + (\mathbf{y} - \boldsymbol{\mu}) / \phi$ ▷ düzeltilmiş yanıt vektörü
 - 8: $\boldsymbol{\beta} = (\mathbf{X}^\top \mathbf{W} \mathbf{X})^{-1} \mathbf{X} \mathbf{W} \mathbf{z}$ ▷ tahmini katsayı vektörü
 - 9: $\boldsymbol{\eta} = \mathbf{X} \boldsymbol{\beta}$
 - 10: $\boldsymbol{\mu} = \Phi(\boldsymbol{\eta})$
 - 11: OldDev = Dev
 - 12: Dev = $2 \sum_{y_i=1} 1 / \ln(\mu_i) + 2 \sum_{y_i=0} 1 / \ln(1 - \mu_i)$
 - 13: Δ Dev = Dev - OldDev
-

Yukarıda verilen algoritmada birbirini izleyen iki sapma istatistiklerinin farkı olan bir yakınsama ölçütünün elde edildiği varsayılarak iterasyonlar durur ve en çok olabilirlik tahmincisi aşağıdaki gibi elde edilir

$$\hat{\boldsymbol{\beta}}_{\text{MLE}} = (\mathbf{X}^\top \widehat{\mathbf{W}} \mathbf{X})^{-1} \mathbf{X} \widehat{\mathbf{W}} \hat{\mathbf{z}}. \quad (4.14)$$

Burada $\hat{\mathbf{z}}$ düzeltilmiş yanıt vektörüdür, öyle ki $\hat{\mathbf{z}} = \boldsymbol{\eta} + (\mathbf{y} - \boldsymbol{\mu}) / \phi$, $\boldsymbol{\mu} = \Phi(\mathbf{X} \hat{\boldsymbol{\beta}}_{\text{MLE}})$, $\boldsymbol{\eta} = \mathbf{X} \hat{\boldsymbol{\beta}}_{\text{MLE}}$ ve $\widehat{\mathbf{W}} = \text{diag} \left(\frac{\phi(\mathbf{x}_i^\top \hat{\boldsymbol{\beta}}_{\text{MLE}})^2}{\Phi(\mathbf{x}_i^\top \hat{\boldsymbol{\beta}}_{\text{MLE}})(1 - \Phi(\mathbf{x}_i^\top \hat{\boldsymbol{\beta}}_{\text{MLE}}))} \right)$. $\boldsymbol{\beta}$ 'nin en çok olabilirlik tahmini asimptotik olarak normal dağılır ve varyans-kovaryans matrisi

$$\text{Cov}(\hat{\boldsymbol{\beta}}_{\text{MLE}}) = (\mathbf{X}^\top \widehat{\mathbf{W}} \mathbf{X})^{-1} \quad (4.15)$$

şeklindedir ve dolayısıyla $\hat{\beta}_{MLE}$ 'nin skaler hata kareler ortalaması

$$\begin{aligned} MSE\left(\hat{\beta}_{MLE}\right) &= E\left(\left(\hat{\beta}_{MLE} - \beta\right)^\top \left(\hat{\beta}_{MLE} - \beta\right)\right) \\ &= tr\left(\left(\mathbf{X}^\top \widehat{\mathbf{W}}\mathbf{X}\right)^{-1}\right) = \sum_{j=1}^p \frac{1}{\lambda_j} \end{aligned} \quad (4.16)$$

ile elde edilir. $tr(\cdot)$ iz operatörünü ve λ_j , ağırlıklı $\mathbf{X}^\top \widehat{\mathbf{W}}\mathbf{X}$ matrisinin özdeğerlerini ifade etmektedir.

4.3. Probit Ridge Tahmin Edicisi

Hoerl ve Kennard (1970a)'ın çalışmalarının ardından Kibria ve Saleh (2012) probit regresyon modeli altında ridge tahmincisini tanımlamıştır. Kibria ve Saleh (2012) probit regresyon modeli için k yanlılık parametresi olmak üzere probit ridge tahmincisini (PRR)

$$\hat{\beta}_{PRR} = \left(\mathbf{X}^\top \widehat{\mathbf{W}}\mathbf{X} + k\mathbf{I}\right)^{-1} \mathbf{X}^\top \widehat{\mathbf{W}}\mathbf{X} \hat{\beta}_{MLE} \quad (4.17)$$

şeklinde önermiştir. Probit regresyon modelde çoklu bağlantı probleminin gözlemlendiği durumlarda MLE'nin kararsız ve değişken bir yapıya sahip olduğu gözlenir. Ayrıca MLE'nin uzunluğu, gerçek parametre vektörüne kıyasla daha fazladır. Bu gibi durumlarda asimptotik olarak yansız olan MLE yerine yanlı tahminciler tercih edilmelidir.

Locking ve ark. (2013), Kibria ve Saleh (2012)'nin önerdiği probit ridge metodundaki yanlılık parametresini (k) tahmin etmek için doğrusal model için önerilen bazı yöntemleri genelleştirmişlerdir. Ayrıca yaptıkları araştırmada kendi önerdikleri tahmincileri Kibria ve Saleh (2012)'in önerdiği ridge tahmincileriyle Monte Carlo simülasyon deneyleri ile karşılaştırmışlardır.

Tezin bu bölümünde, Hoerl ve Kennard (1970a), Singh ve ark. (1986) ve Batah ve ark. (2008)'in çalışmalarında yapıldığı gibi probit regresyon model için geliştirilmiş ridge tahmincisi (GRR) aşağıdaki gibi önerilmiştir

$$\hat{\beta}_{GRR} = \left(\mathbf{X}^\top \widehat{\mathbf{W}}\mathbf{X} + \mathbf{K}\right)^{-1} \mathbf{X}^\top \widehat{\mathbf{W}}\mathbf{X} \hat{\beta}_{MLE}. \quad (4.18)$$

Burada $\mathbf{K} = \text{diag}(k_1, k_2, \dots, k_p)$ her bir açıklayıcı değişken için ridge parametrelerini ihtiva eden köşegen matristir. (4.17)'deki denklemde ridge tahmincisinin aksine, geliştirilmiş tahmincilerin her bir parametre için farklı miktarda büzülme sağladığı açık olarak görülmektedir.

4.3.1. Probit Ridge Tahmincisinin Beklenen Deęeri

Probit ridge tahmincisinin beklenen deęeri

$$\begin{aligned} E(\hat{\beta}_{\text{PRR}}) &= E\left[\left(\mathbf{X}^\top \widehat{\mathbf{W}}\mathbf{X} + k\mathbf{I}\right)^{-1} \mathbf{X}^\top \widehat{\mathbf{W}}\mathbf{X} \hat{\beta}_{\text{MLE}}\right] \\ &= \left(\mathbf{X}^\top \widehat{\mathbf{W}}\mathbf{X} + k\mathbf{I}\right)^{-1} \mathbf{X}^\top \widehat{\mathbf{W}}\mathbf{X} E(\hat{\beta}_{\text{MLE}}) \\ &= \left[\left(\mathbf{X}^\top \widehat{\mathbf{W}}\mathbf{X}\right)^{-1} \left(\mathbf{X}^\top \widehat{\mathbf{W}}\mathbf{X} + k\mathbf{I}\right)\right]^{-1} \beta \\ &= \left[\mathbf{I} + k \left(\mathbf{X}^\top \widehat{\mathbf{W}}\mathbf{X}\right)^{-1}\right]^{-1} \beta \\ &= \left[\mathbf{I} - k \left(\mathbf{X}^\top \widehat{\mathbf{W}}\mathbf{X} + k\mathbf{I}\right)^{-1}\right] \beta \end{aligned}$$

olarak elde edilir.

4.3.2. Probit Ridge Tahmincisinin Yanlılıęı

Probit Ridge tahmincisinin yanlılıęı

$$\begin{aligned} \text{Bias}(\hat{\beta}_{\text{PRR}}) &= E(\hat{\beta}_{\text{PRR}}) - \beta \\ &= \left[\mathbf{I} - k \left(\mathbf{X}^\top \widehat{\mathbf{W}}\mathbf{X} + k\mathbf{I}\right)^{-1}\right] \beta - \beta \\ &= \beta - k \left(\mathbf{X}^\top \widehat{\mathbf{W}}\mathbf{X} + k\mathbf{I}\right)^{-1} \beta - \beta \\ &= -k \left(\mathbf{X}^\top \widehat{\mathbf{W}}\mathbf{X} + k\mathbf{I}\right)^{-1} \beta \end{aligned}$$

eşitlikteki gibi ifade edilir.

4.3.3. Probit Ridge Tahmincisinin Varyans–kovaryans matrisi ve Varyansı

Probit ridge tahmincisinin varyans–kovaryans matrisi

$$\begin{aligned}
Cov\left(\widehat{\beta}_{PRR}\right) &= Cov\left[\left(\mathbf{X}^T\widehat{\mathbf{W}}\mathbf{X}+k\mathbf{I}\right)^{-1}\mathbf{X}^T\widehat{\mathbf{W}}\mathbf{X}\widehat{\beta}_{MLE}\right] \\
&= \left(\mathbf{X}^T\widehat{\mathbf{W}}\mathbf{X}+k\mathbf{I}\right)^{-1}\mathbf{X}^T\widehat{\mathbf{W}}\mathbf{X}Cov\left(\widehat{\beta}_{MLE}\right)\mathbf{X}^T\widehat{\mathbf{W}}\mathbf{X}\left(\mathbf{X}^T\widehat{\mathbf{W}}\mathbf{X}+k\mathbf{I}\right)^{-1} \\
&= \left(\mathbf{X}^T\widehat{\mathbf{W}}\mathbf{X}+k\mathbf{I}\right)^{-1}\mathbf{X}^T\widehat{\mathbf{W}}\mathbf{X}\left(\mathbf{X}^T\widehat{\mathbf{W}}\mathbf{X}\right)^{-1}\mathbf{X}^T\widehat{\mathbf{W}}\mathbf{X}\left(\mathbf{X}^T\widehat{\mathbf{W}}\mathbf{X}+k\mathbf{I}\right)^{-1} \\
&= \left(\mathbf{X}^T\widehat{\mathbf{W}}\mathbf{X}+k\mathbf{I}\right)^{-1}\mathbf{X}^T\widehat{\mathbf{W}}\mathbf{X}\left(\mathbf{X}^T\widehat{\mathbf{W}}\mathbf{X}+k\mathbf{I}\right)^{-1}
\end{aligned}$$

olarak elde edilir. Probit ridge tahmincisinin varyansı yukarıdaki formülle oluşan kovaryans matrisinin köşegen elemanlarının toplamı olarak yazılabileceğinden;

$$Var\left(\widehat{\beta}_{PRR}\right) = tr\left(Cov\left(\widehat{\beta}_{PRR}\right)\right) = \sum_{j=1}^p \frac{\lambda_j}{(\lambda_j + k)^2}$$

elde edilir.

4.3.4. Probit Ridge Tahmincisinin Hata Kareler Ortalaması

Probit ridge tahmincisinin matris MSE'si (MMSE)

$$\begin{aligned}
MMSE\left(\widehat{\beta}_{PRR}\right) &= Cov\left(\widehat{\beta}_{PRR}\right) + Bias\left(\widehat{\beta}_{PRR}\right)Bias\left(\widehat{\beta}_{PRR}\right)^T \\
&= \left(\mathbf{X}^T\widehat{\mathbf{W}}\mathbf{X}+k\mathbf{I}\right)^{-1}\mathbf{X}^T\widehat{\mathbf{W}}\mathbf{X}\left(\mathbf{X}^T\widehat{\mathbf{W}}\mathbf{X}+k\mathbf{I}\right)^{-1} \\
&\quad + k^2\left(\mathbf{X}^T\widehat{\mathbf{W}}\mathbf{X}+k\mathbf{I}\right)^{-1}\beta\beta^T\left(\mathbf{X}^T\widehat{\mathbf{W}}\mathbf{X}+k\mathbf{I}\right)^{-1}
\end{aligned}$$

olarak elde edilir. Probit ridge tahmincisinin skaler MSE'si yukarıdaki formülle oluşan MMSE'nin köşegen elemanlarının toplamı olarak yazılabileceğinden

$$\begin{aligned}
MSE\left(\widehat{\beta}_{PRR}\right) &= tr\left(MMSE\left(\widehat{\beta}_{PRR}\right)\right) = \sum_{j=1}^p \frac{\lambda_j}{(\lambda_j + k)^2} + k^2 \sum_{j=1}^p \frac{\beta_j^2}{(\lambda_j + k)^2} \\
&= \vartheta_1(k) + \vartheta_2(k)
\end{aligned}$$

olarak elde edilir. Burada birinci ifade $\vartheta_1(k)$ probit ridge tahmincisinin toplam varyansını ve ikinci ifade $\vartheta_2(k)$ ise toplam karesel yanlılığı gösterir. Aşağıdaki teoremlerde bu fonksiyonların süreklilik ve monotonluk özellikleri incelenmiştir.

Teorem 4.1 *Probit ridge tahmincisinin toplam varyans fonksiyonu $\vartheta_1(k)$, k 'nin sürekli ve monoton azalan fonksiyonudur.*

İspat Probit ridge tahmincisinin toplam varyans fonksiyonu $\vartheta_1(k)$ 'nin k 'ye göre türevi:

$$\frac{d\vartheta_1(k)}{dk} = -2 \sum_{j=1}^p \frac{\lambda_j}{(\lambda_j + k)^3} < 0$$

olarak elde edilir. $\forall k_1 > k_2$ için;

$$\begin{aligned} \vartheta_1(k_1) - \vartheta_1(k_2) &= \sum_{j=1}^p \frac{\lambda_j}{(\lambda_j + k_1)^2} - \sum_{j=1}^p \frac{\lambda_j}{(\lambda_j + k_2)^2} \\ &= \sum_{j=1}^p \frac{\lambda_j ((\lambda_j + k_2)^2 - (\lambda_j + k_1)^2)}{(\lambda_j + k_1)^2 (\lambda_j + k_2)^2} \\ &= (k_2 - k_1) \sum_{j=1}^p \frac{\lambda_j (2\lambda_j + k_1 + k_2)}{(\lambda_j + k_1)^2 (\lambda_j + k_2)^2} < 0 \end{aligned}$$

olur. $\vartheta_1(k_1) < \vartheta_1(k_2)$ olduğundan dolayı $\vartheta_1(k_1)$ monoton azalan bir fonksiyondur. ■

Teorem 4.2 *Probit ridge tahmincisinin toplam karesel yanlılık fonksiyonu $\vartheta_2(k)$, k 'nin artan bir fonksiyonudur.*

İspat Probit ridge tahmincisinin yanlılık karesinin fonksiyonu $\vartheta_2(k)$ 'nin k 'ye göre türevi:

$$\frac{d\vartheta_2(k)}{dk} = 2k \sum_{j=1}^p \frac{\lambda_j \beta_j^2}{(\lambda_j + k)^3} > 0$$

olarak elde edilir. Dolayısıyla $\vartheta_2(k)$ artan bir fonksiyondur. ■

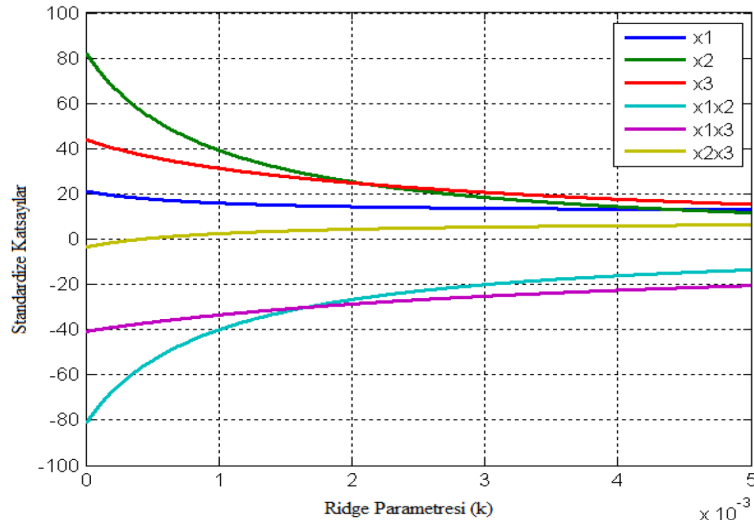
Burada k yanlılık parametresine bağlı olarak elde edilen varyans k 'nin azalan bir fonksiyonu, karesel yanlılık ise k 'nin artan bir fonksiyonudur. Yani k yanlılık parametresi artarken tahminci gerçek parametre değerlerinden uzaklaşmasına rağmen varyans değeri azalacaktır.

