



T.C.
NECMETTİN ERBAKAN
ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ



**BETA REGRESYON MODELİNDE TAHMİN YÖNTEMLERİNİN
KARŞILAŞTIRILMASI**

Pınar GEDİKLİ

YÜKSEK LİSANS TEZİ
Matematik Anabilim Dalı

Nisan - 2024

KONYA

Her Hakkı Saklıdır

TEZ KABUL VE ONAYI

Pınar GEDİKLİ tarafından hazırlanan "*BETA REGRESYON MODELİNDE TAHMİN YÖNTEMLERİNİN KARŞILAŞTIRILMASI*" adlı tez çalışması 19/04/2024 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

Doç. Dr. Ahmet PEKGÖR

Danışman

Prof. Dr. Yasin ASAR

Üye

Dr. Öğr. Üyesi Kadir KARAKAYA

Yukarıdaki sonucu onaylarım.

Prof. Dr. Şerife Yurdagül KUMCU

FBE 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.

Pınar GEDİKLİ

Tarih: 19/04/2024

ÖZET

YÜKSEK LİSANS TEZİ

BETA REGRESYON MODELİNDE TAHMİN YÖNTEMLERİNİN KARŞILAŞTIRILMASI

Pınar GEDİKLİ

Necmettin Erbakan Üniversitesi Fen Bilimleri Enstitüsü

Matematik Anabilim Dalı

Danışman: Prof. Dr. Yasin ASAR

2024, 79 Sayfa

Jüri

Prof. Dr. Yasin ASAR

Doç. Dr. Ahmet PEKGÖR

Dr. Öğr. Üyesi Kadir KARAKAYA

Beta regresyon modelinde çoklu bağlantı problemleri ortaya çıktığında, en çok olabilirlik tahmin edicisinin performansı olumsuz etkilenmektedir. Bu zorlukları aşmak amacıyla, çeşitli tahmin ediciler üzerinde detaylı bir inceleme gerçekleştirilmiştir. Modelin kararlılığını artırmak ve güvenilir sonuçlar elde etmek için, hemen hemen yansız Liu tahmin edicisinden esinlenerek geliştirilen hemen hemen yansız beta Liu tahmin edicisi, beta regresyon modeli için önerilmiştir. Bu tez kapsamında önerilen tahmin ediciler, hem teorik olarak karşılaştırılmış hem de Monte Carlo simülasyon çalışmaları ve gerçek veri uygulamaları ile performansları analiz edilmiştir. Elde edilen nümerik sonuçlar, teorik sonuçların doğruluğunu destekleyerek, önerilen tahmin edicilerin beta regresyon modellerinde çoklu bağlantı problemi ile başa çıkma konusundaki etkinliklerini açıkça ortaya koymaktadır.

Anahtar Kelimeler: Beta regresyon modeli, Çoklu bağlantı problemi, Hemen hemen yansız tahmin edici, Liu tahmin edici, Monte Carlo simülasyon, Ridge tahmin edici.

ABSTRACT

MS THESIS

COMPARISON OF ESTIMATION METHODS IN BETA REGRESSION MODEL

Pınar GEDİKLİ

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

Advisor: Prof. Dr. Yasin ASAR

2024, 79 Pages

Jury

Prof. Dr. Yasin ASAR

Assoc. Prof. Dr. Ahmet PEKGÖR

Assist. Prof. Dr. Kadir KARAKAYA

When multicollinearity problems occur in the beta regression model, the performance of the maximum likelihood estimator is negatively affected. In order to prevent multicollinearity problems, a detailed examination of the estimators was carried out. To increase the reliability of the model and the results, the almost unbiased beta Liu estimator, inspired by the almost unbiased Liu estimator, has been proposed for the beta regression model. The estimators used within the scope of this thesis were compared both theoretically and the estimators' performances were analyzed using the Monte Carlo simulation method. Furthermore, proposed estimators were tested on real data applications. The obtained numerical results support the accuracy of the theoretical results and clearly demonstrate the effectiveness of the proposed estimators in dealing with multicollinearity problems in beta regression models.

Keywords: Beta regression model, Multicollinearity problem, Almost unbiased estimator, Liu estimator, Monte Carlo simulation, Ridge estimator.

ÖNSÖZ

Çalışmamın başından itibaren ilgi ve yardımlarını esirgemeyen, bilimsel katkıları ile bana her zaman yardımcı olan ve yol gösteren tez danışman hocam Sayın Prof. Dr. Yasin ASAR'a içtenlikle teşekkür ederim. Kendisinin değerli yönlendirmeleri, bu tezin kalitesini artırmamda büyük katkı sağlamıştır.

Son olarak, manevi desteklerini hiçbir zaman esirgemeyen, her zaman ve her konuda yanımda olan sevgili eşim Tolga GEDİKLİ'ye ve bugünlere gelmemi sağlayan kıymetli aileme sonsuz teşekkürlerimi sunuyorum.

Pınar GEDİKLİ
KONYA-2024

İÇİNDEKİLER

ÖZET	iv
ABSTRACT	v
ÖNSÖZ	vi
ŞEKİLLER LİSTESİ	ix
ÇİZELGELER LİSTESİ	x
SİMGELER VE KISALTMALAR	xi
1. GİRİŞ	1
2. KAYNAK ARAŞTIRMASI	3
3. ÇOKLU BAĞLANTI PROBLEMİ	7
3.1. Giriş	7
3.1.1. Çoklu Doğrusal Bağlantı Probleminin Belirlenmesi	8
3.1.2. Çoklu Doğrusal Bağlantı Probleminin Giderilmesi	11
3.2. Ridge Tahmin Edicisi	12
3.2.1. Ridge Tahmin Edicisinin Teorik Özellikleri	13
3.3. Liu Tahmin Edicisi	15
3.3.1. Liu Tahmin Edicisinin Teorik Özellikleri	17
3.3.2. d Yanlılık Parametresinin Tahmini	19
3.4. Hemen Hemen Yansız Tahmin Edici	20
3.4.1. Hemen Hemen Yansız Liu Tahmin Edicisi	21
3.4.1.1. Hemen Hemen Yansız Liu Tahmin Edicisinin Teorik Özellikleri	22
3.4.1.2. d Yanlılık Parametresinin Tahmini	25
4. BETA REGRESYON MODELİ	28
4.1. Giriş	28
4.2. En Çok Olabilirlik Tahmin Edicisi	30
4.3. Beta Ridge Tahmin Edicisi	31
4.4. Beta Liu Tahmin Edicisi	34
5. HEMEN HEMEN YANSIZ BETA LIU TAHMİN EDİCİSİ	37
5.1. Hemen Hemen Yansız Beta Liu Tahmin Edicisinin Teorik Özellikleri	38

5.2. <i>d</i> Yanlılık Parametresinin Tahmini	40
5.3. Tahmin Edicilerin Teorik Olarak Karşılaştırılması.....	43
5.3.1. MLE'ye karşı AULE	43
5.3.2. LE'ye karşı AULE	45
6. MONTE CARLO SİMÜLASYON ÇALIŞMASI	47
7. GERÇEK VERİ UYGULAMASI	62
7.1. Gini Verisi.....	62
7.2. Yaşam Endeksi Verisi	65
8. SONUÇ VE ÖNERİLER	71
8.1. Sonuçlar.....	71
8.2. Öneriler	72
KAYNAKLAR	73

ŞEKİLLER LİSTESİ

<u>Şekil</u>	<u>Sayfa</u>
6.1 $\rho = 0.90$ ve $\phi = 5$ için MSE değerlerinin kutu grafiği	52
6.2 $\rho = 0.90$ ve $\phi = 10$ için MSE değerlerinin kutu grafiği	53
6.3 $\rho = 0.90$ ve $\phi = 25$ için MSE değerlerinin kutu grafiği	54
6.4 $\rho = 0.95$ ve $\phi = 5$ için MSE değerlerinin kutu grafiği	55
6.5 $\rho = 0.95$ ve $\phi = 10$ için MSE değerlerinin kutu grafiği	56
6.6 $\rho = 0.95$ ve $\phi = 25$ için MSE değerlerinin kutu grafiği	57
6.7 $\rho = 0.99$ ve $\phi = 5$ için MSE değerlerinin kutu grafiği	58
6.8 $\rho = 0.99$ ve $\phi = 10$ için MSE değerlerinin kutu grafiği	59
6.9 $\rho = 0.99$ ve $\phi = 25$ için MSE değerlerinin kutu grafiği	60
7.1 Gini verisi için, d parametresine bağlı tahmin edicilerin MSE değerlerindeki değişimin grafiği	66
7.2 Yaşam endeksi verisi için, d parametresine bağlı tahmin edicilerin MSE değerlerindeki değişimin grafiği	69

ÇİZELGELER LİSTESİ

<u>Çizelge</u>	<u>Sayfa</u>
6.1 Tahmin Edicilerin Karşılaştırılması	50
6.2 Önerdiğimiz Tahmin Edicilerin RE Tahmin Edicisinden Üstünlüğü	50
6.3 $\phi = 5$ için MSE Değerleri	51
6.4 $\phi = 10$ için MSE Değerleri	61
6.5 $\phi = 25$ için MSE Değerleri	61
7.1 MLE, RE, LE ve önerilen d1, d2, d3, d4 tahmin edicileri için: Gini verisi- nin beta regresyon analizi parametre tahminleri ve MSE değerleri	64
7.2 MLE, RE, LE ve önerilen d1, d2, d3, d4 tahmin edicileri için: Yaşam en- deksi verisinin beta regresyon analizi parametre tahminleri ve MSE de- ğerleri	70

SİMGELER VE KISALTMALAR

Simgeler

y	Bağımlı değişken
x	Bağımsız değişken (Açıklayıcı değişken)
X	Bağımsız değişken matrisi
X'	Bağımsız değişken matrisinin transpozu
I	$n \times n$ boyutlu birim matris
β	Regresyon katsayı vektörü
$\hat{\beta}$	β 'nin tahmin edicisi
p	Açıklayıcı değişken sayısı
ϕ	Dağılım parametresi
n	Örneklem büyüklüğü
μ	Bağımlı değişkenin ortalaması
ρ	Açıklayıcı değişkenler arasındaki korelasyon derecesi
λ	$X'X$ matrisinin özdeğerleri
σ	Modellerdeki hata teriminin varyansı
κ	Koşul sayısı
κ_j	Koşul indeksi
ε	Hata terimi
$\Gamma(\cdot)$	Gamma fonksiyonu
$\psi(\cdot)$	Digamma fonksiyonu
$g(\cdot)$	Monoton türevlenebilir bağlantı fonksiyonu
k	Ridge tahmin edicisinin yanlılık parametresi
d	Liu tahmin edicisinin yanlılık parametresi
Σ	Toplam sembolü
\hat{W}	Ağırlık matrisinin tahmini
$\text{tr}(\cdot)$	İz operatörünü
$d_{optimal}$	Optimal d değeri
$\text{diag}(\cdot)$	Matrisin köşegen elemanları
$E(\hat{\beta})$	$\hat{\beta}$ tahmin edicisinin beklenen değeri
$\text{bias}(\hat{\beta})$	$\hat{\beta}$ tahmin edicisinin yanlılık vektörü
$\text{Var}(\hat{\beta})$	$\hat{\beta}$ tahmin edicisinin toplam varyans matrisi
$\text{Cov}(\hat{\beta})$	$\hat{\beta}$ tahmin edicisinin kovaryans matrisi

Kısaltmalar

BRM	Beta Regresyon Modeli
MLE	En Çok Olabilirlik Tahmin Edici (Maximum Likelihood Estimator)
SE	Stein Tahmin Edici (Stein Estimator)
MSE	Skaler Hata Kareler Ortalaması
MMSE	Matris Hata Kareler Ortalaması
RE	Ridge Tahmin Edici (Ridge Estimator)
LE	Liu Tahmin Edici (Liu Estimator)
AULE	Hemen Hemen Yansız Liu Tahmin Edici (Almost Unbiased Liu Estimator)
IWLS	Tekrarlı Ağırlıklanmış En Küçük Kareler Algoritması
<i>VIF</i>	Varyans Şişirme Faktörü
EKK	En Küçük Kareler
GLE	Genelleştirilmiş Liu Tahmin Edici (Generalized Liu Estimator)

1. GİRİŞ

Beta regresyon modeli (BRM), bağımlı değişkenin sürekli değerler almasının yanı sıra 0 ile 1 arasında tanımlı oranlar, yüzdeler ve kesirler gibi değerlerle sınırlandırıldığı durumlarda kullanılan bir regresyon modelidir. BRM'nin temel amacı, bu tür sınırlamalara uyan verilerle uygun bir regresyon modeli oluşturmaktır. Ancak, bu modele ilişkin parametre tahminleri, genellikle en çok olabilirlik tahmini (MLE: maximum likelihood estimator) yöntemiyle yapılır. BRM'deki açıklayıcı değişkenler arasındaki ilişkilerin karmaşıklığı, MLE'nin doğru ve güvenilir tahminler sunma konusunda zorluklar yaşamasına neden olabilir. Bu zorlukların üstesinden gelmek adına, literatürde farklı yazarlar tarafından ridge ve Liu tahmin edicileri gibi alternatif tahmin ediciler önerilmiştir.

Bu tezin amacı, beta regresyon modeli için Liu tahmin edicisinin detaylı bir incelemesi ve geliştirilmiş bir versiyonunun sunulmasıdır. Liu tahmin edicisi, özellikle MLE'nin zorlandığı durumlarda daha etkili ve daha güvenilir tahminler sağlamak üzere önerilmiştir. Tez kapsamında, hemen hemen yansız beta Liu tahmin edicisinin bazı tahmin parametrelerini türetme süreci detaylı bir şekilde açıklanmış olup ve bu tahmin edicisinin skaler hata kareler ortalaması (MSE) değerindeki d parametresini minimize ederek en iyi performansı sağlayacak $d_{optimal}$ parametre değerini bulma süreci anlatılmıştır. Bu değer, Liu tahmin edicisinin en iyi performansı sergilemesini sağlayacak bir parametre olarak kullanılmıştır.

Daha ileri adımlarda, elde edilen $d_{optimal}$ değeri, d_1 , d_2 , d_3 ve d_4 tahmin edicilerine entegre edilerek yeni tahmin ediciler önerilmiş ve bu tahmin edicilerin performansı teorik açıdan değerlendirilmiştir. Bu değerlendirmede tahmin edicilerin MSE'si bir performans değerlendirme kriteri olarak kabul edilmiştir. Ayrıca, önerilen tahmin edicilerin mevcut ridge tahmin edicisi (RE: ridge estimator), Liu tahmin edicisi (LE: liu estimator) ve en çok olabilirlik tahmin edicisi ile karşılaştırılması, hem teorik olarak hemde simülasyon deneyleri ve gerçek veri uygulamaları ile test edilmiştir. Bu değerlendirmede, tahmin edicilerin MSE değerleri kullanılarak karşılaştırmalar yapılmıştır. Elde edilen sonuçlar, beta

regresyon modelinde önerdiğimiz tahmin edicilerin çoklu bağlantı sorunuyla karşılaştığında parametre tahminlerinde diğer rekabetçi tahmin edicilere göre daha üstün bir performans sergileyip sergilemediğini açıkça ortaya koymuştur.

Sonuçlar, önerilen tahmin edicilerin beta regresyon modelinin parametrelerini tahmin etme konusunda diğer rekabetçi tahmin edicilere göre daha üstün olduğunu göstermiştir. Bu çalışma, özellikle karmaşık ilişkiler içeren beta regresyon modeli uygulamalarında daha doğru ve güvenilir sonuçlar elde etmeye yönelik araştırmalara önemli bilgiler sunarak bu alandaki literatüre yeni bakış açıları getirmiştir. Araştırmacılara, karmaşık yapıdaki veri setlerinde daha etkili tahminler yapabilmeleri için kullanabilecekleri yeni araştırma yöntemi ve yaklaşımlar hakkında değerli bilgiler sunmuştur.

Tezin akışı aşağıdaki gibi özetlenebilir: Bölüm 3'te çoklu bağlantı problemi ele alınarak, bu problemin nasıl belirleneceğine ve nasıl çözüme kavuşturulacağına değinilmiştir. İlgili bölümde ayrıca ridge tahmin edici, Liu tahmin edici, hemen hemen yansız tahmin edici ve hemen hemen yansız Liu tahmin edicisi detaylı bir şekilde incelenmiştir. Ardından, Bölüm 4'te beta regresyon modeli, en çok olabilirlik tahmin edicisi, beta ridge tahmin edicisi ve beta Liu tahmin edicisi tanıtılmıştır. Bölüm 5'te, hemen hemen yansız Liu tahmin edicisinin beta regresyon modelinde ele alınmış hali olan hemen hemen yansız beta Liu tahmin edicisi (AULE: almost unbiased Liu estimator) önerilmiştir. Bu bölümde, AULE'nin teorik özellikleri incelenmiş ve $d_{optimal}$ değerinin türetilmesi ile bulunan d değeri, d_1 , d_2 , d_3 ve d_4 tahmin edicilerine entegre edilerek dört yeni tahmin edici önerilmiştir. Ayrıca ilgili bölümde d parametresine sahip MLE ve LE tahmin edicilerinin önerilen AULE tahmin edicisi ile teorik MSE'lerinin karşılaştırılması yapılmıştır. Bölüm 6'da, önerilen tahmin edicilerin performansını karşılaştırmak için R programlama dili kullanılarak gerçekleştirilen bir Monte Carlo simülasyon deneyi ile sonuçlar hem tablo olarak hemde tablolardan elde edilen kutu grafikleri ile raporlanmıştır. Bu deney, farklı parametre değerleri altında tahmin edicilerin nasıl davrandığını anlamak ve hangi durumlarda daha etkili olduklarını belirlemek amacıyla tasarlanmıştır. Bölüm 7'de, önerilen tahmin yöntemleri R programlama dili kullanılarak iki farklı gerçek veri setine uygulanmış ve MSE değerlerindeki değişimin grafiği çizilerek elde edilen sonuçlar değerlendirilmiştir. Son bölümde, çalışmanın elde edilen sonuçları açıklanmış ve araştırmacılara yönelik çeşitli öneriler sunulmuştur.

2. KAYNAK ARAŞTIRMASI

Doğrusal ve genelleştirilmiş doğrusal modeller için model parametrelerini tahmin etmek için birkaç temel varsayım bulunmaktadır. Bu varsayımlardan biri, açıklayıcı değişkenlerin ilişkisiz olmasıdır, ancak pratikte bu varsayım genellikle yerine getirilemez ve çoklu bağlantı olarak adlandırılan bir duruma neden olur.

Çoklu bağlantı, regresyon analizinde ilk olarak Frisch (1934) tarafından ele alınmıştır. Bu çalışmada, problemin önemi ve regresyon tahminine etkileri ilk defa gösterilmiştir. Daha sonraki dönemlerde, Farrar ve Glauber (1967), Silvey (1969) ve Mackinnon ve Puterman (1988) gibi araştırmacılar da bu konuda çalışmalar yapmışlardır.

Çoklu bağlantı açıklayıcı değişkenlerin bazıları arasında yüksek derecede korelasyon olması anlamına gelir. Bu durum, regresyon katsayılarının varyansının artmasına neden olarak çoklu bağlantı sorununu ortaya çıkarır. Bu sorunu çözmek için birkaç yöntem bulunmaktadır. Bunlar arasında ek veri toplama, ilişkili değişkeni/değişkenleri çıkartma yada ekleme, bağımsız değişkenleri birleştirerek modeli yeniden tanımlama ve yanlış tahmin yöntemlerinin kullanılması yer alır. Ancak, bu yöntemlerden en etkili olanı, modeldeki değişkenleri çıkarmadan regresyon katsayılarını yanlış olarak tahmin etmektir.

Hoerl ve Kennard (1970a) bu problemin üstesinden gelmek için ridge tahmin edicisini önerdi. Ridge tahmin edicisi modele az miktarda yan ekleyip varyansı küçültmek için daha uygun tahminlerde bulunmayı sağlayan bir yapıdadır. Ridge tahmin edicisinin en çok olabirlik tahmin edicisinden farkı, $X'X$ matrisinin köşegen elemanlarına k pozitif değeri eklemektir. Bu k parametresine ridge tahmin edicisinin yanlışlık parametresi denir ve amacı tahmin edicinin varyansını azaltarak mümkün olduğunca az yan oluşturmaktır. Birçok yazar, ridge tahmin edicisinin en küçük MSE'ye sahip olması gereken optimal k değerini elde etmek için çeşitli yöntemler önermiştir. Bu çalışmalardan bazıları şunlardır: Hoerl ve Kennard (1970a), Hoerl ve ark. (1975), McDonald ve Galarneau (1975), Hocking ve ark. (1976), Lawless ve Wang (1976), Gunst (1983), Nomura (1988), Kibria (2003), Khalaf ve Shukur (2005), Alkhamisi ve ark. (2006), Alkhamisi ve Shukur (2007),

Muniz ve ark. (2009), Khalaf (2012), Kibria ve ark. (2012), Muniz ve ark. (2012), Khalaf ve ark. (2013), Dorugade (2014), Asar ve Karaibrahimoğlu (2014), Khalaf ve Iguernane (2014), Bhat (2016), Göktaş ve Sevinç (2016), Asar ve Genç (2017), Lukman ve Olatunji (2018), Muhammad ve Kibria (2020), Ali ve ark. (2021).

Liu (1993) tarafından önerilen Liu tahmin edicisi, çoklu bağlantı durumlarında kullanılabilir alternatif bir tahmin yöntemidir. Bu yöntem, ridge tahmin edicisinin ve en çok olabilirlik tahmin edicisinin eksik yönlerini gidermeyi amaçlayarak her iki tahmin edicinin avantajlarını bir araya getirmeyi hedefler. Liu tahmin edicisinde, $0 < d < 1$ aralığında değişen bir yanlılık parametresi olan d bulunur. Özellikle çok yüksek şiddetli çoklu bağlantı durumlarında ridge tahmin edicisinin yetersiz kaldığı göz önüne alındığında, Liu (2003) daha etkili sonuçlar elde edebilmek için iki parametrelili Liu-tipi tahmin edicisini önermiştir. Bu tahmin edicinin k ve d parametreleri sırasıyla $k > 0$ ve $-\infty < d < \infty$ aralıklarında değer alabilir. Liu tahmin edicisi, ridge tahmin edicisi ile yakından ilişkili bir istatistiksel yöntemdir. Bu yöntemlerde, öngörülen parametre değerlerini dengelemek ve aşırı uyumu kontrol altında tutmak için özel bir düzenleyici terim eklenir. Hem LE hem de RE, MLE'ye göre farklı yaklaşımlar sunarlar. Özellikle regresyon analizinde, çoklu bağlantı sorunlarıyla karşılaşıldığında ve yüksek boyutlu veri setleriyle çalışıldığında tercih edilir. Böylece Liu tahmin edicisi, regresyon analizi ve tahminleme problemlerinde yüksek boyutlu ve karmaşık veri setlerinde daha güvenilir sonuçlar elde etmeyi amaçlayan bir istatistiksel yaklaşım olarak öne çıkar. Bu özellikleri, özellikle çok yüksek şiddetli çoklu bağlantı durumlarında RE'nin yetersiz kaldığı durumlarda LE'nin daha etkili olmasını sağlar. Bu nedenle, LE ve RE, regresyon analizinde ve büyük veri setlerinde başarıyla kullanılabilen önemli tahmin edicilerdir. Literatürde birçok yazar, Liu tahmin edicisinin en küçük MSE'ye sahip olması gereken optimal d değerini ($d_{optimal}$) elde etmek için çeşitli yöntemler önermiştir. Bu çalışmalardan bazıları şunlardır: Månsson ve ark. (2012), Månsson (2012), Månsson ve ark. (2015), Qasim ve ark. (2018), Qasim ve ark. (2020), Qasim et al. (2020).

Beta regresyon analizinin temel varsayımı, bağımlı değişkenin beta dağılımına sahip olmasıdır. Beta regresyon analizi hakkında yapılan çalışmalar arasında, yöntemin kullanım alanlarının belirlenmesi, yöntemin etkinliğinin test edilmesi ve iyileştirilmesi, model seçimindeki kriterlerin belirlenmesi ve tahminlerin doğruluğunun artırılması gibi bir dizi konu incelenmiştir.

Oranlar, orantılar, kesirler veya yüzdeler gibi verilerin beta regresyon analiziyle nasıl ele alınabileceği, ilk olarak Ferrari and Cribari-Neto (2004) tarafından incelenmiştir. Ferrari, 1956 yılına ait bir veri seti üzerinde bu tür verilerin modelleme yöntemlerini uygulamış ve diğer regresyon modelleri ile karşılaştırmıştır. Bu çalışma, beta regresyon analizinin bu tür veri problemlerini nasıl çözebileceğini ve diğer modellere göre sağladığı avantajları ortaya koymayı amaçlamıştır. Bu tür araştırmalar, beta regresyon analizinin pratik uygulamalardaki etkisini değerlendirme ve bu yöntemin avantajlarını vurgulama açısından önemlidir. Cribari-Neto ve Zeileis (2010) yılındaki çalışmasında, R programındaki beta regresyon modelleri sınıfını sağlayan **betareg** paketini kullanarak beta regresyonunun istatistiksel hesaplamalarını ele almıştır. Beta regresyon analizinde, açıklayıcı değişkenler ilişkili olduğunda, bilinmeyen parametrelerin en çok olabilirlik tahmin edicisi ile tahmin edilmesi durumunda regresyon parametrelerinin varyansları çok büyük olmaktadır. Bu tür durumlarda literatürde, BRM'nin çoklu bağlantı sorunlarıyla başa çıkmak için farklı yöntemler önerilmiştir. Qasim et al. (2021) beta ridge regresyon tahmin edicilerini, Karlsson ve ark. (2020) beta Liu tahmin edici olarak bilinen beta regresyon modelinin LE'sini tanıttı. Abonazel ve Taha (2021) beta regresyon modeli için bazı ridge regresyon tahmin edicilerini, Akram ve ark. (2022) beta regresyon modeli için modifiye beta ridge tipi tahmin edicisini, Algamal ve Abonazel (2022) beta regresyon modeli için Liu-tipi tahmin ediciyi, Abonazel ve ark. (2022) beta regresyon modeli için yeni iki parametrelili bir tahmin ediciyi, Abonazel ve Lukman (2022) beta regresyon modeli için Dawoud-Kibria tahmin edicisini, Rasha ve Samah (2022) beta regresyon modeli için genelleştirilmiş yanlı tahmin edicisini, Erkoc ve ark. (2023) beta regresyon modeli için beta Liu tipi tahmin ediciyi, Amin ve ark. (2023) ise beta regresyon modeli için James Stein tahmin edicisi önermişlerdir.

Ancak literatürde beta regresyon modelinde çoklu bağlantı sorunuyla karşılaşıldığında AULE'nin performansı incelenmemiştir. Bu sebeple, bu çalışmanın temel amacı yüksek korelasyonlu değişkenlerin varlığında AULE yöntemini beta regresyon modeli için uyarlayıp, bu tahmin edicinin özelliklerini incelemektir. Dolayısıyla, bu çalışma, beta regresyon modelinde çoklu bağlantı probleminin olduğu durumlarda yeni bir AULE yönteminin önerilmesi bakımından önemli bir katkı sağlamaktadır. Hata kareler ortalaması kriterine dayalı olarak beta

regresyon modeli için önerilen tahmin edicilerin mevcut diğer tahmin yöntemleriyle karşılaştırılmasını amaçlamaktadır.

Beta regresyon analizi ile ilgili uygulamalar son yıllarda giderek artan bir şekilde yaygınlaşmaktadır. Bu analiz, çeşitli uygulama alanlarında kullanılan güçlü bir istatistiksel araç olup, özellikle finansal ve ekonomik verilerin anlaşılması ve değerlendirilmesi için önemli bir rol oynamaktadır. Guolo and Varin (2014) tarafından beta regresyon modeli, sınırlı zaman serisi verilerini analiz etmek için Kanada Google®Flu Trendlerini modellemek üzere uygulanmıştır. Ünlü ve Aktaş (2017), Türkiye'deki illerin refah endeksi gösterge değerleri için beta regresyonunu kullanmıştır. Chen ve ark. (2017), konteyner taşımacılığı sektörü için hizmet nitelikleri ile müşteriye elde tutma olasılığı arasındaki ilişkiyi araştırmak için beta regresyonunu kullanmıştır. Zaman ve ark. (2019), gini katsayısını etkileyen faktörlerin belirlenmesinde beta regresyonunu kullanmıştır. Geissinger ve ark. (2022), doğa biliminde beta regresyonu kullanmıştır. Son zamanlarda, beta regresyon modeli makine öğrenimi alanında da kullanılmaktadır. Espinheira ve Silva (2019), beta regresyon modellerinde değişken seçimi için bazı kriterler önermiştir. Koc ve Akın (2022), beta regresyon modelini ve makine öğrenimini kullanarak lise giriş sınavı başarı oranlarını tahmin etmiştir. Bu örnekler, beta regresyon analizinin çeşitli alanlarda başarıyla kullanıldığını ve geniş bir uygulama yelpazesine sahip olduğunu göstermektedir.

3. ÇOKLU BAĞLANTI PROBLEMİ

Regresyon analizi, veri bilimi ve istatistiksel analizin temel taşıdır. Ancak, bu güçlü analiz yönteminde karşılaşılan çoklu bağlantı (multicollinearity) problemi, sonuçların güvenilirliğini ve yorumlanabilirliğini ciddi şekilde etkileyebilir. Bu bölümde, çoklu bağlantı probleminin ne olduğunu anlatarak, bu zorluğun nasıl tanımlanabileceğini ve nasıl aşılacağı açıklanacaktır.

3.1. Giriş

Çoklu bağlantı, bir regresyon modelinin açıklayıcı değişkenlerinin birbirleriyle güçlü bir lineer ilişkiye sahip olduğu bir durumu ifade eder. Bu durumu matematiksel olarak açıklamak için aşağıdaki bir tane bağımlı değişkeni ve p tane açıklayıcı değişkeni olan çoklu doğrusal regresyon denklemi kullanılabilir

$$y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_p X_p + \varepsilon_i \quad i = 1, 2, \dots, n.$$

Buradaki y_i , i -inci gözlemdeki bağımlı değişken değerini, β_0 , tüm açıklayıcı değişkenler sıfır olduğunda bağımlı değişkene ait beklenen değeri, β_1, \dots, β_p regresyon katsayılarını, X_1, X_2, X_3 ve böylece devam eden X_p 'ler bağımsız değişkenleri, ε_i , i -inci gözlemin hata terimini temsil etmektedir (Kalkan, 2021). Hata terimi gerçek değerler ile model tahminleri arasındaki farkları gösterir. Yukarıdaki çoklu doğrusal regresyon denklemi matrislerle aşağıdaki gibi gösterilebilir (Kalkan, 2021).

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

Bu denklemde

- y : $n \times 1$ boyutlu bağımlı değişkeni temsil eder.
- β : $p \times 1$ boyutlu bilinmeyen katsayı vektörünü ifade eder.
- ε : $n \times 1$ boyutlu hata terimi vektörüdür.

- \mathbf{X} : $n \times p$ boyutlu bağımsız değişkenleri içeren veri matrisini temsil eder.

Çoklu bağlantı problemi, bu açıklayıcı değişkenlerden en az birinin diğerlerinin lineer birleşimi olması durumunda ortaya çıkar. Yani başka bir deyişle, bir değişken diğer değişkenlerle yüksek bir korelasyona sahipse, çoklu bağlantı problemi meydana gelir. Bu tür bir durumda, normal denklemlerin çözümü ya hiç bulunamaz ya da birden fazla çözümü olabilir (Güzey, 2023). Bu durumda regresyon modelinin parametre tahminleri kararlılık ve güvenilirlik eksikliği gösterebilir. Çoklu bağlantı problemi, regresyon katsayılarının tahminlerini zorlaştırabilir ve modelin yorumlanmasını karmaşık hale getirebilir. Bu nedenle, çoklu bağlantı sorunlarını tespit etmek ve ele almak, regresyon analizi sırasında önemlidir.

Eğer bağımsız değişkenler arasında lineer bir ilişki yoksa değişkenlerin "ortogonal" olduğu ifade edilir. Ancak, bu ideal koşul genellikle nadir görülür (Montgomery ve ark., 2012). Bağımsız değişkenler arasındaki lineer ilişkinin zayıf olması durumunda modelin parametre tahminlerini büyük ölçüde etkileyebilir ve bu nedenle ihmal edilebilir (Kuvat, 2018).

