



T.C.
NECMETTİN ERBAKAN ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ



KANTİL REGRESYONDA YANLI TAHMİN
EDİCİLERİN PERFORMANSLARININ
İNCELENMESİ

Nurullah YAMAN

YÜKSEK LİSANS TEZİ

İstatistik Anabilim Dalı

Temmuz-2019
KONYA
Her Hakkı Saklıdır

TEZ KABUL VE ONAYI

Nurullah YAMAN tarafından hazırlanan “Kantil Regresyonda Yanlı Tahmin Edicilerin Performanslarının İncelenmesi” adlı tez çalışması 05/07/2019 tarihinde aşağıdaki jüri tarafından oy birliği / ~~oy çokluğu~~ ile Necmettin Erbakan Üniversitesi Fen Bilimleri Enstitüsü İstatistik Anabilim Dalı’nda YÜKSEK LİSANS TEZİ olarak kabul edilmiştir.

Jüri Üyeleri

Başkan

Dr. Öğr. Üyesi Serkan AKOĞUL

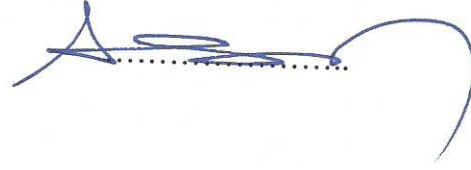
Danışman

Doç. Dr. Murat ERİŞOĞLU

Üye

Prof. Dr. Aşır GENÇ

İmza



Yukarıdaki sonucu onaylarım.

Prof. Dr. Süleyman Savaş DURDURAN
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.



Nurullah YAMAN

Tarih:05.07.2019

ÖZET

YÜKSEK LİSANS TEZİ

KANTİL REGRESYONDA YANLI TAHMİN EDİCİLERİN PERFORMANSLARININ İNCELENMESİ

Nurullah YAMAN

Necmettin Erbakan Üniversitesi Fen Bilimleri Enstitüsü
İstatistik Anabilim Dalı

Danışman: Doç. Dr. Murat ERİŞOĞLU

2019, 73 Sayfa

Jüri

Doç. Dr. Murat ERİŞOĞLU

Prof. Dr. Aşır GENÇ

Dr. Öğr. Üyesi Serkan AKOĞUL

Bu çalışmada aykırı gözlemlerin varlığında en küçük kareler regresyonuna alternatif olarak kullanılan kantil regresyonunda çoklu bağlantı probleminin çözümü ele alınmıştır. Kantil regresyonunda çoklu bağlantı probleminin çözümünde ridge regresyon yaklaşımı kullanılmıştır. Ridge tahminine dayalı kantil regresyonunda bazı yanlılık parametre tahminlerinin performansı hata kareler ortalamasına göre karşılaştırılmıştır. Simülasyon çalışması sonuçlarına göre Hocking, Speed ve Lynn (1976) ile Kibria (2003) tarafından önerilen yanlılık parametre tahmin edicileri daha başarılı bir performans göstermişlerdir.

Anahtar Kelimeler: Çoklu Doğrusal Bağlantı, En Küçük Kareler, Kantil Regresyon, Lineer Regresyon, Ridge Tahmincisi

ABSTRACT

MS THESIS

INVESTIGATING THE PERFORMANCES OF BIASED ESTIMATORS IN QUANTILE REGRESSION

Nurullah YAMAN

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

Advisor: Assoc. Prof. Dr. Murat ERİŞOĞLU

2019, 73 Pages

Jury

Advisor Assoc. Prof. Dr. Murat ERİŞOĞLU

Prof. Dr. Aşır GENÇ

Asst. Prof. Dr. Serkan AKOĞUL

In this study, the solution of the multicollinearity problem was investigated in the quantile regression which is used as an alternative to the least squares regression in case the outliers. The ridge regression approach was used to solve the multicollinearity problem in quantile regression. In the quantile regression based on ridge estimation, the performance of some bias parameter estimates was compared according to the mean error squares. According to the results of the simulation study, the bias parameter estimators proposed by Hocking, Speed and Lynn (1976) and Kibria (2003) showed a more successful performance.

Keywords: Least Squares, Linear Regression, Multiple Linear Correlation, Quantile Regression, Ridge Estimator

ÖNSÖZ

Tez çalışmamın tamamlanmasındaki süreçte hiçbir desteği benden esirgemeyen danışman hocam Sayın Doç. Dr. Murat ERİŞOĞLU'na sonsuz teşekkürlerimi sunarım. Tez çalışmamın konusunun belirlenmesinde bana yol gösteren hocam Sayın Doç. Dr. Yasin ASAR'a teşekkür ederim.

Başta İstatistik Anabilim Dalı Başkanı Prof. Dr. Aşır GENÇ olmak üzere üzerimde emeği olan öğretim üyeleri Sayın Doç. Dr. Ülkü ERİŞOĞLU, Sayın Dr. Öğr. Üyesi Aydın KARAKOCA ve Sayın Dr. Öğr. Üyesi Ahmet PEKGÖR'e teşekkürü bir borç bilirim.

Bu zorlu süreçte beni yalnız bırakmayan, eğitim hayatım boyunca maddi ve manevi destekçilerim olan anneme, babama ve ağabeyim İsmail YAMAN'a ithaf ederim.

Nurullah YAMAN
KONYA-2019

İÇİNDEKİLER

ÖZET	iv
ABSTRACT	v
ÖNSÖZ	vi
İÇİNDEKİLER	vii
ŞEKİLLER DİZİNİ	ix
SİMGELER VE KISALTMALAR.....	x
1. GİRİŞ.....	1
2. KAYNAK ARAŞTIRMASI	5
3. KANTİL REGRESYONU.....	16
3.1. Temel Kavramlar.....	16
3.2. Koşullu Ortalama ve Koşullu Kantiller	20
3.3. Basit Doğrusal Kantil Regresyon Modeli	21
3.4. Çoklu Doğrusal Kantil Regresyon Modeli.....	26
3.5. Kantil Regresyonun Özellikleri.....	29
3.6. Simülasyon Çalışması.....	30
4. RIDGE REGRESYONU	34
4.1. Giriş	34
4.2. Ridge Tahmininin Niteliği	36
4.3. Ridge Tahmininin Özellikleri.....	38
4.4. Ridge Tahmininin Hata Kareler Toplamının Minimumluğu	39
4.5. Ridge Tahmininin Hata Kareler Ortalaması	40
4.6. Ridge Tahmininin Geometrik Yorumu.....	41
5. RIDGE TAHMİNİNE DAYALI KANTİL REGRESYON ANALİZİNDE YANLILIK PARAMETRESİ TAHMİNLERİNİN PERFORMANSLARININ KARŞILAŞTIRILMASI	42
5.1. Giriş	42
5.2. Kantil Regresyonu	44
5.3. Ridge Tahminine Dayalı Kantil Regresyonu	45
5.4. Simülasyon.....	47
5.5. Uygulama	51
5.6. Sonuç.....	54
6. SONUÇLAR VE ÖNERİLER	55
KAYNAKLAR	57
ÖZGEÇMİŞ.....	63

TABLÖLAR DİZİNİ

Tablo 3.1. Chevrolet marka sekiz farklı model araçlarda beygir gücü ve fiyatlar	24
Tablo 3.2. Örnek veri seti ile EKK regresyonu ve $\theta = 0.25, 0.50, 0.75$ için KRM nde parametre tahminleri	25
Tablo 3.3. Farklı örneklem hacimlerinde EKK ve KR nda elde edilen parametre tahminleri ve MSE değerleri.....	31
Tablo 3.4. Hataların lognormal dağılıma sahip olması durumunda farklı örneklem hacimlerinde EKK ve KR nda elde edilen parametre tahminleri ve MSE değerleri	33
Tablo 5.1. Seçili parametre değerlerine göre tahmin edicilerin toplam MSE değerleri ($\theta = 0.25$)....	49
Tablo 5.2. Seçili parametre değerlerine göre tahmin edicilerin toplam MSE değerleri ($\theta = 0.50$)....	49
Tablo 5.3. Seçili parametre değerlerine göre tahmin edicilerin toplam MSE değerleri ($\theta = 0.75$)....	49
Tablo 5.4. Farklı tahmin yöntemlerine göre regresyon katsayı tahminleri, R^2 ve CV değerleri	53

ŞEKİLLER DİZİNİ

Şekil 3.1. Cars93 veri setinde fiyat değişkenine ilişkin değerler için ampirik dağılım fonksiyonu ve onun tersi ampirik kantil fonksiyonunun grafiği	17
Şekil 3.2. Tek değişkenli dağılımlarda Cars93 veri seti için merkezi eğilim ölçüleri ortalama ve çeyrekliklerin karşılaştırılması.....	20
Şekil 3.3. Chevrolet veri seti için saçılım grafiği ve $\theta = 0.5$ için tahmini basit doğrusal KRM	25
Şekil 3.4. EKK regresyonu ve 0.25, 0.50 ve 0.75. kantil değerleri ile KR dan elde edilen regresyon doğruları.....	26
Şekil 3.5. Aykırı gözlem durumunda EKK ve KR nun MSE kriterine göre karşılaştırılması.....	31
Şekil 3.6. $\mu = 0$ ve $\sigma = 0.5$ parametreleri ile Lognormal dağılıma ait olasılık yoğunluk fonksiyonu eğrisi.....	32
Şekil 3.7. Hataların lognormal dağılıma sahip olması durumunda EKK ve KR nun MSE kriterine göre karşılaştırılması.....	33
Şekil 4.1. Ridge parametresi k 'nın farklı değerlerine karşı Ridge tahminine ait varyans, yan ve MSE değerleri (solda), Ridge ve EKK tahminlerinin MSE karşılaştırılması (sağda)	40
Şekil 4.2. Ridge tahmininin geometrik yorumu.....	41
Şekil 5.1. $\theta = 0.25$ için yanlılık parametresi tahminlerine göre elde edilen regresyon katsayı tahminlerinin toplam MSE değerlerinin çizgi grafikleri	50
Şekil 5.2. $\theta = 0.50$ için yanlılık parametresi tahminlerine göre elde edilen regresyon katsayı tahminlerinin toplam MSE değerlerinin çizgi grafikleri	50
Şekil 5.3. $\theta = 0.75$ için yanlılık parametresi tahminlerine göre elde edilen regresyon katsayı tahminlerinin toplam MSE değerlerinin çizgi grafikleri	51
Şekil 5.4. Tekrarlı k katmanlı çapraz doğrulama ile elde edilen CV değerlerine ait çizgi grafiği	53

SİMGELER VE KISALTMALAR

Simgeler

$q(\theta)$: Kantil yoğunluk fonksiyonu

$\rho_{\theta}(\cdot)$: Kayıp fonksiyon

$f(y)$: Olasılık fonksiyonu

$\hat{\beta}(\theta)$: θ . kantil için parametre tahmin vektörü

$\hat{\beta}(k)$: Ridge tahmin edicisi

L_1^2 : Yanlılığın karesi

$\hat{\beta}$: β 'nin EKK tahmini

Q_{θ} : Kantil dağılım fonksiyonu

Kısaltmalar

CV : Çapraz doğruluk hatası

EKK : En küçük kareler

KR : Kantil regresyon

KRM : Kantil regresyon modeli

LAD : Medyan regresyon

LP : Doğrusal programlama

MLE : En çok olabilirlik

MSE : Hata kareler ortalaması

SSE : Hata kareler toplamı

VIF : Varyans büyütme faktörü

1. GİRİŞ

İstatistiksel analizler içerisinde en çok bilinen analizlerden biri de regresyon analizidir. Regresyon analizi iki veya daha fazla değişken arasındaki ilişkinin matematiksel olarak modellenmesine imkân sağlayan güçlü bir istatistiksel yöntemdir. Regresyon terimi ilk defa Francis Galton tarafından 1877'de İngiltere’de sunulan "Tipik Kalıtım Yasaları" konulu bir makalede tatlı bezelyelerin kalıtımı ile ilgili çalışmada kullanılmıştır. Galton çalışmasında tohum olarak kullanılan ebeveyn bezelyelerin çapları ile bu tohumlardan elde edilen yavru bezelyelerin çapları arasında doğrusal bir ilişki belirlemiştir. Çalışmada uzun çaplı ebeveyn bezelyelerin uzun çaplı yavru bezelyelere, kısa çaplı ebeveyn bezelyelerin kısa çaplı yavru bezelyelere sahip olma eğilimlerine karşı yavru bezelyelerin çaplarının ana kitle ortalamasına yaklaştığını kendi ifadesi ile gerilediğine dair bulgular elde etmiştir. Galton 1877 yılındaki çalışmasında “geri döndürme” kelimesine karşılık gelen “revert” terimini kullanmış ama 1885 yılındaki çalışmasında “gerileme” kelimesine karşılık gelen “regress” terimini kullanmıştır. Regresyon analizinin gelişimine ikinci önemli katkı Karl Pearson (1896) çalışması ile sağlanmıştır. Pearson (1896) bu yöntemi daha geniş genel istatistiksel alanlara uygulayıp geliştirmiştir. Bu çalışmalarda bağımlı ve bağımsız değişkenlerin normal dağılım gösterdiği varsayılmaktadır. Bu kısıtlayıcı varsayım R. A. Fisher 1922 ve 1925 yıllarındaki çalışmalarla sadece bağımlı değişkenin koşullu dağılımının normal dağılıma sahip olması şekliyle uygulanarak, regresyon analizinin daha yaygın kullanımına imkân sağlamıştır.

Regresyon terimi ilk kullanımı gerileme kavramını aşmış istatistikçiler ve farklı birçok disiplindeki araştırmacılar tarafından yaygın bir şekilde kullanılan istatistiksel bir yöntem haline gelmiştir. Günümüzde regresyon terimi değişkenler arasında doğrusal bağlantı bulunması ve eğri uydurma ile eş anlamlı hale gelmiştir. Birçok regresyon analizi türü varken, özünde hepsi bir veya birden fazla bağımsız değişkenin bağımlı değişken üzerindeki etkisini incelemektedir.

Gerçek uygulamalarda bağımlı değişken tam olarak açıklayıcı değişkenlerden tahmin edilemez. Bunun yerine, her bir bağımsız değişkenin sabit bir değerine verilen cevap rastgele bir değişken olarak kabul edilir ve bu rastgele değişkenin koşullu merkezi eğilim ölçüsü ile ilgilenilir. En yaygın kullanılan merkezi eğilim ölçüleri ortalama, medyan (ortanca), mod (tepe değer) ve çeyrekliklerdir. Klasik regresyon analizinde koşullu merkezi eğilim ölçüsü olarak ortalama kullanılır. Koşullu ortalamayı kullanan

regresyon modellerinde, her bir bağımsız değişkenin sabit değerleri için bağımlı değişkenin ortalamasıyla bağımlı değişken ile bağımsız değişken arasındaki ilişki açıklanır. Koşullu ortalama fonksiyonu ile modelleme basit doğrusal regresyon, çoklu regresyon, ağırlıklı en küçük kareler kullanılarak farklı varyanslı hatalara sahip modeller ve doğrusal olmayan regresyon modeller dâhil olmak üzere geniş bir regresyon modelleme yaklaşımı ailesinin çekirdeğini oluşturur. Koşullu ortalama cazip özelliklere sahiptir. Uygun koşullar altında koşullu ortalama modellerinin tahmin edicileri olan En Küçük Kareler (EKK) ve En Çok Olabilirlik (MLE) tahmin edicileri etkili istatistiksel özelliklere sahiptirler ve yorumlanmaları basittir (Hao ve Naiman, 2007).

Koşullu ortalama kullanışlı ve yararlı birçok özelliğe sahip olmasına karşın ortalamanın içerdiği dezavantajları taşırlar. Aykırı gözlemlerin varlığında ortalama uygun bir merkezi eğilim ölçüsü olmadığı için regresyon analizinde koşullu ortalamaya dayalı bir modelleme uygun olmaz. Bağımlı değişkenin dağılımının kalın-kuyruklu olduğu ve aykırı değerlerin var olduğu durumlarda, koşullu ortalama modellerinin kullanılması uygun değildir. Ayrıca bağımlı değişkenin dağılımının çarpık bir dağılıma sahip olduğu durumlarda da koşullu ortalama modelleri iyi sonuçlar vermeyebilir.

Koşullu ortalamayı temel alan klasik regresyon modellerinde parametre tahmininde kullanılan EKK ve MLE yöntemlerinin cazip istatistiksel özelliklere sahip olması için hataların normal dağılıma sahip olması, sabit varyans ve bağımsız değişkenler arasında çoklu bağlantı olmaması gibi varsayımların sağlanması gerekmektedir.

Model varsayımlarının sağlanmaması durumunda alternatif regresyon yöntemleri kullanılmaktadır. Aykırı gözlem, hataların normal dağılım göstermemesi ve değişen varyans durumlarında yaygın bir kullanıma sahip alternatif regresyon yöntemlerinden biri de Kantil Regresyon (KR)'udur.

Koenker ve Bassett (1978), klasik doğrusal regresyon modellerinin kısıtlarını ortadan kaldırmak için herhangi bir kantil değerinde bağımlı değişken ile bağımsız değişken ya da değişkenler arasındaki fonksiyonel ilişkinin tahmininde kullanılan ve kantil regresyon olarak adlandırılan bir yöntem geliştirmişlerdir. KR'unun temel özelliği koşullu merkezi eğilim ölçüsü olarak kantilleri kullanmasıdır. Kantiller duyarlı olmayan merkezi eğilim ölçüleri olarak aykırı gözlemlerden etkilenmezler ve bu özelliği ile kantil regresyonunun tahmin edicileri EKK regresyonuna göre aykırı değerlere karşı daha

sağlamdırlar. Ayrıca farklı kantillerin kullanımı ile KR’unda değişkenler arasındaki ilişkinin daha kapsamlı bir analizi elde edilebilir.

KR, EKK regresyonuna ve tipik olarak bağımsız ve bağımlı değişkenler arasındaki ilişkilerin tüm seviyelerde aynı olduğunu varsayan ilgili yöntemlere alternatif oluşturur. EKK regresyonunda amaç, bağımlı değişkene ait gözlem değerleri ile regresyon doğrusu ile tahmin edilen değerleri arasındaki farkların kareleri toplamını ifade eden hata kareler toplamını en küçüklemeektir. Buna karşılık, KR, regresyon doğrusu ile elde edilen tahmin değerleri ile gözlenen değerler arasındaki mesafeleri diferansiyel olarak ağırlıklandırır ve daha sonra ağırlıklandırılmış mesafeleri en aza indirmeye çalışır.

Bu çalışmada aykırı gözlem, hataların normal dağılım göstermemesi ve farklı varyans durumlarında EKK regresyonuna alternatif olarak kullanılan sağlam istatistiksel yöntemlerden kantil regresyonunda çoklu bağlantı problemi için ridge tahminlerinin performansı incelenecektir.

Klasik regresyon analizinde EKK tahmin edicilerinin en küçük varyansa sahip olması için bağımsız değişkenler arasında ilişki olmaması gerekmektedir. Bağımsız değişkenler arasında ilişki olması çoklu bağlantı problemi olarak ifade edilmektedir. Çoklu bağlantı durumunda EKK tahmin edicileri yansızlık özelliğini korumasına karşı en küçük varyansa sahip olma özelliğini yitirirler ve etkin bir tahmin edici olmaktan çıkarlar. Çoklu bağlantı probleminin çözümünde yan kullanarak tahmin edicilerin varyansını küçültme temelli regresyon yöntemleri kullanılabilir. En yaygın kullanılan yanlı tahmin edicilerden biri de ridge tahminidir. Ridge tahmininde X bağımsız değişkenlere ait gözlem değerlerinden oluşan tasarım matrisi olmak üzere EKK yönteminde kullanılan $X'X$ matrisinin köşegen değerlerine yanlılık parametresi eklenerek çoklu bağlantı probleminin etkisi azaltılmaktadır. Ridge regresyon (RR)’unda yanlılık parametresi k karmaşık bir fonksiyona sahiptir. Yanlılık parametresi k , tahminlerin varyansının değişiminde önemli bir rol oynamaktadır. Bu nedenle ridge regresyonunda yanlılık parametresinin seçimi oldukça önemlidir. Yanlılık parametresinin seçimi için literatürde birçok yöntem önerilmiştir.

Çalışmada öncelikle KR’u ve RR’u incelenecektir. Sonrasında aykırı gözlem, hataların normal dağılıma sahip olmaması veya farklı varyans durumlarına ek olarak çoklu bağlantı problemi içeren veri setlerinin analizi için ridge tahminine dayalı kantil regresyonu tanımlanacaktır. Simülasyon çalışması ile ridge tahminine dayalı KR’unda yanlılık parametresi k ’nın tahmini için literatürde yaygın olarak kullanılan bazı tahmin

edicilerin performansı hata kareler ortalaması kriteri ile karşılaştırılacaktır. Buna ek olarak aykırı gözlem ve çoklu bağlantı problemi içeren tobacco veri setinde tekrarlı k katmanlı çapraz doğrulama ile yanlılık parametresi tahminlerinin etkinliği incelenecektir.



2. KAYNAK ARAŞTIRMASI

Hoerl ve Kennard (1970) çalışmalarında, çoklu regresyonda, en küçük hata kareler toplamına dayanan parametre tahminlerinin, tahmin vektörleri ortogonal değilse, hatalı değilse, yetersiz olma ihtimalinin yüksek olduğunu göstermişlerdir. Önerilen, $X'X$ matrisinin köşegenine küçük pozitif sayıların eklenmesine dayanan bir tahmin prosedürü olduğuna değinmişlerdir. Sunulan ridge izleri, iki boyutta ortogonalitenin eklerini gösterme yöntemidir. Daha sonra, daha küçük hata kareler ortalamasıyla yanlış tahminler elde etmek için $X'X$ matrisinin nasıl artırılacağı gösterilmiştir.

Marquardt ve Snee (1975) çalışmalarında, yanlış tahminin veri analizi ve modeldeki kullanımını incelemişlerdir. RR teorisi ve genelleştirilmiş ters regresyonla ilişkisini gözden geçiren bir simülasyon deneyinin sonuçlarını ve pratikte ridge regresyonun kullanımına ilişkin üç örnek ile sunmuşlardır. Değişken seçim prosedürleri, model doğrulama, RR ve genelleştirilmiş ters regresyon hesaplamaya ilişkin yorumlar yer almaktadır. Burada incelenen örnekler, tahmin değişkenleri olduğundan yüksek dereceli korelasyonu RR yöntemi, EKK yönteminden daha iyi tahmin etmekte ve değişken seçimi için güvenli katsayılar ürettiği sonucuna ulaşmışlardır.

Marquardt ve Snee (1975) çalışmalarında, yanlış tahminin veri analizi ve modeldeki kullanımını incelemişlerdir. RR teorisi ve genelleştirilmiş ters regresyonla ilişkisini gözden geçiren bir simülasyon deneyinin sonuçlarını ve pratikte RR'un kullanımına ilişkin üç örnek ile sunmuşlardır. Değişken seçim prosedürleri, model doğrulama, RR ve genelleştirilmiş ters regresyon hesaplamaya ilişkin yorumlar yer almaktadır. Burada incelenen örnekler, tahmin değişkenleri olduğundan yüksek dereceli korelasyonu ridge regresyon yöntemi EKK yönteminden daha iyi tahmin etmekte ve değişken seçimi için güvenli katsayılar ürettiği sonucuna ulaşmışlardır.

Koenker ve Basett (1978) çalışmalarında, konum modelinde sıradan örnek kantilini veren basit bir minimizasyon problemini yeni bir sınıf oluşturan doğrusal modele genelleştirmişlerdir ve bu istatistikleri “regresyon kantili” olarak adlandırmışlardır. Tahmin ediciler KR yönteminde hataların mutlak toplamını en aza indirmektedir. Bazı eş varyans özellikleri ile regresyon kantillerin ortak asimptotik dağılımları ayrıca bu çalışmada yer almıştır.

