T.C.
DİCLE ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

ORTALAMA ÖTELEMELİ SAPAN DEĞER MODELİNDE M-TAHMİN YÖNTEMİ VE KONİK PROGRAMLAMA İLE PARAMETRE TAHMİNİ

Burcu BİLGİÇ UÇAK

TEZ DANIŞMANI: DOÇ. DR. Pakize TAYLAN

YÜKSEK LİSANS TEZİ

MATEMATİK ANABİLİM DALI

DİYARBAKIR
Nisan – 2016

T.C. DİCLE ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ MÜDÜRLÜĞÜ
DİYARBAKIR

Burcu BİLGİÇ UÇAK tarafından yapılan "Ortalama Ötelemeli Sapan Değer Modelinde M-Tahmin Yöntemi Ve Konik Programlama İle Parametre Tahmini" konulu bu çalışma, jürimiz tarafından Matematik Anabilim Dalında YÜKSEK LİSANS tezi olarak kabul edilmiştir.

Jüri Üyeleri

Başkan :

Üye :

Üye :

Tez Savunma Sınavı Tarihi: .../...../.....

Yukarıdaki bilgilerin doğruluğunu onaylarım.

.../...../.....

Prof. Dr. …..

Enstitü Müdürü

TEŞEKKÜR

Yüksek lisans çalışmamın her aşamasındaki sonsuz desteğinden, yol göstericiliğinden, bitmeyen sabrından ve yüreklendirmelerinden dolayı tez danışmanım, saygıdeğer hocam, Doç. Dr. Pakize TAYLAN' a sonsuz teşekkürlerimi sunuyorum.

Çalışmamın uygulama aşamasında yardımlarını esirgemeyen Sayın Fatma YERLİKAYA ÖZKURT' a saygılarımla; çalışmamın arka planında desteklerini ve inançlarını esirgemeyen değerli ailem; anneme, babama, bilgi birikimi ve deneyimlerini esirgemeyen ablam Yrd. Doç. Dr. Fundagül BİLGİÇ' e ve tezimi düzenleme aşamasında takıldığım bütün noktalarda önümü açmama yardımcı olan sevgili kuzenim Diren SARISALTIK YAŞIN' a ve biricik eşim Ali UÇAK' a sevgiyle teşekkür ederim.

İÇİNDEKİLER

Sayfa

TEŞEKKÜR……...…...I

İÇİNDEKİLER…….…..II

ÖZET……..IV

ABSTRACT…….……V

ÇİZELGE LİSTESİ……...VI

ŞEKİL LİSTESİ……...…….VII

KISALTMA VE SİMGELER……..……..VIII

1. GİRİŞ…….……1

2. REGRESYON ANALİZİ……..……5

2.1. Lineer Regresyon Modeli……..……..……5

2.2. En Küçük Kareler Yöntemi…….5

2.3. Ridge Regresyon ……7

2.4. En Küçük Mutlak Küçültme ve Operatör Seçimi (LASSO) ……10

3. KONİK KARESEL PROGRAMLAMA ……15

4. SAPAN DEĞERLER…….21

4.1. Sapan Değerin Tanımı……...21

4.2. Kaldıraç Nokta (Leverage Point)……...23

4.3. Maskeleme ve Gölgeleme Etkileri……24

5.SAPAN DEĞER TEŞHİS YÖNTEMLERİ……..27

5.1. Doğrudan Yöntemler…….27

5.1.1. Basit Kalan Yöntemi……..27

5.1.2. Standartlaştırılmış Kalan Yöntemi…….28

5.1.3. Studentized Kalan Yöntemi ……..30

5.1.4. Cook Mesafesi ……...32

5.2. Doğrudan Olamayan Yöntemler ……...33

5.2.1. Sağlam (Robust) Regresyon ……..33

5.2.1.1. En Küçük Kesilmiş Kareler (LTS) Tahmin Yöntemi ……...34

5.2.1.2. En Küçük Medyan Kareler (LMS) Tahmin Yöntemi ……...34

5.2.1.3. S Tahmin Yöntemi ……35

5.2.1.4. M- Tahmin Yöntemi ……...36

6. ORTALAMA ÖTELEMELİ SAPAN DEĞER (OÖSD) MODELİ……...…41

6.1. Ortalama Ötelemeli Sapan Değer Modeli İçin Parametre Tahmini……..43

6.1.1. M- Tahmin Yöntemi ile OÖSD de Parametre Tahmini ……43

6.1.2. OÖSD Modelinin M- Tahmin Edicisini Elde Etmek İçin Tikhonov Düzenlemesi Yöntemi……46

6.1.3. OÖSD Modelinin M- Tahmin Edicisini Elde Etmek İçin Konik Karesel Programlama ……49

6.1.4. OÖSD Modelinin M- Tahmini İçin LASSO Yöntemi (M_L-OÖSD)……51

7. SAYISAL UYGULAMA VE LM, OÖSD VE CT-OÖSD NİN PERFORMANSLARININ KARŞILAŞTIRILMASI ……..…… 57

7.1. Veri Kümesi ……..57

7.2. Karşılaştırmada Kullanılan Performans Ölçütleri ……58

7.3. Sapan Değer Teşhisi ve Modelin Oluşturulması ……..60

7.4. LM, OÖSD ve CT-OÖSD Modellerinin Performanslarının Sayısal Karşılaştırılması…....62

7.5. Tartışma ve Sonuç……63

8. KAYNAKLAR ……..…….65

ÖZGEÇMİŞ……...…….73

ÖZET

ORTALAMA ÖTELEMELİ SAPAN DEĞER MODELİNDE M-TAHMİN YÖNTEMİ VE KONİK PROGRAMLAMA İLE PARAMETRE TAHMİNİ

YÜKSEK LİSANS TEZİ

Burcu BİLGİÇ UÇAK

DİCLE ÜNİVERSİTESİ

FEN BİLİMLERİ ENSTİTÜSÜ

MATEMATİK ANABİLİM DALI

2016

Bu tez çalışması, sapan değerlerle bozulmuş bir veri kümesi üzerinde, öncelikle sapan değerlerin teşhis edilmesi, ardından sapan değerin sahip olduğu bilgiyi göz ardı etmemek için Ortalama Ötelemeli Sapan Değer (OÖSD) modelin oluşturulmasını amaçlamaktadır. Oluşturulan modelin parametreleri, çok önemli ve direkt olmayan sağlam bir sapan değer yöntemi olan M-tahmin yöntemi kullanılarak tahmin edildi. Doğrusal regresyon modelindeki sapan değer probleminin üstesinden gelmek için, M-tahmin edicilerin sağlamlığını Tikhonov Düzenleme ve En Küçük Mutlak Küçültme Ve Operatör Seçimi (LASSO)'nun kararlılığı ile birleştiren iki yöntem geliştirildi.

Bunun için öncelikle Huber tipi fonksiyon ile M-tahmin yöntemine dayanarak Tikhonov düzenleme ve LASSO problemi OÖSD modeline uyduruldu. Daha sonra bu problemin çözümü için iç noktalar yöntemini kullanan konik karesel programlama (CQP) yöntemi önerildi. Burada amaç, modeli sapan değerlerin olumsuz etkilerinden korumaktır. Ayrıca, verilen problem için en uygun ayar sabitinin nasıl hesaplanacağı üzerine öneri getirildi. Daha sonra oluşturulan modeller amonyağı nitrik aside oksitleyen bir düzeneğin 21 günlük çalışması sonucu elde edilen veri grubuna uygulandı. Bu uygulamada MATLAB ve MOSEK paket programları kullanılmıştır.

Anahtar Kelimeler: Sapan değerler, Tikhonov Düzenleme, LASSO, M-tahmin, Sürekli En iyileme

ABSTRACT

MEAN SHIFT OUTLIER MODELS BY M- ESTIMATION METHOD AND PARAMETER ESTIMATION WITH CONIC PROGRAMMING

MSc THESIS

Burcu BİLGİÇ UÇAK

DEPARTMENT OF MATHEMATICS

INSTITUTE OF NATUREL AND APPLIED SCIENCES

UNIVERSITY OF DICLE

2016

This thesis aims the constitution of Mean Shift Outlier Model (MSOM) on a data set contaminated with outliers, after detection of outliers to not disregard the information possessed by the outliers. The parameters of this model were estimated by using M-estimation method which is a very important and indirect robust outlier detection method. Robustness of M-estimators were combined with the efficiency of Tikhonov Regularization and Least Absolute Shrinkage and Selection Operator (LASSO) to overcome the problem of outliers in linear regression model.

Therefore, firstly Tikhonov Regularization and LASSO problems were applied to the Mean Shift Outlier Model (MSOM) based on Huber type M-estimation method. Then, the conic quadratic programming method that uses the interior point method was proposed for solving this problem. Here, The aim is to protect the model from the negative effects of outliers. Moreover, a proposal was introduced on how the calculation of the optimal tuning constant for the given problem. Then, the established models were applied to the data set which was obtained by 21-day operation of a plant for the oxidation of ammonia to nitric acid. For that application MATLAB and MOSEK software packages were used.

Keywords: Outliers, Tikhonov Regularization, LASSO, M-estimation, Contunious Optimization

ÇİZELGE LİSTESİ

Çizelge No **Sayfa**

Çizelge 4.1. Farklı Tipteki Sapan Değerler ……22

Çizelge 5.1. M-tahmin yönteminde yaygın olarak kullanılan bazı $\rho(x),\psi(x),w(x)$ fonksiyonları ……...37

Çizelge 7.1. Amonyağı nitrik asite oksitleyen bir düzeneğinin 21 günlük çalışma sistemi için elde edilen veriler…….58

Çizelge 7.2. EKK, OÖSD ve CT-OÖSD alternatif modelleri için performans ölçümü…….……63

ŞEKİL LİSTESİ

Şekil No **Sayfa**

Şekil 2.1. En küçük kareler yönteminin grafiksel olarak gösterimi..…….……7

Şekil 2.2. Tikhonov düzenlemesi problemlerinin grafiksel gösterimi …….……..…..8

Şekil 2.3. λ nın seçimi için L-eğrisi..……..…….……...…...10

Şekil 2.4. Lasso problemine ve Ridge problemine ilişkin tahmin şekilleri…..……..12

Şekil 4.1. Farklı tipte gözlem değerleri için serpme diyagramı.…..……....22

Şekil 5.1. Sapan değerin bulunduğu ve bulunmadığı veri kümesi için tahmin edilmiş regresyon doğruları. ……...…….....29

Şekil 7.1. Teşhis edilen sapan değerler……....……61

KISALTMA VE SİMGELER

Kısaltmalar

Adj-R2	:	Düzeltilmiş belirlilik katsayısı
böl.	:	Bölüm
CC	:	Korelasyon Katsayısı
CM_i	:	Cook Mesafesi
CQP	:	İkinci dereceden konik karesel programlama
diag	:	Köşegen elemanlarından oluşan matris
EKK	:	En Küçük Kareler
GCV	:	Genelleştirilmiş Çapraz Doğrulama
GM	:	Genelleştirilmiş M-tahmin edicileri
HKO	:	Hata Kareleri Ortalaması
HKOK	:	Hata Karelerinin Ortalamasının Karekökü
HKT	:	Hata Kareleri Toplamı
IRLS	:	Ardışık olarak ağırlaştırılmış en küçük kareler (*Iteratively Reweighted Least Squares)*
i.e.	:	Diğer bir deyişle
LASSO	:	En Küçük Mutlak Küçültme ve Operatör Seçimi
LM	:	Lineer Model
LMS	:	En Küçük Ortanca Kareler
LTS	:	En Küçük Kesilmiş Kareler

k.s. : Kısıt

M : M tahmin edicileri

$MAD(\,.\,)$: Ortalama Mutlak Sapma

Maks : En büyük

Med : Medyan

Min : En küçük

ML-OÖSD : M-Tahmini için LASSO Yöntemi

OMH : Ortalama Mutlak Hata

OÖSD : Ortalama Ötelemeli Sapan Değer

RMSE : Ortalama Hata Karelerinin Karekökü

RR : Ridge Regresyon

S : S tahmin edicileri

SOCP : İkinci dereceden konik programlama

$\mathrm{sgn}(.)$: İşaret fonksiyon

$Var(\,.\,)$: Varyans

Mantık

= : Eşittir

$\succeq$: Kısmi sıralama veya vektör eşitsizliği

< : Küçüktür

> : Büyüktür

$\geq$: Büyük veya eşittir

$\leq$: Küçük veya eşittir

Matematiksel Simgeler

$\mathbb{N}_0$: Doğal sayılar kümesi

$\mathbb{R}$: Gerçel sayılar kümesi

$\mathbb{R}^n$: n boyutlu gerçel sayılar kümesi

$\bigcup$: Çoklu birleşim

$\bigcap$: Kesişim

$\times$: Kartezyen çarpım

$\|\cdot\|_2$: Öklit normu

İstatistiksel Simgeler

$\boldsymbol{I}_n$: n boyutlu birim matris

H : Şapka matrisi

h_{ii} : Şapka matrisinin köşegen elemanları

H_0 : Başlangıç hipotezi

H_1 : Karşıt hipotez

$\mathbf{L}^n$: n-boyutlu ikinci dereceden veya Lorentz tipi koni

p : Modelde bulunan parametre sayısı

r_i : i. veriye ait kalan değeri

s_i : Standartlaştırılmış kalan

s_i^* : Jackknife (i. dıştan Studentized) kalanlar

$\boldsymbol{y}$: $(n\times 1)$ boyutlu yanıt değişkenlerin vektörü

σ : Standart sapma

$\hat{\sigma}$: Standart sapmanın tahmini

x : Açıklayıcı değişken

$\boldsymbol{X}$: $(n\times p)$ boyutlu bağımsız veya açıklayıcı değişkenlerin tam ranklı matrisi

w : Ağırlık fonksiyonu

$\boldsymbol{\beta}$: $p\times 1$ boyutlu bilinmeyen parametreler vektörü

$\hat{\boldsymbol{\beta}}$: $\boldsymbol{\beta}$ nın EKK tahmini

$\hat{\boldsymbol{\beta}}^{ridge}$: $\boldsymbol{\beta}$ nın ridge tahmini

$\hat{\boldsymbol{\beta}}^{lasso}$: $\boldsymbol{\beta}$ nın LASSO tahmini

β^* : OÖSD modelinde $(p+m)\times 1$ boyutlu parametre vektörü

$\hat{\beta}_M^*$: : β^* ın M-tahmin edicisi

$\hat{\beta}^{*M=lasso}$: OÖSD modelinin LASSO ile M- tahmincisi

λ : Düzenleme parametresi

$\hat{\lambda}$: λ nın tahmin edicisi

$\psi(.)$: Etki fonksiyonu

$w(.)$: Ağırlık fonksiyonu

1. GİRİŞ

İstatistiksel çıkarsamalar gözlemler sonucu elde edilen verilere dayalı olarak oluşturulan regresyon modellerine göre yapılır. Bu modellerden biri doğrusal regresyon modelidir. Bu modelleri içeren çok sayıda bilimsel çalışma mevcuttur (Vovk ve ark. 2009, Biçkici 2007). Bununla beraber doğrusal regresyon modeline dayalı veri analizinin sonuçları model seçimine ve veri kümesindeki sapan değerlere karşı oldukça duyarlı olduğundan, sapan değer teşhisi konusunda çok sayıda bilimsel çalışma yapılmıştır. Bu konudaki çalışmalar sapan değer teşhisini direkt ve dolaylı olmak üzere iki şekilde inceleyerek ele alınmıştır. Bu konuda yapılan ilk çalışmalar, en çoktan en aza aykırı değerlere doğru hem teşhis hem de test anlamında ileri adımda, tek bir sapan değerin teşhis edilmesi üzerine yoğunlaşmıştır (Prescott 1975, Tietjen ve ark. 1973). Ancak, bu şekildeki yöntemlerde gölgeleme ve maskeleme problemleri ile karşılaşılabilir ve sapan değer sayısının birden fazla olması durumunda, bunlara bağlı yapılan testler önemli ölçüde güçlerini kaybeder (Barnett ve Lewis 1984, Hawkins 1980, Kianifard ve Swallow 1989, 1990, Marasinghe 1985). Marasinghe (1985) bu şekildeki bir yöntemi çok adımlı bir yaklaşıma uydurdu. Buna göre, K tane aday sapan değeri seçtikten sonra tek bir sapan değerin teşhisinin dizisel uygulamasını kullanarak, birinin anlamsızlığı belirleninceye kadar her aday sapan değerin aykırılığı test edilir. Paul ve Fung'ın (1991) genelleştirilmiş uç studentized kalanlar yöntemi, K nın daha iyi bir tahminini elde etmek için yöntemin test oranını değiştirdi. Bununla birlikte bu yöntemlerin her ikisinde de aday sapan değer sayısı K ya ihtiyaç duyulur. Bu yöntemlerdeki karşılaşılabilen maskeleme ve gölgeleme problemlerine karşı Paul ve Fung (1991), genelleştirilmiş uç studentized kalanlar yönteminden elde edilen aday sapan değerlerin, en büyük dizisel Cook mesafeli (Cook 1977) K gözlemin kümesi ile birleştiği iki aşamalı bir yöntem önerdi. Birleştirilen bu gruptaki gözlemler en küçük uç değerden en büyük uç değere kadar genelleştirilmiş uç studentized kalanlar yöntemi ile test edilir. Bu konuda farklı bir çalışma Kianifard and Swallow (1989, 1990) tarafından yapılmıştır. Bu çalışmada, birinci adımda tek bir sapan değer teşhisi ile gözlemler sıralanır, ikinci adımda ise sapan değer olarak belirlemek için teşhis edilmiş en küçük K gözlem sayısına dayalı olarak ardışık kalanlar kullanılır.

Yukarıda belirtilen direkt sapan değer yöntemlerinin yanında direkt olmayan yöntemleri içeren önemli çalışmalar da yapılmıştır. Bu çalışmalar, tahmin probleminin sapan değerlerden etkilenmediği, sağlam tahmin yöntemlerine bağlı olarak yapılmıştır. Direkt olmayan sapan değer teşhis yöntemlerinin en çok kullanılan ve önemli çalışmalarından birisi Huber (1973) tarafından yapılmıştır. Bu çalışmada Huber, M-tahmininin tek değişkenli tahmin kavramının regresyona uygulanmasını önerdi. Burada, $\rho(\cdot)$ kalan karelerine göre sapan değerden daha az etkilenen bir fonksiyon olmak üzere,

$$\min_{\boldsymbol{\beta}} \sum \rho(\boldsymbol{r}_i)$$

amaç fonksiyonunun çözümünden elde edilen parametre değerini seçti. M-tahmin ediciler yeterli ve yanıt değeri $\boldsymbol{y}$ nin sıra dışı değerlerine karşı oldukça sağlamdır, ama tek bir yanıltıcı kaldıraç nokta bu kestiricilerin performansını tamamen bozabilir. Bu nedenle, sınırlanmış etki kestiriciler olarak da bilinen ve aşağıdaki promlemin çözümü olan genelleştirilmiş M-tahmin edici tanımlandı (Krasker ve Welsch 1982):

$$\sum_i w(x_i)\psi\left(r_i/\hat{\sigma}w\left(x_i\right)\right)x_i = 0\,.$$

Burada $w(\cdot)$ ağırlık fonksiyonudur. Bu tahminler kaldıraç noktalara karşı daha az duyarlı olmakla birlikte yine de ağırlık fonksiyonu sapan değerlere karşı büyük ölçüde dirençli olmadığı sürece regresyon modelindeki parametre sayısının bir fonksiyonu olan bir kırılma noktasına sahiptir.

Son yıllarda yüksek kırılma yöntemleri üzerine dikkat çekici çalışmalar yapılmaktadır. Bu metotların çoğu, kalanların karelerinin toplamından daha güçlü bir ölçek tahminini en küçüklemeye dayalıdır. Rousseeuw (1984) LMS olarak adlandırılan ve kalanların karelerinin toplamı yerine medyanını en küçükleyen bir yüksek kırılma yöntemi önermiştir. Ancak LMS kestiricileri yüksek derecede yetersiz görülen bir yöntemdir. Buna karşın daha yeterli yüksek kırılma kestiricileri üzerine birçok çalışma yapılmıştır. Bunlardan en dikkat çekeni MM, tau ve GM tahmincilerinin bir sınıfıdır (Simpson ve ark. 1992, Yohai 1987, Yohai ve Zamar 1988). MM tahminciler başlangıç tahmini olarak daha az etkili olan bir yüksek kırılma yöntemi kullanır, daha sonra yeniden ψ azalan fonksiyonuna dayalı bir M- tahmin staratejisi kullanır. Normal hatalar altında kestiricinin yeterliliği ψ fonksiyonunun parametrelerinden birine bağlı olarak

ayarlanabilir. Tau kestiricisi yeterli sağlam ölçek kestiricilerinin genel bir sınıfına ait bir üyenin en küçük değeri olarak tanımlanır. Simpson ve arkadaşlarının (1992) çalıştığı GM kestiricileri β ve σ^2 nın yüksek kırılmalı başlangıç tahminlerine bağlı olan tek adımlı tahminlerdir. Rousseeuw ve Leroy (1987), LMS yöntemiyle bir dereceye kadar sapan değer olarak göz önüne alınan noktaların işaretlenerek ve işaretlenmeyen noktaların basit en küçük kareler yönteminin uygulanmasıyla LMS tahmin edicilerinin yetersizliğinin üstesinden geldi. Önemli sapan değer yöntemlerinden biri de, ortalama ötelemeli sapan (OÖSD) modelidir. Bu model, e_i i. birim vektör olamak üzere,

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e}_i\delta + \boldsymbol{\varepsilon}, \tag{6.1}$$

eşitliğiyle verilir (Rao ve Toutenburg 1999). Bu modelde, i. yanıt değeri y_i ya da $\boldsymbol{x}_i\boldsymbol{\beta}$ değerlerinin sistematik olarak δ kadar $y_i = \boldsymbol{x}_i^T\boldsymbol{\beta} + \varepsilon_i$ modelinden sapma gösterdiği kabul edilir. Böylece i. gözlem değeri geriye kalan gözlemlerden farklı bir yol izleyeceğinden sapan değer olarak değerlendirilir. Bu durum,

$$H_0 : \delta = 0 \quad \left(\text{i.e., } y_i = \boldsymbol{x}_i^T\boldsymbol{\beta}\right)$$

hipotezinin

$$H_1 : \delta \neq 0 \quad \left(\text{i.e., } y_i \neq \boldsymbol{x}_i^T\boldsymbol{\beta}\right),$$

alternatif hipotezine karşı olabilirlik-oran(likelihood ratio) yöntemine dayalı olarak test edilir. Bu hipotez için test istatistiği olarak,

$$F_i = \frac{\left[SSE\left(H_0\right) - SSE\left(H_1\right)\right]/1}{SSE\left(H_1\right)/(n-p-1)}. \tag{6.2}$$

alınır (Rao ve Toutenburg, 1999: 221).