3.1.1. Çoklu Doğrusal Bağlantı Probleminin Belirlenmesi

Açıklayıcı değişken sayısı az olduğunda, bu değişkenler arasındaki ilişkinin varlığını tespit etmek daha kolaydır. Ancak, açıklayıcı değişken sayısı arttığında, bu ilişkileri görmek zorlaşabilir. Çoklu bağlantının varlığını belirlemek için kullanılan birçok yöntem bulunmaktadır (Topal ve ark., 2010). Bu yöntemlerden bazıları şunlardır:

1. **Korelasyon matrisinin incelenmesi:** Korelasyon matrisi, \mathbf{X}_i ve \mathbf{X}_j bağımsız değişkenleri arasındaki ilişkinin derecesini gösteren r_{ij} köşegen elemanlarına sahip $\mathbf{X}'\mathbf{X}$ matrisidir. Eğer iki değişken arasındaki güçlü bir lineer ilişki var ise korelasyon katsayılarının mutlak değeri yani $|r_{ij}|$, 1 veya 1'e yakın olacaktır. Böyle bir durumda çoklu bağlantı sorunu ortaya çıkabilir (Pamukçu, 2010). Eğer bağımsız değişkenler arasındaki korelasyon derecesi yüksekse yani $r_{ij} > 0.80$ olduğunda, çoklu bağlantı sorununun varlığından söz edilmektedir (Neter ve ark., 1990; Gujarati, 1995). Değişken sayısı daha

fazla olduğunda çoklu bağlantı problemi tespiti için korelasyon matrisinin incelenmesi yeterli olmayabilir.

2. **Varyans Şişirme Faktörünün (Variance Inflation Factor-VIF) incelenmesi:** VIF_j , j 'inci varyans şişirme faktörüdür ve $X'X^{-1}$ matrisinin j 'inci köşegen elemanıdır (Marquardt, 1970).

j 'inci değişkene ait VIF değeri aşağıdaki gibi hesaplanır:

$$VIF_j = \frac{1}{1 - R_j^2} \quad 0 < R_j^2 < 1 \quad (3.1)$$

bu formül, bir bağımsız değişkenin diğer bağımsız değişkenlerle olan ilişkisini ölçmek ve çoklu bağlantının varlığını değerlendirmek için kullanılır. Formülde ki R_j^2 , j 'inci değişkenin diğer tüm değişkenlerle olan ilişkisinin belirli bir regresyon modelindeki R^2 değerini ifade eder. Eğer bir j 'inci bağımsız değişkeninin diğer değişkenlerle güçlü bir doğrusal ilişkisi varsa, j 'inci bağımsız değişkenine ait R_j^2 değeri 1'e yaklaşacaktır ve Denklem (3.1)'deki, VIF_j değeri artış gösterecektir.

- Eğer $VIF_j = 1$ ise çoklu bağlantı yoktur.
- Eğer $1 < VIF_j \leq 5$ ise orta şiddetli çoklu bağlantı vardır.
- Eğer $5 < VIF_j \leq 10$ ise yüksek şiddetli çoklu bağlantı vardır.
- Eğer $VIF_j > 10$ ise çok yüksek şiddetli çoklu bağlantı vardır (Karaoğlu, 2019).

3. **Korelasyon matrisinin özdeğerlerinin incelenmesi:** $X'X$ matrisinin özdeğerleri, regresyon modelindeki bağımsız değişkenler arasındaki ilişkinin derecesini anlamak için kullanılan önemli bir ölçüdür. Eğer regresyon analizindeki bağımsız değişkenler arasında güçlü bir doğrusal ilişki varsa, $X'X$ matrisinin özdeğerlerinden bir ya da daha fazlası çok küçük değer alır. Yani, bu matrisin bir veya birden fazla özdeğeri sıfır yada sıfıra yakınsa, modeldeki bağımsız değişkenlerin birbiriyle güçlü bir şekilde ilişkili olduğunu ve bu durumun çoklu bağlantıya neden olduğunu işaret eder (Vinod ve Ullah, 1981).

4. **Korelasyon matrisinin koşul sayısının veya koşul indeksinin incelenmesi:** $X'X$ matrisinin, en büyük ve en küçük özdeğerleri olan λ_{max}

ve λ_{min} kullanılarak koşul sayısı κ ,

$$\kappa = \frac{\lambda_{max}}{\lambda_{min}} \quad (3.2)$$

formülü ile hesaplanır (Vinod ve Ullah, 1981). Bu özdeğer analizi çoklu bağlantıyı belirlemek için kullanılır ve κ değerinin büyüklüğü, çoklu bağlantının şiddeti hakkında bilgi verir. Koşul sayısı κ , $\kappa_1 < 100 < \kappa_2 < 1000 < \kappa_3$ aralıklarda sınıflandırılabilir:

- Eğer $\kappa = \kappa_1$, bu durumda çoklu bağlantı yoktur.
- Eğer $\kappa = \kappa_2$, bu durumda orta şiddetli çoklu bağlantı mevcuttur.
- Eğer $\kappa = \kappa_3$, bu durumda ise çok yüksek şiddetli çoklu bağlantı mevcuttur (Stewart, 1973).

Çoklu bağlantının tespit edilmesinde kullanılan bir diğer özdeğer analizi olan koşul indeksi değeri κ_j ise,

$$\kappa_j = \frac{\lambda_{max}}{\lambda_j} \quad j = 1, 2, \dots, p \quad (3.3)$$

formülü ile hesaplanır. Her bir bağımsız değişken için hesaplanan koşul indeksi değeri, ilgili bağımsız değişkenin diğer bağımsız değişkenlerle olan ilişkisini gösterir. Burada, λ_{max} , özdeğerlerin en büyüğüdür ve λ_j ise j 'inci bağımsız değişkene ait özdeğerdir. Dolayısıyla, bir bağımsız değişkenin koşul indeksi değeri ne kadar büyükse, o değişkenin diğer değişkenlerle olan ilişkisi o kadar güçlüdür. Ayrıca Denklem (3.2) ve Denklem (3.3) karşılaştırıldığında en büyük koşul indeksinin koşul sayısına eşit olduğu görülecektir. Koşul indeksi değeri κ_j , $\kappa_{j1} < 100 < \kappa_{j2} < 1000 < \kappa_{j3}$ aralıklarda sınıflandırılabilir:

- Eğer $\kappa_j = \kappa_{j1}$, bu durumda çoklu bağlantı yoktur.
- Eğer $\kappa_j = \kappa_{j2}$, bu durumda orta şiddetli çoklu bağlantı mevcuttur.
- Eğer $\kappa_j = \kappa_{j3}$, bu durumda ise çok yüksek şiddetli çoklu bağlantı mevcuttur.

5. Korelasyon matrisinin determinantının incelenmesi: Bağımsız değişkenlere ait korelasyon matrisi $X'X$ 'in determinantı, çoklu bağlantının derecesini belirlemek için kullanılır ve bu determinant değeri 0 ile 1 arasında

değer alır. Determinant değeri küçüldükçe yani sifıra yaklaştıkça çoklu bağlantının şiddeti artar (Güzey, 2023).

6. **Regresyon modelinde F ve t -istatistiklerinin incelenmesi:** Regresyon modelinin anlamlılığına dair F istatistiği ve regresyon katsayılarına ilişkin t -istatistikleri incelenir. F testi sonucu modelin anlamlı olduğu durumlarda, katsayılara ilişkin t testleri anlamsız sonuçlar veriyorsa, bu durum çoklu bağlantının bir işareti olabilir (Hines, 1990).
7. **Regresyon katsayılarının işaretlerinin ve büyüklüklerinin incelenmesi:** Parametre tahminlerinin işaretlerinin araştırmacının beklentileriyle ters düşmesi, bağımsız değişkenlerin eklenmesi veya çıkarılması durumunda parametre tahminlerinde önemli değişikliklerin gözlenmesi, çoklu bağlantı varlığının olasılığını düşündürmektedir (Özkale, 2007). Yüksek bir R^2 değeri elde edilmesine rağmen bazı parametre tahminlerinin anlamsız hale gelmesi, çoklu bağlantı sorununun bir göstergesi olabilir.
8. **Standartlaştırılmış X matrisinin bağımsız değişkenlerinin birbirine olan ilişkisinin incelenmesi:** Başka bir yöntem olarak, standartlaştırılmış bir X matrisinin kullanılması durumunda, bağımsız değişkenlerin birbirine olan ilişkisi ölçülebilir. Bu yöntemde $-1 \leq |X'X| \leq 1$ aralığında bulunan $|X'X|$ değeri kullanılır. Bu özellik sağlandığında, matrisdeki bağımsız değişkenler ortogonal hale gelir. $|X'X|$ matrisinin değeri sifıra yaklaştıkça, çoklu bağlantının etkisi artar (Farrar ve Glauber, 1967).

3.1.2. Çoklu Doğrusal Bağlantı Probleminin Giderilmesi

Önceki bölümde açıklanan yöntemlerle çoklu doğrusal bağlantı problemi tespit edilirse, regresyon tahminleri üzerindeki olumsuz etkilerini azaltmak veya ortadan kaldırmak için çözüm yollarına ihtiyaç duyulmaktadır. Bu problemle başa çıkma önerileri, literatürde şu şekilde sunulmuştur:

1. **Örneklem büyüklüğünün artırılması:** Çoklu bağlantı sorununun üstesinden gelme yöntemlerinden biri, örneklem büyüklüğünü artırmaktır. Bu durum, örneklem büyüklüğünün artıkça parametre tahminlerinin

varyansını azaltmaya katkıda bulunabilir. Bu nedenle, varyansın örneklem büyüklüğü ile ters orantılı olduğunu söylemek mümkündür (Büyüküysal, 2010).

2. **Bazı bağımsız değişkenlerin regresyon modeline aşamalı olarak dahil edilmesi veya çıkarılması:** Çoklu doğrusal bağlantı sorunuyla başa çıkmak gerçekten hassas bir denge gerektirir. Aşamalı regresyon yöntemi kullanılırken, değişkenleri modele dahil etme veya çıkarma kararları, gerçekten dikkatlice ele alınması gereken bir konudur. Yanlış değişkenlerin çıkarılması veya gerçekten önemli olan bir değişkenin gözden kaçırılması, modelin doğruluğunu ciddi şekilde etkileyebilir (Güzey, 2023).
3. **Bağımsız değişkenlerin birleştirilmesi:** Bu yöntem birbiriyle yüksek korelasyona sahip olan bağımsız değişkenleri bir araya getirerek yeni bir değişken oluşturmayı içerir. Bu şekilde, regresyon modeline giren değişken sayısı azalır ve çoklu doğrusal bağlantı sorunuda azalabilir. Ancak, bu yöntem genellikle pek tercih edilmez, son çare olarak düşünülür ve diğer yöntemlerle başa çıkılmadığında kullanılır (Güzey, 2023). Çünkü birleştirilen değişkenin aslında modeldeki ilişkileri yeterince temsil edip etmediği belirsiz olabilir. Ayrıca, birleştirme işlemi, aslında önemli olan bilgileri kaybetme riskini taşır. Eğer bağımsız değişkenler arasında karmaşık bir ilişki varsa, bu yöntem modelin doğruluğunu düşürebilir.
4. **Yanlı tahmin yöntemlerinin kullanılması:** Yanlı tahmin yöntemleri, çoklu bağlantı sorununu aşmak ve modelin performansını artırmak amacıyla özel olarak geliştirilmiş önemli araçlardır. Bu regresyon araçları, modelin doğruluğunu artırmak için varyansı azaltmaya çalışırken, yanlılığı artırarak bir denge oluşturmayı hedefler (Güzey, 2023). Yanlılık, tahminlerin gerçek değerlerden sapma eğiliminde olması anlamına gelir, yani bu sapma tahminin gerçek değerden ne kadar uzakta olduğunu belirtir. Bu yanıltahmin yöntemleri şunlardır: Temel Bileşenler tahmin edicisi, Shrunken tahmin edicisi, özdeğerler tahmin edicisi, ridge tahmin edicisi ve Liu tahmin edicisidir. En yaygın olarak kullanılanlar ise ridge ve Liu tahmin edicileridir. Bu çalışmada ise bu yanıltahminciler arasından Liu tahmin edicisi üzerinde durulacaktır.

3.2. Ridge Tahmin Edicisi

Çoklu bağlantı sorununu aşmak için birçok yanlı tahmin yöntemi geliştirilmiştir. Bu yöntemler, modelin doğruluğunu artırmak ve aynı zamanda varyansı kontrol etmek amacıyla kullanılır. Bu yanlı tahmin yöntemlerinden biri, Hoerl ve Kennard (1970a,b) tarafından önerilen ridge tahmin edicisidir. Ridge tahmin edicisi modele az miktarda yan ekleyip varyansı küçülterek daha uygun tahminlerde bulunmayı sağlayan bir yapıdadır. Amacı $\mathbf{X}'\mathbf{X}$ matrisinin köşegen elemanlarına k pozitif değeri eklemektir. k parametresine ridge regresyonun yanlılık parametresi denilmektedir ve amacı tahmin edicinin varyansını azaltıp olabildiğince az yan oluşturmasıdır. Yanlılık parametresi k artıkça, ridge regresyonu katsayıları sıfıra yaklaşır ve bu da modelin daha güçlü hale gelmesini sağlar (Işıklar, 2020).

Ridge tahmin edicisi yani $\hat{\beta}_{RE}$ şu şekilde tanımlanmaktadır:

$$\hat{\beta}_{RE} = (\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1}\mathbf{X}'\mathbf{X}\hat{\beta}_{OLS}. \quad (3.4)$$

Burada \mathbf{X} , bağımsız değişken matrisini; $\hat{\beta}_{OLS}$, en küçük kareler tahmin edicisi ile elde edilen parametre vektörünü; \mathbf{I} , birim matrisi ve k , yanlılık parametresini temsil etmektedir.

3.2.1. Ridge Tahmin Edicisinin Teorik Özellikleri

Bu bölümde ridge tahmin edicisinin teorik özelliklerinden olan beklenen değeri, yanlılık vektörü, varyans-kovaryans matrisi, matris hata kareler ortalaması (MMSE) değeri ve MSE değeri incelenecektir.

Ridge tahmin edicisinin beklenen değeri

$$\begin{aligned} E(\hat{\beta}_{RE}) &= E\left((\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1}\mathbf{X}'\mathbf{X}\hat{\beta}_{OLS}\right) \\ &= E\left((\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1}\mathbf{X}'\mathbf{X}\right) E\left(\hat{\beta}_{OLS}\right) \\ &= (\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1}\mathbf{X}'\mathbf{X}\beta. \end{aligned} \quad (3.5)$$

Ridge tahmin edicisinin yanlılık vektörü

$$\begin{aligned}
Bias(\hat{\beta}_{RE}) &= E(\hat{\beta}_{RE}) - \beta \\
&= (\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1}\mathbf{X}'\mathbf{X}\beta - \beta \\
&= [(\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1}\mathbf{X}'\mathbf{X} - \mathbf{I}] \beta \\
&= [(\mathbf{I} + k(\mathbf{X}'\mathbf{X})^{-1})^{-1} - \mathbf{I}] \beta \\
&= [\mathbf{I} - k(\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1} - \mathbf{I}] \beta \\
&= -k(\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1}\beta.
\end{aligned} \tag{3.6}$$

Ridge tahmin edicisinin kovaryans matrisi

$$\begin{aligned}
Cov(\hat{\beta}_{RE}) &= Cov\left((\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1}\mathbf{X}'\mathbf{X}\hat{\beta}_{OLS}\right) \\
&= ((\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1}\mathbf{X}'\mathbf{X}) Cov(\hat{\beta}_{OLS}) \\
&\quad \times ((\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1}\mathbf{X}'\mathbf{X})' \\
&= \sigma^2(\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1}
\end{aligned} \tag{3.7}$$

şeklindedir. Bu kovaryans matrisinin köşegen elemanlarının toplamı, Liu tahmin edicisinin toplam varyans matrisini verir.

Ridge tahmin edicisinin toplam varyans matrisi:

$$\begin{aligned}
Var(\hat{\beta}_{RE}) &= tr\left(Cov(\hat{\beta}_{RE})\right) \\
&= tr\left(\sigma^2(\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1}\right) \\
&= \sigma^2 \sum_{j=1}^p \frac{\lambda_j}{(\lambda_j + k)^2}
\end{aligned} \tag{3.8}$$

şeklindedir. Ridge tahmin edicisinin matris hata kareler ortalaması:

$$\begin{aligned}
MMSE(\hat{\beta}_{RE}) &= Cov(\hat{\beta}_{RE}) + Bias(\hat{\beta}_{RE})Bias(\hat{\beta}_{RE})' \\
&= \sigma^2(\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1} \\
&\quad + (-k(\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1}\beta)(-k(\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1}\beta)' \\
&= \sigma^2(\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1} \\
&\quad + k^2(\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1}\beta\beta'(\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1}
\end{aligned} \tag{3.9}$$

şeklindedir. Ridge tahmin edicisinin skaler hata kareler ortalaması ise:

$$\begin{aligned}
 MSE(\widehat{\beta}_{RE}) &= tr(MMSE(\widehat{\beta}_{RE})) \\
 &= tr(\sigma^2(\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1} \\
 &\quad + k^2(\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1}\beta\beta'(\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1}) \\
 &= \sigma^2 \sum_{j=1}^p \frac{\lambda_j}{(\lambda_j + k)^2} + k^2 \sum_{j=1}^p \frac{\beta_j^2}{(\lambda_j + k)^2} \quad (3.10)
 \end{aligned}$$

$$= \gamma_1(k) + \gamma_2(k) \quad (3.11)$$

şeklindedir (Hoerl ve Kennard, 1970a,b). Bu denklemde, ilk terim olan $\gamma_1(k)$ toplam varyansı, ikinci terim olan $\gamma_2(k)$ ise toplam karesel yanlılığı ifade eder. Ayrıca σ^2 hata varyansını, λ_j 'ler ridge regresyonunda kullanılan $\mathbf{X}'\mathbf{X}$ matrisinin özdeğerlerini, β_j tahmin edicinin katsayılarını ve k yanlılık parametresini temsil etmektedir.

3.3. Liu Tahmin Edicisi

Çoklu bağlantı problemi durumlarında yanlı tahmin edici seçimi, istatistiksel analizde büyük bir öneme sahiptir. Bu yanlı tahmin edicilerden en yaygın olarak kullanılan tahmin edici ridge tahmin edicisi $\widehat{\beta}_{RE}$ 'dir. Ancak, $\widehat{\beta}_{RE}$ 'nin seçimi oldukça karmaşık olduğundan, bu yanlı tahmin ediciye alternatif olarak çeşitli tahmin ediciler önerilmiştir. Bunlardan biri Stein (1956) tarafından önerilen Stein tahmin edicisi (SE: Stein estimator) ($\widehat{\beta}_s = c\widehat{\beta}$, $0 < c < 1$)'dir. Bu tahmin yönteminin avantajı, c parametresinin SE'nin lineer bir fonksiyonu olmasıdır. Ancak, tüm parametrelerin aynı oranda küçülmesi nedeniyle, bu tahmin edici çoklu bağlantı sorunlarına karşı yeterince etkili olamamıştır.

Liu (1993) yaptığı çalışmada, Hoerl ve Kennard (1970a,b) tarafından önerilen ridge tahmin edicisi $\widehat{\beta}_{RE}$ ve Stein (1956) tarafından önerilen Stein tahmin edicisi $\widehat{\beta}_s$ 'nin farklı bir birleşimini inceleyerek, bu tahmincilerin güçlü yönlerini bir araya getirerek yeni bir yanlı tahmin yöntemi geliştirip daha iyi sonuçlar elde etmeyi amaçlamıştır.

Liu (1993)'nun önerdiği bu yeni tahmin yöntemi, $k = 1$ durumuna ait Ridge tahmin edicisinin $\mathbf{X}'\mathbf{y}$ vektörüne Stein tahmin edicisini ekleyerek alternatif bir tahmin edici önermektedir. Yani, çoklu doğrusal regresyon modeli esas alınarak

$\beta + \varepsilon = d\hat{\beta}$ Stein dönüşümü uygulanmıştır. Özellikle düşük yanlılık parametresi (d) değerleri için, En Küçük Kareler (EKK) tahmin edicisine kıyasla daha düşük MSE değeri elde edilmiştir. Bu durum, Liu (1993)'ün önerdiği tahmin yönteminin daha üstün tahmin performansı sunduğunu göstermektedir. Bu tahmin yöntemi daha sonra Akdeniz ve Kaçiranlar (1995) tarafından Liu tahmin edicisi (LE: Liu Estimator) olarak isimlendirilmiş ve literatürde "Linear Unified" kısaltmasıyla anılmıştır.

Çoklu doğrusal regresyon modeli temel alınarak, β parametresinin Liu tahmin edicisi $\hat{\beta}_{LE}$ aşağıdaki gibi tanımlanmıştır.

Çoklu doğrusal regresyon modeli

$$\mathbf{y} = \mathbf{X}\beta + \varepsilon \quad (3.12)$$

ile ifade edilir. Burada \mathbf{y} , $(n \times 1)$ boyutlu bağımlı değişken vektörünü, \mathbf{X} , $n \times p$ boyutlu bağımsız değişkenleri içeren veri matrisini, β , $p \times 1$ boyutlu regresyon katsayılarının vektörünü ve ε , $(n \times 1)$ boyutlu hata vektörünü temsil eder. Hata terimleri için $E(\varepsilon) = 0$ ve $Cov(\varepsilon) = \sigma^2\mathbf{I}$ varsayımı altında, Denklem (3.12)'deki β regresyon katsayıları parametresini tahmin etmek için en yaygın olarak kullanılan yöntem en küçük kareler tahmin edicisi yöntemidir. EKK yani $\hat{\beta}_{OLS}$ aşağıdaki gibi tanımlanmaktadır

$$\hat{\beta}_{OLS} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}. \quad (3.13)$$

Hem EKK tahmin edicisi hem de kovaryans matrisi, büyük ölçüde $\mathbf{X}'\mathbf{X}$ matrisinin özellikleriyle bağlantılıdır. Eğer $\mathbf{X}'\mathbf{X}$ kötü koşullanmışsa, yani \mathbf{X} 'in sütun vektörleri doğrusal olarak bağımlıysa, EKK tahmin edicisi çeşitli hatalara duyarlı hale gelir. Bu durumda, bazı regresyon katsayıları istatistiksel olarak anlamsız olabilir veya yanlış işarete sahip olabilir. Bunlar, tahmin parametreleri için geniş güven aralıkları ile sonuçlanabilir. Kötü koşullandırılmış bir $\mathbf{X}'\mathbf{X}$ matrisi ile, regresyon parametreleri hakkında güvenilir istatistiksel çıkarımlar yapmak zorlaşır (Alheety ve Kibria, 2009).

β parametre vektörünün Liu tahmin edicisi $\hat{\beta}_{LE}$

$$\hat{\beta}_{LE} = (\mathbf{X}'\mathbf{X} + \mathbf{I})^{-1}(\mathbf{X}'\mathbf{y} + d\hat{\beta}_{OLS}) \quad (3.14)$$

olarak tanımlanmıştır (Liu, 1993). Bu denklemde, d yanlılık parametresini ifade eder ve bu parametre 0 ile 1 arasında sabit bir değere sahiptir. Liu tahmin edicisinde d parametresinin seçimi, ridge tahmin edicisinde kullanılan k

parametresinin seçimine göre az karmaşıktır. Bu durum, Liu tahmin edicisinde d parametresinin doğrusal bir fonksiyon olması sayesinde ortaya çıkar. Bu özellik, Liu tahmin edicisini diğer yöntemlerden ayıran en temel avantajlardan biridir. Ayrıca, Liu tahmin edicisi, EKK tahmin edicisinin doğrusal bir fonksiyonu olarak ifade edilebilecek bir yapıya sahiptir. Yukarıdaki Denklem (3.4) ele alınırsa,

$$\hat{\beta}_{LE} = (\mathbf{X}'\mathbf{X} + \mathbf{I})^{-1}(\mathbf{X}'\mathbf{y} + d\hat{\beta}_{OLS})$$

ve Denklem (3.13): $\hat{\beta}_{OLS} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ bu denklemde yerine koyulursa,

$$\hat{\beta}_{LE} = (\mathbf{X}'\mathbf{X} + \mathbf{I})^{-1}(\mathbf{X}'\mathbf{y} + d(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y})$$

elde edilir. $\mathbf{X}'\mathbf{y}$ ortak paranteze alınırsa,

$$\hat{\beta}_{LE} = (\mathbf{X}'\mathbf{X} + \mathbf{I})^{-1} (\mathbf{I} + d(\mathbf{X}'\mathbf{X})^{-1}) \mathbf{X}'\mathbf{y}$$

$\mathbf{I} = (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{X})$ olarak denklemde yerine yazılırsa,

$$\hat{\beta}_{LE} = (\mathbf{X}'\mathbf{X} + \mathbf{I})^{-1} ((\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{X}) + d(\mathbf{X}'\mathbf{X})^{-1}) \mathbf{X}'\mathbf{y}$$

$(\mathbf{X}'\mathbf{X})^{-1}$ ortak parantezine alırsak,

$$\hat{\beta}_{LE} = (\mathbf{X}'\mathbf{X} + \mathbf{I})^{-1}(\mathbf{X}'\mathbf{X} + d\mathbf{I})\mathbf{X}'\mathbf{y}(\mathbf{X}'\mathbf{X})^{-1}$$

$\mathbf{X}'\mathbf{y}(\mathbf{X}'\mathbf{X})^{-1}$ ifadesi Denklem (3.13)'teki $\hat{\beta}_{OLS}$ eşittir. Eşitliği tekrar düzenlersek Liu tahmin edicisi

$$\hat{\beta}_{LE} = (\mathbf{X}'\mathbf{X} + \mathbf{I})^{-1}(\mathbf{X}'\mathbf{X} + d\mathbf{I})\hat{\beta}_{OLS} \quad (0 < d < 1) \quad (3.15)$$

ifadesine eşittir (Liu, 1993). Burada \mathbf{X} , bağımsız değişken matrisini; $\hat{\beta}_{OLS}$, en küçük kareler tahmin edicisi ile elde edilen parametre vektörünü; \mathbf{I} , birim matrisi; d , $[0, 1]$ aralığında tanımlanan yanlılık parametresini temsil etmektedir.

3.3.1. Liu Tahmin Edicisinin Teorik Özellikleri

Bu bölümde Liu tahmin edicisinin teorik özelliklerinden olan beklenen değeri, yanlılık vektörü, varyans-kovaryans matrisi, matris hata kareler ortalaması değeri ve skaler hata kareler ortalaması değeri incelenecektir.

Liu tahmin edicisinin beklenen değeri:

$$\begin{aligned}
E(\widehat{\beta}_{LE}) &= E\left((\mathbf{X}'\mathbf{X} + \mathbf{I})^{-1}(\mathbf{X}'\mathbf{X} + d\mathbf{I})\widehat{\beta}_{OLS}\right) \\
&= E\left((\mathbf{X}'\mathbf{X} + \mathbf{I})^{-1}(\mathbf{X}'\mathbf{X} + d\mathbf{I})\right) E(\widehat{\beta}_{OLS}) \\
&= (\mathbf{X}'\mathbf{X} + \mathbf{I})^{-1}(\mathbf{X}'\mathbf{X} + d\mathbf{I})\beta
\end{aligned} \tag{3.16}$$

Liu tahmin edicisinin yanlışlık vektörü:

$$\begin{aligned}
Bias(\widehat{\beta}_{LE}) &= E(\widehat{\beta}_{LE}) - \beta \\
&= (\mathbf{X}'\mathbf{X} + \mathbf{I})^{-1}(\mathbf{X}'\mathbf{X} + d\mathbf{I})\beta - \beta \\
&= (\mathbf{X}'\mathbf{X} + \mathbf{I})^{-1}(\mathbf{X}'\mathbf{X} + d\mathbf{I} - \mathbf{I})\beta \\
&= \left((\mathbf{I} + (\mathbf{X}'\mathbf{X})^{-1})^{-1} + d(\mathbf{X}'\mathbf{X} + \mathbf{I})^{-1} - \mathbf{I}\right)\beta \\
&= (\mathbf{I} - (\mathbf{X}'\mathbf{X} + \mathbf{I})^{-1} + d(\mathbf{X}'\mathbf{X} + \mathbf{I})^{-1} - \mathbf{I})\beta \\
&= -(1 - d)(\mathbf{X}'\mathbf{X} + \mathbf{I})^{-1}\beta
\end{aligned} \tag{3.17}$$

Liu tahmin edicisinin kovaryans matrisi ise,

$$\begin{aligned}
Cov(\widehat{\beta}_{LE}) &= Cov\left((\mathbf{X}'\mathbf{X} + \mathbf{I})^{-1}(\mathbf{X}'\mathbf{X} + d\mathbf{I})\widehat{\beta}_{OLS}\right) \\
&= (\mathbf{X}'\mathbf{X} + \mathbf{I})^{-1}(\mathbf{X}'\mathbf{X} + d\mathbf{I})Cov(\widehat{\beta}_{OLS})\left((\mathbf{X}'\mathbf{X} + \mathbf{I})^{-1}(\mathbf{X}'\mathbf{X} + d\mathbf{I})\right)' \\
&= \sigma^2(\mathbf{X}'\mathbf{X} + \mathbf{I})^{-1}(\mathbf{X}'\mathbf{X} + d\mathbf{I})(\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{X} + d\mathbf{I})(\mathbf{X}'\mathbf{X} + \mathbf{I})^{-1} \\
&= \sigma^2(\mathbf{X}'\mathbf{X} + \mathbf{I})^{-2}(\mathbf{X}'\mathbf{X} + d\mathbf{I})^2(\mathbf{X}'\mathbf{X})^{-1}
\end{aligned} \tag{3.18}$$

şeklindedir. Bu kovaryans matrisinin köşegen elemanlarının toplamı, Liu tahmin edicisinin toplam varyans matrisini verir.

Liu tahmin edicisinin toplam varyans matrisi:

$$\begin{aligned}
Var(\widehat{\beta}_{LE}) &= tr\left(Cov(\widehat{\beta}_{LE})\right) \\
&= tr\left(\sigma^2(\mathbf{X}'\mathbf{X} + \mathbf{I})^{-2}(\mathbf{X}'\mathbf{X} + d\mathbf{I})^2(\mathbf{X}'\mathbf{X})^{-1}\right) \\
&= \sigma^2 \sum_{j=1}^p \frac{(\lambda_j + d)^2}{\lambda_j(\lambda_j + 1)^2}
\end{aligned} \tag{3.19}$$

şeklindedir. Burada $\lambda_1, \lambda_2, \dots, \lambda_p > 0$ matris $\mathbf{X}'\mathbf{X}$ 'in özdeğerlerini, $tr(\cdot)$ ise iz operatörünü temsil eder.

Liu tahmin edicisinin matris hata kareler ortalaması değeri:

$$\begin{aligned}
MMSE(\widehat{\boldsymbol{\beta}}_{LE}) &= Cov(\widehat{\boldsymbol{\beta}}_{LE}) + Bias(\widehat{\boldsymbol{\beta}}_{LE})Bias(\widehat{\boldsymbol{\beta}}_{LE})' \\
&= \sigma^2(\mathbf{X}'\mathbf{X} + \mathbf{I})^{-2}(\mathbf{X}'\mathbf{X} + d\mathbf{I})^2(\mathbf{X}'\mathbf{X})^{-1} \\
&\quad + (-(1-d)(\mathbf{X}'\mathbf{X} + \mathbf{I})^{-1}\boldsymbol{\beta})(-(1-d)(\mathbf{X}'\mathbf{X} + \mathbf{I})^{-1}\boldsymbol{\beta})' \\
&= \sigma^2(\mathbf{X}'\mathbf{X} + \mathbf{I})^{-2}(\mathbf{X}'\mathbf{X} + d\mathbf{I})^2(\mathbf{X}'\mathbf{X})^{-1} \\
&\quad + (1-d)^2(\mathbf{X}'\mathbf{X} + \mathbf{I})^{-1}\boldsymbol{\beta}(\mathbf{X}'\mathbf{X} + \mathbf{I})^{-1}\boldsymbol{\beta}' \tag{3.20}
\end{aligned}$$

Liu tahmin edicisinin skaler hata kareler ortalaması değeri ise,

$$\begin{aligned}
MSE(\widehat{\boldsymbol{\beta}}_{LE}) &= tr(MMSE(\widehat{\boldsymbol{\beta}}_{LE})) \\
&= tr(\sigma^2(\mathbf{X}'\mathbf{X} + \mathbf{I})^{-2}(\mathbf{X}'\mathbf{X} + d\mathbf{I})^2(\mathbf{X}'\mathbf{X})^{-1} \\
&\quad + (1-d)^2(\mathbf{X}'\mathbf{X} + \mathbf{I})^{-1}\boldsymbol{\beta}(\mathbf{X}'\mathbf{X} + \mathbf{I})^{-1}\boldsymbol{\beta}') \\
&= \sigma^2 \sum_{j=1}^p \frac{(\lambda_j + d)^2}{\lambda_j(\lambda_j + 1)^2} + (1-d)^2 \sum_{j=1}^p \frac{\beta_j^2}{(\lambda_j + 1)^2} \tag{3.21}
\end{aligned}$$

şeklindedir (Liu, 1993). Bu denklemde σ^2 hata varyansını, λ_j 'ler $\mathbf{X}'\mathbf{X}$ matrisinin özdeğerlerini ve β_j 'ler tahmin edicinin katsayılarını ve d yanlılık parametresini temsil eder. $0 < d < 1$ aralığında olan d yanlılık parametre değerleri için, her σ^2 ve \mathbf{X} değeri kombinasyonları için Liu tahmin edicisi, EKK tahmin edicisine göre daha düşük MSE değeri elde edilir. $d = 1$ olduğunda, Liu tahmin edicisinin MSE değeri EKK tahmin edicisinin MSE değerine ile aynı sonucu verir.