4.3.5. Probit Ridge Tahmincisinin Parametre Seçimi

Probit ridge tahmincisinin analiz sonunda elde edilen kestirimlerinin kararlı bir yapıda olması, seçeceğimiz k yanlılık parametresinin optimum değerde olmasına bağlıdır. Uygun k parametresinin seçilmesi durumunda probit ridge tahmincisinin hata kareler ortalaması, en

çok olabilirlik tahmincisinin hata kareler ortalamasından daha küçük bir değere sahip olacaktır. Araştırmalar sonunda bir öneri olarak sunulan probit ridge regresyon analizi yapılırken k yanlılık parametresini belirleyecek kesin bir yöntem yoktur. Bununla beraber probit ridge regresyonda k parametresinin optimal seçimi için literatürde birçok öneri geliştirilmiştir. Bu yöntemlerden bazıları şu şekildedir:

- Ridge İzi Yöntemi: Çoklu doğrusal bağlantı probleminin varlığında regresyon katsayılarının kestirimi etkilenir. Regresyon modeline veri ekleyip çıkarıldığı zaman tahmincilerde değişiklik gözlenir. Ridge regresyonun en büyük avantajlarından biri üzerinde çalışılan veri için kararsız yapıdaki $\hat{\beta}_{MLE}$ katsayılarını görmeye yardımcı olan ve ridge izi (ridge trace) olarak bilinen grafiksel gösterimin olmasıdır. Hoerl ve Kennard (1970a) 'ın önerdiği ve ridge tahmincisinin grafiksel gösterimi olan ridge izi; regresyon katsayılarını ifade eden $\hat{\beta}_{PRR}$ parametreleri düşey ekseninde, k yanlılık parametreleri yatay ekseninde olacak şekilde iki boyutlu uzayda grafik elde edilmesidir. Ridge izi araştırmacıya hangi katsayıların verilere daha duyarlı olduğu konusunda fikir verir. Ridge izi grafiklerinde yanlı regresyon katsayıları, k yanlılık parametresinin bir fonksiyonu olarak ifade edilir (Hoerl ve Kennard, 1970b).



Şekil 4.2. Ridge İzi.

Standartlaştırılmış katsayı parametreleri küçük k değerleriyle anormal olarak değişir, sonrasında durağan bir hal alır. Katsayı parametrelerinin durağanlaşmaya başladığı bu noktada en küçük k değeri en uygun değer olarak seçilmelidir. En uygun k yanlılık parametresinin seçiminde kullanılan diğer kriterler arasında durağanlığı, makul büyüklüğü, kabul edilebilir hata kareler toplamını sağlayan k sabiti yaklaşımıdır. R

programı yardımıyla elde edilecek k yanlılık parametreleriyle en uygun k değerini yaklaşık olarak bulunabilir. Uygulamada tüm kriterlerin birlikte değerlendirilmesi en sağlıklı yaklaşım olarak kabul edilmektedir. k yanlılık parametresinin seçiminde dikkat edilmesi gereken unsurlar şöyledir:

- Seçilen k parametresinde sistem durağan bir hale gelip sabitleşmeli ve ortogonal sistem özelliklerini taşımalıdır.
- Katsayı parametreleri bağımlı değişken üzerinde meydana gelen değişimi ifade ettiğinden mutlak değerce makul bir büyüklüğe sahip olmalıdır.
- $k = 0$ değerini aldığı anda yanlış işarete sahip olduğu düşünülen katsayılar, uygun k parametresi için işaret değiştirmiş bir yapıda olmalıdır.
- Artık kareler toplamı en küçük değerinden fazla uzaklaşmamalıdır (Albayrak, 2012; Hoerl ve Kennard, 1970a; McDonald ve Galarneau, 1975; Türkay, 1996).
- Lineer regresyon modeli için birçok k parametresi önerilmiştir ve bunların bazıları probit ridge tahmincisi için genelleştirerek verilecektir. İlk olarak Hoerl ve Kennard (1970a)

$$k_1 = \frac{\hat{\sigma}^2}{\hat{\alpha}_{max}^2}$$

eşitliğini önermiştir. Burada $\hat{\sigma}^2$ karesel sapma miktarlar toplamının $(n-p-1)$ serbestlik derecesine bölünmesini ve $\hat{\alpha}_{max}$ en çok olabilirlik tahmincisinin kanonik yapıdaki maksimum tahmini değerlerini ifade eder. Schaefer ve ark. (1984) bu tahmincinin değiştirilmiş bir versiyonunu önermiştir:

$$k_2 = \frac{1}{\hat{\alpha}_{max}^2}.$$

Locking ve ark. (2013), Kibria (2003)'e dayanarak aşağıdaki iki ridge tahmincisini önermiştir

$$k_3 = \frac{\hat{\sigma}^2}{\left(\prod_{j=1}^p \hat{\alpha}_j^2\right)^{\frac{1}{p}}} \quad \text{ve} \quad k_4 = \text{median}(m_j^2).$$

Burada

$$m_j = \sqrt{\frac{\hat{\sigma}^2}{\hat{\alpha}_j^2}}.$$

Locking ve ark. (2013) ayrıca aşağıda verilen tahmincileri önermiştir

$$k_5 = \max\left(\frac{1}{m_j}\right), \quad k_6 = \max(m_j), \quad k_7 = \prod_{j=1}^p \left(\frac{1}{m_j}\right)^{\frac{1}{p}}, \quad k_8 = \prod_{j=1}^p (m_j)^{\frac{1}{p}},$$

$$k_9 = \text{median} \left(\frac{1}{m_j} \right), \quad k_{10} = \text{median} (m_j), \quad k_{11} = \max \left(\frac{1}{q_j} \right), \quad k_{12} = \max (q_j),$$

$$k_{13} = \prod_{j=1}^p \left(\frac{1}{q_j} \right)^{\frac{1}{p}}, \quad k_{14} = \prod_{j=1}^p (m_j)^{\frac{1}{p}}, \quad k_{15} = \text{median} \left(\frac{1}{q_j} \right), \quad k_{16} = \text{median} (q_j).$$

Burada $q_j = \frac{\lambda_{max}}{(n-p)\hat{\sigma}^2 + \lambda_{max}\hat{\alpha}_j^2}$ ve λ_{max} , $\mathbf{X}^\top \widehat{\mathbf{W}} \mathbf{X}$ 'in maksimum özdeğeri olarak tanımlanır.



5. PROBİT JACKKNİFE RİDGE TAHMİN EDİCİSİ

Ridge regresyon tahmin edicisi, çoklu doğrusal bağlantı probleminin varlığında en çok olabilirlik tahmin edicisine alternatif olarak kullanılmasına rağmen yanlış bir tahmin tekniğidir. Bu nedenle Quenouille (1956) ve Tukey (1958) tarafından önerilen jackknife prosedürü kullanılarak ridge tahmin edicisinin sebep olduğu bu sapma miktarı azaltılabilir. Jackknife ridge tahmin edicisi uygulanırken örneklemdaki i . gözlem vektörü silinir ve bu gözlem vektörü olmadan katsayı parametrelerinin tahmin edicisi hesaplanarak en sonda bu tahmin edicilerin ortalaması hesaplanır.

\mathbf{X}_{-i} , $\widehat{\mathbf{W}}_{-i}$ ve $\widehat{\mathbf{z}}_{-i}$; i . satırı silinmiş veri matrisini, i . diyagonal elemanı silinmiş ağırlık matrisini ve i . elemanı silinmiş $\widehat{\mathbf{z}}$ vektörünü gösterebilir. Böylece i . gözlem değeri olmaksızın probit ridge regresyon tahmin edicisi

$$\widehat{\beta}_{\text{PRR}}^{-i} = \left(\mathbf{X}_{-i}^{\top} \widehat{\mathbf{W}}_{-i} \mathbf{X}_{-i} + k\mathbf{I} \right)^{-1} \mathbf{X}_{-i} \widehat{\mathbf{W}}_{-i} \widehat{\mathbf{z}}_{-i} \quad (5.1)$$

denklem (5.1)'teki gibi tanımlanır. Burada $\mathbf{X}_{-i} \widehat{\mathbf{W}}_{-i} \widehat{\mathbf{z}}_{-i} = \mathbf{X} \widehat{\mathbf{W}} \widehat{\mathbf{z}} - \mathbf{x}'_i \omega_i \widehat{z}_i$ 'dir ve öyle ki ω_i , $\widehat{\mathbf{W}}$ matrisinin i . diyagonal elemanını ifade etmektedir. Sherman-Morrison-Woodbury teoreminden (bkz. Sherman ve Morrison (1950) ve Woodbury (1950)) faydalanarak

$$\begin{aligned} \left(\mathbf{X}_{-i}^{\top} \widehat{\mathbf{W}}_{-i} \mathbf{X}_{-i} + k\mathbf{I} \right)^{-1} &= \left(\mathbf{X}^{\top} \widehat{\mathbf{W}} \mathbf{X} + k\mathbf{I} \right)^{-1} \\ &+ \frac{\left(\mathbf{X}^{\top} \widehat{\mathbf{W}} \mathbf{X} + k\mathbf{I} \right)^{-1} \mathbf{x}_i \mathbf{x}'_i \left(\mathbf{X}^{\top} \widehat{\mathbf{W}} \mathbf{X} + k\mathbf{I} \right)^{-1} \omega_i}{1 - \mathbf{x}'_i \left(\mathbf{X}^{\top} \widehat{\mathbf{W}} \mathbf{X} + k\mathbf{I} \right)^{-1} \mathbf{x}_i \omega_i} \end{aligned}$$

denklemini elde edilir ve bazı cebirsel hesaplamalar sonucu

$$\widehat{\beta}_{\text{JRR}}^{-i} = \widehat{\beta}_{\text{PRR}} - \frac{\left(\mathbf{X}^{\top} \widehat{\mathbf{W}} \mathbf{X} + k\mathbf{I} \right)^{-1} \mathbf{x}_i \omega_i \left(\widehat{z}_i - \mathbf{x}'_i \widehat{\beta}_{\text{PRR}} \right)}{1 - \mathbf{x}'_i \left(\mathbf{X}^{\top} \widehat{\mathbf{W}} \mathbf{X} + k\mathbf{I} \right)^{-1} \mathbf{x}_i \omega_i} \quad (5.2)$$

eşitliği elde edilir. Hinkley (1977)'in fikri kullanılarak probit regresyon modeli için jackknife ridge tahmin edicisi (JRR)

$$\begin{aligned} \widehat{\beta}_{\text{JRR}} &= \widehat{\beta}_{\text{PRR}} + \left(\mathbf{X}^{\top} \widehat{\mathbf{W}} \mathbf{X} + k\mathbf{I} \right)^{-1} \sum_{i=1}^n \mathbf{x}_i \omega_i \left(\widehat{z}_i - \mathbf{x}'_i \widehat{\beta}_{\text{PRR}} \right) \\ &= \left(\mathbf{I} + k\mathbf{C}_k^{-1} \right) \widehat{\beta}_{\text{PRR}} \\ &= \left(\mathbf{I} - k^2 \mathbf{C}_k^{-2} \right) \widehat{\beta}_{\text{MLE}} \end{aligned} \quad (5.3)$$

(5.3)'deki gibi önerilmiştir. Burada $\mathbf{C} = \mathbf{X}^\top \widehat{\mathbf{W}}\mathbf{X}$ ve $\mathbf{C}_k = \mathbf{C} + k\mathbf{I}$ 'dir.

Ayrıca Hoerl ve Kennard (1970a), Singh ve ark. (1986) ve Batah ve ark. (2008)'ın ardından probit regresyon model için geliştirilmiş ridge tahmircisi ve geliştirilmiş jackknife ridge tahmircisini sırasıyla aşağıdaki gibi tanımlanabilir

$$\widehat{\beta}_{\text{GRR}} = \left(\mathbf{X}^\top \widehat{\mathbf{W}}\mathbf{X} + \mathbf{K} \right)^{-1} \mathbf{X}^\top \widehat{\mathbf{W}}\mathbf{X} \widehat{\beta}_{\text{MLE}} \quad (5.4)$$

$$\widehat{\beta}_{\text{GJR}} = \left(\mathbf{I} - \mathbf{K}^2 (\mathbf{C} + \mathbf{K})^{-2} \right) \widehat{\beta}_{\text{MLE}}. \quad (5.5)$$

Burada $\mathbf{K} = \text{diag}(k_1, k_2, \dots, k_p)$ her bir tahmirci için ridge parametrelerini ihtiva eden diyagonal matristir. Denklem (5.4) ve (5.5)'te önerilen geliştirilmiş ridge ve geliştirilmiş jackknife ridge tahmircilerinin, ridge ve jackknife ridge tahmin edicilerinin aksine her bir parametre için farklı miktarda büzülme sağladığı açık olarak görülmektedir. JRR, GRR ve GJR tahmin yöntemleri ilk defa bu tez çalışmasında önerilmiştir.