Cessie ve Houwelingen (1992) çalışmalarında, parametre tahminlerini iyileştirmek için ve ileriye dönük tahminler tarafından yapılan hatayı azaltmak için ridge tahmininin lojistik regresyonda nasıl kullanılacağını göstermişlerdir. Bilinmeyen ridge

parametresini seçmenin farklı yollarını araştırmışlardır. RR'u kullanımı, over kanseri olan hastaların deoksiribonükleik asit (DNA) histogramlarının bir fonksiyonu olarak iki yıllık sağkalım olasılıkları için prognostik bir endeks geliştirerek hesaplamışlardır.

Lee ve Tanaka (1999) çalışmalarında, kantil regresyon tekniklerine dayalı yeni aralık regresyon analizi önermişlerdir. Bulanık bir ortamda bir olayın analizi için, iki aralıklı yaklaşım modelini önermişlerdir. Bu yaklaşımla çalışmanın temel amacı üst ve alt yaklaşım modellerinin bulunması ve bunları belirli bir ortamı bulanık bir ortamda temsil etmek için bulanık bir model olarak birleştirilmesi şeklindedir.

Aktaş ve Yılmaz (2003) çalışmalarında, çoklu regresyon analizi uygulamalarında karşılaşılan problemlerden birisi olan çoklu bağlantı problemi olduğundan bahsetmişlerdir. Regresyon modelindeki bağımsız değişkenleri modelden çıkarmadan EKK tahminine nazaran daha küçük hata kareler ortalaması sağlayan, fakat yanlış olan RR tahmini ile son zamanlarda RR tahminine alternatif olarak kullanılan Liu tahmin edicisinin karşılaştırılmasını yapmışlardır.

Kibria (2003) çalışmasında, RR analizinde, ridge parametresi olan k 'nin tahmini önemli bir problem olduğuna ve böyle bir parametreyi tahmin etmek için birçok yöntemin mevcut olduğuna değinmiştir. Çalışmasında bu yöntemlerden bazılarını ele almış ve ayrıca genelleştirilmiş RR yaklaşımına dayanan bazı yeni tahmin ediciler önermiştir. Önerilen tahmin edicilerin performansını en küçük hata kareler ortalaması (MSE) kriterine dayandırarak değerlendirmek için bir simülasyon çalışması yapmıştır. Simülasyon çalışması, belirli koşullar altında önerilen tahmin edicilerinin EKK tahminlerine ve diğer popüler tahmin edicilerle karşılaştırıldığında iyi bir performans sergilediğini göstermektedir. Son olarak sayısal bir örnek analiz etmiş ve sonuçları simülasyon sonuçlarını belli bir dereceye kadar desteklediği sonucuna ulaşmıştır.

Yu ve ark. (2003) çalışmalarında, KR yönteminin ortalama regresyondan daha eksiksiz bir istatistiksel model sunduğunu ve yaygın uygulama alanlarının olduğundan bahsederek bu tekniği incelemişlerdir. KR'un ilaç ve hayatta kalma analizi, finansal ve ekonomik istatistikler ve çevresel modelleme gibi birçok uygulama alanında kullanıldığından söz ederek bunların tahminlerindeki çeşitli yaklaşımları ele almışlardır. KR'un çok faydalı üç istatistiksel konseptte güçlü bağlantıları olduğunu bunların regresyon, sağlamlık ve aşırı değer teorilerini göstermişlerdir. KR'un parlak bir geleceği olduğuna değinmişler ve mevcut çalışma alanlarını ana hatlarıyla belirleyerek ileri de yapılacak olan çalışmalara önerilerde bulunmuşlardır.

Martins ve Pereira (2004) çalışmalarında, KR yöntemini eğitime geri dönüş tahminleri, okullaşma ve ücret eşitliği arasındaki ilişkiyi tahmin etmek kullanılmıştır. 1990 yılına ait 16 ülke için erkek işçilere yönelik verileri çalışmalarında kullanmışlardır. Okula dönüşü daha yetenekli bireyler için gözlemlenebilir özelliklerine bağlı olarak daha yüksek bulmuşlardır. Bu okullaşmanın düzeyler arası ücret eşitsizliği üzerinde olumlu bir yönde etkisi olduğunu göstermişlerdir.

Khalaf ve Shukur (2005) çalışmalarında, Hoerl ve Kennard (1970a) RR tahmin edicisini çoklu bağlantı varlığında normal EKK tahminlerine bir alternatif olarak sunduğunu, tasarım matrisinin kolonları arasında çoklu doğrusallık olduğu zaman ridge parametresini (K) seçmek için yeni bir yaklaşım, MSE kriterini simülasyon teknikleri ile önermişler ve değerlendirmişlerdir. Bu yöntemlerin özelliklerini etkileyebilecek bir dizi faktörün değiştiğine ve bu yaklaşımdan gelen MSE'nin hemen hemen her durumunda Hoerl ve Kennard (1970a) kullanmaya göre daha küçük olduğu görülmüştür. Çalışmada farklı çoklu doğrusallık düzeylerine ek olarak, gözlem sayısı ve hata farklılıklarının değiştiğini Monte Carlo simülasyon yöntemini kullanılarak incelemişlerdir. Her kombinasyon için 10.000 tekrar kullanılmıştır. Metodun değerlendirilmesi, MSE'leri önerdiği metot ile Hoerl ve Kennard (1970a) arasındaki karşılaştırmalarla yapılmıştır.

Saçaklı (2005) çalışmasında, alternatif regresyon modellerini karşılaştırmıştır. Uygulamasında OECD ülkelerinin büyüme oranlarını farklı model tahminleriyle incelemiştir. Büyüme ile ilgili olarak EKK, KR ve M modelleri tahmin edilerek karşılaştırma yapılmıştır. Büyüme ile ilgili kullanılan veri setinde aşırı değerler olduğunu saptamıştır. Karşılaştırma sonucu Robust ve KR modellerinin diğer modellere göre çok daha iyi sonuçlar verdiğini göstermiştir.

Alkhamisi ve ark. (2006) çalışmalarında, klasik EKK regresyonu, açıklayıcı değişkenlerin yüksek oranda korelasyonlu veya çoklu bağlantı olması durumunda büyük MSE tahminlerini ürettiğine ve tasarım matrisinin sütunları arasında çoklu bağlantı bulunduğu (K) ridge parametresini seçmek için dört değişiklik önermişlerdir. Önerilen yeni tahmin ediciler, Khalaf ve Shukur (2005) tarafından önerilenlerin genişletilmiş versiyonlarıdır. Bu tahmin edicilerin özellikleri, Hoerl ve Kennard (1970a) ve MSE kriterini kullanan EKK ile karşılaştırılmıştır. İncelenen tüm tahminciler, özelliklerini etkileyebilecek çeşitli faktörlerin değiştiği belli koşullar altında simülasyon teknikleri kullanılarak değerlendirilmiştir. Ek olarak, önerilen tahmin edicilerden en az birinin ya diğerlerinden daha küçük bir MSE'ye sahip olduğu ya da diğerinin en iyisi olduğu gösterilmiştir.

Alkhamisi ve Shukur (2007) çalışmalarında, yanlı regresyon parametrelerinin tahmin edicileri için bir takım prosedürlerin geliştirildiğine, bu işlemlerden birinin de RR'u olduğunu ve (K) ridge parametresini elde etmek için yeni bir yaklaşım önerildiğini ve daha sonra Monte Carlo simülasyonları ile değerlendirildiğini ifade etmişlerdir. Farklı gözlem sayısı, açıklayıcı değişkenler arasındaki korelasyonun gücü ve hata terimlerinin dağılımı için bir dizi model araştırmışlardır. MSE kriteri, bilinen tahmin edicilerin diğer tanınmış tahmin ediciler ile karşılaştırıldığında performansının değerlendirilmesi için kullanıldığına değinmişlerdir. Belirli koşullar altında, önerilen tahmin edicilerden en az birinin, normal EKK tahminlerinden ve Hoerl ve Kennard (1970a) tahmincisinden (HK) daha küçük bir MSE'ye sahip olduğu göstermişlerdir. Her model için 1000 tekrar yapılmış ve MSE'yi ve EKK için MSE'nin diğer RR tahminleri için MSE'den daha küçük olduğu replikasyon oranını (PR) hesaplamışlardır.

Al-Hassan (2008) çalışmasında, RR'u ile ilgili çeşitli çalışmalar yaparak ridge parametresi için k seçimi ile ilgilenmiştir. İstatistik literatüründe ridge parametresi için birçok algoritma önerilmiştir. Bu makalede, ridge parametresini tahmin etmek için yedi yöntem ele alınmıştır. Bu tahmin edicilerin performansını MSE'nin minimizasyonu kriterine göre değerlendirmek için bir simülasyon çalışması yapılmıştır. Simülasyon çalışması, belirli koşullar altında iki tahmin edicinin neredeyse genel olarak iyi performans gösterdiğini göstermektedir. Bu iki tahmin edici GM ve HKB olduğunu göstermiştir..

Sakallıoğlu ve Kaçıranlar (2008) çalışmalarında, lineer regresyon modelindeki parametre vektörü için yeni bir yanlı tahmin ediciyi tanıtmışlar ve özelliklerini tartışmışlardır. Yeni yanlı tahmin edicinin, MSE kriterine göre, normal EKK tahmincisine, sıradan RR tahminlerine ve Liu tahmin edicisine göre üstün olduğunu göstermişlerdir. Ayrıca yeni yanlı tahmin edicinin performansını Liu (2003)'te önerilen diğer iki özel Liu tipi tahmin ediciyle karşılaştırmışlardır. Bulguları Portland çimentosu üzerinde geniş bir şekilde analiz edilen veri setine dayanan sayısal bir örnekle göstermişlerdir.

Altındağ (2010) çalışmasında, basit regresyon analizi ve çoklu regresyon analizi için KR, LAD ve EKK yöntemlerini incelemiş, uygulamalarda bulunmuştur. Üç farklı uygulama verisi kullanılmıştır. Birinci uygulamasın da Devlet Meteoroloji Müdürlüğü'nün 1999-2006 yılları arasını kapsayan Konya iline ait aylık ortalama bulutluluk ve aylık toplam güneşlenme seri setini kullanmıştır. İkinci uygulamasın da ise Devlet Meteoroloji İşleri'nin 2000-2009 yılları arasını kapsayan veri setini aylık ortalama

hava sıcaklığındaki değişimleri etkileyen etkenler açısından incelemiştir. Üçüncü uygulamasın da da Birkes ve Dodge (1993)'ten elde edilen veriyle apartman dairelerindeki yangın sayısındaki değişimleri etkileyen etmenleri incelemiştir. KR analizi sonuçlarına karşılık gelen MSE ve MAD değerlerini irdelemiştir. En küçük MAD ve MSE değerlerini $\tau = 0.5$ kantilinde gerçekleştirdiğini belirlemiştir.

Topal ve ark. (2010) çalışmalarında, çoklu bağlantı problemi meydana geldiğinde EKK yönteminin sapmasız tahminlerinin olmasına rağmen varyanslarının büyük olmasından kaynaklanan gerçek değerlerinden uzakta olabilmesine değinmişlerdir. Çoklu doğrusal bağlantı probleminin olması durumun da alternatif olarak temel bileşenler regresyonu ve RR'u metotlarının kullanılabileceğini vurgulamışlardır. Araştırmalarında 91 adet farklı yaşlara sahip sazan balıklarından elde edilmiş çeşitli vücut ölçüleri ile karkas ağırlıklarını tahmin eden modelin geliştirilmesini amaçlamışlardır. Sazan balıklarından elde edilen vücut ölçümleri arasında meydana gelen çoklu doğrusal bağlantı probleminin varlığını tespit etmişler ve EKK yöntemi yerine alternatif modeller olan temel bileşenler regresyonu ile RR yöntemlerini uygulamışlardır. Aynı veri seti ile EKK, temel bileşenler regresyonu ve RR'u karşılaştırmışlardır.

Dereny ve Rashwan (2011) çalışmalarında, çoklu doğrusal bağlantı probleminin çözümü için birçok farklı regresyon modellerini tanıtmışlardır. Bu yöntemler, genelleştirilmiş RR'u, direkt RR'u ve standart RR'u yöntemlerini incelemiştir. RR tahmin edicilerin özellikleri ve yanlış RR parametresini seçme yöntemlerine değinmişlerdir. RR yöntemleri ve klasik EKK yöntemi arasındaki karşılaştırmayı yapmak için simülasyon verilerini kullanmışlardır. Çalışmanın sonuçlarına göre; RR'un çoklu doğrusal bağlantı probleminin varlığında EKK yönteminden daha iyi sonuçlar verdiğini göstermişlerdir.

Karakuş (2011) çalışmasında, klasik regresyon yönteminin uygulanabildiği veri setlerine RR yönteminin de uygulanabileceğini ancak çoklu doğrusal bağlantı probleminin olması hâlinde RR yönteminin klasik EKK regresyon yöntemi yerine kullanılabileceğini vurgulamıştır. Ayrıca çalışmasında, çoklu doğrusal bağlantı probleminin giderilmesi için temel bileşen regresyon analizi ve faktör skorları yöntemlerinin alternatif olarak klasik EKK regresyon analizi yöntemi yerine kullanılabileceğini ifade etmiştir.

Toker (2011) çalışmasında, ridge tahmin edicisi ve bu tahmin edici ile ilişkili olan birden çok tahmin ediciyi incelemiştir. Ayrıca çalışmada farklı regresyon modelleri için ridge tahmin edicileri de incelemiştir. Bununla birlikte iki parametreye sahip ridge tahmin

edicisi; EKK tahmin edicisi, Stein tipi tahmin edicisi ve ridge tahmin edicisi ile hata kareler ortalamasının matrisi kriteri altında karşılaştırma yapmıştır. Sözü edilen tahmin edicilerin performansını görmek için de gerçek veri seti ile analiz gerçekleştirmiştir.

Keskin (2012) çalışmasında, 2005-2009 yılları arasını kapsayan Ekonomik Kalkınma İşbirliği Örgütü verilerini kullanarak Avrupa ülkelerindeki beklenen yaşam süreleri, KR ve EKK yöntemlerini incelemiştir. Sonuç olarak da düşük kantillerde KR elde edilen tahminlerin EKK yönteminden elde edilen tahminlere göre çok daha küçük hatalı olduğunu görmüştür. KR modellerinin, yanıt değişkenin uç değerleri ile çalışıldığında daha iyi tahminlerde bulunduğu sonucuna ulaşmıştır.

Zeebari (2012) çalışmasında, ridge tahmin yöntemini medyan regresyona genelleştirmiştir. En küçük mutlak sapma (LAD) kestirim yöntemi, Gauss olmayan ya da asimetrik hata terimlerinin varlığında güçlü olmasına rağmen dikgen olmayan açıklayıcı değişkenler söz konusu olduğunda, ciddi bir çoklu doğrusallık probleminde bozulmaya devam edebileceğine değinmiştir. Önerdiği yöntem, LAD tahmin edicilerinin verimliliğini artıracak ve bu durumda sapmayı azaltarak ve yan için LAD tahmin edicilerinde daha küçük bir MSE elde edilmesi için daha fazla imkan sağlamaktadır. Çalışma ayrıca yeni yöntemin bir uygulamasını ve bir simülasyon çalışmasını da içermektedir. Ek olarak, asimetrik veriler ve aykırı değerler ile açıklayıcı değişkenler arasındaki çoklu doğrusal bağlantı da herhangi bir artış için LAD tahmin yöntemine kıyasla LAD ridge yönteminin göreceli etkinliğinin de artığı sonucuna ulaşmıştır.

Akyol (2013) çalışmasında, 187 ülkenin 2012 yılına ait veri seti üzerinde EKK yöntemi, en küçük mutlak sapma yöntemi ve KR yöntemlerini kullanarak çeşitli tahminlerde bulunmuştur. Uygulamasında kullandığı veri seti normal dağılmadığından, aykırı değerler ve değişen varyans olduğu için KR analiziyle elde edilen sonuçların ve bulguların diğer tahmin ettiği yöntemlere göre istatistiksel olarak anlamlı olduğunu açıklamıştır.

Gökçe (2013) çalışmasında, döviz kurlarında meydana gelen değişimlerin Türkiye’de özel sermaye ile kurulmuş olan bankaların bilançosundaki kalemlere etkisini incelemiştir. Gerçekleştirdiği analizde klasik EKK regresyon modeline göre veri setindeki uç değerleri daha çok dikkate alan KR modelini tercih etmiştir. Bilanço verilerinin değişimini 0.05, 0.25, 0.50, 0.75 ve 0.95 kantillerde en ayrıntılı bir şekilde ölçmüştür.

Çelik ve Selim (2014) çalışmalarında, Türkiye’de kırsal ve kentsel kesimdeki özel ve kamuda çalışan erkekler ve kadınlar arasında meydana gelen gelir farklılıklarını analiz

etmişlerdir. Türkiye İstatistik Kurumu (TÜİK)'nin yapmış olduğu 2011 yılına ait hanehalkı işgücüne dayalı araştırması mikro veri seti ile KR ve EKK yöntemlerini karşılaştırmışlardır.

Elmalı (2014) çalışmasında, 2010-2013 yılları arası esas alınarak illerin ilaç kullanımında etkili olan faktörleri incelemiştir. EKK yönteminin varsayımları sağlanmadığında çoklu doğrusal regresyona alternatif olan robust yöntemlerden olan KR yöntemini kullanmış ve çeşitli kantillerde analizler yapmıştır. Çeşitli karşılaştırma yöntemleri ile negatif binomial regresyon analizi sonuçlarının KR analizi sonuçlarına göre daha etkin sonuçlar da bulunduğunu ifade etmiştir.

Koşan (2014) çalışmasında, 2002-2010 yılları arasını kapsayan İktisadi İşbirliği ve Gelişme Teşkilatı (OECD) üyesi olan ülkelerin dış ticaret hadlerin de etkide bulunan değişkenlerin incelenmesi amacıyla panel KR ve panel regresyon modellerinin tahminlerinde bulunarak sonuçları yorumlamıştır. Panel KR modeli ve panel regresyon modeli için hesapladığı standart hataları karşılaştırmış, panel KR modeli için standart hataların daha küçük olduğunu tespit etmiştir.

Durusu Çiftçi (2015) çalışmasında, 1989-2011 yıllarını kapsayan 40 ülkeye ait panel veri setini kullanarak, panel eş-bütünleşme ve yatay kesit bağımlılığını dikkate alan Kesitler Arası Korelasyon ve Genişletilmiş Ortalama Grup tahmin yöntemleri kullanılarak; analizi yapılan değişkenlerin büyüme üzerindeki etkileri ülkelerin koşullu büyüme dağılımında bulundukları konumlara göre belirlenmesine olanak sağlayan KR yöntemiyle tahminde bulunmuştur.

Karaoğlu (2015) çalışmasında, Türkiye Sağlık Araştırması verilerini kullanmıştır. Eğitim ve Vücut Kitle Endeksi ilişkisi üzerinde durulmuş ve KR modeli kullanılarak Vücut Kitle Endeksi'nin değişik kantillerinde eğitim ve diğer sosyoekonomik nedenlerin nasıl bir rol oynadığı incelenmiştir. KR sonuçlarının yüksek eğitilmiş kişilerin Vücut Kitle Endeksi düzeylerinin yüksek olmasının tehlikelerine karşı daha fazla dikkate aldığını görmüştür.

Türkş (2015) çalışmasında, kantil regresyon yöntemi ile parametrik regresyon yöntemlerinin performanslarını, 1000 tekrar sayısından oluşan simülasyon çalışması ile hem de homojen ve heterojen yapıda olan gerçek sağkalım veri setleri üzerinde karşılaştırmış ve minimum hatalı olan modeli belirlemiştir. Sağkalım analizinde KR yönteminin, hem homojen yapıdaki veri setinde hem de heterojen yapıdaki veri setinde parametrik regresyon yöntemlerine göre daha iyi performans gösterdiği sonucuna ulaşmıştır.

Firinguetti ve ark. (2016) çalışmalarında, çoklu doğrusal bağlantı problemiyle mücadele etmek için kısmi EKK ve RR yöntemlerini incelemişlerdir. Klasik EKK ile olan performanslarını karşılaştırmak için bir simülasyon çalışması yapmışlardır. Değişkenler arasındaki çoklu doğrusal bağlantı derecelerinde, hem kısmi EKK hem de RR tahmin edicilerinde MSE kriteri ve tahmin MSE kriterinde, EKK'ye göre önemli düşüşler sağladığını görmüşlerdir. Ayrıca simülasyon çalışmasında hataların varyansı büyük olduğunda RR yönteminin daha iyi performans gösterdiği ve modelin daha fazla değişken içerdiği durumda ise kısmi EKK tahminlerinin en iyi sonuca ulaştığını görmüşlerdir. Bununla birlikte, RR yönteminin kısmi EKK yöntemine göre avantajının, kısmi EKK yönteminin sağlayamadığı regresyon katsayıları için %95 güven aralığını sağladığı da ifade edilmiştir.

Kibria ve Banik (2016) çalışmalarında, ridge parametresinin tahmini, çoklu doğrusal bağlantı problemini çözmek için yaygın olarak kullanılan RR yönteminde önemli bir problem olduğunu ifade etmişlerdir. Kullanılabilir 28 farklı tahmin edici ve önerilen beş ridge tahmincisi, KB1, KB2, KB3, KB4 ve KB5 üzerine kapsamlı bir çalışma yapmışlardır. Simülasyon çalışması yaparak tahmin edicileri karşılaştırmışlardır. Seçilen ridge tahmin edicilerinden bazılarının, klasik EKK tahmincisine ve mevcut bazı popüler ridge tahminlerine göre iyi performans gösterdiğini, önerilen ridge tahmincilerinden olan KB3'ün en iyisi olduğunu sayısal örneklerle göstermişlerdir.

Uslu (2016) çalışmasında, eğitimin ücretler üzerindeki etkisini incelemiştir. Eğitimin getirisini Türkiye İstatistik Kurumu (TÜİK)'in 2010 yılına ait Kazanç Yapısı Araştırması verisini kullanarak anketel regresyon ve kantil regresyon metotlarıyla tahmin etmiştir. 0.5 (medyan) ücret kantili öncesinde eğitimin getiri olarak erkek ve kadın çalışanlar arasında göze çarpan bir farklılık bulamazken medyan ücret kantili sonrası getirinin kadın çalışanlar yararına daha yüksek olduğunu bulmuştur.

Bayrak Gezim (2017) çalışmasında, 1980-2015 yılları arasını kapsayan 165 ülkenin panel veri setini kullanmıştır. Küresel CO₂ emisyon dağılımının heterojen olması nedeniyle EKK yöntemi yerine KR yöntemini tercih etmiştir. Gecikmeli yanıt değişkeninin etkisini dinamik panel KR yöntemiyle incelemiştir. Ekonomik büyüme, nüfus, endüstriyel faaliyetler ve demokrasinin yüksek ve düşük emisyon düzeylerinde farklı etkilere sahip olduğu bulgusuna ulaşmıştır. Finansal açıklığın emisyon dağılımındaki etkisini anlamsız bulmuştur.

Dorak (2017) çalışmasında, Borsa İstanbul'un imalat sektöründe işlem gören 183 firma ile ilgili 2013-2015 yılları arasındaki verilerini kullanarak bu verileri EKK ve KR

yöntemleriyle incelemiştir. KR ve EKK yöntemleri için aynı veri setini kullanarak farklı sonuçlar elde etmiştir. Veri setinde aykırı gözlemler olduğu için EKK ve KR yöntemlerinden farklı sonuçlar elde edilmesinde etkili olmuştur. Bu farklılıklar dikkate alındığında, KR yönteminin istatistiksel açıdan daha doğru ve anlamlı sonuçların elde edilmesinde etkili olduğu sonucuna ulaşmıştır.

