Dokuz Eylül Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi,
Cilt:25, Sayı:2, Yıl:2010, ss.137-148.

Sağlam Ridge Regresyon Analizi ve Bir Uygulama

Özlem ALPU[1] Hatice ŞAMKAR[2] Ekrem ALTAN[3]

Özet

Çoklu regresyon analizinde doğrusal bağıntı problemi En Küçük Kareler(EKK) tekniğiyle regresyon parametrelerinin tahmininde güvenilir olmayan sonuçlar verir. Diğer taraftan veri kümesinin aykırı değerler içermesi EKK tahmin edicisinin en iyi, yansız, tutarlı olma özelliğini yitirmesine neden olur. Bu iki problemin aynı veri kümesinde yer alması durumunda sağlam tahmin ediciler üzerinde temellenen yanlı teknikler kullanılması önerilir. Bu çalışmada hem x hem de y yönünde aykırı değer içeren bir veri kümesi için sağlam tahmin edicilerden M, En Küçük Kırpılmış Kareler, En Küçük Medyan Kareler, S ve Genelleştirilmiş M üzerinde temellenen ridge regresyon analizi üzerinde durulmuş ve karşılaştırmalı olarak sonuçları incelenmiştir.

Anahtar Kelimeler: Sağlam Regresyon, Ridge Regresyon, Aykırı değer, Çoklu Bağıntı
JEL Sınıflandırma Kodları: C100, C200, C180

Abstract

The problem of multicollinearity in multiple linear regression analysis gives unreliable estimates for regression parameters with least squares methods. In addition, outliers in data set cause to lose characteristics of best, unbiased, consistent of least squares estimation. In the presence of multicollinearity and outliers in the data set, it is suggested using biased methods based on robust estimators. In this study, for the data set including outliers both x and y directed, ridge regression analysis based on some robust techniques (M, Least Trimmed Sums of Squares, Least Median of Squares, S and Generalized M) is applied and the results are considered as comparatively.

Keywords: Robust regression, ridge regression, outlier, multicollinearity
JEL Classification Codes: C100, C200, C180

[1] Yrd.Doç.Dr., Eskişehir Osmangazi Üniversitesi, İstatistik Bölümü, Meşelik Kampüsü, Eskişehir
e-posta:oalpu@ogu.edu.tr
[2] Yrd.Doç.Dr., Eskişehir Osmangazi Üniversitesi, İstatistik Bölümü, Meşelik Kampüsü, Eskişehir
e-posta: hfidan@ogu.edu.tr
[3] Yüksek Lisans Öğrencisi, Eskişehir Osmangazi Üniversitesi, Fen Bilimleri Enstitüsü

1. Giriş

Veri kümesinde aykırı değerlerin var olması ve aynı zamanda çoklu doğrusal bağıntı problemiyle karşılaşılması durumunda bu iki problemle ayrı ayrı uğraşmak yerine problemlerin her ikisiyle aynı anda başedecek teknikler üzerinde çalışmak daha uygun olabilir. Bunun için sağlam tahmin ediciler üzerinde temellenen yanlı tekniklerin kullanılması önerilir. Literatürde Askin ve Montgomery(1980) veri setinde çoklu bağıntı probleminin var olması ve hata terimlerinin normal olmaması durumunda yanlı ve sağlam tekniklere dayalı tahmin edicileri kullanmıştır. Silvapulle(1991) en küçük kareler tahmin edicilerinin yerine M tahmin edicilerinden elde edilen ridge tipi M tahmin edicilerini kullanmış ve yanlılık parametresinin optimal değerinin seçimi için öneride bulunmuştur. Arslan ve Billor (1996) çoklu bağıntı ve aykırı değerlerin varlığında iki alternatif ridge tipi Genelleştirilmiş M (GM) tahmin edicisini geliştirmiş ve aykırı değerlerin etkisini azaltmak amacıyla yanlılık parametresini bu sağlam tahmin edicilerle hesaplayıp, regresyon parametrelerinin sağlam ridge tahmin edicilerini elde etmişlerdir. Pfaffenberger ve Dielman(1990) en küçük mutlak değer(Least Absolute Value-LAV) ve ridge tahmin edicilerinin özelliklerini birleştirerek ridge en küçük mutlak değer (Ridge Least Absolute Value-RLAV) tahmin edicilerini önermiştir.