5.1. Probit Jackknife Ridge Tahmircisinin Beklenen Değeri

Probit jackknife ridge tahmircisinin beklenen değeri

$$\begin{aligned} E\left(\widehat{\beta}_{\text{JRR}}\right) &= E\left[\left(\mathbf{I} + k\left(\mathbf{X}^\top \widehat{\mathbf{W}}\mathbf{X} + k\mathbf{I}\right)^{-1}\right)\widehat{\beta}_{\text{RR}}\right] \\ &= E\left[\left(\mathbf{I} + k\left(\mathbf{X}^\top \widehat{\mathbf{W}}\mathbf{X} + k\mathbf{I}\right)^{-1}\right)\left(\mathbf{I} - k\left(\mathbf{X}^\top \widehat{\mathbf{W}}\mathbf{X} + k\mathbf{I}\right)^{-1}\right)\widehat{\beta}_{\text{MLE}}\right] \\ &= \left[\left(\mathbf{I} + k\left(\mathbf{X}^\top \widehat{\mathbf{W}}\mathbf{X} + k\mathbf{I}\right)^{-1}\right)\left(\mathbf{I} - k\left(\mathbf{X}^\top \widehat{\mathbf{W}}\mathbf{X} + k\mathbf{I}\right)^{-1}\right)\right] E\left(\widehat{\beta}_{\text{MLE}}\right) \\ &= \left[\mathbf{I} - k^2\left(\mathbf{X}^\top \widehat{\mathbf{W}}\mathbf{X} + k\mathbf{I}\right)^{-2}\right]\beta \end{aligned}$$

olarak elde edilir.

5.2. Jackknife Ridge Tahmincisinin Yanlılığı

Probit jackknife ridge tahmincisinin yanlılığı

$$\begin{aligned} Bias(\hat{\beta}_{JRR}) &= E(\hat{\beta}_{JRR}) - \beta \\ &= \left[\mathbf{I} - k^2 (\mathbf{X}^\top \widehat{\mathbf{W}} \mathbf{X} + k\mathbf{I})^{-2} \right] \beta - \beta \\ &= \beta - k^2 (\mathbf{X}^\top \widehat{\mathbf{W}} \mathbf{X} + k\mathbf{I})^{-2} \beta - \beta \\ &= -k^2 (\mathbf{X}^\top \widehat{\mathbf{W}} \mathbf{X} + k\mathbf{I})^{-2} \beta \end{aligned}$$

eşitlikteki gibi ifade edilir.

5.3. Probit Jackknife Ridge Tahmincisinin Varyansı

Probit jackknife ridge tahmincisinin varyans–kovaryans matrisi

$$\begin{aligned} Cov(\hat{\beta}_{JRR}) &= Cov \left[\left(\mathbf{I} - k^2 (\mathbf{X}^\top \widehat{\mathbf{W}} \mathbf{X} + k\mathbf{I})^{-2} \right) \hat{\beta}_{MLE} \right] \\ &= \left[\mathbf{I} - k^2 (\mathbf{X}^\top \widehat{\mathbf{W}} \mathbf{X} + k\mathbf{I})^{-2} \right] Cov(\hat{\beta}_{MLE}) \left[\mathbf{I} - k^2 (\mathbf{X}^\top \widehat{\mathbf{W}} \mathbf{X} + k\mathbf{I})^{-2} \right] \\ &= \left[\mathbf{I} - k^2 (\mathbf{X}^\top \widehat{\mathbf{W}} \mathbf{X} + k\mathbf{I})^{-2} \right] (\mathbf{X}^\top \widehat{\mathbf{W}} \mathbf{X})^{-1} \left[\mathbf{I} - k^2 (\mathbf{X}^\top \widehat{\mathbf{W}} \mathbf{X} + k\mathbf{I})^{-2} \right] \\ &= \left[\mathbf{I} - k^2 (\mathbf{X}^\top \widehat{\mathbf{W}} \mathbf{X} + k\mathbf{I})^{-2} \right] (\mathbf{X}^\top \widehat{\mathbf{W}} \mathbf{X})^{-1} \left[\mathbf{I} - k^2 (\mathbf{X}^\top \widehat{\mathbf{W}} \mathbf{X} + k\mathbf{I})^{-2} \right] \end{aligned}$$

olarak elde edilir. Probit jackknife ridge tahmincisinin varyansı yukarıda elde edilen kovaryans matrisinin köşegen elemanlarının toplamı olarak yazılabileceğinden

$$\begin{aligned} Var(\hat{\beta}_{JRR}) &= tr(Cov(\hat{\beta}_{JRR})) = \sum_{j=1}^p \left(1 - \frac{k^2}{(\lambda_j + k)^2} \right) \frac{1}{\lambda_j} \left(1 - \frac{k^2}{(\lambda_j + k)^2} \right) \\ &= \sum_{j=1}^p \left(\frac{(\lambda_j + k)^2 - k^2}{(\lambda_j + k)^2} \right) \frac{1}{\lambda_j} \left(\frac{(\lambda_j + k)^2 - k^2}{(\lambda_j + k)^2} \right) \\ &= \sum_{j=1}^p \frac{(\lambda_j + 2k) \lambda_j}{(\lambda_j + k)^2} \frac{1}{\lambda_j} \frac{(\lambda_j + 2k) \lambda_j}{(\lambda_j + k)^2} \\ &= \sum_{j=1}^p \frac{\lambda_j (\lambda_j + 2k)^2}{(\lambda_j + k)^4} \end{aligned}$$

şeklinde elde edilir.

5.4. Probit Jackknife Ridge Tahmincisinin Hata Kareler Ortalaması

Probit jackknife ridge tahmincisinin matris MSE'si

$$\begin{aligned} MMSE(\widehat{\beta}_{JRR}) &= Cov(\widehat{\beta}_{JRR}) + Bias(\widehat{\beta}_{JRR}) Bias(\widehat{\beta}_{JRR})^\top \\ &= \left[\mathbf{I} - k^2 (\mathbf{X}^\top \widehat{\mathbf{W}} \mathbf{X} + k\mathbf{I})^{-2} \right] (\mathbf{X}^\top \widehat{\mathbf{W}} \mathbf{X})^{-1} \left[\mathbf{I} - k^2 (\mathbf{X}^\top \widehat{\mathbf{W}} \mathbf{X} + k\mathbf{I})^{-2} \right] \\ &\quad + k^4 (\mathbf{X}^\top \widehat{\mathbf{W}} \mathbf{X} + k\mathbf{I})^{-2} \beta \beta^\top (\mathbf{X}^\top \widehat{\mathbf{W}} \mathbf{X} + k\mathbf{I})^{-2} \end{aligned}$$

olarak elde edilir. Probit jackknife ridge tahmincisinin skaler MSE'si

$$\begin{aligned} MSE(\widehat{\beta}_{JRR}) &= \sum_{j=1}^p \left(1 - \frac{k^2}{(\lambda_j + k)^2} \right) \frac{1}{\lambda_j} \left(1 - \frac{k^2}{(\lambda_j + k)^2} \right) + k^4 \sum_{j=1}^p \frac{\beta_j^2}{(\lambda_j + k)^4} \\ &= \sum_{j=1}^p \left(\frac{(\lambda_j + k)^2 - k^2}{(\lambda_j + k)^2} \frac{1}{\lambda_j} \frac{(\lambda_j + k)^2 - k^2}{(\lambda_j + k)^2} \right) + k^4 \sum_{j=1}^p \frac{\beta_j^2}{(\lambda_j + k)^4} \\ &= \sum_{j=1}^p \frac{\lambda_j (\lambda_j + 2k)^2}{(\lambda_j + k)^4} + k^4 \sum_{j=1}^p \frac{\beta_j^2}{(\lambda_j + k)^4} \end{aligned}$$

olarak elde edilir. Burada birinci ifade probit jackknifed ridge tahmincisinin toplam varyansını ve ikinci ifade toplam karesel yanlılığı gösterir.

Teorem 5.1 *Probit jackknife ridge tahmincisinin toplam varyans fonksiyonu $v_1(k)$, k 'nin azalan bir fonksiyonudur.*

İspat Probit jackknife ridge tahmincisinin toplam varyans fonksiyonu $v_1(k)$ 'nin k 'ye göre türevi:

$$\frac{dv_1(k)}{dk} = -2 \sum_{j=1}^p \frac{\lambda_j (\lambda_j + 2k) (\lambda_j + 3k)}{(\lambda_j + k)^5} < 0$$

olarak elde edilir. Bu durumda $v_1(k)$, k 'nin azalan bir fonksiyonudur. ■

Teorem 5.2 *Probit jackknife ridge tahmincisinin toplam karesel yanlılık fonksiyonu $v_2(k)$, k 'nin artan fonksiyonudur.*

İspat Probit jackknife ridge tahmincisinin yanlılık karesinin fonksiyonu $v_2(k)$ 'nin k 'ye göre türevi:

$$\frac{dv_2(k)}{dk} = 4k^3 \sum_{j=1}^p \frac{\lambda_j \beta_j^2}{(\lambda_j + k)^5} > 0$$

olarak elde edilir. Dolayısıyla $v_2(k)$ artan bir fonksiyondur. ■

Burada k yanlılık parametresine bağlı olarak varyansın k 'nın azalan bir fonksiyonu, karesel yanlılığın ise k 'nın artan bir fonksiyonu olduğu anlaşılmıştır. Yani k yanlılık parametresi artarken tahminci gerçek parametre değerlerinden uzaklaşmasına rağmen varyans değeri azalacaktır.

5.5. MSE Özellikleri ve Tahminciler Arasındaki Teorik

Karşılaştırmalar

Bu bölümde, önerilen tahmincilerin kanonik formdaki MSE ve MMSE fonksiyonları elde edilmiştir. Skaler MSE ve matris MSE fonksiyonları elde edilmeden önce Batah ve ark. (2008), Khurana ve ark. (2014) ve Türkan ve Özel (2018)'i takip ederek modelin kanonik formu elde edilmiştir. Model kanonik forma çevrilirken $\mathbf{T}^\top \mathbf{X}^\top \widehat{\mathbf{W}} \mathbf{X} \mathbf{T} = \mathbf{Z}^\top \widehat{\mathbf{W}} \mathbf{Z} = \mathbf{\Lambda}$ ve $\mathbf{Z} = \mathbf{X} \mathbf{T}$ dönüşümü yapılır. Burada $\mathbf{\Lambda} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_p)$ ve \mathbf{T} sütunları $\mathbf{X}^\top \widehat{\mathbf{W}} \mathbf{X}$ matrisinin $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p > 0$ sıralı özdeğerlerine karşılık gelen normalize edilmiş özvektörlerden oluşan $p \times p$ boyutlu ortogonal matristir. Böylece kanonik haldeki probit regresyon modelin katsayı parametreleri $\boldsymbol{\alpha} = \mathbf{T}^\top \boldsymbol{\beta}$ ile ifade edilebilir. Bu dönüşüm, probit ridge ve probit jackknife ridge tahmincisine uygulanırsa bu iki tahmincinin kanonik formu sırasıyla

$$\widehat{\boldsymbol{\alpha}}_{\text{PRR}} = \mathbf{\Lambda}_k^{-1} \mathbf{\Lambda} \widehat{\boldsymbol{\alpha}}_{\text{MLE}} \quad (5.6)$$

$$\widehat{\boldsymbol{\alpha}}_{\text{JRR}} = (\mathbf{I} - k^2 \mathbf{\Lambda}_k^{-2}) \widehat{\boldsymbol{\alpha}}_{\text{MLE}} \quad (5.7)$$

(5.6) ve (5.7)'deki gibi elde edilir. Burada $\widehat{\boldsymbol{\alpha}}_{\text{MLE}} = \mathbf{T}^\top \widehat{\boldsymbol{\beta}}_{\text{MLE}}$ ve $\mathbf{\Lambda}_k = \mathbf{\Lambda} + k\mathbf{I}$ 'dir. Tahmin edicilerin skaler MSE ve matris MSE fonksiyonlarını elde edebilmek için tahmin edicilerin yanlılık vektörüne ve varyans-kovaryans matrisine gereksinim vardır. $\widehat{\boldsymbol{\alpha}}_{\text{MLE}}$ asimptotik olarak yansızdır ve $\widehat{\boldsymbol{\alpha}}_{\text{MLE}}$ 'nin varyans-kovaryans matrisi Denklem (4.15)'ten $\text{Cov}(\widehat{\boldsymbol{\alpha}}_{\text{MLE}}) = \mathbf{\Lambda}^{-1}$ şeklinde elde edilir. $\widehat{\boldsymbol{\alpha}}_{\text{MLE}}$ 'nin tüm özdeğerleri pozitif olduğundan $\mathbf{\Lambda}^{-1}$ matrisinin de pozitif tanımlı olduğu dikkate alınmalıdır. PRR ve JRR'nin yanlılık vektörleri ve varyans-

kovaryans matrisleri sırasıyla

$$Bias(\hat{\alpha}_{PRR}) = -k\Lambda_k^{-1}\alpha \quad (5.8)$$

$$Cov(\hat{\alpha}_{PRR}) = \Lambda_k^{-1}\Lambda\Lambda_k^{-1} \quad (5.9)$$

$$Bias(\hat{\alpha}_{JRR}) = -k^2\Lambda_k^{-2}\alpha \quad (5.10)$$

$$Cov(\hat{\alpha}_{JRR}) = (\mathbf{I} - k^2\Lambda_k^{-2})\Lambda^{-1}(\mathbf{I} - k^2\Lambda_k^{-2}). \quad (5.11)$$

olarak elde edilir. En çok olabilirlik tahmin edicisinin MMSE fonksiyonu aşağıdaki gibi elde edilir

$$MMSE(\hat{\alpha}_{MLE}) = \Lambda^{-1}. \quad (5.12)$$

(5.8)–(5.11)’deki denklemler kullanılarak probit ridge ve probit jackknife ridge tahmin edicilerinin MMSE fonksiyonları

$$MMSE(\hat{\alpha}_{PRR}) = \Lambda_k^{-1}\Lambda\Lambda_k^{-1} + k^2\Lambda_k^{-1}\alpha\alpha^\top\Lambda_k^{-1} \quad (5.13)$$

$$MMSE(\hat{\alpha}_{JRR}) = (\mathbf{I} - k^2\Lambda_k^{-2})\Lambda^{-1}(\mathbf{I} - k^2\Lambda_k^{-2}) + k^4\Lambda_k^{-2}\alpha\alpha^\top\Lambda_k^{-2} \quad (5.14)$$

olarak ifade edilir. Skaler MSE, MMSE’nin izi olduğundan (5.13) ve (5.14) denklemlerinin izleri alınarak probit ridge ve probit jackknife ridge tahmincilerinin MSE fonksiyonları sırasıyla

$$MSE(\hat{\alpha}_{PRR}) = \sum_{j=1}^p \frac{\lambda_j}{(\lambda_j + k)^2} + k^2 \sum_{j=1}^p \frac{\alpha_j^2}{(\lambda_j + k)^2} \quad (5.15)$$

$$MSE(\hat{\alpha}_{JRR}) = \sum_{j=1}^p \frac{\lambda_j(\lambda_j + 2k)^2}{(\lambda_j + k)^4} + k^4 \sum_{j=1}^p \frac{\alpha_j^2}{(\lambda_j + k)^4} \quad (5.16)$$

şeklinde elde edilir ki bu eşitliklerde birinci terimler tahmincilerin toplam varyansını, ikinci terimler tahmincilerin toplam karesel yanlılığını göstermektedir.

