Çoklu Doğrusal Regresyonda Bazı Sağlam Yöntemlerin İncelenmesi

Refiye Şen

YÜKSEK LİSANS TEZİ

İstatistik Anabilim Dalı

Haziran 2010

Evaluation of Some Robust Methods in Multiple Linear Regression

Refiye Şen

MASTER OF SCIENCE THESIS

Department of Statistics

June 2010

Çoklu Doğrusal Regresyonda Bazı Sağlam Yöntemlerin İncelenmesi

Refiye Şen

Eskişehir Osmangazi Üniversitesi
Fen Bilimleri Enstitüsü
Lisansüstü Yönetmeliği Uyarınca
İstatistik Anabilim Dalı
Uygulamalı İstatistik Bilim Dalında
YÜKSEK LİSANS TEZİ
Olarak Hazırlanmıştır

Danışman: Yrd. Doç. Dr. Özlem ALPU

Haziran 2010

ONAY

İstatistik Anabilim Dalı Yüksek Lisans öğrencisi Refiye Şen'in YÜKSEK LİSANS tezi olarak hazırladığı "Çoklu Doğrusal Regresyonda Bazı Sağlam Yöntemlerin İncelenmesi" başlıklı bu çalışma, jürimizce lisansüstü yönetmeliğin ilgili maddeleri uyarınca değerlendirilerek kabul edilmiştir.

Danışman : Yrd. Doç. Dr. Özlem ALPU

İkinci Danışman : -

Yüksek Lisans Tez Savunma Jürisi:

Üye : Yrd. Doç. Dr. Özlem ALPU

Üye : Doç. Dr. Şenol ERDOĞMUŞ

Üye : Yrd. Doç. Dr. Sema BEHDİOĞLU

Üye : Yrd. Doç. Dr. Hatice ŞAMKAR

Üye : Yrd. Doç. Dr. Serpil TÜRKYILMAZ

Fen Bilimleri Enstitüsü Yönetim Kurulu'nun .... tarih ve .... sayılı kararıyla onaylanmıştır.

Prof. Dr. Nimetullah BURNAK

Enstitü Müdürü

ÇOKLU DOĞRUSAL REGRESYONDA BAZI SAĞLAM YÖNTEMLERİN İNCELENMESİ

REFİYE ŞEN

ÖZET

Bu çalışmada, çoklu doğrusal regresyonda aykırı değer ve etkin gözlemleri inceleyerek, veri setinde varsayım bozulumları veya aykırı değer ve/veya etkin gözlem olması durumunda kullanılabilen bazı sağlam regresyon metotları incelenmiştir.

Çalışmanın birinci bölümünde çoklu doğrusal regresyon modeli, parametre kestirimleri ve varsayımlarına kısaca değinilmiştir.

İkinci bölümde aykırı değerler ve etkin gözlemler açıklanarak, bu gözlemlerin sonuçları, belirlenmesi için kullanılan tek gözleme ve çok gözleme dayalı teşhis ölçüleri üzerinde durulmuştur.

Üçüncü bölümde sağlam regresyondaki temel kavramlar ve x-yönünde ve/veya y-yönünde aykırı değerlerin varlığında kullanılabilen bazı sağlam regresyon metotları incelenmiştir.

Dördüncü bölümde ele alınan veri setlerinden ilkinde hem x hem de y yönünde aykırı değerlerin olduğu, diğerinde sadece y yönünde aykırı değerlerin olduğu iki veri seti üzerinde öncelikle aykırı değer ve/veya etkin gözlemlerin istatistiklerle ve grafiklerle incelenmesi yapılmıştır. Daha sonra sağlam regresyon, parametre kestirim değerleri, ölçek kestirim değerleri ve ağırlıkların incelenmesi gerçekleştirilmiştir.

Çalışmanın son bölümünde ise sonuç ve önerilere yer verilmiştir.

Anahtar Kelimeler: Çoklu doğrusal regresyon, sağlam regresyon, aykırı değer, etkin gözlem

EVALUATION OF SOME ROBUST METHODS IN MULTIPLE LINEAR REGRESSION

REFİYE ŞEN

SUMMARY

In this study, by considering outliers and influential observations in multiple linear regression, some robust regression methods which can be used in the presence of the outliers and/or influential observations are examined.

In the first chapter of the study, multiple linear regression, its parameter estimates and assumptions are presented briefly.