Bu çalışmada sağlam tahmin edicilerden M, En Küçük Kırpılmış Kareler(Least Trimmed Squares-LTS), En Küçük Medyan Kareler(Least Median of Squares- LMS), S ve Genelleştirilmiş M(Generalized M-GM) üzerinde temellenen ridge regresyon analizi ele alınmıştır. Hesaplanan parametre tahminlerinden hangisinin daha iyi olduğuna karar vermek için hata kareler ortalaması kriteri göz önünde tutulmuştur. Silvapulle(1991)'in M tahmin edicisi ve Arslan ve Billor(1996)'un GM tahmin edicileri için önerdiği sağlam ridge tahmin edicilerine ilişkin hata kareler ortalaması formülü bu çalışmada LTS, LMS ve S'ye dayalı sağlam ridge tahmin edicilerine uyarlanarak kullanılmıştır. Uygulama kısmında ise gerçek bir veri seti üzerinde bu metotlar karşılaştırmalı olarak değerlendirilmiştir.

2.Metot

2.1. Ridge Regresyon

Çoklu doğrusal regresyon için aşağıdaki standart modeli düşünelim.

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \qquad (2.1)$$

Burada $\mathbf{X}$ ($n\times p$) boyutlu ve p ranklı bağımsız değişken matrisi, $\boldsymbol{\beta}$ ($p\times 1$) boyutlu bilinmeyen parametre vektörü ve $\boldsymbol{\varepsilon}$, $\mathrm{E}(\boldsymbol{\varepsilon})=0$ ve $\mathrm{Var}(\boldsymbol{\varepsilon})=\sigma^2\mathbf{I_n}$ olan ($n\times 1$) boyutlu hata terimleri vektörüdür. Verilerin standartlaştırılması durumunda $\mathbf{X'X}$ matrisi korelasyon matrisi formundadır.

Çoklu doğrusal bağıntı varlığında (2.1) eşitliğindeki modelin En Küçük Kareler (EKK) tekniğiyle bulunacak parametre tahminleri güvenilmez hale gelir. Bu problemle başedebilmek için Hoerl ve Kennard(1970) tarafından ridge regresyon analizi önerilmiştir.

$\boldsymbol{\beta}$ 'nın yanlı bir tahmin edicisi olan ridge regresyon tahmin edicisi

$$\hat{\boldsymbol{\beta}}_{\mathbf{R}}=(\mathbf{X'X}+k\mathbf{I})^{-1}\mathbf{X'y} \tag{2.2}$$

biçiminde verilir. Burada k yanlılık parametresi olarak bilinir.

(2.2) eşitliğinde verilen ridge regresyon tahmin edicisinin kanonik formu aşağıdaki gibi yazılabilir:

$$\hat{\boldsymbol{\alpha}}_{\mathbf{R}}(k)=(\mathbf{C'C}+k\mathbf{I})^{-1}\mathbf{C'C}\hat{\boldsymbol{\alpha}}$$

Burada $\mathbf{C}=\mathbf{XP}$, $\boldsymbol{\alpha}=\mathbf{P}'\boldsymbol{\beta}$ ve $\mathbf{P}$ ($n\times p$) boyutlu ortogonal bir matris $(\mathbf{P'P}=\mathbf{I})$'tir. $\boldsymbol{\alpha}$ 'nın herhangi bir tahmin edicisi olan $\hat{\boldsymbol{\alpha}}$, EKK tahmin edicisi $\hat{\boldsymbol{\beta}}$ ile $(\hat{\boldsymbol{\beta}}=\mathbf{P}\hat{\boldsymbol{\alpha}})$ bağlantılıdır. Bu durumda $HKO(\hat{\boldsymbol{\alpha}})=HKO(\hat{\boldsymbol{\beta}})$ yazılabilir.