5.5.1. Tahmincilerin Yanlılık(Bias) Karşılaştırmaları

Bu tezde probit regresyon modelde ridge tahmin edicisinin sebep olduğu yanlılığı azaltabilmek için jackknife ridge tahmin edicisi önerilmiştir. Probit ridge tahmincisi ve probit jackknife ridge tahmin edicisinin yanlılık karşılaştırması aşağıdaki teoremden sunulmuştur. Teorem 5.3 ve 5.4’e göre probit jackknife ridge tahmin edicisinin her bir parametresinin mutlak sapması ve toplam karesel yanlılığı probit ridge tahmin edicisine göre daha az olduğu ispatlanmıştır.

Teorem 5.3 *Probit regresyon modelinde jackknife ridge tahmin edicisinin toplam karesel yanlılığı, ridge tahmin edicisinin toplam karesel yanlılığından daha küçüktür, yani $\forall k > 0$*

$$\|Bias(\hat{\alpha}_{JRR})\|^2 < \|Bias(\hat{\alpha}_{PRR})\|^2.$$

İspat Probit ridge tahmin edicisi ile probit jackknife ridge tahmin edicisinin toplam karesel yanlılıklarının farkı alınacak olursa

$$\begin{aligned} \|Bias(\hat{\alpha}_{PRR})\|^2 - \|Bias(\hat{\alpha}_{JRR})\|^2 &= \sum_{j=1}^p \frac{k^2 \alpha_j^2}{(\lambda_j + k)^2} - \frac{k^4 \alpha_j^2}{(\lambda_j + k)^4} \\ &= \sum_{j=1}^p \frac{k^2 \alpha_j^2 (\lambda_j + k)^2 - k^4 \alpha_j^2}{(\lambda_j + k)^4} \\ &= \sum_{j=1}^p \frac{k^2 \alpha_j^2 [(\lambda_j + k)^2 - k^2]}{(\lambda_j + k)^4} \\ &= \sum_{j=1}^p \frac{k^2 \alpha_j^2 \lambda_j (\lambda_j + 2k)}{(\lambda_j + k)^4} \end{aligned}$$

elde edilir ki bu eşitlik $\forall k > 0$ için pozitif bir değere sahiptir. ■

Teorem 5.4 *Probit regresyon modelde jackknife ridge tahmin edicisinin yanlılığının j . bileşenin mutlak değeri, ridge tahmin edicisinin yanlılığının j . bileşenin mutlak değerinden daha küçüktür.*

$$|Bias(\hat{\alpha}_{JRR}^j)| < |Bias(\hat{\alpha}_{PRR}^j)|, \quad \forall k > 0.$$

İspat Yanlılık vektörlerinin j . bileşeninin mutlak değerleri arasındaki fark

$$\begin{aligned} |Bias(\hat{\alpha}_{PRR}^j)| - |Bias(\hat{\alpha}_{JRR}^j)| &= \left| -k(\Lambda + k\mathbf{I})^{-1} \alpha_j - \left| -k^2(\Lambda + k\mathbf{I})^{-2} \alpha_j \right| \right| \\ &= \left| \left[\frac{k}{\lambda_j + k} - \frac{k^2}{(\lambda_j + k)^2} \right] \alpha_j \right| \\ &= \left| \frac{k(\lambda_j + k) - k^2}{(\lambda_j + k)^2} \alpha_j \right| \\ &= \frac{k\lambda_j}{(\lambda_j + k)^2} |\alpha_j| \end{aligned}$$

olarak elde edilir ve bu eşitlik $\forall k > 0$ için pozitif bir değere sahiptir. Burada $|\alpha_j|$, α 'nın j . bileşeninin mutlak değeridir. ■

5.5.2. Tahmincilerin Varyans Karşılaştırmaları

Herhangi bir $\widehat{\beta}^*$ tahmincisinin toplam varyansı, varyans–kovaryans matrisinin izi olarak tanımlanır, öyle ki $Var(\widehat{\beta}^*) = tr(Cov(\widehat{\beta}^*))$ 'dir. En çok olabilirlik tahmin edicisi, probit ridge tahmin edicisi ve probit jackknife ridge tahmin edicisinin toplam varyans karşılaştırması aşağıdaki teoremlerde sunulmaktadır.

Teorem 5.5 *Probit regresyon modelde jackknife ridge tahmin edicisinin toplam varyansı, en çok olabilirlik tahmin edicisinin toplam varyansından daha küçüktür, yani*

$$Var(\widehat{\alpha}_{JRR}) < Var(\widehat{\alpha}_{MLE}), \quad \forall k > 0.$$

İspat MLE tahmin edicisinin toplam varyansı

$$Var(\widehat{\alpha}_{MLE}) = tr(Cov(\widehat{\alpha}_{MLE})) = \sum_{j=1}^p \frac{1}{\lambda_j}$$

ve JRR tahmin edicisinin toplam varyansı

$$Var(\widehat{\alpha}_{JRR}) = tr(Cov(\widehat{\alpha}_{JRR})) = \sum_{j=1}^p \frac{\lambda_j (\lambda_j + 2k)^2}{(\lambda_j + k)^4}$$

olmak üzere, MLE ile JRR tahmin edicilerinin toplam varyansları arasındaki fark

$$\begin{aligned} Var(\widehat{\alpha}_{MLE}) - Var(\widehat{\alpha}_{JRR}) &= \sum_{j=1}^p \frac{1}{\lambda_j} - \sum_{j=1}^p \frac{\lambda_j (\lambda_j + 2k)^2}{(\lambda_j + k)^4} \\ &= \sum_{j=1}^p \frac{(\lambda_j + k)^4}{\lambda_j (\lambda_j + k)^4} - \sum_{j=1}^p \frac{\lambda_j^2 (\lambda_j + 2k)^2}{\lambda_j (\lambda_j + k)^4} \\ &= \sum_{j=1}^p \frac{(\lambda_j + k)^2 (\lambda_j + k)^2}{\lambda_j (\lambda_j + k)^4} - \sum_{j=1}^p \frac{\lambda_j^2 (\lambda_j + 2k)^2}{\lambda_j (\lambda_j + k)^4} \\ &= \sum_{j=1}^p \frac{k^4 + 4k^3 \lambda_j + 2k^2 \lambda_j^2}{\lambda_j (\lambda_j + k)^4} \\ &= k^2 \sum_{j=1}^p \frac{(2\lambda_j^2 + 4k\lambda_j + k^2)}{\lambda_j (\lambda_j + k)^4} \end{aligned}$$

eşitlikteki gibi elde edilir ki bu sonuç $\forall k > 0$ için pozitiftir. ■

Teorem 5.6 *Probit regresyon modelde jackknife ridge tahmin edicisinin toplam varyansı, ridge tahmin edicisinin toplam varyansından daha büyüktür, yani*

$$Var(\widehat{\alpha}_{PRR}) < Var(\widehat{\alpha}_{JRR}) > 0, \quad \forall k > 0.$$

İspat PRR tahmin edicisinin toplam varyansı

$$Var(\hat{\alpha}_{PRR}) = tr(Cov(\hat{\alpha}_{PRR})) = \sum_{j=1}^p \frac{\lambda_j}{(\lambda_j + k)^2}$$

ve JRR tahmin edicisinin toplam varyansı

$$Var(\hat{\alpha}_{JRR}) = tr(Cov(\hat{\alpha}_{JRR})) = \sum_{j=1}^p \frac{\lambda_j (\lambda_j + 2k)^2}{(\lambda_j + k)^4}$$

olmak üzere, JRR ve PRR tahmin edicilerinin toplam varyansları arasındaki fark

$$\begin{aligned} Var(\hat{\alpha}_{JRR}) - Var(\hat{\alpha}_{PRR}) &= \sum_{j=1}^p \frac{\lambda_j (\lambda_j + 2k)^2}{(\lambda_j + k)^4} - \sum_{j=1}^p \frac{\lambda_j}{(\lambda_j + k)^2} \\ &= \sum_{j=1}^p \frac{\lambda_j (\lambda_j + 2k)^2}{(\lambda_j + k)^4} - \sum_{j=1}^p \frac{\lambda_j (\lambda_j + k)^2}{(\lambda_j + k)^4} \\ &= \sum_{j=1}^p \frac{\lambda_j (\lambda_j^2 + 4k\lambda_j + 4k^2)}{(\lambda_j + k)^4} - \sum_{j=1}^p \frac{\lambda_j (\lambda_j^2 + 2k\lambda_j + k^2)}{(\lambda_j + k)^4} \\ &= k \sum_{j=1}^p \frac{\lambda_j (2\lambda_j + 3k)}{(\lambda_j + k)^4} \end{aligned}$$

eşitlikteki gibi elde edilir ki bu sonuç $\forall k > 0$ için pozitiftir. ■

5.5.3. Tahmincilerin MSE ve MMSE Karşılaştırmaları

Tahmin ediciler arasında karşılaştırma yapabilmek için MSE ve MMSE fonksiyonlarından faydalanılabilir. Probit jackknife ridge tahmin edicisinin en çok olabilirlik ve probit ridge tahmin edicisinden MSE ve MMSE kriterleri bakımından daha üstün olduğu koşullarını sağlayan teoremlerin ispatlarında aşağıdaki lemmalardan faydalanılacaktır.

Lemma 5.1 (Farebrother, 1976). \mathbf{M} pozitif tanımlı bir matris ve $\boldsymbol{\theta}$ uygun boyutlu bir vektör olmak üzere; $\mathbf{M} - \boldsymbol{\theta}\boldsymbol{\theta}^\top > \mathbf{0}$ olması için gerek ve yeter koşul $\boldsymbol{\theta}^\top \mathbf{M}^{-1} \boldsymbol{\theta} < 1$ eşitsizliğinin sağlanmasıdır.

Lemma 5.2 (Trenkler ve Toutenburg, 1990). $\hat{\boldsymbol{\beta}}_1$ ve $\hat{\boldsymbol{\beta}}_2$, $\boldsymbol{\beta}$ vektörünün herhangi iki tahmincisi olsun. Ayrıca tahmincilerin varyans–kovaryans matrislerinin farkları $\mathbf{D} = Cov(\hat{\boldsymbol{\beta}}_1) - Cov(\hat{\boldsymbol{\beta}}_2)$ pozitif tanımlı bir matris olsun. $\mathbf{a}_1 = Bias(\hat{\boldsymbol{\beta}}_1)$ ve $\mathbf{a}_2 = Bias(\hat{\boldsymbol{\beta}}_2)$ olmak üzere $MMSE(\hat{\boldsymbol{\beta}}_1) - MMSE(\hat{\boldsymbol{\beta}}_2)$ farkının pozitif tanımlı olması için gerek ve yeter koşul $\mathbf{a}_2^\top (\mathbf{D} + \mathbf{a}_1 \mathbf{a}_1^\top)^{-1} \mathbf{a}_2 < 1$ eşitsizliğinin sağlanmasıdır.

Teorem 5.7 *Probit regresyon modelde jackknife ridge tahmin edicisinin MMSE kriterine göre MLE tahmin edicisinden üstün olması için gerek ve yeter koşul,*

$$k^2 \boldsymbol{\alpha}^\top [2\boldsymbol{\Lambda} + 4k\mathbf{I} + k^2 \boldsymbol{\Lambda}^{-1}]^{-1} \boldsymbol{\alpha} < 1.$$

eşitsizliğinin sağlanmasıdır.

İspat (5.12) ve (5.14) denklemleri kullanılarak MLE ve JRR tahmin edicilerinin MMSE kriterine göre $\Delta_1 = MMSE(\hat{\boldsymbol{\alpha}}_{MLE}) - MMSE(\hat{\boldsymbol{\alpha}}_{JRR})$ farkının pozitif tanımlı olması için gerek ve yeter koşul incelenecek olursa

$$\begin{aligned} \Delta_1 &= \boldsymbol{\Lambda}^{-1} - [(\mathbf{I} - k^2 \boldsymbol{\Lambda}_k^{-2}) \boldsymbol{\Lambda}^{-1} (\mathbf{I} - k^2 \boldsymbol{\Lambda}_k^{-2}) + k^4 \boldsymbol{\Lambda}_k^{-2} \boldsymbol{\alpha} \boldsymbol{\alpha}^\top \boldsymbol{\Lambda}_k^{-2}] \\ &= \boldsymbol{\Lambda}^{-1} - [(\boldsymbol{\Lambda}^{-1} - k^2 \boldsymbol{\Lambda}_k^{-2} \boldsymbol{\Lambda}^{-1}) (\mathbf{I} - k^2 \boldsymbol{\Lambda}_k^{-2}) + k^4 \boldsymbol{\Lambda}_k^{-2} \boldsymbol{\alpha} \boldsymbol{\alpha}^\top \boldsymbol{\Lambda}_k^{-2}] \\ &= \boldsymbol{\Lambda}^{-1} - [\boldsymbol{\Lambda}^{-1} - k^2 \boldsymbol{\Lambda}^{-1} \boldsymbol{\Lambda}_k^{-2} - k^2 \boldsymbol{\Lambda}_k^{-2} \boldsymbol{\Lambda}^{-1} + k^4 \boldsymbol{\Lambda}_k^{-2} \boldsymbol{\Lambda}^{-1} \boldsymbol{\Lambda}_k^{-2} + k^4 \boldsymbol{\Lambda}_k^{-2} \boldsymbol{\alpha} \boldsymbol{\alpha}^\top \boldsymbol{\Lambda}_k^{-2}] \\ &= k^2 \boldsymbol{\Lambda}^{-1} \boldsymbol{\Lambda}_k^{-2} + k^2 \boldsymbol{\Lambda}_k^{-2} \boldsymbol{\Lambda}^{-1} - k^4 \boldsymbol{\Lambda}_k^{-2} \boldsymbol{\Lambda}^{-1} \boldsymbol{\Lambda}_k^{-2} - k^4 \boldsymbol{\Lambda}_k^{-2} \boldsymbol{\alpha} \boldsymbol{\alpha}^\top \boldsymbol{\Lambda}_k^{-2} \\ &= k^2 \boldsymbol{\Lambda}_k^{-2} [\boldsymbol{\Lambda}_k^2 \boldsymbol{\Lambda}^{-1} + \boldsymbol{\Lambda}^{-1} \boldsymbol{\Lambda}_k^2 - k^2 \boldsymbol{\Lambda}^{-1} - k^2 \boldsymbol{\alpha} \boldsymbol{\alpha}^\top] \boldsymbol{\Lambda}_k^{-2} \\ &= k^2 \boldsymbol{\Lambda}_k^{-2} [(\boldsymbol{\Lambda}^2 + 2k\boldsymbol{\Lambda} + k^2) \boldsymbol{\Lambda}^{-1} + \boldsymbol{\Lambda}^{-1} (\boldsymbol{\Lambda}^2 + 2k\boldsymbol{\Lambda} + k^2) - k^2 \boldsymbol{\Lambda}^{-1} - k^2 \boldsymbol{\alpha} \boldsymbol{\alpha}^\top] \boldsymbol{\Lambda}_k^{-2} \\ &= k^2 \boldsymbol{\Lambda}_k^{-2} [\boldsymbol{\Lambda} + 2k + k^2 \boldsymbol{\Lambda}^{-1} + \boldsymbol{\Lambda} + 2k + k^2 \boldsymbol{\Lambda}^{-1} - k^2 \boldsymbol{\Lambda}^{-1} - k^2 \boldsymbol{\alpha} \boldsymbol{\alpha}^\top] \boldsymbol{\Lambda}_k^{-2} \\ &= k^2 \boldsymbol{\Lambda}_k^{-2} [2\boldsymbol{\Lambda} + 4k\mathbf{I} + k^2 \boldsymbol{\Lambda}^{-1} - k^2 \boldsymbol{\alpha} \boldsymbol{\alpha}^\top] \boldsymbol{\Lambda}_k^{-2} \end{aligned}$$

elde edilir. $2\boldsymbol{\Lambda} + 4k\mathbf{I} + k^2 \boldsymbol{\Lambda}^{-1} = \text{diag} \left\{ \frac{2\lambda_j^2 + k\lambda_j + k^2}{\lambda_j} \right\}_{j=1}^p$ pozitif tanımlı olduğundan Δ_1 'inde pozitif tanımlı bir matris olacağı açıktır. Lemma 5.1'den faydalanarak ispat tamamlanır. ■