2. REGRESYON ANALİZİ

2.1. Lineer Regresyon Model

Regresyon analizinin (Rencher ve Schaalje 2008:2) amacı değişkenler arasındaki ilişkiyi belirlemektir. Bu ilişkiyi belirlemek için değişik yöntemler kullanılır. Bu yöntemlerin en yaygın olanlarından birisi doğrusal regresyon yöntemidir (Rao ve Toutenburg 1999: 23). Bu yöntemde açıklayıcı veya bağımsız değişken **X** ve yanıt değişken veya bağımlı değişken arasındaki ilişkinin doğrusal olduğunu öngörür. Standart doğrusal regresyon modeli (Rencher 2000:179-191):

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{2.1}$$

olarak tanımlanır. Burada $\boldsymbol{y}, (n \times 1)$ boyutlu yanıt değişkenlerin vektörü, $\boldsymbol{X}$, $(n \times p)$ boyutlu bağımsız veya açıklayıcı değişkenlerin tam ranklı matrisi, $\boldsymbol{\beta}$, $p \times 1$ boyutlu bilinmeyen parametreler vektörü ve $\boldsymbol{\varepsilon}$, koşullu ortalaması $E(\boldsymbol{\varepsilon}|\boldsymbol{X}) = 0$ ve varyansı σ^2 olan bilinmeyen bir sabit, $\boldsymbol{I}_n$ n boyutlu birim matris olmak üzere , $\mathrm{Var}(\boldsymbol{\varepsilon}|\boldsymbol{X}) = \sigma^2 \boldsymbol{I}_n$ olan $n \times 1$ boyutlu, bağımsız, özdeş dağılımlı rasgele hatalar vektörüdür. Bilinmeyen parametre vektörü $\boldsymbol{\beta}$ aşağıda ifade edilecek olan doğrusal en küçük kareler yöntemi ile tahmin edilir.

2.2. En Küçük Kareler Yöntemi (EKK)

(2.1) denklemi ile ifade edilen doğrusal regresyon modelindeki $\boldsymbol{\beta}$ parametresinin enküçük kareler tahmin edicisi kalan kareler toplamı

$$\begin{aligned} S(\boldsymbol{\beta}) &= \sum_{i=1}^{n} \left(y_i - \boldsymbol{x}_i^T \boldsymbol{\beta} \right)^2 \\ &= \left(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta} \right)^T \left(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta} \right) \\ &= \left\| \boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta} \right\|^2 \end{aligned} \tag{2.2}$$

probleminin minimum değerinden elde edilir. Burada $\boldsymbol{x}_i = \left(x_{i1}, x_{i2}, ..., x_{ip}\right)^T$, i. açıklayıcı değişken vektörüdür. $S(\boldsymbol{\beta})$ gerçel değerli, konveks ve diferansiyellenebilir bir fonksiyon olduğundan her zaman bir minimum değere sahiptir. Eğer $S(\boldsymbol{\beta})$

$$S(\boldsymbol{\beta}) = \boldsymbol{y}^T\boldsymbol{y} + \boldsymbol{\beta}^T\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{\beta} - 2\boldsymbol{\beta}^T\boldsymbol{X}^T\boldsymbol{y} \tag{2.3}$$

şeklinde yazılarak $\boldsymbol{\beta}$ parametresine göre türev alınırsa

$$\frac{\partial S(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = 2\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{\beta} - 2\boldsymbol{X}^T\boldsymbol{y}, \tag{2.4}$$

$$\frac{\partial^2 S(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^2} = 2\boldsymbol{X}^T\boldsymbol{X} \tag{2.5}$$

denklemleri elde edilir. Burada, $\boldsymbol{X}^T\boldsymbol{X}$ negatif olmayan tanımlı bir matristir. (2.4) denklemi doğrusal regresyon modeli için normal denklem olarak adlandırılır ve bu denklemin 0 a eşitlenmesi ile

$$\boldsymbol{X}^T\boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{X}^T\boldsymbol{y} \tag{2.6}$$

elde edilir. Eğer $\boldsymbol{X}$ tam ranklı ise, $\boldsymbol{X}^T\boldsymbol{X}$ tekil olmayan bir matristir ve $S(\boldsymbol{\beta})$ nın $\hat{\boldsymbol{\beta}}$ çözümü ve uydurulmuş model

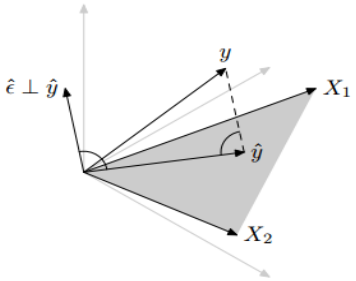
$$\begin{aligned} \hat{\boldsymbol{\beta}} &= \left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\boldsymbol{y} \\ \hat{\boldsymbol{y}} &= \boldsymbol{X}\hat{\boldsymbol{\beta}} \end{aligned} \tag{2.7}$$

olarak elde edilir ve bu çözüm tektir.

$p \times n$ boyutlu $\boldsymbol{X}$ matrisini sütun uzayı olarak şöyle tanımlayalım:

$$R(\boldsymbol{X}) = \left\{\theta : \theta = \boldsymbol{X}\boldsymbol{\beta}, \boldsymbol{\beta} \in \mathbb{R}^n\right\}. \tag{2.8}$$

Burada $R(\boldsymbol{X})$, $\mathbb{R}^n$ in altuzayıdır. Eğer $\boldsymbol{x} \in \mathbb{R}^n$ için $\|\boldsymbol{x}\| = \left(\boldsymbol{x}^T\boldsymbol{x}\right)^{1/2}$ olduğu göz önüne alınırsa, en küçük kareler ilkesi, $\theta \in R(\boldsymbol{X})$ için $\|y - \theta\|$ nin minimize edilmesi ile aynıdır. EKK yönteminin geometrik özellikleri, p=3 ve n=2 özel durumu için $\theta \in R(\boldsymbol{X})$ olmak üzere aşağıdaki şekilde verilmiştir (Rao ve Toutenburg 1999:25).



Şekil 2.1. Enküçük kareler yönteminin

grafiksel olarak gösterimi (Rao ve Toutenburg 1999)

$n \geq p$ olmak üzere σ^2 nin yansız tahmin edicisi $\hat{\sigma}^2$,

$$\hat{\sigma}^2 = \boldsymbol{y}^T \left(\boldsymbol{I} - \boldsymbol{H} \right) \boldsymbol{y} / \left(n - p \right) \tag{2.9}$$

burada $\boldsymbol{H} = \boldsymbol{X} \left(\boldsymbol{X}^T \boldsymbol{X} \right)^{-1} \boldsymbol{X}^T$ şapka operatörüdür. (Rencher ve Schaalje 2000:230)

2.3. Ridge Regresyon

(2.2) $S(\boldsymbol{\beta})$ amaç fonksiyonu lineer denklem sistemidir. Bu lineer denklem sistemi düzensiz ve tutarsız, yani kötü koşullu olabilir. Bu durumda enküçük kareler çözümü büyük bir norma sahip olur ki bu da çözümü anlamsız veya kararsız yapar. Sistemi iyi koşullu duruma getirmek ve sistemin kararlı bir çözümünü elde etmek için düzenleme tekniklerinden birini kullanmak gerekir. Bu tekniklerin en önemlilerinden birisi Ridge regresyon olarak bilinen Tikhonov düzenlemesidir. Bu düzenlemede (2.2) amaç fonksiyonu $S(\boldsymbol{\beta})$ ya bir karesel ceza terimi eklenir. Bu durumda $\boldsymbol{\beta}$ nın ridge tahmini $\hat{\boldsymbol{\beta}}^{ridge}$ tutarlılık ilkesi altında,

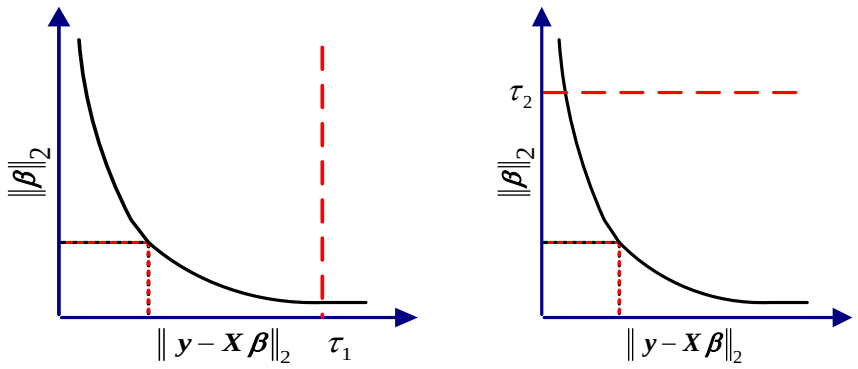
$$\begin{aligned} &\min \ \left\| \boldsymbol{\beta} \right\|_2 \\ &k.s \left\| \boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta} \right\|_2 \leq \tau_1, \end{aligned} \tag{2.10}$$

veya

$$\begin{aligned} &\min \ \left\| \boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta} \right\|_2 \\ &k.s \ \left\| \boldsymbol{\beta} \right\|_2 \leq \tau_2 \end{aligned} \tag{2.11}$$

minimizasyon problemlerinden birinin çözümünden elde edilir.

Burada, τ_1 gözlemlerin hata düzeyi için bir üst sınırdır. (2.10) minimizasyon probleminde τ_1 artarken uygun modellerin kümesi genişler ve $\|\boldsymbol{\beta}\|_2$ nin minimum değeri azalır. Böylece τ_1 e karşılık $\|\boldsymbol{\beta}\|_2$ minimum değerlerinin eğrisi çizilir. Bu durum şekil 2.2. de sol taraftaki şekilde gösterilmektedir.



Şekil 2.2. Tikhonov düzenlemesi problemlerinin grafiksel gösterimi (Aster ve ark. 2005)

(2.11) minimizasyon problemi için de aynı karşılaştırma yapılabilir. Bu problemde, τ_2 azalırken, uygun çözümlerin kümesi küçülür ve $\|\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}\|_2$ nin minimum değeri artar. Tekrar τ_2 ayarlanır, $\|\boldsymbol{\beta}\|_2$ ve $\|\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}\|_2$ nin optimal değerlerinin eğrisi çizilir (Aster ve ark. 2005: 90).

Tikhonov düzenlemesi problemi ile veriyi uydurmak için gerekli model özellikleri göz önüne alınırken gereksiz olan model özellikleri düzenleme ile ortadan kaldırılır. Tikhonov düzenlemesinin üçüncü formu aşağıda verilen etkisi azaltılmış en küçük kareler problemidir:

$$\min \ \|\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}\|_2^2+\lambda\|\boldsymbol{\beta}\|_2^2. \qquad (2.12)$$

Bu problem, (2.11) problemine lagrange çarpanları yönteminin uygulanması ile elde edilir (Aster ve ark. 2005: 50). Burada $\lambda\,(\lambda>0)$ küçültme ve ceza parametresi katsayıların büyüklüğünü kontrol ederek ölçümlere uygunluğun yani küçük hata kareleri ve modelin karmaşıklığı arasında bir denge sağlar (Beck ve Teboulle 2009). $\lambda=0$ olması durumunda $\boldsymbol{\beta}$ nın ridge tahmini $\hat{\boldsymbol{\beta}}^{ridge}$ en küçük kareler tahmini $\hat{\boldsymbol{\beta}}$ yı

verir, $\lambda \to \infty$ durumunda ise $\hat{\boldsymbol{\beta}}^{ridge} = 0$ olur. Görülebileceği gibi (2.10) ve (2.11) çözümleri, kısıtları sağlayacak λ parametresinin ayarlanarak (2.12) nin kullanılması ile elde edilebilir. (2.12) probleminin çözümü katsayı matrisi $\boldsymbol{X}$ in tekil değer ayrışımının kullanılması ile

$$\hat{\beta}^{ridge} = \left(\boldsymbol{X}^T\boldsymbol{X} + \lambda \boldsymbol{I}\right)^{-1}\boldsymbol{X}^T\boldsymbol{y} \tag{2.13}$$

olarak elde edilir. Burada τ_1, τ_2 ve λ uygun olarak seçildiğinde (2.10), (2.11) ve (2.12) problemleri aynı çözümleri verir. En iyi çözümü verecek λnın tahmin edicisi $\hat{\lambda}$ genelleştirilmiş çapraz doğrulama (GCV) (Golub ve ark. 1979) yöntemi ile bulunur. Bu yöntemde $A(\lambda) = \boldsymbol{X}\left(\boldsymbol{X}^T\boldsymbol{X} + n\lambda\boldsymbol{I}\right)^{-1}\boldsymbol{X}^T$ olmak üzere $\hat{\lambda}$,

$$V(\lambda) = \tfrac{1}{n}\left\|\left(\boldsymbol{I} - \boldsymbol{A}(\lambda)\right)\boldsymbol{y}\right\|^2 \Big/ \left[\tfrac{1}{n}\dot{I}z\left(\boldsymbol{I} - \boldsymbol{A}(\lambda)\right)\right]^2 \tag{2.14}$$

ifadesini minimum yapan değer olarak alınır. $\boldsymbol{X\beta}$ nın tahmini için hata karelerinin ortalaması $T(\lambda)$,

$$T(\lambda) = \tfrac{1}{n}\left\|\boldsymbol{X\beta} - \boldsymbol{X}\hat{\boldsymbol{\beta}}(\lambda)\right\|^2 \tag{2.15}$$

şeklindedir. $\boldsymbol{g} = \boldsymbol{X\beta}$ olmak üzere $T(\lambda)$nın ortalaması

$$ET(\lambda) = \tfrac{1}{n}\left\|\left(\boldsymbol{I} - \boldsymbol{A}(\lambda)\right)\boldsymbol{g}\right\|^2 + \tfrac{\sigma^2}{n}\,\text{İz}\boldsymbol{A}^2(\lambda) \tag{2.16}$$

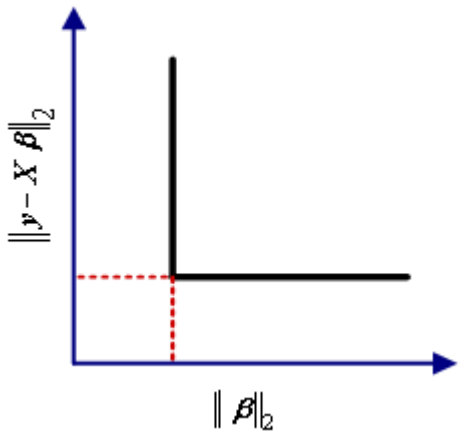
olarak elde edilir. Burada *n* nin büyük değerleri için, λnın GCV tahmini aynı zamanda, σ^2 nin tahminine ihtiyaç olmaksızın $ET(\lambda)$ yi minimum yapan değerdir. Bunun sonucu olarak GCV *n-p* farkının küçük olduğu veya modelin,

$$y_i = \sum_{j=1}^{p} x_{ij}\beta_j + \varepsilon_j,\ i = 1,2,...,n \tag{2.17}$$

olarak yazılabildiği bazı durumlarda kullanılabilir (Golub ve ark. 1979).

Bununla birlikte, log-log ölçek grafiği çizildiğinde $\|\boldsymbol{\beta}\|_2$ ve $\|\boldsymbol{y} - \boldsymbol{X\beta}\|_2$ nin uygun değerlerinin eğrisi *L şeklindedir ve bu* eğri *L*-eğrisi olarak adlandırılır. *L*- eğrisi şekil 2.3. te verilmiştir. Bu şekilde, $\|\boldsymbol{\beta}\|_2$ ve $\|\boldsymbol{y} - \boldsymbol{X\beta}\|_2$ nın en uygun değeri eğrinin köşe

noktasında yer alır. Bir yeterlilik sınırı olan *L*-eğrisi üst sınır rolündeki bir parametrenin kapalı bir fonksiyonu olarak göz önüne alınabilir ve bunun gibi birkaç parametrenin olması durumunda yeterli bir yüzey elde edilir.



Şekil 2.3. λ nın seçimi için L-eğrisi

Bu tip kapalı fonksiyonlar 'transversality' olarak adlandırılan kendilerine özgü koşullar altında genel olarak mevcuttur (Jongen ve ark. 1986). Tutarsızlık ilkesine ek olarak λ nin değerini belirlemek için kullanılan diğer kriter *L*-eğrisi kriteridir. Bu kriterde çözümü *L*-eğrisinin köşesinde veya en yakınında veren λ değeri seçilir (Aster ve ark. 2005: 91)

2.4. En Küçük Mutlak Küçültme ve Operatör Seçimi (LASSO)

Bu yöntem de Tikhonov düzenlemesi gibi bir ceza terimine bağlı olarak uygulanır. Tikhonov düzenlemesinde ceza terimi L_2 normuna, LASSO'da ise L_1 normuna bağlı olarak yazılır. Yani, Tikhonov düzenlemesinde küçültme $\|\boldsymbol{\beta}\|_2^2$ ile yapılırken LASSO'da $\|\boldsymbol{\beta}\|_1^1 := \sum_{j=0}^{p} |\beta_j|$ olmak üzere $\|\boldsymbol{\beta}\|_1^1$ ile yapılır. LASSO yönteminin ridge regresyona göre üstün tarafları vardır. Ridge regresyonu β_j katsayılarını küçülterek tutarlı çözümler sağlayan sürekli bir yöntemdir. Bununla birlikte bu yöntemde β_j katsayıları tam olarak sıfıra eşitlenemediğinden kolay yorumlanabilir modeller elde edilemez. LASSO yönteminde ise bazı β_j katsayıları küçülürken bazıları ise sıfıra olarak elde edilir. Bunun sonucu olarak da daha kolay yorunlanabilir modellerin elde edilmesine olanak tanır. Ancak bu iki yöntem çözüm açısından karşılaştırılacak olursa,

ridge regresyon probleminin LASSO'ya göre daha kolay elde edilecek bir çözüme sahiptir.

$\boldsymbol{\beta}$ parametresinin LASSO tahmini $\hat{\boldsymbol{\beta}}^{lasso}$, $\|\boldsymbol{\beta}\|_1^1 := \sum_{j=1}^{p}|\beta_j|$ olmak üzere

$$\begin{aligned}\hat{\boldsymbol{\beta}}^{lasso} &= \min_{\beta}\left\{\frac{1}{2}\sum_{i=1}^{N}\left(y_i-\sum_{j=1}^{p}x_{ij}\beta_j\right)^2+\lambda\sum_{j=1}^{p}|\beta_j|\right\}\\ &= \min_{\beta}\tfrac{1}{2}\|\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}\|_2^2+\lambda\|\boldsymbol{\beta}\|_1^1\end{aligned} \tag{2.18}$$

probleminin çözümünden elde edilir. Burada λ ceza parametresidir ve ridge regresyonda olduğu gibi $\lambda\to\infty$ yaklaşırken (2.18) nin optimal çözümü 0' a yaklaşır. Ancak (2.18) probleminin çözümü tek değildir. (2.18) LASSO problemi homotopi (Garrigues ve Ghaoui 2008) ve conveks en iyileme (Osborne ve ark. 2000) yöntemlerinin kullanılması ile çözülebilir. Homotopi yönteminde çözüm dizisel olarak elde edilen $(y_i,\boldsymbol{x}_i)$ değerlerine bağlı olan bir algoritma yardımı ile elde edilir. Bu algoritma $\boldsymbol{\beta}^{(n)}$, *n*-tane gözlem değerinin kullanılması ile elde edilen LASSO çözümü ve yeni gözlem değeri $(y_{n+1},\boldsymbol{x}_{n+1})\in\mathbb{R}\times\mathbb{R}^m$ nin edilmesinde sonra bulunan $\boldsymbol{\beta}^{(n+1)}$ çözümlerine bağlı olarak ve $\boldsymbol{\beta}^{(n)}$ den $\boldsymbol{\beta}^{(n+1)}$ e bir homotopikliği hesaplayan en iyileme problemine bağlı olarak kullanılır. Bu algoritmada bir önceki çözüm $\boldsymbol{\beta}^{(n)}\neq 0$ ve $\boldsymbol{\beta}^{(n+1)}\neq 0$ olduğu noktalar biribirine yakın olduğunda metodu anlamlı ve etkili yapan iyi bir başlangıç noktası olarak göz önüne alınır. Bu algoritma aşağıdaki en iyileme problemine bağlı olarak oluşturulur:

$$\boldsymbol{\beta}(t,\lambda)=\min_{\beta}\frac{1}{2}\left\|\begin{pmatrix}\boldsymbol{X}\\ t\boldsymbol{x}_{n+1}^T\end{pmatrix}\boldsymbol{\beta}-\begin{pmatrix}\boldsymbol{y}\\ ty_{n+1}\end{pmatrix}\right\|_2^2+\lambda\|\boldsymbol{\beta}\|_1 \tag{2.19}$$

burada $\boldsymbol{\beta}^{(n)}=\boldsymbol{\beta}(0,\lambda_n)$ ve $\boldsymbol{\beta}^{(n+1)}=\boldsymbol{\beta}(1,\lambda_{n+1})$ dir. Bu algoritmada $\beta^{(n)}$ den $\beta^{(n+1)}$ e doğru giden bir yol iki adımda hesaplanır:

Adım 1: $t=0$ alarak düzenleme (regularization) parametresi $\boldsymbol{\beta}^{(n)}$ den $\boldsymbol{\beta}^{(n+1)}$ e değiştirilir. Bu değişim $\boldsymbol{\beta}^{(n)}$ ile $\boldsymbol{\beta}^{(n+1)}$ arasındaki düzenleme yolunun hesaplanması

anlamına gelir. Çözüm yolu parçalı doğrusaldır (Efron ve ark 2004, Malioutov ve ark. 2005, Osborne 1992).

Adım 2: $\lambda = \lambda_{n+1}$ alınarak t parametresi 0'dan 1'e değiştirilir. Buradaki düzenleme parametresi λ nın en iyi değeri, bir gözlem değerinin gözlem grubunun dışında bırakıldığı çapraz doğrulama yöntemi ile elde edilir (Golub ve ark. 1979: 217).

Yukarıda belirtildiği gibi LASSO problemi, konveks en iyileme problemleri, özellikle de ikinci dereceden konik programların (SOCP) kullanılması ile çözülebilir (Lobo ve ark. 1998). Bu problemler,

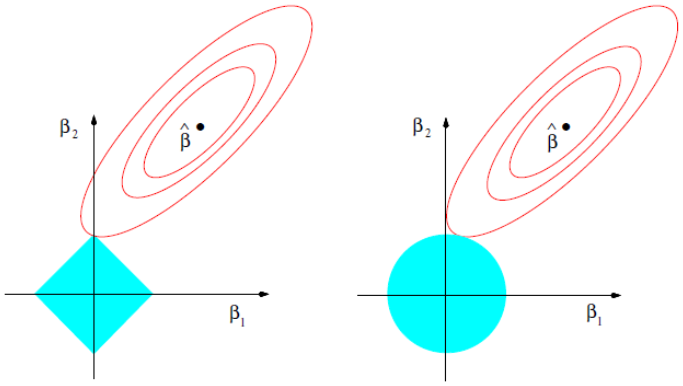
$$\begin{aligned} &\min_{\boldsymbol{\beta}} \|\boldsymbol{\beta}\|_1 \\ &k.s.\ \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|_2 \le \tau_1 \end{aligned} \tag{2.20}$$

veya

$$\begin{aligned} &\min_{\boldsymbol{\beta}} \tfrac{1}{2}\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|_2^2 \\ &k.s.\ \|\boldsymbol{\beta}\|_1 \le \tau_2 \end{aligned} \tag{2.21}$$

şeklindedir. Bu problemler konveks programların çözümüne en iyi şekilde yanıt veren iç noktalar yöntemi ile çözülebilir (Nesterov ve Nemirovski 1994:67-80).

$\hat{\boldsymbol{\beta}}$ değerini elde etmek için kullanılan LASSO ve Ridge regresyon problemlerine ilişkin hata fonksiyonunun eşyükselti eğrileri kısıt fonksiyonlarını oluşturduğu bölge aşağıdaki şekil 2.4. ile gösterilmiştir (Tibshirani 1996).