Kırdemir (2017) çalışmasında, EKK yöntemi ile regresyon analizi gerçekleştirmiş daha sonra ise RR analizi ile elde ettiği sonuçları yorumlamıştır. Uygulama da çoklu doğrusal bağlantı problemi olduğunu tespit etmiş ve modelden veri çıkarmayı ya da eklemeyi istemediğinden analiz için yanlı bir tahmin edici olan RR modelini kullanmıştır. Sonuç olarak güçlü çoklu doğrusal bağlantı problemi olduğunda RR yönteminin geçerli, tutarlı, kuramsal ve durağan beklentilere uygun şekilde tahminlerde bulunduğunu görmüştür.

Özocaklı (2017) çalışmasında, yetersiz beslenme yaygınlığına sahip olan ülkeler arasından seçilen 52 ve 80 ülkeli iki örneklem ile 2000-2015 yılları arasını kapsayan iki ayrı model belirlemiştir. Analiz yöntemi olarak da KR yöntemini kullanmıştır. Tahmin, katsayılarının anlamlı farklı olup olmadığını (0.05, 0.10,...,0.95, 0.97, 0.98, 0.99) kantillerinde gerçekleştirmiştir.

Tongal (2017) çalışmasında, PISA 2015 yılı Türkiye veri setini kullanarak bazı değişkenlerin (EMOSUPS, EPIST, DISCLISCI, UNFAIRTEACHER, HISEI, HISCED, HEDRES, ESCS, CULTPOSS) Fen okuryazarlığı puanına etkisinin bazı bölgelere (Güneydoğu Anadolu, Akdeniz ve İstanbul) göre değişim gösterip göstermediğini incelemiştir. Analizinde KR yöntemini kullanmıştır. Ebeveynlerin duygusal desteği değişkeninin Türkiye genelin 0.4 ile 0.7 kantilleri aralığı haricinde diğer bölgeler için istatistiksel olarak anlamsız olduğunu belirtmiştir.

Yıldırım (2017) çalışmasında, Türkiye İstatistik Kurumu (TÜİK)'in 2003-2013 yılları arasında yapmış olduğu hanehalkı bütçe anketi veri setini kullanmış ve Türkiye'de hanehalkı tasarruf yönünü ekonometrik yöntemlerle araştırmıştır. Analiz doğrultusunda KR yöntemiyle düşük, orta ve yüksek gelire sahip hanehalkı tasarruf yönelimini karşılaştırmıştır. Faiz katsayısını tahmin edilen tüm kantiller de negatif tasarruf yönünde sonuçlanmış ve tahmin edilen tüm koşullu kantil katsayıları negatif bulmuştur. Katsayıları anlamlı olarak elde etmiştir.

Çamurlu (2018) çalışmasında, 2000-2017 yıllarını kapsayan Üretici Fiyat Endeksi(ÜFE), ÜFE(-2) Dönem Gecikmesi ve Beklenti veri setini kullanmıştır. Veri seti Bootstrap yöntemini kullanarak belirli seviyede veri sayılarını arttırmış KR ve doğrusal

regresyon yöntemlerinin sonuçlarını MSE'nin karekökü ve ortalama mutlak sapma değerleri ile karşılaştırarak hangi yöntemin modeli en iyi tahmin ettiğini belirlemiştir. Sonuç olarak ortalama mutlak sapma ve MSE'nin karekökü değerlerini karşılaştırdığında minimum değeri doğrusal regresyon yönteminin sağladığını görmüştür. İkinci model olarak da KR'un (Q_2) veri seti için uygun bir model olduğunu belirlemiştir.

Demirez (2018) çalışmasında, PISA 2015 Almanya, Singapur ve Türkiye veri setlerini kullanarak bazı değişkenlerin (SCIEACT, SCIEEFF, COOPERATE, SMINS, OUTHOURS, CPSVALUE, ESCS, WEALTH, IBTEACH, TDTEACH, ICTRES) Fen okuryazarlığı puanına etkisinin söz konusu ülkelere göre değişim gösterip göstermediğini incelemiştir. Bu değişkenlerin etkisi düşük, orta yüksek puanlı öğrencilerde farklı olup olmadığını araştırmıştır. Araştırmasında KR yöntemi ve EKK regresyon yöntemini kullanmıştır.

Kuvat (2018) çalışmasında, sağlam parametre geliştirmede RR'un parametrelerinden yararlanmıştır. Çoklu doğrusal bağlantılı serilerdeki en iyi olan ridge parametresini belirlemek için tarama yöntemini kullanmış ve tarama sonucunda en iyi olan ridge parametresinin literatürde bulunan ridge parametreleri ile fonksiyonel ilişkisini araştırmıştır. EKK yöntemine alternatif olan RR için sağlam bir parametre geliştirmiştir. Çoklu doğrusal bağlantı problemi olan modellerde RR yönteminin kullanılması bu parametre için fayda sağlayacağını belirtmiştir.

Lukman ve Ayinde (2018) çalışmalarında, etkili gözlemlerin regresyon modellerinin performansı üzerinde büyük bir sorun oluşturduğunu ve klasik EKK yöntemini kullanmışlardır. Literatürde Cook uzaklığı ve DFFITS dahil olmak üzere farklı etkili istatistikleri tanıtmışlardır. Bu durumda gerçekleştirilecek önlemlerin etkinliği, lineer regresyonda çok doğrusal bağlantının varlığından etkilenmekte olduğuna ve bununla birlikte, her iki probleminde ortak bir regresyon modelinde bulunabileceğinden bahsetmişlerdir. Özkale ve Kaçıranlar (2007) tarafından tanımlanan iki parametrelili Liu-Ridge tahmin edicisine (TPE) dayandırılan yeni tanısallık önlemlerin mevcut olduğuna ve bunlara alternatif olarak önermişlerdir. TPE için etkili olayların tespiti için yaklaşık silme formülleri önermişlerdir. Son olarak iki gerçek veri seti üzerinde tanısallık önlemleri göstermişlerdir.

Mori ve Suzuki (2018) çalışmalarında, çok değişkenli regresyonda kare risk altında maksimum olasılık tahmin edicisine hâkim olan genelleştirilmiş ridge tahminlerine ve çok değişkenli lineer regresyonda Kullback-Leibler riskine dayanan yeni bir model seçimi kriterini sunmuşlardır. Model seçim kriterleri yansızlık, tutarlılık ve

düzgün olarak minimum varyans özelliklerine sahip olduğuna değinmişlerdir. Tutarlılık $p / n \rightarrow c$ 'nin asimptotik yapısı altında kanıtlamışlar, bura n , örneklem büyüklüğüdür ve p , yanıt değişkenlerinin parametre büyüklüğünü temsil etmektedir. Önerdikleri tahmin edici sınıfı, model gerçek modeli içermese dahi, kare riski altında maksimum olasılık tahmin edicisine hâkimdir. Deneyisel sonuçlar ise model seçim kriterlerinin risklerin maksimum olasılık tahmin edicisine göre daha küçük olduğuna ve önerilen kriterin bazı koşullar altında gerçek modeli belirlediğini göstermektedir.

Özdemir (2018) çalışmasında, kırılğan beşli olarak da isimlendirilmiş olan Hindistan, Brezilya, Güney Afrika, Türkiye ve Endonezya için enflasyonun kalıcılığı üzerinde araştırma yapmıştır. Enflasyonun kalıcılığını her bir kantil üzerinde ve birikimli şekilde kantil birim kök testleri ile analiz etmiştir. Türkiye için enflasyon serisinin hepsinde ve birinci kantilinde (0.1) durağan olduğunu geriye kalan kantillerde (0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9) ise birim kök olduğu sonucuna ulaşmıştır.

Unat (2018) çalışmasında, Türkiye İstatistik Kurumu (TÜİK)'in 2005-2016 yılları arasında Hanehalkı Bütçe veri setini kullanarak hanehalkı tüketim eğiliminin yapısının harcama gruplarına göre araştırmasını yapmıştır. Geleneksel tüketim fonksiyonundan yola çıkarak gelir ve tüketim ilişkisinin doğrusal regresyon modelleriyle araştırıldığını belirtmiş ve yöntemin varsayımları sağlamaması hâlinde hanehalkı gelirleri ile hanehalkı tüketim harcamaları arasındaki ilişkiyi KR modeliyle araştırmıştır. Hanehalkının düşük tüketim eğilimli olduğunu ve kantil regresyonun tüm kantilleri için en düşükten en yükseğe doğru gidildiğinde tüketim eğiliminde artış olduğunu gözlemlemiştir.

Alakaya (2019) çalışmasında, farklı büyüklüklerde örneklem için aykırı değerler üreterek EKK yöntemi ile KR yönteminin performanslarını değerlendirmiştir. İlk önce aykırı gözlem olmadığı hâlinde analizler gerçekleştirmiş sonuçları değerlendirmiş sonra ise aykırı gözlem olduğunda sonuçları tekrar değerlendirmiştir. Aykırı gözlem olduğunda R^2 değerleri, EKK ve KR yöntemlerinde aykırı gözlem olmadığı hâline göre artış göstermiştir. Örneklem büyüklüklerinin hepsinde %5 ve %10 aykırı değer bulunduğunda en yüksek R^2 değeri KR için $\tau = 0.75$ 'te elde etmiştir. %20 aykırı değer olması hâlinde en yüksek R^2 değeri, kantil regresyon için $\tau = 0.25$ 'te elde etmiştir. MAD ve MSE değerlerinde istatistiksel olarak önemli bir farklılık gözlemlememiştir.

3. KANTİL REGRESYONU

Bu bölümde KR'nu ile ilgili temel kavramlar, basit ve çoklu KR modeli anlatılacaktır. Bu bölümün anlatımında genel olarak Davino ve ark. (2013) den yararlanılmıştır. Bölüm sonunda simülasyon çalışması ile EKK ve KR yöntemleri MSE kriterine göre karşılaştırılmıştır.

3.1. Temel Kavramlar

Bu bölümde ortalama, medyan, kantil, koşullu ortalama ve koşullu kantil kavramları açıklanacaktır.

Kantil Regresyon (KR) ile En Küçük Kareler (EKK) regresyonu arasındaki temel farkı açıklayabilmek için her iki yöntemdeki amaç fonksiyonlarını göz önünde bulundurmak yararlıdır.

Merkezi eğilim ölçüsü olarak en yaygın kullanılan ortalamanın en önemli özelliklerinden biri terimlerin ortalamadan olan farklarının kareleri toplamının minimum olmasıdır. Bu özellik göz önünde bulundurularak ortalama, farkların kareleri toplamını en aza indiren merkezi eğilim ölçüsü c olarak tanımlanır ve Y bir rasgele değişken olmak üzere aşağıdaki eşitlik ile gösterilir.

$$\mu = \underset{c}{\operatorname{argmin}} E(Y - c)^2 \quad (3.1)$$

Diğer bir merkezi eğilim ölçüsü olan medyan, küçükten büyüğe sıralanmış bir veri setinde seriyi iki eşit parçaya bölen tam ortadaki değer olarak tanımlanır. Duyarlı bir merkezi eğilim ölçüsü olan ortalamanın aykırı gözlemlerin varlığında veriyi temsil yeteneği olmadığından ortalamaya alternatif olarak veriyi temsil etmede genellikle medyan kullanılır. Terimlerin medyandan olan mutlak farklarının toplamı minimumdur. Medyanın bu özelliği kullanılarak medyan mutlak farkların toplamını en aza indiren merkezi eğilim ölçüsü c olarak tanımlanır ve aşağıdaki eşitlik ile gösterilir.

$$Q_2 = \underset{c}{\operatorname{argmin}} E|Y - c| \quad (3.2)$$

Ortalama ve medyan için örneklemden elde edilen tahmin ediciler örneklem ortalaması ve örneklem medyanı sırasıyla $\bar{y}$ ve q_2 ile gösterilir.

Kantiller, tek değişkenli dağılımlarda belirli konumlara karşılık gelen gözlem değerleri olarak tanımlanır. Yani θ . kantil $P(Y \leq y) = \theta$ eşitliğini sağlayan y değeridir. Y rastgele değişkeninin y değerine eşit veya daha küçük olma olasılığı olarak tanımlanan birikimli dağılım fonksiyonu

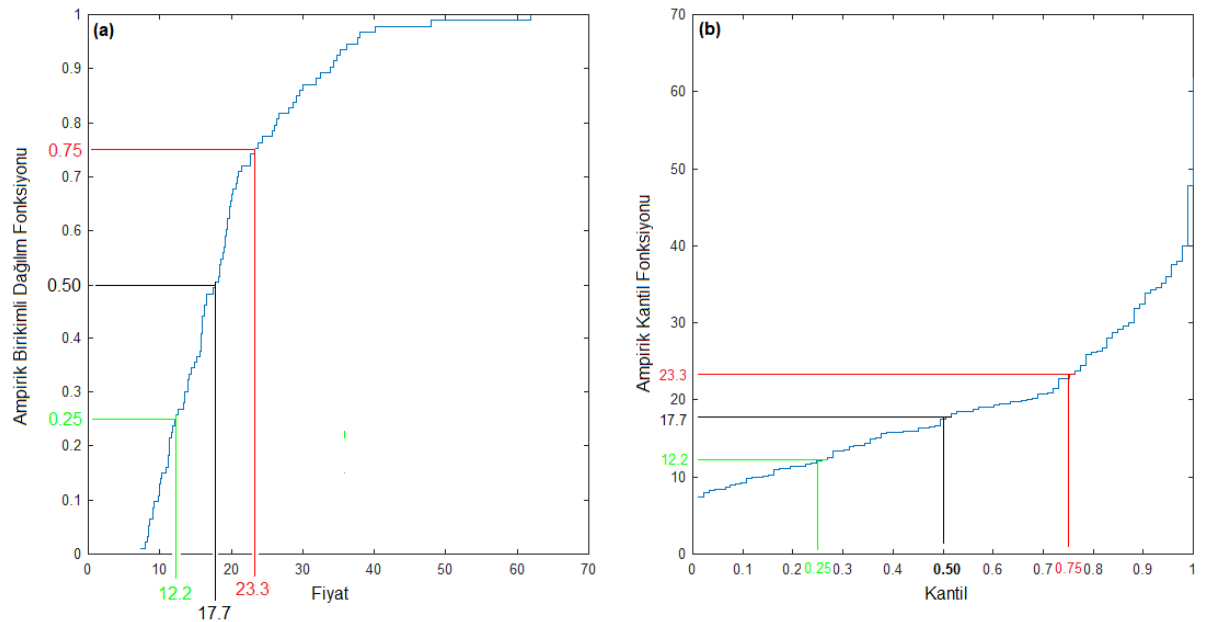
$$F_Y(y) = F(y) = P(Y \leq y) \quad (3.3)$$

şeklinde tanımlanır.

Kantil fonksiyonu ise tek değişkenli bir dağılımda birikimli dağılım fonksiyonunun tersi olarak belirli bir olasılığa karşı gelen gözlem değerini veren

$$Q_Y(\theta) = Q_\theta = F_Y^{-1}(\theta) = \inf\{y : F(y) > \theta\} \quad (3.4)$$

fonksiyon olarak tanımlanır. Eşitlik (3.4.)'te yer alan θ olasılık değerini göstermektedir ve $0 \leq \theta \leq 1$ olmalıdır. Q_θ gösterimi ise θ olasılığına karşılık gelen kantil değerini göstermektedir. Şekil 3.1'de Cars93 veri setinde fiyat değişkenine ilişkin değerler için ampirik dağılım fonksiyonu ve onun tersi ampirik kantil fonksiyonunun grafiği verilmiştir.



Şekil 3.1. Cars93 veri setinde fiyat değişkenine ilişkin değerler için ampirik dağılım fonksiyonu ve onun tersi ampirik kantil fonksiyonunun grafiği

Kantil fonksiyonunun türevi alınarak kantil yoğunluk fonksiyonu $q(\theta) = \frac{\partial Q_Y(\theta)}{\partial \theta}$ elde edilir. Kantil fonksiyonu ile kantil yoğunluk fonksiyonu arasındaki ilişki $q(\theta) = \frac{1}{f(Q_Y(\theta))}$ eşitliği ile gösterilir. $Q_Y(\theta)$ fonksiyonu azalmayan bir fonksiyon olduğundan türevi de negatif değildir ve her zaman $0 < \theta < 1$ olur.

Kantiller, ortalama ve medyadaki gibi dağılımın mutlak farkları toplamını en aza indiren belirli dağılım merkezi

$$Q_\theta = \underset{c}{\operatorname{argmin}} E[\rho_\theta(Y - c)] \quad (3.5)$$

olarak da tanımlanabilir (Hao and Naiman 2007). $\rho_\theta(.)$ kayıp fonksiyonu olarak tanımlanır ve

$$\rho_\theta(y) = [\theta - I(y < 0)]y = [(1 - \theta)I(y \leq 0) + \theta I(y > 0)]|y| \quad (3.6)$$

eşitliğiyle gösterilir. Gerçekte $\rho_\theta(.)$ kayıp fonksiyonu asimetrik bir mutlak kayıp fonksiyonudur ve mutlak farkların ağırlıklı bir toplamını ifade eder. Bu fonksiyon negatif farklar için $(1 - \theta)$ ağırlığını pozitif farklar için θ ağırlığını kullanır.

Olasılık fonksiyonu $f(y)$ ve Y kesikli bir rastgele değişken olmak üzere kantil tanımında kullanılan minimizasyon problemi

$$Q_\theta = \underset{c}{\operatorname{argmin}} \left\{ (1 - \theta) \sum_{y \leq c} |y - c| f(y) + \theta \sum_{y > c} |y - c| f(y) \right\} \quad (3.7)$$

şeklinde gösterilebilir. Aynı kriter, sürekli rastgele değişken için

$$Q_\theta = \underset{c}{\operatorname{argmin}} \left\{ (1 - \theta) \int_{-\infty}^c |y - c| f(y) d(y) + \theta \int_c^{+\infty} |y - c| f(y) d(y) \right\} \quad (3.8)$$

şeklinde gösterilir.

Tek deęişken durumunda kantillerin tanımında kullanılan Eşitlik (3.5) ile gösterilen minimizasyon probleminin çözümü (Koenker, 2005) için öncelikle medyanın tanımında kullanılan Eşitlik (3.2)'deki minimizasyon probleminin çözümü yararlı olacaktır.

Y sürekli bir rastgele deęişken ve tanım aralığı $-\infty < y < \infty$ olsun. Y rastgele deęişkeninin c 'den mutlak farklarının beklenen deęeri

$$\begin{aligned} E|Y - c| &= \int_{-\infty}^{\infty} |y - c|f(y)d(y) \\ &= \int_{-\infty}^c (c - y)f(y)d(y) + \int_c^{\infty} (y - c)f(y)d(y) \end{aligned} \quad (3.9)$$

şeklinde gösterilir. Mutlak deęer dışbükey bir fonksiyon olduęu için Eşitlik (3.9)'un c 'ye göre kısmı türevi sıfıra eşitlenerek beklenen deęeri minimum yapan c deęeri belirlenebilir.

$$\begin{aligned} \frac{\partial}{\partial c} E|Y - c| &= \int_{-\infty}^c f(y)d(y) - \int_c^{\infty} f(y)d(y) = 0 \\ \frac{\partial}{\partial c} E|Y - c| &= F(c) - (1 - F(c)) = 0 \\ F(c) &= \frac{1}{2} \end{aligned} \quad (3.10)$$

Eşitlik (3.10)'da elde edilen $F(c) = \frac{1}{2}$ eşitlięi c deęerinin kitleyi iki eşit parçaya ayırdıęını ifade etmektedir. Dolayısıyla mutlak farkların beklenen deęerini minimum yapan c deęeri medyan deęeridir ve $c = Q_2$ olur.

Yukarıdaki çözüm, $E|Y - c|$ 'nin iki bileşenini sırasıyla sabit θ ve $(1 - \theta)$ ile çarpılmasıyla deęişmez. Bu, kantillerin genel tanımında kullanılan Eşitlik (3.5) ile ifade edilen minimizasyon probleminin çözümüne imkan sağlar. Yani, Eşitlik (3.5)'teki minimizasyon probleminin çözümü benzer şekilde

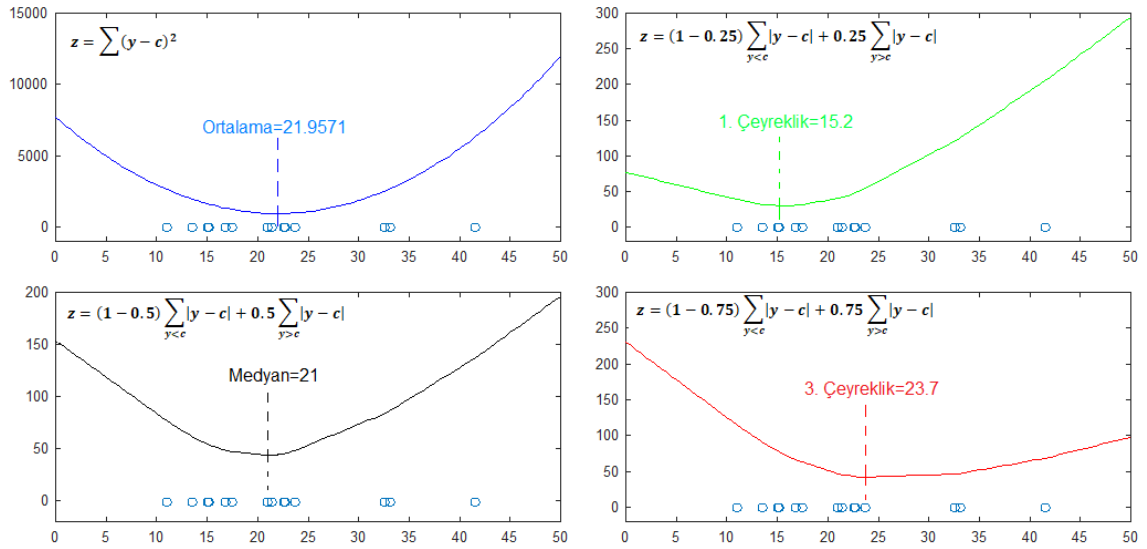
$$\frac{\partial}{\partial c} E[\rho_\theta(Y - c)] = \theta \int_{-\infty}^c f(y) d(y) - (1 - \theta) \int_c^{\infty} f(y) d(y) = 0$$

$$\frac{\partial}{\partial c} E[\rho_\theta(Y - c)] = (1 - \theta)F(c) - \theta(1 - F(c)) = 0 \quad (3.11)$$

$$F(c) - \theta F(c) - \theta + \theta F(c) = 0$$

$$F(c) = \theta \quad c = Q_\theta$$

olarak elde edilir. Şekil 3.2’de *Cars93* veri setinde sportif araçlar için oluşturulan veri setinde merkezi eğilim ölçüleri ortalama ve çeyrekliklerin hesaplanmasında kullanılan amaç fonksiyonlarının karşılaştırılması verilmiştir.



Şekil 3.2. Tek değişkenli dağılımlarda *Cars93* veri seti için merkezi eğilim ölçüleri ortalama ve çeyrekliklerin karşılaştırılması.

3.2. Koşullu Ortalama ve Koşullu Kantiller

Bağımlı değişken Y ve onu açıklamada kullanılacak bağımsız değişkenlerin kümesi X olmak üzere ortalama fonksiyonunun tanımında kullanılan Eşitlik (3.1) koşullu ortalama fonksiyonunun tanımı için

$$\hat{\mu}(\mathbf{x}_i, \boldsymbol{\beta}) = \underset{\mu}{\operatorname{argmin}} E(Y - \mu(\mathbf{x}_i, \boldsymbol{\beta}))^2 \quad (3.12)$$

şeklinde genelleştirilebilir. Eşitlikte yer alan $\mu(\mathbf{x}_i, \boldsymbol{\beta}) = E(Y|\mathbf{X} = \mathbf{x}_i)$ koşullu ortalama fonksiyonudur. Eşitlik (3.12)'de yer alan koşullu ortalama fonksiyonunun doğrusal olması durumunda $\mu(\mathbf{x}_i, \boldsymbol{\beta}) = \mathbf{x}_i' \boldsymbol{\beta}$ olmak üzere

$$\hat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} E(Y - \mathbf{x}_i' \boldsymbol{\beta})^2 \quad (3.13)$$

doğrusal regresyon modelinde EKK tahminlerinin elde edildiği amaç fonksiyonuna dönüşür.