3.3.2. d Yanlılık Parametresinin Tahmini

$MSE(\widehat{\boldsymbol{\beta}}_{LE})$ değeri d parametresine bağlı bir fonksiyondur. $MSE(\widehat{\boldsymbol{\beta}}_{LE}) = f(d)$ olsun. d parametresine göre birinci türevi alındığında ve $f'(d) = 0$ eşitliği çözüldüğünde, optimal d değeri şu şekilde bulunur:

$$f'(d) = \frac{\partial(MSE(\widehat{\boldsymbol{\beta}}_{LE}))}{\partial d} = 0 \tag{3.22}$$

Denklem (3.21)'deki LE'nin MSE'si şu şekilde düzenlenebilir.

$$MSE(\widehat{\boldsymbol{\beta}}_{LE}) = d^2 \sum_{j=1}^p \frac{\sigma^2 + \lambda_j \beta_j^2}{\lambda_j(\lambda_j + 1)^2} + 2d \sum_{j=1}^p \frac{\sigma^2 \lambda_j - \lambda_j \beta_j^2}{\lambda_j(\lambda_j + 1)^2} + \sum_{j=1}^p \frac{\sigma^2 \lambda_j^2 + \lambda_j \beta_j^2}{\lambda_j(\lambda_j + 1)^2}$$

Bu fonksiyonun d parametresine göre birinci türevi alınırsa,

$$\frac{\partial(MSE(\widehat{\beta}_{LE}))}{\partial d} = 2d \sum_{j=1}^p \frac{\sigma^2 + \lambda_j \beta_j^2}{\lambda_j (\lambda_j + 1)^2} + 2 \sum_{j=1}^p \frac{\sigma^2 \lambda_j - \lambda_j \beta_j^2}{\lambda_j (\lambda_j + 1)^2} = 0$$

elde edilir. Buna göre LE'nin minimum MSE'sini verecek optimal d değeri,

$$d = \frac{\sum_{j=1}^p \frac{\beta_j^2 - \sigma^2}{(\lambda_j + 1)^2}}{\sum_{j=1}^p \frac{\sigma^2 + \lambda_j \beta_j^2}{\lambda_j (\lambda_j + 1)^2}} \quad (3.23)$$

olarak bulunur.

Liu tahmin edicisinin yanlılık parametresi olan d matris şeklinde ifade edilebilir (Akdeniz ve Kaçiranlar, 1995). Bu özelliği sayesinde parametre tahminlerindeki yanlılığı düzeltmeye odaklanarak geleneksel Liu tahmin edicisini genişletir. Liu tahmin edicisinin geliştirilmiş halini Akdeniz ve Kaçiranlar (1995) ve Akdeniz ve Duran (2012) çalışmalarında kullanmıştır.

Genelleştirilmiş Liu tahmin edicisi (GLE: Generalized Liu Estimator), d matrisini $0 < d_j < 1$ ve $j = 1, 2, \dots, p$ koşullarını sağlayacak şekilde genişleterek aşağıdaki formül ile ifade edilir:

$$\widehat{\beta}_{GLE} = (\mathbf{X}'\mathbf{X} + \mathbf{I})^{-1} (\mathbf{X}'\mathbf{X} + \begin{bmatrix} d_1 & 0 & \dots & 0 \\ 0 & d_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & d_p \end{bmatrix}) \widehat{\beta}_{OLS}$$

ve matrisi

$$D = \begin{bmatrix} d_1 & 0 & \dots & 0 \\ 0 & d_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & d_p \end{bmatrix}$$

olarak tekrar düzenlersek,

$$\widehat{\beta}_{GLE} = (\mathbf{X}'\mathbf{X} + \mathbf{I})^{-1} (\mathbf{X}'\mathbf{X} + D) \widehat{\beta}_{OLS}$$

şeklinde ifade edilebilir.

Liu (1993) tarafından önerilen optimal d parametresi, LE durumunda $\widehat{\beta}(d = 1) = \widehat{\beta}$ özelliğini sergiler. Ancak, daha sonraki birçok çalışmada, çeşitli d parametresi tahminleri sunulmuştur. Yapılan bu araştırmalarda görüleceği üzere, d parametresi β_j^2 değerlerine bağlıdır. Alt bölümde hemen hemen yansız tahmin edicinin tanımı ve Liu tahmin edicisinin hemen hemen yansız hali tanıtılacaktır.

3.4. Hemen Hemen Yansız Tahmin Edici

Beta regresyon modelinde çoklu bağlantı problemi ortaya çıktığında, en çok olabilirlik tahmin edicisinin istenen performansı sağlamadığı bu duruma çözüm olarak, önceki bölümlerde, yanlı tahmincilerin kullanılması önerilmişti. Farklı regresyon türlerinde kullanılan hemen hemen yansız tahmin ediciler, özellikle çoklu bağlantı probleminin etkili olduğu durumlar için geliştirilmiştir. Bu tahmin ediciler, modelin performansını artırmak amacıyla modele az miktarda yan ekleyerek varyansı azaltmayı ve MSE'yi minimize etmeyi hedefler. Hemen hemen yansız tahmin edicilerin, yanlı tahmincilere kıyasla avantajı, eklenen yanın daha küçük olması ve bu sayede daha etkili tahmin sonuçları elde edebilmesidir. Hemen hemen yansız tahmin edicinin tanımı, aşağıda gibi açıklanmaktadır.

Tanım 3.1 $\hat{\beta}$, β parametresinin yanlı bir tahmin edicisi olarak kabul edilsin. $\hat{\beta}$ 'nin yanlılık vektörü, $Bias(\hat{\beta}) = E(\hat{\beta}) - \beta = \mathbf{W}\beta$ olarak verilirse, ki bu da $E(\hat{\beta} - \mathbf{W}\beta) = \beta$ anlamına gelir. O zaman $\tilde{\beta} = \hat{\beta} - \mathbf{W}\hat{\beta} = (\mathbf{I} - \mathbf{W})\hat{\beta}$ tahmin edicisine, $\hat{\beta}$ 'nin yanlı tahmin edicisine dayalı olarak elde edilen hemen hemen yansız tahmin edicisi denir (Xu ve Yang, 2011).

Bu tanım, regresyon modelindeki tahmin edicisinin yanını kullanarak yanlılık düzeltilmesini gerçekleştirir ve bu sayede tahmin edicinin hemen hemen yansız olmasını sağlar. Yanlılık düzeltilmesi, regresyon tahminlerinin gerçek değerlere daha yakın olması amacını taşır. Alt bölümlerde, Tanım 3.1'i kullanarak tahmin edicilerin hemen hemen yansız halleri tanıtılacaktır.

3.4.1. Hemen Hemen Yansız Liu Tahmin Edicisi

Doğrusal regresyon modelinde Akdeniz ve Kaçıranlar (1995) tarafından öne sürülen genelleştirilmiş liu tahmin edicisinden hareketle, hemen hemen yansız genelleştirilmiş liu tahmin edicisini tanımladılar. Daha sonra Akdeniz'in çalışmasından ilham alarak Alheety ve Kibria (2009) çoklu doğrusal regresyon modeli için hemen hemen yansız Liu tahmin edicisi (AULE)'yi geliştirmiştir. Bu tahmin edici AULE'nin yanını azaltarak daha güvenilir tahminler elde etmeyi amaçlamaktadır. AULE'nin türetilişi aşağıdaki gibidir.

Liu tahmin edicisinin yanlışlık vektörü, Denklem (3.17)'de aşağıdaki gibi verilmiştir

$$Bias(\hat{\beta}_{LE}) = -(1-d)(\mathbf{X}'\mathbf{X} + \mathbf{I})^{-1}\beta.$$

Bu yanlışlık vektörünü kullanarak, Kadiyala (1984), Ohtani (1986) ve Varathan ve Wijekoon (2021) gibi yukarıdaki Tanım 3.1'i takip eden makaleler incelenerek yanlışlık düzeltmeleri yapılmıştır. Bu düzeltmelerin sonucunda, hemen hemen yansız Liu tahmin edicisi (AULE: almost unbiased Liu estimator):

$$\begin{aligned}\hat{\beta}_{AULE} &= \hat{\beta}_{LE} - Bias(\hat{\beta}_{LE}) \\ &= \hat{\beta}_{LE} - (-(1-d)(\mathbf{X}'\mathbf{X} + \mathbf{I})^{-1}\beta)\end{aligned}$$

olarak elde edilmiştir. Burada, $Bias(\hat{\beta}_{LE})$ formülündeki β yerine $\hat{\beta}_{LE}$ yazılacaktır (bkz. Tanım 3.1). Bu durumda, AULE'nin türetilişi aşağıda daha ayrıntılı olarak verilmiştir

$$\begin{aligned}\hat{\beta}_{AULE} &= \hat{\beta}_{LE} - \left(-(1-d)(\mathbf{X}'\mathbf{X} + \mathbf{I})^{-1}\hat{\beta}_{LE} \right) \\ &= \hat{\beta}_{LE} + (1-d)(\mathbf{X}'\mathbf{X} + \mathbf{I})^{-1}\hat{\beta}_{LE} \\ &= (\mathbf{I} + (1-d)(\mathbf{X}'\mathbf{X} + \mathbf{I})^{-1})\hat{\beta}_{LE} \\ &= (\mathbf{I} + (1-d)(\mathbf{X}'\mathbf{X} + \mathbf{I})^{-1})(\mathbf{X}'\mathbf{X} + \mathbf{I})^{-1}(\mathbf{X}'\mathbf{X} + d\mathbf{I})\hat{\beta}_{OLS} \\ &= (\mathbf{I} + (1-d)(\mathbf{X}'\mathbf{X} + \mathbf{I})^{-1})(\mathbf{I} - (1-d)(\mathbf{X}'\mathbf{X} + \mathbf{I})^{-1})\hat{\beta}_{OLS} \\ &= (\mathbf{I} - (1-d)^2(\mathbf{X}'\mathbf{X} + \mathbf{I})^{-2})\hat{\beta}_{OLS}.\end{aligned}$$

Yukarıda $\hat{\beta}_{LE}$ yerine Denklem (3.15): $\hat{\beta}_{LE} = (\mathbf{X}'\mathbf{X} + \mathbf{I})^{-1}(\mathbf{X}'\mathbf{X} + d\mathbf{I})\hat{\beta}_{OLS}$ yazılmıştır. Hemen hemen yansız Liu tahmin edicisi, en küçük kareler tahmin edicisi $\hat{\beta}_{OLS}$ üzerine inşa edilmiştir ve $\hat{\beta}_{AULE}$ olarak adlandırılmıştır. Yani kısaca AULE,

$$\hat{\beta}_{AULE} = (\mathbf{I} - (1-d)^2(\mathbf{X}'\mathbf{X} + \mathbf{I})^{-2})\hat{\beta}_{OLS}, \quad 0 < d < 1 \quad (3.24)$$

ifadesine eşittir (Alheety ve Kibria, 2009). Burada \mathbf{X} , bağımsız değişken matrisini; $\hat{\beta}_{OLS}$, en küçük kareler tahmin edicisi ile elde edilen parametre vektörünü; \mathbf{I} , birim matrisi ve d , $[0, 1]$ aralığında tanımlanan AULE'nin yanlışlık parametresini temsil etmektedir.

3.4.1.1. Hemen Hemen Yansız Liu Tahmin Edicisinin Teorik Özellikleri

Bu bölümde hemen hemen yansız Liu tahmin edicisinin teorik özelliklerinden olan beklenen değeri, yanlışlık vektörü, varyans-kovaryans matrisi, matris hata kareler ortalaması değeri ve skaler hata kareler ortalaması değeri incelenecektir.

AULE'nin beklenen değeri

$$\begin{aligned}
 E(\widehat{\beta}_{AULE}) &= E\left(\mathbf{I} - (1-d)^2(\mathbf{X}'\mathbf{X} + \mathbf{I})^{-2}\widehat{\beta}_{OLS}\right) \\
 &= E\left(\mathbf{I} - (1-d)^2(\mathbf{X}'\mathbf{X} + \mathbf{I})^{-2}\right) E(\widehat{\beta}_{OLS}) \\
 &= \mathbf{I} - (1-d)^2(\mathbf{X}'\mathbf{X} + \mathbf{I})^{-2}\beta,
 \end{aligned} \tag{3.25}$$

AULE'nin yanlışlık vektörü

$$\begin{aligned}
 Bias(\widehat{\beta}_{AULE}) &= E(\widehat{\beta}_{AULE}) - \beta \\
 &= (\mathbf{I} - (1-d)^2(\mathbf{X}'\mathbf{X} + \mathbf{I})^{-2})\beta - \beta \\
 &= (\mathbf{I} - (1-d)^2(\mathbf{X}'\mathbf{X} + \mathbf{I})^{-2} - \mathbf{I})\beta \\
 &= -(1-d)^2(\mathbf{X}'\mathbf{X} + \mathbf{I})^{-2}\beta
 \end{aligned} \tag{3.26}$$

ve AULE'nin kovaryans matrisi ise,

$$\begin{aligned}
 Cov(\widehat{\beta}_{AULE}) &= Cov\left(\mathbf{I} - (1-d)^2(\mathbf{X}'\mathbf{X} + \mathbf{I})^{-2}\widehat{\beta}_{OLS}\right) \\
 &= \left((\mathbf{I} - (1-d)^2(\mathbf{X}'\mathbf{X} + \mathbf{I})^{-2}) Cov(\widehat{\beta}_{OLS})\right. \\
 &\quad \left. \times ((\mathbf{I} - (1-d)^2(\mathbf{X}'\mathbf{X} + \mathbf{I})^{-2}))'\right) \\
 &= (\mathbf{I} - (1-d)^2(\mathbf{X}'\mathbf{X} + \mathbf{I})^{-2})\sigma^2(\mathbf{X}'\mathbf{X})^{-1} \\
 &\quad \times (\mathbf{I} - (1-d)^2(\mathbf{X}'\mathbf{X} + \mathbf{I})^{-2})' \\
 &= \sigma^2(\mathbf{I} - (1-d)^2(\mathbf{X}'\mathbf{X} + \mathbf{I})^{-2})(\mathbf{X}'\mathbf{X})^{-1} \\
 &\quad \times (\mathbf{I} - (1-d)^2(\mathbf{X}'\mathbf{X} + \mathbf{I})^{-2})
 \end{aligned} \tag{3.27}$$

şeklindedir. Bu kovaryans matrisinin köşegen elemanlarının toplamı, AULE'nin toplam varyansını verir. AULE'nin toplam varyansı

$$\begin{aligned}
Var(\widehat{\boldsymbol{\beta}}_{AULE}) &= tr\left(Cov(\widehat{\boldsymbol{\beta}}_{AULE})\right) \\
&= tr(\sigma^2(\mathbf{I} - (1-d)^2(\mathbf{X}'\mathbf{X} + \mathbf{I})^{-2})(\mathbf{X}'\mathbf{X})^{-1} \\
&\quad \times (\mathbf{I} - (1-d)^2(\mathbf{X}'\mathbf{X} + \mathbf{I})^{-2})) \\
&= \sigma^2 \sum_{j=1}^p \left(1 - \frac{(1-d)^2}{(\lambda_j + 1)^2}\right) \frac{1}{\lambda_j} \left(1 - \frac{(1-d)^2}{(\lambda_j + 1)^2}\right) \\
&= \sigma^2 \sum_{j=1}^p \frac{(\lambda_j + d)^2(\lambda_j + 2 - d)^2}{(\lambda_j + 1)^4} \tag{3.28}
\end{aligned}$$

olarak elde edilir. Burada $\lambda_1, \lambda_2, \dots, \lambda_p > 0$ matris $\mathbf{X}'\mathbf{X}$ 'in özdeğerlerini, $tr(.)$ ise iz operatörünü temsil eder.

AULE'nin matris hata kareler ortalaması

$$\begin{aligned}
MMSE(\widehat{\boldsymbol{\beta}}_{AULE}) &= Cov(\widehat{\boldsymbol{\beta}}_{AULE}) + Bias(\widehat{\boldsymbol{\beta}}_{AULE})Bias(\widehat{\boldsymbol{\beta}}_{AULE})' \\
&= \sigma^2(\mathbf{I} - (1-d)^2(\mathbf{X}'\mathbf{X} + \mathbf{I})^{-2})(\mathbf{X}'\mathbf{X})^{-1} \\
&\quad \times (\mathbf{I} - (1-d)^2(\mathbf{X}'\mathbf{X} + \mathbf{I})^{-2}) + (-(1-d)^2(\mathbf{X}'\mathbf{X} + \mathbf{I})^{-2}\boldsymbol{\beta}) \\
&\quad \times (-(1-d)^2(\mathbf{X}'\mathbf{X} + \mathbf{I})^{-2}\boldsymbol{\beta})' \\
&= \sigma^2(\mathbf{I} - (1-d)^2(\mathbf{X}'\mathbf{X} + \mathbf{I})^{-2})(\mathbf{X}'\mathbf{X})^{-1} \\
&\quad (\mathbf{I} - (1-d)^2(\mathbf{X}'\mathbf{X} + \mathbf{I})^{-2}) + (1-d)^4(\mathbf{X}'\mathbf{X} + \mathbf{I})^{-4}\boldsymbol{\beta}\boldsymbol{\beta}' \tag{3.29}
\end{aligned}$$

ve AULE'nin skaler hata kareler ortalaması ise,

$$\begin{aligned}
MSE(\widehat{\boldsymbol{\beta}}_{AULE}) &= tr(MMSE(\widehat{\boldsymbol{\beta}}_{AULE})) \\
&= tr(\sigma^2(\mathbf{I} - (1-d)^2(\mathbf{X}'\mathbf{X} + \mathbf{I})^{-2})(\mathbf{X}'\mathbf{X})^{-1} \\
&\quad \times (\mathbf{I} - (1-d)^2(\mathbf{X}'\mathbf{X} + \mathbf{I})^{-2}) + (1-d)^4(\mathbf{X}'\mathbf{X} + \mathbf{I})^{-4}\boldsymbol{\beta}\boldsymbol{\beta}') \\
&= \sigma^2 \sum_{j=1}^p \left(1 - \frac{(1-d)^2}{(\lambda_j + 1)^2}\right) \frac{1}{\lambda_j} \left(1 - \frac{(1-d)^2}{(\lambda_j + 1)^2}\right) \\
&\quad + (1-d)^4 \sum_{j=1}^p \frac{\beta_j^2}{(\lambda_j + 1)^4} \\
&= \sigma^2 \sum_{j=1}^p \frac{(\lambda_j + d)^2(\lambda_j + 2 - d)^2}{\lambda_j(\lambda_j + 1)^4} + (1-d)^4 \sum_{j=1}^p \frac{\beta_j^2}{(\lambda_j + 1)^4} \tag{3.30}
\end{aligned}$$

$$= \gamma_1(d) + \gamma_2(d) \tag{3.31}$$

olarak bulunur (Alheety ve Kibria, 2009). Bu denklemde d , AULE'nin yanlılık parametresini simgelerken, σ^2 hata varyansını, λ_j 'ler $\mathbf{X}'\mathbf{X}$ matrisinin özdeğerlerini ve β_j 'ler tahmin edicinin katsayılarını temsil eder. Denklem (3.31)'deki ilk terim $\gamma_1(d)$ toplam varyans, ikinci terim $\gamma_2(d)$ ise toplam karesel yanlılıktır. AULE tahmin edicisinin performansının iyi olması için yani birinci terimdeki azalmanın, ikinci terimdeki artışı aşacak şekilde bir d değeri bulmak gerekmektedir.

3.4.1.2. d Yanlılık Parametresinin Tahmini

$MSE(\widehat{\beta}_{AULE})$ d parametresine bağlı bir fonksiyondur. $MSE(\widehat{\beta}_{AULE}) = w(d)$ olsun. d parametresine göre birinci türevi alındığında

$$w'(d) = \frac{\partial(MSE(\widehat{\beta}_{AULE}))}{\partial d} = 0 \quad (3.32)$$

ve $w'(d) = 0$ denklemi çözüldüğünde optimal d değeri şu şekilde elde edilir:

Denklem (3.30)'deki AULE'nin MSE'si şu şekilde düzenlenebilir.

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

Denklem (3.32)'deki türev fonksiyonu

$$\begin{aligned} &\frac{\partial(MSE(\widehat{\beta}_{AULE}))}{\partial d} \\ &= \sum_{j=1}^p \frac{1}{\lambda_j(\lambda_j + 1)^4} (2\sigma^2(\lambda_j + d)(\lambda_j + 2 - d)^2 - 2\sigma^2(\lambda_j + 2 - d)(\lambda_j + d)^2) \\ &\quad - \sum_{j=1}^p \frac{1}{\lambda_j(\lambda_j + 1)^4} (4(1 - d)^3 \lambda_j \beta_j^2) \\ &= \sum_{j=1}^p \frac{1}{\lambda_j(\lambda_j + 1)^4} (4\sigma^2(\lambda_j + d)(\lambda_j + 2 - d)(1 - d)) - \sum_{j=1}^p \frac{1}{\lambda_j(\lambda_j + 1)^4} \\ &\quad \times (4(1 - d)^3 \lambda_j \beta_j^2) \end{aligned}$$

$$\begin{aligned}
&= \sum_{j=1}^p \frac{1}{\lambda_j(\lambda_j + 1)^4} (\sigma^2(\lambda_j + d)(\lambda_j + 2 - d)) - \sum_{j=1}^p \frac{1}{\lambda_j(\lambda_j + 1)^4} \\
&\quad \times ((1 - d)^2 \lambda_j \beta_j^2) \\
&= \sum_{j=1}^p \frac{1}{\lambda_j(\lambda_j + 1)^4} (\sigma^2(\lambda_j + d)(\lambda_j + 2 - d) - (1 - d)^2 \lambda_j \beta_j^2) = 0 \quad (3.33)
\end{aligned}$$

olarak elde edilir. Buradaki $(\sigma^2(\lambda_j + d)(\lambda_j + 2 - d) - (1 - d)^2 \lambda_j \beta_j^2)$ ifadesini $h_j(d)$ olarak tanımlarsak, $h_j(d) = (\sigma^2(\lambda_j + d)(\lambda_j + 2 - d) - (1 - d)^2 \lambda_j \beta_j^2) = 0$ olarak yazılabilir. Dolayısıyla, $h_j(d)$ fonksiyonu, d değişkeninin karesel bir fonksiyonu olarak ifade edilebilir:

$$h_j(d) = d^2 \underbrace{(-\sigma^2 - \lambda_j \beta_j^2)}_a + d \underbrace{(2\sigma^2 + 2\lambda_j \beta_j^2)}_b + \underbrace{(\sigma^2 \lambda_j^2 + 2\sigma^2 \lambda_j - \lambda_j \beta_j^2)}_c = 0.$$

Bu fonksiyonun kökleri diskriminant ($\Delta = b^2 - 4ac$) kullanılarak elde edilir. $h_j(d)$ 'nin diskriminantı,

$$\begin{aligned}
\Delta_{h_j(d)} &= (2\sigma^2 + 2\lambda_j \beta_j^2)^2 - 4(-\sigma^2 - \lambda_j \beta_j^2)(\sigma^2 \lambda_j^2 + 2\sigma^2 \lambda_j - \lambda_j \beta_j^2) \\
&= 4\sigma^2 \lambda_j \beta_j^2 + 8\sigma^2 \lambda_j^2 \beta_j^2 + 4\sigma^2 \lambda_j^3 \beta_j^2 + 4\sigma^4 + 8\sigma^4 \lambda_j + 4\sigma^4 \lambda_j^2 \\
&= 4\sigma^2 (\lambda_j \beta_j^2 + 2\lambda_j^2 \beta_j^2 + \lambda_j^3 \beta_j^2 + \sigma^2 + 2\sigma^2 \lambda_j + \sigma^2 \lambda_j^2) \\
&= 4\sigma^2 [\lambda_j \beta_j^2 (\lambda_j + 1)^2 + \sigma^2 (\lambda_j + 1)^2] \\
&= 4\sigma^2 (\lambda_j + 1)^2 (\sigma^2 + \lambda_j \beta_j^2) > 0
\end{aligned}$$

olarak bulunur. Diskriminantın pozitif olması, karesel denklemin iki reel kökü olduğunu gösterir. Diskriminant kullanılarak $h_j(d)$ 'nin kökleri, d_{1j} ve d_{2j} olarak adlandırılır ve bu kökler $d_{1j,2j} = \left(\frac{-b \pm \sqrt{\Delta}}{2a} \right)$ formülü ile elde edilir. Elde edilen kökler şu şekildedir:

$$d_{1j,2j} = 1 \mp \sqrt{\frac{\sum_{j=1}^p \frac{\sigma^2}{\lambda_j(\lambda_j+1)^2}}{\sum_{j=1}^p \frac{\sigma^2 + \lambda_j \beta_j^2}{\lambda_j(\lambda_j+1)^4}}}.$$

AULE'nin minimum MSE'sini verecek optimal d değeri,

$$d_{AULE} = 1 - \sqrt{\frac{\sum_{j=1}^p \frac{\sigma^2}{\lambda_j(\lambda_j+1)^2}}{\sum_{j=1}^p \frac{\sigma^2 + \lambda_j \beta_j^2}{\lambda_j(\lambda_j+1)^4}}}$$

olarak seçilir. d_{AULE} , bilinmeyen parametreler β_j 'ler, σ^2 bağlı olduğundan, bunları yansız tahmin edicilerle değiştiririz ve tahmin edilen optimal d değeri,

$$\hat{d}_{AULE} = 1 - \sqrt{\frac{\sum_{j=1}^p \frac{\hat{\sigma}^2}{\lambda_j(\lambda_j+1)^2}}{\sum_{j=1}^p \frac{\hat{\sigma}^2 + \lambda_j \hat{\beta}^2}{\lambda_j(\lambda_j+1)^4}}} \quad (3.34)$$

olarak bulunur (Alheety ve Kibria, 2009).



4. BETA REGRESYON MODELİ

4.1. Giriş

Regresyon analizi aşamasında kullandığımız bağımlı değişkenin dağılımı hakkında bilgi sahibi olmadan regresyon modelini kuramayız. Bu yüzden bağımlı değişkenin değer aralığı ve dağılımı bizim için önemlidir. Ferrari and Cribari-Neto (2004) tarafından tanıtılan beta regresyon modeli, bağımlı değişkenin değer aralığı $(0, 1)$ aralığı ile kısıtlı olduğunda ve normal dağılıma uygun olmadığı durumlarda kullanılır. Bu yöntem, bağımlı değişkenin beta dağılımına uymasını varsayar ve bu dağılımın parametrelerini tahmin ederek bağımsız değişkenin bağımlı değişken üzerindeki etkisini ölçer. BRM $(0, 1)$ aralığında tanımlı oranlar, orantılar, kesirler veya yüzdeler şeklindeki verileri analiz ederken yaygın olarak kullanılır (Qasim et al., 2021). Örneğin; ülkelerin işsizlik oranları, her ülke için gini indeksi, üniversitelerin mezuniyet oranları veya tıbbi konularda vücut yağ yüzdesi gibi çok sayıda verileri regresyon analizinde bağımlı değişken olarak seçip beta regresyon analizi ile modelleyebiliriz (Qasim et al., 2021).

y_1, y_2, \dots, y_n beta dağılımına sahip bir rasgele değişken olmak üzere, beta dağılımının olasılık yoğunluk fonksiyonu,

$$f(y; a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} y^{a-1} (1-y)^{b-1}; \quad 0 < y < 1 \quad (4.1)$$

şeklinde yazılır. Burada y beta dağılımlı bir değişken olup, $y \sim Beta(a, b)$ 'dir ve a ve b ($a, b > 0$) dağılımın parametreleridir. Bu parametrelerin değerleri, dağılımın şeklini ve merkezini belirler. Ayrıca, $\Gamma(\cdot)$ gamma fonksiyonunu temsil eder.

Beta dağılımının beklenen değeri şu şekildedir:

$$E(y) = \frac{a}{a+b}.$$

Beta dağılımının varyansı şu şekildedir:

$$Var(y) = \frac{ab}{(a+b)^2(a+b+1)}.$$

Bu ifadeler, beta dağılımının temel istatistiksel özelliklerini açıklar ve bu dağılımın, a ve b parametrelerine bağlı olarak nasıl değiştiğini gösterir.

Ferrari and Cribari-Neto (2004), Denklem (4.1) için başka bir parametrelendirme önermiştir. $\mu = \frac{a}{a+b}$ ve $\phi = a + b$ olarak tanımlanırsa ve gerekli dönüşümler yapılırsa a ve b , μ ve ϕ 'ye bağlı olarak $a = \mu\phi$ ve $b = (1 - \mu)\phi$ olarak ifade edilebilir. Bu durumda, y beta dağılımlı ($y \sim Beta(\mu\phi, (1 - \mu)\phi)$) sürekli bir rastgele değişkendir ve y , μ ve ϕ sırasıyla $0 < y < 1$, $0 < \mu < 1$ ve $\phi > 0$ olacak şekilde bu parametre dönüşümleri ile beta dağılımının olasılık yoğunluk fonksiyonu tekrar yazılırsa,

$$f(y; \mu, \phi) = \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma((1 - \mu)\phi)} y^{\mu\phi-1} (1 - y)^{(1-\mu)\phi-1}; 0 < y < 1 \quad (4.2)$$

şekline ifade edilebilir. Bu eşitlikte yer alan $\Gamma(\cdot)$ gamma fonksiyonunu, μ bağımlı değişkenin ortalamasını ve ϕ dağılım (dispersion) parametresini temsil eder.

Beta dağılımının beklenen değeri,

$$E(y) = \mu$$

ve varyansı,

$$Var(y) = \frac{\mu(1 - \mu)}{(1 + \phi)}$$

şeklinde dir.

Beta regresyon modelinde logit, probit, log-log, tamamlayıcı log-log ve cauchy gibi çeşitli bağlantı fonksiyonları kullanılabilir. Alternatif bağlantı fonksiyonları bulunsada, genellikle yaygın olarak kullanılan bağlantı fonksiyonu, Ferrari and Cribari-Neto (2004) tarafından $g(\mu) = \log(\mu/(1 - \mu))$ olarak verilen logit bağlantı fonksiyonudur. Bu bağlantı fonksiyonu μ_i , açıklayıcı değişkenlere bağlı olarak aşağıdaki gibi ifade edilir:

$$g(\mu_i) = \log\left(\frac{\mu_i}{1 - \mu_i}\right) = \sum_{j=1}^p \mathbf{x}_{ij} \boldsymbol{\beta}_j = \mathbf{x}'_i \boldsymbol{\beta} = \boldsymbol{\eta}_i. \quad (4.3)$$

Burada, $g(\cdot)$ monoton türevlenebilir bir bağlantı fonksiyonunu, \mathbf{x}'_i $n > p$ olacak şekilde $n \times p$ boyutlu matrisi temsil eder. Öyle ki $i = 1, 2, \dots, n$ için $\mathbf{x}'_i = [x_1, x_2, \dots, x_n]'$, i . gözlem vektörünü, $j = 1, 2, \dots, p$ için $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)'$ ise $p \times 1$ boyutunda bilinmeyen parametrelerin vektörünü ve $\boldsymbol{\eta}$ lineer kestiriciyi temsil eder. Ayrıca

$$\mu_i = \frac{\exp(\mathbf{x}'_i \boldsymbol{\beta})}{1 + \exp(\mathbf{x}'_i \boldsymbol{\beta})} \quad (4.4)$$

şeklinde. BRM'nin bilinmeyen regresyon katsayısını tahmin etmek için en çok olabilirlik tahmin edicisi kullanılır (Qasim et al., 2021).