Şekil 2.4. LASSO problemine (sol) ve Ridge problemine (sağ) ilişkin tahmin şekilleri (Tibshirani 1996)

Şekil 2.4.te görüldüğü gibi $p = 2$ özel durumu için LASSO problemine ilişkin kısıt bölge $|\beta_1| + |\beta_2| \leq t$ bir eşkenar dörtgen ridge problemine ilişkin kısıt bölge $\beta_1^2 + \beta_2^2 \leq t$ ise $\sqrt{t}$ yarıçaplı bir daire şekildedir. Ayrıca bu şekilde LASSO ve ridge yöntemlerine ilişkin elipsoid olan hata fonksiyonlarının eşyükselti eğrileri verilmiştir.

3. KONİK KARESEL PROGRAMLAMA

n-boyutlu ikinci dereceden veya Lorentz tipi koni (dondurma-ice-cream),

$$\boldsymbol{L}^n = \left\{ \boldsymbol{x} = \left(x_1, x_2, ..., x_n\right)^T \in \mathbb{R}^n \mid x_n \geq \sqrt{x_1^2 + ...x_{n-1}^2} \right\} \quad (n \geq 2).$$

olarak tanımlanır. Bir konik karesel problem

$$\begin{aligned} &\min_x \boldsymbol{c}^T \boldsymbol{x} \\ &k.s \quad \boldsymbol{A}\boldsymbol{x} - \boldsymbol{b} \in \boldsymbol{K} \end{aligned} \tag{3.1}$$

şeklinde bir en iyilime problemidir. Burada $\boldsymbol{K}$ yukarıda tanımlanan ikinci dereceden konilerin aşağıdaki şekilde direkt çarpımıdır:

$$\begin{aligned} \mathbf{K} &= \mathbf{L}^{n_1} \times \mathbf{L}^{n_2} ... \times \mathbf{L}^{n_k} \\ &= \left\{ \begin{pmatrix} \boldsymbol{y}[1] \\ \boldsymbol{y}[2] \\ ... \\ \boldsymbol{y}[k] \end{pmatrix} \mid \boldsymbol{y}[i] \in \mathbf{L}^{n_i} \ \ (i = 1, 2, ..., k) \right\}. \end{aligned} \tag{3.2}$$

(3.2) den görülebileceği gibi bir konik karesel program doğrusal amaç fonksiyonlu ve kısıt fonksiyonları aşağıdaki gibi sonlu sayıda ikinci dereceden koniler olan bir en iyileme problemidir:

$$\boldsymbol{A}_i \boldsymbol{x} - \boldsymbol{b}_i \in \mathbf{L}^{n_i} \qquad (i = 1, 2, ..., k).$$

Burada

$$[\boldsymbol{A}, \boldsymbol{b}] = \begin{bmatrix} [\boldsymbol{A}_1 ; \boldsymbol{b}_1] \\ [\boldsymbol{A}_2 ; \boldsymbol{b}_2] \\ ... \\ [\boldsymbol{A}_k ; \boldsymbol{b}_k] \end{bmatrix}$$

(3.2) deki *y* nin parçalanmasına karşılık gelen $[\boldsymbol{A}, \boldsymbol{b}]$ veri matrisinin parçalanmasıdır. Böylece konik karesel program aşağıdaki şekilde yazılabilir:

$$\begin{aligned} &\min_{x} \ \boldsymbol{c}^T \boldsymbol{x}, \\ &\boldsymbol{A}_i \boldsymbol{x} - \boldsymbol{b}_i \in \mathbf{L}^{n_i} \ \ (i = 1, 2, ..., k) \end{aligned} . \tag{3.3}$$

Bazı durumlarda, $\boldsymbol{A}_i \boldsymbol{x} - \boldsymbol{b}_i \in \mathbf{L}^{n_i}$ bağıntısı kısmi sıralama anlamına gelen vektör eşitsizliği formunda da yazılabilir. Bu formlar $\boldsymbol{A}_i \boldsymbol{x} - \boldsymbol{b}_i \succeq_{\mathbf{L}^{n_i}} 0$ veya $Ax_i \succeq_{\mathbf{L}^{n_i}} \boldsymbol{b}_i$ şeklindedir. Daha genel olarak, bu tarz bir gösterim ve kısmi sıralama herhangi bir sonlu boyutlu **E** Öklid uzayında kullanılabilir. Burada iyi vektör eşitsizliği ‘‘$\succeq$’’ tam olarak negatif olmayan vektörlerin bir $\mathbf{K} = \left\{ a \in \mathbf{E} \middle| \ a \succeq 0 \right\}$ kümesi olarak tanımlanır. Burada $a \succeq b \Leftrightarrow a - b \succeq 0$ $(\Leftrightarrow a - b \in \mathbf{K})$ şeklindedir. Bununla birlikte **K** kümesi isteğe bağlı yazılamadığı gibi sivri konveks koni olmak zorundadır. **E** deki her sivri konveks koni **E** üzerinde ‘‘$\geq_K$’’ ile verilen $a \geq_{\mathrm{K}} b \Leftrightarrow a - b \geq_{\mathrm{K}} 0 \Leftrightarrow a - b \in \mathbf{K}$ şeklindeki kısmi sıralamayı ifade eder (Taylan ve ark. 2007). $\left[\boldsymbol{A}_i ; \boldsymbol{b}_i \right]$ veri matrisi

$$\left[\boldsymbol{A}_i ; \boldsymbol{b}_i \right] = \begin{bmatrix} \boldsymbol{D}_i & \boldsymbol{d}_i \\ \boldsymbol{p}_i^T & q_i \end{bmatrix}$$

şeklinde parçalanırsa, (3.3) problemi

$$\begin{aligned} &\min_{x} \ \boldsymbol{c}^T \boldsymbol{x}, \\ &\left\| \boldsymbol{D}_i \boldsymbol{x} - \boldsymbol{d}_i \right\|_2 \leq \boldsymbol{p}_i^T \boldsymbol{x} - q_i \ \ (i = 1, 2, ..., k) \end{aligned} \tag{3.4}$$

olarak yazılabilir. Burada $\boldsymbol{D}_i$, $(n_i - 1) \times (\dim x)$ tipli matris, $\boldsymbol{d}_i$ sütun vektörlerinin boyutu $\boldsymbol{D}_i$ matrislerinin sütun vektörleri ile, p_i ler $\boldsymbol{x}$ vektörleri ile aynı boyutlu sütun vektörleridir, q_i ler ise gerçel değerlerdir ve $\left\| \cdot \right\|_2$ öklit normunu göstermektedir. Bu form, konik problemin en açık formudur. (3.2) eşitliği gerçekten bir koniyi göstermektedir ve $\boldsymbol{K}$ duali $\boldsymbol{K}_*$ olmak üzere $\boldsymbol{K}_* = \boldsymbol{K}$ dir (Taylan ve ark. 2007). (3.1) en iyileme probleminin dual problemi

$$\boldsymbol{A}^T \boldsymbol{\lambda} = \boldsymbol{c}, \ \boldsymbol{\lambda} \in \boldsymbol{K} \ \text{iken} \ \max_{\lambda} \ \boldsymbol{b}^T \boldsymbol{\lambda} \ \text{dir.} \ \left(i = 1, 2, ..., k \right) \tag{3.5}$$

şeklindedir. Eğer $\boldsymbol{\lambda}$ vektörü m_i boyutlu $\boldsymbol{\lambda}_i$ blokları ile $\boldsymbol{\lambda} := (\boldsymbol{\lambda}_1^{\mathrm{T}}, \boldsymbol{\lambda}_2^{\mathrm{T}}, ..., \boldsymbol{\lambda}_{\mathrm{K}}^{\mathrm{T}})^{\mathrm{T}}$ şeklinde parçalanırsa, dual problemi aşağıdaki gibi ifade edilir (Lobo ve ark. 1998):

$$\begin{aligned} &\underset{\lambda_1,\ldots,\lambda_n}{\text{maks}} \sum_{i=1}^{k} \boldsymbol{b}_i^T \boldsymbol{\lambda}_i \\ &\text{k.s.} \quad \sum_{i=1}^{k} \boldsymbol{A}_i^T \boldsymbol{\lambda}_i = \boldsymbol{c}, \boldsymbol{\lambda}_i \in \boldsymbol{L}^{n_i} \end{aligned} \tag{3.6}$$

Eğer ν_i skaler bileşenli $\boldsymbol{\lambda}_i = (\boldsymbol{\kappa}_i^T, \nu_i)^T$ vektörü ve '' $\geq_{\mathbf{L}^n} 0$ ''ifadesinin anlamı göz önüne alınırsa (3,4) dual problemi aşağıdaki formda yazılır:

$$\begin{aligned} &\underset{(\mu_i),(\nu_i)}{\text{maks}} \sum_{i=1}^{k} \left[\boldsymbol{\kappa}_i^T \boldsymbol{d}_i + \nu_i q_i \right], \ (i = 1, 2, \ldots, k) \\ &\text{k.s.} \quad \sum_{i=1}^{k} \left[D_i^T \boldsymbol{\kappa}_i + \nu_i p_i \right] = \boldsymbol{c}, \ \|\boldsymbol{\kappa}_i\|_2 \leq \nu_i \end{aligned} \tag{3.7}$$

(3.7) deki tasarım değerleri $\boldsymbol{\kappa}_i$ sütun vektörleridir, $\boldsymbol{d}_i$ vektörleri ile aynı boyutludur ve ν_i (i = 1,2,..., k) dir. (3.4) ve (3.7) ile verilen en iyileme problemleri bir konik karesel problemin ve onun dualinin standart formlarıdır (Ben-Tal ve Nemirovski 2001, Alizadeh ve Goldfarb 2003).

Bazen uygulamalarda ortaya çıkan en iyileme problemleri kendi standart formlarında olmayabilir. Esas problemin her zaman bir standart en iyileme problemi tarafından tanımlanması oldukça önemlidir (Taylan ve ark. 2007). Genel olarak, en iyileme problemleri aşağıdaki formda verilir

$$\min_{x} \ f(\boldsymbol{x}), \boldsymbol{x} \in X \, . \tag{3.8}$$

Burada, f bir "kayıp fonksiyonu" (loss function) ve X kümesi tasarım vektörlerinden oluşur ve genel olarak aşağıdaki şekilde verilir:

$$X = \bigcap_{i=1}^{n} X_i \, . \tag{3.9}$$

Burada $g_j(\boldsymbol{x})$, j. kısıt fonksiyonu olmak üzere, her X_i birçok durumda aşağıda belirtilen küme tarafından temsil edilen belirli bir tasarım kısıtlaması için kabul edilebilir vektörlerin bir kümesidir.

$$X_i \ = \ \left\{ \boldsymbol{x} \in \mathbb{R}^n \mid \ g_j(\boldsymbol{x}) \leq 0 \right\} \tag{3.10}$$

(3.8) deki amaç fonksiyonu f her zaman doğrusal kabul edilir, aksi durumda amaç fonksiyonu kısıtların listesine taşınabilir ve eşdeğer problem, $\hat{X} := \{(\boldsymbol{x},t) \mid \boldsymbol{x} \in X,\ t \geq f(\boldsymbol{x})\}$ olmak üzere

$$\min_{t,x} \quad t,$$
$$\text{k.s. } (t,\ \boldsymbol{x}) \in \hat{X}$$

olarak yazılabilir. Bu form oldukça faydalıdır, örneğin, $f(x)$ Öklid normu olarak verilebilir (karesel olmayan). $f(x)$ in kareler toplamı olması durumunda, yani, karesel Öklid normu olması durumunda, $t^2 \geq f(x),\ t \geq 0$ şeklindeki form tercih edilir ki bu da Lorentz tipi koninin tanımına uygundur. Bu form problemin yeni bir formda yazılmasını sağlar ve esas probleme eşdeğerdir. Böylece, orijinal problemin aşağıdaki şekilde görüleceğini varsayabiliriz:

$$\boldsymbol{x} \in X := \bigcap_{i=1}^{n} X_i \text{ olmak üzere}$$

$$\min_{x} \ \boldsymbol{c}^T \boldsymbol{x}.$$

X in standart formda olduğunu belirlemek amacıyla, aynı yapının farklı formlarını içeren bir tür sözlüğe gereksinim vardır. Böyle bir sözlük konik karesel programlar için de üretilmiştir. Böylece, bir X kümesi verildiğinde $\|\boldsymbol{D}\boldsymbol{x} - \boldsymbol{d}\|_2 \leq \boldsymbol{p}^T \boldsymbol{x} - q$ gibi konik karesel eşitsizlikleri olarak yazılıp yazılmayacağı belirlenebilir. Kısaca, X , (3.11) probleminin çözüm kümesinin *x*-uzayı üzerine izdüşümü olacak şekilde $\boldsymbol{x} \in \mathbb{R}^n$ ve ekli *u* değişkenlerini içeren

$$\boldsymbol{A}_j \begin{pmatrix} \boldsymbol{x} \\ \boldsymbol{u} \end{pmatrix} - \boldsymbol{b}_j \succeq_{L_{m_j}} 0 \tag{3.11}$$

şeklindeki çok sayıda sonlu vektör eşitsizliklerinin bir sistemi mevcutsa problem konik karesel program olarak yazılabilir. Bu da şu anlama gelir: $x \in X$ olması için gerek ve yeter koşulun *x* in

$$\boldsymbol{A}_j \begin{pmatrix} \boldsymbol{x} \\ \boldsymbol{u} \end{pmatrix} - \boldsymbol{b}_j \succeq_{L_{m_j}} 0$$

sisteminin $(x;u)$ çözümüne genişlemesi anlamına gelir, yani,

$$x \in X \quad \Leftrightarrow \qquad \exists u : \boldsymbol{A}_j \begin{pmatrix} \boldsymbol{x} \\ u \end{pmatrix} - \boldsymbol{b}_j \succeq_{L_{m_j}} 0 \quad (j = 1, 2, ..., N) .$$

(3.11) olarak ifade edilen her *X kümesi* konik karesel gösterim (*CQG*) olarak adlandırılır (Taylan ve ark. 2007).

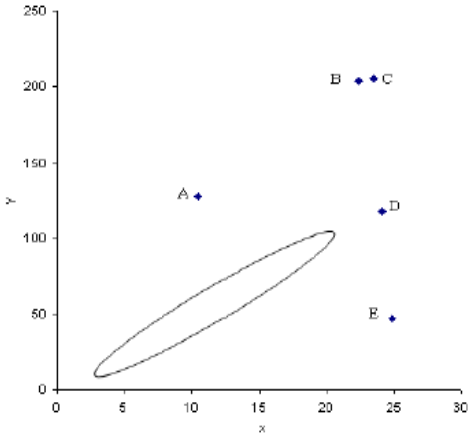
4. SAPAN DEĞERLER

4.1. Sapan Değerin Tanımı

İstatistikte tutarlı bir analiz elde etmeye yönelik ilk adım sapan gözlem değerlerini belirlemektir. Sapan değerler genel olarak hata veya gürültü olarak kabul edilse de önemli bilgi taşıyor olabilirler.

Sapan değerler, model belirlemeye öncülük edebilecek yanlı parametre tahminlerine ve doğru olmayan sonuçlara yol açan verilerdir. Bu nedenle istatistiksel modelleme ve analizden önce bunların teşhis edilmesi gerekir (Williams ve ark. 2002, Liu ve ark. 2004).

Hawkins (1980:1) sapan değeri "diğer gözlemlerden oldukça fazla sapma gösteren ve farklı bir mekanizma tarafından üretildiği konusunda şüphe uyandıran bir gözlem" olarak, Barnett ve Lewis (1994:584) "içinde bulunduğu örneklemin diğer üyelerinden önemli derecede ayrılmış görünen bir gözlem olarak veya bir veri kümesinde gözlemlerin çoğunun sahip olduğu dağılıma veya modele uymayan gözlemler" olarak ifade etmiştir. Benzer şekilde Johnson ve Wichern (1992:187) "bir veri kümesinde bulunan ve veri kümesinin geri kalanıyla tutarsızlık gösteren bir gözlem" şeklinde tanımlar. Yukarıdaki tanımlardan da anlaşılacağı gibi sapan değerler verilerin çoğunluğu tarafından oluşturulan doğrusal kalıba uymayan gözlemlerdir. Uygulamada hem yanıt değişken hem de açıklayıcı değişkende veya her ikisinde aynı anda bir sapan değere rastlamak mümkündür (Rousseuw ve Van Aelst 1999). Bu durum şekil 4.1. de detaylı olarak gösterilmiştir (Adnan ve Mohamad 2003).



Şekil 4.1. Farklı tipte gözlem değerleri için serpme diyagramı (Adnan ve Mohamad 2003)

Yukarıdaki şekilde elips veri kümesinin büyüklüğünü göstermektedir. A, B, C noktaları *Y*-yönünde sapan değerlerdir, çünkü *y* değerleri geri kalan verilerden önemli ölçüde farklıdır ve bunlar aynı zamanda kalan sapan değerlerdir. B, C ve D noktaları *X*-yönünde sapan değerlerdir, yani, kaldıraç noktalarıdır. Bununla birlikte *X*-ekseni yönünde sapma gösteren D noktası kalan sapan değeri değildir. B ve C noktaları kaldıraç noktalarıdır ve kalan sapan değerleridir. A noktası *X*-ekseninde iç bölgededir (*X*-ekseni yönünde sapmayan değerdir) ama bir kalan sapan değeridir. E noktası *Y*-ekseninde iç bölgededir (inlier) ve aynı zamanda kalan sapan değerdir. Bütün bu çıkarsamalar çizelge 4.1. de özetlenmiştir:

Çizelge 4.1. Farklı tipteki sapan değerler

Noktalar	*Y*-yönünde sapan	*X*-yönünde sapan	Kalan sapan
A	*	-	*
B	*	*	*
C	*	*	*
D	*	*	-
E	-	*	*

4.2. Kaldıraç Nokta (Leverage Point)

Yukarıda da belirtildiği gibi regresyonda önemli bir rol oynayan sapan değerlere hem yanıt değişkende hem de açıklayıcı değişkende rastlamak mümkündür. Yanıt değişkenindeki sapan değerler model başarısızlığına neden olur. Böyle gözlemler "*outlier*" yani "*sapan değer"* veya "*aykırı değer"* olarak adlandırılır. Açıklayıcı değişkenlerdeki sapan değerler (*x*- yönünde sapma gösteren de denebilir) "*leverage points*" veya "*kaldıraç noktalar"* olarak adlandırılır. Yanıt değişkenlerde sapan değer olmasa bile kaldıraç noktalar da oluşturulan regresyon modelini etkileyebilir. Genel olarak regresyon teorisinde kaldıraç noktalar iyi ve kötü olmak üzere ikiye ayrılır. *İyi kaldıraç nokta*, *X* açıklayıcı değerleri arasında sıra dışı olarak diğer gözlemlerden daha büyük veya daha küçük değere sahip olan ama regresyon eğrisini etkilemeyen noktadır. Yani, gözlem değerlerinin noktasal gösteriminde, gözlem kitlesinden nispeten uzağa gider ama çoğu gözlem noktasının etrafında toplandığı regresyon doğrusuna oldukça yakındır. İyi bir kaldıraç nokta, çoğunlukla, ilişkilendirilen noktaların ne derece çarpık görüntü verdiği konusunda sınırlı bir etkiye sahiptir. Bu tip noktalar, regresyon katsayılarının hassasiyetini artırır (Rousseeuw ve Zomeren 1990).

Kötü bir kaldıraç nokta ise merkezlenmiş (centered) gözlemlerin kitlesinin civarındaki regresyon çizgisinin uzağında yer alır. Başka bir deyişle, kötü bir kaldıraç nokta, açıklayıcı değişkenler arasında aykırı bir değere sahip olan regresyon sapan değeridir. Kötü kaldıraç nokta, küçük bir kırılma noktalı (breakdown point) tahmin kullanıldığında, regresyon çizgisinin eğiminin tahminini büyük ölçüde etkiler. Kötü kaldıraç noktaları regresyon katsayılarının doğruluğunu azaltır (Rousseeuw ve Zomeren 1990).

Kaldıraç nokta, *şapka* matrisi $\boldsymbol{H} = \boldsymbol{X}^T(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T$ nın köşegen elemenlarının kullanılması ile ölçülür. Bu ölçüm aşağıdaki adımlarda yapılır:

i. Şapka matrisi *Y* yi tahmin edilen değerlere dönüştürür.

ii. Şapka matrisinin köşegenleri (h_{ii}) hangi değerlerin sapan değer olup olmadığını gösterir. Bu nedenle köşegenler kaldıracın ölçüleridir.

iii. Kaldıraç iki limit tarafından sınırlanır, bunlar $1/n$ ve 1 dir. Kaldıraç 1 e ne kadar yakınsa o kadar değer kazanır.

iv. Şapka matrisinin *iz*i modeldeki değişkenlerin sayısını verir.

v. Kaldıraç $2p/n$ den büyük olduğunda yüksek kaldıraç olduğu belirtilir (Belsley 1980:262). Küçük boyutlu örneklemler için Vellman and Welsch (1981) kriter olarak $2p/n$ yerine $3p/n$ oranını önermiştir.

Kaldıraç noktaları genel olarak regresyon doğrusu üzerinde güçlü bir etkiye sahiptir. İyi kaldıraç noktalar olarak yanlış tanımlanabilecek bütün noktaların ihmal edilmesi kötü kaldıraç noktalarının etkilerini engeller. İstatistiksel sonuçların sağlıklı olabilmesi açısından kullanılacak kaldıraç noktalarının "iyi" kaldıraç noktalar olduklarından emin olmak gerekir. İyi kaldıraç noktaları bir nedenle de olsa "iyi" olarak adlandırılır. Bazı durumlarda ihmal edilen kaldıraç noktaları açıklayıcı değişken kümesine eklenir. Çünkü *x*-lerin geniş aralığı üzerinde regresyon fonksiyonunun daha iyi anlaşılmasını sağlayabilirler ve parametre tahminlerindeki standart hataların küçültülmesine yardımcı olabilirler. Bütün iyi kaldıraç noktaları dışarıda tutmak tahmin edilmeye çalışılan regresyon fonksiyonunu ve gözlemlerden elde edilen veri hakkında değerli olabilecek bilgiyi kaybetmek anlamına gelir.

4.3. Maskeleme ve Gölgeleme Etkileri

Sapan değerler parametre tahminini iyileştirmeye yardımcı olarak kullanıldığında, elde edilen tahminler kötü bir şekilde yanlılıkla sonuçlanabilir. Bu da, maskeleme (masking) ve gölgeleme (swamping) denilen istenmeyen iki duruma sebep olur (McCann 2006). *Maskeleme*, bir dizi sapan değerin başka bir dizi sapan değerin varlığı sebebiyle belirlenemediği durumda oluşur, *gölgeleme* ise sapan değer olmayan gözlemlerin bir grup sapan değerin varlığı nedeniyle sapan değermiş gibi yanlış tanımlanmasıdır (Atkinson 1986, Fung 1993, Hadi 1992, Hadi ve Simonoff 1993).

Sapan değerleri belirlemek için kullanılan birçok metot veriyi temiz ve olası sapan değerler kümesi olmak üzere iki kümeye ayırmayı amaçlar (Hadi 1992, Hadi ve Simonoff 1993). Burada temiz küme kavramı için iki yorumlama vardır. Birincisi; temiz noktalar, *y*-yönünde dikey (vertical outliers) ya da *x*-yönünde yatay yani kaldıraç (hem iyi hem kötü) sapan değer içermeyen bütün noktalardır. Diğer yorumlama ise bütün temiz noktaların dikey sapan değerleri dışındaki bütün noktalar olduğu şeklindedir. Her iki yorumun da hem avantaj hem de dezavantajları vardır.