Ridge tahmin edicisinin hata kareler ortalaması ise

$$HKO(\hat{\alpha}_R(k))=\sigma^2\sum_{i=1}^{n}\lambda_i(\lambda_i+k)^{-2}+\sum_{i=1}^{n}\left(\frac{k\alpha_i}{\lambda_i+k}\right)^2 \tag{2.3}$$

biçiminde hesaplanır ve ridge tahmin edicisinin HKO'nın EKK tahmin edicisinin HKO'dan daha küçük olacak şekilde bir yanlılık parametresi k 'nın (k>0) her zaman var olacağı Hoerl ve Kennard (1970) tarafından gösterilmiştir.

Literatürde k'nın seçimi için çeşitli metotlar önerilmiştir. Bu çalışmada uygun k değerini belirlemek için Hoerl, Kennard ve Baldwin(HKB) (1975) tarafından önerilen (2.4) eşitliğinde verilen formül kullanılacaktır:

$$k=\frac{p.\hat{\sigma}^2}{\hat{\alpha}'\hat{\alpha}} \tag{2.4}$$

2.2. Sağlam Tahmin

Sağlam sözcüğü istatistiksel varsayımlardan küçük sapmalara duyarsız istatistik işlemleri için kullanılmaktadır. Doğrusal regresyon analizinde β ’ların tahmini için en çok kullanılan teknik EKK tekniğidir. Ancak, bağımlı ve bağımsız değişkendeki aykırı değerler EKK için ciddi bir sorundur. Bu tekniğe alternatif olarak aykırı değerlerden çok fazla etkilenmeyen tahmin edicileri kullanan sağlam regresyon yöntemleri önerilmektedir.

Sağlam tekniklerin teorik anlamda performansını tanımlamak için etkinlik, bozulma(kırılma) ve sınırlı etkili olma özellikleri kullanılır (Simpson ve Montgomery, 1998).

En popüler sağlam tahmin edicilerden bazıları M, LTS, LMS, S ve GM tahmin edicileridir. M tahmin edici etkindir. LTS, LMS ve S tahmin edicileri yüksek bozulma noktasına sahiptir. GM tahmin edicisi etkin ve sınırlı etkili olma özelliğine sahip çok aşamalı bir tahmin edicidir.

2.2.1. Yüksek Bozulma Noktalı Tahmin Ediciler

Yüksek bozulma noktalı regresyon tahmin edicileri çok sayıda aykırı değerin varlığında güvenilir tahminler elde etmek için geliştirilmiştir. Bu tahmin ediciler %50 bozulma noktasına sahiptir ve dirençli(resistant) tahmin ediciler olarak bilinir. Yüksek bozulma noktalı tahmin ediciler hem x hem de y yönündeki aykırı değerlerin varlığı durumunda güvenilir parametre tahminleri verir (Hawkins ve Olive, 2002).

2.2.1.1. LMS Tahmin Edicileri

Rousseeuw (1984) tarafından bulunan ve Rousseeuw ve Leroy (1987) tarafından geliştirilen LMS metodu artık kareler toplamı yerine artık karelerin medyanının en küçüklenmesi fikrine dayanmaktadır. LMS tahmin edicileri;

$$\underset{\hat{\beta}}{\text{en küçük}}(\text{medyan } e_i^2)$$

olarak verilir, burada $e_i = y_i - x_i\hat{\beta}$ ’dır.

LMS tahmin edicisinin bozulma noktası %50’dir ve ölçek tahmini aşağıdaki gibi belirlenir:

$$\hat{\sigma} = \sqrt{\frac{\sum_{i=1}^{n} w_i . e_i^2}{\sum_{i=1}^{n} w_i - p}}$$

Burada

$$w_i = \begin{cases} 1 & \left| \mathrm{e_i} / s^0 \right| \leq 2.5 \\ 0 & \text{diğer durumlarda} \end{cases}$$

ve

$$s^0 = 1.4826 \left(1 + \frac{5}{n-p} \right) \sqrt{\underset{i}{med}\ \mathrm{e_i^2}}$$

olarak hesaplanır.