4.2. En Çok Olabilirlik Tahmin Edicisi

MLE yöntemi, tahmin edicileri bulmak için sıkça kullanılan bir istatistiksel yöntemdir. Bu yöntemin uygulanabilmesi için ilk olarak log-olabilirlik fonksiyonu oluşturulur.

Beta regresyon modelinde log-olabilirlik fonksiyonu:

$$l(\boldsymbol{\beta}) = \sum_{i=1}^n l_i(\mu_i, \phi) = \sum_{i=1}^n \left[\log(\Gamma(\phi)) - \log(\Gamma(\mu_i \phi)) - \log(\Gamma(1 - \mu_i) \phi) + (\phi \mu_i - 1) \log(y_i) + ((1 - \mu_i) \phi - 1) \log(1 - y_i) \right] \quad (4.5)$$

şeklinde. MLE tahmin edicisi Denklem (4.5)'deki log-olabilirlik fonksiyonunu maksimum yapan vektördür. Bu yüzden fonksiyon $\boldsymbol{\beta}$ ve ϕ parametrelerine bağlı olarak maksimize edilir. Ancak log-olabilirlik fonksiyonu doğrusal olmadığından, parametre tahminlerini elde etmek için tekrarlayan bir algoritma kullanılması gerekmektedir. Bu nedenle, skor fonksiyonları, log-olabilirlik fonksiyonunun $\boldsymbol{\beta}$ ve ϕ parametrelerine göre sırasıyla türevi alınarak elde edilir. Elde edilen $\mathbf{S}_{\boldsymbol{\beta}}(\boldsymbol{\beta}, \phi)$ için skor fonksiyonu şu şekildedir:

$$\mathbf{S}_{\boldsymbol{\beta}}(\boldsymbol{\beta}, \phi) = \phi \mathbf{X}' \mathbf{T} (\mathbf{y}^* - \boldsymbol{\mu}^*). \quad (4.6)$$

Burada

$$\mathbf{T} = \text{diag}\left(\frac{1}{g'(\mu_1)}, \dots, \frac{1}{g'(\mu_n)}\right), \quad \mathbf{y}^* = (y_1^*, \dots, y_n^*)', \quad \boldsymbol{\mu}^* = (\mu_1^*, \dots, \mu_n^*)',$$

$$y_i^* = \log\left(\frac{y_i}{1 - y_i}\right) \quad \text{ve} \quad \mu_i^* = \psi\left(\mu_i \left(\frac{1 - \sigma_i^2}{\sigma_i^2}\right)\right) - \psi\left((1 - \mu_i) \left(\frac{1 - \sigma_i^2}{\sigma_i^2}\right)\right)$$

şeklinde tanımlanır. $\psi(\cdot)$ digamma fonksiyonunu ifade eder ve $\boldsymbol{\beta}$ parametre vektörünün tahmin edilmesi için kullanılan tekrarlı ağırlıklanmış en küçük kareler (IWLS) algoritması veya Fisher skor algoritmasıdır (Espinheira ve Silva, 2019). IWLS algoritmasının parametre güncelleme adımı şu şekildedir:

$$\boldsymbol{\beta}^{(r+1)} = \boldsymbol{\beta}^{(r)} + \left(\mathbf{I}_{\boldsymbol{\beta}\boldsymbol{\beta}}^{(r)}\right)^{-1} \mathbf{S}_{\boldsymbol{\beta}}(\boldsymbol{\beta}, \phi).$$

Bu algoritmada, $\mathbf{S}_{\boldsymbol{\beta}}(\boldsymbol{\beta}, \phi)$, Denklem (4.6) içinde tanımlanan skor fonksiyonudur ve $\mathbf{I}_{\boldsymbol{\beta}\boldsymbol{\beta}}^{(r)}$ Fisher bilgi matrisidir (Espinheira ve Silva, 2019). $\boldsymbol{\beta}$ 'nin başlangıç değeri en

küçük kareler yöntemiyle elde edilebilirken, her bir dağılım parametresinin başlangıç değeri şu şekildedir:

$$\hat{\phi}_i = \frac{\hat{\mu}_i(1 - \hat{\mu}_i)}{\hat{\sigma}_i^2}. \quad (4.7)$$

Burada $\hat{\mu}$ ve $\hat{\sigma}_i^2$ değerleri lineer regresyondan elde edilir. $r = 0, 1, 2, \dots$ için ardışık tahminler arasındaki fark belirli bir sabitten daha küçük olduğunda yakınsama gerçekleşir ve MLE tahmin edicisi β aşağıdaki gibi ifade edilir:

$$\hat{\beta}_{MLE} = (\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X})^{-1}\mathbf{X}'\widehat{\mathbf{W}}\mathbf{z}. \quad (4.8)$$

Burada \mathbf{X} açıklayıcı değişken matrisini, \mathbf{z} bağımlı değişken vektörü, öyle ki $\mathbf{z} = \widehat{\mathbf{W}}^{-1}\widehat{\mathbf{T}}(\mathbf{y}^* - \boldsymbol{\mu}^*)$ ve $\widehat{\mathbf{W}} = \text{diag}(\hat{w}_1, \dots, \hat{w}_n)$ şeklinde tanımlanan ağırlık matrisidir. Burada $\widehat{\mathbf{W}}$ ve $\widehat{\mathbf{T}}$, MLE tahmin edicisindeki \mathbf{W} ve \mathbf{T} matrisleridir (Ferrari and Cribari-Neto, 2004). Ayrıca \hat{w}_i şu şekildedir:

$$\hat{w}_i = \frac{(1 - \hat{\sigma}_i^2)}{\hat{\sigma}_i^2} \left\{ \psi' \left(\frac{\hat{\mu}_i(1 - \hat{\sigma}_i^2)}{\hat{\sigma}_i^2} \right) + \psi' \left(\frac{(1 - \hat{\mu}_i)(1 - \hat{\sigma}_i^2)}{\hat{\sigma}_i^2} \right) \right\} \frac{1}{g'(\hat{\mu}_i)^2}$$

β 'nin MLE tahmin edicisi asimptotik olarak normal dağılır.

$\hat{\beta}_{MLE}$ 'nin beklenen değeri:

$$E(\hat{\beta}_{MLE}) = \beta.$$

$\hat{\beta}_{MLE}$ 'nin kovaryans matrisi:

$$\text{Cov}(\hat{\beta}_{MLE}) = \phi^{-1} (\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X})^{-1} \quad (4.9)$$

şekindedir. Dolayısıyla $\hat{\beta}_{MLE}$ 'nin hata kareler ortalaması değeri aşağıdaki gibi elde edilir

$$\begin{aligned} \text{MSE}(\hat{\beta}_{MLE}) &= E \left((\hat{\beta}_{MLE} - \beta)' (\hat{\beta}_{MLE} - \beta) \right) \\ &= E \left(\hat{\beta}_{MLE} - \beta \right)' E \left(\hat{\beta}_{MLE} - \beta \right) \\ &= \text{tr} \left[\phi^{-1} (\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X})^{-1} \right] \\ &= \phi^{-1} \sum_{j=1}^p \frac{1}{\lambda_j}. \end{aligned} \quad (4.10)$$

Burada λ_j , matris $\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X}$ 'in j . özdeğerlerini, $\text{tr}(\cdot)$ iz operatörünü ve ϕ^{-1} hata varyansını ifade etmektedir.

4.3. Beta Ridge Tahmin Edicisi

Ağırlıklı hata kareler toplamını minimize etmek amacıyla $\widehat{\beta}$, β vektörünün bir tahminicisi olduğunu varsayalım. Ağırlıklı hata kareler toplamı şu şekilde tanımlanabilir

$$\begin{aligned}\Theta &= (\mathbf{y} - \widehat{\beta})'(\mathbf{y} - \widehat{\beta}) \\ &= (\mathbf{y} - \mathbf{X}\widehat{\beta}_{MLE})'(\mathbf{y} - \mathbf{X}\widehat{\beta}_{MLE}) + (\widehat{\beta} - \widehat{\beta}_{MLE})'\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X}(\widehat{\beta} - \widehat{\beta}_{MLE}) \\ &= \Theta_{min} + \Theta(\widehat{\beta}).\end{aligned}$$

Burada, $\Theta(\widehat{\beta}) > 0$ ise, $\widehat{\beta}_{MLE}$ yerine $\widehat{\beta}$ konulduğunda ağırlıklı hata kareleri toplamındaki artışı ifade eder. Beta ridge tahmin edicisi, Hoerl ve Kennard (1970a) tarafından önerilen bir kısıtlama altında elde edilir. Yani, $(\widehat{\beta} - \widehat{\beta}_{MLE})'\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X}(\widehat{\beta} - \widehat{\beta}_{MLE}) = \Theta_0$ kısıtlamasına tabi olarak $\widehat{\beta}'\widehat{\beta}$ uzunluğunun minimize edilmesiyle elde edilir.

Minimize edilerek toplam hatayı ifade eden Q fonksiyonu aşağıdaki gibi tanımlanır:

$$\begin{aligned}\text{Minimize}(Q) &= (\mathbf{y} - \widehat{\beta})'(\mathbf{y} - \widehat{\beta}) \\ &= (\mathbf{y} - \mathbf{X}\widehat{\beta}_{MLE})'(\mathbf{y} - \mathbf{X}\widehat{\beta}_{MLE}) + (\widehat{\beta} - \widehat{\beta}_{MLE})'\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X}(\widehat{\beta} - \widehat{\beta}_{MLE}). \\ Q &= \widehat{\beta}'\widehat{\beta} + \frac{1}{k} \left((\widehat{\beta} - \widehat{\beta}_{MLE})'\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X}(\widehat{\beta} - \widehat{\beta}_{MLE}) - \Theta_0 \right) \quad (4.11)\end{aligned}$$

burada $(1/k)$ Lagrangian çarpanıdır. Denklem (4.11)'in $\widehat{\beta}$ 'ye göre türevi alınarak sıfıra eşitlenirse

$$\frac{\partial Q}{\partial \widehat{\beta}} = 2\widehat{\beta} + \frac{\left(2\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X}(\widehat{\beta} - \widehat{\beta}_{MLE}) - \Theta_0 \right)}{k} = 0$$

elde edilir. Bu ifade çözümlenerek $\widehat{\beta}$ yalnız bırakılınca, beta ridge tahmin edicisi şu şekilde elde edilir (Abonazel ve Taha, 2021).

$$\widehat{\beta}_{RE} = (\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X} + k\mathbf{I})^{-1}\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X}\widehat{\beta}_{MLE} \quad (4.12)$$

burada \mathbf{X} , bağımsız değişken matrisini; $\widehat{\beta}_{MLE}$, en çok olabilirlik tahmin edicisi ile elde edilen beta parametre vektörünü; $\widehat{\mathbf{W}}$, ağırlık matrisinin tahminini; \mathbf{I} , birim matrisi ve k , $k > 0$ ile tanımlı yanlışlık parametresini temsil etmektedir.

Beta ridge tahmin edicisinin beklenen değeri

$$\begin{aligned}
 E(\hat{\beta}_{RE}) &= E\left((\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X} + k\mathbf{I})^{-1}\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X}\hat{\beta}_{MLE}\right) \\
 &= E\left((\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X} + k\mathbf{I})^{-1}\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X}\right) E\left(\hat{\beta}_{MLE}\right) \\
 &= (\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X} + k\mathbf{I})^{-1}\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X}\beta.
 \end{aligned} \tag{4.13}$$

Beta ridge tahmin edicisinin yanlılık vektörü

$$\begin{aligned}
 Bias(\hat{\beta}_{RE}) &= E(\hat{\beta}_{RE}) - \beta \\
 &= (\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X} + k\mathbf{I})^{-1}\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X}\beta - \beta \\
 &= \left[(\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X} + k\mathbf{I})^{-1}\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X} - \mathbf{I}\right] \beta \\
 &= \left[\left(\mathbf{I} + k(\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X})^{-1}\right)^{-1} - \mathbf{I}\right] \beta \\
 &= \left[\mathbf{I} - k(\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X} + k\mathbf{I})^{-1} - \mathbf{I}\right] \beta \\
 &= -k(\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X} + k\mathbf{I})^{-1}\beta.
 \end{aligned} \tag{4.14}$$

Beta ridge tahmin edicisinin kovaryans matrisi

$$\begin{aligned}
 Cov(\hat{\beta}_{RE}) &= Cov\left((\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X} + k\mathbf{I})^{-1}\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X}\hat{\beta}_{MLE}\right) \\
 &= \left((\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X} + k\mathbf{I})^{-1}\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X}\right) Cov(\hat{\beta}_{MLE}) \\
 &\quad \times \left((\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X} + k\mathbf{I})^{-1}\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X}\right)' \\
 &= \phi^{-1}(\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X} + k\mathbf{I})^{-1}\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X}(\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X} + k\mathbf{I})^{-1}
 \end{aligned} \tag{4.15}$$

olarak elde edilir. Denklem (4.9): $Cov(\hat{\beta}_{MLE}) = \phi^{-1}(\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X})^{-1}$ yukarıdaki eşitlikte yerine yazıldığında, $Cov(\hat{\beta}_{RE})$ Denklem (4.15)'deki gibi elde edilir. Bu kovaryans matrisinin köşegen elemanlarının toplamı, beta ridge tahmin edicisinin toplam varyansını verir. Beta ridge tahmin edicisinin toplam varyansı

$$\begin{aligned}
 Var(\hat{\beta}_{RE}) &= tr\left(Cov(\hat{\beta}_{RE})\right) \\
 &= tr\left(\phi^{-1}(\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X} + k\mathbf{I})^{-1}\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X}(\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X} + k\mathbf{I})^{-1}\right) \\
 &= \phi^{-1} \sum_{j=1}^p \frac{\lambda_j}{(\lambda_j + k)^2}
 \end{aligned} \tag{4.16}$$

şeklindedir. Beta ridge tahmin edicisinin matris hata kareler ortalaması:

$$\begin{aligned}
MMSE(\widehat{\beta}_{RE}) &= Cov(\widehat{\beta}_{RE}) + Bias(\widehat{\beta}_{RE})Bias(\widehat{\beta}_{RE})' \\
&= \phi^{-1}(\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X} + k\mathbf{I})^{-1}\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X}(\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X} + k\mathbf{I})^{-1} \\
&\quad + (-k(\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X} + k\mathbf{I})^{-1}\boldsymbol{\beta})(-k(\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X} + k\mathbf{I})^{-1}\boldsymbol{\beta})' \\
&= \phi^{-1}(\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X} + k\mathbf{I})^{-1}\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X}(\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X} + k\mathbf{I})^{-1} \\
&\quad + k^2(\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X} + k\mathbf{I})^{-1}\boldsymbol{\beta}\boldsymbol{\beta}'(\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X} + k\mathbf{I})^{-1} \quad (4.17)
\end{aligned}$$

olarak elde edilir. Beta ridge tahmin edicisinin skaler hata kareler ortalaması ise:

$$\begin{aligned}
MSE(\widehat{\beta}_{RE}) &= tr(MMSE(\widehat{\beta}_{RE})) \\
&= tr(\phi^{-1}(\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X} + k\mathbf{I})^{-1}\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X}(\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X} + k\mathbf{I})^{-1} \\
&\quad + k^2(\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X} + k\mathbf{I})^{-1}\boldsymbol{\beta}\boldsymbol{\beta}'(\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X} + k\mathbf{I})^{-1}) \\
&= \phi^{-1} \sum_{j=1}^p \frac{\lambda_j}{(\lambda_j + k)^2} + k^2 \sum_{j=1}^p \frac{\beta_j^2}{(\lambda_j + k)^2} \\
&= \gamma_1(k) + \gamma_2(k) \quad (4.18)
\end{aligned}$$

şeklindedir (Abonazel ve Taha, 2021). Burada, ilk terim $\gamma_1(k)$ toplam varyansı, ikinci terim $\gamma_2(k)$ ise toplam karesel yanlılığı ifade eder. Ayrıca k parametresi, beta ridge tahmin edicisinin yanlılık parametresini simgelerken, ϕ^{-1} hata varyansını, λ_j 'ler $\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X}$ matrisinin özdeğerlerini ve β_j terimi ise modelin j 'inci bağımsız değişkenine ait tahmin katsayısını temsil eder.

4.4. Beta Liu Tahmin Edicisi

Bir beta regresyon modelinde, beta Liu tahmin edicisi Karlsson ve ark. (2020) tarafından

$$\widehat{\beta}_{LE} = (\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X} + \mathbf{I})^{-1}(\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X} + d\mathbf{I})\widehat{\beta}_{MLE} \quad (4.19)$$

olarak tanımlanmıştır. Burada \mathbf{X} , bağımsız değişken matrisini; $\widehat{\beta}_{MLE}$, en çok olabilirlik tahmin edicisi ile elde edilen beta parametre vektörünü; $\widehat{\mathbf{W}}$, ağırlık matrisinin tahminini; \mathbf{I} , birim matrisi; d , $[0, 1]$ aralığı ile tanımlı yanlılık parametresini temsil etmektedir.

Beta Liu tahmin edicisinin beklenen değeri

$$\begin{aligned}
E(\widehat{\beta}_{LE}) &= E\left((\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X} + \mathbf{I})^{-1}(\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X} + d\mathbf{I})\widehat{\beta}_{MLE}\right) \\
&= E\left((\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X} + \mathbf{I})^{-1}(\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X} + d\mathbf{I})\right) E(\widehat{\beta}_{MLE}) \\
&= (\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X} + \mathbf{I})^{-1}(\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X} + d\mathbf{I})\beta,
\end{aligned} \tag{4.20}$$

beta Liu tahmin edicisinin yanlışlık vektörü

$$\begin{aligned}
Bias(\widehat{\beta}_{LE}) &= E(\widehat{\beta}_{LE}) - \beta \\
&= (\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X} + \mathbf{I})^{-1}(\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X} + d\mathbf{I})\beta - \beta \\
&= \left(\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X} + \mathbf{I}\right)^{-1}(\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X} + d\mathbf{I}) - \mathbf{I} \beta \\
&= \left(\left(\mathbf{I} + (\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X})^{-1}\right)^{-1} + d(\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X} + \mathbf{I})^{-1} - \mathbf{I}\right) \beta \\
&= \left(\mathbf{I} - (\mathbf{X}'\mathbf{X} + \mathbf{I})^{-1} + d(\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X} + \mathbf{I})^{-1} - \mathbf{I}\right) \beta \\
&= -(1 - d)(\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X} + \mathbf{I})^{-1}\beta
\end{aligned} \tag{4.21}$$

ve beta Liu tahmin edicisinin kovaryans matrisi ise,

$$\begin{aligned}
Cov(\widehat{\beta}_{LE}) &= Cov\left((\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X} + \mathbf{I})^{-1}(\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X} + d\mathbf{I})\widehat{\beta}_{MLE}\right) \\
&= (\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X} + \mathbf{I})^{-1}(\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X} + d\mathbf{I})Cov(\widehat{\beta}_{MLE}) \\
&\quad \times \left((\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X} + \mathbf{I})^{-1}(\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X} + d\mathbf{I})\right)' \\
&= \phi^{-1}(\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X} + \mathbf{I})^{-1}(\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X} + d\mathbf{I})(\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X})^{-1} \\
&\quad \times (\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X} + d\mathbf{I})(\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X} + \mathbf{I})^{-1}
\end{aligned} \tag{4.22}$$

şeklindedir. Denklem (4.9)'da verilen $Cov(\widehat{\beta}_{MLE}) = \phi^{-1}(\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X})^{-1}$ yukarıdaki denklemde yerine yazıldığında, $Cov(\widehat{\beta}_{LE})$ Denklem (4.22)'deki gibi elde edilir. Bu kovaryans matrisinin köşegen elemanlarının toplamı, beta Liu tahmin edicisinin varyans matrisini verir. Beta Liu tahmin edicisinin toplam varyansı

$$\begin{aligned}
Var(\widehat{\beta}_{LE}) &= tr\left(Cov(\widehat{\beta}_{LE})\right) \\
&= tr(\phi^{-1}(\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X} + \mathbf{I})^{-1}(\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X} + d\mathbf{I})(\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X})^{-1} \\
&\quad \times (\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X} + d\mathbf{I})(\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X} + \mathbf{I})^{-1}) \\
&= \phi^{-1} \sum_{j=1}^p \frac{(\lambda_j + d)^2}{\lambda_j(\lambda_j + 1)^2}
\end{aligned} \tag{4.23}$$

şeklindedir. Burada $\lambda_1, \lambda_2, \dots, \lambda_p > 0$ matris $\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X}$ 'in özdeğerlerini temsil eder.

Beta Liu tahmin edicisinin matris hata kareler ortalaması

$$\begin{aligned}
MMSE(\widehat{\boldsymbol{\beta}}_{LE}) &= Cov(\widehat{\boldsymbol{\beta}}_{LE}) + Bias(\widehat{\boldsymbol{\beta}}_{LE})Bias(\widehat{\boldsymbol{\beta}}_{LE})' \\
&= \phi^{-1}(\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X} + \mathbf{I})^{-1}(\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X} + d\mathbf{I})(\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X})^{-1} \\
&\quad \times (\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X} + d\mathbf{I})(\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X} + \mathbf{I})^{-1} \\
&\quad + (-(1-d)(\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X} + \mathbf{I})^{-1}\boldsymbol{\beta})(-(1-d)(\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X} + \mathbf{I})^{-1}\boldsymbol{\beta})' \\
&= \phi^{-1}(\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X} + \mathbf{I})^{-2}(\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X} + d\mathbf{I})^2(\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X})^{-1} \\
&\quad + (1-d)^2(\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X} + \mathbf{I})^{-1}\boldsymbol{\beta}(\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X} + \mathbf{I})^{-1}\boldsymbol{\beta}' \tag{4.24}
\end{aligned}$$

ve beta Liu tahmin edicisinin skaler hata kareler ortalaması değeri,

$$\begin{aligned}
MSE(\widehat{\boldsymbol{\beta}}_{LE}) &= tr(MMSE(\widehat{\boldsymbol{\beta}}_{LE})) \\
&= tr(\phi^{-1}(\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X} + \mathbf{I})^{-2}(\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X} + d\mathbf{I})^2(\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X})^{-1} \\
&\quad + (1-d)^2(\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X} + \mathbf{I})^{-1}\boldsymbol{\beta}(\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X} + \mathbf{I})^{-1}\boldsymbol{\beta}') \\
&= \phi^{-1} \sum_{j=1}^p \frac{(\lambda_j + d)^2}{\lambda_j(\lambda_j + 1)^2} + (1-d)^2 \sum_{j=1}^p \frac{\beta_j^2}{(\lambda_j + 1)^2} \tag{4.25}
\end{aligned}$$

şeklindedir (Karlsson ve ark., 2020). Ayrıca d parametresi, beta Liu tahmin edicisinin yanlılık parametresini simgelerken, ϕ^{-1} hata varyansını, λ_j 'ler $\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X}$ matrisinin özdeğerlerini ve β_j terimi ise modelin j 'inci bağımsız değişkenine ait tahmin katsayısını temsil eder.

5. HEMEN HEMEN YANSIZ BETA LIU TAHMİN EDİCİSİ

Yanlı tahmin ediciler arasında öne çıkan Liu (1993) tarafından önerilen Liu tahmin edici ve hemen hemen yansız tahmin ediciler arasında öne çıkan Akdeniz ve Kaçiranlar (1995) ve Alheety ve Kibria (2009) tarafından önerilen hemen hemen yansız Liu tahmin edicisi literatürde incelenmiştir. Yapılan literatür incelemesinde, hemen hemen yansız Liu tahmin edicisinin beta regresyon modelinde ele alınmamış olduğu gözlemlenmiştir.

Bu çalışma, beta regresyon modelinde çoklu bağlantı sorunuyla karşılaşıldığında etkili tahminler yapabilen ancak literatürde daha önce tanımlanmamış bir tahmin ediciyi tanıttir. Literatürde adı geçmeyen bu yeni tahmin edici, beta regresyon modelinde kullanılmak üzere hemen hemen yansız beta Liu tahmin edicisi olarak adlandırılmış ve AULE olarak kısaltılmıştır.

Beta regresyon modelinde, beta Liu tahmin edicisinin yanlılık vektörü, Denklem (4.21)'de aşağıdaki gibi verilmiştir:

$$Bias(\hat{\beta}_{LE}) = -(1 - d)(\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X} + \mathbf{I})^{-1}\beta.$$

Bu yanlılık vektörünü kullanarak, Kadiyala (1984), Ohtani (1986) ve Varathan ve Wijekoon (2021) gibi Tanım 4.1'i takip eden makaleler incelenerek yan düzeltmeleri yapılmıştır. Bu düzeltmelerin sonucunda, hemen hemen yansız Liu tahmin edicisi:

$$\begin{aligned}\hat{\beta}_{AULE} &= \hat{\beta}_{LE} - Bias(\hat{\beta}_{LE}) \\ &= \hat{\beta}_{LE} - \left(-(1 - d)(\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X} + \mathbf{I})^{-1}\beta \right)\end{aligned}$$

olarak elde edilmiştir. Ancak burada β yerine $\hat{\beta}_{LE}$ yazılarak, yeni hemen hemen

yansız beta Liu tahmin edicisi $\widehat{\beta}_{AULE}$ aşağıdaki adımlarla elde edilir:

$$\begin{aligned}
\widehat{\beta}_{AULE} &= \widehat{\beta}_{LE} - \left(-(1-d)(\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X} + \mathbf{I})^{-1}\widehat{\beta}_{LE} \right) \\
&= \widehat{\beta}_{LE} + (1-d)(\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X} + \mathbf{I})^{-1}\widehat{\beta}_{LE} \\
&= \left(\mathbf{I} + (1-d)(\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X} + \mathbf{I})^{-1} \right) \widehat{\beta}_{LE} \\
&= \left(\mathbf{I} + (1-d)(\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X} + \mathbf{I})^{-1} \right) (\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X} + \mathbf{I})^{-1} (\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X} + d\mathbf{I}) \widehat{\beta}_{MLE} \\
&= \left(\mathbf{I} + (1-d)(\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X} + \mathbf{I})^{-1} \right) \left(\mathbf{I} - (1-d)(\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X} + \mathbf{I})^{-1} \right) \widehat{\beta}_{MLE} \\
&= \left(\mathbf{I} - (1-d)^2(\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X} + \mathbf{I})^{-2} \right) \widehat{\beta}_{MLE}.
\end{aligned}$$

Yukarıda $\widehat{\beta}_{LE}$ yerine Denklem (4.19): $\widehat{\beta}_{LE} = (\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X} + \mathbf{I})^{-1}(\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X} + d\mathbf{I})\widehat{\beta}_{MLE}$ yazılmıştır. Yani kısaca AULE,

$$\widehat{\beta}_{AULE} = \left(\mathbf{I} - (1-d)^2(\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X} + \mathbf{I})^{-2} \right) \widehat{\beta}_{MLE}, \quad d \in (-\infty, +\infty) \quad (5.1)$$

ifadesine eşittir. Burada \mathbf{X} , bağımsız değişken matrisini; $\widehat{\beta}_{MLE}$, en çok olabilirlik tahmin edicisi ile elde edilen beta parametre vektörünü; $\widehat{\mathbf{W}}$, ağırlık matrisinin tahminini; \mathbf{I} , birim matrisi; d , $(-\infty, +\infty)$ aralığı ile tanımlı AULE'nin yanlılık parametresini temsil etmektedir (Alheety ve Kibria, 2009).

5.1. Hemen Hemen Yansız Beta Liu Tahmin Edicisinin Teorik Özellikleri

Bu bölümde hemen hemen yansız beta Liu tahmin edicisi AULE'nin teorik özelliklerinden olan beklenen değeri, yanlılık vektörü, varyans-kovaryans matrisi, matris hata kareler ortalaması (MMSE) ve skaler hata kareler ortalaması (MSE) değerleri incelenecektir.

AULE'nin beklenen değeri:

$$\begin{aligned}
E(\widehat{\beta}_{AULE}) &= E \left(\mathbf{I} - (1-d)^2(\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X} + \mathbf{I})^{-2} \widehat{\beta}_{MLE} \right) \\
&= E \left(\mathbf{I} - (1-d)^2(\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X} + \mathbf{I})^{-2} \right) E(\widehat{\beta}_{MLE}) \\
&= \mathbf{I} - (1-d)^2(\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X} + \mathbf{I})^{-2}\beta,
\end{aligned} \quad (5.2)$$

AULE'nin yanlışlık vektörü:

$$\begin{aligned}
Bias(\widehat{\beta}_{AULE}) &= E(\widehat{\beta}_{AULE}) - \beta \\
&= (\mathbf{I} - (1-d)^2(\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X} + \mathbf{I})^{-2})\beta - \beta \\
&= \left(\mathbf{I} - (1-d)^2(\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X} + \mathbf{I})^{-2} - \mathbf{I} \right) \beta \\
&= -(1-d)^2(\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X} + \mathbf{I})^{-2}\beta,
\end{aligned} \tag{5.3}$$

AULE'nin kovaryans matrisi ise,

$$\begin{aligned}
Cov(\widehat{\beta}_{AULE}) &= Cov\left(\mathbf{I} - (1-d)^2(\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X} + \mathbf{I})^{-2}\widehat{\beta}_{MLE}\right) \\
&= \left(\mathbf{I} - (1-d)^2(\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X} + \mathbf{I})^{-2}\right) Cov(\widehat{\beta}_{MLE}) \\
&\quad \times \left(\mathbf{I} - (1-d)^2(\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X} + \mathbf{I})^{-2}\right)' \\
&= (\mathbf{I} - (1-d)^2(\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X} + \mathbf{I})^{-2})\phi^{-1}(\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X})^{-1} \\
&\quad \times \left(\mathbf{I} - (1-d)^2(\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X} + \mathbf{I})^{-2}\right)' \\
&= \phi^{-1}(\mathbf{I} - (1-d)^2(\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X} + \mathbf{I})^{-2})(\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X})^{-1} \\
&\quad \times (\mathbf{I} - (1-d)^2(\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X} + \mathbf{I})^{-2})
\end{aligned} \tag{5.4}$$

olarak bulunur. Denklem (4.9): $Cov(\widehat{\beta}_{MLE}) = \phi^{-1}(\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X})^{-1}$ denklemde yerine yazıldığında, $Cov(\widehat{\beta}_{AULE})$ Denklem (5.4)'deki gibi elde edilir. Bu kovaryans matrisinin köşegen elemanlarının toplamı, AULE'nin toplam varyans matrisini verir.

AULE'nin toplam varyans matrisi:

$$\begin{aligned}
Var(\widehat{\beta}_{AULE}) &= tr\left(Cov(\widehat{\beta}_{AULE})\right) \\
&= tr(\phi^{-1}(\mathbf{I} - (1-d)^2(\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X} + \mathbf{I})^{-2})(\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X})^{-1} \\
&\quad \times (\mathbf{I} - (1-d)^2(\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X} + \mathbf{I})^{-2})) \\
&= \phi^{-1} \sum_{j=1}^p \left(1 - \frac{(1-d)^2}{(\lambda_j + 1)^2}\right) \frac{1}{\lambda_j} \left(1 - \frac{(1-d)^2}{(\lambda_j + 1)^2}\right) \\
&= \phi^{-1} \sum_{j=1}^p \frac{(\lambda_j + d)^2(\lambda_j + 2 - d)^2}{\lambda_j(\lambda_j + 1)^4}
\end{aligned} \tag{5.5}$$

olarak bulunur. Burada $\lambda_1, \lambda_2, \dots, \lambda_p > 0$ matris $\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X}$ 'in özdeğerlerini, $tr(\cdot)$ ise iz operatörünü temsil eder.