In the second chapter, by explaining outliers and influential observations, sources of these observations, its results and single-case and multiple-case diagnostics are explained.

In the third chapter, the basics of the robust regression, some robust regression methods in the presence of outliers in the x-direction and/or y-direction are discussed.

In the fourth chapter, two data sets are taken: one of them has outliers both in the x- and y-direction while the other one has only outliers in the y-direction. Firstly, outliers and/or influential observations are examined statistically and graphically. Then, robust regression parameter estimates, scale estimates and its weights are discussed.

In the last chapter of the study, results and recommendations are given.

Keywords: Multiple linear regression, robust regression, outlier, influence observation.

TEŞEKKÜR

Yüksek lisans eğitimim ve tez çalışmam boyunca bana yol gösteren danışman hocam Yard.Doç.Dr. ÖZLEM ALPU' ya tüm sevgi ve saygımlarımla teşekkür ediyorum.

Bana güler yüzlerini eksik etmeyen ve zaman zaman yol gösteren bölümümüzdeki tüm değerli hocalarıma ayrıca teşekkür ediyorum.

Tüm hayatım boyunca maddi manevi desteğini şevkatini ve sevgisini esirgemeyen değerli annem FATMA ŞEN' e tüm sevgi ve saygımlarımla teşekkür ediyorum.

İÇİNDEKİLER

Sayfa

İÇİNDEKİLER (devam)

Sayfa

İÇİNDEKİLER (devam)

Sayfa

İÇİNDEKİLER (devam)

Sayfa

İÇİNDEKİLER (devam)

ŞEKİLLER DİZİNİ

ŞEKİLLER DİZİNİ (devam)

Şekil **Sayfa**

ŞEKİLLER DİZİNİ (devam)

Şekil **Sayfa**

TABLOLAR DİZİNİ

BÖLÜM 1

GİRİŞ

Regresyon terimi ilk olarak ingiliz antropolog ve istatistikçi Sir Francis Galton tarafından ebeveynlerin boyları ile çocuklarının boylarının arasındaki ilişkiyi araştırdığı bir çalışmada kullanılmıştır (Kardaun, 2005). Galton çalışmasında regresyon terimini çocukların genlerinde yer alan boy ortalamasına yaklaşma eğilimi gösterdiğini ifade etmek için kullanmıştır. Bu terim regresyonun günümüz teriminden oldukça farklıdır. Literatürde yaygın bir şekilde kullanılan regresyon terimi için regresyon analizi bağımlı değişken ile bir veya daha fazla bağımsız değişkenlerin bağlantısını inceleyerek bilinen (sabit, bağımsız değişkenin) değerlerden yola çıkıp bağımlı değişkenin ortalama değerini önkestirmeyle ve/veya kestirmeyle ilgilenir (Gujarati, 1999). Regresyon analizine ilişkin hesapsal sürecin temeli ise 19. yy' ın başlarında Gauss ve Legendre' nin yazılarına kadar gider. Yazılarında gözlemlerin saçılımına yaklaşık bir tanım vermek amacıyla bir doğru veya düzlemi belirlemede En Küçük Kareler (EKK) metodunu tanımlamışlardır. Regresyon analizinde olabilirliğe dayalı metodolojiyi R.A.Fisher 20.yy' ın ilk çeyreğinde vermiştir. Bu metodolojiyi özetleyen Draper ve Smith (1966), Scheffe (1959) ve Graybill (1961) gibi yazarlar da bu yüzyılın ilk yarısında yer almaktadır. Bu metotlar her alandaki araştırmalarda gelişmelere önemli katkı sağlamış ve halen yaygın bir şekilde kullanılmaktadır (Hocking, 2003). Regresyon analizinin uygulamaları tıp, biyoloji, tarım, ekonomi, mühendislik, sosyoloji ve jeolojiyi de içeren pek çok bilim alanında görülebilir.

Regresyon analizinin amaçları şu şekilde sıralanabilir:

1- Bağımlı değişken y ve bağımsız değişkenler $X_1, X_2, \ldots, X_k$ arasındaki nedensel ilişkinin kurulması,

2- $X_1, X_2, \ldots, X_k$ değerlerine dayalı bağımlı değişken değerlerinin önkestirimi,

3- Bağımlı ve bağımsız değişkenler arasındaki nedensel ilişkiyi daha doğru ve etkin bir şekilde belirleyebilmek amacıyla $X_1, X_2, \ldots, X_k$ değişkenlerinden hangilerinin bağımlı değişkeni açıklamada daha önemli olduğunun belirlenmesidir.