Benzer bir yaklaşımla kantillerin tanımında kullanılan Eşitlik (3.5) genel koşullu kantil fonksiyonunun tanımı için

$$\hat{q}_Y(\theta, \mathbf{X}) = \underset{Q_Y(\theta, \mathbf{X})}{\operatorname{argmin}} E[\rho_\theta(Y - Q_Y(\theta, \mathbf{X}))]^2 \quad (3.14)$$

genelleştirilebilir. Eşitlikte yer alan $Q_Y(\theta, \mathbf{X}) = Q_\theta(Y|\mathbf{X} = \mathbf{x})$ genel koşullu kantil fonksiyonudur. Benzer şekilde koşullu kantil fonksiyonunun doğrusal olması durumunda

$$\hat{\beta}(\theta) = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} E[\rho_\theta(Y - \mathbf{X}\boldsymbol{\beta})] \quad (3.15)$$

doğrusal regresyon modelinde KR da θ . kantil için parametre tahminlerinin elde edildiği amaç fonksiyonuna dönüşür. Eşitlikte yer alan $\hat{\beta}(\theta)$ gösterimi θ . kantil için parametre tahmin vektörünü ifade etmektedir.

3.3. Basit Doğrusal Kantil Regresyon Modeli

Çarpık dağılımlarda medyan, ortalamaya göre daha uygun bir merkezi eğilim ölçüsüdür. Bu nedenle, konum değişimlerinin modellenmesinde koşullu ortalamayı temel alan klasik doğrusal regresyon yerine koşullu medyanyı temel alan medyan regresyonu kullanılmalıdır.

Koşullu medyan fonksiyonunu temel alan medyan regresyonu ilk olarak Boscovich tarafından 18. yüzyılın ortalarında önerilmiş ve daha sonra Laplace ve Edgeworth tarafından geliştirilmiştir. Medyan regresyonu bağımsız değişkenlerin değerlerindeki değişimin koşullu medyan üzerindeki etkisini tahmin eder. Bu nedenle dağılım çarpık olsa bile merkezi konumu iyi bir şekilde temsil eder.

Koenker ve Bassett (1978) hem konum hem de şekil değişimlerini modellemek için medyan regresyon modelinin daha genel bir formu olan ve koşullu kantilleri temel alan kantil regresyon modelini önerdiler.

Bir bağımlı ve bir bağımsız değişken arasındaki basit doğrusal kantil regresyon modeli (KRM)

$$y_i = \beta_0^{(\theta)} + \beta_1^{(\theta)} x_i + \varepsilon_i^{(\theta)} \quad (3.16)$$

eşitliği ile gösterilir. Eşitlikte yer alan θ , $0 < \theta < 1$ olmak üzere θ . kantili göstermektedir. Klasik doğrusal regresyon analizinde hata teriminin sıfır ortalamaya sahip olması varsayımı altında x_i verildiğinde y_i 'nin koşullu ortalaması $E(y_i|x_i) = \beta_0 + \beta_1 x_i$ olur. Benzer şekilde KR analizinde x_i verildiğinde y_i 'nin koşullu θ . Kantili

$$Q^{(\theta)}(y_i|x_i) = \beta_0^{(\theta)} + \beta_1^{(\theta)} x_i \quad (3.17)$$

şeklinde tanımlanır. Eşitlik (3.17) x_i 'nin belirli bir değerine karşı gelen θ . kantil değerinin $\beta_0^{(\theta)}$ ve $\beta_1^{(\theta)}$ parametreleri yardımıyla tanımlandığını ifade etmektedir. $\beta_1^{(\theta)}$ parametresi x_i 'nin değeri bir birim değiştiğinde θ . kantil değerinde meydana gelecek değişimi ifade eder. KR analizinde $Q^{(\theta)}(y_i|x_i) = \beta_0^{(\theta)} + \beta_1^{(\theta)} x_i + Q^{(\theta)}(\varepsilon_i^{(\theta)}) = \beta_0^{(\theta)} + \beta_1^{(\theta)} x_i$ eşitliğinin sağlanması için $\beta_0^{(\theta)} + \beta_1^{(\theta)} x_i$ sabit olduğundan dolayı hata teriminin θ . kantil değerinin 0 olması gerekir.

Wagner (1959), en küçük mutlak sapma regresyonunun amaç fonksiyonunun doğrusal bir programlama tekniği olarak formüle edilebileceğini ve uygun yöntem ve algoritmalarla yararlanarak etkin bir şekilde çözüldüğünü kanıtladı. Koenker ve Bassett (1978), koşullu kantillerin, kantillerin asimetric fonksiyonları gibi ağırlıklar kullanılarak

mutlak sapmaların ağırlıklı toplamını en aza indiren bir optimizasyon işlevi ile tahmin edilebileceğini gösterdi.

Basit doğrusal KRM nde $\theta = 0.50$ için parametrelerin tahmininde L_1 kriterine göre veriyi temsil edecek en uygun doğruyu belirleme başka bir ifade ile parametrelerin tahmini Eşitlik (3.18)'de tanımlanan minimizasyon probleminin çözümüne karşılık gelmektedir.

$$\min_{\beta_0^{(0.5)} \beta_1^{(0.5)}} \sum_{i=1}^n \left| \beta_0^{(0.5)} + \beta_1^{(0.5)} x_i - y_i \right| \quad (3.18)$$

Eşitlik (3.18)'de tanımlanan amaç fonksiyonu doğrusal olmasa da basit bir düzenleme ile doğrusal hale getirilir.

$$\begin{aligned} \text{Amaç Fonksiyonu} & : z_{min} = \sum_{i=1}^n \varepsilon_i^{(0.5)} \\ \text{Kısıtlar} & : \varepsilon_i^{(0.5)} \geq \beta_0^{(0.5)} + \beta_1^{(0.5)} x_i - y_i \quad i = 1, \dots, n \\ & \varepsilon_i^{(0.5)} \geq -(\beta_0^{(0.5)} + \beta_1^{(0.5)} x_i - y_i) \quad i = 1, \dots, n \end{aligned}$$

Yukarıda tanımlı doğrusal programlama modelinde kısıtlar $i = 1, \dots, n$ için

$$\begin{aligned} \varepsilon_i^{(0.5)} & \geq \max \left\{ \beta_0^{(0.5)} + \beta_1^{(0.5)} x_i - y_i, -(\beta_0^{(0.5)} + \beta_1^{(0.5)} x_i - y_i) \right\} \\ & = \left| \beta_0^{(0.5)} + \beta_1^{(0.5)} x_i - y_i \right| \end{aligned}$$

şeklinde de gösterilebilir. Tanımlanan doğrusal programlama modelinin simpleks algoritması veya farklı optimizasyon algoritmalarıyla çözümünden $\beta_0^{(0.5)}$ ve $\beta_1^{(0.5)}$ parametrelerinin tahmini elde edilir.

Basit doğrusal KRM nde parametrelerin tahmininde doğrusal programlama modelinin oluşturulmasına örnek olması için *cars93* veri setinde Chevrolet marka sekiz farklı model araçlarda beygir gücü ile fiyattaki değişimini açıklamaya çalışalım. Oluşturulan örnek veri seti Tablo 3.1'de verilmiştir.

Tablo 3.1. Chevrolet marka sekiz farklı model araçlarda beygir gücü ve fiyatlar

Model	Cavalier	Corsica	Camaro	Lumina	Lumina_APV	Astro	Caprice	Corvette
Fiyat (y_i)	13.4	11.4	15.1	15.9	16.3	16.6	18.8	38
Beygir gücü (x_i)	110	110	160	110	170	165	170	300

Basit doğrusal KRM nde parametrelerin tahmini için doğrusal programlama modeli

Amaç Fonksiyonu: $z_{min} = \sum_{i=1}^8 \varepsilon_i^{(0.5)}$

Kısıtlar: $-\beta_0^{(0.5)} - 110\beta_1^{(0.5)} - \varepsilon_1^{(0.5)} \leq -13.4$

$$-\beta_0^{(0.5)} - 110\beta_1^{(0.5)} - \varepsilon_2^{(0.5)} \leq -11.4$$

$$-\beta_0^{(0.5)} - 160\beta_1^{(0.5)} - \varepsilon_3^{(0.5)} \leq -15.1$$

$$-\beta_0^{(0.5)} - 110\beta_1^{(0.5)} - \varepsilon_4^{(0.5)} \leq -15.9$$

$$-\beta_0^{(0.5)} - 170\beta_1^{(0.5)} - \varepsilon_5^{(0.5)} \leq -16.3$$

$$-\beta_0^{(0.5)} - 165\beta_1^{(0.5)} - \varepsilon_6^{(0.5)} \leq -16.6$$

$$-\beta_0^{(0.5)} - 170\beta_1^{(0.5)} - \varepsilon_7^{(0.5)} \leq -18.8$$

$$-\beta_0^{(0.5)} - 300\beta_1^{(0.5)} - \varepsilon_8^{(0.5)} \leq -38.0$$

$$\beta_0^{(0.5)} + 110\beta_1^{(0.5)} - \varepsilon_1^{(0.5)} \leq 13.4$$

$$\beta_0^{(0.5)} + 110\beta_1^{(0.5)} - \varepsilon_2^{(0.5)} \leq 11.4$$

$$\beta_0^{(0.5)} + 160\beta_1^{(0.5)} - \varepsilon_3^{(0.5)} \leq 15.1$$

$$\beta_0^{(0.5)} + 110\beta_1^{(0.5)} - \varepsilon_4^{(0.5)} \leq 15.9$$

$$\beta_0^{(0.5)} + 170\beta_1^{(0.5)} - \varepsilon_5^{(0.5)} \leq 16.3$$

$$\beta_0^{(0.5)} + 165\beta_1^{(0.5)} - \varepsilon_6^{(0.5)} \leq 16.6$$

$$\beta_0^{(0.5)} + 170\beta_1^{(0.5)} - \varepsilon_7^{(0.5)} \leq 18.8$$

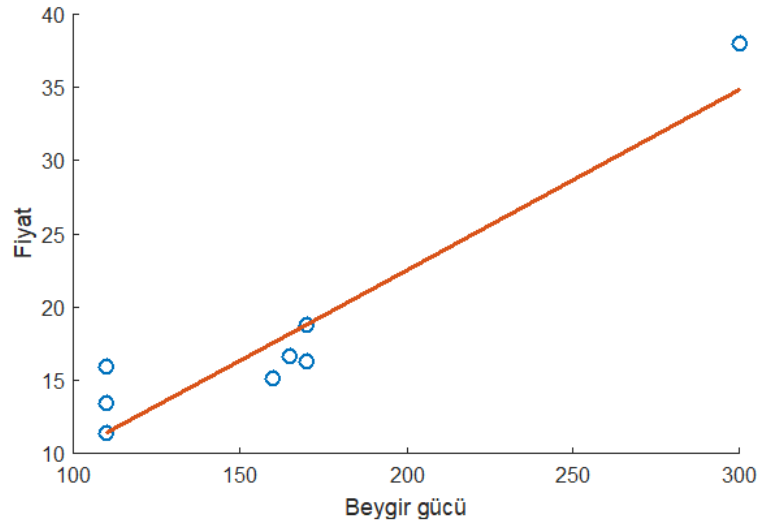
$$\beta_0^{(0.5)} + 300\beta_1^{(0.5)} - \varepsilon_8^{(0.5)} \leq 38.0$$

$$\varepsilon_1^{(0.5)}, \varepsilon_2^{(0.5)}, \varepsilon_3^{(0.5)}, \varepsilon_4^{(0.5)}, \varepsilon_5^{(0.5)}, \varepsilon_6^{(0.5)}, \varepsilon_7^{(0.5)}, \varepsilon_8^{(0.5)} \geq 0.$$

şeklinde oluşturulur. Doğrusal programlama modelinin çözümü ile

$$\hat{\beta}^{(0.5)} = \begin{bmatrix} -2.1667 \\ 0.1233 \end{bmatrix}$$

elde edilir. Örnek veri seti için saçılım grafiği ve tahmini basit doğrusal KRM Şekil 3.3'te verilmiştir.

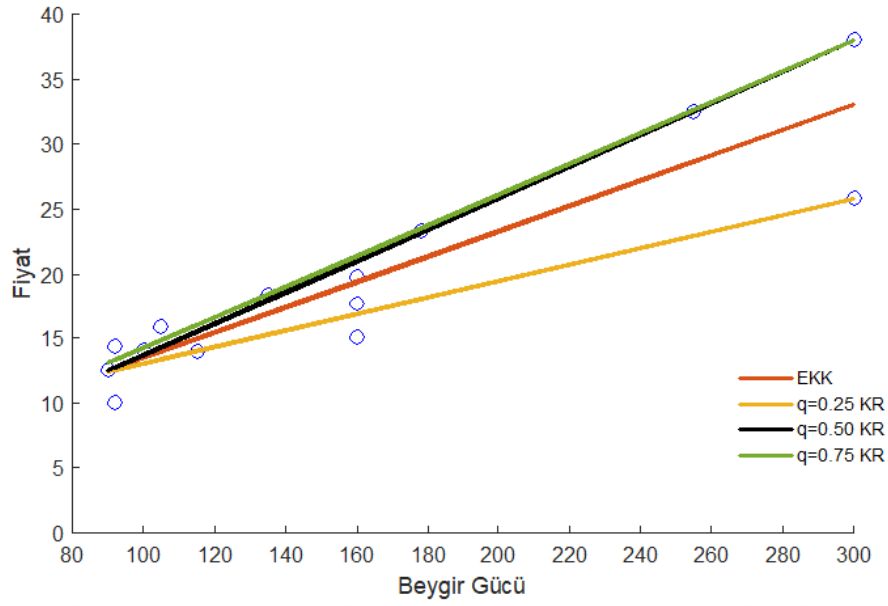


Şekil 3.3. Chevrolet veri seti için saçılım grafiği ve $\theta = 0.5$ için tahmini basit doğrusal KRM

Cars93 veri setinde sportif araçlar için fiyat değişkeni bağımlı değişken, beygir gücü bağımsız değişken olarak alındığında EKK regresyonu ve $\theta = 0.25, 0.50, 0.75$ kantil değerleri ile KR dan elde edilen regresyon doğrularının karşılaştırılması Şekil 3.4'te verilmiştir. Örnek veri seti ile EKK regresyonu ve $\theta = 0.25, 0.50, 0.75$ kantil değerleri ile KR için parametre tahminleri Tablo 3.2'de verilmiştir.

Tablo 3.2. Örnek veri seti ile EKK regresyonu ve $\theta = 0.25, 0.50, 0.75$ için KRM nde parametre tahminleri

Beygir Gücü	160	300	300	105	115	90	160	92	255	100	92	160	135	178
Fiyat	15.1	38	25.8	15.9	14	12.5	19.8	10	32.5	14.1	14.4	17.7	18.4	23.3
	EKK				$\theta = 0.25$ KR				$\theta = 0.50$ KR				$\theta = 0.75$ KR	
$\hat{\beta}_0$	3.6940				6.6649				1.5909				2.3637	
$\hat{\beta}_1$	0.0980				0.0638				0.1212				0.1188	



Şekil 3.4. EKK regresyonu ve 0.25, 0.50 ve 0.75. kantil değerleri ile KR dan elde edilen regresyon doğruları

3.4. Çoklu Doğrusal Kantil Regresyon Modeli

Çoklu doğrusal KRM

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta}(\theta) + \epsilon \quad (3.19)$$

eşitliği ile tanımlanır. Modelde yer alan $\mathbf{y}$ bağımlı değişkene ait $n \times 1$ boyutlu gözlem vektörü, $\mathbf{X}$ bağımsız değişkenlerin gözlemlenen değerlerinden oluşan $n \times p$ boyutlu tasarım matrisi, $\boldsymbol{\beta}(\theta)$ θ . koşullu kantil için $p \times 1$ bilinmeyen parametreler vektörü ve ϵ , $n \times 1$ boyutlu hata vektörünü göstermektedir.

Medyan regresyonu $\theta = 0.5$ için kantil regresyonunun özel bir halidir. Medyan regresyonu en küçük sapmalar regresyonu olarak da ifade edilir. Medyan regresyonunda koşullu medyan fonksiyonunun bilinmeyen parametrelerinin tahmini

$$\min_{\boldsymbol{\beta}} \sum_{i=1}^n |y_i - x_i \boldsymbol{\beta}| \quad (3.20)$$

minimizasyon probleminin çözümünden elde edilir. Negatif olmayan farklar için

$$\begin{aligned} s_1 &= |\mathbf{y} - \mathbf{X}\boldsymbol{\beta}|_+ \\ s_2 &= |\mathbf{X}\boldsymbol{\beta} - \mathbf{y}|_+ \end{aligned} \quad (3.21)$$

tanımlamaları yapıldığında Eşitlik (3.20)'deki minimizasyon problemi

$$\min_{\beta} \{ \mathbf{1}' \mathbf{s}_1 + \mathbf{1}' \mathbf{s}_2 \mid \mathbf{y} = \mathbf{X}\beta + \mathbf{s}_1 - \mathbf{s}_2, \{\mathbf{s}_1, \mathbf{s}_2\} \in \mathbb{R}_+^n \}$$

$$\mathbf{B} = |\mathbf{X} - \mathbf{X}\mathbf{I} - \mathbf{I}|$$

$$\psi = \begin{bmatrix} [\beta]_+ \\ [-\beta]_+ \\ |\mathbf{y} - \mathbf{X}\beta|_+ \\ |\mathbf{X}\beta - \mathbf{y}|_+ \end{bmatrix}$$

$$\mathbf{d} = \begin{bmatrix} \mathbf{0}_p \\ \mathbf{0}_p \\ \mathbf{1}_n \\ \mathbf{1}_n \end{bmatrix}$$

şeklinde formüle edilebilir. Yeniden formülleme ile minimizasyon problemi standart bir doğrusal programlama problemine dönüşür. Böyle bir doğrusal programlama probleminin primal formülasyonu

$$\min_{\psi} \mathbf{d}'\psi$$

$$\text{Kısıtlar: } \mathbf{B}\psi = \mathbf{y}$$

$$\theta \geq 0$$

ve dual karşılığı

$$\max_{\mathbf{d}} \mathbf{y}'\mathbf{z}$$

$$\text{Kısıtlar: } \mathbf{B}'\mathbf{z} \leq \mathbf{d}$$

şeklinde oluşturulur.

Doğrusal programlamanın temel sonucunu akılda bulundurarak, bu tür bir küçültme probleminin çözümlerinin simpleksin köşelerinde basit bir pozisyonla aranması gerektiği teoremi olarak göz önüne alındığında, yukarıdaki problem aşağıdaki gibi yeniden yapılandırılabilir:

$$\max_{\mathbf{z}} \{ \mathbf{y}'\mathbf{z} \mid \mathbf{X}'\mathbf{z} = \mathbf{0}, \mathbf{z} \in [-1, +1]^n \}$$

Gerçekte bu eşitlik,

$$\mathbf{X}'\mathbf{z} = \mathbf{0}$$

şeklinde tanımlıdır. Eşitliğin her iki tarafı 1/2 ile çarpılır

$$\frac{1}{2}\mathbf{X}'\mathbf{z} = \mathbf{0}$$

ve eşitliğin her iki tarafına $\frac{1}{2}\mathbf{X}'\mathbf{1}$ eklenirse

$$\frac{1}{2}\mathbf{X}'\mathbf{z} + \frac{1}{2}\mathbf{X}'\mathbf{1} = \frac{1}{2}\mathbf{X}'\mathbf{1}$$

$$\mathbf{X}' \underbrace{\left(\frac{1}{2}\mathbf{z} + \frac{1}{2}\mathbf{1}\right)}_{\boldsymbol{\eta}} = \underbrace{\frac{1}{2}\mathbf{X}'\mathbf{1}}_{\mathbf{b}} \quad (3.22)$$

formülasyonu elde edilir. Dual problemdeki eşitlik

$$\max_{\mathbf{J}} \{\mathbf{y}'\mathbf{J} | \mathbf{X}'\mathbf{J} = \mathbf{b}, \mathbf{J} \in [0,1]^n\}$$

şeklinde gösterilir.

Eşitlik (3.22)'de $1/2$ teriminin görevi nötr gibi görünse de kantil regresyonundaki genellemenin anahtarıdır. Kantil regresyonundaki amaç fonksiyonunun

$$\min_{\boldsymbol{\beta}(\theta)} \sum_{i=1}^n \rho_{\theta}(y_i - \mathbf{x}_i\boldsymbol{\beta}(\theta)) \quad (3.23)$$

dual yapıdaki formülasyonu benzer adımları gerçekleştirerek

$$\max_{\mathbf{z}} \{\mathbf{y}'\mathbf{z} | \mathbf{X}'\mathbf{z} = (1 - \theta)\mathbf{X}'\mathbf{1}, \quad \mathbf{z} \in [0,1]^n\}$$

şeklinde elde edilir. Eşitlikte yer alan $(1 - \theta)$ terimi medyan regresyonunda ki $1/2$ terimi ile aynı işlevi görür.

3.5. Kantil Regresyonun Özellikleri

- i. EKK regresyonu koşullu ortalama fonksiyonunu kullanırken KR yönteminde koşullu kantil fonksiyonu kullanılır.
- ii. KR yönteminde $\min_{b \in \mathbb{R}^p} \sum_{i=1}^n \rho_{\theta}(y_i - x_i^T b)$ minimizasyonu, doğrusal programlama (LP) şeklindedir ve bu yaklaşım tahmini kolaylaştırmaktadır (Cade ve Richards, 2006)
- iii. Kantiller monoton dönüşümlere imkân verir. Herhangi bir $h(\cdot)$ monoton fonksiyonu için $Q_{h(y)/x}(\theta/x) = h(Q_{h(y)/x}(\theta/x))$ şeklindedir.
- iv. Kantiller y 'deki aşırı değerlere karşı sağlam (robust) olduğundan dolayı aykırı değer durumunda alternatif olarak kullanılan KR yöntemi sağlam istatistiksel bir yöntemdir.
- v. Hata teriminin dağılımı normal dağılım olmadığında, KR tahminleri EKK tahminlerine göre daha etkin olabilir.
- vi. Farklı varyans probleminde EKK tahminleri güvenilir değildir. Bu durumda, EKK yöntemine alternatif olarak KR yöntemi kullanılmaktadır (John ve Nduka, 2009).
- vii. KR yönteminin amaç fonksiyonu için elde edilecek katsayı vektörü, bağımlı değişkendeki aykırı değerlere karşı duyarlı değildir.
- viii. Farklı kantillerden elde edilen sonuçların farklı olması, bağımlı değişkeninin koşullu dağılım fonksiyonunun farklı kantil değerlerinde açıklayıcı değişkenlerdeki değişimlere farklı tepki vermesi şeklinde yorumlanabilir (Saçaklı, 2005).
- ix. Sağlam istatistiksel yöntem olan L-tahminleri KR tahminlerinin doğrusal kombinasyonuna dayanmaktadır ve EKK tahminlerinden genel olarak daha etkilidirler (Buchinsky, 1998).
- x. KR analizinde $\theta = 0.5$ olması hâlinde Medyan Regresyon (LAD) analizi elde edilmektedir (Koenker ve Bassett, 1978).
- xi. EKK regresyonunda koşullu ortalama, LAD yöntemlerinde koşullu medyan fonksiyonlarının tahmini gerçekleştirilirken, KR'nda ise bağımlı değişkenin dağılımının rastgele bir noktasındaki kantil değerleri için koşullu kantil fonksiyonu kullanılarak parametre tahminleri gerçekleştirilmektedir (Zietz ve ark., 2008).