Aykırı maskelemeyi bulmak için en eski yöntemler leave-one-out silme teknikleridir (Belsley ve ark. 1980). Bu yöntemlerde bir veri noktası silinerek ilgili varyans istatistikleri, geriye kalan noktaların kullanılması ile hesaplanır. Bu tek noktanın kullanılmaması tahminlerin bu nokta tarafından etkilenmemesini sağlar. Bu ise, bu noktanın verinin geri kalanı ile aynı eğilimde olup olmadığı konusunda daha objektif bir bilgi verir. Bu tip yöntemler sapan değerlerin sayısı az ve dağınık olması durumunda sağlıklı sonuçlar verir, ancak aksi durumda iyi sonuçlar elde edilmeyebilir. Örneğin, bir çift benzer sapan değerin mevcut olduğu durumda, leave-one-out silme tekniğinin kullanılması ile bu iki sapan değerden sadece biri silinmiş olur. İkinci nokta hala tahminleri etkilemeye devam edecektir (McCann 2006). Bu nedenle belirlenen sapan değerleri veri kümesinden atmak kullanışlı bir yöntem değildir.

5. SAPAN DEĞER TEŞHİS YÖNTEMLERİ

Olası bir sapan değeri teşhis etmek için çeşitli yaklaşımlar vardır. Bunlar $\boldsymbol{\beta}$ parametresinin EKK tahmincisi $\hat{\boldsymbol{\beta}}$ olmak üzere $r_i = y_i - \boldsymbol{x}_i^T \hat{\boldsymbol{\beta}}$ kalanlarını kullanan *doğrudan yaklaşımlar* ve *doğrudan olmayan yaklaşımlar* olmak üzere iki kategoriye ayrılır.

5.1. Doğrudan Yöntemler

Verilen bir probleme uydurulan doğrusal modelin uygunluğunu değerlendirmek için, hatalar ile ilgili varsayımların makul olup olmadığını belirlemek gerekir. ε hataları gözlemlenebilir olmadığından kalanlar kullanılarak dolaylı şekilde incelenir (Cook ve Weisberg 1982:11).

Regresyon eşitliğine dayalı olarak tahmin edilen değer ile gerçek gözlenen değer arasındaki fark, *kalan* olarak adlandırılır. Bunlar aşağıdaki bölümlerde detaylı olarak ele alınacaktır.

5.1.1. Basit Kalan Yöntemi

Basit kalan yönteminde sapan değer teşhisi $r_i = y_i - \boldsymbol{x}_i^T \hat{\boldsymbol{\beta}}$ kalan değerine ve bunun grafiğinin regresyon çizgisinin veriye ne kadar iyi uydurulduğunu ve özellikle örnek eğri için sistematik bir uyumsuzluk olup olmadığını kontrol etme amacıyla kullanılır (Cook ve Weisberg 1982:11). Basit şekliyle kalan $r_i = \left(y_i - \hat{y}_i\right)$ şeklinde hesaplanır. r_i, (Y) yanıt değişkeninin ölçeğini korur. Çözüm kalanın varyansının bir tahmini tarafından standartlaştırılmasıdır. $Var(y_i) = \sigma^2$ varyansı hata kareleri ortalaması (HKO) ile tahmin edilir. Hata kareleri ortalaması bir model veya tahminci tarafından öngörülen değerler ile gerçek gözlem değerleri arasındaki farkın bir ölçüsüdür ve $\sqrt{n^{-1}\sum_{i=1}^{n}\left(r_i\right)^2}$ şeklinde hesaplanır (Willmott ve Matsuura 2005). Fakat sadece y_i den daha fazla anlam içeren $r_i = \left(y_i - \hat{y}_i\right)$ kalan terimi için varyans $Var(r_i) = \sigma^2(r_i) = \sigma^2\left(1 - h_{ii}\right)$ şeklindedir. Burada standartlaştırılmış kalan s_i kullanılmış

olur. h_{ii} ifadesi $\boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T$ matrisinin *i*. köşegen elemanını gösterir (Cook ve Weisberg 1982:12):

$$\boldsymbol{H} = \begin{pmatrix} h_{11} & h_{12} & . & . & . & h_{1n} \\ h_{21} & h_{21} & & & & . \\ . & & . & & & . \\ . & & & . & & . \\ . & & & & . & . \\ h_{n1} & . & . & . & . & h_{nn} \end{pmatrix}$$

Görsel olarak tanımlarsak, bir tek noktayı yukarıya doğru çekip regresyon çizgisinin nasıl bir yol izlediği ölçülür. Gözlenmiş değer y_i ve kestirilmiş değer $\hat{y}_i$ olmak üzere, belirlenmiş bir x_i, y_i değerini bir miktar yerinden oynatır. Eğer $\hat{y}_i$ de y_i kadar hareket ederse y_i nin regresyonu sürdürecek potansiyele sahip olduğu açıktır, dolayısıyla y_i bir kaldıraçtır.

Eğer $\hat{y}_i$ neredeyse hiç hareket etmiyorsa o zaman y_i nin regresyonu sürdürme şansı yoktur. Diğer bir deyişle h_{ii} değeri "kaldıraç" için bir ölçüdür. Sadece *x*-değerlerine bağlıdır ve

$$h_{ii} = \frac{d\hat{y}_i}{dy_i} \tag{5.1}$$

formülüyle hesaplanır (Cook ve Weisberg 1982:14).

5.1.2. Standartlaştırılmış Kalan Yöntemi

Bu yöntem, yanıt vektörü ile uydurulmuş yanıt vektörünün farkı olarak nitelenen basit kalanın standart sapmaya bölümüne bağlı olarak, $\hat{\sigma}^2 = \boldsymbol{r}^T\boldsymbol{r}/n - p$ olmak üzere

$$s_i = \frac{r_i}{\sqrt{\hat{\sigma}^2}} = \frac{r_i}{\hat{\sigma}} \ , \tag{5.2}$$

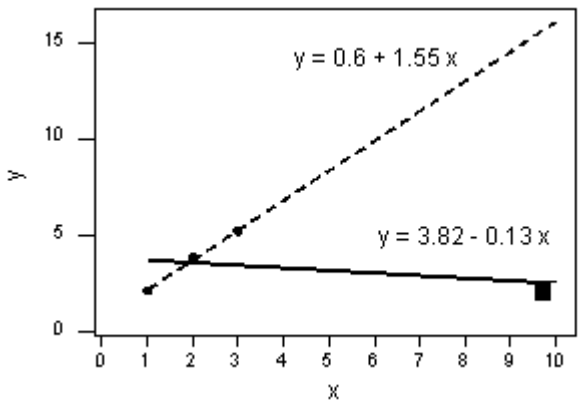
olarak tanımlanır (Chatterjee ve Hadi 1988:73). Başka bir ifade ile hatanın standart sapmaya oranı ile elde edilir:

$$s_i = \frac{\hat{y}_i - y_i}{\hat{\sigma}} . \tag{5.3}$$

Bu yönteme göre *i.* yanıt değeri için $s_i > 3.5$ veya $s_i < -3.5$ ise *i.* yanıt değeri sapan değerdir. Eğer standart kalanlar $|s_i| < 3.5$ aralığında ise sapan değer olarak göz önüne alınmayacaktır.

Standartlaştırılmış kalanlar, kalanların "standart ölçüm" e göre karşılaştırılmalarını sağlar. $s_i = \pm 2$, sıra dışı bir durumun söz konusu olduğunu, $s_i = \pm 3$ olağan üstü bir durumun bir söz konusu olduğunu ve $s_i = \pm 4$ olmaması gereken, dışarıdan bir etkilenmenin söz konusu olduğunu gösterir. Standartlaştırılmış kalanlar (s_i), $y_i - \hat{y}_i$ ile başlar, eğer y_i gerçekten bir sapan değer ise regresyon çizgisini kendisine doğru çeker ve kendi kalan tahminini etkiler ki bu da bir problem oluşturur. Bu durumda, y_i değeri dışarıda (hariç) tutularak, regresyon doğrusu $r_i = \left(y_i - \hat{y}_{i,(-i)}\right)$ kalanına bağlı olarak uydurulur. Burada $\hat{y}_{i,(-i)}$ ifadesi y_i değerini içermeden tahmin edilmiş bir regresyon doğrusunu göstermektedir. Bu durum Örnek 5.1. de ele alınmıştır.

Örnek 5.1. *n=4* veri noktası için oluşturulmuş şekil 5.1. de, düz çizgi bütün veri noktalarının kullanılması ile tahmin edilen regresyon doğrusunu gösterirken, kesikli doğru sadece 3 veri noktasının kullanılması ile tahmin edilen regresyon doğrusunu göstermektedir.



Şekil 5.1. Sapan değerin bulunduğu ve bulunmadığı veri kümesi için tahmin edilmiş regresyon doğruları (Rao ve Toutenburg 1920)

Şekil 5.1. de görüldüğü gibi kare ile gösterilen nokta bir sapan değerdir ve beklendiği gibi tahmin edilen regresyon doğrusunu kendisine doğru çekmektedir. Bu nokta ihmal edildiğinde tahmin edilmiş regresyon doğrusu bu noktadan uzağa geri

sıçrar. Kare ile gösterilen sapan değerin ($y_4 = 2.1$) veri kümesinden çıkarılmasından sonra tahmin edilen regresyon doğrusuna göre belirlenen tahmin değeri $\hat{y}_{(4)} = 0.6 + 1.55x$ olur. x=10 durumu için $y_4 = 2.1$ iken bu gözlem değerinin silinmesi ile $\hat{y}_{(4)} = 16.1$ olarak tahmin edilir. Böylece, kare veri noktası için silinmiş kalan $r_4 = 2.1 - 16.1 = -14$ dir (Rao ve Toutenburg 1920:218).

5.1.3. Studentized Kalan Yöntemi

Kalanlara bağlı olarak kullanılan yöntemlerden biri de Studentized kalanlar yöntemidir. Bu yöntem, aynı zamanda bir standartlaştırılmış jackknifed kalanı olarak da adlandırılır ve yanıt değişkenin herhangi bir gözlem değerinin sapan değer olup olmadığını değerlendirmek için kullanışlı bir yöntemdir. Studentized kalan yöntemi, içten ve dıştan studentized yöntemleri olmak üzere ikiye ayrılır (Hadi ve Simonoff 1993).

İçten studentized yöntemi için, sıradan enküçük kareler kalanı kullanılır. Hata terimlerinin tahminleri $\mathrm{Var}(\hat{\boldsymbol{\varepsilon}}) = (\boldsymbol{I} - \boldsymbol{H})\sigma^2$ ile ilişkili olduğundan i. kalan $\hat{r}_i$ ve $\hat{\sigma}^2 = \dfrac{\hat{\boldsymbol{\varepsilon}}^T \hat{\boldsymbol{\varepsilon}}}{N - p}$ olmak üzere

$$s_i = \frac{\hat{r}_i}{\hat{\sigma}\sqrt{1 - h_{ii}}}, \quad i = 1, \ldots, n \tag{5.4}$$

olarak standartlaştırılır. Behnken ve Draper (1972) bir değerin sapan değer olup olmadığını teşhis etmek için, test istatiği olarak

$$R_n = \text{maks}|s_i|$$

ve kritik değer olarak,

$$C_n = \left[(n - p)F/(n - p - 1 + F)\right]^{1/2}$$

ifadesini önerdi. Burada F ,1; $n - p - 1$ serbestlik dereceli F dağılımının $\%(1 - \alpha/n)$ noktasıdır. Bu yöntemde veri kümesinin en fazla bir sapan değer kapsadığı ve hangi değerin sapan değer olduğunun bilinmediği kabul edilir. Bu yöntem aşağıdaki iki adımda özetlenir:

i) $P\left\{R_n > C_n \,\middle|\, \text{en fazla bir sapan değer}\right\} \le \alpha$

ii) $R_n > C_n$ ise *i*. gözlem değeri sapan dağer olarak belirlenir.

Bu değer uç gözlem değerlerinin yüksek kaldıraç noktarı için özel bir değerdir. Bu nedenle, *i*. gözlem değeri *n* boyutlu veri kümesinden çıkarılarak, *i*. değerin uydurulmuş değeri $\hat{y}_{-i}$, geriye kalan *n*-1 gözlem değerinin kullanılması ile hesaplanır. Gözlenen y_i değeri ile $\hat{y}_{-i}$ değeri arasındaki farka *i*. durum için silinmiş artıkları denir ve aşağıdaki şekilde gösterilir:

$$d_i = y_i - \hat{y}_{-i}\,. \tag{5.5}$$

Buna göre indirgenmiş kalan kareler ve *i*. gözlem değerinin silinmesi ile elde edilen $\boldsymbol{\beta}_i$ nin tahmini $\hat{\boldsymbol{\beta}}_{-i}$

$$\hat{\boldsymbol{\beta}}_{-i} = \hat{\boldsymbol{\beta}} - \frac{\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\boldsymbol{x}_i\hat{r}}{1-h_{ii}}\ , \tag{5.6}$$

$$HKT_{-i} = HKT - \frac{\hat{r}_i^2}{1-h_{ii}} \tag{5.7}$$

olarak elde edilir. Burada HKT ve $HKT_{-i} = \hat{\boldsymbol{r}}_{-\boldsymbol{i}}^T\hat{\boldsymbol{r}}_{-\boldsymbol{i}}$ sırası ile *i*. gözlem değerinin göz önüne alındığı ve silindiği durumlara göre hesaplanan hata kareleri toplamlarıdır. *i*. durum için silinmiş artık d_i ve d_i nin tahmin edilen varyansı,

$$d_i = \frac{r_i}{1-h_{ii}}\ , \tag{5.8}$$

$$Var\left(d_i\right) = \frac{HKO_{-i}}{1-h_{ii}} \tag{5.9}$$

olarak ifade edilir. Burada $\mathrm{HKO}_{-\mathrm{i}}$, *i*. durumun olmadığı hata kareler ortalamasıdır.

Studentized kalanlar, bağımsız olmamakla birlikte *n*-*p*-1 serbestlik dereceli *t* - dağılımına sahiptir [Belsley 1980: 33]. Eğer,

$$\frac{d_i}{\sqrt{Var\left(d_i\right)}} \tag{5.10}$$

istatistiği s_i^* ile gösterilirse, s_i^*,

$$s_i^* = \frac{r_i}{\sqrt{\hat{\sigma}_{-i}^2(1-h_{ii})}} \quad , i = 1,2,...,n \ . \tag{5.11}$$

olarak elde edilir (Atkinson 1983). $\hat{\sigma}_{-i}$, i. gözlem değeri $(y_i, \boldsymbol{x}_i)$nin veri kümesinden silinmesi ile elde edilen standart sapma olmak üzere,

$$\hat{\sigma}_{-i}^2 = \frac{\boldsymbol{y}_{-i}^T(\boldsymbol{I} - \boldsymbol{H}_{-i})\boldsymbol{y}_{-i}}{(n-p-1)} \tag{5.12}$$

olarak elde edilir. Burada $\boldsymbol{H}_{-i}$ ve $\boldsymbol{y}_{-i}$, i. gözlem değeri $(y_i, \boldsymbol{x}_i)$nin silinmesi ile elde edilen şapka matrisi ve yanıt vektörüdür. s_i^* kalanı Atkinson (1983) tarafından Jackknife (*i.* dıştan Studentized) kalanlar olarak adlandırılır. σ_i, i. kalanın standart sapması olmak üzere, σ_i nin tahmini $\hat{\sigma}_i$

$$\hat{\sigma}_i = \hat{\sigma}_{-i}\sqrt{1-h_{ii}} \tag{5.13}$$

şeklindedir. Bu durumda $(\boldsymbol{y}_{-i}, \boldsymbol{X}_{-i})$, i. gözlem değerinin silinmesi sonucu kalan veri matrisi olsun. Eğer $rank(\boldsymbol{X}_{-i}) = p$ ve $\boldsymbol{r}$, $N(0, \sigma^2 I_n)$ dağılımına sahip bağımsız rasgele değişkenler ise Jackknife kalanlar $-s_i^*$, $n-p-1$serbestlik dereceli student $-t_{n-p-1}$ dağılımına sahiptir. Eğer $rank(\boldsymbol{X}_{-i}) = p-1$, ise s_i^* kalanları tanımlı değildir (Beckman ve Trussel 1974). Ayrıca s_i^* ve s_i arasında, r_i *nin* kesin olarak artan fonksiyonu olan aşağıdaki ilişki yazılabilir:

$$s_i^* = s_i\sqrt{\frac{n-p-1}{n-p-r_i^2}}\,. \tag{5.14}$$

Test istatistiği $\max|s_i^*|$, maksimum studentized teste eşdeğerdir (Beckman ve Trussell 1974).

5.1.4. Cook Mesafesi Yöntemi

Cook Mesafesi (CM_i) yöntemi, regresyon parametreleri olan $\hat{\beta}$ ların arasındaki farka dayanan ve *i.* veri noktasının kümeden çıkarılması ile elde edilen $\hat{\beta}_{-i}$ tahmininin

nasıl şekillendiğini inceleyen bir etki ölçüsüdür (McDonald 2002). Cook Mesafesi yöntemini aşağıdaki eşitlik ile ifade edilir (Cook ve Weisberg 1982:116):

$$CM_{-i}^2 = \frac{\left(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{-i}\right)^T \boldsymbol{X}^T \boldsymbol{X} \left(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{-i}\right)}{p\hat{\sigma}^2} = \frac{\left(\hat{\boldsymbol{y}} - \hat{\boldsymbol{y}}_{-i}\right)^T \left(\hat{\boldsymbol{y}} - \hat{\boldsymbol{y}}_{-i}\right)}{p\hat{\sigma}^2} \tag{5.19}$$

burada $\hat{\sigma}^2$ hata varyansının tahmin edicisi olmak üzere

$$\hat{\sigma}^2 = \sum_{i=1}^{n} r_i^2 / n - p,$$

şeklinde, $\hat{y}_j$, *j.* gözlem için tam regresyon modelinden elde edilen tahmin, $\hat{y}_{j(i)}$, *i.* gözlemin ihmal edilmesi ile uydurulmuş modelden bulunan *j* gözlemin tahminidir.

CM_i için güven aralığı $CM_{-i}^2 > F_{p,n-p,1-\alpha}$ şeklindedir (McDonald 2002). CM_i nin büyük değeri (yaklaşık olarak 1.0) *i.* gözlemin sapan değer olduğunu gösterir (Chatterjee ve ark. 2000).

Bununla beraber, Cook mesafesi yöntemi için eşdeğer bir eşitlik aşağıdaki şekilde verilebilir:

$$CM_i = \left(\frac{1}{p}\right)\left(\frac{h_{ii}}{1-h_{ii}}\right)\left(\frac{r_i^2}{\hat{\sigma}^2\left(1-h_{ii}\right)}\right). \tag{5.20}$$

5.2. Doğrudan Olmayan Yöntemler

5.2.1. Sağlam (Robust) Regresyon

Robust (sağlam) sözcüğü, küçük sapmalardan etkilenmeyen istatistik işlemleri olarak tanımlanmaktadır (Alpu ve ark. 2010). Doğrusal regresyon analizinde β parametrelerinin tahmini için yaygın olarak kullanılan yöntem EKK tekniğidir. Ancak, bağımlı ve bağımsız değişkendeki kaldıraç ve sapan değerlerin varlığı EKK tekniğinin tahmin gücünü büyük ölçüde azaltır. Bu nedenle EKK tekniğine alternatif olarak sapan değerlerden çok fazla etkilenmeyen tahmin edicileri kullanan sağlam regresyon yöntemleri önerilir (Alpu ve ark. 2010).

Sağlam tekniklerin teorik anlamda performansını tanımlamak için etkinlik, bir tahmincinin keyfi olarak büyük değerler almasına neden olabilecek sapan değerlerin en küçük yüzdesi olarak tanımlanan kırılma noktası (Ben-Gal 2005) ve sınırlı etkili olma özellikleri kullanılır (Simpson ve Montgomery 1998). En popüler sağlam tahmin edicilerden bazıları; maksimum olabilirlik yönteminin genişletilmiş hali olan *M* tahmin edici (Susanti 2008), En Küçük Kesilmiş Kareler (Least Trimmed Squares, LTS) tahmin edicisi (Rousseeuw ve Leroy 1987:120), En Küçük Medyan Kareler (Least Median Square, LMS) tahmin edicisi (Rousseeuw 1984), *S* tahmin edicisi (Rousseeuw ve Yohai 1984) ve Genelleştirilmiş M tahmin edicileridir (Wiens 1996). *M*- tahmin edicileri belirtilen tahmin edicileri ile karşılaştırıldığında nispeten daha etkin olandır (Susanti 2008). *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.

5.2.1.1. En Küçük Kesilmiş Kareler (LTS*)* Tahmin Yöntemi

Rousseeuw (1984) tarafından önerilen LTS tahmin edicisi bir *p*-vektörü olarak tanımlanır. Bu tahmin edici,

$$Q_{LTS}(\boldsymbol{\theta}) = \sum_{i=1}^{h} r_i^2$$

olmak üzere, aşağıdaki en iyileme probleminin çözümünden elde edilir:

$$\hat{\boldsymbol{\theta}}_{\text{LTS}} = \min_{\boldsymbol{\theta}} Q_{\text{LTS}}(\boldsymbol{\theta}). \tag{5.21}$$

Burada $r_1^2 \le r_2^2 \le ... \le r_n^2$ sıralı karesel kalanlar $r_i^2 = (y_i - \mathbf{x}_i^T\boldsymbol{\theta})^2, i = 1,2,...,n$ ve h üst sınırı $(n/2)+1 \le h \le (3n+p+1)/4$ aralığında tanımlanır. Bu tahmin tahmin edici için kırılma (breakdown) değeri $(n-h)/n$ olarak alınır.

5.2.1.2. En Küçük Medyan Kareler (LMS) Tahmin Yöntemi

LMS yöntemi, EKK tekniğinde olduğu gibi hata karelerinin toplamını minimize etmek yerine kareleri alınmış hata terimlerinin medyan değerini minimize eden aşağıdaki en iyileme probleminin çözümünden elde edilir (Rousseeuw 1984):

$$\min_{\hat{\theta}} \underset{i}{med}\, r_i^2. \tag{5.22}$$

Bu öneri özellikle Hampel'ın (1975) bir fikrine dayanmaktadır. LMS tahmin edicisinin kırılma noktası %50 dir fakat düşük yeterliliğe sahiptir.

Rousseeuw ve Van Zomeren (1990) regresyon sapan değerleri teşhis etmek için LMS tahmincisini kullanmayı önermiştir. Bu yöntem LMS regresyon ile ilişkili kalanları hesaplayarak başlar:

$$s = 1.4826\left(1+\frac{5}{n-p-1}\right)\sqrt{M_r}\,. \tag{5.23}$$

Burada M_r, $r_1^2,...,r_n^2$ kalan karelerinin medyanı, p ise modelde bulunan parametre sayısıdır. $\left(y_i, x_{i1},..., x_{ip}\right)$ noktası, ilgili standartlaştırılmış kalan büyük ise bir regresyon sapan değeri olarak belirlenir. Özellikle, Rousseeuw ve Van Zomeren (1990) i. vektörü $\left|r_i\right|/s > 2.5$ ise bir regresyon sapan değeri olarak teşhis eder.