AULE'nin matris hata kareler ortalaması:

$$\begin{aligned}
MMSE(\widehat{\boldsymbol{\beta}}_{AULE}) &= Cov(\widehat{\boldsymbol{\beta}}_{AULE}) + Bias(\widehat{\boldsymbol{\beta}}_{AULE})Bias(\widehat{\boldsymbol{\beta}}_{AULE})' \\
&= \phi^{-1}(\mathbf{I} - (1-d)^2(\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X} + \mathbf{I})^{-2})(\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X})^{-1} \\
&\quad \times (\mathbf{I} - (1-d)^2(\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X} + \mathbf{I})^{-2}) \\
&\quad + (1-d)^4(\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X} + \mathbf{I})^{-4}\boldsymbol{\beta}\boldsymbol{\beta}', \tag{5.6}
\end{aligned}$$

AULE'nin skaler hata kareler ortalaması değeri ise,

$$\begin{aligned}
MSE(\widehat{\boldsymbol{\beta}}_{AULE}) &= tr(MMSE(\widehat{\boldsymbol{\beta}}_{AULE})) \\
&= tr(\phi^{-1}(\mathbf{I} - (1-d)^2(\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X} + \mathbf{I})^{-2})(\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X})^{-1}(\mathbf{I} - (1-d)^2 \\
&\quad \times (\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X} + \mathbf{I})^{-2} + (1-d)^4(\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X} + \mathbf{I})^{-4}\boldsymbol{\beta}\boldsymbol{\beta}') \\
&= \phi^{-1} \sum_{j=1}^p \left(1 - \frac{(1-d)^2}{(\lambda_j + 1)^2}\right) \frac{1}{\lambda_j} \left(1 - \frac{(1-d)^2}{(\lambda_j + 1)^2}\right) \\
&\quad + (1-d)^4 \sum_{j=1}^p \frac{\beta_j^2}{(\lambda_j + 1)^4} \\
&= \phi^{-1} \sum_{j=1}^p \frac{(\lambda_j + d)^2(\lambda_j + 2 - d)^2}{\lambda_j(\lambda_j + 1)^4} + (1-d)^4 \sum_{j=1}^p \frac{\beta_j^2}{(\lambda_j + 1)^4} \tag{5.7}
\end{aligned}$$

$$= \gamma_1(d) + \gamma_2(d) \tag{5.8}$$

olarak bulunur. Bu denklemde d parametresi, AULE'nin yanlılık parametresini simgelerken, ϕ^{-1} hata varyansını, λ_j 'ler $\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X}$ matrisinin özdeğerlerini ve β_j terimi ise modelin j 'inci bağımsız değişkenine ait tahmin katsayısını temsil eder. Denklem (5.8)'deki ilk terim $\gamma_1(d)$ toplam varyansı, ikinci terim $\gamma_2(d)$ ise toplam karesel yanlılığı ifade eder. AULE tahmin edicisinin performansının iyi olması için yani birinci terimdeki azalmanın, ikinci terimdeki artışı aşacak şekilde bir d değeri bulmak gerekmektedir.

5.2. d Yanlılık Parametresinin Tahmini

AULE'nin d parametresine bağlı olarak değişen MSE fonksiyonunu ele alalım. MSE, d değerine bağlı olarak değişir ve en uygun d değeri, AULE'nin minimum MSE'sini verir. Bu bağlamda, AULE'nin MSE değeri ile yanlılık parametresi d arasında matematiksel bir ilişkiyi ele alınarak, optimal bir d değeri

seçilerek, MSE'nin minimize edilmesi amaçlanmaktadır. Bu nedenle, AULE için d değerini tahmin etmek amacıyla yeni tahmin yöntemleri öneriyoruz. İlk olarak, AULE'nin optimal d değerini bulmak için Denklem (5.7)'de ki MSE fonksiyonunun d parametresine göre türevi alınır. Bu işlem, aşağıdaki denklemlerle ifade edilir:

$$\frac{\partial(MSE(\hat{\beta}_{AULE}))}{\partial d} = 0. \quad (5.9)$$

Bu ifade, $MSE(\hat{\beta}_{AULE})$ fonksiyonunun d parametresine göre birinci türevinin sıfıra eşitlenmesini ifade eder. Gerekli matematiksel işlemler sonucunda elde edilen ifade, optimal d değerini bulmak için kullanılır. $MSE(\hat{\beta}_{AULE})$, d parametresine göre düzenlenmiş bir formda aşağıdaki gibi ifade edilebilir:

$$\begin{aligned} MSE(\hat{\beta}_{AULE}) &= \sum_{j=1}^p \frac{1}{\lambda_j(\lambda_j + 1)^4} (\phi^{-1}(\lambda_j + d)^2 (\lambda_j + 2 - d)^2) \\ &\quad + \sum_{j=1}^p \frac{1}{\lambda_j(\lambda_j + 1)^4} (1 - d)^4 (\lambda_j \beta_j^2). \end{aligned}$$

Denklem (5.9)'daki türev ifadesinin çözümü şu şekildedir:

$$\begin{aligned} &\frac{\partial(MSE(\hat{\beta}_{AULE}))}{\partial d} \\ &= \sum_{j=1}^p \frac{1}{\lambda_j(\lambda_j + 1)^4} (2\phi^{-1}(\lambda_j + d)(\lambda_j + 2 - d)^2 - 2\phi^{-1}(\lambda_j + 2 - d)(\lambda_j + d)^2) \\ &\quad - \sum_{j=1}^p \frac{1}{\lambda_j(\lambda_j + 1)^4} (4(1 - d)^3 \lambda_j \beta_j^2) \\ &= \sum_{j=1}^p \frac{1}{\lambda_j(\lambda_j + 1)^4} (4\phi^{-1}(\lambda_j + d)(\lambda_j + 2 - d)(1 - d)) - \sum_{j=1}^p \frac{1}{\lambda_j(\lambda_j + 1)^4} \\ &\quad \times (4(1 - d)^3 \lambda_j \beta_j^2) \\ &= \sum_{j=1}^p \frac{1}{\lambda_j(\lambda_j + 1)^4} (\phi^{-1}(\lambda_j + d)(\lambda_j + 2 - d)) - \sum_{j=1}^p \frac{1}{\lambda_j(\lambda_j + 1)^4} \\ &\quad \times ((1 - d)^2 \lambda_j \beta_j^2) \\ &= \sum_{j=1}^p \frac{1}{\lambda_j(\lambda_j + 1)^4} (\phi^{-1}(\lambda_j + d)(\lambda_j + 2 - d) - (1 - d)^2 \lambda_j \beta_j^2) = 0 \quad (5.10) \end{aligned}$$

olarak elde edilir. Burada $g_j(d) = (\phi^{-1}(\lambda_j + d)(\lambda_j + 2 - d) - (1 - d)^2 \lambda_j \beta_j^2) = 0$ olarak yazılabilir. Dolayısıyla, $g_j(d)$ fonksiyonu, d değişkeninin karesel bir fonksiyonu olarak aşağıdaki gibi ifade edilebilir:

$$g_j(d) = d^2 \underbrace{(-\phi^{-1} - \lambda_j \beta_j^2)}_a + d \underbrace{(2\phi^{-1} + 2\lambda_j \beta_j^2)}_b + \underbrace{(\phi^{-1} \lambda_j^2 + 2\phi^{-1} \lambda_j - \lambda_j \beta_j^2)}_c = 0.$$

Bu fonksiyonun kökleri diskriminant yöntemi kullanılarak elde edilir. $g_j(d)$ 'nin diskriminantı

$$\begin{aligned}
\Delta_{g_j(d)} &= (2\phi^{-1} + 2\lambda_j\beta_j^2)^2 - 4(-\phi^{-1} - \lambda_j\beta_j^2)(\phi^{-1}\lambda_j^2 + 2\phi^{-1}\lambda_j - \lambda_j\beta_j^2) \\
&= 4\phi^{-1}\lambda_j\beta_j^2 + 8\phi^{-1}\lambda_j^2\beta_j^2 + 4\phi^{-1}\lambda_j^3\beta_j^2 + 4\phi^{-2} + 8\phi^{-2}\lambda_j + 4\phi^{-2}\lambda_j^2 \\
&= 4\phi^{-1}(\lambda_j\beta_j^2 + 2\lambda_j^2\beta_j^2 + \lambda_j^3\beta_j^2 + \phi^{-1} + 2\phi^{-1}\lambda_j + \phi^{-1}\lambda_j^2) \\
&= 4\phi^{-1}[\lambda_j\beta_j^2(\lambda_j + 1)^2 + \phi^{-1}(\lambda_j + 1)^2] \\
&= 4\phi^{-1}(\lambda_j + 1)^2(\phi^{-1} + \lambda_j\beta_j^2) > 0
\end{aligned}$$

olarak bulunur. Diskriminantın pozitif olması, karesel denklemin iki reel kökü olduğunu gösterir. Bu kökler, d_{1j} ve d_{2j} olarak adlandırılır ve elde edilen kökler şu şekildedir:

$$\begin{aligned}
d_{1j,2j} &= \frac{-2\phi^{-1} - 2\lambda_j\beta_j^2 \pm \sqrt{4\phi^{-1}(\lambda_j + 1)^2(\phi^{-1} + \lambda_j\beta_j^2)}}{-2\phi^{-1} - 2\lambda_j\beta_j^2} \\
&= 1 \pm (\lambda_j + 1) \frac{\sqrt{\phi^{-1}(\phi^{-1} + \lambda_j\beta_j^2)}}{\phi^{-1} + \lambda_j\beta_j^2} \\
&= 1 \pm (\lambda_j + 1) \sqrt{\frac{\phi^{-1}}{\phi^{-1} + \lambda_j\beta_j^2}}.
\end{aligned}$$

Bu denklemin çözümüyle elde edilen d değeri, AULE'nin minimum MSE'sini verecek optimal d değeridir. Simülasyon ve gerçek veri uygulamalarında AULE'nin d değeri bu köklerden yararlanarak aşağıdaki gibi seçilir. Algamal ve Asar (2020) 'deki makalesinde bu seçim $d_{optimal}$ olarak adlandırılmıştır.

$$d_{optimal} = 1 + (\lambda_j + 1) \sqrt{\frac{\phi^{-1}}{\phi^{-1} + \lambda_j\beta_j^2}} \quad (5.11)$$

$d_{optimal}$ 'in, bilinmeyen parametreleri β_j ve ϕ^{-1} 'e bağlı olduğundan, bunları yansız tahmin edicilerle değiştiririz ve AULE'nin optimal d değeri,

$$\hat{d}_{optimal} = 1 + (\lambda_j + 1) \sqrt{\frac{\hat{\phi}^{-1}}{\hat{\phi}^{-1} + \lambda_j\hat{\beta}_j^2}} \quad (5.12)$$

olarak bulunur. Ayrıca, Alkhamisi ve ark. (2006), Muniz ve ark. (2009) ve Asar ve

Genç (2017)'den ilham alarak elde ettiğimiz $\hat{d}_{optimal}$ değerini d1, d2, d3 ve d4 tahmin edicilerine entegre ederek aşağıda tanımlanan dört yeni tahmin ediciyi öneriyoruz:

$$d1 = \min \left(1 + (\lambda_j + 1) \sqrt{\frac{\hat{\phi}^{-1}}{\hat{\phi}^{-1} + \lambda_j \hat{\beta}_j^2}} \right),$$

$$d2 = \text{mean} \left(1 + (\lambda_j + 1) \sqrt{\frac{\hat{\phi}^{-1}}{\hat{\phi}^{-1} + \lambda_j \hat{\beta}_j^2}} \right),$$

$$d3 = \text{median} \left(1 + (\lambda_j + 1) \sqrt{\frac{\hat{\phi}^{-1}}{\hat{\phi}^{-1} + \lambda_j \hat{\beta}_j^2}} \right),$$

$$d4 = \text{harmmean} \left(1 + (\lambda_j + 1) \sqrt{\frac{\hat{\phi}^{-1}}{\hat{\phi}^{-1} + \lambda_j \hat{\beta}_j^2}} \right).$$

Burada d1, d2, d3, d4 sırasıyla minimum, ortalama, medyan ve harmonik ortalama değerlerini ifade eder. Bu tahmin ediciler, λ_j , ϕ ve β_j değerlerine bağlı olarak hesaplanır.

5.3. Tahmin Edicilerin Teorik Olarak Karşılaştırılması

Bu bölümde, beta regresyon modeli için önerdiğimiz AULE tahmin edicisinin, MLE ve LE tahmin edicilerine karşı teorik olarak karşılaştırıyoruz. Bu karşılaştırmalar yapılırken aşağıdaki lemmadan yararlanılacaktır.

Lemma 5.1 (Trenkler and Toutenburg, 1990). $\hat{\beta}_1$ ve $\hat{\beta}_2$, β 'nin iki tahmin edicisi olsun. Bu tahmin edicilerin varyans-kovaryans matrislerinin farkı $\mathbf{D} = \text{Cov}(\hat{\beta}_1) - \text{Cov}(\hat{\beta}_2)$ pozitif tanımlı bir matris olsun ve ayrıca $\mathbf{a}_1 = \text{Bias}(\hat{\beta}_1)$ ve $\mathbf{a}_2 = \text{Bias}(\hat{\beta}_2)$ olarak verilsin. O halde $\mathbf{a}_2'(\mathbf{D} + \mathbf{a}_1\mathbf{a}_1')^{-1}\mathbf{a}_2 < 1$ ancak ve ancak $\text{MMSE}(\hat{\beta}_1) - \text{MMSE}(\hat{\beta}_2)$ pozitif tanımlı bir matristir.

5.3.1. MLE'ye karşı AULE

Teorem 5.1 *BRM altında $\mathbf{b}_{AULE} = Bias(\hat{\beta}_{AULE})$ olarak kabul edilsin. O zaman $\forall j = 1, 2, \dots, p$ için $(1 - d)^2 (-\phi^{-1} - \lambda_j \beta_j^2) + 2\phi^{-1}(\lambda_j + 1)^2 > 0$ olursa $MSE(\hat{\beta}_{MLE}) - MSE(\hat{\beta}_{AULE}) > 0$ olur.*

İspat MLE ve AULE tahmin edicilerinin MSE fonksiyonları arasındaki fark, Denklem (4.10) ve (5.7) kullanılarak aşağıdaki gibi elde edilir.

$$\begin{aligned}
& MSE(\hat{\beta}_{MLE}) - MSE(\hat{\beta}_{AULE}) \\
&= \phi^{-1} \sum_{j=1}^p \frac{1}{\lambda_j} - \phi^{-1} \sum_{j=1}^p \frac{(\lambda_j + d)^2 (\lambda_j + 2 - d)^2}{\lambda_j (\lambda_j + 1)^4} - (1 - d)^4 \sum_{j=1}^p \frac{\beta_j^2}{(\lambda_j + 1)^4} \\
&= \sum_{j=1}^p \frac{\phi^{-1} [-(1 - d)^4 + 2(1 - d)^2 + (1 - d)^2 (2\lambda_j^2 + 4\lambda_j)]}{\lambda_j (\lambda_j + 1)^4} - \sum_{j=1}^p \frac{(1 - d)^4 \lambda_j \beta_j^2}{\lambda_j (\lambda_j + 1)^4} \\
&= \sum_{j=1}^p (1 - d)^2 \frac{(1 - d)^2 (-\phi^{-1} - \lambda_j \beta_j^2) + 2\phi^{-1} (\lambda_j + 1)^2}{\lambda_j (\lambda_j + 1)^4} \tag{5.13}
\end{aligned}$$

Burada $\forall j = 1, 2, \dots, p$ için $(1 - d)^2 (-\phi^{-1} - \lambda_j \beta_j^2) + 2\phi^{-1} (\lambda_j + 1)^2 > 0$ olursa beta regresyon modelinde $0 < d < 1$ için önerdiğimiz AULE tahmin edicisinin teorik MSE'sinin, MLE tahmin edicisinden üstün olduğu kanıtlanmıştır. ■

Teorem 5.2 *BRM altında $0 < d < 1$ iken $\forall j = 1, 2, \dots, p$ için $2\lambda_j^2 + 2\lambda_j + 1 + 2d - d^2 > 0$ olduğunda, AULE, MMSE kriterine göre MLE'den daha iyi performansa sahiptir ancak ve ancak*

$$\mathbf{b}'_{AULE} (\mathbf{H}_1)^{-1} \mathbf{b}_{AULE} < 1.$$

Burada $\mathbf{H}_1 = Cov(\hat{\beta}_{MLE}) - Cov(\hat{\beta}_{AULE})$ şeklindedir ve \mathbf{b}_{AULE} AULE'nin yan vektörünü ifade eder.

İspat MLE ve AULE tahmin edicilerinin kovaryansları arasındaki fark ele alınırsa,

$$\begin{aligned}
\mathbf{H}_1 &= Cov(\hat{\beta}_{MLE}) - Cov(\hat{\beta}_{AULE}) \\
&= \phi^{-1} \mathbf{Q} \text{diag} \left\{ \frac{1}{\lambda_j} - \frac{(\lambda_j + d)^2 (\lambda_j + 2 - d)^2}{\lambda_j (\lambda_j + 1)^4} \right\}_{j=1}^p \mathbf{Q}' \\
&= \phi^{-1} \mathbf{Q} \text{diag} \left\{ \frac{(\lambda_j + 1)^4 - (\lambda_j + d)^2 (\lambda_j + 2 - d)^2}{\lambda_j (\lambda_j + 1)^4} \right\}_{j=1}^p \mathbf{Q}' \\
&= \phi^{-1} \mathbf{Q} \text{diag} \left\{ \frac{(1 - d)^2 (2\lambda_j^2 + 2\lambda_j + 1 + 2d - d^2)}{\lambda_j (\lambda_j + 1)^4} \right\}_{j=1}^p \mathbf{Q}' \tag{5.14}
\end{aligned}$$

Burada $\forall j = 1, 2, \dots, p$ için $2\lambda_j^2 + 2\lambda_j + 1 + 2d - d^2 > 0$ olduğunda \mathbf{H}_1 pozitif tanımlı olur. Böylece Lemma 5.1 kullanılarak

$$\mathbf{b}'_{AULE} (\mathbf{H}_1)^{-1} \mathbf{b}_{AULE} < 1$$

elde edilir. Böylece ispat tamamlanmış olur. ■

5.3.2. LE'ye karşı AULE

Teorem 5.3 *BRM altında $\mathbf{b}_{AULE} = Bias(\widehat{\boldsymbol{\beta}}_{AULE})$ olarak kabul edilsin. O zaman $\forall j = 1, 2, \dots, p$ için $\phi^{-1}(\lambda_j + d)(d - 2\lambda_j - 3) + (1 - d)(\lambda_j + 2 - d)\lambda_j\boldsymbol{\beta}_j^2 > 0$ olursa $MSE(\widehat{\boldsymbol{\beta}}_{LE}) - MSE(\widehat{\boldsymbol{\beta}}_{AULE}) > 0$ olur.*

İspat LE ve AULE tahmin edicilerinin MSE fonksiyonları arasındaki fark, Denklem (4.25) ve (5.7) kullanılarak aşağıdaki gibi elde edilir.

$$\begin{aligned} & MSE(\widehat{\boldsymbol{\beta}}_{LE}) - MSE(\widehat{\boldsymbol{\beta}}_{AULE}) \\ &= \phi^{-1} \sum_{j=1}^p \frac{(\lambda_j + d)^2}{\lambda_j(\lambda_j + 1)^2} + (1 - d)^2 \sum_{j=1}^p \frac{\boldsymbol{\beta}_j^2}{(\lambda_j + 1)^2} \\ &\quad - \phi^{-1} \sum_{j=1}^p \frac{(\lambda_j + d)^2(\lambda_j + 2 - d)^2}{\lambda_j(\lambda_j + 1)^4} - (1 - d)^4 \sum_{j=1}^p \frac{\boldsymbol{\beta}_j^2}{(\lambda_j + 1)^4} \\ &= \sum_{j=1}^p \frac{\phi^{-1}(\lambda_j + d)^2[(\lambda_j + 1)^2 - (\lambda_j + 2 - d)^2] + (1 - d)^2\lambda_j\boldsymbol{\beta}_j^2[(\lambda_j + 1)^2 - (1 - d)^2]}{\lambda_j(\lambda_j + 1)^4} \\ &= \sum_{j=1}^p (1 - d)(\lambda_j + d) \frac{\phi^{-1}(\lambda_j + d)(d - 2\lambda_j - 3) + (1 - d)(\lambda_j + 2 - d)\lambda_j\boldsymbol{\beta}_j^2}{\lambda_j(\lambda_j + 1)^4} \end{aligned} \tag{5.15}$$

Burada $\forall j = 1, 2, \dots, p$ için $\phi^{-1}(\lambda_j + d)(d - 2\lambda_j - 3) + (1 - d)(\lambda_j + 2 - d)\lambda_j\boldsymbol{\beta}_j^2 > 0$ olursa beta regresyon modelinde $0 < d < 1$ için önerdiğimiz AULE tahmin edicisinin teorik MSE'sinin, LE tahmin edicisinden üstün olduğu kanıtlanmıştır. ■

Teorem 5.4 *BRM altında $0 < d < 1$ iken $\forall j = 1, 2, \dots, p$ için $(2\lambda_j - d + 3)(d - 1) > 0$ olduğunda, AULE, MMSE kriterine göre LE'den daha iyi performansa sahiptir ancak ve ancak*

$$\mathbf{b}'_{AULE} \left(\mathbf{H}_2 + \mathbf{b}_{LE}\mathbf{b}'_{LE} \right)^{-1} \mathbf{b}_{AULE} < 1.$$

Burada $\mathbf{H}_2 = Cov(\widehat{\boldsymbol{\beta}}_{LE}) - Cov(\widehat{\boldsymbol{\beta}}_{AULE})$ şeklindedir ve \mathbf{b}_{AULE} ile \mathbf{b}_{LE} sırasıyla AULE ile LE'nin yan vektörlerini ifade eder.

İspat LE ve AULE tahmin edicilerinin kovaryansları arasındaki fark ele alınırsa,

$$\begin{aligned}
\mathbf{H}_2 &= Cov(\hat{\beta}_{LE}) - Cov(\hat{\beta}_{AULE}) \\
&= \phi^{-1} \mathbf{Q} \text{diag} \left\{ \frac{(\lambda_j + d)^2}{\lambda_j(\lambda_j + 1)^2} - \frac{(\lambda_j + d)^2(\lambda_j + 2 - d)^2}{\lambda_j(\lambda_j + 1)^4} \right\}_{j=1}^p \mathbf{Q}' \\
&= \phi^{-1} \mathbf{Q} \text{diag} \left\{ \frac{(\lambda_j + d)^2 [(2\lambda_j - d + 3)(d - 1)]}{\lambda_j(\lambda_j + 1)^4} \right\}_{j=1}^p \mathbf{Q}' \quad (5.16)
\end{aligned}$$

elde edilir. Burada $\forall j = 1, 2, \dots, p$ için $(2\lambda_j - d + 3)(d - 1) > 0$ olduğunda \mathbf{H}_2 pozitif tanımlı olur. Böylece Lemma 5.1 kullanılarak

$$\mathbf{b}'_{AULE} \left(\mathbf{H}_2 + \mathbf{b}_{LE} \mathbf{b}'_{LE} \right)^{-1} \mathbf{b}_{AULE} < 1$$

elde edilir. Böylece ispat tamamlanmış olur. ■

Bir sonraki bölümde, önerilen tahmin edicilerin uygulanabilirliğini ve etkinliğini değerlendirmek amacıyla R programlama dilini kullanarak bir Monte Carlo simülasyon deneyi gerçekleştirilmiştir. Bu deney, önerilen tahmin edicilerin farklı senaryolar altında nasıl performans gösterdiğini ve diğer tahmin modelleriyle karşılaştırıldığında ne tür avantajlar sunduğunu ortaya koymaktadır.

6. MONTE CARLO SİMÜLASYON ÇALIŞMASI

Bu bölümde, önerilen yeni tahmin edicimizin performansını farklı senaryolar altında incelemek için bir Monte Carlo simülasyon deneyi gerçekleştirilmiştir. McDonald ve Galarneau (1975) çalışmasını takip ederek, açıklayıcı değişkenleri (bağımsız değişkenleri) elde etmek için aşağıdaki denklem kullanılmıştır:

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

Bu denklemde, w_{ij} standart normal dağılımdan üretilmiş rastgele sayılar olup, korelasyon gücünü değiştirmek ve çoklu bağlantılı bir yapı oluşturmak için kullanılmıştır. Burada p açıklayıcı değişken sayısını, ρ ise herhangi iki açıklayıcı değişken arasındaki korelasyon derecesini temsil etmektedir.

Beta regresyon analizinde, bağımlı değişkenin beta dağılımını taklit etmesi amacıyla $[0,1]$ aralığında beta dağılımlarına sahip uygun bir link fonksiyonu kullanılmaktadır. Bu nedenle, simülasyon tasarımında bağımlı değişken, aşağıdaki denklemle ifade edilen logit link fonksiyonu kullanılarak beta dağılımından türetilmiştir:

$$y_i = \log \left(\frac{\mu_i}{1 - \mu_i} \right) \\ = (\beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip})^{-1}, \quad i = 1, 2, \dots, n \quad \text{ve} \quad j = 1, 2, \dots, p$$

Burada y_i yanıt değişkenini (bağımlı değişkeni), $x_{i1}, x_{i2}, \dots, x_{ip}$ ise i . gözlem için açıklayıcı değişken sayısını (bağımsız değişkenleri) ve $\beta_1, \beta_2, \dots, \beta_p$ regresyon katsayılarını temsil eder. Regresyon analizinde, parametrelerin seçimi genellikle verilerle iyi uyum sağlamak ve modelin doğruluğunu artırmak amacıyla gerçekleştirilir. Ancak, çok sayıda bağımsız değişkenin (p) olduğu durumlarda, model karmaşık hale gelir ve aşırı uyum (overfitting) sorunu ortaya çıkabilir. Bu bağlamda, regresyon parametrelerinin $\sum_{j=1}^p \beta_j^2 = 1$ şartını sağlayacak şekilde seçilmesi, regresyon analizinde sıkça kullanılan bir kısıtlama yöntemidir (Kibria, 2003). Bu kısıtlama, çalışmada kullanılan parametrelerin etkin bir şekilde tahmin edilmesine ve ilişkilerin kapsamlı bir şekilde analiz edilmesine olanak tanımaktadır.

Farklı örneklem büyüklükleri, açıklayıcı değişken sayıları, korelasyon dereceleri ve dağılım parametreleri gibi çeşitli faktörlerin değiştirildiği Monte Carlo simülasyon deneyi sonuçları analiz edilmiştir. Bu simülasyonlar kapsamında, korelasyon dereceleri $\rho = 0.90, 0.95, 0.99$ olarak belirlenmiş; örneklem büyüklükleri $n = 50, 100, 200$ olarak seçilmiş; açıklayıcı değişken sayıları $p = 4, 8, 12$ olarak değişmiştir. Simülasyon çalışması için gereken açıklayıcı değişkenler önceden üretilmiş ve simülasyon süreci ilerledikçe sabit tutulmuştur. Ayrıca, dağılım parametreleri $\phi = 5, 10, 25$ olarak ayarlanmıştır.

Simülasyonda elde edilen verileri ρ , n , p ve ϕ değerlerinin farklı kombinasyonları için 1000 kez tekrar edilmiştir. Bu tekrarlama işlemi, elde ettiğimiz sonuçların daha güvenilir ve istatistiksel açıdan anlamlı olmasını sağlamıştır. Önerilen tahmin edicilerin performansını değerlendirmek için, değerlendirme kriteri olarak MSE değeri kullanılmıştır. MSE, tahmin edilen değerler ile gerçek değerler arasındaki farkların karesinin ortalamasını ölçen bir metriktir ve aşağıdaki formülle tanımlanmıştır:

$$MSE(\hat{\beta}) = \frac{\sum_{i=1}^R (\hat{\beta}_i - \beta)' (\hat{\beta}_i - \beta)}{R}.$$

Burada R , 1000 olarak ayarlanan simülasyonun toplam tekrar sayısını, β gerçek değerleri, $\hat{\beta}_i$ tahmin edilen değerleri ifade eder. $(\hat{\beta}_i - \beta)$ ise önerilen tahmin edici ve diğer tahmin edicilerin i 'inci tekrarda tahmin edilen değer ile gerçek değer arasındaki farkı temsil etmektedir. MSE, regresyon analizinde tahminlerin gerçek değerlere ne kadar yakın olduğunu ölçmek için kullanılan bir hata ölçüsüdür. Düşük MSE değerleri, modelin tahminlerin gerçek değerlere daha yakın olduğunu ve bu nedenle modelin iyi performans gösterdiğini gösterir. Yüksek MSE değerleri ise modelin tahminlerinin gerçek değerlere göre daha fazla sapma gösterdiğini ve bu nedenle modelin düşük performans gösterdiğini gösterir.

Tüm hesaplama çalışmaları, R programlama dili içindeki betareg R paketi yardımıyla gerçekleştirilmiştir (Cribari-Neto ve Zeileis, 2010). Farklı tahmin yöntemlerinin MSE değerleri Çizelge 6.3, Çizelge 6.4 ve Çizelge 6.5'te farklı dağılım parametreleri ($\phi = 5$ için Çizelge 6.3, $\phi = 10$ için Çizelge 6.4 ve $\phi = 25$ için Çizelge 6.5) altında gerçekleştirilen simülasyon sonuçları, aşağıda özetlenmiştir: Çizelge 6.3, Çizelge 6.4 ve Çizelge 6.5'te diğer tüm faktörler sabit tutulduğunda,

- Açıklayıcı değişken sayısı olan " p " değeri artıkça, önerdiğimiz tahmin edicilerin MSE değerleri artmıştır. Çünkü değişken sayısının artması, modelin karmaşıklığını artırabilir ve tahmin edicilerin performansını olumsuz etkileyebilir. Ama Çizelge 6.4'de $n = 200$ ve $\rho = 0.90$ olduğunda d_2 tahmin edicisinin MSE sonucunda önce artış sonra azalış gözlemlenmiştir.
- Gözlem sayısı olan " n " değeri artıkça, önerdiğimiz tahmin edicilerinin MSE değerlerinin azaldığı gözlenmiştir. Büyük bir n değeri, örneklem büyüklüğünü artırarak daha güvenilir ve istatistiksel olarak daha sağlam tahminlere olanak tanır. Ancak, bu olumlu eğilime istisna olarak, Çizelge 6.3'de d_2 tahmin edicisinin bağımsız değişken sayısı $p = 12$ olduğu durumda MSE değerinde bir artış gözlemlenmiştir. Bu özel durum, d_2 tahmin edicisinin bağımsız değişken sayısının artmasının, modelin karmaşıklığını ve tahmin edicilerin performansını olumsuz yönde etkileyebileceğini işaret etmektedir.
- Dağılım parametresi olan " ϕ " değeri artıkça, tahmin edilen değerlerin gerçek değerlerinden sapma olasılığı artar ve bu da tahmin hatalarının artmasına neden olabilir. Böyle durumlarda, MSE değerinin artması beklenir. Ancak, yaptığımız analizler ve önerdiğimiz tahmin edicilerin performans değerlendirmeleri göstermiştir ki, önerdiğimiz tüm tahmin edicilerde MSE değerlerinde genellikle bir azalma gözlenmiştir. Bu durum, modelin tahminlerinin gerçek değerlere daha yakın olduğunu ve modelin daha iyi performans sergilediğini gösterir.
- Korelasyon derecesi olan " ρ " değeri artıkça, önerdiğimiz tahmin edicilerin MSE değerlerinin arttığı gözlemlenmiştir. Bu durum, yüksek korelasyonda tahmin edicilerin performansının olumsuz etkilenebileceği ve doğru tahminlerin zorlaşabileceği anlamına gelir.
- Genel olarak d_1 , d_2 , d_3 , d_4 tahmin edicisi MLE tahmin edicisinden, d_3 ve d_4 tahmin edicisi LE tahmin edicisinden daha iyi sonuçlar vermiştir. Bu durum, önerdiğimiz tahmin edicilerin MLE ve LE tahmin edicisine göre daha güvenilir ve doğru tahminler yapabildiğini göstermektedir. Bu karşılaştırmayı özetlemek için aşağıdaki tablo oluşturulmuştur:

Bu tabloya ek olarak, Çizelge 6.3’de bağımsız değişken sayısı $p=12$ iken

Durum	Tahmin Ediciler	Üstünlük Durumu
Tüm durumlarda	d1, d2, d3, d4	MLE’den üstün
Tüm durumlarda	d3 ve d4	LE’den üstün

Çizelge 6.1. Tahmin Edicilerin Karşılaştırılması

korelasyon derecesi $\rho = 0.95$ olduğu durumda $n = 100$ ve $n = 200$ için ve korelasyon derecesi $\rho = 0.99$ olduğu durumda $n = 200$ için LE tahmin edicisi, d2 tahmin edicisine göre daha düşük MSE değerine sahiptir. Yani LE tahmin edicisi d2 tahmin edicisinden üstün gelmiştir. Diğer tüm durumlarda önerdiğimiz d2 tahmin edicisi LE’den üstün gelmiştir.