3.6. Simülasyon Çalışması

Kantil regresyonu aykırı gözlem veya hataların normal dağılıma sahip olmaması durumlarında EKK regresyonuna alternatif olarak kullanılan sağlam bir istatistiksel yöntem olarak bilinmektedir. Bu bölümde aykırı gözlem, hataların lognormal dağılımına sahip olması durumlarında $\theta = 0.25, 0.50$ ve 0.75 değerleri için KR ile EKK regresyonu MSE kriterine göre karşılaştırılacaktır.

$$MSE = \frac{1}{10000} \sum_{i=1}^{10000} \sum_{j=1}^4 (\beta_j - \hat{\beta}_{ij})^2 \quad (3.24)$$

Simülasyon çalışmasında bağımsız değişken sayısı $p = 4$ alınmıştır. Bağımsız değişken değerleri $\mu = [0 \ 0 \ 0 \ 0]$ ve $\Sigma = \mathbf{I}_{4 \times 4}$ olmak üzere çok değişkenli normal dağılımdan üretilmiştir. Bilinmeyen parametre vektörü β , $(X'X)$ matrisinin en büyük özdeğerine karşılık gelen öz vektörden oluşturulmuştur. Bağımlı değişken değerleri

$$y = X\beta + \varepsilon \quad (3.25)$$

eşitliğine göre oluşturulmuştur. Hata terimi ε 'nin üretilmesinde iki farklı durum göz önünde bulundurulmuştur.

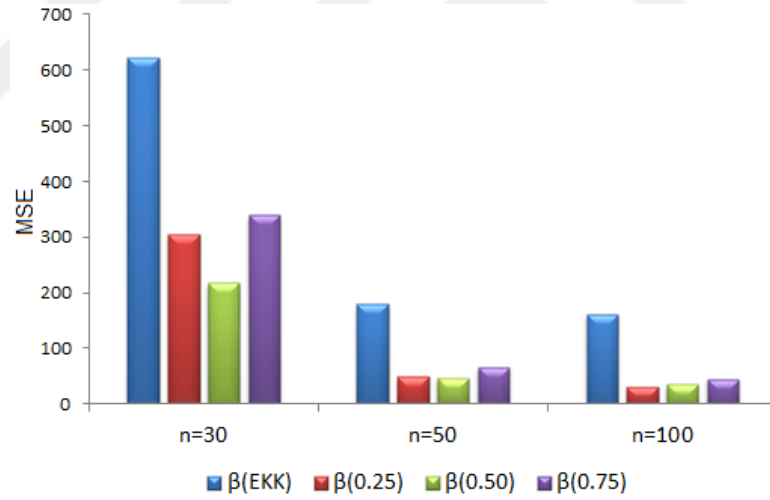
Durum I: Aykırı gözlem durumunda.

Burada hata terimi öncelikle $\varepsilon \sim N(0,1)$ olacak şekilde üretilmiştir. Aykırı gözlem oluşturmak için bağımlı değişken değerleri arasından rastgele seçilen iki gözlem değeri $y_i^* = y_i + 5$ ile aykırı gözlem haline getirilmiştir. Simülasyon çalışması örneklem hacmi 30, 50 ve 100 olmak üzere 10000 tekrar ile gerçekleştirilmiştir.

Farklı örneklem hacimlerinde EKK ile $\theta = 0.25, 0.50$ ve 0.75 değerleri için KR nda elde edilen parametre tahminleri ve MSE değerleri Tablo 3.3'te verilmiştir. EKK ve KR nda elde edilen MSE değerlerinin grafiksel karşılaştırılması Şekil 3.5'te verilmiştir.

Tablo 3.3. Farklı örneklem hacimlerinde EKK ve KR nda elde edilen parametre tahminleri ve MSE değerleri

	β_{EKK}	$\beta_{0.25}$	$\beta_{0.50}$	$\beta_{0.75}$
$n = 30$	-50.2691	-26.4230	-20.9376	-30.5293
	-0.1500	-0.0508	0.0204	-0.0608
	0.0472	-0.0148	-0.0007	0.0218
	-0.0129	-0.0567	-0.0798	-0.0443
	1.5182	1.2648	1.2181	1.3305
MSE	621.7814	305.3702	216.8449	339.3751
$n = 50$	-26.3134	-7.3754	-6.8885	-9.1342
	0.0817	-0.0330	-0.0777	-0.0686
	0.0980	0.0918	0.0827	0.0783
	-0.1448	-0.1190	-0.0993	-0.1087
	1.2592	1.0633	1.0719	1.1043
MSE	179.1727	49.2960	45.6388	65.6220
$n = 100$	-25.5926	-5.4619	-5.8303	-7.5545
	-0.0744	-0.0602	-0.0156	-0.0187
	0.0642	0.0699	0.0747	0.0806
	-0.1535	-0.1105	-0.0957	-0.1152
	1.2540	1.0466	1.0613	1.0867
MSE	161.6757	31.9075	34.9732	44.8052

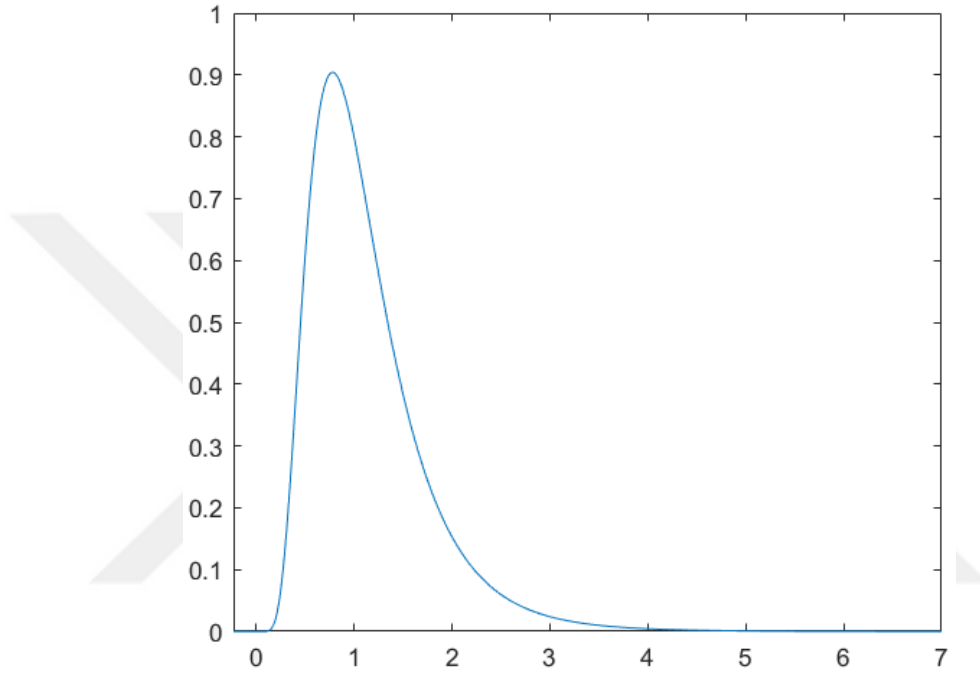


Şekil 3.5. Aykırı gözlem durumunda EKK ve KR nun MSE kriterine göre karşılaştırılması

Tablo 3.3 ve Şekil 3.5 incelendiğinde aykırı gözlem durumunda MSE kriterine göre KR yönteminin EKK yöntemine göre daha başarılı olduğu görülmüştür. Örneklem hacminin 30 ve 50 olduğu durumlarda en küçük MSE değeri $\theta = 0.50$ kantil değeri için elde edilmiştir. Örneklem hacminin 100 olduğu durumda ise en küçük MSE değeri $\theta = 0.25$ kantil değeri için elde edilmiştir.

Durum II: Hataların lognormal dağılımına sahip olması.

Bu simülasyon çalışmasında hatalar $\mu = 0$, $\sigma = 0.5$ parametreleri ile Lognormal dağılımından üretilmiştir. Hatalar sağa çarpık bir dağılım göstermektedir. $\mu = 0$ ve $\sigma = 0.5$ parametreleri ile Lognormal dağılıma ait olasılık yoğunluk fonksiyonu eğrisi Şekil 3.6'da verilmiştir.



Şekil 3. 6. $\mu = 0$ ve $\sigma = 0.5$ parametreleri ile Lognormal dağılıma ait olasılık yoğunluk fonksiyonu eğrisi

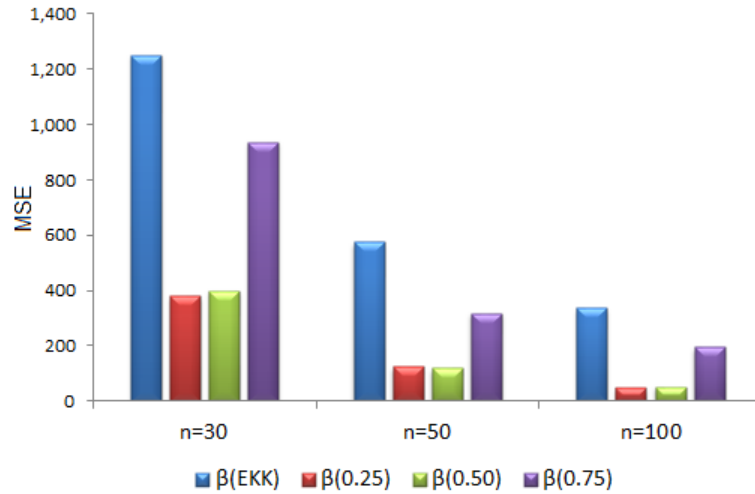
Örneklem hacmi 30, 50 ve 100 olacak şekilde 10000 tekrar ile gerçekleştirilen simülasyon çalışmasından elde edilen sonuçlar Tablo 3.4'te verilmiştir. Farklı örneklem hacimlerine göre incelenen regresyon yöntemleri ile elde edilen MSE değerlerine ait grafiksel gösterim Şekil 3.7'de verilmiştir.

Tablo 3.4 ve Şekil 3.7 incelendiğinde beklenildiği gibi örneklem hacmi arttıkça MSE değerlerinin azaldığı görülmüştür. Hataların sağa çarpık bir dağılıma sahip olması durumunda EKK yönteminin KR yöntemine göre daha büyük MSE değerlerine sahip olduğu görülmüştür. KR yönteminde 30 ve 100 örneklem hacimleri için en küçük MSE değeri $\theta = 0.25$ koşullu kantil değeri için elde edilirken 50 örneklem hacminde en küçük MSE değeri $\theta = 0.50$ koşullu kantil değeri için elde edilmiştir. Hataların normal dağılıma

sahip olmaması durumunda KR yöntemi EKK yönteminden MSE kriterine göre daha başarılı bulunmuştur.

Tablo 3.4. Hataların lognormal dağılıma sahip olması durumunda farklı örneklem hacimlerinde EKK ve KR nda elde edilen parametre tahminleri ve MSE değerleri

	β_{EKK}	$\beta_{0.25}$	$\beta_{0.50}$	$\beta_{0.75}$
$n = 30$	3.8569	0.1836	1.0882	2.6390
	0.0077	0.0567	0.0595	0.0804
	0.0497	0.0251	0.0109	-0.0333
	-0.1168	-0.1064	-0.1166	-0.1544
	0.9943	1.0062	1.0072	1.0123
MSE	1251	379.8436	399.9454	936.9336
$n = 50$	4.2162	1.4569	2.4786	4.0453
	-0.0261	0.0385	0.0153	-0.0356
	0.0609	0.0419	0.0389	0.0221
	-0.1002	-0.1012	-0.0985	-0.0947
	0.9921	0.9931	0.9923	0.9989
MSE	574.543	125.5643	118.3894	317.2325
$n = 100$	4.1547	1.0040	2.6131	3.0309
	0.0098	0.0234	0.0266	0.0281
	0.0543	0.0568	0.0518	0.0549
	-0.1019	-0.0886	-0.0956	-0.1050
	0.9925	0.9983	0.9892	1.0029
MSE	335.7397	49.2302	52.8745	194.3076



Şekil 3.7. Hataların lognormal dağılıma sahip olması durumunda EKK ve KR nun MSE kriterine göre karşılaştırılması

4. RIDGE REGRESYONU

4.1. Giriş

Çoklu regresyon analizinde bağımsız (açıklayıcı) değişkenler arasında anlamlı istatistiksel ilişkiler olması çoklu doğrusal bağıntı problemi olarak ifade edilir. Çoklu doğrusal bağıntı probleminde EKK tahminleri en küçük varyansa sahip olma özelliklerini yitirirler ve gerçekleştirilen istatistiksel testler anlamsız sonuçlar üretebilirler. Bu durumda EKK yöntemine alternatif olarak yanlı tahmin yöntemleri kullanılabilir. Çoklu doğrusal bağlantı durumunda en yaygın kullanılan yanlı tahmin yöntemi ridge tahmin yöntemidir. Bu bölümde öncelikle Birkes ve Dodge (2011) temel alınarak ridge regresyonunun EKK tahmininden nasıl daha doğru tahminler elde ettiği anlatılacak ve daha sonrasında ridge regresyonunun teorik özellikleri gösterilecektir.

Ridge regresyonu Arthur Hoerl (1962) tarafından kimya mühendisliği dergisindeki bir makalede yayınlandı. Arthur Hoerl (1962), regresyon analizinde açıklayıcı değişkenler arasında ilişki olması durumunda EKK tahminlerinin genellikle veriyi oluşturan süreç bağlamında anlamlı olmadığını ifade etmiş ve daha iyi tahminler elde etmek için bir yöntem önermiştir.

Tabii ki, sadece bir açıklayıcı değişkenli basit regresyonda, açıklayıcı değişkenler arasındaki ilişki ile ilgili bir problem yoktur, ancak yine de, ridge regresyonunun nasıl daha doğru tahminler üretmeye çalıştığını açıklamak için basit regresyon örneğine bakmak öğreticidir.

Ridge regresyonunda ilk adım genellikle açıklayıcı değişkenlerin standartlaştırılmasıdır. Standartlaştırma üç şekilde faydalı olabilir. İlk olarak, $X'X$ matrisinin ters çevrilmesinde yuvarlama hatasını azaltır. İkincisi, değişkenlerin ve regresyon katsayılarının yorumlanabilirliğini artırabilir. Üçüncüsü, standartlaştırma farklı değişkenlerin regresyon katsayılarını doğrudan karşılaştırma imkânı verir. Y : akciğer kapasitesi (cl), X_1 : boy uzunluğu (cm) ve X_2 : ağırlık (kg) olmak üzere tahmini regresyon modeli $\hat{Y} = -171 + 1.92X_1 + 0.286X_2$ olsun. Ölçü birimleri farklı olduğundan dolayı 1.92 ve 0.286 regresyon katsayı tahminlerini karşılaştırmak doğru değildir. Açıklayıcı değişkenler standartlaştırıldığında, Z_1 ve Z_2 standardize edilmiş değişkenleri göstermek üzere tahmini regresyon modeli $\hat{Y} = 192 + 12.9Z_1 + 3.28Z_2$ olsun. Standartlaştırma

sonrası 12.9 ve 3.28 regresyon katsayı tahminleri karşılaştırılabilir ve akciğer kapasitesindeki farklılıkların, ağırlıktaki farklılıklardan ziyade boydaki farklılıklardan kaynaklandığı sonucuna varılabilir.

X değişkenini standartlaştırmak için $\bar{x}$ ve s_x sırasıyla x_i 'lerin ortalaması ve standart sapması olmak üzere $z_i = (x_i - \bar{x})/s_x$ eşitliği kullanılır.

Standartlaştırılmamış X değerleri ile elde edilen basit doğrusal regresyon modeli $Y = \alpha + \beta X + e$ eşitliği ile gösterilir. Standartlaştırma sonrası basit doğrusal regresyon modelini $Y = \mu + \gamma Z + e$ eşitliği ile gösterelim.

Regresyon katsayıları μ ve γ 'nın EKK tahminleri, z_i 'ler standartlaştırıldığı için $\bar{z} = 0$ olur ve formüller basitleşir ve $\hat{\gamma}_{EKK} = \frac{\sum z_i y_i}{\sum z_i^2}$, $\hat{\mu}_{EKK} = \bar{y}$ olarak elde edilir.

Ridge tahminleri için $\hat{\mu} = \bar{y}$ alınır ve

$$\hat{\gamma} = \frac{\sum z_i y_i}{\sum z_i^2 + k} \quad (4.1)$$

olur. Burada $k = \hat{\sigma}_{EKK}^2 / \hat{\gamma}_{EKK}^2$ ve $\hat{\sigma}_{EKK}$ EKK regresyonunda σ 'nın tahminini göstermektedir.

Hata Kareler Ortalaması. Eşitlik (4.1)'in $\hat{\gamma}_{EKK}$ 'dan nasıl daha doğru bir tahmin üretebileceğini açıklamadan önce, “doğru” ile ne ifade edilmek istendiğinin belirtmesi gerekir. Eğer $\hat{\gamma}$ doğru ise bir rassal değişken olarak düşünüldüğünde γ 'ya yakın olma eğilimindedir yani γ ile $\hat{\gamma}$ arasındaki uzaklık küçük olma eğilimindedir. Bir $\hat{\gamma}$ tahmin değerinin doğruluğunu ölçmenin en yaygın yolu γ ile $\hat{\gamma}$ arasındaki uzaklığın karesinin beklenen değeri olan hata kareler ortalamasıdır $MSE(\hat{\gamma}) = E((\gamma - \hat{\gamma})^2)$. $MSE(\hat{\gamma})$ değeri ne kadar küçük olursa $\hat{\gamma}$ o kadar doğru olur.

Hata kareler ortalaması MSE 'nin kullanışlı bir açılımı Eşitlik (4.2)'de verilmiştir.

$$MSE(\hat{\gamma}) = Var(\hat{\gamma}) + (E(\hat{\gamma}) - \gamma)^2 \quad (4.2)$$

Yani, $\hat{\gamma}$ 'nın hata kareler ortalaması $\hat{\gamma}$ 'nın varyansı ile tahminin yanının karesinin toplamıdır. Bu nedenle, tahminin hata kareler ortalamasının küçük olması için tahminin, eşzamanlı olarak küçük varyansa ve küçük yanlılığa sahip olması gerekir.

EKK tahmini $\hat{\gamma}_{EKK}$ yansızdır yani $E(\hat{\gamma}) = \gamma$ 'dır. Ridge tahmini yanlı bir tahmin edicidir fakat daha küçük varyansa sahiptir. Kare yanlılığındaki artışın, varyanstaki azalmadan daha az olması beklenir, böylece ridge tahmini daha küçük hata kareler ortalamasına sahip olur.

Eşitlik (4.1)'in Gerekçesi. Ridge tahmini $\hat{\gamma}_{EKK}$ tahmin değerini "küçülterek" tahminin doğruluğunu iyileştirmeye çalışır. $\hat{\gamma}_{EKK}$ tahmin değerini küçültmek için 0 ile 1 arasında yer c sayısı ile çarpma işlemi yapılır, böylece $\hat{\gamma} = c\hat{\gamma}_{EKK}$ elde edilir. Bu işlem $|c\hat{\gamma}_{EKK}| < |\hat{\gamma}_{EKK}|$ olduğu için küçültme olarak isimlendirilir. $\hat{\gamma}_{EKK}$ 'nın hata kareler ortalaması yansız bir tahmin olduğu için $v = \sigma^2 / \sum z_i^2$ varyans değerine eşittir.

$$MSE(c\hat{\gamma}_{EKK}) = c^2v + (c - 1)^2\gamma^2 \quad (4.3)$$

En küçük hata kareler ortalaması değerinin elde edileceği c değeri, Eşitlik (8.3)'ün c 'ye göre türevini sıfıra eşitleyerek $c = \gamma^2 / (\gamma^2 + v)$ olarak elde edilir. Elde edilen eşitlikte bilinmeyen parametreler γ ve σ yerine onların EKK tahminleri $\hat{\gamma}_{EKK}$ ve $\hat{\sigma}_{EKK}$ yazılarak $\hat{c}$ elde edilir. Gerekli dönüşümler sonrası $\hat{c} = \sum z_i^2 / (\sum z_i^2 + \hat{\sigma}_{EKK}^2 / \hat{\gamma}_{EKK}^2)$ elde edilir ve $c\hat{\gamma}_{EKK}$ ridge tahmini Eşitlik (4.1)'deki ridge tahmini ile aynıdır.

Eşitlik (4.3) yalnızca c rasgele olmayan bir sayı ise geçerlidir. Rastgele bir c seçtiğimizde rastgele $\hat{\gamma}_{EKK}$ ve $\hat{\sigma}_{EKK}$ değerlerini içeren $\hat{\gamma} = c\hat{\gamma}_{EKK}$ 'nın hata kareler ortalamasını belirlemek zordur ve aslında rastgele hataların dağılımının şekline bağlıdır. Bununla birlikte, önceki paragrafta verilen tartışma en azından $\hat{\gamma}$ 'nın $\hat{\gamma}_{EKK}$ 'dan daha doğru olabileceğini göstermektedir. Ancak, yalnızca γ sıfıra yakınsa, $\hat{\gamma}$ tahmininin $\hat{\gamma}_{EKK}$ tahmininden daha doğru olduğu bulunmuştur.

4.2. Ridge Tahmininin Niteliği

Klasik çoklu doğrusal regresyon modelinde olayla ilgili açıklayıcı değişkenler, deney düzenlemesi etkisi olmadan veya yalnızca deney düzenleme yöntemiyle veriler elde edilse dahi bu sefer de olayın karakterindeki matematiksel ve fiziksel kısıtlar sebebiyle birbirleriyle ilişkili durumda olabilirler. Regresyon analizinde çoklu doğrusal bağıntının meydana gelmesi ile sorundaki sebep-sonuç ilişkisini sunan parametrelerin

tahmin edilmesiyle sonuçların duyarlı olması beklenemez. Şayet $X'X$ korelasyon matrisinin değerleri birim matrise yakın sonuçlu olursa EKK tahminleri bu durumda güvenilir sonuçlar verir. Lakin $X'X$ korelasyon matrisinin değerinin birim matris olamaması, EKK tahminlerinin Varyans Büyütme Faktörü (VIF) değerlerinin büyük olmasına neden olur. Bu yüzden parametre tahminlerinin hataları da fazlalaşmış olur.

Açıklayıcı değişkenlerin birbirleriyle olan ilişkileri, EKK tahminlerinin varyanslarını fazlalaştırmaktadır. Modeldeki değişkenlerin her biri için VIF değerleri; modelin katsayılarının varyansları üstündeki basit korelasyonların totaldeki etkisini göstermektedir. Çoklu doğrusal bağlantı olması hâlinde, korelasyon matrisinin tersinin köşegenleri, ayrı ayrı bütün tahminlerin diğer tahminlerle gerçekleşen çoklu korelasyonunda sonsuz durumda olur. EKK tahminleri bu hâlde yansız tahminler arasında en küçük varyansa sahip olma özelliklerini yitirirler (Marquardt ve Snee, 1975). Zira çoklu doğrusal bağıntı gerçek β ile $\hat{\beta}$ değerleri arasında sapmaya sebep olur. $\hat{\beta}$ 'nin β 'dan sapması olacak şekilde yanlılığın karesi,

$$L_1^2 = (\hat{\beta} - \beta)'(\hat{\beta} - \beta) \quad (4.4)$$

ifade edilirse, L_1^2 'nin beklenen değeri de

$$E(L_1^2) = \sigma^2 \text{İz}(X'X)^{-1} \quad (4.5)$$

şeklinde olur. İz burada, kare matrisin esas köşegenlerinin elemanlarının toplamıdır $X'X$ matrisinin öz değerleri $\lambda_1 > \lambda_2 > \dots > \lambda_k$ biçiminde gösterildiğinde, $\hat{\beta}$ 'nin β 'ya olan uzaklığının karesinin beklenen değeri;

$$E(L_1^2) = \sigma^2 \sum_{j=1}^k \lambda_j^{-1}$$

biçiminde elde edilir (Mahajan ve ark., 1977).