5.2.1.3. S Tahmin Yöntemi

S tahmin yöntemi, M- tahmin yönteminin geliştirilmesiyle elde edilmiştir (Susanti ve ark. 2014). Bu yöntemde medyanın zayıflığını gidermek için kalanların standart sapması kullanılır (Rousseeuw ve Yohai 1984). S tahmin edicisi, en küçük sağlam ölçek tahmin edicisi $\hat{\sigma}_s$ nın belirlenmesi ve

$$\min \sum_{i=1}^{n} \rho\left(\frac{y_i - \sum_{i=1}^{n} \boldsymbol{x}_i^T \boldsymbol{\beta}}{\hat{\sigma}_s}\right) \tag{5.24}$$

yukarıdaki en iyileme problemini sağlayan,

$$\hat{\boldsymbol{\beta}}_s = \min_{\beta}\ \hat{\sigma}_s\left(r_1, r_2,..., r_n\right) \tag{5.25}$$

olarak tanımlanır (Salibian ve Yohai 2006) . Burada $\hat{\sigma}_s = \sqrt{\frac{1}{nK}\sum_{i=1}^{n} w_i r_i^2}$, $K = 0{,}199$, $w_\sigma\left(u_i\right) = \frac{\rho\left(u_i\right)}{u_i^2}$ ve başlangıçtaki tahmin $\hat{\sigma}_s = \frac{med\left|r_i - med\left(r_i\right)\right|}{0{,}6745}$ olarak alınır. (5.24)

probleminin çözümü, ρ nun $\boldsymbol{\beta}$ ya göre aşağıda verilen diferensiyelinin alınması ile elde edilir:

$$\sum_{i=1}^{n} x_{ij}\psi\left(\frac{y_i-\sum_{j=0}^{p} x_{ij}\beta_j}{\hat{\sigma}_s}\right)=0 \qquad j=0,1,...,p\,. \tag{5.26}$$

Burada ψ , ρ nun türevi olan ve aşağıda verilen fonksiyonu,

$$\psi(u_i)=\rho'(u_i)=\begin{cases} u_i\left[1-\left(\frac{u_i}{c}\right)^2\right]^2 & ,\ |u_i|\le c \\ 0 & ,\ |u_i|>c \end{cases}, \tag{5.27}$$

w_i ise $u_i=\frac{r_i}{\sigma_s}$ ve $c=1,547$ olmak üzere

$$w_i(u_i)=\begin{cases} \left[1-\left(\frac{u_i}{c}\right)^2\right]^2 & ,\ |u_i|\le c \\ 0 & ,\ |u_i|>c \end{cases}$$

şeklinde verilen, ardışık olarak ağırlaştırılmış en küçük kareler (*Iteratively Reweighted Least Squares, IRLS*) ağırlık fonksiyonudur. Bu nedenle, (5.26) problemi ardışık olarak yeniden ağırlıklandırılmış en küçük kareler yöntemi kullanılarak çözülebilir.

5.2.1.4. M-Tahmin Yöntemi

En popüler sağlam tahmin yöntemlerinden biri olan M tahmin yöntemi, En Çok Olabilirlik Tahmin yönteminin bir genişlemesidir (Susanti ve ark. 2014). Bu yöntem, uygun olmayan bazı verilerin kullanılmamasını mümkün kılar ve kalanlara bağlı olarak elde edilir. $\boldsymbol{r_i}$, *i*. veriye ait kalanı, yani *i*. gözlem ve bu gözlemin uydurulmuş değeri arasındaki farkı göstermek üzere, standart EKK yöntemi $\sum_i r_i^2$ değerini en küçük yapmaya çalışır, ancak veri kümesinde sapan değer veya değerler mevcutsa EKK tahmin edicisi kararsızlık gösterir. Veri kümesindeki sapan değer enküçükleme problemini güçlü bir şekilde etkilediğinden tahmin edilen parametrenin doğruluğunu olumsuz bir şekilde etkiler. M-tahmin ediciler, EKK tahminde kullanılan r_i^2 karesel

kalanları, sapan değerlerin etkisini azaltacak kalanların bir fonksiyonu ile değiştirme fikrine dayanır. Bu nedenle M-tahmin edici,

$$\min_{\boldsymbol{\beta}} \sum \rho(\boldsymbol{r}_i) \tag{5.28}$$

en iyileme probleminin çözümünden elde edilir. Burada ρ simetrik, pozitif tanımlı ve 0 noktasında tek bir minimum değere sahip olan bir fonksiyondur. Yani ρ aşağıdaki özelliklere sahiptir (Huber 1964):

$(i)\ \rho(\boldsymbol{r}) \geq 0$

$(ii)\ \rho(\boldsymbol{0}) = 0$

$(iii)\ \rho(\boldsymbol{r}) = \rho(-\boldsymbol{r})$

$(iv) |\boldsymbol{r_i}| > |\boldsymbol{r_i}'|$ için $\rho(\boldsymbol{r_i}) \geq \rho(\boldsymbol{r_i}')$.

Bu fonksiyonun tercih edilmesinin sebebi r_i^2 ye göre daha yavaş artmasıdır. Uygulamada kullanılan bazı ρ fonksiyonları çizelge 5.1. de verilmiştir:

Çizelge 5.1. M- tahmin yönteminde yaygın olarak kullanılan bazı $\rho(x), \psi(x), w(x)$ fonksiyonları

Fonksiyonun Adı		$\rho(x)$	$\psi(x)$	$w(x)$
L_2		$x^2/2$	x	1
L_1		$\lvert x\rvert$	$Sgn(x)$	$1/\lvert x\rvert$
$L_2 - L_1$		$2\left(\sqrt{1+x^2/2}-1\right)$	$x/\sqrt{1+x^2/2}$	$1/\sqrt{1+x^2/2}$
Huber	$\lvert x\rvert \leq k$ *ise*	$x^2/2$	x	1
	$\lvert x\rvert \geq k$ *ise*	$k/(\lvert x\rvert - k/2)$	$k\,Sgn(x)$	$k/\lvert x\rvert$
L_p		$\lvert x\rvert^v/v$	$Sgn(x)\lvert x\rvert^{v-1}$	$\lvert x\rvert^{v-2}$

Çizelge 5.1. de kullanılan fonksiyonların olumlu ve olumsuz yönleri vardır. L_2 uzunluğunu kullanan En Küçük Kareler tahmin yönteminde kullanılan etki fonksiyonu sınırlı olmadığından bu yöntemle elde edilen tahmin ediciler sağlam değildir. L_1 uzunluğuna (Mutlak Değer) bağlı olarak elde edilen tahmin edicileri $\rho = |x|$ fonksiyonunun x'te kesin konveks olmamaları nedeni ile kararlı (stable) değildir. Gerçekten de, $x = 0$ da $\rho = |x|$ fonksiyonunun ikinci dereceden türevi sınırlı olmadığından etkin olmayan bir çözüme neden olabilir. $L_2 - L_1$ tahmin edicileri hem L_1 tahmin edicilerinin büyük hataların etkisini azaltması hem de L_2 tahmin edicilerinin konveks olması avantajlarına sahiptir. L_p, en küçük güçler fonksiyonu, bir fonksiyon ailesini temsil eder. $v = 2$ ve $v = 1$ özel durumları sırası L_2 ve L_1 fonksiyonlarını verir. $v > 0$ sayısı ne kadar küçük olursa büyük hataların $\boldsymbol{\beta}$ tahmini üzerinde etkisi o denli küçük olacaktır. Kısaca, v nin değerleri sapan değerden etkilenmeyen bir tahmin ediciyi sağlayacak şekilde ölçülü olmalıdır. $v = 2$ ve $v = 1$ özel durumları sağlam tahmin edicini elde edilmesi için en iyi durumlardır. $1 < v < 2$ durumu sıkıntılı durumlara neden olan sıfır kalanları vereceğinden, bu durumu ele almak oldukça zordur (William 1983:103-117).

Huber tipi M-tahminler istatistiksel olarak (Gaussian hatalı modelde) L_1 regresyona göre daha sağlamdır, aynı zamanda sapan y_i lere göre de daha sağlamdır. (5.28) probleminin çözümü, amaç fonksiyonu $\sum \rho(\mathbf{r}_i)$ nin $\boldsymbol{\beta}$ ya göre türevin alınarak 0 a eşitlenmesi ile bulunur (Huber 1981: 44). Bu durumda $\boldsymbol{\beta}$ nın M-tahmin edicisi $\hat{\boldsymbol{\beta}}_M$, p tane eşitlikten oluşan

$$\sum_i \psi(r_i) \frac{\partial r_i}{\partial \beta_j} = 0, \quad j=1,2,...,p \tag{5.29}$$

denklem sisteminin çözümünden elde edilir. Burada $\psi(x) := \partial\rho(x)/\partial x$ türev ifadesine *etki fonksiyonu* denir. Bir sağlam M-tahmincisini elde etmek için kullanılan bu fonksiyonun sınırlı olması gerekir. Eğer $w(x) := \psi(x)/x$ şeklinde bir ağırlık fonksiyonu tanımlarsak (5.29) eşitliği

$$\sum_i w(r_i) r_i \frac{\partial r_i}{\partial \beta_j} = 0, \quad j=1,2,...,p, \tag{5.30}$$

eşitliğine dönüşür. Bu ise tam olarak,

$$\min_{\boldsymbol{\beta}} \sum_i w\left(r_i^{(k-1)}\right) r_i^2 \tag{5.31}$$

şeklinde verilen IRLS yöntemine ilişkin en iyileme problemidir. Burada k iterasyon sayısını gösterir. $w\left(r_i^{(k-1)}\right)$ ağırlığı bir sonraki iterasyonda kullanılması için yeniden hesaplanır. Bununla birlikte, bir sağlam M-tahmincisini elde etmek için kullanılan etki fonksiyonunun sınırlı olması ve sağlam M-tahmin edicisinin tek olması için ρ-fonksiyonunun konveks olması gerekir.

6. ORTALAMA ÖTELEMELİ SAPAN DEĞER (OÖSD) MODELİ

Sapan değer belirleme yöntemleri çalışılırken, *Ortalama Ötelemeli Sapan Değer* modeli ihmal edilmemelidir. Çünkü bu model ilgili sapan değerin ihmal edilmesi sonucu elde edilen uydurulmuş model ile aynı kalanlara sahiptir (Kima ve ark. 1974). OÖSD modeli aşağıdaki şekilde ifade edilir (Rao ve Toutenburg 1999: 220):

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e}_i\delta + \boldsymbol{\varepsilon}, \tag{6.1}$$

Burada e_i i. birim vektördür. Bu modelde, y_i ya da $\boldsymbol{x}_i^T\boldsymbol{\beta}$ değerlerinin sistematik olarak δ kadar $y_i = \boldsymbol{x}_i^T\boldsymbol{\beta} + \varepsilon_i$ modelinden sapma gösterdiği kabul edilir. Böylece i. gözlem değeri olan $(y_i, \boldsymbol{x}_i^T\boldsymbol{\beta})$ geriye kalan gözlemlerden farklı bir yol izler ve bu nedenle $(y_i, \boldsymbol{x}_i\boldsymbol{\beta})$ sapan değer olarak değerlendirilir. Bu durum,

$$H_0 : \delta = 0 \quad \left(\text{i.e., } y_i = \boldsymbol{x}_i^T\boldsymbol{\beta}\right)$$

hipotezinin

$$H_1 : \delta \neq 0 \quad \left(\text{i.e., } y_i \neq \boldsymbol{x}_i^T\boldsymbol{\beta}\right),$$

alternatif hipotezine karşı olabilirlik-oran (likelihood ratio) yöntemine dayalı olarak test edilir. Bu hipotez için test istatistiği olarak,

$$F_i = \frac{\left[HKT\left(H_0\right) - HKT\left(H_1\right)\right]/1}{HKT\left(H_1\right)/(n-p-1)}. \tag{6.2}$$

alınır (Rao ve Toutenburg 1999: 221). Burada, $HKT\left(H_0\right)$, n gözlemin tamamını içeren $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta}$ modeline göre elde edilen kalan kareler toplamı, $HKT(H_1)$ ise $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + e_i\delta + \boldsymbol{\varepsilon}$ modeline göre elde edilen kalan kareler toplamını ifade eder. Bunlar aşağıdaki ifadelerle verilir:

$$HKT\left(H_0\right) = \sum_{i=1}^{n}\left(y_i - \boldsymbol{x}_i^T\boldsymbol{\beta}\right)^2 = \boldsymbol{y}^T\left(\boldsymbol{I} - \boldsymbol{H}\right)\boldsymbol{y} = \left(n - p\right)\hat{\sigma}^2,$$

$$HKT\left(H_1\right) = \sum_{i=1}^{n}\left(y_i - \boldsymbol{x}_i^T\boldsymbol{\beta} - \delta_i\right)^2,$$

burada δ_i sapan değer değişkeninin durumuna göre ya δ ya da sıfır olacaktır. Aynı zamanda $HKT(H_0)$ ile $HKT(H_1)$ arasında

$$HKT(H_1) = HKT(H_0) - \frac{\hat{\varepsilon}_i^2}{1-h_{ii}} \ , \tag{6.3}$$

şeklinde bir ilişki vardır. Burada $\hat{\varepsilon} = (\boldsymbol{I} - \boldsymbol{H})\boldsymbol{y}$, $\hat{\varepsilon}_i = e_i^T \hat{\varepsilon}$ ve $\boldsymbol{e}_i^T \boldsymbol{H} \boldsymbol{e}_i = h_{ii}$ dir. *i*. gözlem $(y_i, \boldsymbol{x}_i)$ ihmal edildiğinde, σ_i^2 standart sapmanın tahmin edicisi $s_{-i}^2 = \boldsymbol{y}_{-i}^T (\boldsymbol{I} - \boldsymbol{H}_{-i}) \boldsymbol{y}_{-i} / n - p - 1$ şeklinde ifade edilir. Burada $\boldsymbol{H}_{-i}$ ve $\boldsymbol{y}_{-i}$ sırasıyla, *i*. gözlemin çıkarılmasından sonra elde edilen şapka matrisi ve yanıt vektörü, σ_i ise *i*. kalanın standart sapmasıdır. Eğer σ_i nin tahmini, $\hat{\sigma}_i = s_{-i}\sqrt{1-h_{ii}}$, (6.2) eşitliğinde yazılırsa , s_i^* dıştan studentized kalan (Rao ve Toutenburg 1999:219) olmak üzere test istatistiği aşağıdaki şekilde yazılabilir:

$$F_i = \frac{\hat{\varepsilon}_i^2}{s_{-i}^2 (1-h_{ii})} = \left(s_i^*\right)^2 \quad (i = 1, 2, ..., n).$$

Beckman ve Trusell'ın (1974) teoremine göre, eğer $\text{rank}(\boldsymbol{X}_{-i}) = p$ ve $\boldsymbol{\varepsilon} \sim N(0, \sigma^2 \boldsymbol{I}_n)$ ise dıştan Studentized kalanlar $s_i^* (i = 1, ..., n)$, $n - p - 1$serbestlik dereceli student t_{n-p-1} dağılımına sahiptir. Eğer $\text{rank}(\boldsymbol{X}_{-i}) = p - 1$ ise, s_i^* kalanı tanımlı olmayacaktır. Eğer $\text{rank}(\boldsymbol{X}_{-i}) = p$ olduğu kabul edilirse, Beckman ve Trusell (1974) teoreminin ilk özelliğine dayanarak $F_i = \left(s_i^*\right)^2$ test istatistiğinin sırasıyla, H_0 varsayımı altında merkezi $F_{1,n-p-1}$ dağılımına sahip olduğu ve H_1 varsayımı altında merkezi olmayan $F_{1,n-p-1}\left(\delta^2 (1-h_{ii}) \sigma^2\right)$ dağılımına sahip olduğu söylenebilir. Merkezi olmama parametresi $\delta^2 (1-h_{ii}) \sigma^2$, h_{ii} arttıkça azalacaktır. Yani, sapan değerlerin teşhisi h_{ii} büyüdükça zorlaşır.

6.1. Ortalama Ötelemeli Sapan Değer Modeli İçin Parametre Tahmini

6.1.1. M-Tahmin Yöntemi ile OÖSD de Parametre Tahmini

Bu bölümde ortalama ötelemeli sapan değer modelindeki parametreleri tahmin etmek için M- tahmin yöntemini kullanılacaktır. Hadi ve Simonoff (1993), (6.2) ile verilen F_i test istatistiği, Cook mesafesi (Cook, 1979) veya Studentized kalanlar (Beckman ve Trusell 1974) gibi direkt yöntemlerden birini kullanarak $y_i = \boldsymbol{x}_i^T \boldsymbol{\beta} + \varepsilon_i$ modelinden δ_i kadar sistematik olarak sapmaya uğrayan m tane y_i gözlemine sahip olduğumuzu düşünelim, $m < n$. Yani m tane gözlem sapan değer olsun. Buna göre Ortalama Ötelemeli Sapan Değer (OÖSD) modeli aşağıdaki şekilde yazılabilir (Taylan ve ark. 2014):

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{E}\boldsymbol{\delta} + \boldsymbol{\varepsilon}. \tag{6.4}$$

$\boldsymbol{X}$ tam ranklı açıklayıcı bağımsız veya açıklayıcı gözlemlerin $n \times p$ tipinde matrisi, $\boldsymbol{E}$, m göstergeç değişkenli $n \times m$ tipinde bir matris ve $\boldsymbol{\delta}$ göstergeç değişkenlerin regresyon katsayılarından oluşan $m \times 1$ tipinde vektördür. Buna göre, OÖSD modeli

$$\boldsymbol{y} = \boldsymbol{X}^*\boldsymbol{\beta}^* + \boldsymbol{\varepsilon}, \tag{6.5}$$

şeklinde gösterilebilir. Burada $\boldsymbol{X}^* = (\boldsymbol{X} \mid \boldsymbol{E})$ şeklindedir, yani, $\boldsymbol{X}$ ve $\boldsymbol{E}$ matrislerinden oluşan $n \times (p+m)$ tipinde blok matris, $\boldsymbol{\beta}^*$ da $\boldsymbol{\beta}$ ve $\boldsymbol{\delta}$ vektörlerinden oluşan $\boldsymbol{\beta}^* = (\boldsymbol{\beta}^T \mid \boldsymbol{\delta}^T)^T$ şeklindeki $(p+m) \times 1$ tipinde vektördür. OÖSD modeli için M-tahmin edicisi, i. kalan değeri $r_i^* = y_i - \boldsymbol{x}_i\boldsymbol{\beta} - \delta_i$ olmak üzere,

$$G(\boldsymbol{r}^*) = \sum_{i=1}^{n} \rho\left(r_i^*\right) \tag{6.6}$$

olmak üzere

$$\min_{\boldsymbol{\beta}^*} G(\boldsymbol{r}^*) \tag{6.7}$$

en iyileme probleminin çözümünden elde edilir. M- tahmin edicisini elde etmek için gereken konveks ve diferensiyellenebilen ρ fonksiyonu olarak,

$$\rho\left(r_i^*\right)=\begin{cases}\dfrac{r_i^{*2}}{2} & , \quad \left|r_i^*\right| \le k \\ k\left|r_i^*\right|-\dfrac{k^2}{2} & , \quad \left|r_i^*\right| > k\end{cases} \tag{6.8}$$

şeklindeki Huber tipi fonksiyon kullanılacaktır (Huber 1981). Burada *k*, $(k>0)$ ayar sabiti olarak adlandırılır. *k* nın küçük değerleri, hataların normal dağıldığı durumda daha düşük etkinliğe rağmen sapan değerlere daha dirençli kestiriciler verir. Genel olarak, ayar sabiti yüksek etkinliği verecek bir değer olarak alınır. Huber tipi fonksiyon için,

$k=1.345\sigma$ değeri, hataların normal dağıldığı durumda %95 düzeyinde bir etkinlik üreterek sapan değere karşı bir koruma sağlar. Burada σ hataların standart sapmasıdır. *k* değerini kullanabilmek için σ nın bilinmesi gerekir. Kalanların standart sapması için tercihen yayılımın güçlü bir ölçüsü kullanılır. Bu ölçü için genel olarak, Med:= Orta değer olmak üzere,

$$\sigma=\frac{\text{Med}\left|\boldsymbol{r}^*\right|}{0.6745} \tag{6.9}$$

değeri alınır (Fox ve Weisberg 2010:5). Bununla beraber ayar sabiti,

$$\sigma = Med\left(\left|r^*\right|\right)+4MAD\left(\left|r^*\right|\right) \tag{6.10}$$

olmak üzere, Verboon (1992) tarafından $k=\frac{2}{3}\sigma$ olarak tanımlanır.

(6.7) probleminin çözümü için aşağıdaki iki durum göz önüne alınır:

i) $\left|r_i^*\right| \le k$ olsun. Bu durumda $\rho\left(r_i^*\right)=\dfrac{r_i^{*2}}{2}$ olarak göz önüne alınır ve (6.6) ifadesi,

$$\begin{aligned} G(\boldsymbol{\beta}_i^*) &= \tfrac{1}{2}\left(\boldsymbol{y}-\boldsymbol{X}^*\boldsymbol{\beta}^*\right)^{\boldsymbol{T}}\left(\boldsymbol{y}-\boldsymbol{X}^*\boldsymbol{\beta}^*\right) \\ &= \tfrac{1}{2}\left\|\boldsymbol{y}-\boldsymbol{X}^*\boldsymbol{\beta}^*\right\|^2, \end{aligned}$$

olarak yazılacağından, (6.7) problemi

$$\min_{\boldsymbol{\beta}^*} \tfrac{1}{2}\left\|\boldsymbol{y}-\boldsymbol{X}^*\boldsymbol{\beta}^*\right\|^2 \tag{6.11}$$

problemine dönüşür. Bu problem en küçük kareler tahmin yöntemi ile çözülür (Rao ve Toutenburg 1999:25). Yani. $\left|r_i^*\right| \le k$ olduğu durumda $\boldsymbol{\beta}^*$ parametresinin M tahmin

edicisi aynı zamanda en küçük kareler (EKK) tahmin edicisidir ve bu tahmin edici $\rho(r_i^*)$ fonksiyonuna dayalı aşağıdaki *n* tane eşitliğin çözümü olan $\hat{\boldsymbol{\beta}}_{\boldsymbol{M}}^*$ vektörüdür:

$$\sum_i \psi(r_i^*)\frac{\partial r_i^*}{\partial \boldsymbol{\beta}_i^*} = 0,\ i = 1,...,n\ . \tag{6.12}$$

$\rho\left(r_i^*\right) = r_i^{*2}/2$ olduğundan etki fonksiyonu $\psi\left(r_i^*\right) = r_i^*$ olarak, ağırlık fonksiyonu ise $w(r_i^*) := \psi(r_i^*)/r_i^* = 1$ olarak elde edilir.