2.2.1.2. LTS Tahmin Edicileri

LTS Rousseeuw (1984) tarafından sunulan ve EKK'ya benzer bir en küçükleme işlemiyle elde edilen tahmin edicidir.

$$\underset{\hat{\beta}}{\text{en küçük}} \sum_{i=1}^{h} (e^2)_{i:n}$$

Burada $(e^2)_{1:n} \leq (e^2)_{2:n} \leq \ldots \leq (e^2)_{n:n}$ sıralı artık karelerin ve *h* toplama dahil edilen artık karelerin sayısıdır. Bu formülün EKK ile tek farkı en büyük kareli artıkların toplama işlemi dışında bırakılmasıdır(Rousseeuw ve Leroy, 1987).

2.2.1.3. S Tahmin Edicileri

Rousseeuw ve Yohai(1984) tarafından geliştirilen S tahmin edicileri ölçek istatistiğinin en küçüklenmesi fikrine dayandırılmıştır. Regresyon parametrelerinin S tahmini

$$\underset{\hat{\beta}}{\text{en küçük}}\, s(e_1(\beta), \ldots, e_n(\beta))$$

ve ölçek tahmin edicisi

$$\hat{\sigma} = s(e_1(\hat{\beta}), \ldots, e_n(\hat{\beta}))$$

şeklinde tanımlanır. Yayılma (dispersion) $s(e_1(\beta),...,e_n(\beta))$ olmak üzere $\frac{1}{n}\sum_{i=1}^{n}\rho\left(\frac{e_i}{s}\right)$ =K ifadesinin çözümü olarak tanımlanmıştır. $\rho(.)$ artık fonksiyonudur ve Rousseeuw ve Yohai(1984) aşağıdaki fonksiyonun kullanımını önermiştir:

$$\rho(x)=\begin{cases}\frac{x^2}{2}-\frac{x^4}{2c^2}+\frac{x^6}{6c^4} & |x|\leq c\\ \frac{c^2}{6} & |x|>c\end{cases}$$

S tahmin edicilerinin %50 bozulma noktasına sahip olmasını sağlayabilmek için c sabiti 1.548 olarak seçilir (Rousseeuw ve Leroy, 1987).

2.2.2. M Tahmin Edicileri

Huber (1964) tarafından önerilen bu tahmin tekniği, artıkların kareleri yerine artıkların başka bir fonksiyonunu en küçükleme fikrine dayanır.

M tahmin edicisinin amaç fonksiyonu

$$\underset{\hat{\beta}}{\text{en küçük}}\sum_{i=1}^{n}\rho(e_i)$$

biçiminde verilir (Rousseeuw ve Leroy, 1987). Huber'ın önerdiği $\rho(e)$ fonksiyonu

$$\rho(e)=\begin{cases}e^2 & -k\leq e\leq k\\ 2k|e|-k^2 & |e|>k\end{cases}$$

biçimindedir; burada k=1,5$\hat{\sigma}$ ve en çok kullanılan ölçek tahmin edicisi

$\hat{\sigma}$=1,483MAD

olarak alınır.

MAD= $medyan\,|\,e_i-medyan(e_i)\,|$

şeklinde tanımlanır(Huber, 1981).

M tahmin edicileri y yönündeki aykırı değerlere karşı sağlam tahmin edicilerdir.

2.2.3. GM Tahmin Edicileri

Bu tahmin edicilerin amacı bazı ağırlık fonksiyonlarını kullanarak x yönündeki aşırı değerlerin etkisini sınırlamaktır.