Çoklu doğrusal bağıntı varken EKK yönteminde tahminde bulunulduğunda, tahminlerin hata kareler toplamı, hesaplamasında öz değerlerden yararlanarak sağlayan yukarıdaki eşitlik, bir veya birden çok öz değerın sıfır ya da sıfıra yakın olması, $\hat{\beta}$ 'nin

gerçek değer β 'dan saplamalarının beklenen değerinin artmasına sebep olacaktır. Bu sebepler regresyon katsayılarının tahmin edilmesi için, açıklayıcı değişkenlerin birbirlerine olan etkilerini minimuma indirmek ve kararlı katsayı tahminleri bulmak için yanlı tahminlerden Ridge tahmini kullanılmalıdır.

4.3. Ridge Tahmininin Özellikleri

Ridge yöntemi için bugüne kadar birçok çalışma yapılmıştır. Bu yöntem EKK yöntemiyle gerçekleştirilen yön ve değerce istatistiksel bekleyişleri karşılamayan parametre tahminlerinde bulunduğu çoklu doğrusal bağlantı sorununun olduğu hâllerde yararlı ve uygun olan bir yöntemdir.

Yanlı tahminlerde bulunmasına karşın varyansı küçültmesinden dolayı tercih sebebi olan Ridge tahmininin önerilme amaçları şunlardır:

- i. Çoklu regresyon modelinde birbirleriyle bağıntılı olan açıklayıcı değişkenlerin en küçük kareler tahminlerinden daha az varyanslı β tahminlerinin bulunmasında;
- ii. Güçlü çoklu doğrusal bağlantı nedeniyle regresyon katsayılarında meydana gelen kararsızlıkların grafik üzerinde gösteriminde;
- iii. Modelde gerekli olmayan değişkenlerin modelden çıkarılmasında kullanılır (Hoerl ve Kennard, 1970).

EKK yönteminde izlenen aşamalar ridge tahmininde birden fazla tekrarlanmaktadır. Ridge tahmininin EKK yönteminden farklılığı “ k ” ridge parametresinin oluşudur. 0 ile 1 aralığında değer alan her k Ridge parametresi için hesaplanan tahminler arasından, istenilen kriterleri taşıyanlar belirlenir. Ridge tahmini $X'X$ matrisinin köşegenlerine pozitif olan küçük k değerleri ilave ederek koşul sayısını küçültmeyi hedefler.

Çoklu doğrusal bağlantı olması hâlinde, yanlılığın tanımlanmış birtakım derecelerde tahminlerin varyanslarını düşürecek ridge tahmini Eşitlik (4.6)'daki gibi tanımlanır:

$$\hat{\beta}(k) = (X'X + kI)^{-1}X'Y, \quad k \geq 0 \quad (4.6)$$

Ridge tahmininin optimum k değeri, EKK tahmincisinden daha küçük olan hata kareler ortalamasının bulunmasını sağlayan değerdir ve burada k 'nin tercihi tahminin performansını etkiler. Genelleştirilmiş ridge tahmini $K = \text{diag}(k_i)$, $i=1,2,\dots,p$ şeklinde olmak üzere;

$$\hat{\beta}(K) = (X'X + KI)^{-1}X'Y \quad (4.7)$$

tanımlanır (Akdeniz ve Erol, 2003).

Ridge tahmininin özellikleri:

- i. $\hat{\beta}(k)$, yarıçapı $\hat{\beta}(k)$ 'nin uzunluğunda merkezi orijin olan küre üzerindeki MSE'nin değerini minimum yapar.
- ii. k 'nin artan fonksiyonudur MSE değeri.
- iii. $\hat{\beta}(k)'\hat{\beta}(k) < \hat{\beta}'\hat{\beta}$ ve $k \rightarrow \infty$ olduğunda $\hat{\beta}(k)'\hat{\beta}(k) \rightarrow 0$ 'dır.
- iv. Ridge tahmini $k = 0$ iken $\hat{\beta}(k) = \hat{\beta}_{EKK}$ olacağından ridge tahmini ile EKK tahmini eşit olur. $k \rightarrow \infty$ olduğunda ise $\hat{\beta}(k)$ sıfır vektörüne dönüşerek bu vektör geçersiz olur (Özkale, 2007).

4.4. Ridge Tahmininin Hata Kareler Toplamının Minimumluğu

Ridge tahmininin hata kareler toplamının (SSE) minimum olduğunu göstermek amacıyla başka bir deyişle $\hat{\beta}$ ile $\hat{\beta}(k)$ 'nin tahminlerinin SSE'lerini karşılaştırmak üzere $\hat{\beta}(k)$ 'nin β 'ya olan karesel uzaklığı incelendiğinde;

$$L_1^2 = (\hat{\beta}(k) - \beta)(\hat{\beta}(k) - \beta) \quad (4.8)$$

$$E(L_1^2) = \sigma^2 \sum_{j=1}^k \lambda_j^{-1} \quad (4.9)$$

$$E(L_1^2(k)) = \sigma^2 \sum_{j=1}^k \lambda_j(1 + k) + k\beta'(X'X + kI)^{-2}\beta \quad (4.10)$$

Sonuç olarak; $E(L_1^2(k)) < E(L_1^2)$ olacak biçimde $k > 0$ olduğunda her zaman elde edilebilir.

4.5. Ridge Tahmininin Hata Kareler Ortalaması

Akdeniz ve Erol (2003), Ridge tahmininin MSE değerini Eşitlik (4.11)'deki gibi tanımlamıştır:

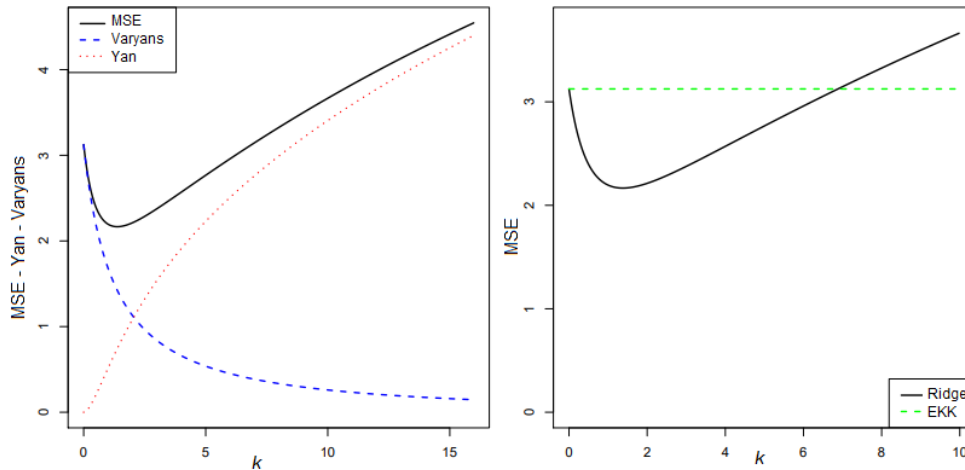
$$MSE(\hat{\beta}(k)) = tr(M(\hat{\beta}(k)), \beta) = \sum_{i=1}^p \frac{\lambda_i \sigma^2 + k^2 a_i^2}{(\lambda_i + k)^2} \quad (4.11)$$

Başka bir ifadeyle;

$$\begin{aligned} MSE(\hat{\beta}(k)) &= Var(\hat{\beta}(k)) + (\hat{\beta}(k)'deki\ yanlılık)^2 \\ &= \sigma^2 tr((X'X + kI)^{-1} X'X (X'X + kI)^{-1}) \\ &= \gamma_1(k) + \gamma_2(k) \end{aligned} \quad (4.12)$$

formülüyle elde edilir.

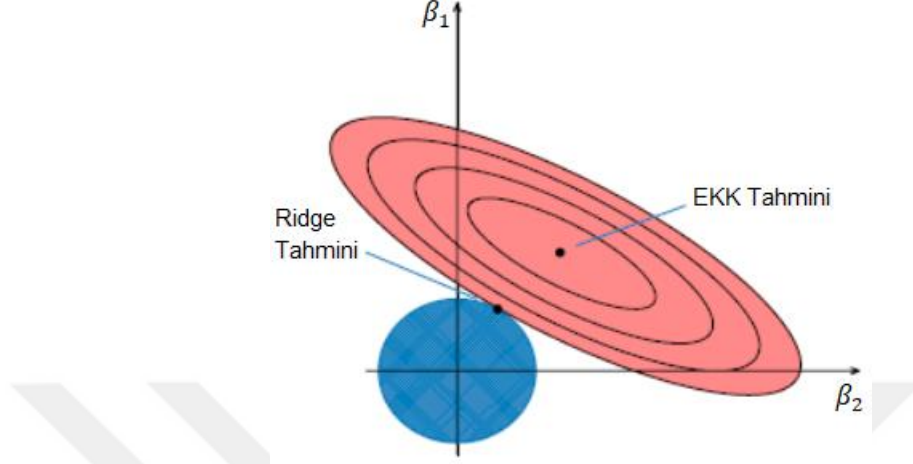
Farklı ridge parametresi k değerlerine göre Ridge tahmini için MSE değerinin bias ve varyans bakımından grafiksel gösterimi ile ridge ve EKK tahminlerini MSE değerlerinin karşılaştırılması Şekil 4.1'de verilmiştir



Şekil 4.1. Ridge parametresi k 'nın farklı değerlerine karşı Ridge tahminine ait varyans, yan ve MSE değerleri (solda), Ridge ve EKK tahminlerinin MSE karşılaştırılması (sağda)

4.6. Ridge Tahmininin Geometrik Yorumu

Ridge tahmininin geometrik yorumu Şekil 4.2’de verilmiştir.



Şekil 4.2. Ridge tahmininin geometrik yorumu

Şekilde yer alan elipsler artık kareler toplamını temsil etmektedir. EKK tahmini artık kareler toplamının minimum olduğu en küçük elipsin merkezine karşılık gelmektedir. Ridge regresyonunda $p = 2$ için ceza terimi $\sum_{j=1}^p \beta_j^2$ bir daireye karşılık gelmektedir. Ridge regresyonunda hem elipsin boyutu hem de daire mümkün olduğunca en küçüklenmeye çalışılmaktadır. Ridge tahmini elips ile dairenin kesiştiği noktaya karşılık gelmektedir. Daha büyük β değerleri seçilerek daha küçük artık kareler toplamı elde edilebilir ancak bu durumda ceza terimi büyüyecektir. Ridge regresyonunda artık kareler toplamı kötü küçük β 'lar tercih edilir.

5. RIDGE TAHMİNİNE DAYALI KANTİL REGRESYON ANALİZİNDE YANLILIK PARAMETRESİ TAHMİNLERİNİN PERFORMANSLARININ KARŞILAŞTIRILMASI

5.1. Giriş

Çoklu doğrusal regresyon modeli,

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (5.1)$$

eşitliği ile ifade edilir. Eşitlikte yer alan $\mathbf{y}$: $n \times 1$ boyutlu gözlem vektörünü, $\mathbf{X}$: $n \times p$ boyutlu standartlaştırılmış açıklayıcı değişkenlerin $n \times p$ boyutlu bilinen tasarım matrisini, $\boldsymbol{\beta}$: $p \times 1$ boyutlu bilinmeyen regresyon katsayılar vektörünü ve $\boldsymbol{\varepsilon}$: $n \times 1$ boyutlu 0 ortalama vektörü ve $\sigma^2 \mathbf{I}_n$ varyans kovaryans matrisi ile çok değişkenli normal dağılıma sahip rastgele hata vektörünü göstermektedir. Regresyon katsayılarının tahmininde en yaygın kullanılan yöntem hataların kareleri toplamını,

$$\sum_{i=1}^n (y_i - \mathbf{x}_i \boldsymbol{\beta})^2 \quad (5.2)$$

en küçüklemeyi amaçlayan EKK yöntemidir. Regresyon katsayılar vektörü $\boldsymbol{\beta}$ 'nin EKK tahmini

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \quad (5.3)$$

ile elde edilir. Model varsayımlarının geçerli olması durumunda EKK tahminleri yansız ve en küçük varyansa sahip tahmin edicidir. Çoklu doğrusal regresyon modelinde açıklayıcı değişkenlerin karşılıklı ilişkisiz olduğu varsayılır. Ancak uygulamada açıklayıcı değişkenler arasında yüksek ya da güçlü doğrusal ilişkiler olabilir ve bu durum çoklu bağlantı problemine neden olur. Çoklu doğrusal bağlantı probleminde EKK tahminlerinin yansızlık özelliği korunsun bile en küçük varyansa sahip olma yani etkinlik özelliği bozulur. Çoklu doğrusal bağlantı problemi, regresyon katsayılarının tahminlerinin varyansının büyük olmasına ve buna bağlı olarak istatistiksel çıkarımların yanlış sonuçlar üretmesine neden olmaktadır (Pfaffenberger ve Dielman, 1990). Çoklu

doğrusal bağlantı probleminin çözümünde en yaygın kullanılan yöntemlerden biri Hoerl ve Kennard (1970) tarafından önerilen ridge regresyonudur. Çoklu doğrusal bağlantı probleminde, ridge regresyonu yan ekleyerek varyansı daha küçük tahminler elde etme için kullanılır. Regresyon katsayılar vektörü β 'nın tahmininde,

$$\sum_{i=1}^n (y_i - \mathbf{x}_i \beta)^2 + k \beta' \beta \quad (5.4)$$

şeklinde ifade edilen fonksiyonu en küçüklenmeyi amaçlayan ridge tahmini,

$$\hat{\beta}(k) = (\mathbf{X}'\mathbf{X} + k\mathbf{I}_p)^{-1} \mathbf{X}'\mathbf{y} \quad (5.5)$$

eşitliği ile elde edilir. Eşitlikte yer alan ridge parametresi k yanlılık parametresi olarak da ifade edilmektedir. Yanlılık parametresi tahminlerin varyansının değişiminde önemli bir rol oynamaktadır. Bu nedenle yanlılık parametresinin seçimi oldukça önemlidir (Suhail ve Kibria, 2019).

Bu çalışmada aykırı gözlem, hata terimlerinin normal olmayan bir dağılım göstermesi veya farklı varyanslılık gibi durumlarda kullanılan alternatif regresyon yöntemlerinden kantil regresyon analizinde çoklu doğrusal bağlantı probleminin çözümü incelenmiştir. Çoklu doğrusal bağlantı probleminin çözümünde ridge tahminine dayalı kantil regresyonu kullanılmıştır. Yanlılık parametresinin tahmini için önerilen bazı tahmin yöntemlerinin ridge tahminine dayalı kantil regresyondaki performansları hata kareler ortalaması ile karşılaştırılmıştır. Çalışmanın ikinci bölümünde kantil regresyonu, üçüncü bölümünde ridge parametresine dayalı kantil regresyonu verilmiştir. Çalışmanın dördüncü bölümünde, seçili parametre değerlerine göre yanlılık parametresinin tahmininde kullanılan tahmin edicilerin performansları simülasyon çalışması ile incelenmiştir. Çalışmanın beşinci bölümünde aykırı gözlem ve çoklu doğrusal bağlantı problemi içeren tobacco veri setinde tekrarlı k katmanlı çapraz doğrulama ile yanlılık parametre tahminlerinin etkinliği karşılaştırılmıştır. Son olarak çalışmadan elde edilen bulgular sonuç bölümünde verilmiştir.

5.2. Kantil Regresyonu

Hata terimlerinin normal olmayan bir dağılıma sahip olması veya aykırı gözlemlerin olması durumunda EKK tahminlerinin alternatifi olarak kullanılan sağlam (robust) yöntemlerden biri de kantil regresyonudur (Yavuz ve Aşık, 2017). Kantil regresyonu aykırı gözlemlere veya hata terimlerinin normal olmayan bir dağılıma sahip olması durumlarına karşı hassas olmadığından dolayı sağlam yöntem olarak ifade edilmektedir. EKK yönteminde açıklayıcı değişkenler ile cevap değişkeni arasındaki ilişki, açıklayıcı değişkenler bilindiğinde cevap değişkeninin koşullu ortalaması olarak modellenir (Chen, 2005). Koenker ve Bassett (1978) tarafından önerilen kantil regresyon yönteminde ise koşullu kantil ile modelleme gerçekleştirilir (Koenker ve Hallock, 2001). Kantil regresyon yönteminde hata terimlerinin varyans yapısına ilişkin herhangi bir varsayım bulunmamaktadır (Baur ve ark., 2004). Bundan dolayı değişen varyans durumunda da kantil regresyonu, EKK yöntemi için alternatif bir yöntem olarak kullanılmaktadır. Kantil regresyonunda,

$$\sum_{i=1}^n \rho_{\theta}(y_i - \mathbf{x}_i \boldsymbol{\beta}) \quad (5.6)$$

eşitliği ile ifade edilen amaç fonksiyonu en küçüklenmeye çalışılır. Eşitlikte yer alan $\rho_{\theta}(\cdot)$ fonksiyonu,

$$\rho_{\theta}(\cdot) = \begin{cases} \theta(y_i - \mathbf{x}_i \boldsymbol{\beta}) & \text{eğer } (y_i - \mathbf{x}_i \boldsymbol{\beta}) \geq 0 \\ (\theta - 1)(y_i - \mathbf{x}_i \boldsymbol{\beta}) & \text{eğer } (y_i - \mathbf{x}_i \boldsymbol{\beta}) < 0 \end{cases} \quad (5.7)$$

seçilen θ ($0 < \theta < 1$) kantil değerine göre bir mutlak değer fonksiyonu olarak değerlendirilebilir. Kantil regresyonunda Eşitlik (5.6) ile ifade edilen amaç fonksiyonunun analitik çözümü yoktur. Kantil regresyonunda katsayılarının tahmini için iteratif algoritmalar veya doğrusal programlama yaklaşımı kullanılır (Altındağ, 2010). Regresyon katsayılarının tahmini için kullanılan algoritmalarda yaygın olarak katsayılar başlangıç tahmin vektörü olarak EKK yöntemi ile elde edilen regresyon katsayılar tahmin vektörü kullanılmaktadır.

5.3. Ridge Tahminine Dayalı Kantil Regresyonu

Kantil regresyon yönteminde çoklu doğrusal bağlantı probleminin çözümünde yanlı tahmin yöntemleri uygulanabilir (Bager, 2018; Zaikarina ve ark., 2016; Zeebari, 2012). Ridge tahminine dayalı kantil regresyon yönteminde Eşitlik (5.8) ile belirtilen amaç fonksiyonu en küçüklenmeye çalışılır.

$$\sum_{i=1}^n \rho_{\theta}(y_i - \mathbf{x}_i \boldsymbol{\beta}) + k \boldsymbol{\beta}' \boldsymbol{\beta} \quad (5.8)$$

Ridge tahminine dayalı kantil regresyon yönteminde, katsayılar tahmin vektörünün elde edilmesinde kullanılan algoritmalarda başlangıç katsayılar tahmin vektörü olarak çoklu regresyonda ridge yöntemi ile elde edilen katsayılar vektörü kullanılabilir. Ridge yaklaşımında katsayıların tahmininde k yanlılık parametresinin seçimi oldukça önemlidir. Literatürde yanlılık parametresinin seçimi için birçok yöntem önerilmiştir. Önerilen yöntemler Eşitlik (5.1) ile ifade edilen regresyon modelinin,

$$\mathbf{y} = \mathbf{Z}\boldsymbol{\alpha} + \boldsymbol{\varepsilon} \quad (5.9)$$

eşitliği ile ifade edilen kanonik formuna dayalıdır. Eşitlikte yer alan $\mathbf{Z} = \mathbf{X}\mathbf{D}$, $\mathbf{D}'\mathbf{D} = \boldsymbol{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_p)$ ve $\boldsymbol{\alpha} = \mathbf{D}'\boldsymbol{\beta}$ olur. Burada $\lambda_1, \dots, \lambda_p$ değerleri $\mathbf{X}'\mathbf{X}$ matrislerinin özdeğerlerini göstermektedir ve $\mathbf{D}$ ortogonal bir matristir. Literatürde önerilen bazı yanlılık parametresi tahmin değerleri aşağıda verilmiştir.

Hoerl ve Kennard (1970) yanlılık parametresi k tahmini için,

$$\hat{k}_{HK} = \frac{\hat{\sigma}^2}{\hat{\alpha}_{\text{enb}}^2} \quad (5.10)$$

eşitliğini önermişlerdir. Burada $\hat{\sigma}^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 / (n - p)$ ve $\hat{\alpha}_{\text{enb}}$ en büyük $\hat{\alpha}$ değerini göstermektedir.

Hoerl, Kennard ve Baldwin (1975) yanlışlık parametresinin tahmininde, her $\hat{\alpha}_i$ için elde edilecek $\hat{k}_i = \frac{\hat{\sigma}^2}{\hat{\alpha}_i^2}$ tahminlerinin harmonik ortalamasının kullanımını önerdiler.

$$\hat{k}_{HKB} = \frac{p\hat{\sigma}^2}{\sum_{i=1}^p \hat{\alpha}_i^2} = \frac{p\hat{\sigma}^2}{\hat{\alpha}'\hat{\alpha}} \quad (5.11)$$

Lawless ve Wang (1976) yanlışlık parametresinin tahmininde bayesci bir yaklaşım kullanarak

$$\hat{k}_{LW} = \frac{p\sigma^2}{\sum_{i=1}^p \lambda_i \hat{\alpha}_i^2} = \frac{p\sigma^2}{\hat{\alpha}'\mathbf{X}'\mathbf{X}\hat{\alpha}} \quad (5.12)$$

eşitliğini önermişlerdir.

Hocking, Speed ve Lynn (1976) yanlışlık parametresinin tahmini için

$$\hat{k}_{HSL} = \hat{\sigma}^2 \frac{\sum_{i=1}^p (\lambda_i \hat{\alpha}_i)^2}{(\sum_{i=1}^p \lambda_i \hat{\alpha}_i^2)^2} \quad (5.13)$$

eşitliğini önermişlerdir.

Kibria (2003) yanlışlık parametresinin tahmininde, $\hat{k}_i = \frac{\hat{\sigma}^2}{\hat{\alpha}_i^2}$ tahminlerinin aritmetik ortalaması ($\hat{k}_{AM}$), geometrik ortalaması ($\hat{k}_{GM}$) ve medyanını ($\hat{k}_{MED}$) kullanmayı önermiştir.

$$\hat{k}_{AM} = \frac{1}{p} \sum_{i=1}^p \frac{\hat{\sigma}^2}{\hat{\alpha}_i^2} \quad (5.14)$$

$$\hat{k}_{GM} = \frac{\hat{\sigma}^2}{(\prod_{i=1}^p \hat{\alpha}_i^2)^{\frac{1}{p}}} \quad (5.15)$$

$$\hat{k}_{MED} = \text{Median} \left\{ \frac{\hat{\sigma}^2}{\hat{\alpha}_i^2} \right\}, \quad i = 1, 2, \dots, p \quad (5.16)$$

Khalaf ve Shukur (2005), yanlışlık parametresinin tahmini için Hoerl ve Kennard (1970) tarafından önerilen $\hat{k}_{HK}$ tahmininin modifikasyonuna dayanarak

$$\hat{k}_{KS} = \frac{\lambda_{\max} \hat{\sigma}^2}{(n - p) \hat{\sigma}^2 + \lambda_{\max} \hat{\alpha}_{\max}^2} \quad (5.17)$$

tahminini önermişlerdir.