Ağırlıklandırılmış EKK iteratif yöntemine ilişkin çözülen eşitlik sistemi $\min \sum_i w\left(r_i^{*(k-1)}\right) r_i^*$ şeklindedir. Burada *k* iterasyon sayısını gösterir. $w\left(r_i^{*(k-1)}\right)$ ağırlığı bir sonraki iterasyonda kullanılması için yeniden hesaplanır. $\hat{\boldsymbol{\beta}}^*$ ın EKK tahmini, kalanların karelerinin toplamı olan

$$\text{EKK} =: \sum_{i=1}^{n}\left(y_i - \boldsymbol{x}_i^{*T}\boldsymbol{\beta}^*\right)^2 = \left(\boldsymbol{y} - \boldsymbol{X}^*\boldsymbol{\beta}^*\right)^T\left(\boldsymbol{y} - \boldsymbol{X}^*\boldsymbol{\beta}^*\right) = \left\|\boldsymbol{y} - \boldsymbol{X}^*\boldsymbol{\beta}^*\right\|^2 \tag{6.13}$$

ifadesini en küçük yapan $\boldsymbol{\beta}^*$ değeri olarak alınır. Bu değer en iyi değerdir ve buna bağlı olarak ve *y* nin tahmin edilen değeri aşağıdaki şekilde elde edilir:

$$\begin{aligned} \hat{\boldsymbol{\beta}}^* &= (\boldsymbol{X}^{*T}\boldsymbol{X}^*)^{-1}\boldsymbol{X}^{*T}\boldsymbol{y}, \\ \hat{\boldsymbol{y}} &= \boldsymbol{X}^*\hat{\boldsymbol{\beta}}^* = \boldsymbol{X}^*(\boldsymbol{X}^{*T}\boldsymbol{X}^*)^{-1}\boldsymbol{X}^{*T}\boldsymbol{y} \end{aligned}\ . \tag{6.14}$$

ii) $\left|r_i^*\right| > k$ olsun. Bu durumda, Huber tipi M- tahmin edicisi için (6.7) problemi,

$$\rho(r_i^*) = k\left|r_i^*\right| - \frac{k^2}{2}$$

olacağından,

$$\min \sum_i \left(k\left|y_i - \boldsymbol{x}_i^{*T}\boldsymbol{\beta}^*\right| - \frac{1}{2}k^2\right)$$

problemine dönüşür. Bu problem yinelemeli yeniden ağırlıklandırılmış en küçük kareler yöntemi ile çözülebilir. Buna göre etki fonksiyonu $\psi(r_i^*) = k.\text{sgn}(r_i^*)$ ve ağırlık fonksiyonu $w(r_i^*) = k/\left|r_i^*\right|$ olacaktır. Bu problem, $\boldsymbol{W}$ matrisi köşegen elemanları

$w(r_i^*)$ olan ağırlık matrisi olmak üzere $\boldsymbol{X}^{*T}\boldsymbol{W}\boldsymbol{X}^*\boldsymbol{\beta}^* = \boldsymbol{X}^{*T}\boldsymbol{W}\boldsymbol{y}$ olarak yazılabilir. Buradan hareketle,

$$\hat{\boldsymbol{\beta}}^* = (\boldsymbol{X}^{*T}\boldsymbol{W}\boldsymbol{X}^*)^{-1}\boldsymbol{X}^{*T}\boldsymbol{W}\boldsymbol{y} \tag{6.15}$$

olarak bulunur.

6.1.2. OÖSD Modelinin M-Tahmin Edicisini Elde Etmek İçin Tikhonov Düzenlemesi Yöntemi

Sağlam regresyonun amacı sapan değerlerden etkilenmeyen çözümler üretmektir (Huber 1981:1). Yukarıda da (böl. 5.2.1.4.) açıklandığı gibi en önemli sağlam yöntemlerden biri M-tahmin edicisidir. Bununla beraber, M-tahmin edicisi elde edilirken kötü koşullu bir problemle (düzensiz veya kararsız) karşılaşılabilir. Bu nedenle regresyon modelindeki parametrelerin M-tahmin edicisi araştırılırken Tikhonov düzenlemesinin göz önüne alınması çözümü sapan değere karşı daha güçlü yapacaktır. Bu nedenle modeldeki parametrelerin M-tahmin edicisini elde etmek için Tikhonov düzenlemesi göz önüne alınacaktır (Nemirovski 2002). OÖSD modelinde M-tahmincisini elde etmek için, Huber kriterinin kullanılması ile oluşturulan (6.11) problemi için Tikhonov düzenlemesi problemi,

$$\begin{aligned} &\min \ \left\|\boldsymbol{\beta}^*\right\|_2 \\ &k.s.\ \tfrac{1}{2}\left\|\boldsymbol{y}-\boldsymbol{X}^*\boldsymbol{\beta}^*\right\|_2 \le \tau_1, \end{aligned} \tag{6.16}$$

veya

$$\begin{aligned} &\min \ \tfrac{1}{2}\left\|\boldsymbol{y}-\boldsymbol{X}^*\boldsymbol{\beta}^*\right\|_2 \\ &k.s.\ \left\|\boldsymbol{\beta}^*\right\|_2 \le \tau_2. \end{aligned} \tag{6.17}$$

şeklindedir. Tikhonov düzenlemesi problemi, veriyi uydurmak için gerekli olan model özelliklerini kullanıp, karşıt olarak gereksiz olanları düzenleme ile modelden kaldırarak modeli küçültür. Bu düzenleme probleminin üçüncü formu Lagrange çarpanlarının uygulanması ile elde edilen etkisi azaltılmış en küçük kareler problemidir, bu problem

$$\min \ \left\|\boldsymbol{y}-\boldsymbol{X}^*\boldsymbol{\beta}^*\right\|_2^2 + \lambda^2\left\|\boldsymbol{\beta}^*\right\|_2^2, \tag{6.18}$$

şeklinde yazılır. Bu problem, (6.16) problemine Lagrange çarpanları eniyileme yönteminin (Aster ve ark. 2005:90) uygulanması ile elde edilir. Burada, λ^2 ceza parametresidir. Bu parametre uyumun iyiliği ve modelin karmaşıklığı arasında bir denge sağlar. (6.16) ve (6.17) problemlerinin çözümü, (6.18) denkleminin kullanılarak kısıtlar sağlanıncaya kadar uygun λ parametresinin seçimi ile elde edilir, bununla beraber τ_1, τ_2 ve λ parametrelerine uygun seçimi ile bu üç problem aynı çözüme sahiptir.

Ayrıca, $\left\|\boldsymbol{\beta}^*\right\|_2$ ın $\left\|\boldsymbol{y}-\boldsymbol{X}^*\boldsymbol{\beta}^*\right\|_2$ e karşı grafiği log-log scale programının kullanılması ile çizildiğinde, bunların en iyi değerlerinin eğrisi genel olarak ‘‘L” şeklindedir (Hansen 1992).

Uyuşmazlık (discrepancy) ilkesine ek olarak, λ nın değerini seçmek için diğer bir popüler kriter de yukarıda ifade edilen *L-eğrisi* dir. Burada, *L-eğrisi*nin köşesine en yakın çözümü veren λ değeri seçilir (Aster ve ark. 2005:91).

Daha önce belirtildiği gibi sağlam regresyonun amacı sapan değerlerin etkisini sınırlayıcı veya sapan değerlere karşı dirençli sonuçlar sağlamaktır. Bu nedenle MSOM için M- tahmin edicisini elde etmek için Tikhonov düzenlemesi problemini kullanarak, bu probleme modern konveks en iyileme tekniklerinden Konik Karesel Programlama (CQP) ve En Küçük Mutlak Küçültme ve Operatör Seçimi (LASSO) yöntemlerini kullanarak alternatif çözümler göz önüne alınacaktır.

$\left|r_i^*\right| > k$ olduğu durumda Huber tipi M- tahmin edicisinin, $\rho(r_i^*) = k\left|r_i^*\right| - \dfrac{k^2}{2}$ şeklinde olduğu belirtilmişti. Bu problemin çözümü için Tikhonov düzenlemesi aşağıdaki şekilde yapılabilir: $r_i^* \geq 0$ olduğu durumda $\rho(r_i^*)$,

$$\begin{aligned}
\rho(r_i^*) &= k(y_i - \frac{k}{2} - \boldsymbol{x}_i^{*T}\boldsymbol{\beta}_i^*) \\
&= k(y_i^* - \boldsymbol{x}_i^{*T}\boldsymbol{\beta}_i^*) \\
&= \frac{k}{y_i^* - \boldsymbol{x}_i^{*T}\boldsymbol{\beta}_i^*}(y_i^* - \boldsymbol{x}_i^{*T}\boldsymbol{\beta}_i^*)^2 \\
&= w_i(y_i^* - \boldsymbol{x}_i^{*T}\boldsymbol{\beta}_i^*)^2 \\
&= \sqrt{w_i}^2(y_i^* - \boldsymbol{x}_i^{*T}\boldsymbol{\beta}_i^*)^2
\end{aligned} \tag{6.19}$$

olarak yazılabilir. $r_i^* < 0$ olduğu durumda ise $\rho(r_i^*)$,

$$\begin{aligned}\rho(r_i^*) &= k(-y_i - \frac{k}{2} + \boldsymbol{x}_i^{*T}\boldsymbol{\beta}_i^*) \\ &= -k(y_i + \frac{k}{2} - \boldsymbol{x}_i^{*T}\boldsymbol{\beta}_i^*) \\ &= -k(y_i^* - \boldsymbol{x}_i^{*T}\boldsymbol{\beta}_i^*) \\ &= \frac{-k}{y_i^* - \boldsymbol{x}_i^{*T}\boldsymbol{\beta}_i^*}(y_i^* - \boldsymbol{x}_i^{*T}\boldsymbol{\beta}_i^*)^2\end{aligned} \tag{6.20}$$

olarak elde edilir. Burada $y_i^* - \boldsymbol{x}_i^{*T}\beta_i^* < 0$ olduğu varsayımı altında $\rho(r_i^*)$,

$$\begin{aligned}\rho(r_i^*) &= w_i(y_i^* - \boldsymbol{x}_i^{*T}\boldsymbol{\beta}_i^*)^2 \\ &= \sqrt{w_i}^2(y_i^* - \boldsymbol{x}_i^{*T}\boldsymbol{\beta}_i^*)^2\end{aligned} \tag{6.21}$$

olarak yazılır. Yukarıdaki eşitliklerde,

$$y_i^* := \begin{cases} y_i - \frac{k}{2} & r_i^* \geq 0 \\ y_i + \frac{k}{2} & r_i^* < 0 \end{cases} \quad \text{ve} \quad w_i := \begin{cases} \frac{k}{r_i^*} & r_i^* \geq 0 \\ \frac{-k}{r_i^*} & r_i^* < 0 \end{cases} \tag{6.22}$$

şeklinde tanımlanmıştır. $\boldsymbol{W} = \left(diag\sqrt{w_i}\right)$ olmak üzere , $G(\boldsymbol{\beta}_i^*)$,

$$\begin{aligned}G(\boldsymbol{\beta}_i^*) &= (\boldsymbol{y}^* - \boldsymbol{X}^{*T}\boldsymbol{\beta}^*)^T\boldsymbol{W}^T\boldsymbol{W}(\boldsymbol{y}^* - \boldsymbol{X}^{*T}\boldsymbol{\beta}^*) \\ &= (\boldsymbol{W}\boldsymbol{y}^* - \boldsymbol{W}\boldsymbol{X}^{*T}\boldsymbol{\beta}^*)^T(\boldsymbol{W}\boldsymbol{y}^* - \boldsymbol{W}\boldsymbol{X}^{*T}\boldsymbol{\beta}^*) \\ &= (\boldsymbol{y}^{**} - \boldsymbol{A}\boldsymbol{\beta}^*)^T(\boldsymbol{y}^{**} - \boldsymbol{A}\boldsymbol{\beta}^*) \\ &= \left\|\boldsymbol{y}^{**} - \boldsymbol{A}\boldsymbol{\beta}^*\right\|^2\end{aligned} \tag{6.23}$$

olarak elde edilir. Burada $\boldsymbol{y}^{**} = \boldsymbol{W}\boldsymbol{y}^*$ ve $\boldsymbol{A} = \boldsymbol{W}\boldsymbol{X}^{*T}$ şeklindedir. Bu problem için yukarıda da ifade edildiği gibi üç şekilde Tikhonov düzenlemesi problemi yazılabilir. Bu problem için Tikhonov düzenlemesi probleminin aşağıdaki 3. formu kullanılacaktır:

$$\min \left\|\boldsymbol{y}^{**} - \boldsymbol{A}\boldsymbol{\beta}^*\right\|_2^2 + \kappa^2\left\|\boldsymbol{\beta}^*\right\|_2^2, \tag{6.24}$$

Burada $\lambda = \kappa^2$ birinci ve ikinci bölüm arasındaki koordinasyonu sağlayan ceza parametresidir (Aster ve ark. 2005:91). Bu problem, $\boldsymbol{A}$ katsayı matrisinin tekil değer ayrışımı kullanılarak, MATLAB programı ile kolayca çözülebilir (Hansen 1994).

Bununla birlikte bu probleme sürekli en iyileme problemlerinin çözüm yöntemi olan konik karesel programlama yöntemi (CQP) ile bir alternatif çözüm üretilecektir.

6.1.3. OÖSD Modelinin M-Tahmin Edicisini Elde Etmek İçin Konik Karesel Programlama

Bu bölümde, OÖSD modelinin M-tahmin edicisini elde etmek için 6.1.1. de elde edilen problem, konveks en iyileme tekniklerinden olan ve iç noktalar yöntemi ile çözülebilen *Konik Karesel Programlama* ile çözülecektir. Konik karesel programlama ve iç noktalar yöntemi güçlü metotları ve gelişmiş teorileri ile iyi yapılanmış forma sahip olduklarından nümerik olarak gelişmiş ve kullanışlı yöntemlerdir. Bu şekildeki matematiksel yaklaşım ile Genelleştirilmiş Çapraz Değerlendirme Yöntemi gibi modelden bağımsız yaklaşımlara bir alternatif sunarak, kullanılan yöntemlerin model oluşturmadaki performanslarını karşılaştırmak için, ortalama hata karelerinin kökü (RMSE), Korelasyon katsayısı (CC), Belirleme katsayısı (R^2) gibi istatistiksel performans kriterleri kullanılacaktır.

En iyileme probleminde ve iç noktalar yönteminde kullanılan M parametresi performans ve karşılaştırma yöntemlerine bağlı olarak seçilecektir. (6.24) te yazılan Tikhonov düzenlemesi problemi uygun bir M sınırına bağlı olarak konik karesel problem olarak aşağıdaki şekilde yazılabilir:

$$\begin{aligned} &\min \left\| \boldsymbol{A}\boldsymbol{\beta}^{*} - \boldsymbol{y}^{**} \right\|_2^2 \\ &\text{k.s. } \left\| \boldsymbol{\beta}^{*} \right\|_2^2 \le M. \end{aligned} \qquad (6.25)$$

Burada, λ ve M in uygun seçimleri için (6.24) ve (6.25) problemlerini aynı çözümlere sahiptir. Bu çalışmada, M parametresi deneme ve yanılma yolu ile seçilecektir. Daha sonraki çalışmalarda çeşitli algoritmalar göz önüne alınarak bu parametre belirlenebilir. (6.25) ifadesinde de görülebileceği gibi problem en küçük kareler amaç fonksiyonu $\left\| \boldsymbol{A}\boldsymbol{\beta}^{*} - \boldsymbol{y}^{**} \right\|_2^2$ ve negatif olmayan $-\left\| \boldsymbol{\beta}^{*} \right\|_2^2 + M$ şeklindeki kısıt (k.s.) fonksiyonundan oluşur. Eğer (6.25) en iyileme problemi göz önüne alınırsa, bu problemdeki amaç fonksiyonu doğrusal değildir. Dolayısı ile doğrusal olmayan bu amaç fonksiyonu problemin kısıtları kümesine eklenir ve bu problem eşdeğer olarak,

$$\begin{aligned} &\min_{t,\boldsymbol{\beta}^*} \quad t \\ &\text{k.s.} \quad \left\|\boldsymbol{A}\boldsymbol{\beta}^* - \boldsymbol{y}^{**}\right\|_2^2 \le t^2, \quad t \ge 0, \\ &\qquad\qquad \left\|\boldsymbol{\beta}^*\right\|_2^2 \le M, \end{aligned} \tag{6.26}$$

veya

$$\begin{aligned} &\min_{t,\boldsymbol{\beta}^*} \quad t \\ &\text{k.s} \quad \left\|\boldsymbol{A}\boldsymbol{\beta}^* - \boldsymbol{y}^{**}\right\|_2 \le t, \\ &\qquad\qquad \left\|\boldsymbol{\beta}^{**}\right\|_2 \le \sqrt{M}. \end{aligned} \tag{6.27}$$

formlarında yazılabilir. k, çözülecek en iyileme problemine göre tanımlanan kısıt fonksiyonlarının sayısı olmak üzere CQP nin genel formu (Taylan ve ark. 2007),

$$\begin{aligned} &\min_{\boldsymbol{u}} \quad \boldsymbol{c}^T\boldsymbol{u} \\ &k.s. \quad \left\|\boldsymbol{D}_i\boldsymbol{u} - \boldsymbol{d}_i\right\|_2 \le \boldsymbol{p}_i^T\boldsymbol{u} - q_i \quad (i = 1,2,...,k). \end{aligned} \tag{6.28}$$

şeklindedir. Eğer (6.27) deki en iyileme problemi CQP nin genel formu ile karşılaştırılırsa, bu problemin aşağıdaki özelliklerle bir CQP problemi olduğu görülür:

$$\boldsymbol{c} = \begin{pmatrix} 1 & | & \boldsymbol{0}_{p+m}^T \end{pmatrix}^T, \boldsymbol{u} = (t, \boldsymbol{\beta}^{*T})^T, \ \boldsymbol{D}_1 = \begin{pmatrix} \boldsymbol{0}_{n\times 1} & | & \boldsymbol{A}_{n\times(p+m)} \end{pmatrix}, \boldsymbol{d}_1 = \boldsymbol{y}^{**},$$

$$\boldsymbol{p}_1 = (1,0,...,0)_{1\times(p+m+1)}^T, \ q_1 = 0, \boldsymbol{D}_2 = \begin{pmatrix} \boldsymbol{0}_{(p+m)\times 1} & | & \boldsymbol{I}_{(p+m)\times(p+m)} \end{pmatrix}, \ \boldsymbol{d}_2 = 0_{(p+m)\times 1},$$

$\boldsymbol{p}_2 = \boldsymbol{0}_{(p+m+1)\times 1}$ ve $q_2 = -\sqrt{M}$.

Bu problem için eniyilik (optimality) koşullarını yazmak için öncelikle (6.27) problemini aşağıdaki gibi yeniden düzenlenir:

$$\begin{aligned} &\min_{t,\boldsymbol{\beta}^*} \quad t \\ &k.s. \quad \boldsymbol{\chi} = \begin{pmatrix} \boldsymbol{0}_{n\times 1} & \boldsymbol{A} \\ 1 & \boldsymbol{0}_{(p+m)\times 1}^T \end{pmatrix}\begin{pmatrix} t \\ \boldsymbol{\beta}^* \end{pmatrix} + \begin{pmatrix} -\boldsymbol{y}^{**} \\ 0 \end{pmatrix}, \\ &\qquad \boldsymbol{\eta} = \begin{pmatrix} \boldsymbol{0}_{(p+m)\times 1} & \boldsymbol{I}_{(p+m)\times(p+m)} \\ \boldsymbol{0} & \boldsymbol{0}_{(p+m)\times 1}^T \end{pmatrix}\begin{pmatrix} t \\ \boldsymbol{\beta}^* \end{pmatrix} + \begin{pmatrix} \boldsymbol{0}_{(p+m)\times 1} \\ \sqrt{M} \end{pmatrix}, \\ &\qquad \boldsymbol{\chi} \in L^{n+1}, \ \boldsymbol{\eta} \in L^{p+m+1}, \end{aligned} \tag{6.29}$$

burada L^{n+1}, L^{p+m+1} sırasıyla $(N+1)$ ve $(p+m+1)$-boyutlu *ikinci dereceden veya Lorentz* konileridir. (6.26) daki primal probleme karşılık gelen *dual problem*,

$$
\begin{aligned}
&\text{maks} \quad \begin{pmatrix} \boldsymbol{y}^{**} \\ 0 \end{pmatrix}^T \boldsymbol{\omega}_1 + \begin{pmatrix} \mathbf{0}_{(p+m)\times 1} \\ -\sqrt{M} \end{pmatrix} \boldsymbol{\omega}_2 \\
&\text{k.s.} \quad \begin{pmatrix} \mathbf{0}_{n\times 1}^T & 1 \\ A^T & \mathbf{0}_{(p+m)\times 1}^T \end{pmatrix} \boldsymbol{\omega}_1 + \begin{pmatrix} \mathbf{0}_{(p+m)\times 1}^T & 0 \\ \boldsymbol{I}_{(p+m)\times(p+m)} & \mathbf{0}_{(p+m)\times 1} \end{pmatrix} \boldsymbol{\omega}_2 = \begin{pmatrix} 1 \\ \mathbf{0}_{(p+m)\times 1} \end{pmatrix}, \qquad (6.30) \\
&\qquad\qquad \boldsymbol{\omega}_1 \in L^{n+1},\ \boldsymbol{\omega}_2 \in L^{p+m+1}.
\end{aligned}
$$

şeklindedir.

Bununla beraber, $(t, \boldsymbol{\beta}^*, \boldsymbol{\chi}, \boldsymbol{\eta}, \boldsymbol{\omega}_1, \boldsymbol{\omega}_2)$ ın en iyi primal–dual çözüm olması ancak ve ancak aşağıdaki koşulların sağlanması ile mümkündür (Alizadeh ve Goldfarb 2003, Lobo ve ark. 1998):

$$
\begin{aligned}
&\boldsymbol{\chi} = \begin{pmatrix} \mathbf{0}_{n\times 1} & \boldsymbol{A} \\ 1 & \mathbf{0}_{(p+m)\times 1}^T \end{pmatrix} \begin{pmatrix} t \\ \boldsymbol{\beta}^* \end{pmatrix} + \begin{pmatrix} -\boldsymbol{y}^{**} \\ 0 \end{pmatrix}, \\
&\boldsymbol{\eta} = \begin{pmatrix} \mathbf{0}_{(p+m)\times 1} & \boldsymbol{I}_{(p+m)\times(p+m)} \\ 0 & \mathbf{0}_{(p+m)\times 1}^T \end{pmatrix} \begin{pmatrix} t \\ \boldsymbol{\beta}^* \end{pmatrix} + \begin{pmatrix} \mathbf{0}_{(p+m)\times 1} \\ \sqrt{M} \end{pmatrix}, \qquad (6.31) \\
&\begin{pmatrix} \mathbf{0}_{n\times 1}^T & 1 \\ A^T & \mathbf{0}_{(p+m)\times 1}^T \end{pmatrix} \boldsymbol{\omega}_1 + \begin{pmatrix} \mathbf{0}_{(p+m)\times 1}^T & 0 \\ \boldsymbol{I}_{(p+m)\times(p+m)} & \mathbf{0}_{(p+m)\times 1} \end{pmatrix} \boldsymbol{\omega}_2 = \begin{pmatrix} 1 \\ \mathbf{0}_{(p+m)\times 1} \end{pmatrix} \\
&\boldsymbol{\omega}_1^T \boldsymbol{\chi} = 0, \quad \boldsymbol{\omega}_2^T \boldsymbol{\eta} = 0, \\
&\boldsymbol{\omega}_1 \in L^{n+1}, \quad \boldsymbol{\omega}_2 \in L^{p+m+1}, \ \boldsymbol{\chi} \in L^{n+1}, \quad \boldsymbol{\eta} \in L^{p+m+1}.
\end{aligned}
$$

Konik karesel problemler gibi "iyi yapılandırılmış" konveks problemler, ilk kez Karmakar (1984) tarafından tanıtılan *İç Noktalar metodları* ile çözülebilir (Regener 2001, Nemirovski ve Todd 2009). Bu yöntemler, problemin bütün yapısını kullanması, en iyi zamanda en iyi çözümün elde edilmesi anlamına gelen karmaşıklığın daha iyi sınırlarını sağlaması ve çok daha iyi bir pratik performans göstermesi nedeni ile çok büyük bir avantaja sahiptir.