Teorem 5.8 *Probit regresyon modelde ridge tahmin edicisinin MMSE kriterine göre jackknife ridge tahmin edicisinden üstün olması için gerek ve yeter koşul*

$$k\boldsymbol{\alpha}^\top \boldsymbol{\Lambda}_k (2\boldsymbol{\Lambda}^2 + 3k\boldsymbol{\Lambda} + k^3 \boldsymbol{\alpha} \boldsymbol{\alpha}^\top)^{-1} \boldsymbol{\Lambda}_k \boldsymbol{\alpha} < 1, \quad \forall k > 0$$

eşitsizliğinin sağlanmasıdır.

İspat (5.13) ve (5.14) denklemleri kullanılarak PRR ve JRR tahmin edicilerinin MMSE kriterine göre $\Delta_2 = MMSE(\hat{\boldsymbol{\alpha}}_{JRR}) - MMSE(\hat{\boldsymbol{\alpha}}_{PRR})$ farkının pozitif tanımlı olması için gerek ve yeter koşul incelenecek olursa;

$$\Delta_2 = \mathbf{D} + k^4 \boldsymbol{\Lambda}_k^{-2} \boldsymbol{\alpha} \boldsymbol{\alpha}^\top \boldsymbol{\Lambda}_k^{-2} - k^2 \boldsymbol{\Lambda}_k^{-1} \boldsymbol{\alpha} \boldsymbol{\alpha}^\top \boldsymbol{\Lambda}_k^{-1} \quad (5.17)$$

burada $\mathbf{D} = Cov(\widehat{\boldsymbol{\alpha}}_{\text{JRR}}) - Cov(\widehat{\boldsymbol{\alpha}}_{\text{PRR}})$ olmak üzere,

$$\begin{aligned}
\mathbf{D} &= (\mathbf{I} - k^2 \boldsymbol{\Lambda}_k^{-2}) \boldsymbol{\Lambda}^{-1} (\mathbf{I} - k^2 \boldsymbol{\Lambda}_k^{-2}) - \boldsymbol{\Lambda}_k^{-1} \boldsymbol{\Lambda} \boldsymbol{\Lambda}_k^{-1} \\
&= (\boldsymbol{\Lambda}^{-1} - k^2 \boldsymbol{\Lambda}_k^{-2} \boldsymbol{\Lambda}^{-1}) (\mathbf{I} - k^2 \boldsymbol{\Lambda}_k^{-2}) - \boldsymbol{\Lambda}_k^{-1} \boldsymbol{\Lambda} \boldsymbol{\Lambda}_k^{-1} \\
&= \boldsymbol{\Lambda}^{-1} - k^2 \boldsymbol{\Lambda}_k^{-2} \boldsymbol{\Lambda}^{-1} - k^2 \boldsymbol{\Lambda}^{-1} \boldsymbol{\Lambda}_k^{-2} + k^4 \boldsymbol{\Lambda}_k^{-2} \boldsymbol{\Lambda}^{-1} \boldsymbol{\Lambda}_k^{-2} - \boldsymbol{\Lambda}^{-1} \boldsymbol{\Lambda} \boldsymbol{\Lambda}^{-1} \\
&= \boldsymbol{\Lambda}_k^{-2} [\boldsymbol{\Lambda}_k^2 \boldsymbol{\Lambda}^{-1} \boldsymbol{\Lambda}_k^2 - k^2 \boldsymbol{\Lambda}^{-1} \boldsymbol{\Lambda}_k^2 - k^2 \boldsymbol{\Lambda}_k^2 \boldsymbol{\Lambda}^{-1} + k^4 \boldsymbol{\Lambda}^{-1} - \boldsymbol{\Lambda}_k \boldsymbol{\Lambda} \boldsymbol{\Lambda}_k] \boldsymbol{\Lambda}_k^{-2} \\
&= \boldsymbol{\Lambda}_k^{-2} [(\boldsymbol{\Lambda}^2 + 2k\boldsymbol{\Lambda} + k^2) \boldsymbol{\Lambda}^{-1} (\boldsymbol{\Lambda}^2 + 2k\boldsymbol{\Lambda} + k^2) - k^2 \boldsymbol{\Lambda}^{-1} (\boldsymbol{\Lambda}^2 + 2k\boldsymbol{\Lambda} + k^2) \\
&\quad - k^2 (\boldsymbol{\Lambda}^2 + 2k\boldsymbol{\Lambda} + k^2) \boldsymbol{\Lambda}^{-1} + k^2 \boldsymbol{\Lambda}^{-1} - (\boldsymbol{\Lambda} + k) \boldsymbol{\Lambda} (\boldsymbol{\Lambda} + k)] \boldsymbol{\Lambda}_k^{-2} \\
&= \boldsymbol{\Lambda}_k^{-2} [(\boldsymbol{\Lambda} + 2k + k^2 \boldsymbol{\Lambda}^{-1}) (\boldsymbol{\Lambda}^2 + 2k\boldsymbol{\Lambda} + k^2) - k^2 \boldsymbol{\Lambda} - 2k^3 - k^4 \boldsymbol{\Lambda}^{-1} - k^2 \boldsymbol{\Lambda} \\
&\quad - 2k^3 - k^4 \boldsymbol{\Lambda}^{-1} + k^4 \boldsymbol{\Lambda}^{-1} - (\boldsymbol{\Lambda}^2 + k\boldsymbol{\Lambda}) (\boldsymbol{\Lambda} + k)] \boldsymbol{\Lambda}_k^{-2} \\
&= \boldsymbol{\Lambda}_k^{-2} [\boldsymbol{\Lambda}^3 + 2k\boldsymbol{\Lambda}^2 + k^2 \boldsymbol{\Lambda} + 2k\boldsymbol{\Lambda}^2 + 4k^2 \boldsymbol{\Lambda} + 2k^3 + k^2 \boldsymbol{\Lambda} + 2k^3 + k^4 \boldsymbol{\Lambda}^{-1} - k^2 \boldsymbol{\Lambda} \\
&\quad - 2k^3 - k^4 \boldsymbol{\Lambda}^{-1} - k^2 \boldsymbol{\Lambda} - 2k^3 - k^4 \boldsymbol{\Lambda}^{-1} + k^4 \boldsymbol{\Lambda}^{-1} - \boldsymbol{\Lambda}^3 - \boldsymbol{\Lambda}^2 k - \boldsymbol{\Lambda}^2 k - \boldsymbol{\Lambda} k^2] \boldsymbol{\Lambda}_k^{-2} \\
&= \boldsymbol{\Lambda}_k^{-2} [2k\boldsymbol{\Lambda}^2 + 3k^2 \boldsymbol{\Lambda}] \boldsymbol{\Lambda}_k^{-2}
\end{aligned}$$

pozitif tanımlı bir matristir. $k^4 \boldsymbol{\Lambda}_k^{-2} \boldsymbol{\alpha} \boldsymbol{\alpha}^\top \boldsymbol{\Lambda}_k^{-2}$ pozitif tanımlı olduğundan $\mathbf{D} + k^4 \boldsymbol{\Lambda}_k^{-2} \boldsymbol{\alpha} \boldsymbol{\alpha}^\top \boldsymbol{\Lambda}_k^{-2}$ matrisi de pozitif tanımlı olacaktır. Bu aşamadan sonra Lemma 5.1 uygulanacak olursa

$$\begin{aligned}
\boldsymbol{\alpha}^\top \boldsymbol{\Lambda}_k^{-1} k [\boldsymbol{\Lambda}_k^{-2} (2k\boldsymbol{\Lambda}^2 + 3k^2 \boldsymbol{\Lambda}) \boldsymbol{\Lambda}_k^{-2}]^{-1} k \boldsymbol{\Lambda}_k^{-1} \boldsymbol{\alpha} &< 1 \\
\boldsymbol{\alpha}^\top \boldsymbol{\Lambda}_k^{-1} k [\boldsymbol{\Lambda}_k^2 (2k\boldsymbol{\Lambda}^2 + 3k^2 \boldsymbol{\Lambda})^{-1} \boldsymbol{\Lambda}_k^2] k \boldsymbol{\Lambda}_k^{-1} \boldsymbol{\alpha} &< 1 \\
k^2 \boldsymbol{\alpha}^\top \boldsymbol{\Lambda}_k (2k\boldsymbol{\Lambda}^2 + 3k^2 \boldsymbol{\Lambda})^{-1} \boldsymbol{\Lambda}_k \boldsymbol{\alpha} &< 1 \\
k^3 \boldsymbol{\alpha}^\top \boldsymbol{\Lambda}_k (2\boldsymbol{\Lambda}^2 + 3k\boldsymbol{\Lambda})^{-1} \boldsymbol{\Lambda}_k \boldsymbol{\alpha} &< 1
\end{aligned}$$

gerek ve yeter şartı için $\boldsymbol{\Delta}_2$ 'nin pozitif tanımlı bir matris olacağı açıktır. ■

Teorem 5.9 *Probit regresyon modelde en çok olabilirlik tahmin edicisi ve jackknife ridge tahmin edicisinin MSE kriterine göre karşılaştırmaları aşağıdaki iki duruma göre yapılır:*

(i) $\forall j = 1, 2, \dots, p$ için λ_j , $\mathbf{X}^\top \widehat{\mathbf{W}} \mathbf{X}$ matrisinin özdeğerleri olsun. Eğer $1 - \lambda_j \alpha_j^2 > 0$ ise $k > \max(k_{2j})$ olacak şekilde sabit k değerleri için $MSE(\widehat{\boldsymbol{\alpha}}_{\text{MLE}}) - MSE(\widehat{\boldsymbol{\alpha}}_{\text{JRR}}) > 0$ sağlanır.

(ii) $\forall j = 1, 2, \dots, p$ için λ_j , $\mathbf{X}^\top \widehat{\mathbf{W}} \mathbf{X}$ matrisinin özdeğerleri olsun. Eğer $1 - \lambda_j \alpha_j^2 < 0$ ise $\min(k_{1j}) > k > 0$ olacak şekilde sabit k değerleri için $MSE(\widehat{\boldsymbol{\alpha}}_{\text{MLE}}) - MSE(\widehat{\boldsymbol{\alpha}}_{\text{JRR}}) > 0$ eşitsizliğinin sağlanır. Burada

$$k_{1j} = \frac{-2\lambda_j - \lambda_j \sqrt{2(1 + \lambda_j \alpha_j^2)}}{1 - \lambda_j \alpha_j^2} \text{ ve } k_{2j} = \frac{-2\lambda_j + \lambda_j \sqrt{2(1 + \lambda_j \alpha_j^2)}}{1 - \lambda_j \alpha_j^2}.$$

İspat Probit regresyon modelde en çok olabilirlik tahmin edicisi ile jackknife ridge tahmin edicisinin MSE fonksiyonları arasındaki fark, (4.16) ve (5.16) denklemleri kullanılarak

$$\begin{aligned}
F_1 &= \sum_{j=1}^p \frac{1}{\lambda_j} - \sum_{j=1}^p \frac{\lambda_j (\lambda_j + 2k)^2}{(\lambda_j + k)^4} - k^4 \sum_{j=1}^p \frac{\alpha_j^2}{(\lambda_j + k)^4} \\
&= \sum_{j=1}^p \frac{(\lambda_j + k)^4}{\lambda_j (\lambda_j + k)^4} - \sum_{j=1}^p \frac{\lambda_j^2 (\lambda_j + 2k)^2}{\lambda_j (\lambda_j + k)^4} - \sum_{j=1}^p \frac{k^4 \lambda_j \alpha_j^2}{\lambda_j (\lambda_j + k)^4} \\
&= \sum_{j=1}^p \frac{2k^2 \lambda_j^2 + 4k^3 \lambda_j + k^4 - k^4 \lambda_j \alpha_j^2}{\lambda_j (\lambda_j + k)^4} \\
&= \sum_{j=1}^p \frac{k^2 [k^2 (1 - \lambda_j \alpha_j^2) + 4\lambda_j k + 2\lambda_j^2]}{\lambda_j (\lambda_j + k)^4} \\
&= \sum_{j=1}^p \frac{k^2}{\lambda_j (\lambda_j + k)^4} f_j(k)
\end{aligned}$$

elde edilir, öyle ki, $f_j(k) = (1 - \lambda_j \alpha_j^2) k^2 + 4\lambda_j k + 2\lambda_j^2$ 'dir. $\forall f_j(k)$ 'nın pozitif değer aldığı yerleri göstermek için iki durum söz konusudur. İlk olarak $1 - \lambda_j \alpha_j^2 > 0$ olduğu ve $f_j(k)$ 'nin k parametresine bağlı kuadratik fonksiyon olduğu kabul edildiği takdirde, diskriminant yöntemi kullanılarak bu fonksiyonun iki kökü elde edilebilir ki bu kökler

$$k_{1j} = \frac{-2\lambda_j - \lambda_j \sqrt{2(1 + \lambda_j \alpha_j^2)}}{1 - \lambda_j \alpha_j^2} \text{ ve } k_{2j} = \frac{-2\lambda_j + \lambda_j \sqrt{2(1 + \lambda_j \alpha_j^2)}}{1 - \lambda_j \alpha_j^2}.$$

olur. Eşitliklerde $k_{1j} < 0$ ve $k_{2j} < 0$ olduğu açıktır. Bundan dolayı $f_j(k)$, 0'dan büyük k değerleri için pozitiftir. İkinci durumda $1 - \lambda_j \alpha_j^2 < 0$ olduğu ve $f_j(k)$ 'nin iki kökü göz önünde bulundurulacak olursa $k_{1j} > 0$ ve $k_{2j} < 0$ olduğu gözlemlenir; o halde $f_j(k)$, k_{1j} 'den küçük k değerleri için pozitiftir. $\min(k_{1j}) > k > 0$ olan sabit k değerleri için $F_1 > 0$ 'dir ve böylece ispat tamamlanır. ■