5.4. Simülasyon

Bu bölümde ridge tahminine dayalı kantil regresyonunda yanlışlık parametresi k tahminlerinin performansını karşılaştırmak için bir simülasyon çalışması gerçekleştirilmiştir. Simülasyon çalışması için çoklu doğrusal bağlantı problemi ve aykırı gözlemler içeren yapay veri setleri oluşturulmuştur. Regresyon modelinde yer alan açıklayıcı değişkenler çoklu doğrusal bağlantı içerecek şekilde

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

eşitliği ile üretilmiştir (Gibbons, 1981; McDonald ve Galarneau, 1975). Eşitlikte yer alan ρ açıklayıcı değişkenler arasındaki korelasyon katsayısını, w_{ij} gösterimi ise standart normal dağılımından üretilen rasgele değeri göstermektedir. Simülasyon çalışmasında açıklayıcı değişkenler, $X'X$ matrisi korelasyon formunda olacak şekilde standartlaştırılmıştır. β regresyon katsayılar vektörü MSE değerini en küçükleyecek şekilde $X'X$ matrisinin en büyük özdeğerine karşılık gelen özvektör olarak seçilmiştir (Kibria, 2003). Çoklu regresyon modelinde cevap değişkeninin değerleri

$$y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + \varepsilon_i \quad (5.19)$$

eşitliği ile oluşturulmuştur. Hata terimleri ε_i , $N(0, \sigma^2)$ dağılımından üretilmiştir. Veri setlerinde aykırı gözlem oluşturmak için iki cevap değişkeninin değeri $y_i^* = y_i + 10\sigma$ eşitliği ile dönüştürülmüştür.

Böylece simülasyon çalışmasında %20, %6.7 ve %2 aykırı gözlem oranları ile çalışılmıştır. Bu çalışmada açıklayıcı değişken sayısı $p = 4$ alınmıştır ve diğer model parametreleri aşağıdaki gibi belirlenmiştir.

$\theta = 0.25, 0.50$ ve 0.75

$n = 10, 30$ ve 100

$\sigma = 0.2$ ve 0.5

$\rho^2 = 0.95$ ve 0.99

Ridge tahminine dayalı kantil regresyonda yanlılık parametresinin tahmininde Eşitlik (5.10 – 17) arasında tanımlanan sekiz tahmin edici kullanılmıştır. 10000 tekrar ile gerçekleştirilen simülasyon çalışmasında, yanlılık parametresinin tahminlerinin performansları, yanlılık parametresi tahminine dayalı elde edilen tahmin edicilerin toplam hata kareler ortalaması

$$MSE(\hat{\beta}_k) = \frac{1}{10000} \sum_{i=1}^{10000} \sum_{j=1}^4 (\beta_j - \hat{\beta}_{k,ij})^2 \quad (5.20)$$

kriterine göre değerlendirilmiştir. Seçili parametre değerlerine göre gerçekleştirilen simülasyon çalışmasından elde edilen regresyon katsayısı tahminlerinin toplam MSE değerleri Tablo (5.1 – 3)’te verilmiştir.

Tablo (5.1 – 3)’teki değerler incelendiğinde, açıklayıcı değişkenler arasındaki ρ korelasyon katsayısı değeri büyüdüğünde beklenildiği gibi tahmin edicilerin toplam MSE değerlerinin büyüdüğü, aynı şekilde σ^2 varyans değeri arttırıldığında tahmin edicilerin toplam MSE değerlerinin arttığı görülmüştür. Simülasyon çalışmasında örneklem hacmi büyüdükçe genel olarak tahmin edicilerin toplam MSE değerleri azalmıştır. Bu sonuçlar simülasyon çalışmasının başarılı olduğu göstermektedir.

Aykırı gözlem ve çoklu doğrusal bağlantı problemi içeren veri setlerinde klasik kantil regresyon yöntemi ile ridge tahminine dayalı kantil regresyon yöntemi karşılaştırıldığında ridge tahminine dayalı kantil regresyonun daha başarılı olduğu görülmüştür. Ridge tahminine dayalı kantil regresyonda yanlılık parametresi k ’nın tahminlerine göre elde edilen regresyon katsayısı tahminlerinin toplam MSE değerlerine göre grafiksel karşılaştırılması Şekil (5.1-3)’te verilmiştir.

Tablo 5.1. Seçili parametre değerlerine göre tahmin edicilerin toplam MSE değerleri ($\theta = 0.25$)

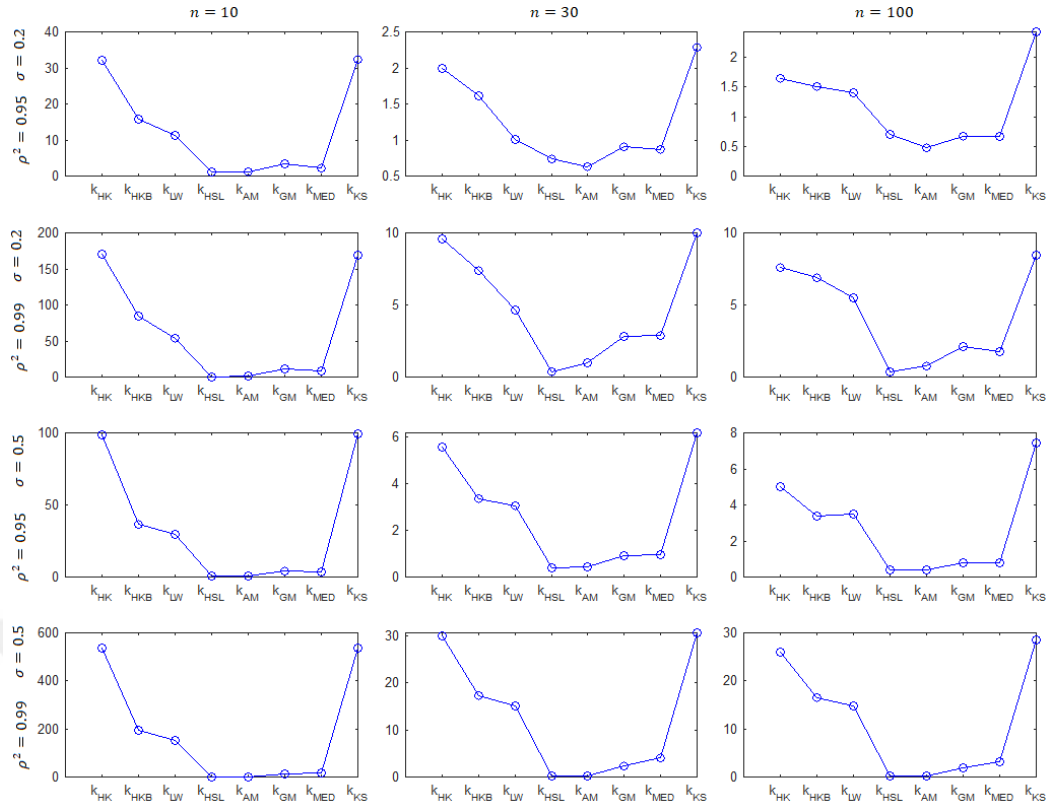
σ	ρ^2	n	Kantil	$\hat{k}_{HK}$	$\hat{k}_{HKB}$	$\hat{k}_{LW}$	$\hat{k}_{HSL}$	$\hat{k}_{AM}$	$\hat{k}_{GM}$	$\hat{k}_{MED}$	$\hat{k}_{KS}$
0.2	0.95	10	69.203	32.065	15.825	11.35	1.059	1.252	3.374	2.169	32.288
0.2	0.95	30	7.659	1.997	1.616	1.008	0.734	0.634	0.903	0.874	2.284
0.2	0.95	100	6.574	1.650	1.514	1.402	0.706	0.472	0.674	0.665	2.427
0.2	0.99	10	360.09	170.14	84.709	54.17	0.531	1.697	11.389	8.943	168.91
0.2	0.99	30	38.840	9.600	7.371	4.647	0.382	0.968	2.811	2.860	9.983
0.2	0.99	100	33.557	7.578	6.871	5.503	0.377	0.746	2.134	1.785	8.442
0.5	0.95	10	548.23	98.841	36.356	29.79	0.633	0.803	4.041	3.536	99.134
0.5	0.95	30	43.866	5.555	3.366	3.060	0.398	0.430	0.924	0.952	6.164
0.5	0.95	100	40.663	5.035	3.373	3.509	0.386	0.395	0.775	0.776	7.444
0.5	0.99	10	2803.55	536.08	194.38	151.63	0.391	0.577	11.639	18.51	535.57
0.5	0.99	30	249.34	29.997	17.354	15.063	0.228	0.300	2.285	4.017	30.601
0.5	0.99	100	211.48	26.013	16.595	14.825	0.244	0.268	1.958	3.214	28.510

Tablo 5.2. Seçili parametre değerlerine göre tahmin edicilerin toplam MSE değerleri ($\theta = 0.50$)

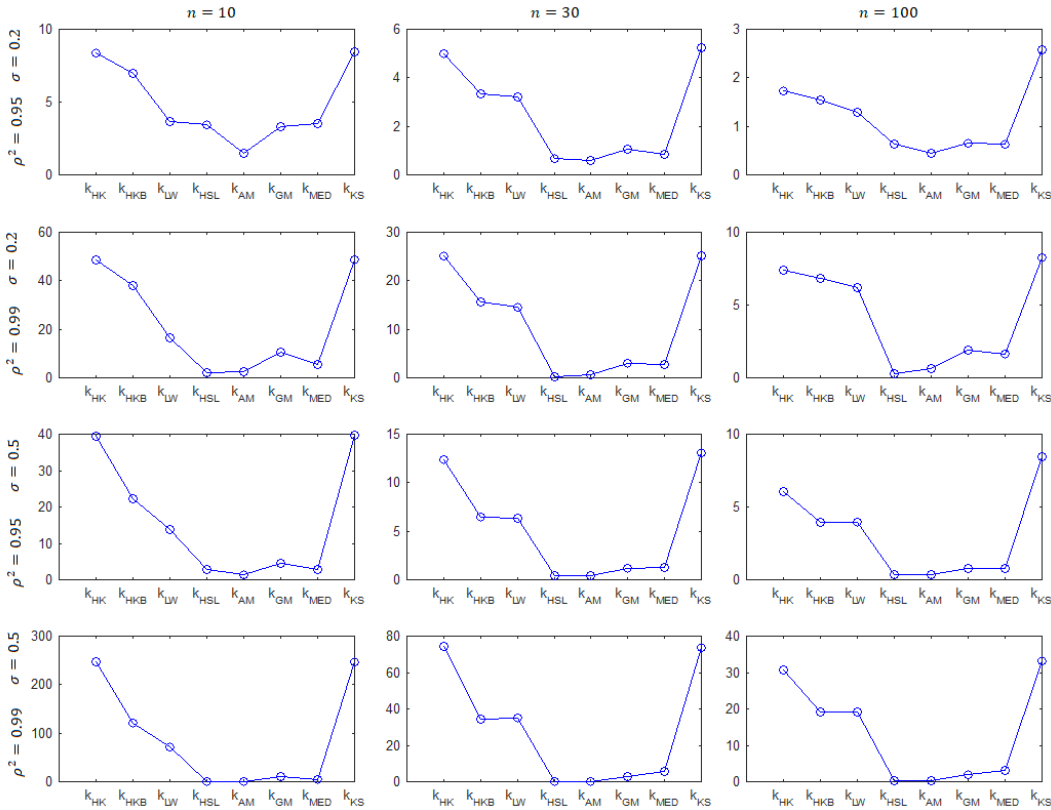
σ	ρ^2	n	Kantil	$\hat{k}_{HK}$	$\hat{k}_{HKB}$	$\hat{k}_{LW}$	$\hat{k}_{HSL}$	$\hat{k}_{AM}$	$\hat{k}_{GM}$	$\hat{k}_{MED}$	$\hat{k}_{KS}$
0.2	0.95	10	18.787	8.361	6.976	3.647	3.426	1.461	3.318	3.527	8.437
0.2	0.95	30	12.619	5.004	3.348	3.214	0.658	0.591	1.048	0.858	5.235
0.2	0.95	100	7.926	1.725	1.549	1.289	0.641	0.444	0.649	0.628	2.581
0.2	0.99	10	106.74	48.623	38.075	16.519	2.009	2.627	10.717	5.582	48.794
0.2	0.99	30	69.483	25.069	15.622	14.486	0.270	0.712	3.043	2.819	25.090
0.2	0.99	100	38.817	7.421	6.842	6.219	0.304	0.632	1.906	1.652	8.271
0.5	0.95	10	170.66	39.485	22.263	13.811	2.811	1.307	4.543	2.788	39.793
0.5	0.95	30	66.503	12.411	6.489	6.315	0.431	0.412	1.218	1.266	13.055
0.5	0.95	100	49.380	6.056	3.944	3.979	0.348	0.368	0.789	0.786	8.443
0.5	0.99	10	971.28	246.86	121.45	71.103	1.171	1.014	11.069	5.280	246.49
0.5	0.99	30	411.58	74.240	34.472	35.220	0.233	0.258	2.963	5.596	73.629
0.5	0.99	100	249.94	30.741	19.086	19.166	0.240	0.250	1.868	3.209	33.124

Tablo 5.3. Seçili parametre değerlerine göre tahmin edicilerin toplam MSE değerleri ($\theta = 0.75$)

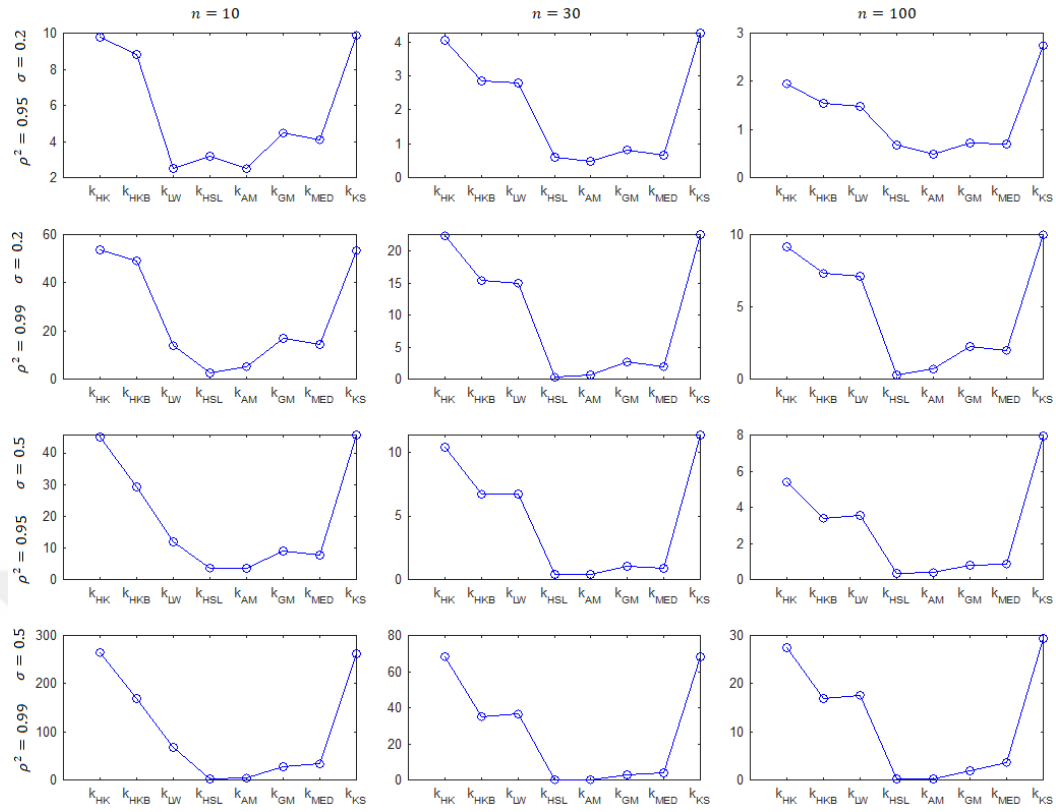
σ	ρ^2	n	Kantil	$\hat{k}_{HK}$	$\hat{k}_{HKB}$	$\hat{k}_{LW}$	$\hat{k}_{HSL}$	$\hat{k}_{AM}$	$\hat{k}_{GM}$	$\hat{k}_{MED}$	$\hat{k}_{KS}$
0.2	0.95	10	33.079	9.761	8.836	2.517	3.187	2.488	4.456	4.115	9.866
0.2	0.95	30	13.065	4.051	2.847	2.797	0.604	0.474	0.826	0.672	4.266
0.2	0.95	100	6.680	1.954	1.546	1.485	0.678	0.484	0.717	0.702	2.737
0.2	0.99	10	178.08	53.774	49.181	13.837	2.588	5.244	16.816	14.392	53.423
0.2	0.99	30	77.039	22.415	15.387	15.036	0.258	0.641	2.662	1.982	22.617
0.2	0.99	100	34.002	9.165	7.299	7.105	0.301	0.686	2.238	2.008	9.970
0.5	0.95	10	280.76	44.962	29.306	11.893	3.696	3.596	8.945	7.701	45.612
0.5	0.95	30	71.010	10.409	6.696	6.713	0.391	0.397	1.051	0.903	11.353
0.5	0.95	100	41.055	5.393	3.404	3.551	0.357	0.377	0.809	0.872	7.933
0.5	0.99	10	1490.7	265.19	169.52	68.173	1.621	3.350	28.096	33.232	262.37
0.5	0.99	30	455.34	68.480	35.191	36.406	0.226	0.264	2.749	3.924	68.383
0.5	0.99	100	212.82	27.476	16.888	17.643	0.265	0.251	1.969	3.709	29.330



Şekil 5.1. $\theta = 0.25$ için yanlışlık parametresi tahminlerine göre elde edilen regresyon katsayı tahminlerinin toplam MSE değerlerinin çizgi grafikleri



Şekil 5.2. $\theta = 0.50$ için yanlışlık parametresi tahminlerine göre elde edilen regresyon katsayı tahminlerinin toplam MSE değerlerinin çizgi grafikleri



Şekil 5.3. $\theta = 0.75$ için yanlışlık parametresi tahminlerine göre elde edilen regresyon katsayı tahminlerinin toplam MSE değerlerinin çizgi grafikleri

Ridge tahminine dayalı kantil regresyonunda yanlışlık parametresi k 'nın tahmin edicileri karşılaştırıldığında Hocking, Speed ve Lynn (1976) tarafından önerilen $\hat{k}_{HSL}$ ve Kibria (2003) tarafından önerilen $\hat{k}_{AM}$ tahminlerinin daha başarılı olduğu gözlemlenmiştir. Hoerl ve Kennard (1970) tarafından önerilen $\hat{k}_{HK}$ ile Khalaf ve Shukur (2005) tarafından önerilen $\hat{k}_{KS}$ tahminleri diğer yanlışlık tahmin edicilerine göre daha yüksek MSE değerine sahip olmuş ve incelenen k yanlışlık tahmin ediciler arasında en kötü performansı göstermiştir.

5.5. Uygulama

Tobacco verisi çoklu doğrusal bağlantı ve aykırı değer problemi taşıyan bir veri setidir (Myers, 1990). Tobacco verisi bir cevap değişkeni ve dört açıklayıcı değişken içeren 30 birimden oluşan bir veri setidir. Bu veri setinde kantil regresyonu ile ridge tahminine dayalı kantil regresyonunun performansı tekrarlı k katmanlı çapraz doğrulama tekniği ile karşılaştırılacaktır. Uygulamada tobacco verisi standartlaştırılmıştır.

Tekrarlı k katmanlı çapraz doğrulama tekniğinde, veri seti rastgele k parçaya bölünür ve bölünen her parça sırayla test verisi, geri kalan k-1 parçadan oluşan veri seti de eğitim verisi olur. Böylece oluşturulan her k parça test verisi olarak kullanılmış olur. Eğitim verisi ile oluşturulan modelin etkinliği test verisinde ölçülür. Verinin rastgele k parçaya ayrılmasındaki rastgeleliğin model belirlemedeki etkisini azaltmak için bu işlemler tekrarlanarak tekrarlı k katmanlı çapraz doğrulama gerçekleştirilmiş olur. Veri seti k parçaya bölündükten sonra her parçadaki gözlem sayısı r olmak üzere test verisindeki toplam hata değeri

$$E(k) = \sum_{i=1}^r (y_{i,\text{test}} - \hat{y}_{i,\text{test}})^2 \quad (5.21)$$

eşitliği ile elde edilir. Eşitlikte yer alan $\hat{y}_{i,\text{test}}$ değeri k-1 parçadan oluşan eğitim verisi ile elde edilen parametre tahmin değerlerine göre test verisindeki i. cevap değişkeninin tahmini değerini göstermektedir. Veri setinin bölünmesi ile oluşan k parçanın her biri test verisi olarak kullanıldığından k tane toplam hata değeri hesaplanır. Tekrar sayısı t olmak üzere çapraz doğruluk hatası

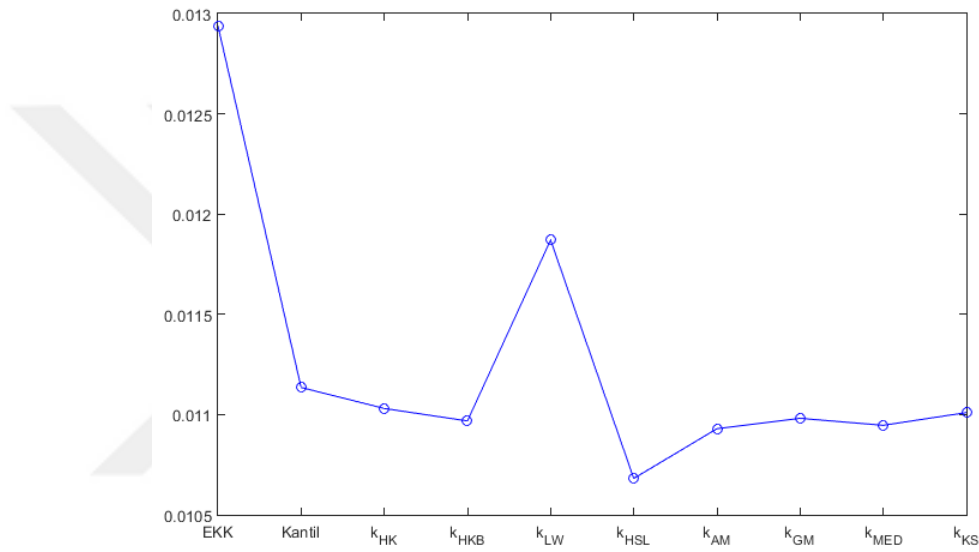
$$CV = \frac{1}{t} \sum_{i=1}^t \frac{1}{k} \sum_{m=1}^k E_i(m) \quad (5.22)$$

eşitliği ile elde edilir. Elde edilen değerin küçük olması parametre tahmin yönteminin performansının iyi olduğu anlamına gelir.

Uygulamada, tekrarlı k katmanlı çapraz doğrulama tekniğinde k=5 ve tekrar sayısı 1000 alınmıştır. Tobacco verisi için EKK, Kantil ve ele alınan yanlılık parametresi tahminlerine göre ridge tahminine dayalı kantil regresyon analizlerinden elde edilen katsayı tahminleri, çoklu belirlilik katsayısı R^2 ve tekrarlı k katmanlı çapraz doğrulama tekniği ile elde edilen CV değerleri Tablo (5.4)'te verilmiştir. Tekrarlı k katmanlı çapraz doğrulama tekniği ile elde edilen CV değerlerine ait çizgi grafiği Şekil (5.4)'te verilmiştir.