6.1.4. OÖSD Modelinin M-Tahmini için LASSO Yöntemi (M_L-OÖSD)

Tikhonov düzenlemesinde kullanılan, $\left\|\boldsymbol{\beta}^*\right\|_2^2$ ceza fonksiyonu kullanılabilecek küçültme yöntemlerinden sadece birisidir. Olumlu ve olumsuz özellikleri ile çeşitli ceza fonksiyonları yazılabilir. Bu şekildeki ceza fonksiyonlarının en genel formu

‘‘*Bridge*’’olarak adlandırılan $\sum_{j=1}^{p}\left|\beta_j^*\right|^{\gamma}$ dir (Frank ve Friedman 1993). Tikhonov düzenlemesinde kullanılan ceza fonksiyonu bridge’ in $\gamma = 2$ özel durumudur. Bridge ceza fonksiyonunun $\gamma = 1$ özel durumu, Tibshirani (1996) tarafından kısaca LASSO (Least Absolute Shrinkage and Selection Operator) olarak adlandırılan en küçük mutlak küçültme ve operator seçim yöntemidir. Bu ceza fonksiyonunun Tikhonov düzenlemesinde kullanılan ceza fonksiyonuna göre üstün tarafları vardır. LASSO tipi ceza kullanıldığında, bazı regresyon katsayıları tam olarak 0 elde edilir, bu durumda elde edilen modellerin daha iyi yorumlanmasını sağlar. Ridge tipi ceza kullanıldığında katsayılar küçülür ancak tam olarak 0 olarak elde edilemezler. Dolayısı ile LASSO da elde edilen modele göre daha karmaşık bir model elde edilebileceğinden, modelin yorumu da güçleşebilir. Bu nedenle OÖSD modelindeki parametrelerin tahmini için LASSO ceza fonksiyonu kullanılacaktır.

(6.5) eşitliğindeki β^* parametresinin M-tahmin edicisi olan $\hat{\beta}_M^*$, $\rho\left(r^*\right)$ fonksiyonuna ve $\sum_{i=1}^{n}\rho\left(y_i - \boldsymbol{x}_i^{*T}\boldsymbol{\beta}^*\right)$ ifadesini enküçük yapan β^* değerine dayanır.

OÖSD modelinin M- tahmini için LASSO problemi;

$$\min_{\beta^*}\left\{\sum_{i=1}^{n}\rho\left(y_i - \boldsymbol{x}_i^{*T}\boldsymbol{\beta}^*\right) + \lambda\sum_{j=0}^{p}\left|\beta_j^*\right|\right\} \tag{6.32}$$

şeklinde yazılır. Burada $\beta_{(p+m)x1}^* = \left[\beta^T \;\; \delta^T\right]^T$ olmak üzere, $\hat{\beta}^{M=lasso}$, (6.32) probleminin çözümüdür. Burada elde edilecek çözüm Huber tipi $\rho(r)$ fonksiyonuna bağlı olduğundan çözüm $\left|r_i^*\right| > k$ ve $\left|r_i^*\right| \le k$ durumlarına bağlı olarak elde edilir. $\left|r_i^*\right| \le k$ olduğu durum göz önüne alınırsa (6.32),

$$\begin{aligned}\min_{\boldsymbol{\beta}^*}\sum_{i=1}^{n}\rho\left(y_i - \boldsymbol{x}_i^{*T}\boldsymbol{\beta}^*\right) + \lambda\sum_{j=0}^{p}\left|\beta_j^*\right| &= \min_{\boldsymbol{\beta}^*}\ \tfrac{1}{2}\left\|\boldsymbol{y} - \boldsymbol{X}^*\boldsymbol{\beta}^*\right\|^2 + \lambda\sum_{j=0}^{p}\left|\beta_j^*\right| \\ &= \min_{\boldsymbol{\beta}^*} G(\boldsymbol{\beta}_i^*) + \lambda\sum_{j=0}^{p}\left|\beta_j^*\right|\end{aligned} \tag{6.33}$$

formuna dönüşür. En küçük kareler problemine LASSO ceza teriminin eklenmesi ile çözümün sağlamlığı pekiştirilir. $\left|r_i^*\right| > k$ olması durumunda ise $r_i^* \geq 0$ ve $r_i^* < 0$ durumlarına göre problem göz önüne alınır. $r_i^* \geq 0$ olduğunda problem,

$$\begin{aligned}\hat{\beta}^{*M=lasso} &= \min_{\beta^*} \sum_{i=1}^{n} \sqrt{w_i}^2 (y_i^* - \boldsymbol{x}_i^{*T} \beta_i^*)^2 + \lambda \sum_{j=0}^{p} \left|\beta_j^*\right| \\ &= \min_{\beta^*} \left\| \boldsymbol{y}^{**} - \boldsymbol{A}\boldsymbol{\beta}^* \right\|^2 + \lambda \sum_{j=0}^{p} \left|\beta_j^*\right| \\ &= \min_{\beta^*} G(\boldsymbol{\beta}_i^*) + \lambda \sum_{j=0}^{p} \left|\beta_j^*\right| \\ &= \min_{\beta^*} G(\boldsymbol{\beta}_i^*) + \lambda \left\| \boldsymbol{\beta}_i^* \right\|_1\end{aligned} \tag{6.34}$$

olarak yazılır. Burada, $\left\| \boldsymbol{\beta}^* \right\|_1 := \sum_{j=0}^{p} \left|\boldsymbol{\beta}_j^*\right|$, y_i^* ve w_i, (6.22) daki gözlem ve ağırlık değerleri, $\left\| \boldsymbol{\beta}^* \right\|_1$, $\boldsymbol{\beta}^*$ ın L_1 normu, $\lambda > 0$ ise düzenleme parametresidir. Eğer (6.34) probleminde $\left|\beta_j^*\right| = \alpha_j^2$ olarak alınırsa problem,

$$\begin{aligned}\hat{\beta}^{*M=lasso} &= \min_{\beta^*} G(\boldsymbol{\beta}_i^*) + \lambda \left\| \boldsymbol{\beta}_i^* \right\|_1 \\ &= \min_{\beta^*} G(\boldsymbol{\beta}_i^*) + \lambda \left\| \boldsymbol{\alpha} \right\|_2^2 \\ &= \min_{\beta^*} \left\| \boldsymbol{y}^{**} - \boldsymbol{A}\boldsymbol{\beta}^* \right\|^2 + \lambda \left\| \boldsymbol{\alpha} \right\|_2^2\end{aligned} \tag{6.35}$$

olarak elde edilir. Bu en iyileme problemi de karesel programlamalardan biri olan konik karesel programlama ile çözülebilir (Taylan ve ark. 2014). Bunun için (6.35) problemi,

$$\begin{aligned}&\min \quad \left\| \boldsymbol{A}\boldsymbol{\beta}^* - \boldsymbol{y}^{**} \right\|_2^2 \\ &\text{k.s..} \left\| \boldsymbol{\alpha} \right\|_2^2 \leq M_{\alpha.}\end{aligned}$$

şeklinde veya eşdeğer olarak,

$$\begin{aligned}&\min_{t,\boldsymbol{\beta}^*} \quad t \\ &\text{k.s.} \quad \left\| \boldsymbol{A}\boldsymbol{\beta}^* - \boldsymbol{y}^{**} \right\|_2^2 \leq t^2, \quad t \geq 0, \\ &\qquad\qquad \left\| \boldsymbol{\alpha} \right\|_2^2 \leq M_{\alpha},\end{aligned} \tag{6.36}$$

olarak yazılabilir. Burada, M_α parametresi, $\left|\beta_j^*\right| = \alpha_j^2$ değerlerine bağlı olarak, deneme ve yanılma yolu ile seçilecektir. Bu problem konik karesel programlamanın standart formu ile karşılaştırıldığında,

$$\boldsymbol{c} = \left(1 \mid \mathbf{0}_{p+m}^T \middle| \mathbf{0}_{p+m}^T\right)^T, \boldsymbol{u} = (t, \boldsymbol{\beta}^{*T}, \boldsymbol{\alpha}^T)^T, \ \boldsymbol{D}_1 = \left(\mathbf{0}_{n\times 1} \mid A_{n\times(p+m)} \middle| \mathbf{0}_{n\times(p+m)}\right), \boldsymbol{d}_1 = \boldsymbol{y}^{**},$$

$$\boldsymbol{p}_1 = (1,0,...,0)_{(2p+2m+1\times 1)}^T, \ q_1 = 0, \boldsymbol{D}_2 = \left(\mathbf{0}_{(p+m)\times 1} \middle| \mathbf{0}_{(p+m)\times(p+m)} \middle| \boldsymbol{I}_{(p+m)\times(p+m)}\right), \boldsymbol{d}_2 = 0_{(2p+2m+1)\times 1}$$

$$\boldsymbol{p}_2 = \mathbf{0}_{(2p+2m+1)\times 1} \text{ ve } q_2 = -\sqrt{M_\alpha}$$

özellikleri ile bir konik karesel programlama olduğu görülür. Tikhonov düzenlemesi ve LASSO için yazılan konik karesel program problemlerin karşılaştırdıgımızda, Tikhonov düzenlemesindeki ikinci kısıt denklemi $\boldsymbol{\beta}^*$ a bağlı iken LASSO daki ikinci kısıt denklemi *i*. bileşeni $\left|\beta_j^*\right| = \alpha_j^2$ olan $\boldsymbol{\alpha}$ vektörüne bağlıdır. Bu durum problemin en iyilik durumunu da etkileyeceği gibi çözümü de zorlaştırabilir. Bu problem için en iyilik koşularını oluşturmak için problem yeniden,

$$\begin{aligned} &\min_{t,\boldsymbol{\beta}^*} \quad t \\ &k.s. \quad \boldsymbol{\chi} = \begin{pmatrix} \mathbf{0}_{n\times 1} & \boldsymbol{A}_{n\times(p+m)} & \mathbf{0}_{n\times(p+m)} \\ 1 & \mathbf{0}_{1\times(p+m)}^T & \mathbf{0}_{1\times(p+m)}^T \end{pmatrix} \begin{pmatrix} t \\ \boldsymbol{\beta}^* \\ \boldsymbol{\alpha} \end{pmatrix} + \begin{pmatrix} -\boldsymbol{y}^{**} \\ 0 \end{pmatrix}, \\ &\qquad \boldsymbol{\eta} = \begin{pmatrix} \mathbf{0}_{(p+m1)\times 1} & \mathbf{0}_{(p+m)\times(p+m)} & \boldsymbol{I}_{(p+m)\times(p+m)} \\ \mathbf{0}_{1\times 1} & \mathbf{0}_{1\times(p+m)}^T & \mathbf{0}_{1\times(p+m)}^T \end{pmatrix} \begin{pmatrix} t \\ \boldsymbol{\beta}^* \\ \boldsymbol{\alpha} \end{pmatrix} + \begin{pmatrix} \mathbf{0}_{(p+m)\times 1} \\ \sqrt{M} \end{pmatrix}, \\ &\qquad\qquad \boldsymbol{\chi} \in L^{n+1}, \ \boldsymbol{\eta} \in L^{p+m+1}, \end{aligned} \tag{6.37}$$

olarak düzenlenir. Burada L^{n+1}, L^{p+m+1} sırasıyla $(N+1)$ ve $(p+m+1)$-boyutlu *ikinci dereceden veya Lorents* konileridir. (6.37) teki primal probleme karşılık gelen *dual problem*,

$$\text{maks} \quad \begin{pmatrix} \boldsymbol{y}^{**} \\ 0 \end{pmatrix}^T \boldsymbol{\omega}_1 + \begin{pmatrix} \mathbf{0}_{(p+m)\times 1} \\ -\sqrt{M_\alpha} \end{pmatrix} \boldsymbol{\omega}_2$$

$$\text{k.s.} \quad \begin{pmatrix} \mathbf{0}_{1\times n}^T & 1 \\ \boldsymbol{A}_{(p+m)\times n}^{\boldsymbol{T}} & \mathbf{0}_{(p+m)\times 1}^T \\ \mathbf{0}_{(p+m)\times n}^T & \mathbf{0}_{(p+m)\times 1}^T \end{pmatrix} \boldsymbol{\omega}_1 + \begin{pmatrix} \mathbf{0}_{1\times(p+m)}^T & 0 \\ \mathbf{0}_{(p+m)\times(p+m)} & \mathbf{0}_{(p+m)\times 1}^T \\ \boldsymbol{I}_{(p+m)\times(p+m)} & \mathbf{0}_{(p+m)\times 1}^T \end{pmatrix} \boldsymbol{\omega}_2 = \begin{pmatrix} 1 \\ \mathbf{0}_{(p+m)\times 1} \\ \mathbf{0}_{(p+m)\times 1} \end{pmatrix}, \qquad (4.26)$$

$$\boldsymbol{\omega}_1 \in L^{n+1},\ \boldsymbol{\omega}_2 \in L^{p+m+1}.$$

olarak elde edilir.

OÖSD modeli için $\hat{\beta}^{*M=lasso}$ tahminini konik karesel programlama ile çözdüğümüzden dolayı, $(t, \boldsymbol{\beta}^*, \boldsymbol{\alpha}, \boldsymbol{\chi}, \boldsymbol{\eta}, \boldsymbol{\omega}_1, \boldsymbol{\omega}_2)$ın en iyi primal–dual çözümünü elde etmek için konik karesel programlamanın en iyilik koşullarının sağlanması gerekir. Bu da aşağıdaki koşulların sağlanması ile mümkündür:

$$\boldsymbol{\chi} = \begin{pmatrix} \mathbf{0}_{n\times 1} & \boldsymbol{A}_{n\times(p+m)} & \mathbf{0}_{n\times(p+m)} \\ 1 & \mathbf{0}_{1\times(p+m)}^T & \mathbf{0}_{1\times(p+m)}^T \end{pmatrix} \begin{pmatrix} t \\ \boldsymbol{\beta}^* \\ \boldsymbol{\alpha} \end{pmatrix} + \begin{pmatrix} -\boldsymbol{y}^{**} \\ 0 \end{pmatrix},$$

$$\boldsymbol{\eta} = \begin{pmatrix} \mathbf{0}_{(p+m1)\times 1} & \mathbf{0}_{(p+m)\times(p+m)} & \boldsymbol{I}_{(p+m)\times(p+m)} \\ \mathbf{0}_{1\times 1} & \mathbf{0}_{1\times(p+m)}^T & \mathbf{0}_{1\times(p+m)}^T \end{pmatrix} \begin{pmatrix} t \\ \boldsymbol{\beta}^* \\ \boldsymbol{\alpha} \end{pmatrix} + \begin{pmatrix} \mathbf{0}_{(p+m)\times 1} \\ \sqrt{M} \end{pmatrix},$$

$$\begin{pmatrix} \mathbf{0}_{1\times n}^T & 1 \\ \boldsymbol{A}_{(p+m)\times n}^{\boldsymbol{T}} & \mathbf{0}_{(p+m)\times 1}^T \\ \mathbf{0}_{(p+m)\times n}^T & \mathbf{0}_{(p+m)\times 1}^T \end{pmatrix} \boldsymbol{\omega}_1 + \begin{pmatrix} \mathbf{0}_{1\times(p+m)}^T & 0 \\ \mathbf{0}_{(p+m)\times(p+m)} & \mathbf{0}_{(p+m)\times 1}^T \\ \boldsymbol{I}_{(p+m)\times(p+m)} & \mathbf{0}_{(p+m)\times 1}^T \end{pmatrix} \boldsymbol{\omega}_2 = \begin{pmatrix} 1 \\ \mathbf{0}_{(p+m)\times 1} \\ \mathbf{0}_{(p+m)\times 1} \end{pmatrix}$$

$$\boldsymbol{\omega}_1^T \boldsymbol{\chi} = 0, \quad \boldsymbol{\omega}_2^T \boldsymbol{\eta} = 0,$$

$$\boldsymbol{\omega}_1 \in L^{n+1}, \quad \boldsymbol{\omega}_2 \in L^{p+m+1}, \ \boldsymbol{\chi} \in L^{n+1}, \quad \boldsymbol{\eta} \in L^{p+m+1}.$$

Bu problem de iç noktalar yöntemi ile çözülebilir.

7. SAYISAL UYGULAMA VE LM, OÖSD VE CT-OÖSD NİN PERFORMANSLARININ KARŞILAŞTIRILMASI

7.1. Veri Kümesi

Yukarıda teorik olarak açıklanan çalışmanın sayısal uygulaması için kullanılacak veri amonyağı nitrik aside oksitleyen bir düzeneğin 21 günlük çalışma sistemini göstermektedir (Daniel ve Wood 1999:60-61). Burada X_1 açıklayıcı değişkeni düzeneğin hava akımını, X_2 açıklayıcı değişkeni nitrik oksit emilim kulesine ters akıntıyla giren soğutma suyunun sıcaklığı, X_3 açıklayıcı değişkeni ise emilen sıvıdaki nitrik asit konsantrasyonudur. Bu açıklayıcı değişkenin değerlerinin hesaplarda kolay kullanılması için esas değerlerden 50 çıkarılarak ve daha sonra 10 ile çarpılarak kodlama yapılmıştır. Burada y, yanıt değişkeni de emilmeyen nitrik oksit nedeniyle içeriye giren kayıp amonyak yüzdesinin 10 katıdır. Bu değer nitrik oksit veriminin dolaylı bir ölçümüdür. Bu veriler aşağıdaki Çizelge 7.1. de verilmiştir:

Çizelge 7.1. Amonyağı nitrik aside oksitleyen bir düzeneğinin 21 günlük çalışma sistemi için elde edilen veriler (Daniel ve Wood 1999)

Gözlem No	HavaAkımı X_1	Soğutma Suyu Giriş Sıcaklığı X_2	Asit Konsantrasyonu X_3	Kütle Kaybı y
1	80	27	89	42**(4.)**
2	80	27	88	37
3	75	25	90	37**(3.)**
4	62	24	87	28**(2.)**
5	62	22	87	18
6	62	23	87	18
7	62	24	93	19
8	62	24	93	20
9	58	23	87	15
10	58	18	89	14
11	58	18	89	14
12	58	17	88	13
13	58	18	82	11
14	58	19	93	12
15	50	18	89	8
16	50	18	86	7
17	50	19	72	8
18	50	19	79	8
19	50	20	80	9
20	56	20	82	15
21	70	20	91	15**(1.)**

7.2. Karşılaştırmada Kullanılan Performans Ölçütleri

Bu çalışmada yapılan uygulamanın amacı oluşturulan LM, OÖSD ve CT-OÖSD modelleri arasından veri kümesindeki sapan değer bakımından en iyisini belirlemektir. Bu nedenle, LM, OÖSD ve CT-OÖSD modellerine uygulanan performans ölçütleri karşılaştırılır. Çalışmada kullanılan performans ölçütleri genel olarak kullanılan ifadeleri ile aşağıda verilmektedir:

Ortalama Mutlak Hata (OMH): Ortalama mutlak hata ölçütü, hatanın büyüklüğünün ortalamasıdır (Willmott ve Matsuura 2005). Bu ölçüt,

$$\text{OMH} = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \hat{y}_i\right|$$

olarak tanımlanır. Burada, y_i , i. gözlem değeri için yanıtın gerçek değeri, $\hat{y}_i$ ise tahmin edilen değeridir. Bu ölçütün küçük değerleri daha iyi modeller verir.

Hata Karelerinin Ortalamasının Karekökü (HKOK): HKOK ölçütü hata karelerinin ortalamalarının kareköküdür ve

$$\text{HKOK}=\sqrt{\frac{1}{n-p}\sum_{i=1}^{n}(y_i-\hat{y}_i)^2},$$

olarak tanımlanır (Moisen ve Frescino 2002). Bu ölçüt hata varyansının bir yansız tahmin edicisini gösterir ve bu ölçüt hem artıklara hem de modelde kullanılan açıklayıcı değişkenlerin sayısına bağlıdır. Bu ölçütün küçük değerleri de OMH gibi daha iyi modeller verir.

Korelasyon Katsayısı: Korelasyon katsayısı iki değişken arasındaki ortaklık ölçüsüdür ve -1 ile +1 arasında değerler alır. Eğer iki değişken arasında kusursuz bir doğrusal ilişki varsa korelasyon katsayısı ya -1 ya da +1 olacaktır. İşareti, değişkenler arasındaki doğrusal ilişkinin yönüne bağlıdır. Eğer korelasyon katsayısı 0 olursa bu iki değişken arasında doğrusal bir ilişki yok demektir, yani, ilişkisizdir. Korelasyon katsayısı aşağıda verilen şekilde hesaplanır (Chung 2007):

$$r=\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2\sum_{i=1}^{n}(y_i-\bar{y})^2}}.$$

Burada $\bar{x}$ ve $\bar{y}$ örneklem ortalamalarını göstermektedir.

Çoklu Belirleme Katsayısı (R^2): R^2 belirleme katsayısıdır. R^2 değerleri artarken, model için daha iyi bir uygunluk elde edilir. Bu ölçütün matematiksel eşitliği aşağıdaki şekildedir (Rao 1999:55):

$$R^2=1-\frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{\sum_{i=1}^{n}(y_i-\bar{y}_i)^2}.$$

Düzeltilmiş R^2 (*Düz-R^2*): Bu değer modeldeki belirleyicilerin sayısını hesaplamak için kullanılır. Bu ölçüt farklı sayıdaki bağımsız değişkenleri içeren modellerin

karşılaştırılması için faydalıdır. Düz-R^2 ölçütünün büyük değerleri veriye uydurulan modelin daha iyi olduğu anlamına gelir. Bu ölçüt için matematiksel eşitlik aşağıdaki şekildedir (Neter ve ark. 1996):

$$\text{Düz-}R^2 = 1 - \left(\frac{\sum_{i=1}^{N} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{N} (y_i - \bar{y}_i)^2} \right) \left(\frac{N-1}{N-p-1} \right),$$

Burada $N - p - 1 \neq 0$ Burada p regresyon eşitliğindeki parameter sayıs, N de kullanılan gözlem sayısıdır.

7.3. Sapan Değer Teşhisi ve Modelin Oluşturulması

Sapan değerlerin teşhisinden önce, bu uygulamanın ilk aşaması olarak, sapan değerlerin model üzerine etkisini görebilmek için sapan değerleri de içeren veri kümesinin ve en küçük kareler yönteminin kullanılması ile (2.7) denklemi ile verilen uydurulmuş doğrusal regresyon model belirlendi. Bu model,

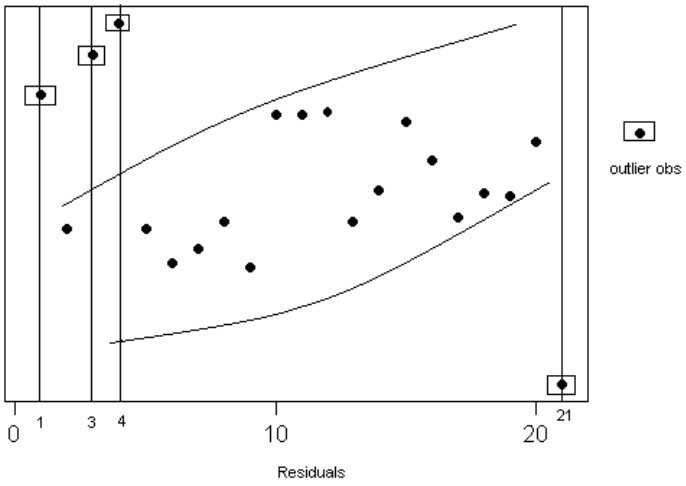
$$y = -42{,}1 + 0{,}712\, X_1 + 1{,}26\, X_2 - 0{,}116\, X_3 \qquad \text{(EKK)}$$

şeklindedir.