Teorem 5.10 *Probit regresyon modelde jackknife ridge tahmin edicisi MSE kriterine göre ridge tahmin edicisinden daha üstün olması için, $0 < \max(k_{3j}) < k$ olması gerekir, öyle ki burada*

$$k_{3j} = \frac{3 - \lambda_j \alpha_j^2 + \sqrt{(3 - \lambda_j \alpha_j^2)^2 + 16\lambda_j \alpha_j^2}}{4\alpha_j^2}$$

İspat Probit regresyon modelde ridge tahmin edicisi ile jackknife ridge tahmin edicisinin

MSE fonksiyonları arasındaki fark (5.15) ve (5.16) denklemleri kullanılarak

$$\begin{aligned}
F_2 &= \sum_{j=1}^p \frac{\lambda_j}{(\lambda_j + k)^2} + k^2 \sum_{j=1}^p \frac{\alpha_j^2}{(\lambda_j + k)^2} - \sum_{j=1}^p \frac{\lambda_j (\lambda_j + 2k)^2}{(\lambda_j + k^4)} - k^4 \sum_{j=1}^p \frac{\alpha_j^2}{(\lambda_j + k^4)} \\
&= \sum_{j=1}^p \frac{\lambda_j (\lambda_j + k)^2 + k^2 \alpha_j^2 (\lambda_j + k)^2 - \lambda_j (\lambda_j + 2k)^2 - k^4 \alpha_j^2}{(\lambda_j + k^4)} \\
&= \sum_{j=1}^p \frac{2\alpha_j^2 \lambda_j k^3 - (3\lambda_j - \alpha_j^2 \lambda_j^2) k^2 - 2\lambda_j^2 k}{(\lambda_j + k^4)} \\
&= \sum_{j=1}^p \frac{k \lambda_j}{(\lambda_j + k)^4} g_j(k)
\end{aligned}$$

elde edilir ve burada $g_j(k) = 2\alpha_j^2 k^2 + (3 - \lambda_j \alpha_j^2)k - 2\lambda_j$ 'dir. $g_j(k)$ 'nin k parametresine bağlı kuadratik fonksiyon olduğu kabul edildiği takdirde, diskriminant yöntemi kullanılarak bu fonksiyonun iki kökü elde edilir ki bu kökler

$$k_{3j} = \frac{3 - \lambda_j \alpha_j^2 + \sqrt{(3 - \lambda_j \alpha_j^2)^2 + 16\lambda_j \alpha_j^2}}{4\alpha_j^2} \text{ ve } k_{4j} = \frac{3 - \lambda_j \alpha_j^2 - \sqrt{(3 - \lambda_j \alpha_j^2)^2 + 16\lambda_j \alpha_j^2}}{4\alpha_j^2}.$$

dür. Eşitliklerde $k_{3j} > 0$ ve $k_{4j} < 0$ olduğu gözlemlenir. O halde $g_j(k)$, k_{3j} 'den büyük k değerleri için pozitifdir. Yani $k > \max(k_{3j}) > 0$ olduğu zaman F_2 pozitif değer alır ve ispat tamamlanır. ■

5.6. Yanlılık Parametresinin Seçimi

Probit regresyon modelde ridge ve jackknife ridge tahmin edicisi ile yapılacak olan kestirimlerin kararlılığı, yanlılık parametresi k için optimal değer belirlenmesine bağlıdır. En uygun k yanlılık parametresinin belirlenmesi durumunda, ridge ve jackknife ridge tahmin edicilerinin MSE değeri MLE tahmin edicisinin MSE değerinden daha küçük olacaktır. $\hat{\alpha}_{\text{GJR}}$ 'nin MSE fonksiyonu minimize edilerek en uygun k değeri elde edilir.

$$MSE(\hat{\alpha}_{\text{GJR}}) = \sum_{j=1}^p \frac{(2k_j + \lambda_j)^2 \lambda_j}{(\lambda_j + k_j)^4} + \sum_{j=1}^p \frac{k_j^4 \alpha_j^2}{(\lambda_j + k_j)^4}$$

olmak üzere denklemin k_j 'ye göre türevi alınır

$$\frac{\partial}{\partial k_j} MSE(\hat{\alpha}_{\text{GJR}}) = \sum_{j=1}^p \frac{4\lambda_j k_j (k_j^2 \alpha_j^2 - 2k_j - \lambda_j)}{(\lambda_j + k_j)^5}$$

elde edilir ve bu denklem 0'a eşitlenirse

$$k_j = \frac{1 + \sqrt{1 + \lambda_j \alpha_j^2}}{\alpha_j^2}. \quad (5.18)$$

olur. Yanlılık parametresinin değeri tahmin ediciyi etkilediğinden harmonik ortalamanın yanı sıra Kibria (2003) doğrusal regresyon modelde k yanlılık parametresini tahmin etmek için medyan, aritmetik ortalama ve geometrik ortalamanın kullanılmasını önermiştir. Aynı şekilde bu çalışmada

$$k_{ao} = \frac{\left(\sum_{j=1}^p \frac{1 + \sqrt{1 + \lambda_j \alpha_j^2}}{\alpha_j^2} \right)}{p}$$

$$k_{go} = \left(\prod_{j=1}^p \frac{1 + \sqrt{1 + \lambda_j \alpha_j^2}}{\alpha_j^2} \right)^{\frac{1}{p}}$$

ve

$$k_{med} = \text{median} \left\{ \frac{1 + \sqrt{1 + \lambda_j \alpha_j^2}}{\alpha_j^2} \right\}_{j=1}^p$$

olmak üzere k parametresini tahmin etmek için aritmetik ortalama, geometrik ortalama ve medyan fonksiyonları kullanılarak yeni tahmin ediciler önerilmiştir. Ayrıca k_j 'lerin maksimum değeri alınarak;

$$k_{max} = \text{max} \left\{ \frac{1 + \sqrt{1 + \lambda_j \alpha_j^2}}{\alpha_j^2} \right\}.$$

tahmin edicisi önerilmiştir.

5.7. Monte Carlo Simülasyonu

Bu bölümde, çoklu bağlantı probleminin varlığında probit regresyon model için en çok olabilirlik tahmin edicisine alternatif olarak sunulan tahmin edicilerin hata kareler ortalamasına göre performansları bir Monte Carlo simülasyon çalışması ile incelenir ve sonuçları karşılaştırılmıştır.

İlk olarak, McDonald ve Galarneau (1975)'in önerdiği şekilde açıklayıcı değişkenler

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

denklemini kullanılarak üretilir, burada w_{ij} bağımsız standart normal dağılıma sahip rasgele sayıları, ρ^2 herhangi iki açıklayıcı değişken arasındaki korelasyonu ve p açıklayıcı değişken

sayısını gösterir. Ayrıca $\mathbf{X}^\top \widehat{\mathbf{W}} \mathbf{X}$ korelasyon matrisi şeklinde olup açıklayıcı değişkenler standartlaştırılır.

Regresyon analizinde hata kareler ortalaması β , σ^2 , k ve p 'nin fonksiyonu olarak kabul edilecek olursa β 'yi, $\mathbf{X}^\top \mathbf{X}$ matrisinin en büyük özdeğerine karşılık gelen normalleştirilmiş özvektör olarak alınır. Bağımlı gizli değişkenin n gözlem değerleri

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

eşitliğinden elde edilir. Burada ε_i 'ler $N(0, 1)$ dağılımından üretilen rasgele sayılardır.

Bu çalışmada açıklayıcı değişkenler arasındaki çoklu bağlantı miktarının yüksek olması istendiğinden $\rho^2 = 0.9, 0.95$ ve 0.99 olarak alınır. Açıklayıcı değişkenler simülasyon çalışmasının başında üretilmiş ve süreç boyunca sabit olacak şekilde ayarlanmıştır. Açıklayıcı değişkenlerin sayısı $p = 4$ ve 8 ; ve örneklem hacmi $n = 100, 150, 200$ ve 400 olacak şekilde seçilmiştir. n, p ve ρ^2 'nin her bir değeri için simülasyon çalışması 1000 defa tekrarlanmıştır.

Tahmin edicilerin performansını karşılaştırabilmek için hata kareler ortalaması ve mutlak yanlılık kriterleri hesaplanmıştır. Herhangi bir $\widehat{\beta}^*$ tahmin edicinin simüle edilmiş hata kareler ortalaması ve karesel yanlılığı

$$MSE(\widehat{\beta}^*) = \frac{1}{1000} \sum_{r=1}^{1000} (\widehat{\beta}^* - \beta)_r^\top (\widehat{\beta}^* - \beta)_r \quad (5.19)$$

$$SB(\widehat{\beta}^*) = (\overline{\widehat{\beta}^*} - \beta)^\top (\overline{\widehat{\beta}^*} - \beta), \quad \overline{\widehat{\beta}^*} = \frac{1}{1000} \sum_{r=1}^{1000} {}_r \widehat{\beta}_j^* \quad (5.20)$$

olarak elde edilmiştir. Burada $(\widehat{\beta}^* - \beta)_r$ tahmin ve gerçek parametre vektörleri arasındaki farkı ve ${}_r \widehat{\beta}_j^*$ simülasyonun r . tekrarındaki $\widehat{\beta}_j^*$ 'in tahmin değerini gösterir. Tüm hesaplamalar R Programlama Dili (R Development Core Team, 2019) kullanılarak gerçekleştirilmiştir.

Farklı tahmin yöntemlerinin simüle edilmiş MSE değerleri Tablo 5.1 ve 5.2'de verilmiştir. Bu tablolara göre şu sonuçlar elde edilmiştir:

- Yaklaşık olarak tüm durumlarda, probit ridge ve probit jackknife ridge tahmin edicisi hata kareler ortalamasına göre MLE'den daha iyi performans gösterir. Bununla birlikte k_{max}^{RR} ve k_{max}^{JRR} tahminleri ile hesaplanan ridge ve jackknife ridge tahmin edicileri, $n = 200$ ve 400 , $\rho = 0.90$ olduğu durumda MLE'den daha yüksek MSE değerine sahip birkaç durum bulunur.

- Genel olarak aynı yöntemle elde edilen k tahminlerini kullanılarak üretilen probit jackknife ridge tahmin edicisinin MSE değerleri, probit ridge tahmin edicisinin MSE değerlerinden daha düşüktür. $\rho = 0.90$ ile 0.95 korelasyon değerlerinde k_{med}^{RR} ile k_{go}^{JRR} , $\rho = 0.99$ değerinde ise k_{max}^{JRR} en iyi performans sergileyen tahmin edicilerdir ve bu değerlerde hesaplanan tahminlerin JRR'deki MSE değeri, PRR'dekine kıyasla daha küçüktür.
- Düşük veya orta dereceli çoklu doğrusallık varlığında k 'yi elde etmek için PRR tahmin edicisinde medyan ve JRR tahmin edicisinde geometrik ortalama fonksiyonunun kullanılması en iyi sonuçları verir. Ayrıca çoklu bağlantının derecesinin yüksek olduğu durumlarda PRR tahmin edicisinde k_{med} ve JRR tahmin edicisinde k_{max} kullanılarak tahmin ediciler üretilmelidir. ($\rho=0.99$ gibi)
- Örneklem büyüklüğü n değerinin artırılması, tahmin edicilerin MSE değerlerinde bir azalma sağlar. En çok olabilirlik tahmin edicisi bu durumdan önemli ölçüde etkilenir. Örneklem boyutu çok büyüdüğünde yüksek korelasyon faktörleri dışında diğer tahmincilerle yakın bir performans sergiler.
- Çoklu bağlantının derecesinin artması durumunda bundan en olumsuz etkilenen tahmin edici, en çok olabilirlik tahmin edicisidir. Diğer tahmin ediciler için de iki grup oluşturulur. Birinci grup, çoklu bağlantının derecesi arttıkça MSE değeri azalan k_{ao}^{RR} , k_{go}^{RR} , k_{med}^{RR} , k_{max}^{RR} , k_{ao}^{JRR} ve k_{max}^{JRR} parametrelerinden oluşan ridge tipi tahmin edicilerden oluşur. İkinci grup, çoklu bağlantının derecesi arttıkça MSE değeri artan k_{go}^{JRR} ve k_{med}^{JRR} 'den üretilen ridge tahmin edicilerinden oluşur.
- Son olarak, açıklayıcı değişkenlerin sayısının artması tahmin edicilerin MSE değerinde artışa neden olur. Ayrıca bu olumsuz durumdan en çok etkilenen tahmin edici MLE'dir.

Tahmincilerin simüle edilen karesel yanlılığı Tablo 5.3 ve 5.4'te verilmiştir. Bu tablolara göre aşağıdaki sonuçlar elde edilmiştir:

- Jackknife ridge tahmin edicisinin karesel yanlılığı her durumda ridge tahmin edicisinden daha küçüktür; öyle ki her iki tahmin edici de k 'yi seçmek için aynı tahmin yöntemini kullanır.
- k 'yi tahmin etmek için k_{med} tahmin edici kullanıldığında hem probit jackknife ridge

tahmin edicisi hem de probit ridge tahmin edicisi en düşük karesel yanlılık değerine sahiptir.

- Maksimum fonksiyon, k tahmininde kullanıldığında diğer yöntemlerden daha fazla karesel yanlılık üretir.
- Çoklu bağlantının derecesinin artırılması tüm durumlarda tahmin ediciler üzerinde olumlu bir etki yapar.
- Açıklayıcı değişkenlerin sayısının artırılması, tahmin ediciler üzerinde açıkça olumsuz bir etkiye neden olur.