GM tahmin edicileri aşağıdaki biçimde verilen normal denklemlerin çözümü ile elde edilir:

$$\sum_{i=1}^{n} \pi_i \psi\left(\frac{y_i - x_i' \hat{\beta}}{s \pi_i} \right) x_i = 0$$

Burada π_i ağırlıkları ifade eder ve tanımlanan eşitlik Schweppe tarafından geliştirilmiştir.

Literatürde GM'in bileşenlerindeki (amaç fonksiyonu, başlangıç tahmini, ölçek tahmini, ağırlık fonksiyonu vb. gibi) farklılıktan dolayı farklı GM tahmin yaklaşımları önerilmiştir. Bunlardan biri Walker (1984)'ın GM tahmin yaklaşımıdır. Walker'ın GM yaklaşımı büyük artıklı gözlemlere düşük ağırlıklar veren Schweppe amaç fonksiyonunu kullanır. Başlangıç tahmini için EKK tekniği ve ölçek tahmini olarak da iteratif olmayan MAD önerilmiştir. Final tahminleri ise iteratif olarak yeniden ağırlıklandırılmış EKK tekniği kullanılarak elde edilir (Wisnowski, Montgomery ve Simpson, 2001).

2.3. Sağlam Ridge Tahmin Edicileri

Sağlam ridge tahmin edicileri

$$\hat{\boldsymbol{\beta}}_{SağlamRidge} = (\mathbf{X}'\mathbf{X} + k^{*}\mathbf{I})^{-1}\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}}_{Sağlam} \tag{2.4}$$

eşitliğiyle verilir, burada $\hat{\boldsymbol{\beta}}_{Sağlam}$ sağlam tahmin edicilerden elde edilen parametre tahminlerini gösterir. Çalışmada M, GM, LTS, LMS ve S sağlam tahmin edicileri kullanılmıştır. Yanlılık parametresi değeri sağlam teknikler için

$$k^{*} = \frac{p.\hat{\sigma}^2_{Sağlam}}{\hat{\boldsymbol{\beta}}'_{Sağlam}\hat{\boldsymbol{\beta}}_{Sağlam}} \tag{2.5}$$

eşitliği ile belirlenir. Burada p bağımsız değişken sayısı, $\hat{\sigma}^2_{Sağlam}$ sağlam regresyondan elde edilen ölçek tahmin edicisidir.

M ve GM tahmin edicilerine dayalı sağlam ridge tahmin edicilerinin hata kareler ortalaması

$$HKO(\hat{\boldsymbol{\alpha}}_{Sağlam}(k^*)) = \sum_{i=1}^{n} \lambda_i(\lambda_i + k^*)^{-2}\Omega_{ii} + \sum_{i=1}^{n}\left(\frac{k^*\alpha_i}{\lambda_i + k^*}\right)^2$$

biçiminde elde edilir, burada $\boldsymbol{\Omega}$ *(pxp)* boyutlu kov($\hat{\boldsymbol{\alpha}}_{\text{Sağlam}}$) matrisidir (Silvapulle(1991), Arslan ve Billor(1996)). Bu çalışmada LTS, LMS ve S tahmin edicilerine dayalı sağlam ridge tahmin edicilerinin hata kareler ortalaması için de aynı formül uyarlanmıştır.