Bu çalışmanın amacı çoklu doğrusal regresyonda aykırı değer ve etkin gözlemlerin varlığında bazı sağlam regresyon metotlarını incelemektir. Bu amaçla öncelikle ilk bölümde çoklu doğrusal regresyon modeli ve varsayımlarına kısaca değinilecektir. İkinci bölümde etkin gözlemler ve aykırı değerler hakkında bilgi verilip, ortaya çıkış nedenleri, belirlenme yolları aktarılacaktır. Üçüncü bölümde etkin gözlem ve aykırı değerlerin varlığında önerilen metotlardan sağlam regresyon metotları, sağlam regresyonda kullanılan temel kavramlar üzerinde durulacaktır. Dördüncü bölümde yukarıda anlatılan konular biri özgün veri seti olmak üzere iki ayrı veri seti üzerinde uygulanarak yorumlamalar gerçekleştirilecektir. Son bölümde sonuç ve önerilere yer verilecektir.

1.1 Çoklu Doğrusal Regresyon Modeli

Çoklu doğrusal regresyon modeli sayısal bir bağımlı değişken ve birkaç bağımsız değişken arasındaki ilişkinin fonksiyonel şeklini verir. Çoklu doğrusal regresyon modelinin genel formu aşağıdaki gibidir:

$$y = \beta_0 + \beta_1 X_1 + ... + \beta_k X_k + \varepsilon \,, \quad \varepsilon \sim N(0, \sigma^2) \qquad (1.1)$$

Burada y bağımlı değişken, $X_1, ..., X_k$ bağımsız değişkenler ve ε ise hata terimidir. Modelin bağımlı değişkeninin β katsayıları ile doğrusal bir ilişkiye sahip olmasından dolayı gerçekte "doğrusal" olduğuna dikkat edilmelidir (Montgomery, et al., 2001).

Bağımlı ve bağımsız değişkenlerimiz için n gözleme sahip olduğumuzu düşünelim. Bu durumda y_i bağımlı değişkenin *i.* gözlem değeri ve x_{ij} , *j.* bağımsız değişkenin *i.* gözlem değerini göstermek üzere, i=1,.......,n ve j=1,.........,k, matris gösterimindeki regresyon modeli şu şekilde ifade edilmektedir:

$$\begin{bmatrix} y_1 \\ y_2 \\ \cdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1k} \\ 1 & x_{21} & \cdots & x_{2k} \\ & & \cdots & \\ 1 & x_{n1} & \cdots & x_{nk} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \cdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \cdots \\ \varepsilon_n \end{bmatrix} \tag{1.2}$$

veya

$$\mathbf{y} = \mathbf{X\beta} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim N(0, \boldsymbol{\sigma}^2\mathbf{I}) \tag{1.3}$$

Burada **y** *(nx1)* boyutlu bağımlı değişken vektörü **X** *nx(k+1)* boyutlu tasarım matrisi, **β** *(k+1)x1* boyutlu regresyon katsayıları vektörü ve **ε** ise *(nx1)* boyutlu hata vektörüdür (GroB, 2003).

1.2 En Küçük Kareler Tekniğinin (EKK) Varsayımları

En küçük kareler regresyon modelinin geçerliliği bazı varsayımlara bağlıdır. Regresyon modelinde belli başlı varsayımlar yerine getirilmediği zaman modelin kestirimi yanlı olabilir (Yaffee, 2002). Dolayısıyla çoklu doğrusal regresyonda parametre kestirimi için aşağıdaki model varsayımlarının sağlanması gerekir. Doğrusal regresyon modelinde hata terimleri üzerine dört temel varsayım söz konusudur. Bunlar;

1- Hata terimleri birbirinden bağımsızdır.

2- Hata terimleri normal dağılımlıdır.

3- Hata terimleri sıfır ortalamaya sahiptir.

4- Hata terimleri varyansları sabittir (aynı varyanslılık-eş varyanslılık olarak bilinir).

Ayrıca modelde ifade edilmeyen dört varsayım daha söz konusudur. Bunlar;

1- Bağımlı, bağımsız değişkenler sayısaldır,

2- Bağımsız değişkenlerin deterministik (belirli) ve hatasız ölçüldüğü varsayılır,

3- Bağımlı değişkenin de hatasız ölçüldüğü varsayılır,

4- Bağımlı değişken fonksiyonunun verilen forma sahip olduğu varsayılır.