Tablo 5.4. Farklı tahmin yöntemlerine göre regresyon katsayı tahminleri, R^2 ve CV değerleri

	EKK	Kantil	$\hat{k}_{HK}$	$\hat{k}_{HKB}$	$\hat{k}_{LW}$	$\hat{k}_{HSL}$	$\hat{k}_{AM}$	$\hat{k}_{GM}$	$\hat{k}_{MED}$	$\hat{k}_{KS}$
$\hat{\beta}_1$	1.50739	1.50151	1.27887	1.15996	0.34391	0.40921	0.36151	1.10929	0.34930	1.28448
$\hat{\beta}_2$	-0.52107	-0.52475	-0.54479	-0.55466	-0.16386	-0.40675	-0.34839	-0.48318	-0.34573	-0.54454
$\hat{\beta}_3$	-0.84160	-0.88288	-0.73362	-0.67789	0.20556	-0.03318	-0.31753	-0.71234	-0.30900	-0.73434
$\hat{\beta}_4$	0.82171	0.87158	0.96596	1.03944	0.60515	1.01369	1.28394	1.06697	1.28541	0.96081
R^2	0.95720	0.95717	0.95697	0.95668	0.94799	0.95262	0.95290	0.95641	0.95280	0.95698
CV	0.01294	0.01114	0.01103	0.01097	0.01187	0.01068	0.01093	0.01098	0.01095	0.01101

**Şekil 5.4.** Tekrarlı k katmanlı çapraz doğrulama ile elde edilen CV değerlerine ait çizgi grafiği

Tablo (5.4) ve Şekil (5.4) incelendiğinde Hocking, Speed ve Lynn (1976) tarafından önerilen $\hat{k}_{HSL}$ yanlışlık parametre tahminine dayalı kantil regresyon yöntemi ile en küçük CV değeri elde edilmiştir. Simülasyon sonuçları ile uyumlu bir şekilde CV kriterine göre en başarılı iki yanlışlık tahmin yöntemi Hocking, Speed ve Lynn (1976) tarafından önerilen $\hat{k}_{HSL}$ ve Kibria (2003) tarafından önerilen $\hat{k}_{AM}$ tahmin edicileridir. Çoklu doğrusal bağlantı ve aykırı gözlem problemleri içeren veri setinde EKK tahminleri en yüksek CV değeri ile en başarısız tahmin edici olmuştur.

5.6. Sonuç

Bu çalışmada çoklu bağlantı problemi ve aykırı gözlem varlığında kantil regresyonunda ridge tahminine dayalı çözüm incelenmiştir. Literatürde yaygın kullanıma sahip sekiz yanlılık tahmin edicisinin performansı tahmin edicilerin toplam MSE kriterine göre değerlendirilmiştir. Simülasyon ve gerçek veri seti ile gerçekleştirilen uygulama sonuçları çoklu bağlantı problemi ve aykırı gözlem varlığında kantil regresyonunda ridge tahmin yaklaşımının kullanılabileceğini göstermiştir.

Simülasyon çalışması sonucunda Hocking, Speed ve Lynn (1976) tarafından önerilen $\hat{k}_{HSL}$ ve Kibria (2003) tarafından önerilen $\hat{k}_{AM}$ yanlılık tahmin edicileri en başarılı tahmin ediciler olarak belirlenmiştir. Tobacco veri setinde tekrarlı k katmanlı çapraz doğrulama tekniği ile gerçekleştirilen karşılaştırma sonucunda en başarılı yanlılık tahmin edicisi $\hat{k}_{HSL}$ olmuştur. En başarısız tahmin edici ise EKK tahmini olmuştur.

6. SONUÇLAR VE ÖNERİLER

Hataların normal dağılmadığı veri setlerinde, eğikliğe ve aykırı değerlere EKK yönteminden daha az hassas olan KR yöntemi tercih edilmektedir. Açıklayıcı değişkenler arasında çoklu bağlantının varlığı hâlinde ise klasik tahmin yöntemleri istenilen performansı gösterememektedirler. Değişkenler arasında çoklu doğrusal bağlantı problemi olması durumunda tahmin edicilerin istenilen performansı gösterebilmesi için kullanılan yöntemler arasında yanlı tahmin ediciler yer almaktadır. Yanlı tahmin edicilerden birisi de ridge tahminidir.

Bu çalışmada regresyon analizi için gerekli varsayımlar sağlanmadığında EKK tahmin edicisinin alternatifi olan KR yönteminin tercih edildiği, EKK tahmin edicisinin minimum varyanslı olmasına rağmen açıklayıcı değişkenler arasında çoklu doğrusal bağlantı olması hâlinde ise yansızlık özelliğini yitirdiğinden dolayı yanlı tahmin edicilerden olan ridge tahmininin performansı incelenmiştir.

Çalışmada öncelikle aykırı gözlem durumunda ve hataların normal dağılmaması durumunda EKK yöntemi ile KR yöntemi simülasyon çalışmasıyla karşılaştırılmıştır. Aykırı gözlem durumunda MSE kriterine göre KR yönteminin EKK yöntemine göre daha başarılı olduğu görülmüştür. Örneklem hacminin 30 ve 50 olduğu durumlarda en küçük MSE değeri $\theta = 0.50$ kantil değeri için elde edilmiştir. Örneklem hacminin 100 olduğu durumda ise en küçük MSE değeri $\theta = 0.25$ kantil değeri için elde edilmiştir.

Hataların normal dağılmaması durumu için hatalar $\mu = 0$, $\sigma = 0.5$ parametreleri ile Lognormal dağılımından üretilmiştir. Simülasyon çalışması sonuçlarına göre beklenildiği gibi KR yönteminin EKK yöntemine göre daha başarılı olduğu görülmüştür. KR yönteminde bütün örneklem hacimleri için en küçük MSE değeri $\theta = 0.25$ kantil değeri için elde edilmiştir.

Simülasyon çalışmasında ridge tahminine dayalı kantil regresyonunda yanlılık parametresi olan k tahminlerinin performansları karşılaştırılmıştır. Çalışma aykırı gözlemler ve çoklu doğrusal bağlantı problemi olan simülasyon üzerinde 10000 tekrar ile gerçekleştirilmiştir. Klasik kantil regresyon yöntemi ile ridge tahminine dayalı kantil

regresyon yöntemi karşılaştırılmış olup ridge tahminine dayalı kantil regresyon yönteminin daha iyi performans gösterdiği görülmüştür.

Sonraki çalışmalarda KR yönteminde ridge tahmininin yerine diğer yanlı tahmin edicilerin performansları değerlendirilebilir.



KAYNAKLAR

- Akdeniz, F., & Erol, H., 2003, Mean Squared Error Matrix Comparisons of Some Biased Estimators in Linear Regression, *Çukurova Üniversitesi*, 32(12), 2389–2413.
- Aktaş, C., & Yılmaz, V., 2003, Çoklu Bağıntılı Modellerde Liu ve Ridge Regresyon Kestiricilerinin Karşılaştırılması, *Anadolu Üniversitesi Bilim ve Teknoloji Dergisi*, 4(2), 189-194.
- Akyol, K., 2013, Kantil Regresyon Modeli Yardımıyla Ülkelerin İnsani Gelişmişlik İndeksi Üzerinde Etkili Olan Faktörlerin İncelenmesi, Yüksek Lisans Tezi, Atatürk Üniversitesi Sosyal Bilimler Enstitüsü İşletme Anabilim Dalı, Erzurum.
- Alakaya, D., 2019, Kantil Regresyon ve Doğrusal Regresyon Yöntemlerinin Performansını Etkileyen Faktörlerin İncelenmesi, Yüksek Lisans Tezi, Mersin Üniversitesi Sağlık Bilimleri Enstitüsü Biyoistatistik ve Tıbbi Bilişim Anabilim Dalı, Mersin.
- Al-Hassan, Y. M.i 2008, A Monte Carlo Evaluation of Some Ridge Estimators, *J. J. Appl. Sci.*, 10(2), pp. 101-110.
- Alkhamisi, M., Khalaf, G., & Shukur, G., 2006, Some Modifications for Choosing Ridge Parameters, *Communications in Statistics—Theory and Methods*, 35: 2005–2020.
- Alkhamisi, M. A., & Shukur, G., 2007, A Monte Carlo Study of Recent Ridge Parameters, *Communications in Statistics—Simulation and Computation*, 36: 535–547.
- Altındağ, İ., 2010, Quantile Regresyon ve Bir Uygulama, Yüksek Lisans Tezi, Selçuk Üniversitesi Fen Bilimleri Enstitüsü, Konya.
- Bager, A. S., 2018, Ridge Parameter in Quantile Regression Models: An Application in Biostatistics, *International Journal of Statistics and Applications*, 8(2) 72-78.
- Baur, D., Saisana, M., & Schulze, N., 2004, Modelling the effects of meteorological variables on ozone concentration: a quantile regression approach, *ozone Atmospheric Environment*, 38(28) 4689-4699.
- Bayrak Gezdim, S., 2017, Küresel CO2 Emisyonunun Belirleyicilerinin Analizi: Dinamik Panel Kantil Regresyon Modeli, Doktora Tezi, Uludağ Üniversitesi Sosyal Bilimler Enstitüsü Ekonometri Anabilim Dalı, Bursa.
- Birkes, D., and Dodge, Y., 2011, Alternative Methods of Regression, Vol:190, John Wiley & Sons.
- Buchinsky, M., 1998, Recent Advences in Quantile Regression Models: A partical Guideline for Empirical Research, *The Journal of Human Resources*, 33(1), 88-126.

- Cade, B. S., & Richards, J. D., 2006, A Permutation Test For Quantile Regression, *Journal of Agricultural, Biological, and Environmental Statistics*, 11(1), pp.106–126.
- Chen, C., 2005, An introduction to quantile regression and the QUANTREG procedure, *In Proceedings of the Thirtieth Annual SAS Users Group International Conference*, SAS Institute Inc. Cary, NC. 213-30.
- Çamurlu, S., 2018, Kantil Regresyon Analizinde Bootstrap Tahmini, Yüksek Lisans Tezi, Cumhuriyet Üniversitesi Sosyal Bilimler Enstitüsü Ekonometri Anabilim Dalı, Sivas.
- Çelik, O., & Selim, S., 2014, Türkiye’de Kamu ve Özel Sektör Ücret Farklılıklarının Kantil Regresyon Yaklaşımı ile Analizi, *Yönetim ve Ekonomi*, 21(1) 205-232.
- Davino, C., Furno, M., and Vistocco, D., 2013, Quantile Regression: Theory and Applications, Vol:988, *John Wiley & Sons*,
- Demirez, G., 2018, Bazı değişkenlerin Fen Başarı Puanına Etkisi: PISA 2015 Türkiye, Singapur ve Almanya Örneği, Yüksek Lisans Tezi, Akdeniz Üniversitesi Eğitim Bilimleri Enstitüsü İlköğretim Anabilim Dalı, Antalya.
- Dereny, M. E., & Rashwan, N. I., 2011, Solving Multicollinearity Problem Using Ridge Regression Models, *Int. J. Contemp. Math. Sciences*, 6(12), 585 – 600.
- Dorak, Ö., 2017, Kantil Regresyon ve En Küçük Kareler Yöntemlerinin Karşılaştırılması: Bir Uygulama Denemesi, Yüksek Lisans Tezi, Eskişehir Anadolu Üniversitesi Sosyal Bilimler Enstitüsü İşletme Anabilim Dalı, Eskişehir.
- Durusu Çiftçi, D., 2015, Finansal Gelişme ve Ekonomik Büyüme İlişkisi: Bir Genişletilmiş Solow Büyüme Modeli ve Ampirik Uygulama, Doktora Tezi. Pamukkale Üniversitesi Sosyal Bilimler Enstitüsü İktisat Anabilim Dalı, Denizli.
- Elmalı, K., 2014, Kantil Regresyon ve Negatif Binomial Regresyon ile İllerde Kullanılan İlaç Sayısına Etki Eden Faktörlerin İncelenmesi, Yüksek Lisans Tezi, Atatürk Üniversitesi Sosyal Bilimler Enstitüsü Ekonometri Anabilim Dalı, Erzurum.
- Firinguetti, L., Kibria, G., & Araya, R., 2016, Study of Partial Least Squares and Ridge Regression Methods, *Communications in Statistics - Simulation and Computation*, 46(8), 6631-6644.
- Fisher, R. A., 1922, Teorik istatistiklerin matematiksel temelleri üzerinde, *Londra Kraliyet Topluluğu'nun Felsefi İşlemleri, Seri A, bir matematiksel veya fiziksel Karakter olan makaleler İçeren*, 222 (594-604), 309-368.
- Fisher, R. A., (1925, Temmuz), İstatistiksel tahmin teorisi, *Gelen Cambridge Felsefe Society Matematiksel İlerleme* (Cilt. 22, No. 5, s. 700-725), Cambridge Üniversitesi Basını.
- Galton, F., 1877, Typical laws of heredity, Royal Institution of Great Britain.

- Gibbons, D. G., 1981, A Simulation Study of Some Ridge Estimators, *Journal of the American Statistical Association*, 76(1) 131-139.
- Gökçe, A., 2013, Türk Bankacılık Sisteminde Döviz Kuru Değişimlerinin Bankaların Mali Tablolarına Etkileri, Doktora Tezi, Pamukkale Üniversitesi Sosyal Bilimler Enstitüsü İşletme Anabilim Dalı, Denizli.
- Hao, L., & Naiman, D. Q., 2007, *Quantile Regression*, The United States of America: Sage Publication.
- Hocking, R. R., Speed, F. M., & Lynn, M. J., 1976, A class of biased estimators in linear regression, *Technometrics*, 18(4) 55-67.
- Hoerl, A. E., 1962, Application of ridge analysis to regression problems, *Chemical Engineering Progress*, Vol. 58, no. 3, pp. 54-59.
- Hoerl, A. E., & Kennard, R. W., 1970, Ridge Regression: Biased Estimation for Nonorthogonal Problems, *Technometrics*, 12(1), 55-67.
- Hoerl, A. E., Kennard, R. W., & Baldwin K. F., 1975, Ridge regression: Some simulations. *Communications in Statistics*, 4(2) 105-123.
- John, O. O., & Nduka, E. C., 2009, Quantile Regression Analysis as a Robust Alternative to Ordinary Least Squares, *Scientia Africana*, 8(2), pp 61-65.
- Karakuş, S., 2011, Çoklu Bağlantı (Multicollinearity) Sorunu ve Giderilmesine Yönelik Alternatif Yaklaşımlar, Yüksek Lisans Tezi, Yüzüncü Yıl Üniversitesi Fen Bilimleri Enstitüsü Zootekni Anabilim Dalı, Van.
- Karaoğlu, H. D., 2015, Essays on the Education Gradient of Health in Turkey, Doktora Tezi, Orta Doğu Teknik Üniversitesi Sosyal Bilimler Enstitüsü Ekonomi Anabilim Dalı, Ankara.
- Keskin, B., 2012, Sağlam bir Çıkarsama Yöntemi: Kantik Regresyon, Yüksek Lisans Tezi, Muğla Sıtkı Koçman Üniversitesi Fen Bilimleri Enstitüsü, Muğla.
- Khalaf, G., & Shukur, G., 2005, Choosing Ridge Parameter for Regression Problems, *Communications in Statistics—Theory and Methods*, 34: 1177–1182.
- Kırdemir, Ç., 2017, Ridge Regresyon Yönteminin Farklı Paket Programlarıyla Uygulanması, Yüksek Lisans Tezi, Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü İstatistik Anabilim Dalı, İstanbul.
- Kibria, G. B., 2003, Performance of Some New Ridge Regression Estimators, *Communications in Statistics-Simulation and Computation*, 32(2), pp. 419-435.
- Kibria, G. B., & Banik, S., 2016, Some Ridge Regression Estimators and Their Performances, *Journal of Modern Applied Statistical Methods*, 15(1), pp. 206-238.

- Koenker, R., & Basset, G., 1978, Regression Quantiles, *Econometrica* 46 (1) 33-50.
- Koenker, R., & Hallock, K. F., 2001, Quantile Regression, *Journal of Economic Perspectives* 15 (4) 143-156.
- Koenker, R., 2005, *Quantile Regression*. The United States of America : Cambridge University Press.
- Koşan, N. İ., 2014, OECD Ülkelerinde Dış Ticaret Hadlerini Etkileyen Değişkenlerin Panel Kantil Regresyon Modelleri ile İncelenmesi, Doktora Tezi, Marmara Üniversitesi Sosyal Bilimler Enstitüsü Ekonometri Anabilim Dalı, İstanbul.
- Kuvat, A., (2018), Ridge Regresyonda Sağlam Parametre Bulma, Yüksek Lisans Tezi, Muğla Sıtkı Koçman Üniversitesi Fen Bilimleri Enstitüsü İstatistik Anabilim Dalı, Muğla.
- Lawless, J. F., & Wang, P., 1976, A simulation study of ridge and other regression estimators, *Communications in Statistics – Theory and Methods*, 5(4) 307-323.
- Le Cessie, S., & Van Houwelingen, J. C., 1992, Ridge Estimators in Logistic Regression, *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 41(1), pp. 191-201.
- Liu K (2003) Using Liu-type estimator to combat collinearity. *Commun Stat Theory Methods* 32:1009–1020
- Lee, H., & Tanaka, H., 1999, Upper and lower approximation models in interval regression using regression quantile techniques, *European Journal of Operational Research*, 116, pp. 653-666.
- Lukman, A. F., & Ayinde, K., 2018, Detecting Influential observations in Two-Parameter Liu-Ridge Estimator, *Journal of Data Science* , 207-218.
- Mahajan, V., Jain, A. K., & Bergier , M., 1977, Parameter Estimation in Marketing Models in the Presence of Multicollinearity: An Application of Ridge Regression, *Journal of Marketing Research*, 14(4), pp. 586-591.
- Marquardt, D. W., & Snee, R. D., 1975, Ridge Regression in Practice, *The American Statistician*, 29(1), 3-20.
- Martins, P. S., & Pereira, P. T., 2004, Does education reduce wage inequality? Quantile regression evidence from 16 countries, *Labour Economics*, 11, pp. 355-371.
- McDonald, G. C., & Galarneau, D. I., 1975, A Monte Carlo Evaluation of Some Ridge-Type Estimators, *Journal of the American Statistical Association*, 70 (350) 407-416.

- Mori, Y., & Suzuki, T., 2018, Generalized Ridge Estimator and Model Selection Criteria in Multivariate Linear Regression, *Journal of Multivariate Analysis*, 165: 243-261.
- Myers, R. H., 1990, Classical and modern regression with applications, Second Edition, Belmont, CA: Duxbury press.
- Özdemir, M., 2018, Kantil Birim Kök Testleri: Kırılgan Beşli Ülkelerinde Enflasyonun Kalıcılığının Analizi, Yüksek Lisans Tezi, İstanbul Üniversitesi Sosyal Bilimler Enstitüsü Ekonometri Anabilim Dalı, İstanbul.
- Özkale, M. R., 2007, Çoklu İç İlişki İle İlgili Problemler, Doktora Tezi. Çukurova Üniversitesi Fen Bilimleri Enstitüsü, Adana.
- Özocaklı, D., 2017, Gıda Güvensizliği ve Makro Ekonomik Belirleyicileri (Kantil Regresyon Analizi), Yüksek Lisans Tezi, Gaziantep Üniversitesi Sosyal Bilimler Enstitüsü İktisat Anabilim Dalı, Gaziantep.
- Pearson, K., 1896, VII. Mathematical contributions to the theory of evolution,—III. Regression, heredity, and panmixia, *Philosophical Transactions of the Royal Society of London, Series A, containing papers of a mathematical or physical character*, (187), 253-318.
- Pfaffenberger, R. C., & Dielman, T. E., 1990, A comparison of regression estimators when both multicollinearity and outliers, içinde: K. D. Lawrence, J. L. Arthur (Ed.) *Robust Regression: Analysis and applications*, Marcer Dekker Inc. New York and Basel, 243-270.
- Saçaklı, İ., 2005, Kantil Regresyon ve Alternatif Regresyon Modelleri ile Karşılaştırılması, Yüksek Lisans Tezi, Marmara Üniversitesi Sosyal Bilimler Enstitüsü Ekonometri Anabilim Dalı, İstanbul.
- Sakallıoğlu, S., & Kaçıranlar, S., 2008, A New Biased Estimator Based on Ridge Estimation, *Stat Papers*, 49: 669–689.
- Suhail, M. S., & Kibria, B. M. G., 2019 Quantile based estimation of biasing parameters in ridge regression model, *Communications in Statistics-Simulation and Computation*.
- Toker, S., 2011, Ridge Regresyon Tahmin Edici ve İlişkili Tahmin Ediciler, Yüksek Lisans Tezi, Çukurova Üniversitesi Fen Bilimleri Enstitüsü İstatistik Anabilim Dalı, Adana.
- Tongal, A., 2017, PISA 2015 Türkiye Verilerine Göre Bazı Değişkenlerin Fen Testi Başarı Puanına Etkisinin Kantil Regresyon Yöntemi İle İncelenmesi, Yüksek Lisans Tezi, Akdeniz Üniversitesi Eğitim Bilimleri Enstitüsü İlköğretim Anabilim Dalı, Antalya.
- 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.

- Türkış, C., 2015, Sağkalım Analizinde Kantil Regresyon ve Parametrik Regresyon Modellerinin Performanslarının Karşılaştırılması, Yüksek Lisans Tezi, Adnan Menderes Üniversitesi Sağlık Bilimleri Enstitüsü Biyoistatistik Anabilim Dalı, Aydın.
- Unat, E., 2018, Gelir ve Tüketim İlişkisinin İstikrarı: Harcama Gruplarına ve Zamana Göre Kantil Regresyon Modelden Kanıtlar, Yüksek Lisans Tezi, Balıkesir Üniversitesi Sosyal Bilimler Enstitüsü Uluslararası Ticaret ve Pazarlama Anabilim Dalı, Balıkesir.
- Uslu, E. E., 2016, Türkiye'de Beşeri Sermaye Kazançları: Ekonomik Faaliyet ve Mesleğe Göre Eğitimin Getirisi, Doktora Tezi, Atatürk Üniversitesi Sosyal Bilimler Enstitüsü Ekonometri Anabilim Dalı, Erzurum.
- Wagner, H. M., 1959. Linear programming Techniques for Regression Analysis, *Journal of American statistical Association*, 54(285), 206-212.
- Yavuz, A. A., & Aşık, E. G., 2017, Kantil Regresyon, *Uluslararası Mühendislik Araştırma ve Geliştirme Dergisi*, 9(2) 137-146.
- Yıldırım, Z., 2017, Kantil Regresyon ve Sansürlü Modellerle Türkiye'de Hanehalkı Tasarruf Eğilimi: Mikroekonomik Analiz, Doktora Tezi, Dokuz Eylül Üniversitesi Sosyal Bilimler Enstitüsü Ekonometri Anabilim Dalı, İzmir.
- Yu, K., Lu, Z., & Stander, J., 2003, Quantile regression: applications and current research areas, *The Statistician*, 52(3) pp. 331–350.
- Zaikařina, H., Djuraidah, A., & Wigena, A. H., 2016, Wigena, Lasso and Ridge Quantile Regression using Cross Validation to Estimate Extreme Rainfall, *Global Journal of Pure and Applied Mathematics*, 12 (3) 3305–3314.
- Zeebari, Z., 2012, Developing Ridge Estimation Method for Median Regression, *Journal of Applied Statistics*, 39(12), pp. 2627–2638.
- Zietz, J., Zietz, E. N., & Sirmans, G. S., 2008, Determinants of House Prices: A Quantile Regression Approach, *The Journal of Real Estate Finance and Economics*, 37(4), 317-333.

ÖZGEÇMİŞ

KİŞİSEL BİLGİLER

Adı Soyadı : Nurullah YAMAN
Uyruğu : T.C.
Doğum Yeri ve Tarihi : Selçuklu-25.11.1991
Telefon : 05071924735
e-mail : nurullahyaman42@gmail.com

EĞİTİM

Derece	Adı, İlçe, İl	Bitirme Yılı
Lise :	Meram Anadolu Ticaret Meslek Lisesi Bilişim Teknolojileri Alanı, Meram, Konya	2009
Üniversite :	Selçuk Üniversitesi Teknik Bilimler Meslek Yüksekokulu Bilgisayar Programcılığı, Konya	2012
	Selçuk Üniversitesi Fen Fakültesi İstatistik Anabilim Dalı, Konya	2015
Yüksek Lisans :	Necmettin Erbakan Üniversitesi Fen Bilimleri Enstitüsü İstatistik Anabilim Dalı, Konya	2019

İŞ DENEYİMLERİ

Yıl	Kurum	Görevi
2018-2019	Konya Meram Gökyurt Ortaokulu	Matematik öğretmeni

UZMANLIK ALANI : İstatistik

YABANCI DİLLER : İngilizce