3. Uygulama

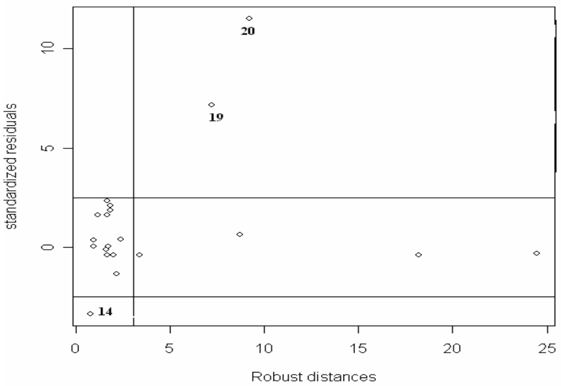
Çalışmanın uygulama kısmında Türkiye'deki turizm verileri kullanılmıştır. Turizmde talebin doğması ve gelişmesini pek çok etmene bağlamak mümkün olmakla birlikte ekonomik ve ekonomik olmayan faktörler şeklinde ikili bir ayrıma tabi tutmak olasıdır (Dinçer, 1993). Gelen turistlerin ihtiyaçlarını karşılayabilecek nitelikte konaklama tesis sayısı, turistlerin gelir durumu, turizm tesislerindeki oda sayısı, ulaşım olanakları ve belgeli yat işletme sayısı gibi etkenler ekonomik faktörleri oluşturur. Literatürde turizm talebini etkileyen faktörlerin belirlenmesi ve bu talebe ilişkin en uygun modelin belirlenmesi yönünde çok sayıda çalışma bulunmaktadır. Bu çalışmada turizm talebini en iyi açıklayan modeli bulmak yerine bu talepte aykırı değerlerin ve çoklu bağıntının olması durumunda kullanılabilecek sağlam bazı teknikler incelenmiştir. Bu amaçla Türkiye'ye gelen turist sayısı değişkenini açıklayabileceği düşünülen ekonomik faktörler arasından tesis sayısı, turizm tesislerindeki oda sayısı ve belgeli yat işletme sayısı bağımsız değişkenler olarak alınmıştır.

Türkiye'de turizm sektörünün gelişimi 1980'li yıllarda hız kazanmaya başladığından çalışmada 1986 yılı ile 2005 yılı arasındaki veriler kullanılmıştır. Veri setindeki gelen turist sayısı değişkenine ait veriler T.C. Başbakanlık Türkiye İstatistik Kurumu, diğer değişkenlere ait veriler ise T.C. Kültür ve Turizm Bakanlığı'nın Turizm İstatistikleri verilerinden derlenmiştir.

Çalışmada veriler standartlaştırılarak kullanılmış ve $\mathbf{X'X}$ matrisi için özdeğerler $\lambda_1 = 2{,}2130$, $\lambda_2 = 0{,}7809$ ve $\lambda_3 = 0{,}0061$ olarak belirlenmiştir. Bu özdeğerler yardımıyla hesaplanan koşul sayısı (361,6) çoklu doğrusal bağıntının varlığını göstermektedir.

Aynı zamanda veri setindeki aykırı değerlerin varlığını sağlam artık grafiklerinde görmek mümkündür. Şekil 1' den $(y_{14}, x_{14}), (y_{19}, x_{19})$

ve (y_{20}, x_{20}) gözlemlerinin x ve y yönünde aykırı değerler oldukları söylenebilir.



Şekil 1. Turizm verileri için sağlam artık grafiği

EKK sonucu elde edilen çoklu doğrusal regresyon katsayıları ve EKK tahmin edicilerine dayalı olarak elde edilen ridge regresyon katsayıları Tablo 1'de verilmiştir.

Tablo 1. EKK ve ridge regresyon parametre tahmin değerleri, yanlılık katsayısı ve artık kareler ortalaması

Katsayı ve Ölçek Tahminleri	EKK	RIDGE
$\hat{\beta}_1$	-0,7857	0,3887
$\hat{\beta}_2$	1,7490	1,3715
$\hat{\beta}_3$	-0,0843	-0,1402
AKO	0,5748	0,3071
	$\hat{\sigma}$ =0,059	k=0,0028

Çoklu doğrusal bağıntı ve aykırı değer varlığında sağlam ridge regresyon tahminlerini elde etmek için öncelikle sağlam regresyon tahmin değerleri hesaplanmış ve bu tahmin değerleri (2.5) eşitliğinde kullanılarak yanlılık parametre değerleri(k^*) bulunmuştur. Sağlam ridge regresyon tahmin değerleri Tablo 2'de verilmiştir.