- Bu tablonun yanı sıra, farklı korelasyon dereceleri (ρ) ve dağılım parametreleri (ϕ) için yapılan karşılaştırmalar sonucunda bazı öne çıkan durumlar tespit edilmiştir:

Korelasyon derecesi $\rho = 0.90$ dağılım parametresi $\phi = 25$ olduğunda d1, d2, d3 ve d4 tahmin edicileri, korelasyon derecesi $\rho = 0.95$ dağılım parametresi $\phi = 25$ olduğunda d2, d3 ve d4 tahmin edicileri, korelasyon derecesi $\rho = 0.90$ dağılım parametresi $\phi = 10$ olduğunda d2, d3 ve d4 tahmin edicileri, korelasyon derecesi $\rho = 0.90$ dağılım parametresi $\phi = 5$ olduğunda ise d4 tahmin edicisinin RE tahmin edicisine göre üstün olduğu durumlar belirtilmiştir. Bu karşılaştırmayı özetlemek için aşağıdaki tablo oluşturulmuştur:

Durum	Korelasyon (ρ)	Dağılım Parametresi (ϕ)	Tahmin Ediciler	Üstünlük Durumu
1	0.90	25	d1, d2, d3, d4	RE’den üstün
2	0.95	25	d2, d3, d4	RE’den üstün
3	0.90	10	d2, d3, d4	RE’den üstün
4	0.90	5	d4	RE’den üstün

Çizelge 6.2. Önerdiğimiz Tahmin Edicilerin RE Tahmin Edicisinden Üstünlüğü

Bu durum, belirli parametre değerleri altında önerilen tahmin edicilerin, RE tahmin edicisine göre daha iyi performans gösterebileceğini göstermektedir.

Aşağıda yer alan Çizelge 6.3, Çizelge 6.4 ve Çizelge 6.5’te gösterilen sonuçlara dayanarak, her bir durumu daha iyi anlamak ve sonuçları görselleştirmek adına R programlama dili kullanılarak kutu grafikleri oluşturulmuştur. Bu

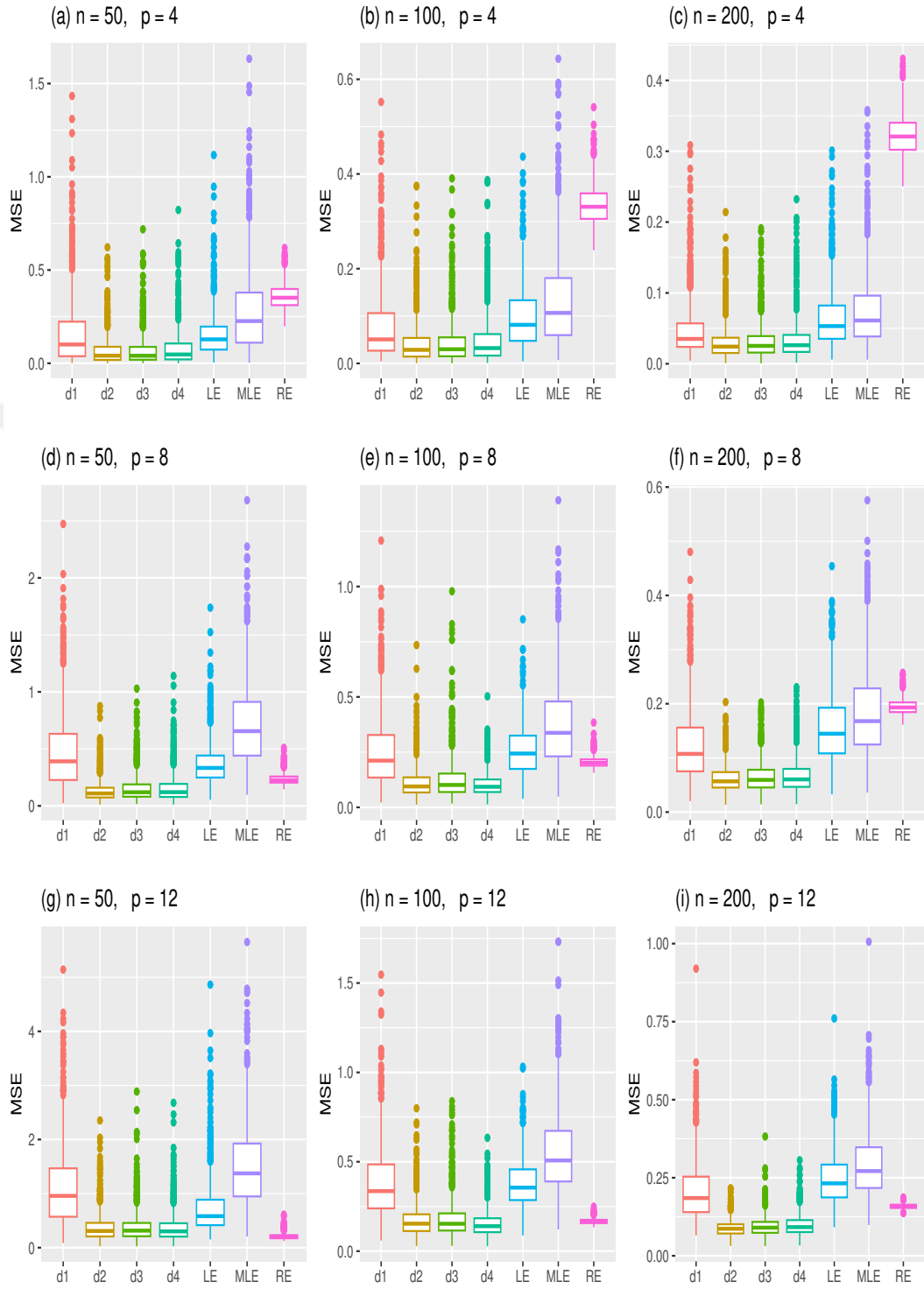
Çizelge 6.3. $\phi = 5$ için MSE Değerleri

p	4			8			12		
n	50	100	200	50	100	200	50	100	200
$\rho = 0.90$									
MLE	0.2772	0.1332	0.0746	0.5515	0.2443	0.1428	2.6143	1.1883	0.6179
RE	0.3592	0.3357	0.3228	0.3598	0.3348	0.3249	0.3575	0.3338	0.3247
LE	0.1527	0.0981	0.0646	0.2619	0.1421	0.1066	1.8451	0.6166	0.2758
d1	0.1620	0.0788	0.0484	0.3174	0.1372	0.0851	1.4779	0.6446	0.3402
d2	0.0704	0.0418	0.0300	0.0816	0.0390	0.0320	0.4830	0.5870	0.5333
d3	0.0704	0.0437	0.0318	0.0988	0.0453	0.0331	0.4035	0.1779	0.0999
d4	0.0834	0.0479	0.0337	0.1137	0.0544	0.0419	0.2610	0.0894	0.0540
$\rho = 0.95$									
MLE	0.7184	0.3702	0.1838	1.3927	0.6788	0.3290	6.5018	3.2342	1.4600
RE	0.2359	0.2049	0.1948	0.2382	0.2059	0.1975	0.2381	0.2094	0.1992
LE	0.3721	0.2577	0.1563	0.6216	0.3702	0.2415	4.6299	1.6495	0.5602
d1	0.4751	0.2518	0.1255	0.9179	0.4398	0.2129	4.1010	1.9941	0.8588
d2	0.1327	0.1108	0.0615	0.2025	0.1949	0.0842	1.7443	2.6364	1.8660
d3	0.1570	0.1297	0.0656	0.2451	0.1726	0.0909	0.9467	0.4342	0.2151
d4	0.1603	0.1055	0.0673	0.2391	0.1487	0.0887	0.6762	0.2844	0.1538
$\rho = 0.99$									
MLE	1.5052	0.5524	0.2919	2.9041	1.0260	0.5360	17.9796	4.8678	2.4259
RE	0.2108	0.1672	0.1588	0.2110	0.1702	0.1646	0.2028	0.1708	0.1692
LE	0.7526	0.3774	0.2470	1.4994	0.5333	0.3947	15.0303	2.4093	0.8824
d1	1.1066	0.3839	0.2059	2.0918	0.6846	0.3658	12.9828	3.1083	1.5275
d2	0.3750	0.1742	0.0887	0.5597	0.2910	0.1277	4.6934	2.0622	2.3690
d3	0.3811	0.1820	0.0945	0.5833	0.2399	0.1384	2.3924	0.5992	0.3418
d4	0.3738	0.1536	0.0994	0.5819	0.2165	0.1412	2.0531	0.4760	0.2864

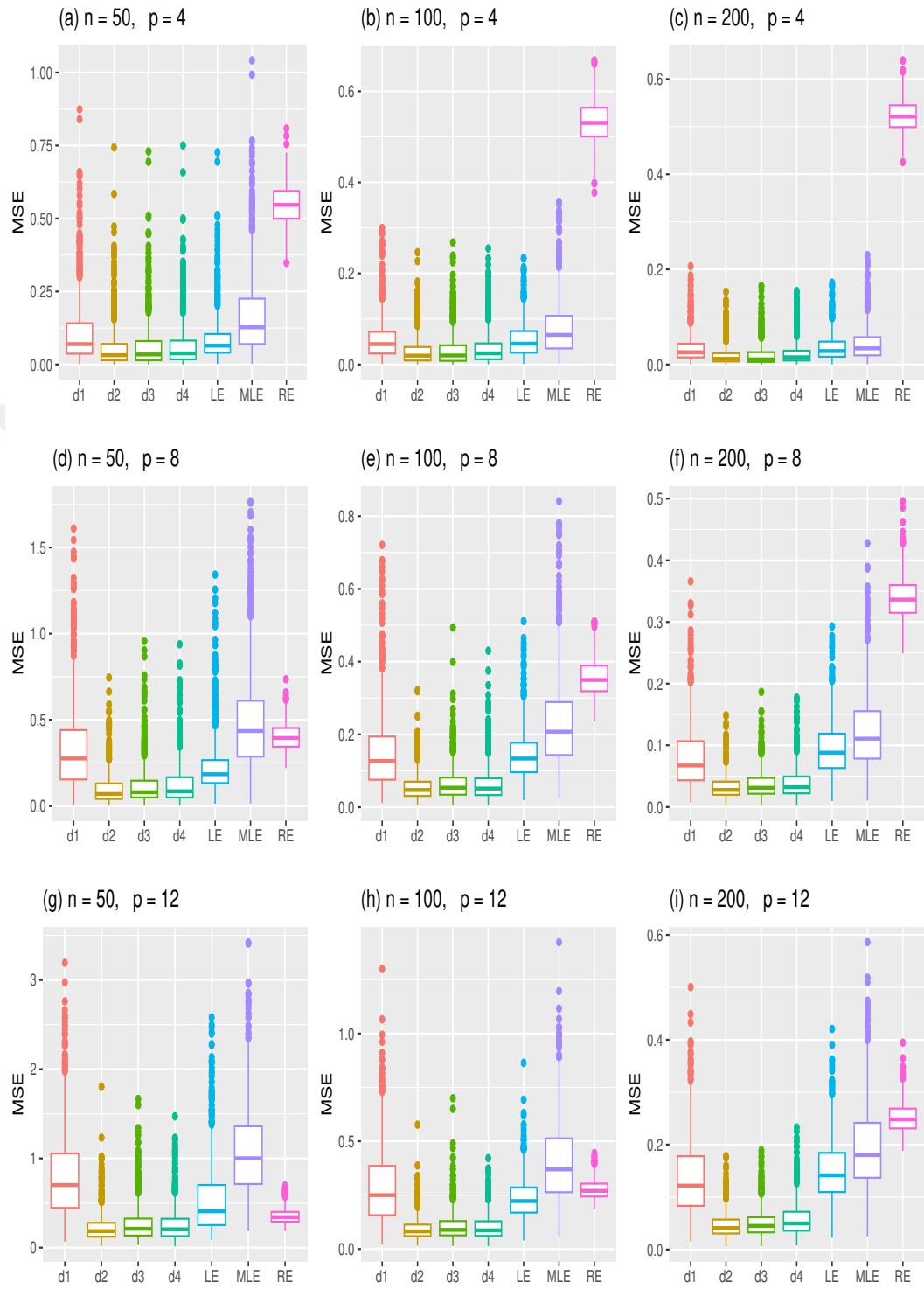
grafikler, MSE değerlerinin dağılımını göstererek, her bir durumdaki tahminciler arasındaki performans farklılıklarını daha açık bir şekilde ortaya koymaktadır. Grafikler, ρ değerlerine ve ϕ değerlerine göre ayrılmıştır.

Öncelikle, $\rho = 0.90$ durumu için oluşturulan kutu grafikleri incelendiğinde, $\phi = 5$, $\phi = 10$ ve $\phi = 25$ için MSE değerlerinin farklı p ve n kombinasyonlarında nasıl dağıldığı görülmektedir (Şekil 6.1, Şekil 6.2, Şekil 6.3). Ardından, $\rho = 0.95$ durumu için hazırlanan kutu grafiklerinde, aynı şekilde ϕ değerlerine göre MSE değerlerinin p ve n 'ye bağlı olarak nasıl dağıldığı gösterilmektedir (Şekil 6.4, Şekil 6.5, Şekil 6.6). Son olarak, $\rho = 0.99$ durumu için oluşturulan kutu grafiklerinde, farklı ϕ değerleri altında elde edilen MSE değerlerinin p ve n 'ye bağlı olarak nasıl değiştiği incelenmektedir (Şekil 6.7, Şekil 6.8, Şekil 6.9).

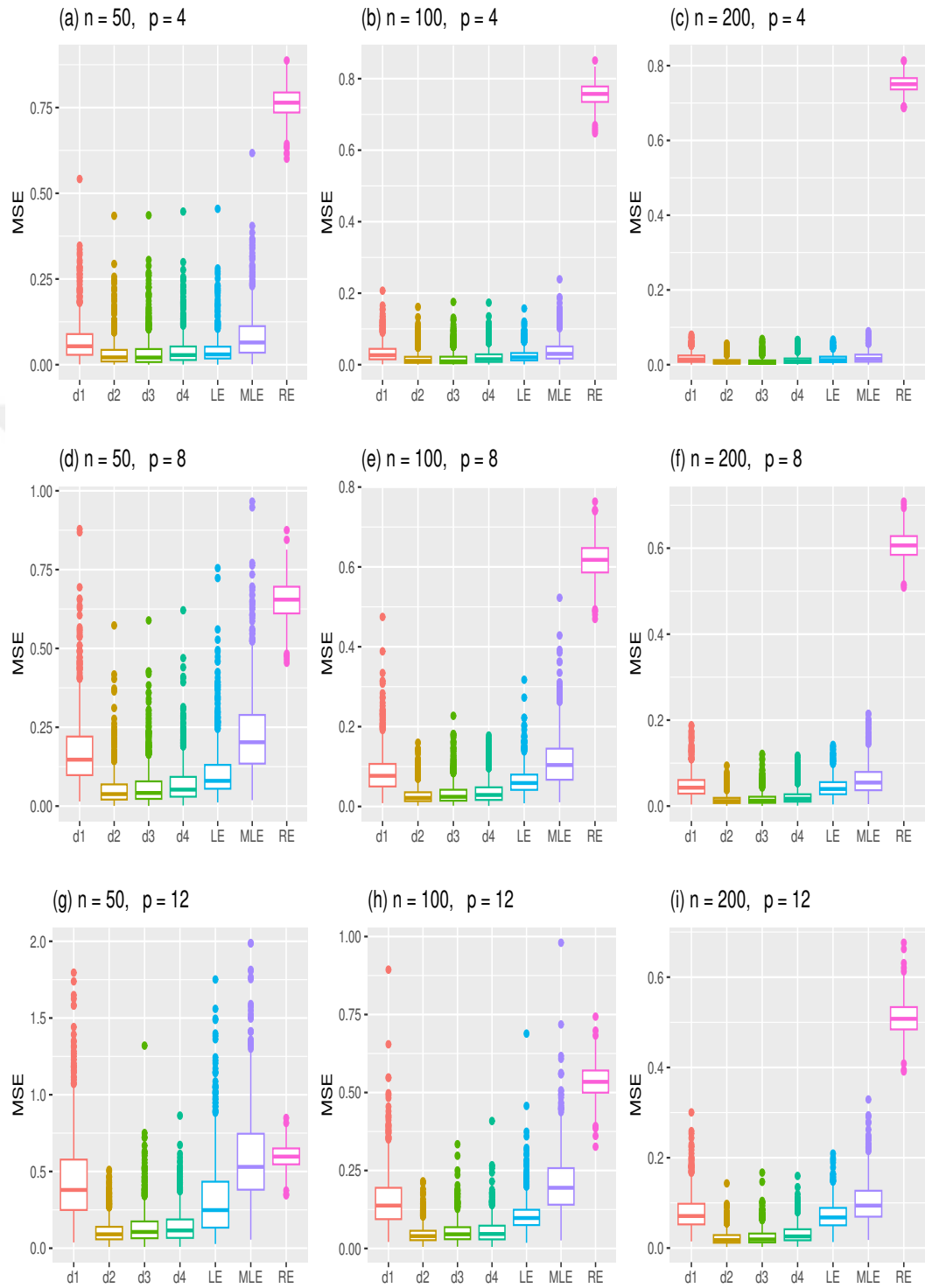
Bu şekilde hem tablolardan elde edilen nicel sonuçlar hem de görsel analiz bir araya getirilerek, Monte Carlo simülasyon sonuçlarının kapsamlı bir değerlendirilmesi yapılmıştır.



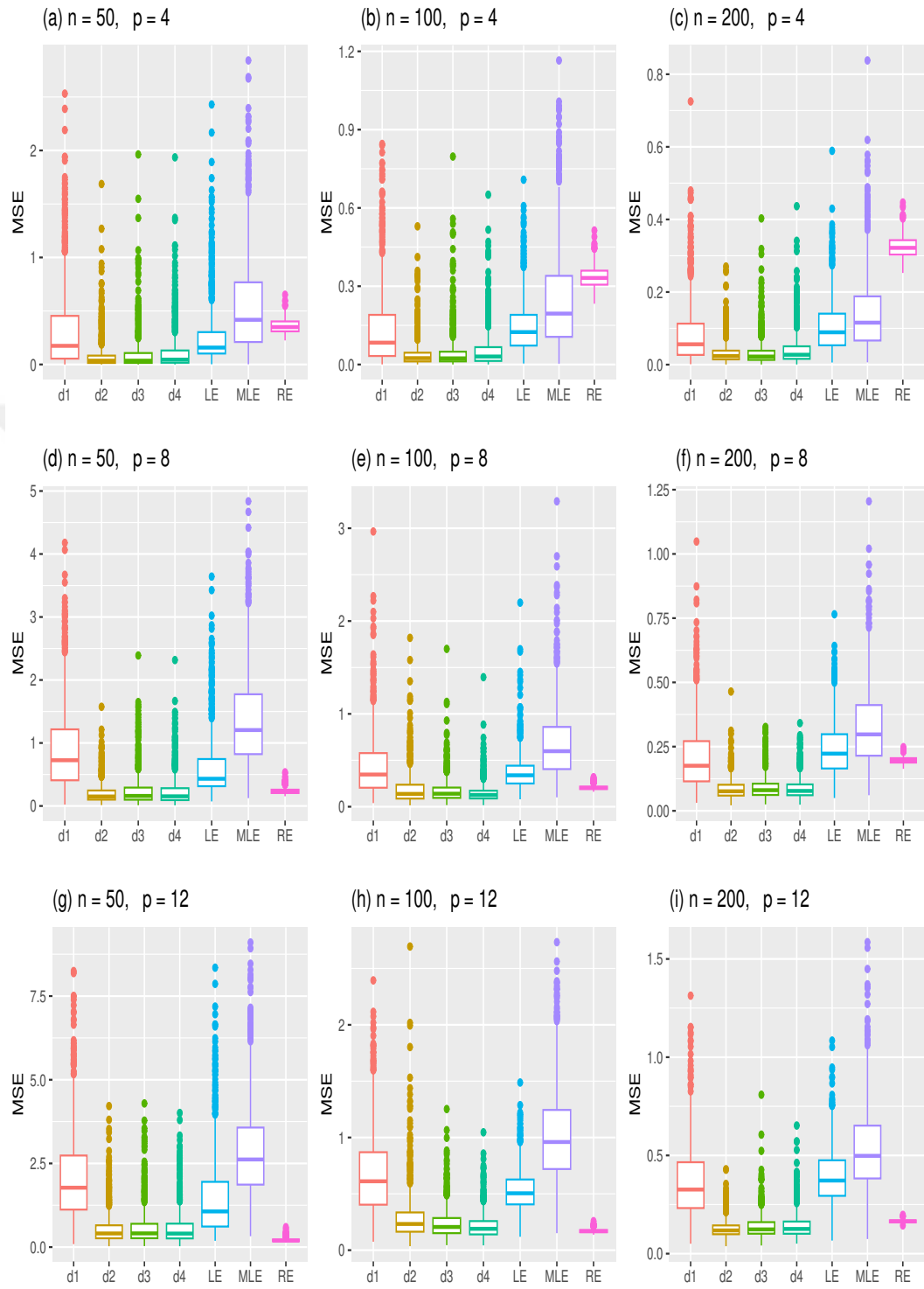
Şekil 6.1. $\rho = 0.90$ ve $\phi = 5$ için MSE değerlerinin kutu grafiği



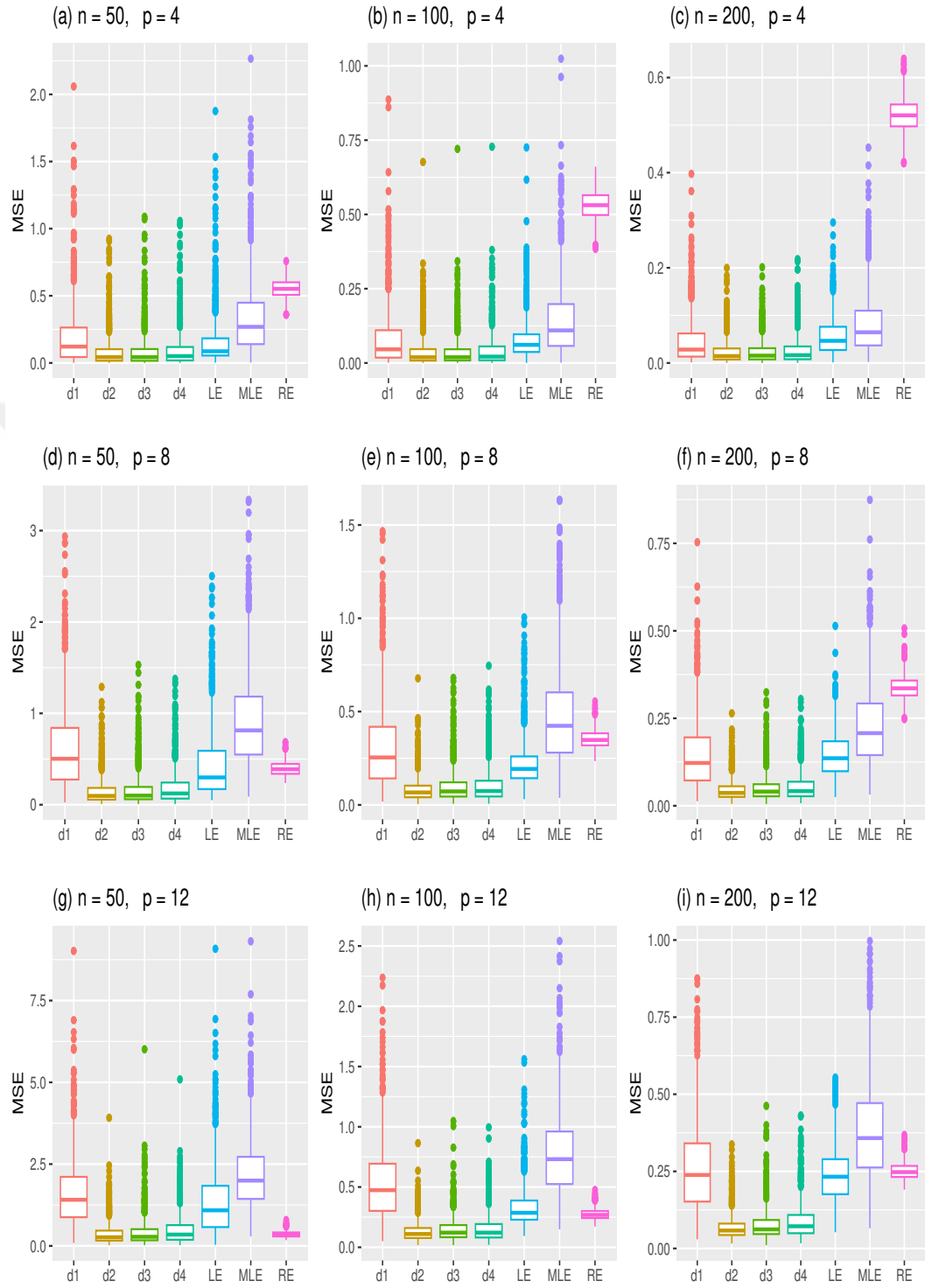
Şekil 6.2. $\rho = 0.90$ ve $\phi = 10$ için MSE değerlerinin kutu grafiği



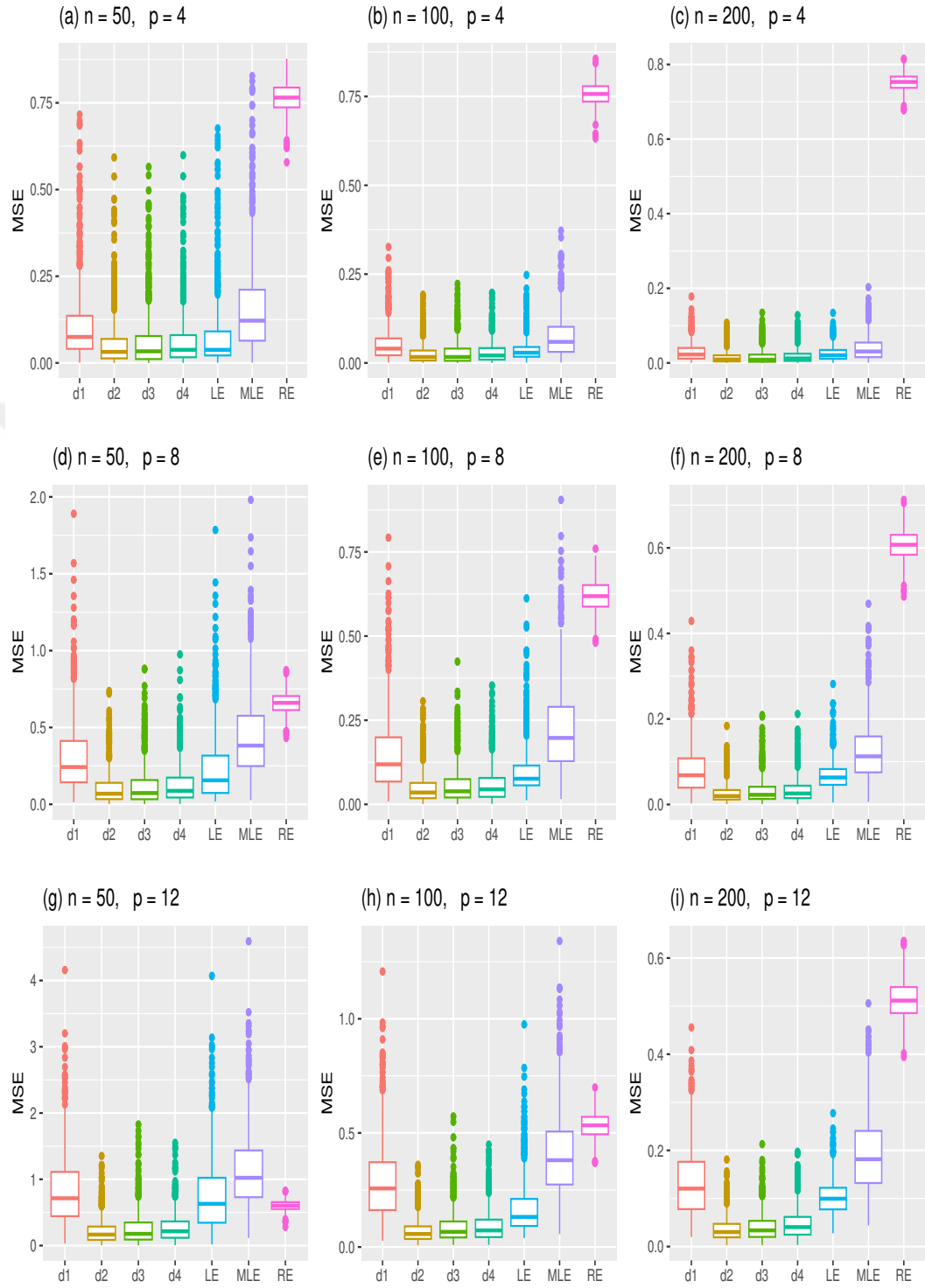
Şekil 6.3. $\rho = 0.90$ ve $\phi = 25$ için MSE değerlerinin kutu grafiği



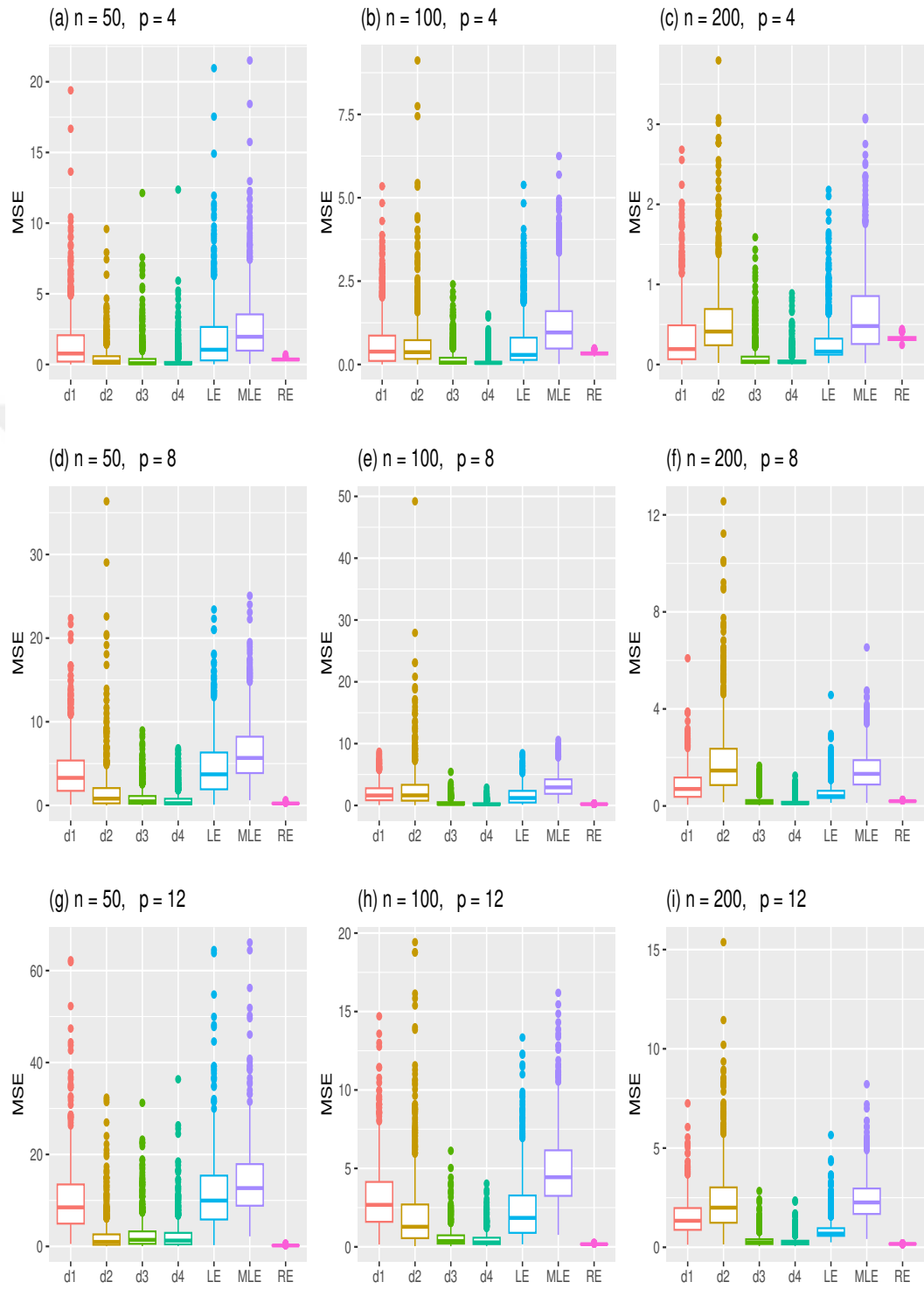
Şekil 6.4. $\rho = 0.95$ ve $\phi = 5$ için MSE değerlerinin kutu grafiği