Tablo 5.1. $p = 4$ 'iken tahmincilerin simülasyon sonucu elde edilen MSE değerleri.

n	MLE	RR				JRR			
		k_{ao}^{RR}	k_{go}^{RR}	k_{med}^{RR}	k_{max}^{RR}	k_{ao}^{JRR}	k_{go}^{JRR}	k_{med}^{JRR}	k_{max}^{JRR}
$\rho = 0.90$									
100	0.8685	0.4839	0.2677	0.2576	0.6467	0.3766	0.2037	0.2321	0.5209
150	0.4409	0.4597	0.2535	0.2428	0.6172	0.3588	0.1858	0.2143	0.4928
200	0.2827	0.4849	0.2437	0.2425	0.6452	0.3673	0.1473	0.1707	0.5179
400	0.1239	0.4491	0.2058	0.2170	0.6199	0.3191	0.1033	0.1271	0.4828
$\rho = 0.95$									
100	1.5941	0.4108	0.2186	0.2145	0.5657	0.3476	0.2309	0.2585	0.4520
150	0.8210	0.3902	0.2042	0.2024	0.5434	0.3216	0.1937	0.2291	0.4284
200	0.5348	0.4050	0.1920	0.1894	0.5620	0.3166	0.1491	0.1827	0.4402
400	0.2483	0.3835	0.1588	0.1613	0.5487	0.2830	0.1013	0.1282	0.4216
$\rho = 0.99$									
100	8.0379	0.2660	0.2149	0.1855	0.3513	0.3430	0.5432	0.4358	0.2850
150	4.1195	0.2766	0.2030	0.1998	0.3540	0.3627	0.4774	0.4686	0.2978
200	2.6782	0.2576	0.1458	0.1365	0.3502	0.2927	0.3058	0.2890	0.2863
400	1.1959	0.2161	0.1001	0.0990	0.3177	0.2207	0.1858	0.1946	0.2487

Tablo 5.2. $p = 8$ 'iken tahmincilerin simülasyon sonucu elde edilen MSE değerleri.

n	MLE	RR				JRR			
		k_{ao}^{RR}	k_{go}^{RR}	k_{med}^{RR}	k_{max}^{RR}	k_{ao}^{JRR}	k_{go}^{JRR}	k_{med}^{JRR}	k_{max}^{JRR}
$\rho = 0.90$									
100	1072.2101	0.6988	0.3661	0.3199	0.8749	0.5643	0.2284	0.2392	0.7925
150	1.6456	0.6782	0.3378	0.2782	0.8659	0.5417	0.2055	0.2238	0.7772
200	0.8775	0.6535	0.2953	0.2386	0.8484	0.5161	0.1605	0.1700	0.7547
400	0.4137	0.6346	0.2723	0.2137	0.8383	0.4967	0.1375	0.1306	0.7421
$\rho = 0.95$									
100	24807.8273	0.6480	0.2925	0.2595	0.8417	0.5224	0.2209	0.2466	0.7505
150	3.3759	0.5990	0.2543	0.2258	0.8149	0.4682	0.2056	0.2796	0.7107
200	1.7571	0.5473	0.2069	0.1803	0.7812	0.4145	0.1584	0.2187	0.6673
400	0.8273	0.5368	0.1902	0.1564	0.7764	0.3989	0.1185	0.1538	0.6604
$\rho = 0.99$									
100	44089.8078	0.4348	0.2387	0.2687	0.6538	0.3791	0.5442	0.6372	0.5326
150	17.5000	0.3910	0.2072	0.2234	0.6138	0.3655	0.5468	0.6230	0.4937
200	8.8946	0.3651	0.1637	0.1883	0.5868	0.3412	0.4282	0.5403	0.4708
400	4.2505	0.3447	0.1157	0.1303	0.5825	0.2922	0.2727	0.3693	0.4599

Tablo 5.3. $p = 4$ 'iken tahmincilerin simülasyon sonucu elde edilen karesel yan değerleri.

n	RR				JRR			
	k_{ao}^{RR}	k_{go}^{RR}	k_{med}^{RR}	k_{max}^{RR}	k_{ao}^{JRR}	k_{go}^{JRR}	k_{med}^{JRR}	k_{max}^{JRR}
$\rho = 0.90$								
100	0.40215	0.19923	0.17019	0.59330	0.20709	0.04547	0.03509	0.40317
150	0.37000	0.18350	0.15173	0.55737	0.18400	0.04098	0.03048	0.36500
200	0.40300	0.18682	0.16805	0.59169	0.21773	0.04447	0.04007	0.40718
400	0.37281	0.16013	0.15675	0.56654	0.19341	0.03554	0.03850	0.37995
$\rho = 0.95$								
100	0.30681	0.13258	0.11428	0.49320	0.13521	0.01975	0.01582	0.30184
150	0.28586	0.12621	0.10485	0.46736	0.12478	0.02055	0.01593	0.27936
200	0.30578	0.12627	0.10329	0.48947	0.14541	0.02271	0.01783	0.30498
400	0.28836	0.10655	0.09110	0.47507	0.13586	0.01783	0.01618	0.29414
$\rho = 0.99$								
100	0.12161	0.03117	0.03269	0.25069	0.03423	0.00079	0.00066	0.10683
150	0.12074	0.03168	0.02713	0.24268	0.03799	0.00181	0.00144	0.10915
200	0.11917	0.02966	0.02504	0.23615	0.04107	0.00197	0.00152	0.11017
400	0.09820	0.02348	0.01888	0.20586	0.03137	0.00128	0.00108	0.09181

Tablo 5.4. $p = 8$ 'iken tahmincilerin simülasyon sonucu elde edilen karesel yan değerleri.

n	RR				JRR			
	k_{ao}^{RR}	k_{go}^{RR}	k_{med}^{RR}	k_{max}^{RR}	k_{ao}^{JRR}	k_{go}^{JRR}	k_{med}^{JRR}	k_{max}^{JRR}
$\rho = 0.90$								
100	0.66884	0.33034	0.26488	0.86603	0.49098	0.12941	0.08875	0.76793
150	0.64377	0.29983	0.21556	0.85678	0.45640	0.10044	0.05317	0.75051
200	0.61245	0.26078	0.18529	0.83570	0.42292	0.07782	0.04139	0.71972
400	0.58813	0.23684	0.16792	0.82316	0.39769	0.06560	0.03440	0.70215
$\rho = 0.95$								
100	0.60400	0.24423	0.19495	0.82665	0.42465	0.08277	0.05833	0.71235
150	0.54868	0.20156	0.14003	0.79744	0.35635	0.05011	0.02539	0.66620
200	0.48804	0.15939	0.10598	0.75755	0.29371	0.03180	0.01491	0.61084
400	0.47632	0.14972	0.09657	0.75127	0.28234	0.02805	0.01237	0.60225
$\rho = 0.99$								
100	0.34289	0.08386	0.07959	0.60692	0.19466	0.02387	0.02306	0.44240
150	0.27958	0.04742	0.03771	0.55551	0.13553	0.00439	0.00295	0.38069
200	0.24671	0.03405	0.02325	0.51710	0.11500	0.00242	0.00134	0.34490
400	0.23335	0.02804	0.01521	0.51164	0.10394	0.00138	0.00055	0.33680

5.8. Gerçek Veri Uygulaması

Bu bölümde, ele alınan tahmin yöntemlerinin uygulanabilirliğini göstermek için İnan ve Erdoğan (2013) tarafından yanlı tahmin yöntemleri ile birlikte lojistik regresyon modeli kullanılarak da analiz edilen gerçek bir veri seti kullanılmıştır. Bağımlı değişken, Asya finansal krizi sırasında Türkiye’de faaliyet gösteren ticari bankalardır (1999). Tasarruf Mevduatı Si-gorta Fonu (TMSF) tarafından devralınan bankalar bir diğer bankalar sıfır olarak kodlanır. Bu analizde dikkate alınan 42 banka vardır. Ayrıca aşağıda açıklanan finansal oranlar şeklinde 7 tane açıklayıcı değişken bulunur: X1:(Özkaynak+ Toplam Gelir)/(Mevduat+ Mev-duat Fonu), X2: Net İşletme Sermayesi / Toplam Varlıklar, X3: Sorunlu Krediler / Toplam Krediler, X4: Likidite Varlıkları / Toplam Varlıklar, X5: Likidite Varlıkları / (Mevduat + Mevduat Dışı Fonlar), X6: Faiz Gelirleri / Faiz Giderleri ve X7: Faiz Dışı Gelirler / Faiz Dışı Giderler.

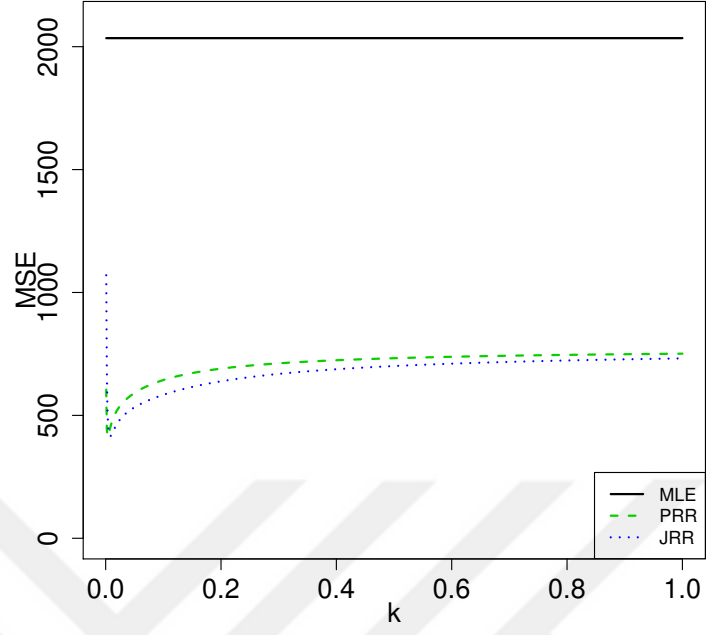
Bu verilere probit regresyon modeli uygulanarak ilgili parametreler maksimum olabirlik, probit ridge ve probit jackknife ridge tahmin yöntemleri kullanılarak tahmin edilmiştir. $\mathbf{X}^T \widehat{\mathbf{W}} \mathbf{X}$ matrisinin özdeğerleri $\lambda_1 = 0.3817$, $\lambda_2 = 0.2832$, $\lambda_3 = 0.1457$, $\lambda_4 = 0.0495$, $\lambda_5 = 0.0299$, $\lambda_6 = 0.0022$ ve $\lambda_7 = 0.0007$ olarak hesaplanır. Çoklubağlantı derecesinin bir ölçüsü olan koşul indeksi, bu veri kümesinde yüksek derecede bir çoklu doğrusallık sorunu olduğunu gösteren $k = \sqrt{\frac{\lambda_{max}}{\lambda_{min}}} = 579.5762$ olarak hesaplanmıştır. Önerilen yöntemler kullanılarak k parametresinin elde edilen tahmini değerleri $k_{ao} = 0.8298$, $k_{go} = 0.1115$, $k_{med} = 0.0503$ ve $k_{max} = 3.3649$ olarak bulunmuştur.

Bu yöntemlerin tahmini katsayıları, standart hataları, tahmini MSE değerleri ve karesel yan değerleri Tablo 5.5’de verilmiştir. Tablo 5.5’e göre maksimum olabirlik tahmincisinin katsayılarının işaretlerinin ve büyüklüklerinin, diğerlerine kıyasla çoklubağlantıdan açıkça olumsuz etkilendiği sonucuna varılabilir. Maksimum olabirlik tahminlerinin standart hataları, yanlı tahmincilerinkine göre oldukça yüksektir. Maksimum olabirlik tahmin edicisinin tahmini teorik MSE’si de diğer tahmincilerin MSE değerlerine göre çok büyük olduğu gözlenmiştir. Ayrıca, probit ridge ve probit jackknife ridge tahmin edicilerinin işaretleri ve büyüklükleri birbirine benzer yapıda olup, maksimum olabirlik tahmin edicisinden daha kararlı bir yapıya sahiptirler. Yanlı tahmin edicilerin standart hataları, maksimum olabirlik tahmin edicisinin standart hatalarından çok daha düşüktür. Probit ridge tahmin edicisinin standart hataları, probit jackknifed ridge tahmin edicisinden daha düşüktür ve bu beklenen bir durumdur, çünkü Teorem (5.6) göre probit ridge tahmin edicisinin varyansı her zaman probit jackknifed ridge tahmin edicisinin varyansından daha düşüktür. Tahmin edicilerin toplam varyansları Tablo 5.5’te verilmiştir ve önerilen her k tahmin yöntemi için probit ridge tahmin edicisinin toplam varyansının, probit jackknife ridge tahmin edicisinin toplam varyansından daha düşük olduğu gözlenir. Bununla birlikte, yanlı tahmin edicilerin karesel yan değerleri (SB) dikkate alınırca, durum farklıdır, yani probit jackknife ridge tahmin edicisinin karesel yan değerleri (SB) probit ridge tahmin edicisinden daha düşüktür. Bu durum Teorem (5.3)’te verilen sonuç ile uyuşmaktadır. Ayrıca, k parametrelerinin farklı değerlerine karşı tahmin edicilerin MSE değerlerinin ve karesel yan değerlerinin (SB) grafikleri sırasıyla Şekil (5.1) ve (5.2)’de verilmiştir. Şekil (5.1)’e göre probit jackknife ridge tahmin edicisinin MSE’si, $0 < k < 1$ parametresinin farklı değerleri için her zaman probit ridge tahmin edicisinin MSE’sinden daha azdır. Benzer şekilde, Şekil (5.2)’den probit jackknife ridge tahmin edicisinin karesel yan değerlerinin (SB) $0 < k < 1$ parametresinin farklı değerleri için her

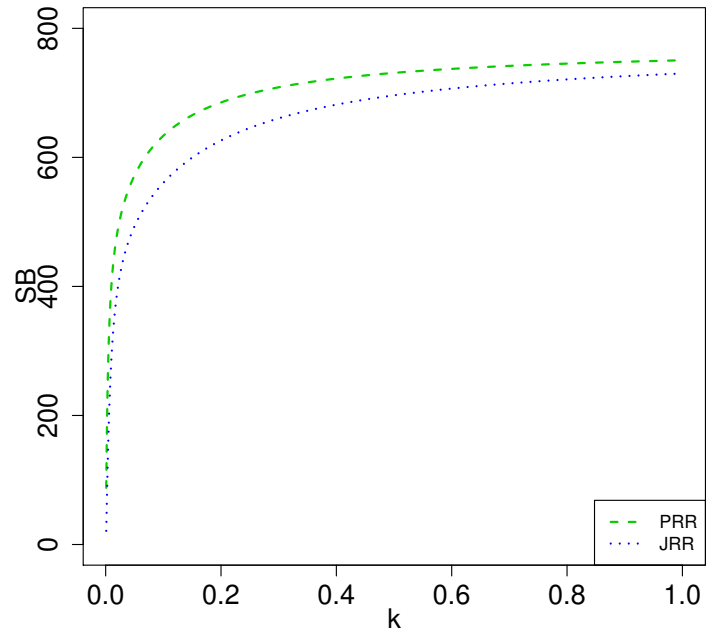
zaman probit ridge tahmin edicisinin karesel yan değerlerinden daha az olduğu görülebilir.