Tablo 2. Sağlam ridge regresyon parametre tahmin değerleri, yanlılık katsayısı ve artık kareler ortalaması

	M	GM	LTS	LMS	S
$\hat{\beta}_1$	0,3725	0,3485	0,3487	0,0890	0,1995
$\hat{\beta}_2$	1,3210	1,2724	1,2590	0,9418	1,0277
$\hat{\beta}_3$	-0,1082	-0,1360	-0,0696	-0,2952	-0,0020
k^*	0,0025	0,0026	0,0016	0,0023	0,0026
AKO	0,2226	0,2160	0,1393	0,1008	0,1305

Tablo 2'ye göre, bulunan tüm sağlam tahmin edicilere dayalı ridge parametre tahmin değerlerinin işaretleri beklentilerimiz doğrultusunda gerçekleşmiştir. Tesis sayısı ve turizm tesislerindeki oda sayısı değişkenleri ülkemize gelen turist sayısını pozitif yönde etkilerken, belgeli yat işletme sayısı değişkeni 2000 yılı ekonomik krizinin etkisiyle verilerdeki azalma neticesinde negatif yönde etki göstermektedir. Bununla birlikte literatürde turizm talebini etkileyen faktörlerin belirlenmesi ve bu talebe ilişkin en uygun modelin belirlenmesi yönünde çok sayıda çalışma bulunmaktadır. Yapılan çalışmalarda modellere ilişkin varsayım bozulumları (otokorelasyon, değişen varyanslılık ve çoklu bağıntı gibi) araştırılırken verilerde aykırı değer olabileceği göz ardı edilmektedir. Bu çalışmada turizm talebini en iyi açıklayan modeli bulmak yerine bu talepte aykırı değerlerin de olması durumunda hangi tekniklerin kullanılabileceği üzerinde durulmuştur. Bu amaçla çalışmada alınan değişkenlerin modeli açıklayan en uygun değişkenler olmasından ziyade aykırı değer içeren ve aynı zamanda çoklu bağıntının olduğu değişkenlerin çalışmaya alınması hedeflenmiştir.

Tablo 2'de verilen sağlam ridge regresyon parametre tahmin değerlerinin artık kareler ortalaması Tablo 1'de verilen EKK'ya dayalı ridge regresyon parametre tahmin değerlerinin AKO'sından oldukça küçük bulunmuştur. Buna göre veri seti aykırı değer içerdiğinde yanlılık parametre değerinin belirlenmesinde sağlam regresyon tekniklerinin kullanılmasının daha etkin sonuçlar verdiği görülmektedir. Diğer taraftan, veri setindeki aykırı değerlerin niteliğinin (sayısı, büyüklüğü, konumu vb.) bu sağlam tekniklerden hangisinin benimseneceği konusunda yapılacak tercihte önemli olduğu açıktır.

4. Sonuç ve Tartışma

Bu çalışmada aykırı değer ve çoklu doğrusal bağıntının var olduğu bir veri setinde yanlılık parametre değerinin belirlenmesinde çeşitli sağlam teknikler (M, GM, LTS, LMS, S) kullanılmış, bu tekniklerden bulunan $\hat{\beta}_{Sağlam}$ ve $\hat{\sigma}_{Sağlam}$ değerleriyle k^* yanlılık parametresi hesaplanmış ve daha sonra sağlam ridge parametre tahmin değerleri elde edilmiştir.

Kullanılan sağlam tekniklerin başarısı aykırı değerlerin x uzayının içinde veya dışında olmasından, büyüklüğünden ve yoğunluğundan etkilenmektedir. Bu nedenle çıkan tekniklerin performansı bu bileşenler dikkate alınarak yorumlanmalıdır.

Çalışmada yer verilen LTS, LMS ve S tahmin edicileri yüksek bozulma noktasına (%50) sahip tahmin ediciler olduklarından bu özellik açısından aynı sınıfta yer alırlar. Ayrıca bu tahmin ediciler x yönündeki aykırı değerlere karşı sağlam tahmin edicilerdir.