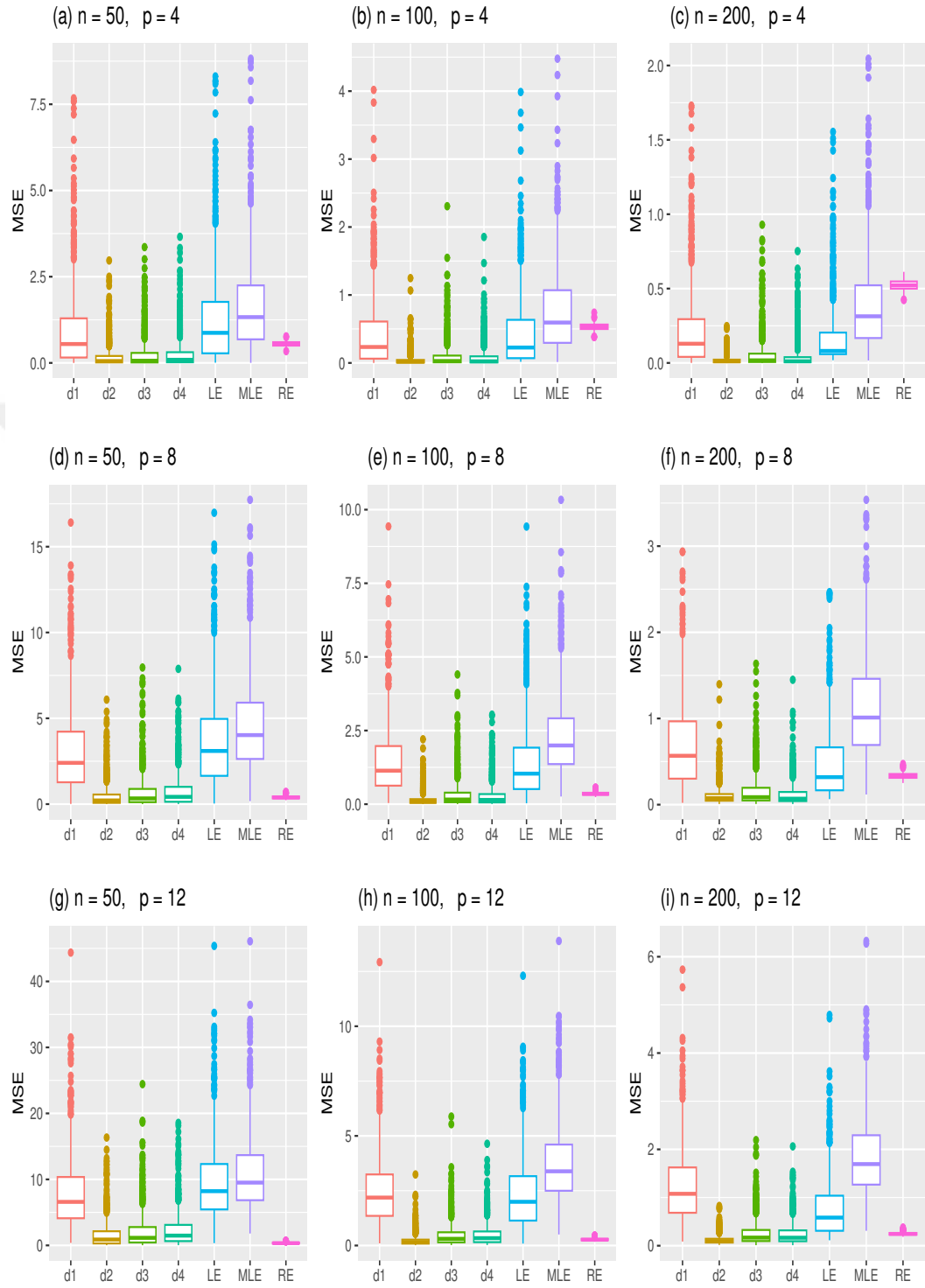
Şekil 6.5. $\rho = 0.95$ ve $\phi = 10$ için MSE değerlerinin kutu grafiği



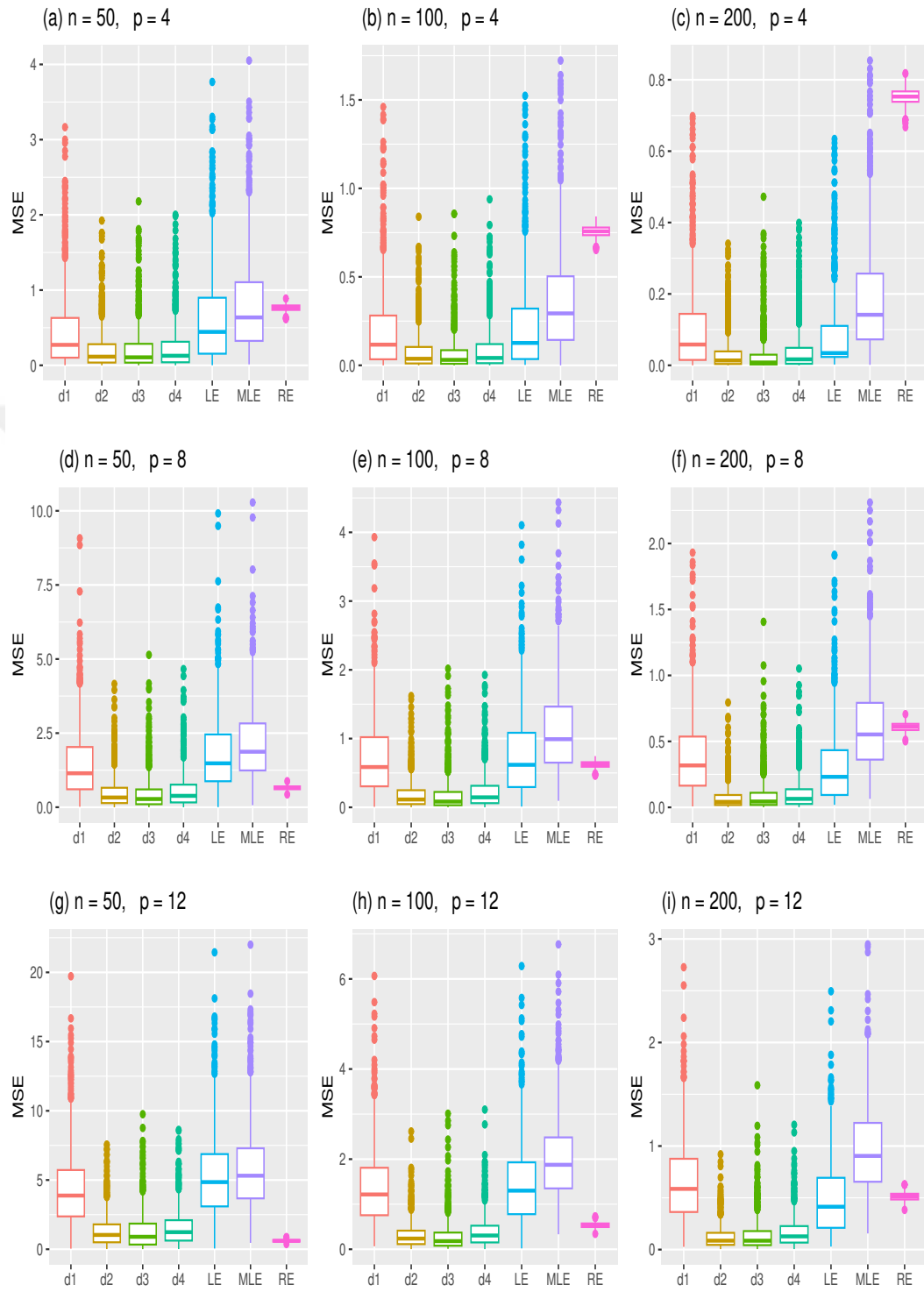
Şekil 6.6. $\rho = 0.95$ ve $\phi = 25$ için MSE değerlerinin kutu grafiği



Şekil 6.7. $\rho = 0.99$ ve $\phi = 5$ için MSE değerlerinin kutu grafiği



Şekil 6.8. $\rho = 0.99$ ve $\phi = 10$ için MSE değerlerinin kutu grafiği



Şekil 6.9. $\rho = 0.99$ ve $\phi = 25$ için MSE değerlerinin kutu grafiği

Çizelge 6.4. $\phi = 10$ için MSE Değerleri

p	4			8			12		
n	50	100	200	50	100	200	50	100	200
$\rho = 0.90$									
MLE	0.1707	0.0800	0.0438	0.3415	0.1463	0.0819	1.6609	0.7581	0.3897
RE	0.5479	0.5325	0.5221	0.5535	0.5322	0.5214	0.5520	0.5320	0.5235
LE	0.0881	0.0546	0.0360	0.1598	0.0776	0.0564	1.2429	0.4338	0.1743
d1	0.1104	0.0558	0.0337	0.1983	0.0821	0.0474	0.9314	0.4175	0.2133
d2	0.0564	0.0295	0.0185	0.0846	0.0369	0.0240	0.1821	0.0484	0.0195
d3	0.0634	0.0326	0.0194	0.0844	0.0366	0.0247	0.2462	0.1102	0.0599
d4	0.0656	0.0353	0.0224	0.0978	0.0428	0.0276	0.2628	0.0909	0.0422
$\rho = 0.95$									
MLE	0.4880	0.2282	0.1217	0.9273	0.4732	0.2311	4.5900	2.2792	1.1140
RE	0.4015	0.3552	0.3382	0.3968	0.3531	0.3384	0.3965	0.3564	0.3351
LE	0.2355	0.1437	0.0947	0.4411	0.2248	0.1465	3.6540	1.4060	0.4667
d1	0.3353	0.1511	0.0810	0.6187	0.3114	0.1479	3.0469	1.4655	0.6787
d2	0.0996	0.0562	0.0326	0.1496	0.0848	0.0452	0.4848	0.1568	0.1069
d3	0.1177	0.0653	0.0372	0.1671	0.0993	0.0507	0.7279	0.3341	0.1613
d4	0.1272	0.0643	0.0389	0.1930	0.1035	0.0552	0.7661	0.2899	0.1233
$\rho = 0.99$									
MLE	1.0868	0.4038	0.1948	2.2026	0.7775	0.3810	10.7398	3.7025	1.8456
RE	0.3525	0.2760	0.2515	0.3564	0.2747	0.2513	0.3544	0.2737	0.2484
LE	0.5452	0.2346	0.1511	1.3566	0.3381	0.2398	9.4732	2.3408	0.7748
d1	0.8051	0.2872	0.1369	1.6290	0.5368	0.2639	7.9085	2.4830	1.2319
d2	0.2275	0.0950	0.0469	0.3770	0.1292	0.0681	1.6834	0.2415	0.1215
d3	0.2624	0.1058	0.0511	0.4334	0.1505	0.0772	2.1843	0.4878	0.2658
d4	0.2600	0.1027	0.0581	0.4850	0.1530	0.0885	2.3885	0.4915	0.2481

Çizelge 6.5. $\phi = 25$ için MSE Değerleri

p	4			8			12		
n	50	100	200	50	100	200	50	100	200
$\rho = 0.90$									
MLE	0.0822	0.0389	0.0192	0.1588	0.0744	0.0387	0.8006	0.3639	0.1840
RE	0.7638	0.7559	0.7508	0.7640	0.7558	0.7521	0.7650	0.7568	0.7525
LE	0.0418	0.0251	0.0153	0.0776	0.0371	0.0249	0.6244	0.2208	0.0865
d1	0.0670	0.0339	0.0175	0.1085	0.0528	0.0287	0.4617	0.2034	0.1023
d2	0.0329	0.0167	0.0090	0.0560	0.0258	0.0148	0.2130	0.0768	0.0309
d3	0.0350	0.0167	0.0086	0.0620	0.0285	0.0158	0.2142	0.0685	0.0278
d4	0.0405	0.0217	0.0121	0.0640	0.0310	0.0179	0.2351	0.0883	0.0387
$\rho = 0.95$									
MLE	0.2299	0.1138	0.0610	0.4450	0.2239	0.1236	2.1611	1.1185	0.6114
RE	0.6530	0.6174	0.6053	0.6573	0.6197	0.6071	0.6564	0.6230	0.6092
LE	0.1076	0.0646	0.0436	0.2308	0.0993	0.0677	1.7947	0.7567	0.3096
d1	0.1744	0.0864	0.0479	0.3102	0.1491	0.0814	1.4604	0.7235	0.3877
d2	0.0535	0.0280	0.0154	0.1016	0.0481	0.0264	0.4902	0.1843	0.0741
d3	0.0634	0.0331	0.0175	0.1183	0.0569	0.0315	0.4761	0.1773	0.0910
d4	0.0720	0.0368	0.0216	0.1267	0.0596	0.0335	0.5662	0.2292	0.1048
$\rho = 0.99$									
MLE	0.5916	0.2075	0.1018	1.1192	0.4096	0.1930	5.8849	2.0182	0.9716
RE	0.5995	0.5350	0.5094	0.6026	0.5330	0.5136	0.6082	0.5331	0.5120
LE	0.3231	0.1054	0.0719	0.7443	0.1722	0.1033	5.3898	1.4560	0.4954
d1	0.4473	0.1541	0.0785	0.8234	0.2879	0.1338	4.4136	1.3840	0.6600
d2	0.1105	0.0454	0.0224	0.2175	0.0724	0.0368	1.3668	0.3167	0.1215
d3	0.1416	0.0543	0.0248	0.2565	0.0881	0.0416	1.3264	0.2936	0.1422
d4	0.1435	0.0557	0.0316	0.2757	0.0918	0.0476	1.6068	0.3963	0.1720

7. GERÇEK VERİ UYGULAMASI

Bu bölümde, önerdiğimiz yeni tahmin edicilerin pratikte ne kadar etkili ve uygulanabilir olduğunu daha detaylı bir şekilde değerlendirmek amacıyla, orta şiddetli ve çok yüksek şiddetli çoklu bağlantı problemlerine sahip iki farklı gerçek veri seti üzerinde uygulama gerçekleştirilmektedir. Her bir bağımsız değişken için tahmin edicilerin parametre tahmin değerleri, Çizelge 7.1 ve Çizelge 7.2’de verilmiştir. Ayrıca MLE, RE, LE tahmin edicileri ve AULE tahmin edicisine entegre ederek önerdiğimiz d_1 , d_2 , d_3 , d_4 tahmin edicilerinin tahmini MSE değerleri sırasıyla Denklemler (4.10), (3.10), (3.21) ve (5.7) kullanılarak hesaplanmıştır. Ayrıca, $0 < d < 1$ aralığı ile sınırlı olan farklı d değerleri için MLE, LE ve AULE tahmin edicilerinin MSE sonuçları Şekil 7.1 ve Şekil 7.2’de verilmiştir. Bu tablolar ve grafikler, önerilen tahmin edicilerin çoklu bağlantı problemleriyle başa çıkma yeteneklerini ayrıntılı bir şekilde analiz etmemize olanak tanımaktadır.

7.1. Gini Verisi

Dünyadaki 63 ülke için derlenen 2012 yılına ait gini endeksi verisi ilk olarak Zaman ve ark. (2019) tarafından beta regresyon analizinde kullanılmıştır. Bu ülkeler aşağıdaki gibidir:

Arjantin, Ermenistan, Avusturya, Beyaz Rusya, Belçika, Bolivya, Brezilya, Bulgaristan, Burkina Faso, Kamerun, Kolombiya, Kosta Rika, Hırvatistan, Kıbrıs, Çek Cumhuriyeti, Danimarka, Ekvador, El Salvador, Estonya, Finlandiya, Fransa, Gürcistan, Almanya, Yunanistan, Guatemala, Honduras, Macaristan, İzlanda, İrlanda, İtalya, Kırgızistan Cumhuriyeti, Letonya, Litvanya, Lüksemburg, Makedonya, Malta, Moritanya, Meksika, Moldova, Moğolistan, Karadağ, Hollanda, Nikaragua, Nijer, Norveç, Panama, Paraguay, Peru, Polonya, Portekiz, Romanya, Sırbistan, Slovakya, Slovenya, İspanya, İsveç, İsviçre, Tacikistan, Türkiye, Ukrayna, Birleşik Krallık, Uruguay, Vietnam.

Gini katsayısı 0 ile 1 arasında bir değer alır. 0, tam eşitlik durumunu

(herkesin eşit gelire sahip olduğu durumu) ifade ederken, 1, tam eşitsizlik durumunu (tek bir kişinin tüm gelire sahip olduğu durumu) ifade eder. Dolayısıyla, gini katsayısının değeri ne kadar yüksekse, gelir dağılımındaki eşitsizlik o kadar büyük demektir. Veride, gini katsayısının en yüksek olduğu ülkenin Honduras, en düşük olduğu ülkenin ise Norveç olduğunu göstermektedir.

Gini endeksi verisi, gini katsayısı (y) bağımlı değişkeni ve 7 açıklayıcı değişken ile ülkeler arasındaki gelir dağılımındaki eşitsizliği karşılaştırmamıza olanak tanır. Bu açıklayıcı değişkenler aşağıdaki gibidir:

- Hükümetin Eğitim Harcamaları
- Elektriğe Erişim
- İşsizlik
- Doğrudan Yabancı Yatırım
- Enflasyon
- Gayri Safi Milli Hasıla
- Vergi Geliri

Burada belirtilen değişkenler, tahmin edicinin performansını değerlendirmek ve gerçek veri setindeki sonuçları analiz etmek için kullanılacak olan girdi verileridir. Veriye beta regresyon modeli uygulanmış ve bu analizde logit bağlantı fonksiyonu tercih edilmiştir.

Veri setindeki örneklem büyüklüğü $n = 63$ olarak belirlenmiştir. $X'WX$ matrisinin özdeğerleri şu şekildedir: $\lambda_1 = 14.6888011$, $\lambda_2 = 0.4543023$, $\lambda_3 = 0.2814306$, $\lambda_4 = 0.2573243$, $\lambda_5 = 0.2093808$, $\lambda_6 = 0.1705625$, $\lambda_7 = 0.1649526$ ve $\lambda_8 = 0.1131911$. Bu özdeğerler kullanılarak hesaplanan koşul sayısı, $\kappa = \sqrt{\frac{\lambda_{max}}{\lambda_{min}}} = 11.39166$ olarak bulunmuştur. Elde edilen bu κ değeri, orta şiddetli çoklu bağlantı probleminin var olduğunu göstermektedir (Stewart, 1973).

Veriler The World Bank Databank internet sitesinden alınmıştır; bkz. (Zaman ve ark., 2019). Bu veriyi kullanarak, Bölüm 5.2'de AULE tahmin edicisine entegre ederek önerilen d yanlılık parametreleri d_1 , d_2 , d_3 ve d_4 'ün performansı Çizelge 7.1'de verilmiştir. Ayrıca, önerdiğimiz AULE tahmin edicisinin farklı d değerlerine göre LE ve MLE tahmin edicilerine karşı

üstünlüğünü göstermek için Şekil 7.1'deki grafik çizilmiştir. Tüm hesaplamalar ve grafikler R programlama dili kullanılarak gerçekleştirilmiştir.

Çizelge 7.1. MLE, RE, LE ve önerilen d1, d2, d3, d4 tahmin edicileri için: Gini verisinin beta regresyon analizi parametre tahminleri ve MSE değerleri

	MLE	RE	LE	d1	d2	d3	d4
(Intercept)	-0.6159	-0.2039	-0.6103	-0.6150	-0.6108	-0.6092	-0.6113
Hükümetin Eğitim Harcamaları	-0.1358	-0.0027	-0.1257	-0.1296	-0.1013	-0.0908	-0.1047
Elektriğe Erişim	-1.0023	-0.0096	-0.9052	-0.9253	-0.5749	-0.4445	-0.6169
İşsizlik	0.1181	-0.0034	0.0957	0.0917	-0.0288	-0.0737	-0.0144
Doğrudan Yabancı Yatırım	-0.0380	-0.0020	-0.0383	-0.0415	-0.0573	-0.0632	-0.0554
Enflasyon	0.2802	0.0044	0.2572	0.2651	0.1964	0.1709	0.2047
Gayri Safi Milli Hasıla	0.3917	0.0057	0.3588	0.3696	0.2693	0.2320	0.2813
Vergi Geliri	-0.8152	-0.0069	-0.7343	-0.7492	-0.4494	-0.3378	-0.4854
MSE	0.9126	2.0675	0.7501	0.7686	0.5973	0.7361	0.5760

Çizelge 7.1'deki tabloda intercept değerleri, regresyon denklemindeki bağımsız değişkenlerin değerleri sıfır olduğunda, bağımlı değişkenin beklenen değerini temsil eden sabit terimdir. Aynı tabloda yer alan MSE değerleri ise, bir regresyon modelinin tahminlerinin gerçek değerlerden ne kadar sapma gösterdiğini ölçen bir metriktir. Düşük MSE değerleri daha iyi bir model performansını göstermektedir. Buna göre Çizelge 7.1 ve Şekil 7.1'den elde edilen sonuçlar şu şekildedir:

- Çizelge 7.1 incelendiğinde MSE sonuçlarında en yüksek değer RE tahmin edicisine, en düşük değer önerdiğimiz d4 tahmin edicisine ait olduğu görülmektedir. Bu durum, d4 tahmin edicisinin modelin gerçek değerlere daha yakın tahminler yapma eğiliminde olduğunu göstermektedir.
- Çizelge 7.1 incelendiğinde önerdiğimiz tüm tahmin edicilerin (d1, d2, d3, d4), MLE ve RE tahmin edicilerine kıyasla daha düşük MSE değerine sahip olduğu görülmektedir. Bu durum, önerdiğimiz tüm tahmin edicilerin orta şiddetli çoklu bağlantı varlığında MLE ve RE tahmin edicilerine göre verilere daha iyi uyum sağladığı ve daha doğru tahminler yaptığını göstermektedir.
- Çizelge 7.1 incelendiğinde önerdiğimiz d2, d3, d4 tahmin edicileri, LE tahmin edicisine göre daha düşük MSE değerlerine sahiptir. Bu durum, orta şiddetli çoklu bağlantı varlığında önerdiğimiz d2, d3, d4 tahmin edicilerinin LE tahmin edicisinden üstün geldiğini göstermektedir.

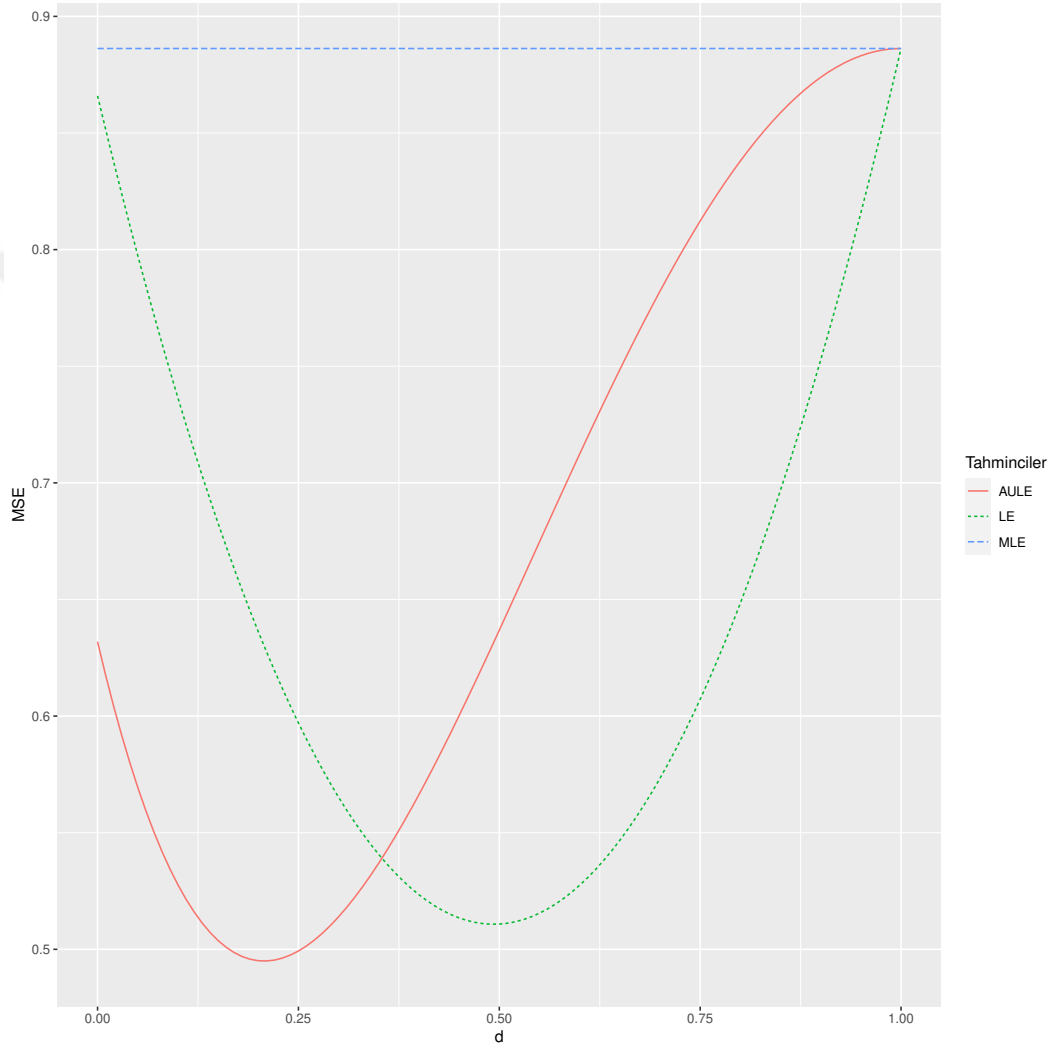
- Çizelge 7.1'deki MLE, RE, LE ve önerdiğimiz d1, d2, d3 ve d4 tahmin edicilerinin açıklayıcı değişken katsayılarının işaretleri incelendiğinde, işsizlik değişkeni hariç diğer tüm değişkenlerin her birinin işaretinin aynı olduğu görülmektedir. Yani bir açıklayıcı değişkenin katsayısı pozitifse, diğerleri de pozitif; eğer bir açıklayıcı değişkenin katsayısı negatifse, diğerleri de negatif yönde değişmektedir. İşsizlik değişkeninde ise MLE, LE ve d1 tahmin edicileri pozitif katsayılı yani işsizlik artıkça gini katsayısı artma eğilimindeyken, RE, d2, d3 ve d4 tahmin edicileri negatif katsayılı yani işsizlik artıkça gini katsayısı azaltma eğilimindedir. Dolayısıyla, işsizlik değişkeni üzerinden yapılan bu analiz ile önerdiğimiz d2, d3 ve d4 tahmin edicilerinin MLE ve LE tahmin edicilerine göre daha iyi performans gösterdiği görülmektedir.
- Çizelge 7.1'deki intercept değerleri incelendiğinde, MLE, LE, d1, d2, d3 ve d4 tahmin edicilerinde intercept değerleri birbirine oldukça yakınken, RE tahmin edicisinin farklı bir değere sahip olduğu görülmektedir.
- Şekil 7.1'deki $0 < d < 1$ aralığı ile sınırlı olan farklı d değerlerine göre MLE, LE ve AULE tahmin edicilerinin MSE sonuçları incelendiğinde, önerdiğimiz AULE tahmin edicisinin yaklaşık olarak (0,15 - 0,25) aralığında en düşük MSE değerine sahip olduğu gözlemlenmiştir.

Sonuç olarak, önerdiğimiz d1, d2, d3, ve d4 tahmin edicileri, modelin genel performansını artırarak diğer tahmin edicilere göre avantaj sağlamaktadır.

7.2. Yaşam Endeksi Verisi

Türkiyedeki 81 ilin 2015 yılına ait yaşam endeksi gösterge değerleri verisi ilk olarak Ünlü ve Aktaş (2017) tarafından daha sonra Abonazel ve Taha (2021) tarafından beta regresyon analizinde kullanılmıştır. Yaşam endeksi değeri 0 ile 1 arasında olduğundan, 1'e yakın değerler daha iyi bir yaşam standardını ifade etmektedir. Verilerde yaşam endeksinin en yüksek olduğu ilin Sinop, en düşük olduğu ilin ise Tunceli olduğunu göstermektedir.

Yaşam endeksi verisi, yaşam memnuniyeti endeksi olarak ölçülen mutluluk düzeyi (y) bağımlı değişkeni ve çeşitli kültürel, sosyal ve ekonomik değişkenleri



Şekil 7.1. Gini verisi için, d parametresine bağlı tahmin edicilerin MSE değerlerindeki değişimin grafiği

temel alan 40 açıklayıcı değişken ile iller arasındaki refahı karşılaştırmamıza olanak tanır. Bu açıklayıcı değişkenler Çizelge 7.2' deki tabloda belirtilmiştir. Burada belirtilen değişkenler, tahmin edicinin performansını değerlendirmek ve gerçek veri setindeki sonuçları analiz etmek için kullanılacak olan girdi verileridir. Veriler Türkiye İstatistik Kurumu'ndan alınmıştır; bkz. (Ünlü ve Aktaş, 2017). Yine Ünlü ve Aktaş (2017) tarafından yapılan analizler doğrultusunda, beta regresyon modelinin verilere en iyi uyumu sağladığı sonucuna ulaşılmıştır. Dolayısıyla, veriye beta regresyon modeli uygulanmış ve bu analizde logit bağlantı fonksiyonu tercih edilmiştir.

Veri setindeki örneklem büyüklüğü $n = 81$ olarak belirlenmiştir. $X'WX$ matrisinin özdeğerleri şu şekildedir: $\lambda_1 = 18.951752335$, $\lambda_2 = 3.126234862$, $\lambda_3 = 1.788048651$, $\lambda_4 = 0.609585185$, $\lambda_5 = 0.420495829$, $\lambda_6 = 0.389115259$, $\lambda_7 = 0.352512384$, $\lambda_8 = 0.270249927$, $\lambda_9 = 0.250526533$, $\lambda_{10} = 0.227153982$, $\lambda_{11} = 0.198523262$, $\lambda_{12} = 0.192025467$, $\lambda_{13} = 0.162053285$, $\lambda_{14} = 0.141479903$, $\lambda_{15} = 0.137757882$, $\lambda_{16} = 0.134563170$, $\lambda_{17} = 0.113431762$, $\lambda_{18} = 0.094019641$, $\lambda_{19} = 0.085131093$, $\lambda_{20} = 0.082136335$, $\lambda_{21} = 0.072276008$, $\lambda_{22} = 0.061443355$, $\lambda_{23} = 0.058501642$, $\lambda_{24} = 0.052933783$, $\lambda_{25} = 0.046412726$, $\lambda_{26} = 0.043512764$, $\lambda_{27} = 0.037162672$, $\lambda_{28} = 0.032486602$, $\lambda_{29} = 0.030788259$, $\lambda_{30} = 0.025925572$, $\lambda_{31} = 0.022833278$, $\lambda_{32} = 0.022160926$, $\lambda_{33} = 0.017868218$, $\lambda_{34} = 0.016183101$, $\lambda_{35} = 0.013826934$, $\lambda_{36} = 0.012023834$, $\lambda_{37} = 0.008671777$, $\lambda_{38} = 0.007361710$, $\lambda_{39} = 0.006430204$, $\lambda_{40} = 0.006027418$ ve $\lambda_{41} = 0.003372699$. Bu özdeğerler kullanılarak hesaplanan koşul sayısı, $\kappa = \sqrt{\frac{\lambda_{max}}{\lambda_{min}}} = 74.96109$ olarak bulunmuştur. Elde edilen bu κ değeri, çok yüksek çoklu bağlantı probleminin var olduğunu göstermektedir (Stewart, 1973).