Uygulamamızın ikinci adımı olarak, sapan değerleri belirlemek için direkt olan sapan değer yöntemlerinden Cook uzaklığı (Cook 1979) ve Studentized artıklar yöntemleri kullanıldı. Bunun sonucunda, 21. gözlem değeri için Studentized kalan değerinin standartlaştırılmış kalan değerinden önemli ölçüde büyük olduğu, ayrıca, bu gözlem değeri için ölçülen Cook uzaklığının da en büyük değere sahip olduğu gözlendi. Bu durum da 21. gözlem değerinin gözlem grubundan silinmesinin tahmin hatasının küçülteceğini ve iyi bir tahmin sağlayacağını gösterir. 21. gözlem değerinin silinmesinden sonra tekrar bir doğrusal model oluşturuldu. Oluşturulan bu model, bir önceki doğrusal model ile karşılaştırıldığında ikinci modelin performansının, kullanılan tüm performans ölçümlerine göre, daha iyi olduğu saptandı. Bununla beraber, 21. gözlem değerinin sapan değer olarak göz önüne alındığı OÖSD modeli oluşturularak, 21. gözlem değerinin silinmediği durumda oluşturulan doğrusal model ile karşılaştırıldığında, OÖSD modelinin daha daha iyi performansa sahip olduğu görüldü. Sonraki en büyük basit, studentized ve standartlaştırılmış kalan değeri ve Cook uzaklığı

4.gözlem değerinde elde edildi. Bu teşhisten sonra yeniden doğrusal model ve OÖSD model oluşturularak önceki doğrusal model ve OÖSD modeli karşılaştırıldı. Bu karşılaştırma, 4.gözlem değerinin sapan değer olarak göz önüne alınması sonucu elde edilen LM ve OÖSD modellerinin performanslarının daha iyi olduğunu ve daha iyi tahminler ürettiğini gösterdi. Aynı işlem, sırası ile 3. ve 1. gözlem değerleri için yapıldı.

Uygulamada kullanılan veri kümesi içinde, bu yöntemler sonucu bulunan sapan değerler çizelge 7.1. de gölgeli olarak vurgulanmıştır. Bunların grafiksel gösterimi Şekil 7.1.de gösterilmiştir.



Şekil 7.1. 1, 3, 4, 21 nolu gözlem değerleri sapan değerdir

Son olarak, sırası ile 21. , 4., 3. ve 1. gözlem değerlerinin aynı anda sapan değer olarak göz önüne alınarak, bu gözlem değerlerine sırası ile δ_1, δ_2, δ_3, δ_4 artmaları verilerek, OÖSD modeli,

$$y = -38.3 + 0.800\,X_1 + 0.554\,X_2 - 0.0553\,X_3 + 6.24\,\delta_1 + 6.40\,\delta_2 + 8.20\,\delta_3 - 8.77\,\delta_4 \qquad \text{(OÖSD)}$$

olarak elde edildi. Oluşturulan bu modelin önceki modellerle karşılaştırılması sonucu, bütün performans ölçütlerine göre en iyi performansa sahip olduğu görüldü.

Uygulamanın 3. adımında yukarıda oluşturulan OÖSD modeli için uydurulmuş modelin, Huber tipi fonksiyonun kullanılması ile M-tahmin edicisi bulundu. Burada Huber tipi M-tahmin edicisini elde eltmek için, OÖSD modeline ilişkin, (6.22) deki r_i^*

kalanlarının mutlak değerleri ve w_i ağırlıkları hesaplandı. Hesaplanan r_i^* değerlerinin mutlak değerleri

$$k = 1.345 \frac{\text{Med}|\boldsymbol{r}^*|}{0.6745}$$

degeri ile karşılaştırıldı. $|r_i^*| > k$ durumu için w_i ağırlıkları (6.22) de verilen eşitlik ile hesaplandı, $|r_i^*| \le k$ olanlar için ise $w_i = 1$ olarak hesaplandı. $|r_i^*| > k$ durumu için köşegen elemanları $\sqrt{w_i}$ olan, ağırlık matrisi $\boldsymbol{W} = \left(diag\sqrt{w_i}\right)$ hesaplandı. Bu matrisin hem OÖSD modelindeki yanıt vektörü $\boldsymbol{y}^*$ hem de açıklayıcı değişken matrisi $\boldsymbol{X}^*$ ile çarpılması sonucu yeni yanıt vektörü $\boldsymbol{y}^{**}$ ve yeni açıklayıcı değişken matrisi $\boldsymbol{A}$ elde edilerek (6.24) teki Tikhonov düzenlemesi problemi yazıldı. Bu problemin çözümü konik karesel en iyileme problemi olarak çözüldü. Bu problemin çözümü için MOSEK paket programı kullanıldı. Bu çözüm sonucu uydurulmuş model,

$$\begin{aligned} y = {} & -36.8679 + 0.8028\, X_1 + 0.5452\, X_2 - 0.0715\, X_3 + \\ & 6.2811\, \delta_1 + 6.4581\, \delta_2 + 8.2228\, \delta_3 - 8.7159\, \delta_4 \end{aligned} \qquad \text{(CT-OÖSD)}$$

olarak elde edildi.

7.4. LM, OÖSD ve CT-OÖSD Modellerinin Performanslarının Sayısal Karşılaştırılması

CT-OÖSD modeli oluşturulurken *k*-sabitinin değişik değerleri kullanıldı ve model performansları karşılaştırıldı. En iyi performansı veren model, k= 1.5 değeri için yani $\text{Med}|\boldsymbol{r}^*|/0.6745 = 1.12$ değeri için elde edildi. Performans karşılaştırmaları MATLAB paket programı kullanılarak yapıldı. LM, OÖSD ve CT-OÖSD modellerinin oluşturulmasından sonra, bu modellerin performanslarının sayısal karşılaştırılması için bölüm 7.2. de tanımlanan performans kriterleri hesaplandı. Elde edilen sonuçlar çizelge 7.2. de verilmektedir.

Çizelge 7.2. EKK, OÖSD ve CT-OÖSD alternatif modelleri için performans ölçümü

İstatistik Değerleri	LM için Değer	OÖSD için Değer	CT-OÖSD için değer
Hata Karelerinin Ortalamasının Karekökü	3.2858	1.2905	1.2776
Korelasyon katsayısı	0.9547	0.9948	0.9949
R^2	0.9114	0.9895	0.9898
Düzeltilmiş R^2	0.8957	0.9839	0.9978

Çizelge 7.2. de de görülebileceği gibi, X_1, X_2, X_3 değişkenlerini içeren ve (1,3,4,21) nolu gözlem değerlerinin sapan değer olduğu göz önüne alınarak uydurulan OÖSD ve CT-OÖSD modelleri bütün performans kriterlerine göre LM'e göre daha iyi sonuçlar vermektedir. Bununla beraber, OÖSD ve CT-OÖSD modelleri karşılaştırıldığında, bütün performans kriterlerine gore CT-OÖSD modeli OÖSD modeline göre daha iyi sonuçlara sahiptir. Bunun nedeni, X_1, X_2, X_3 değişkenlerini içeren ve (1,3,4,21) nolu gözlem değerlerinin sapan değer olduğu göz önüne alınarak uydurulan OÖSD modeline Tikhonov düzenlemesi uygulanarak model için gerekli olmayan özelliklerin ortadan kaldırılması ve önemli bir konveks programlama olan konik karesel programlamanın iyi yapılandırılmış konveks problem olması, amaç ve kısıt fonksiyonun tüm yapısını kullanması, pratikte çok iyi karmaşıklık sınırları ve iyi sonuçlar vermesidir.

7.5. Tartışma ve Sonuç

Veri kümesinde bir veya daha fazla sapan değer olması durumunda en küçük kareler yöntemi zayıf kalmaktadır. OÖSD modelinin Huber tipi M-tahmin edicisinin Tikhonov düzenlemesi ve LASSO ile elde edilmesi OÖSD modelinin oluşturulmasında alternatif ve etkili bir yöntem olmuştur. Ayrıca bu çalışma göstermiştir ki, modern conveks eniyileme yöntemleri ile kötü koşullu istatistiksel problemleri iyi koşullu problemlere dönüştüren Tikhonov düzenlemesi ve LASSO istatistiksel problemleri arasında güçlü ve faydalı ilişkiler kurulabilir.

8. KAYNAKLAR

KİTAP

Aster, R., Borchers, B. and Thurber C. 2005. *Parameter Estimation and Inverse Problems*. Oxford: Elsevier Inc. Academic Press.

Barnett V. and Lewis T. 1984. *Outliers in Statistical Data*. (2nd Edition). New York: John Wiley.

Belsley, D. A., Kuh, E., and Welsch, R. E. 1980. *Regression Diagnostics.* New Jersey: Wiley&Sons, Inc.

Chatterjee, S. and Hadi, A. S. 1988. *Sensitivity analysis in linear regression*. New York: John Wiley & Sons, pp. xiv + 315.

Chatterjee, S., Hadi, A.S. and Price, B. 2000. *Regression Analysis by Example* (3rd Edition). New York: Wiley. p 104.

Cook, R. D. and Weisberg, S. 1982. *Residuals and Influence in Regression*. New York: Chapman and Hall Ltd.

Daniel, C. and Wood, F. S. 1999. *Fitting Equations to Data: Computer Analysis of Multifactor Data*. (2nd Edition). Toronto: John Wiley&Sons, INC.

Hastie, T., Tibshirani, R. and Friedman, J.H. 2001. *The Elements of Statistical Learning*. New York: Springer-Verlag.

Hawkins, D. 1980. *Identification of Outliers*, Chapman and Hall.

Huber, P. J. 1981. Robust Statistics. New York: Wiley&Sons.

Johnson, R. and Wichern, D. W. 1992. *Applied Multivariate Statistical Analysis*. (6th Edition) New Jersey: Prentice Hall.

Jongen, H. Th., Jonker P. and Twilt, F. 1986. Nonlinear Optimization in $\mathbb{R}^n$: II. Transversality, Flows, Parametric Aspects. Frankfurt am Main., Bern, New York: *Peter Lang Verlag.*

Nesterov, Y. and Nemirovski, A. 1993. *Interior Point Methods in Convex Programming,* SIAM.

Nesterov, Y. and Nemirovski A. 1994. *Interior-Point Polynomial Algorithms in Convex Programming*. SIAM s.67-80.

Neter, J., Kutner, M., Wasserman, W. and Nachtsheim, C. 1996. *Applied Linear Statistical Models*. Boston: WCB/McGrawHill.

Rao, C. R. and Toutenburg, H. 1999. *Linear Models Least Squares And Alternatives*. New York, Inc.: Springer-Verlag, (2nd Edition) s.232.

Rencher, A.C. 2000, *Linear models in Statistics,*. New York: John Wiley and Sons, Inc..

Rencher, A. C. and Schaalje G. B. 2008. *Lineer Models in Statistics*. New Jersey: John Wiley&Sons, Inc.

Rey, W. J. J. 1983. *Introduction to Robust and Quasi-Robust Statistical Methods*. Berlin, Heidelberg: Springer-Verlag.

Rousseeuw, P. J. And Leroy, A. M. 1987. *Robust Regression and Outlier Detection*. New York: John Wiley&Sons Inc. p 329.

Tikhonov, A.N. and Arsenin ,V.Y. 1977. *Solution of III-Posed Problems*. Washington, DC: V. H. Winston.

DERGİ

Alizadeh, F. and Goldfarb, D. 2003. Second-order cone programming. *Mathematical Programming*, Vol. 95, Issue 1, Ser. B, p. 3-51.

Alpu, Ö., Şamkar, H. ve Altan, E. 2010. Sağlam Ridge Regresyon Analizi ve Bir Uygulama, *İzmir, Dokuz Eylül Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, Cilt:25, Sayı:2, ss.137-148.

Atkinson, A. C. 1986. Masking Unmasked. Biometrika, 73(3):533-541.

Atkinson, A.C. 1983. Diagnostics, regression analysis and shifted power transformation. *Technometrics*, 25, 23-34.

Adnan, R. and Mohamad M. N. 2003. Multiple Outliers Detection Procedures in Linear Regression. *Matematika*, Jilid 19, No. 1, pages. 29–45.

Beck A. and Teboulle M. 2009. A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems. *SIAM J. Imaging Sciences,* Vol. 2, No. 1, pp. 183–202.

Beck, A., Teboulle M., A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems, *SIAM J. Imaging Sciences* Vol. 2, No. 1, pp. 183–202.

Beckman, R.J. and Trussell, H.J. 1974. The distribution of an arbitrary Studentized residual and effects of updating in multiple regression, *J. Amer. Statistic. Assoc.*, 69, 199-201.

Behnken, D.W and Draper, N.R. 1972. Residuals and their variance patterns. *Technometrics*, 14, 101-111.

Ben-Tal, A. and Nemirovski, A. 2001. Lectures on Modern Convex Optimization: Analysis, Algorithms and Engineering Applications. Philadelphia: *MPS-SIAM Series on Optimization*, SIAM.

Chung, M. 2007. Correlation coefficient. In N. Salkind (Ed.), Encyclopedia of measurement and statistics. (pp. 190-192). Thousand Oaks, CA: SAGE Publications, Inc.

Cook, R. D. 1977. Detection of Influental Observations in Linear Regression. *Technometrics*, 19, 15-18.

Cook, R. D. 1979. Influential observations in linear regression, *J. Amer. Statistic. Assoc.*, 74, 169-174.

Frank, I. E. and Friedman, J. H. 1993. AQ statistical view of chemometrics tools. *Technometrics*, 35, 109-148.

Fung, W. K. 1993. Unmasking outliers and leverage points: A confirmation. *Journal of the American Statistical Association*, 88(422):515-519.

Garrigues, P. J. and Ghaoui L.E. 2008. An homotopy Algorithm for the Lasso with Online Observations. Status: In *Proc. NIPS*.

Golub, G. H., Heath, M. and Wahba, G. 1979. Generalized Cross-Validation As A Method For Choosing A Good Ridge Parameter. *Technometrics*, Vol. 21, No. 2, May.

Hadi, A. S. 1992. Identifying multiple outliers in multivariate data. *Journal of Royal Statistical Society*, Series B, 54(3):761-771.

Hadi, A. S. and Simonoff, J. S. 1993. Procedures for the Identification of Multiple Outlier in Linear Models. *Journal of the American Statistical Association*, 88, Issue 424, 1264-1272.

Hampel, F. R. 1975. Beyond Location Parameters: Robust Concepts and Methods, *Bulletin of the International Statistical Institute*, 46, 375- 382.

Hansen, P. C. 1992. Analysis of discrete ill-posed problems by means of the L-curve. *SIAM Review*, 34(4):561–580.

Hansen, P. C. 1994. Regularization tools: A MATLAB package for analysis and solution of discrete ill-posed problems. *Numerical Algorithms*, 6(I–II):1–35.

Huber, P. J. 1964. Robust Estimation of a Location Parameter. *Annals of Mathematical Statistics*, 35:73— 101.

Huber, P.J. 1973. Robust regression: Asymptotics, conjectures and Monte Carlo, *Ann. Stat.*, 1, 799-821.

Kianifard, F. and Swallow W. H. 1989. Using Recursive Residuals, Calculated an Adaptively Ordered Observations to Identify Outliers in Linear Regression. *Biometrics*, 45, 571-585.

Kianifard, F. and Swallow, W. H. 1990. A Monte Carlo Comparison of Five Procesures for Identifying Outliers in Linear Regression, Communications in Statistics. *Part A-Theory and Methods*, 19, 1913-1938.

Kima, S.S., Parkb, S.H. and Krzanowskic, W.J. 1974. Simultaneous variable selection and outlier identification in linear regression using the mean-shift outlier model, *J. Appl. Stat.*, 35, 3, 283–291.

Krasker, W. S. and Welsch, R. E. 1982. Efficient Bounded-Influence Regression Estimation. *Journal of the American Statistical Association*, 77, 595-604.

Liu, H., Shah, S. and Jiang, W. 2004. On-line Outlier Detection and Data Cleaning. *Computers and Chemical Engineering* 28, 1635-1647.

Lobo, M.S., Vandenberghe, L., Boyd, S. and Lebret, H. 1998. Applications of second-order cone programming, *Linear Algebra Appl.*, 284, 1-3 p. 193-228.

McDonald, B. 2002. A Teaching Note on Cook's Distance – A Guideline. *Res. Lett. Inf. Math. Sci*., 3, 127-128.

Marasinghe, M. G. 1985. A Multistage Procedure for Detecting Several Outliers in Linear Regression. *Technometrics*, 27, 395-399.

Moisen, G.G. and Frescino, T.S. 2002. Comparing ve modelling techniques for predicting forest characteristics, *Ecological Modelling*, pp. 209-225.

Nemirovski, A.S. and Todd, M.J. 2009. Interior Point Methods for Optimization. *Acta Numerica* pp1001-DOI:10.1017/S0962492904.

Osborne, M. R., Presnell, B. and Turlach, B. A. 2000. *On* the LASSO and its Dual, *Journal of Computational and Graphical Statistics*, Vol. 9, No. 2 pp. 319-337.

Paul S. and Fung K. Y. 1991. A Generalized Extreme Studentized Residual Multple-Outlier-Detection Procedure in Linear Regression. *Technometrics*, 33, 339-348.

Prescott, P. 1975. An Approximation Test for Outliers in Linear Models, *Technometrics*, 17, 129-132.

Renegar, J. 2001. A Mathematical View of Interior-Point Methods in Convex Programming. *PS-SIAM Ser. Optim.*, 3, Philadelphia :SIAM.

Rousseeuw P. J. 1984. Least Median of Squares Regression. *Journal of the American Statistical Association*, Volume 79, Issue 338, Pages 871-880.

Rousseeuw, P. J., and Yohai, V. J. 1984. "Robust Regression by Means of S-Estimators," in Robust and Nonlinear Time Series, eds. J. Franke, W. Hardle, and D. Martin, Lecture Notes in Statistics, 26, Berlin: Springer-Verlag, pp. 256–272.

Rousseeuw,P.J. and Van Zomeren, B. C. 1990. Unmasking Multivariate Outliers and Leverage Points. *Journal of the American Statistical Association* 85, 633-639.

Rousseeuw, P.J. and Van Aelst S. 1999. Positive-breakdown robust methods in computer vision. *Computing Science and Statistics*, 31, 451-460.

Salibian-Barrera, M. and Yohai, V. J. 2006. A Fast Algoritm for S-Regression Estimates. *Journal of Computational and Graphical Statistics*, 15, No. 2, 414-427.

Simpson, D. G., Ruppert, D. and Carroll, R. J. 1992. On One-Step GM Estimation and Stability of Inferences in Linear Regression. *Journal of the American Statistical Association*, 87, 439-450.

Simpson, J. R. ve Montgomery, D.C. 1998. A Robust Regression Technique Using Compound Estimation. *Naval Research Logistics*, 45 Issue 2, 125-139.

Susanti, Y. 2008. Estimasi M dan sifat-sifatnya pada Regresi Linear Robust, *Jurnal Math-Info*, 1, No. 11, 8-16.

Susanti, Y., Pratiwi, H., Sulistijowati, S. H. and Liana, T. 2014. M Estimation, S Estimation, And MM Estimation In Robust Regression. *International Journal of Pure and Applied Mathematics*, Volume 91 No. 3, 349-360.

Taylan, P., Yerlikaya-Özkurt, F. and Weber G.W. 2014. An Approach to Mean Shift Outlier Model by Tikhonov Regularization and Conic Programming. *Intelligent Data Analysis* 18 79–94.

Taylan, P., Weber, G.-W. and Beck, A. 2007. New approaches to regression by generalized additive models and continuous optimization for modern applications in finance, science and technology. *Optimization* 56, 5-6 :1-24.

Taylan, P., Weber, G. W. and Yerlikaya, F. 2010. A new approach to multivariate adaptive regression spline by using Tikhonov regularization and continuous optimization. TOP *Journal of the Spanish Society of Statistics and Operations Research*, 18, 2, 377- 395.

Taylan, P., Yerlikaya, F. and Weber, G. W. 2010. An approach to the mean shift outlier model by Tikhonov regularization and conic programming. *Intelligent Data Analysis* 18 (2014) 79–94.

Tibshirani, R. 1996. Regression Shrikage and selection via the laso, *Journal of Royal Statistical Society*, Series B, (Methological), Vol. 58 Issue 1, 267-288.

Vovk, V., Nouretdinov, I. and Gammerman, A. 2009. On-line Predictive Linear Regression. *The Annals of Statistics*, Vol. 37, No. 3, 1566–1590.

Weber, G. W., Batmaz, İ., Köksal, G., Taylan, P. and Yerlikaya-Özkurt, F. 2012. CMARS: A new contribution to nonparametric regression with multivariate adaptive regression splines supported by continuous optimization. *J. Inverse Problems in Science and Engineering*, 20, 371-400.

Wiens, D. P. 1996. Asymptotics of generalized M-estimation of regression and scale with fixed carriers, in an approximately linear model. *Statistics&Probability Letters* 30, 271-285.

Willmott, C. J. and Matsuura K. 2005. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Research*, Vol. 30: 79-82.

Yohai, V. J. 1987. High Breakdown-Point and High Efficiency Robust Estimates for Regression. *The Annals of Statistics*, 15, 642-656.

Yohai, V. J. and Zamar, R. H. 1988. High Breakdown-Point Estimates of Regression by Means of the Minimization of an Efficient Scale. *Journal of the American Statistical Association*, 83, 406-413.

TEZ

Biçkici, B. 2007. Çok Değişkenli Varyans Analizi Ve Çoklu Doğrusal Regresyon Analizinin Uygulamalı Olarak Karşılaştırılması. Erzurum: Atatürk Üniversitesi FBE, Zootekni ABD. Yüksek Lisans Tezi.

McCann, L. 2006. Robust Model Selection and Outlier Detection in Linear Regression**,** Thesis (Ph. D.) Massachusetts Institute of Technology, Sloan School of Management, Operations Research Center.

Yerlikaya, F. 2008. A New Contribution to Nonlinear Robust Regression and Classification with MARS and Its Application to Data Mining for Quality Control in Manufacturing, MSc. Dissertation, Middle East Technical University, Ankara.

KONGRE-SEMPOZYUM-KONERANS

Ben-Gal, I., Outlier detection, In: Maimon O. and Rockach L. (Eds.) Data Mining andKnowledge Discovery Handbook: A Complete Guide for Practitioners and Researchers," Kluwer Academic Publishers, 2005. ISBN 0-387-24435-2.

Fox, J. and Weisberg, S. 2010. Robust Regression Appendix to An R and S-PLUS Companion to Applied Regression Sage, Thousand Oaks, CA, second edition.

Nemirovski, A. 2002. Lectures on modern convex optimization. Israel Institute of Technology, C.O.R.E. Summer School on Modern Convex Optimization August 26-30.

Taylan, P. and Kaya, A. 2009. Outlier problems in linear regression, *Book of Abstracts, 23rd European Conference on Operational Research*, Bonn, 139 (http://www.euro-2009.de/).

Verboon, P. 1992. M-estimators in Multiple Regression With Optimal Scaling. University of Leiden: Department of Data Theory.

Williams, G. J. 2002. Baxter R. A., He H. X., Hawkins S., Gu L., ”A Comparative Study of RNN for Outlier Detection in Data Mining,” IEEE International Conference on Data-mining (ICDM’02), Maebashi City, Japan, CSIRO Technical Report CMIS-02/102.

PAKET PROGRAMLAR

MATLAB, http://mathworks.com/products/matlab/tryit.html

MOSEK, a commercial software for CQP, http://www.mosek.com/index.php?id=99

ALINTI

Belsley, D.A., Kuh, E., and Welsch, R.E. (1980). Regression Diagnostics: Identifying

Influential Data and Sources of Collinearity. New York: John Wiley & Sons. Cited in Long S.

and Fox, J. Modern Methods of Data Analysis (s.262)

ÖZGEÇMİŞ

1986 yılında Diyarbakır'da doğdum. İlköğrenimimi Yenişehir İlkokulu'nda, ortaokul ve lise öğrenimimi Diyarbakır Anadolu Lisesi'nde tamamladım. 2004 yılında Dicle Üniversitesi İlköğretim Matematik Öğretmenliği Bölümünde lisans eğitimime başladım ve 2008 yılında mezun oldum. Mezuniyetimden itibaren farklı ilköğretim okullarında görev yaptım. Görevime Çelebi Eser Ortaokulunda devam etmekteyim.

Burcu BİLGİÇ UÇAK