M ve GM tahmin edicileri ise LTS, LMS ve S tahmin edicilerine göre daha düşük bozulma noktalarına (M için 1/n, ve GM için 1/p) sahip olduklarından bu özellik açısından aynı sınıf içinde değerlendirilebilirler. Ancak M tahmin edicileri y yönündeki aykırı değerlere karşı sağlam iken, GM tahmin edicileri x yönündeki aykırı değerlere karşı da sağlamdır.

Bu çalışmadaki veriler hem x hem de y yönünde aykırı değerler içermektedir. Ayrıca veri setinde x yönünde birden fazla aykırı değer olduğu gözükmektedir ve bu aykırı değerlerden biri aşırı büyüktür. Bu koşullar altında sağlam tahmin edicilere dayalı ridge regresyon tahmin edicilerinin performansları kıyaslandığında (HKO bakımından) LTS, LMS ve S tahmin edicilerinin daha iyi performans göstermeleri beklenen bir durumdur.

Bu çalışmada GM tahminlerinin başlangıç değerleri EKK tekniğiyle elde edilmiştir. Eğer başlangıç değeri olarak yüksek kırılma noktalı tahmin edicilerden biri seçilirse GM tahmin edicilerine dayalı ridge regresyon tahminlerinin performansının daha iyi olması beklenmektedir.

Kaynakça

Arslan O. ve N. Billor (1996), "Robust ridge regression estimation based on the GM estimators", *Journal of Math. and Comp. Sci.(Math Ser.)*, 9:1, 1-9.

Askin, G.R. ve D.C. Montgomery (1980), "Augmented Robust estimators", *Techonometrics*, 22, 333-341.

Dinçer, M.Z. (1993), *Türkiye Ekonomisi ve Türkiye Ekonomisinde Turizm*, Filiz Kitabevi: İstanbul.

Hawkins, D.M. ve D.J. Olive (2002), "In consistency of resampling algorithms for high breakdown regression estimators and a new algorithm", *Journal of the American Statistical Association*, 97:136-148.

Hoerl A.E. ve R.W. Kennard (1970), "Ridge regression: Biased estimation for nonorthogonal problems", *Technometrics*, 12(1), 55-67.

Hoerl A.E., R.W. Kennard ve K.F. Baldwin (1975), "Ridge Regression: Some simulations", *Commun.Statist. A-Theory Methods*, 4, 105-123.

Huber P.J. (1964), "Robust Estimation of a Location Parameter", *Ann. Math. Stat.,* 35, 73-101.

Huber P.J. (1981), *Robust Statistics*, Wiley: New York.

Pfaffenberger, R.C. ve T.E. Dielman (1990), *A comparison of regression estimators when both multicollinearity and outliers are present*. In Robust Regression (ed. Lawrence and Arthur), 243-270.

Rousseeuw P.J. (1984), "Least Median of Squares Regression", *Journal of the American Statistical Association*, 79, 871-880.

Rousseeuw P.J. ve V.J. Yohai (1984), "Robust Regression by Means of S-Estimators", *Robust and Nonlinear Time Series Analysis*, eds. J.Franke, W. Hardle, and D. Martin, Springer-Verlag: Heidelberg, Germany, 256-272.

Rousseeuw P.J. ve A.M. Leroy (1987), *Robust Regression and Outlier Detection*, Wiley: NewYork.

Silvapulle M.J. (1991), "Robust ridge regression based on an M estimator", *Austral.J. Statist*., 33(3), 319-333.

Simpson J.R. ve D.C. Montgomery (1998), "A Robust Regression Technique Using Compound Estimation", Naval Research Logistics, 45, 125-139.

Walker E. (1984), "Influence, Collinearity, and Robust Estimation in Regression", unpublished Ph.D. dissertation, Department of Statistics, Virginia Polytechnic Institute.

Wisnowski J.W., D.C. Montgomery ve J.R. Simpson (2001), "A Comparative Analysis of Multiple Outlier Detection Procedures in the Linear Regression Model", *Computational Statistics and Data Analysis*, 36, 351-382.