Tablo 5.5. Banka Verileri için tahmin edicilere ait katsayılar, standart hatalar, MSE değerleri ve karesel yan değerleri

	MLE	RR				JRR			
		AO	GO	MED	MAX	AO	GO	MED	MAX
β_1	-5.5292	0.5220	2.7563	4.3496	0.1358	1.0086	4.6057	6.3715	0.2691
β_2	23.6603	0.6263	3.3989	5.5605	0.1619	1.2146	5.7728	8.5370	0.3213
β_3	0.4133	-0.1862	-0.4437	-0.4041	-0.0572	-0.3224	-0.4718	-0.2786	-0.1099
β_4	-5.3917	0.1508	0.1360	-0.0899	0.0529	0.2370	-0.0581	-0.4486	0.0990
β_5	5.7454	0.2783	0.8970	1.1813	0.0845	0.4898	1.2644	1.5128	0.1623
β_6	8.3994	-0.1073	-0.2032	0.1204	-0.0301	-0.1951	-0.0276	0.7851	-0.0590
β_7	7.0814	0.3184	1.2082	1.6896	0.0892	0.5864	1.7778	2.3397	0.1744
$se(\beta_1)$	18.1274	0.5100	1.2527	1.4303	0.1649	0.8593	1.5359	1.5967	0.3130
$se(\beta_2)$	14.1514	0.4781	1.3483	1.5959	0.1459	0.8346	1.7292	1.8365	0.2804
$se(\beta_3)$	2.1240	0.3913	1.4840	1.9477	0.1087	0.7242	2.1273	2.4472	0.2130
$se(\beta_4)$	24.2834	0.2530	1.3817	2.2300	0.0651	0.4917	2.3387	3.3540	0.1293
$se(\beta_5)$	28.2072	0.2012	1.2230	2.1570	0.0509	0.3954	2.1872	3.5093	0.1014
$se(\beta_6)$	8.6016	0.0567	0.4146	0.8984	0.0140	0.1132	0.8212	1.7587	0.0280
$se(\beta_7)$	6.4855	0.0309	0.2288	0.5039	0.0076	0.0618	0.4562	1.0013	0.0152
MSE	2034.7560	747.0036	651.6882	591.4870	766.1907	724.8255	592.1442	533.2047	759.4252
SB		746.2531	642.4700	572.4141	766.1234	722.4514	571.1335	493.6332	759.1752
VAR	2034.7560	0.7505	9.2182	19.0729	0.0674	2.3741	21.0107	39.5715	0.2500



Şekil 5.1. Tahmincilerin MSE'lerinin k 'ya göre grafikleri (Banka verisi).



Şekil 5.2. Tahmincilerin karesel yan değerlerinin k 'ya göre grafikleri (Banka verisi).

6. SONUÇ VE ÖNERİLER

6.1. Sonuçlar

Bu çalışmada probit regresyon modelinde çoklu doğrusal bağlantı problemine çözüm olarak yeni bir jackknife ridge tahmin edicisi önerilmiştir. Jackknife prosedürünün amacı yanlılığı azaltmaktır, bu nedenle teorik ve sayısal olarak probit jackknife ridge tahmin edicisinin probit ridge tahmin edicisinden daha düşük bir yanlılığa sahip olduğu kanıtlanmıştır. Ayrıca yanlılık parametresi k 'nin bazı tahmin edicileri önerilerek yeni yöntemin performansının, hem hata kareler ortalaması hem de karesel yanlılık açısından diğerlerinden daha iyi hale geldiği gösterilmiştir.

6.2. Öneriler

Probit regresyon modelinde çoklu doğrusal bağlantı problemine çözüm olarak ikinci derece jackknife ridge tahmin edicisi tanımlanarak bu tahmin edicinin performansı teorik ve sayısal yöntemlerle var olan diğer tahmin edicilerle kıyaslanabilir.

KAYNAKLAR

- Albayrak, A. S., 2012, Çoklu doğrusal bağlantı halinde en küçük kareler tekniğinin alternatifli yanlı tahmin teknikleri ve bir uygulama, *Uluslararası Yönetim İktisat ve İşletme Dergisi*, 1(1), 105-126.
- Aldrich, J. H., Nelson, F. D., and Adler, E. S., 1984, Linear probability, logit, and probit Models (No. 45), *Beverly Hills: Sage*.
- Alp, A., 2007, İstatistiksel çalışmalarda probit analizi ve uygulama alanları, Yüksek Lisans Tezi, *Dicle Üniversitesi Fen Bilimleri Enstitüsü*, Diyarbakır.
- Arıcan, E., 2016, Lojistik regresyon modelinde çoklu iç ilişki problemi, Doktora Tezi, *Çukurova Üniversitesi Fen Bilimleri Enstitüsü*, Adana.
- Batah, F. S. M., Ramanathan, T. V. and Gore, S. D., 2008, The efficiency of modified jackknife and ridge type regression estimators: a comparison, *Surveys in Mathematics Applications*, 3, 111–122.
- Büyükuysal, Ç., 2010, Ridge regresyon analizi ve bir uygulama, Yüksek Lisans Tezi, *Uludağ Üniversitesi Sağlık Bilimleri Enstitüsü*, Bursa.
- Cafrı, R., 2009, Adana ilinde yoksulluğun analizi: Sınırlı bağımlı değişkenli modellerle bir inceleme, Yüksek Lisans Tezi, *Çukurova Üniversitesi Sosyal Bilimler Enstitüsü*, Adana.
- Demirci, M. A., 2014, Ridge regresyonda sapma parametresi k'nın elde edilmesinde genetik algoritma yaklaşımı, Yüksek Lisans 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 Series B (Methodological)*, 38(3), 248–250.
- Farrar, D. E., and Glauber, R. R., 1967, Multicollinearity in regression analysis: the prob-

- lem revisited, *The Review of Economic and Statistics*, 92–107.
- Frisch, R., 1934, Statistical confluence analysis by means of complete regression systems, *Universitetets Okonomiske Instituut*, Oslo.
- Gençer, Ö., 2016, Alkol-tütün kullanımını etkileyen değişkenlerin bivariate probit regresyon ile analizi, Yüksek Lisans Tezi, *Atatürk Üniversitesi Sosyal Bilimler Enstitüsü*, Erzurum.
- Greene, W. H., 2018, *Econometric Analysis*, 8th Edition, *Pearson*, New York.
- Gruber, M. H., 1991, The efficiency of jackknifed and usual ridge type estimators: A comparison, *Statistics and Probability letters*, 11(1), 49–51.
- Gujarati, D. N., 1988, *Basic Econometrics*, U.S.A: McGraw-Hill Book Company.
- Hardin, J. W. and Hilbe, J. M., 2007, *Generalized Linear Models and Extensions* (2th Edition), College Station: Stata press.
- Hinkley, D. V., 1977, Jackknifing in unbalanced situations, *Technometrics*, 19(3), 285–292.
- Hoerl, A. E., Kennard, R. W., 1970a, Ridge regression: Biased estimation for nonorthogonal problems, *Technometrics*, 12(1), 55–67.
- Hoerl, A. E., Kennard, R. W., 1970b, Ridge regression: applications to nonorthogonal problems, *Technometrics*, 12(1), 69–82.
- Hoerl, A. E., Kannard, R. W., Baldwin, K. F., 1975, Ridge regression: some simulations, *Communications in Statistics-Theory and Methods*, 4(2), 105–123.
- İnan, D., ve Erdoğan, B. E., 2013, Liu-type logistic estimator, *Communications in Statistics-Simulation and Computation*, 42(7), 1578–1586.
- İşyar, Y., 1994, *Ekonometrik Modeller* (2. Baskı), Uludağ Üniversitesi Güçlendirme Vakfı, Bursa.

- Khurana, M., Chaubey, Y. P., Chandra, S., 2014, Jackknifing the ridge regression estimator: A revisit, *Communications in Statistics-Theory and Methods*, 43(24), 5249–5262.
- Kırdemir, Ç., 2017, Ridge regresyon yönteminin farklı paket programlarıyla uygulanması, Yüksek Lisans Tezi, *Yıldız Teknik 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.
- Kibria, B. G., Saleh, A. M. E., 2012, Improving the estimators of the parameters of a probit regression model: A ridge regression approach, *Journal of Statistical Planning and Inference*, 142(6), 1421–1435.
- Kibria, B. G., Månsson, K., and Shukur, G., 2012, Performance of some logistic ridge regression estimators, *Computational Economics*, 4
- Kuvat, A., 2018, Ridge regresyonda sağlam parametre bulma, Yüksek Lisans Tezi, *Muğla Sıtkı Koçman Üniversitesi Fen Bilimleri Enstitüsü*, Muğla.
- Locking, H., Månsson, K., Shukur, G., 2013, Performance of some ridge parameters for probit regression: With application to Swedish job search data, *Communications in Statistics–Simulation and Computation*, 42:3, 698–710.
- Marquardt, D. W., 1970, Generalized inverses, ridge regression, biased linear estimation, and nonlinear estimation, *Technometrics*, 12(3), 591–612.
- Marquardt, D. W., Snee, R. D., 1975, Ridge regression in practice, *The American Statistician*, 29(1), 3–20.
- 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, Introduction to linear regression analysis (Vol. 821), *John Wiley and Sons*. New York.

- Nomura, M., and Ohkubo, T., 1985, A note on combining ridge and principal component regression, *Communications in Statistics–Theory and Methods*, 14(10), 2489–2493.
- Nyquist, H., 1988, Applications of the jackknife procedure in ridge regression, *Computational Statistics and Data Analysis*, 6(2), 177–183.
- Ohtani, K., 1986, On small sample properties of the almost unbiased generalized ridge estimator, *Communications in Statistics–Theory and Methods*, 15(5), 1571–1578.
- Özarıcı, Ö., 1996, Farklı not sistemlerinde öğrencinin başarılı olma olasılığının probit regresyon analiziyle değerlendirilmesi, Yüksek Lisans Tezi, *Osmangazi Üniversitesi Fen Bilimleri Enstitüsü*, Eskişehir.
- Özkale, M. R., 2007, Çoklu iç ilişki ile ilgili problemler, Doktora Tezi, *Çukurova Üniversitesi Fen Bilimleri Enstitüsü*, Adana.
- Özkale, M. R., 2008, A jackknifed ridge estimator in the linear regression model with heteroscedastic or correlated errors, *Statistics & Probability Letters*, 78(18), 3159–3169.
- Özkale, M. R. and Arıcan, E., 2019, A first-order approximated jackknifed ridge estimator in binary logistic regression, *Computational Statistics*, 34(2), 683–712.
- Pamukçu, E., 2010, Sistolik kan basıncını etkileyebilecek faktörlerin ridge regresyon analizi ile incelenmesi ve çoklu doğrusal bağıntı problemi, Yüksek Lisans Tezi, *Fırat Üniversitesi Fen Bilimleri Enstitüsü*, Elazığ.
- 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/>.
- Quenouille, M. H., 1956, Notes on bias in estimation, *Biometrika*, 43(3/4), 353–360.
- Sakallıoğlu, S. and Kaçıranlar, S., 2008, A new biased estimator based on ridge estimation, *Statistical Papers*, 49(4), 669–689.
- Schaefer, R. L., Roi, L. D., and Wolfe, R. A., 1984, A ridge logistic estimator, *Communi-*

- cations in Statistics–Theory and Methods*, 13(1), 99–113.
- Sherman, J. and Morrison, W. J., 1950, Adjustment of an inverse matrix corresponding to a change in one element of a given matrix, *The Annals of Mathematical Statistics*, 21(1), 124–127.
- Singh, B. and Chaubey, Y. P., Dwivedi, T. D., 1986, An almost unbiased ridge estimator, *Sankhya: The Indian Journal of Statistics* 48, Series B, 342–346.
- Singh, B. and Chaubey, Y. P., 1987, On some improved ridge estimators, *Statistische Hefte*, 28(1), 53–67.
- Stewart, G. W., 1973, Introduction to Matrix Computations, *Academic Press*. New York.
- Trenkler, G. and Toutenburg, H., 1990, Mean squared error matrix comparisons between biased estimators–An overview of recent results, *Statistical Papers*, 31(1), 165–179.
- Tukey, J., 1958, Bias and confidence in not quite large samples, *Ann. Math. Statist.*, 29, 614.
- Türkan, S. and Özel, G., 2018, A jackknifed estimators for the negative binomial regression model, *Communications in Statistics–Simulation and Computation*, 47:6, 1845–1865.
- Türkay, G. S., 1996, Ridge regresyon yöntemiyle TOFAŞ Firması'nın (1975–1994) yılları arası otomobil talep miktarı analizi, Yüksek Lisans Tezi, *Anadolu Üniversitesi Sosyal Bilimler Enstitüsü*, Eskişehir.
- Woodbury M. A., 1950, Inverting modified matrices, Princeton University Princeton, NJ.: Technical Report 42, Statistical Research Group.
- Yıldırım, N., 2010, En küçük kareler, ridge regresyon ve robust regresyon yöntemlerinde analiz sonuçlarına aykırı değerlerin etkilerinin belirlenmesi, Yüksek Lisans Tezi, *Çukurova Üniversitesi Fen Bilimleri Enstitüsü*, Adana.



ÖZGEÇMİŞ

KİŞİSEL BİLGİLER

Adı Soyadı : Kadriye KILINÇ
Uyruğu : T.C.
Doğum Yeri ve Tarihi : KONYA-1992
Telefon :0546 789 2399
e-mail : kilinckadriye10@gmail.com

EĞİTİM

Derece	Adı	Bitirme Yılı
Lise	Akşehir Anadolu Lisesi, Akşehir, Konya	2009
Lisans	Gazi Üniversitesi, Ankara	2014
Yüksek Lisans	Necmettin Erbakan Üniversitesi, Konya	2020

İŞ DENEYİMLERİ

Yıl	Kurum	Görevi
2014-Devam Ediyor	M.E.B	Öğretmen

YABANCI DİLLER

İngilizce

YAYINLAR

Makale

Asar, Y., Kılınç, K. (2020). A Jackknifed Ridge Estimator in Probit Regression Model, Statistics, <https://doi.org/10.1080/02331888.2020.1775597>.

Bildiri

Kılınç, K (2018). Performance Of Linear Unified Estimator In Probit Regression. International Conference on Mathematical Studies and Applications, 4-6 Ekim, Karaman.