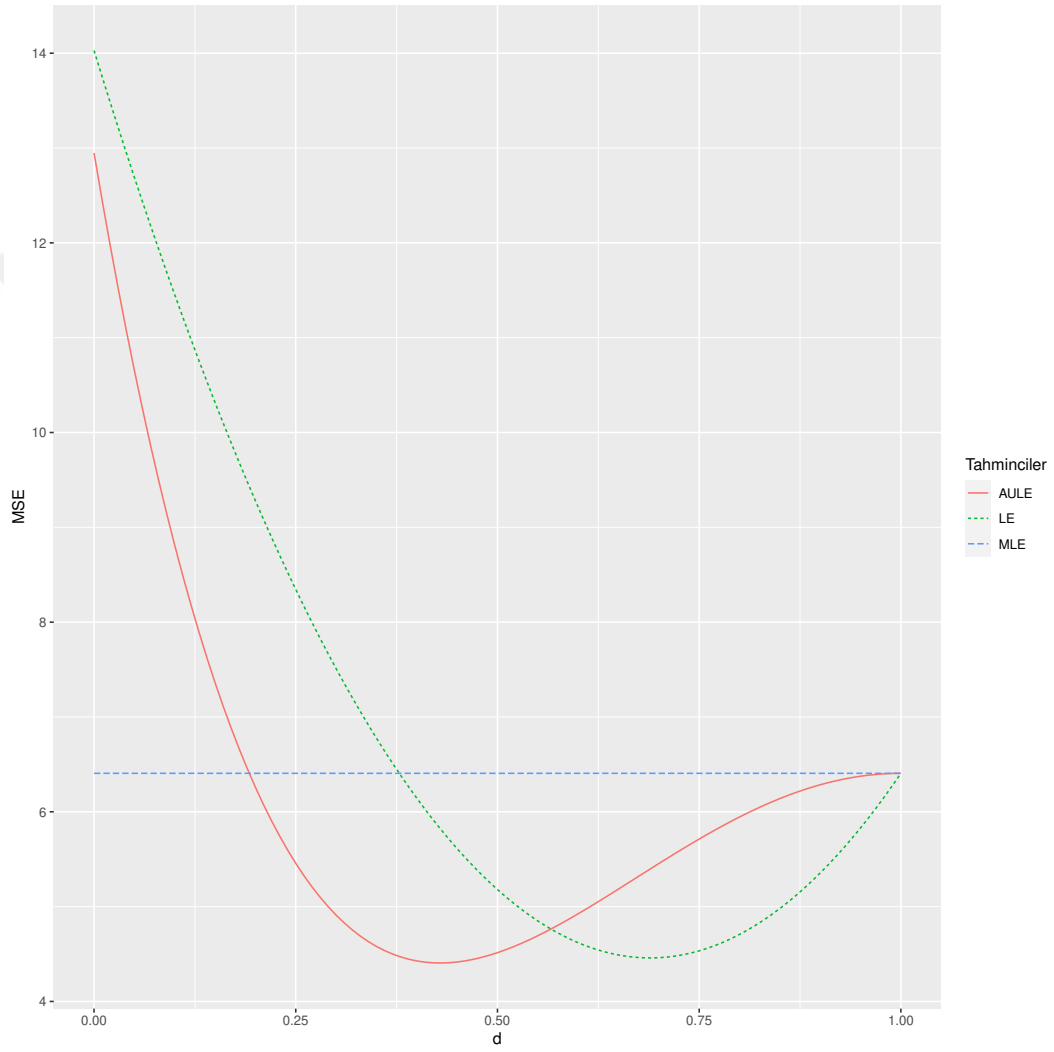
Bu veriyi kullanarak, Bölüm 5.2'de AULE'ye entegre ederek önerilen d yanlılık parametreleri $d1$, $d2$, $d3$ ve $d4$ 'ün performansı Çizelge 7.2'de verilmiştir. Ayrıca, önerdiğimiz AULE tahmin edicisinin farklı d değerlerine göre LE ve MLE tahmin edicilerine karşı üstünlüğünü göstermek için Şekil 7.2'deki grafik çizilmiştir. Tüm hesaplamalar ve grafikler R programlama dili kullanılarak gerçekleştirilmiştir.

Çizelge 7.2'deki tablodaki Intercept değerleri, regresyon denkleminde bağımsız değişkenlerin değerleri sıfır olduğunda, bağımlı değişkenin beklenen değerini temsil eden sabit terimdir. Aynı tabloda yer alan MSE değerleri ise, bir regresyon modelinin tahminlerinin gerçek değerlerden ne kadar sapma gösterdiğini

ölçen bir metriktir. Düşük MSE değerleri daha iyi bir model performansını göstermektedir. Buna göre Çizelge 7.2 ve Şekil 7.2'den elde edilen sonuçlar şu şekildedir:

- Çizelge 7.2 incelendiğinde MSE sonuçlarında en yüksek değer RE tahmin edicisine, en düşük değer önerdiğimiz d4 tahmin edicisine ait olduğu görülmektedir. Bu durum, d4 tahmin edicisinin modelin gerçek değerlerine daha yakın tahminler yapma eğiliminde olduğunu göstermektedir.
- Çizelge 7.2 incelendiğinde önerdiğimiz tüm tahmin edicilerin (d1, d2, d3, d4), MLE, RE ve LE tahmin edicilerine kıyasla daha düşük MSE değerine sahip olduğu görülmektedir. Bu durum, önerdiğimiz tüm tahmin edicilerin çok yüksek çoklu bağlantı varlığında MLE, RE ve LE tahmin edicilerine göre verilere daha iyi uyum sağladığı ve daha doğru tahminler yaptığını göstermektedir.
- Çizelge 7.2 incelendiğinde RE tahmin edicisi dışındaki (MLE, LE, d1, d2, d3, d4) tüm tahmin edicilerin, açıklayıcı değişken katsayılarının işaretlerinin aynı olduğu görülmektedir. Yani, bir açıklayıcı değişkenin katsayısı pozitifse, diğerleri de pozitif; eğer bir açıklayıcı değişkenin katsayısı negatifse, diğerleri de negatif yönde değişmektedir.
- Çizelge 7.2'deki intercept değerleri incelendiğinde, MLE, LE, d1, d2, d3 ve d4 tahmin edicilerinde intercept değerleri birbirine oldukça yakınken, RE tahmin edicisi farklı bir değere sahip olduğu görülmektedir.
- Şekil 7.2'deki $0 < d < 1$ aralığı ile sınırlı olan farklı d değerlerine göre MLE, LE ve AULE tahmin edicilerinin MSE sonuçları incelendiğinde, önerdiğimiz AULE tahmin edicisinin (0,25 - 0,50) aralığında en düşük MSE değerine sahip olduğu gözlemlenmiştir.

Sonuç olarak, önerdiğimiz d1, d2, d3, ve d4 tahmin edicileri, modelin genel performansını artırarak diğer tahmin edicilere göre avantaj sağlamaktadır.



Şekil 7.2. Yaşam endeksi verisi için, d parametresine bağlı tahmin edicilerin MSE değerlerindeki değişimin grafiği

Çizelge 7.2. MLE, RE, LE ve önerilen d1, d2, d3, d4 tahmin edicileri için: Yaşam endeksi verisinin beta regresyon analizi parametre tahminleri ve MSE değerleri

	MLE	RE	LE	d1	d2	d3	d4
(Intercept)	0.4637	0.1658	0.4633	0.4633	0.4592	0.4596	0.4599
Fert başına düşen oda sayısı	0.0686	0.0030	0.0682	0.0671	0.0498	0.0516	0.0526
Konutun içinde tuvalet mevcudiyeti oranı	1.0601	0.0039	1.0501	1.0285	0.6774	0.7149	0.7349
Konutun kalitesinde problem yaşayanların oranı	0.3192	-0.0036	0.3155	0.3061	0.1606	0.1761	0.1844
İstihdam oranı	-0.4322	-0.0003	-0.4284	-0.4216	-0.3031	-0.3157	-0.3225
İşsizlik oranı	-0.4960	-0.0042	-0.4913	-0.4800	-0.3023	-0.3212	-0.3314
Ortalama günlük kazanç	-0.1215	-0.0026	-0.1206	-0.1196	-0.0989	-0.1011	-0.1023
İşinden memnuniyet oranı	0.6324	0.0073	0.6271	0.6165	0.4403	0.4591	0.4691
Kişi başına düşen tasarruf mevduatı	0.0874	-0.0032	0.0860	0.0822	0.0243	0.0305	0.0338
Orta ve üstü gelir grubundaki hanelerin oranı	-0.6466	-0.0007	-0.6401	-0.6247	-0.3805	-0.4065	-0.4204
Temel ihtiyaçlarını karşılayamadığını beyan eden hanelerin oranı	-0.4231	-0.0070	-0.4201	-0.4155	-0.3311	-0.3401	-0.3449
Bebek ölüm hızı	-0.6270	-0.0004	-0.6205	-0.6045	-0.3535	-0.3803	-0.3946
Doğuşta beklenen yaşam süresi	-0.2619	-0.0042	-0.2602	-0.2584	-0.2194	-0.2236	-0.2258
Hekim başına düşen müraعات sayısı	0.3894	0.0028	0.3857	0.3769	0.2373	0.2521	0.2601
Sağlığında memnuniyet oranı	0.2112	0.0109	0.2113	0.2164	0.2739	0.2677	0.2645
Kamunun sağlık hizmetlerinden memnuniyet oranı	1.0050	0.0076	0.9958	0.9756	0.6478	0.6828	0.7014
Okul öncesi eğitimde (3-5 yaş) net okullaşma oranı	-0.0857	-0.0009	-0.0854	-0.0865	-0.0956	-0.0947	-0.0941
TEOG sistemi yerleştirmeye esas puan ortalaması	-1.8887	0.0001	-1.8699	-1.8269	-1.1385	-1.2119	-1.2511
YGS puan ortalaması	-0.3026	-0.0023	-0.3003	-0.2968	-0.2320	-0.2389	-0.2426
Fakülte veya yüksekokul mezunlarının oranı	0.3763	-0.0036	0.3719	0.3611	0.1927	0.2106	0.2202
Kamunun eğitim hizmetlerinden memnuniyet oranı	0.4646	0.0072	0.4607	0.4520	0.3113	0.3263	0.3344
PM10 istasyon değerleri ortalaması (hava kirliliği)	0.0223	0.0014	0.0225	0.0241	0.0441	0.0419	0.0408
Km ² 'ye düşen orman alanı	-0.3601	-0.0007	-0.3567	-0.3498	-0.2347	-0.2469	-0.2535
Atık hizmeti verilen nüfusun oranı	-1.1065	-0.0047	-1.0960	-1.0722	-0.6900	-0.7308	-0.7525
Sokaktan gelen gürültü problemi yaşayanların oranı	-0.6051	-0.0089	-0.6003	-0.5915	-0.4399	-0.4561	-0.4647
Belediyenin temizlik hizmetlerinden memnuniyet oranı	-0.4038	-0.0014	-0.4005	-0.3955	-0.3039	-0.3137	-0.3189
Cinayet oranı (bir milyon kişide)	-0.6072	-0.0072	-0.6026	-0.5954	-0.4647	-0.4787	-0.4861
Ölümlü ve yaralanmalı trafik kazası sayısı (bin kişide)	0.5508	0.0036	0.5456	0.5340	0.3473	0.3672	0.3779
Gece yalnız yürürken kendini güvende hissedendenlerin oranı	0.5601	0.0091	0.5557	0.5478	0.4115	0.4260	0.4338
Kamunun asayiş hizmetlerinden memnuniyet oranı	-1.3095	0.0054	-1.2959	-1.2645	-0.7629	-0.8164	-0.8449
Mahalli idareler seçimlerine katılım oranı	0.5284	0.0036	0.5236	0.5133	0.3446	0.3626	0.3722
Siyasi partilere üyelik oranı	-0.2322	0.0019	-0.2297	-0.2241	-0.1343	-0.1439	-0.1490
Sendika/dernek faaliyetleri ile ilgili olanların oranı	0.3295	-0.0006	0.3260	0.3173	0.1820	0.1965	0.2042
İnternet abone sayısı (yüz kişide)	-0.2092	-0.0009	-0.2072	-0.2028	-0.1310	-0.1386	-0.1427
Kanalizasyon ve şebeke suyuna erişim oranı	0.9058	-0.0036	0.8962	0.8727	0.5031	0.5425	0.5636
Havalimanına erişim oranı	0.3644	-0.0042	0.3604	0.3517	0.2107	0.2257	0.2337
Belediyenin toplu taşıma hizmetlerinden memnuniyet oranı	-0.9402	-0.0001	-0.9311	-0.9117	-0.5942	-0.6281	-0.6462
Sinema ve tiyatro seyirci sayısı (yüz kişide)	0.3611	-0.0029	0.3571	0.3476	0.1968	0.2129	0.2215
Bin kişi başına düşen alışveriş merkezi alanı	0.4313	-0.0006	0.4272	0.4192	0.2853	0.2996	0.3072
Sosyal ilişkilerinden memnuniyet oranı	0.2554	0.0098	0.2545	0.2555	0.2569	0.2567	0.2567
Sosyal hayatından memnuniyet oranı	0.3960	0.0082	0.3934	0.3899	0.3228	0.3299	0.3338
MSE	6.4059	15.8187	6.2805	6.0068	4.5608	4.4526	4.4206

8. SONUÇ VE ÖNERİLER

8.1. Sonuçlar

Alheety ve Kibria (2009), çoklu bağlantı problemlerinin üstesinden gelmek için hemen hemen yansız Liu tahmin edicisi (AULE)'yi tanıttı. Bu çalışma AULE'ye dayalı hemen hemen yansız Liu tahmin edicisinin beta regresyon modelinde ele alınmış hali AULE'ye entegre ederek yeni tahmin ediciler önerilmiştir. Tez kapsamında, hemen hemen yansız beta Liu tahmin edicisinin bazı tahmin parametrelerinin türetme süreci detaylı bir şekilde açıklanmış ve bu tahmin edicisinin teorikteki MSE değerindeki d 'yi minimize ederek en iyi performansı sağlayacak $d_{optimal}$ parametre değerini bulma süreci anlatılmıştır. Önerilen tahmin edicilerin performansı, hem simülasyon çalışması hem de gerçek veri uygulamalarıyla değerlendirilmiştir. Önerdiğimiz yeni tahmin ediciler hemen hemen yansız olmakla kalmaz, aynı zamanda bazı koşullar altında minimum MSE'yi sağlama özelliğine sahiptir.

Daha sonra önerilen d_1 , d_2 , d_3 , d_4 tahmin edicilerin performansı diğer yaygın olarak kullanılan tahmin ediciler olan RE, MLE, LE ile karşılaştırmalı olarak gösterilmesi için hem bir simülasyon çalışması yapılmış hemde gerçek veri uygulaması üzerinden test edilmiştir. Tüm bu karşılaştırmalar sonucunda şunlar söylenebilir:

Simülasyon sonucuna göre, çoklu bağlantı durumlarında önerilen d_1 , d_2 , d_3 , d_4 tahmin edicileri MLE tahmin edicisine göre başarılı sonuçlar elde ettiği ve yine önerilen d_3 ve d_4 tahmin edicilerinin LE tahmin edicisine göre etkili sonuçlar sunduğu gösterilmiştir. İstisnai durumlar dışında, d_2 tahmin edicisinin de LE tahmin edicisine göre daha iyi sonuçlar verdiği de görülmüştür. Bunun yanı sıra, önerilen tahmin edicilerin belirli özel durumlarda RE'ye göre başarılı sonuçlar sunduğu da belirlenmiştir.

İki farklı gerçek veri uygulamasına göre, orta şiddetli çoklu bağlantı varlığında önerdiğimiz d_1 tahmin edicisi MLE ve RE tahmin edicilerinden,

önerdiğimiz d2, d3, d4 tahmin edicileri de MLE, RE ve LE tahmin edicilerinden üstün gelmiştir. Yüksek çoklu bağlantı varlığında ise önerdiğimiz d1, d2, d3, d4 tahmin edicileri MLE, RE ve LE tahmin edicilerinden üstün gelmiştir.

Bu çalışma, beta regresyon modeli için AULE tahmin edicisinin detaylı incelemesi ve geliştirilmiş versiyonunun sunulmasıyla, literatüre yeni bir bakış açısı getirmektedir. Araştırmacılara, karmaşık yapıdaki veri setlerinde daha etkili tahminler yapabilmeleri için kullanabilecekleri yeni bir araştırma yöntemi ve yaklaşım sunmaktadır.

8.2. Öneriler

Bu çalışma temel alınarak önerilen tahmin edicilerin kullanılması, özellikle finans, sağlık, ekonomi gibi alanlarda beta regresyon modellerinin uygulandığı durumlarda, daha güvenilir ve doğru sonuçlara ulaşmak isteyen araştırmacılar için önemli bir alternatif olabilir. Ayrıca, gelecekteki çalışmalarda daha fazla gerçek veri seti analizi ve farklı uygulama alanları üzerinde bu tahmin edicilerin performansının test edilmesi önerilmektedir. Bu, önerilen tahmin edicilerin genel geçerliliğini sağlamak ve farklı koşullar altında nasıl performans gösterdiklerini anlamak açısından değerli olacaktır. Bu önerilerin, beta regresyon modellerinin gelecekteki uygulamalarını daha iyi anlamamıza ve geliştirmemize katkı sağlaması umulmaktadır.

KAYNAKLAR

- Abonazel, M. R., Algamal, Z. Y., Awwad, F. A., Taha, I. M., (2022). A new two-parameter estimator for beta regression model: method, simulation, and application. *Frontiers in Applied Mathematics and Statistics*, 7, 780322.
- Abonazel, M. R., Dawoud, I., Awwad, F. A., Lukman, A. F., (2022). Dawoud-Kibria estimator for beta regression model: simulation and application. *Frontiers in Applied Mathematics and Statistics*, 8, 775068.
- Algamal, Z. Y., Abonazel, M. R., (2022). Developing a Liu-type estimator in beta regression model. *Concurrency and Computation: Practice and Experience*, 34(5), e6685.
- Alkhamisi, M., Khalaf, G., Shukur, G., (2006). Some modifications for choosing ridge parameters. *Communications in Statistics-Theory and Methods*, 35(11), 2005-2020.
- Amin M., Ashraf H., Bakouch HS, Qarmalah N., (2023). James Stein Estimator for the Beta Regression Model with Application to Heat-Treating Test and Body Fat Datasets. 12(6), 526.
- Akdeniz Duran, E., ve Akdeniz, F., (2012). Efficiency of the modified jackknifed Liu-type estimator. *Statistical Papers*, 53(2), 265-280.
- Akdeniz, F., Kaçiranlar, S., (1995). On the almost unbiased generalized liu estimator and unbiased estimation of the bias and mse, *Communications in Statistics-Theory and Methods*, 24(7), 1789-1797.
- Akram, M. N., Amin, M., Elhassanein, A., Ullah, M. A., (2022). A new modified ridge-type estimator for the beta regression model: simulation and application. *AIMS Math*, 7(1), 1035-1057.
- Algamal, Z. Y., ve Asar, Y., (2020). Liu-type estimator for the gamma regression model. *Communications in Statistics-Simulation and Computation*, 49(8), 2035-2048.
- Alheety, M. I., Kibria, B. G., (2009). On the Liu and almost unbiased Liu estimators in the presence of multicollinearity with heteroscedastic or correlated errors. *Surveys in Mathematics and its Applications*, 4, 155-167.
- Ali, S., Khan, H., Shah, I., Butt, M. M., Suhail, M., (2021). A comparison of some new and old robust ridge regression estimators. *Communications in Statistics-Simulation and Computation*, 50(8), 2213-2231.
- Alkhamisi, M. A., Shukur, G., (2007). A Monte Carlo study of recent ridge parameters. *Communications in Statistics-Simulation and Computation*, 36(3), 535-547.

- Alkhamisi, M., Khalaf, G., Shukur, G., (2006). Some modifications for choosing ridge parameters. *Communications in Statistics - Theory and Methods*, 35(11), 2005-2020.
- Asar, Y., Karaibrahimoğlu, A., Genç, A., (2014). Modified ridge regression parameters: A comparative Monte Carlo study. *Hacettepe Journal of Mathematics and Statistics*, 43(5), 827-841.
- Asar, Y., Genç, A., (2017). A note on some new modifications of ridge estimators. *Kuwait Journal of Science*, 44(3).
- Bhat, S. S., (2016). A comparative study on the performance of new ridge estimators. *Pakistan Journal of Statistics and Operation Research*, 317-325.
- Büyükuysal M. Ç., (2010). Ridge Regresyon Analizi ve Bir Uygulama, Yüksek Lisans Tezi, Uludağ Üniversitesi Sağlık Bilimleri Enstitüsü, Bursa, 18-19.
- Chen, K. K., Chiu, R. H., Chang, C. T., (2017). Using beta regression to explore the relationship between service attributes and likelihood of customer retention for the container shipping industry. *Transportation Research Part E: Logistics and Transportation Review*, 104, 1-16.
- Cribari-Neto, F., Zeileis, A., (2010). Beta regression in R. *Journal of statistical software*, 34, 1-24.
- Dorugade, A. V. (2014). New ridge parameters for ridge regression. *Journal of the Association of Arab Universities for Basic and Applied Sciences*, 15(1), 94-99.
- Erkoç, A., Ertan, E., Algamal, Z. Y., Akay, K. U., (2023). The beta Liu-type estimator: simulation and application. *Hacettepe Journal of Mathematics and Statistics*, 52(3), 828-840.
- Espinheira, P. L., da Silva, L. C. M., Silva, A. D. O., Ospina, R., (2019). Model selection criteria on beta regression for machine learning. *Machine Learning and Knowledge Extraction*, 1(1), 427-449.
- Espinheira, P. L., Ferrari, S. L., Cribari-Neto, F., (2008). On beta regression residuals. *Journal of Applied Statistics*, 35(4), 407-419.
- Ferrari, S., Cribari-Neto, F., (2004). Beta regression for modelling rates and proportions. *Journal of Applied Statistics*, 31(7), 799-815.
- Farrar, Donald E., ve Glauber, Robert R., (1967). Multicollinearity in Regression Analysis: The Problem Revisited. *The Review of Economics and Statistics*, 49(1), 92-107.
- Frisch, R. (1934). *Statistical Confluence Analysis by means of Complete Regression Systems*. (Oslo : Universitetets Okonomiske Institutt. 1934. Pp.

192.)

Geissinger, E. A., Khoo, C. L., Richmond, I. C., Faulkner, S. J., Schneider, D. C., (2022). A case for beta regression in the natural sciences. *Ecosphere*, 13(2), e3940.

Göktaş, A., Sevinç, V., (2016). Two new ridge parameters and a guide for selecting an appropriate ridge parameter in linear regression. *Gazi University Journal of Science*, 29(1), 201-211.

Gujarati, D.N. (1995). *Basic Econometrics*, 3rd Ed., Mc Graw-Hill, New York, 1002.

Gunst, R.F., (1983). Regression Analysis with Multicollinear Predictor Variables: Definition, Detection and Effects, *Communation and Statistics- Theory and Methods*, 12(19), 2217-2260.

Guolo A, Varin C., (2014). Beta regression for time series analysis of bounded data, with application to Canada Google Flu Trends. *Ann Appl Stat* 8(1), 74-88.

Güzey, T., (2023). Çoklu Bağlantı Durumunda Tutarlı Değişen Varyans Kovaryans Matrisine Dayalı Tahmin Ediciler. Yüksek Lisans Tezi, Gazi Üniversitesi, Fen Bilimleri Enstitüsü, Ankara.

Hines, W. W., (1990). *Probability and Statistics in Engineering and Management Science*, New York: John Wiley and Sons, Inc, 50-63

Hoerl, A. E. ve 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., Kennard, R. W., ve Baldwin, K. F., (1975). Ridge regression: some simulation. *Communications in Statistics*, 4, 105-123.

Hocking, R. R., Speed, F. M., and Lynn, M. J., (1976). A class of biased estimators in linear regression, *Technometrics*, 18, 425-438.

Işıklar, M., (2020). Çoklu Lineer Regresyon Modelinde Liu Tipi GM Tahmin Edicisi. Yüksek Lisans Tezi, Eskişehir Osmangazi Üniversitesi, Fen Bilimleri Enstitüsü, Eskişehir.

Kadiyala, K., (1984). A class of almost unbiased and efficient estimators of regression coefficients. *Economics Letters*, 16(3-4), 293-296.

Kalkan F., (2021). Çoklu Bağlantı Probleminde Ridge Parametrelerinin

Sağlıklarının İncelenmesi, Yüksek Lisans Tezi, Necmettin Erbakan Üniversitesi Fen Bilimleri Enstitüsü, Konya.

Karaoğlu E., (2019). Regresyon Analizinde Çoklu Doğrusal Bağlantı Probleminin İncelenmesi: Temel Bileşenler, Yüksek Lisans Tezi, Ondokuz Mayıs Üniversitesi Fen Bilimleri Enstitüsü, Samsun.

Karlsson, P., Månsson, K., Kibria, B. M. G., (2020). A Liu estimator for the beta regression model and its application to chemical data. *Journal of Chemometrics*, 34(10), e3300.

Kibria, B. G., (2003). Performance of some new ridge regression estimators. *Communications in Statistics-Simulation and Computation*, 32(2), 419-435.

Kibria, B. G., Månsson, K., Shukur, G., (2012). Performance of some logistic ridge regression estimators. *Computational Economics*, 40, 401-414.

Khalaf, G., (2012). A proposed ridge parameter to improve the least square estimator. *Journal of Modern Applied Statistical Methods*, 11(2), 15.

Khalaf, G., Iguernane, M., (2014). Ridge regression and ill-conditioning, *Journal of Modern Applied Statistical Methods*, 13(2), 18.

Khalaf, G., Månsson, K., Shukur, G., (2013). Modified ridge regression estimators. *Communications in Statistics-Theory and Methods*, 42(8), 1476-1487.

Khalaf, G., ve Shukur, G., (2005). Choosing ridge parameter for regression problems. *Communications in Statistics-Theory and Methods*, 34(5), 1177-1182.

Koc, T., ve Akın, P., (2022). Estimation of high school entrance examination success rates using machine learning and beta regression models. *Journal of Intelligent Systems: Theory and Applications*, 5(1), 9-15.

Kuvat, A., (2018). Ridge regresyonda sağlam parametre bulma, Yüksek Lisans Tezi, Muğla Sıtkı Koçman Üniversitesi Fen Bilimleri Enstitüsü, Muğla.

Lawless, J. F., ve Wang, P., (1976). A simulation study of ridge and other regression estimators. *Communications in Statistics-Theory and Methods*, 5(4), 307-323.

Liu, K., (1993). A new class of biased estimate in linear regression. *Communications in Statistics-Theory and Methods*, 22(2), 393-402.

Liu, K., (2003). Using Liu-type estimator to combat collinearity. *Communications in Statistics-Theory and Methods*, 32(5), 1009-1020.

- Lukman, A. F., Olatunji, A., (2018). Newly proposed estimator for ridge parameter: an application to the Nigerian economy. *Pakistan Journal of Statistics*, 34(2), 91-98.
- Mackinnon, M. J., ve Puterman, M. L., (1988). 'Examples of Collinearity in Gamma Regression Models', Biostatistical Consulting Service Research Paper #1, Research Division, BC Children's Hospital.
- Marquardt, D. W., (1970). Generalized Inverses, Ridge Regression, Biased Linear Estimation, and Nonlinear Estimation. *Technometrics*, 12, 591-612
- Månsson, K., (2012). Developing a Liu estimator for the negative binomial regression model: method and application, *Journal of Statistical Computation and Simulation*, 83(9), 1773-1780.
- Månsson, K., Kibria, B. G., Shukur, G., (2012). On Liu estimators for the logit regression model. *Economic Modelling*, 29(4), 1483-1488.
- Månsson, K., Kibria, B. G., Shukur, G., (2015). Performance of some weighted Liu estimators for logit regression model: an application to Swedish accident data. *Communications in Statistics-Theory and Methods*, 44(2), 363-375.
- McDonald, G. C., ve Galarneau, D. I., (1975). A Monte Carlo evaluation of some ridge-type estimators. *Journal of the American Statistical Association*, 70(350), 407-416
- Mohamed R. Abonazel ve Ibrahim M. Taha, (2021). Beta ridge regression estimators: simulation and application, *Communications in Statistics - Simulation and Computation*, 1-13.
- Montgomery, D. C., Peck, E. A., Vining, G. G., (2012). *Doğrusal Regresyon Analizine Giriş (Çev. Editörü: M. Aydın Erar)*, Nobel Yayınevi: Ankara.
- Montgomery, D. C., Peck, E. A., Vining, G. G., (2001). *Introduction to Linear Regression Analysis*, John Wiley and Sons, Inc.
- Muhammad S., Sohail Chand B. M., Golam Kibria, (2020). Quantile based estimation of biasing parameters in ridge regression model, *Communications in Statistics-Simulation and Computation*. 49(10), 2732-2744.
- Muniz, G., Kibria, B. G., (2009). On some ridge regression estimators: An empirical comparisons. *Communications in Statistics—Simulation and Computation*®, 38(3), 621-630.
- Muniz, G., Kibria, B. M., Shukur, G., (2012). On developing ridge regression parameters: a graphical investigation.
- Neter, J., Wasserman, W., Kutner, M.H., (1990). *Applied Linear Regression*

Models. Irwn Inc, Boston.

- Nomura M., (1988). On the almost unbiased ridge regression estimation. *Communications in Statistics-Simulation and Computation* 1988;17(3), 729-743.
- Ohtani, K., (1986). On small sample properties of the almost unbiased generalized ridge estimator. *Communications in Statistics-Theory and Methods*, 15(5), 1571-1578.
- Özkale, M. R., (2007). Çoklu iç ilişki ile ilgili problemler, Doktora Tezi, Çukurova Üniversitesi, Fen Bilimleri Enstitüsü, Adana.
- Qasim, M., Amin, M., Amanullah, M., (2018). On the performance of some new Liu parameters for the gamma regression model. *Journal of Statistical Computation and Simulation*, 88(16), 3065-3080.
- Qasim, M., Amin, M., Omer, T., (2020). Performance of some new Liu parameters for the linear regression model. *Communications in Statistics-Theory and Methods*, 49(17), 4178-4196.
- Qasim, M., Kibria, B. M. G., Månsson, K., Sjölander, P., (2020). A new Poisson Liu regression estimator: method and application. *Journal of Applied Statistics*, 47(12), 2258-2271.
- Qasim, M., Månsson, K., Kibria, B. M. G., (2021). On some beta ridge regression estimators: method, simulation and application. *Journal of Statistical Computation and Simulation*, 91(9), 1699-1712.
- Pamukçu, E., (2010). Sistolik Kan Basıncını Etkileyebilecek Faktörlerin Ridge Regresyon Analizi ile İncelenmesi ve Çoklu Bağlantı Problemi. Yüksek Lisans Tezi, Fırat Üniversitesi, Fen Bilimleri Enstitüsü, İstatistik Anabilim Dalı, Elazığ.
- Rasha A. Farghali, Samah M. Abo-El-Hadid, (2022). Generalized Biased Estimator for Beta Regression Model: Simulation and Application. *Mathematics and Statistics*, 10(4), 833-843.
- Silvey, S. D., (1969). Multicollinearity and Imprecise Estimation. *Journal of the Royal Statistical Society: Series B (Methodological)*, 31(3), 539-552.
- Stewart, G. W., (1973). *Introduction to Matrix Computations*, Academic Press. New York.
- Stein, C., (1956). *Proceeding of the Third Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, Berkeley, 197-206.
- Topal, M., Eydurhan, E., Yağanoğlu, A. M., Sönmez, A. Y., Keskin, S.

(2010). Çoklu Doğrusal Bağlantı Durumunda Ridge ve Temel Bileşenler Regresyon Analiz Yöntemlerinin Kullanımı. Atatürk Üniversitesi Ziraat Fakültesi Dergisi, 41(1), 53-57.

Trenkler, G. ve Toutenburg, H. (1990). Mean squared error matrix comparisons between biased estimators—An overview of recent results. *Statistical Papers*, 31(1), 165–179.

Ünlü, H., Aktaş, S., (2017). Beta regression for the indicator values of well-being index for provinces in Turkey. *Journal of Engineering Technology and Applied Sciences*, 2(2), 101-111.

Varathan, N. ve Wijekoon, P., (2021). Modified almost unbiased Liu estimator in logistic regression. *Communications in Statistics: Simulation and Computation*, 50(11), 3530-3546.

Vinod, H. D. and Ullah, A., (1981). *Recent Advances in Regression Methods*. Marcel Dekker, New York, 361p.

Zaman, T., Dündar, E., ve Aydın, S. (2019). Gini katsayısını etkileyen faktörlerin beta regresyon yöntemi yardımı ile belirlenmesi. *Erzincan University Journal of Science and Technology*, 12(1), 235-240.

Xu, J., Yang, H., (2011). More on the bias and variance comparisons of the restricted almost unbiased estimators, *Communications in Statistics-Theory and Methods*, 40(22), 4053-4064.