Uygulamada ilave bir varsayım da tasarım matrisi üzerine yapılır. Diğer bir deyişle $(\mathbf{X'X})^{-1}$ matrisinin varolduğu varsayılır. Bu durum ancak ve ancak $\mathbf{X'X}$ matrisinin determinantının sıfırdan farklı olduğu yani tekil olmayan matris olması durumunda gerçekleşir. Regresyon modelinde bu varsayım bağımsız değişkenlerin birbirleriyle doğrusal bağımsız olduğunu ifade etmeye eşdeğerdir. Bununla birlikte uygulamada sıklıkla bağımsız değişkenlerin yaklaşık doğrusal bağımlı olduklarına rastlanır. Bir başka anlatım tarzıyla bağımsız değişkenler arasında güçlü veya tam bir ilişki yoktur. Bu varsayımın sağlanmaması çoklu bağıntı (bağlantı) olarak bilinir (Koutsoyiannis, 1989).

1.3 Çoklu Doğrusal Regresyon Modeli Parametrelerinin Kestirimleri

Doğrusal regresyon modelinde β regresyon katsayılarının $\hat{\beta}_{LS}$, EKK kestiricisini bulmak için hataların karelerinin toplamını minimize etmek amaçlanır ve aşağıdaki eşitlikte ifade edilir:

$$\min HKO = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{n}\hat{\varepsilon}_i^2 = \left\|\mathbf{y} - \mathbf{x}\hat{\boldsymbol{\beta}}\right\|^2 = (\mathbf{y} - \mathbf{x}\hat{\boldsymbol{\beta}})^T(\mathbf{y} - \mathbf{x}\hat{\boldsymbol{\beta}}) \qquad (1.4)$$

burada $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + ... + \hat{\beta}_k x_{ik}$ değeri bağımsız değişkenlerin *i.* gözlem değerlerine karşılık gelen bağımlı değişkenin önkestirilmiş değeridir. $\hat{\varepsilon}_i = y_i - \hat{y}_i$ farkı *i.* hata teriminin kestirimidir ve *i.* artık olarak isimlendirilir. Böylece $\hat{\boldsymbol{\beta}}_{\mathbf{LS}}$, $\dfrac{\partial HKO}{\partial \hat{\boldsymbol{\beta}}} = 0$ eşitliğinden veya aşağıda verilen normal denklemlerin çözümünden elde edilir.

$$\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y} \qquad (1.5)$$

Regresyon katsayılarının en küçük kareler kestiricisi ($\hat{\boldsymbol{\beta}}_{\mathbf{LS}}$), $\mathbf{X'X}$ matrisinin tekil olmayan matris olması koşuluyla şu şekilde ifade edilmektedir.

$$\hat{\boldsymbol{\beta}}_{\mathbf{LS}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \qquad (1.6)$$

Regresyon modeli için σ^2 varyansının kestirimi aşağıdaki eşitlikten elde edilir;

$$s^2 = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-k'} = \frac{\sum_{i=1}^{n}\hat{\varepsilon}_i^2}{n-k'} \qquad (1.7)$$

burada k' parametre sayısını göstermektedir.

1.4 Çoklu Doğrusal Regresyon Modeli Katsayılarıyla İlgili Çıkarsamalar

Bir regresyon modelinde $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}_{\mathbf{LS}}$ en küçük kareler kestiricisi $\boldsymbol{\beta}$ ortalama vektörü ve $\boldsymbol{\sigma}^2(\mathbf{X}'\mathbf{X})^{-1}$ varyans-kovaryans matrisi ile çok değişkenli normal dağılıma sahiptir. Belirli bir bağımsız değişkenin bağımlı değişken üzerinde etkili olup olmadığını görmek için ilgili regresyon katsayılarının sıfıra eşitliği test edilir.

Regresyon modelinde $H_0 : \beta_i = 0$' a karşı $H_1 : \beta_i \neq 0$ hipotezlerinin testi için aşağıdaki test istatistiği hesaplanır.

$$t = \frac{\hat{\beta}_i}{\sqrt{Var(\hat{\beta}_i)}} = \frac{\hat{\beta}_i}{s\sqrt{d_i}} \qquad (1.8)$$

Burada d_i, $(\mathbf{X}'\mathbf{X})^{-1}$ matrisinin *i.* köşegen elemanıdır. Bu test istatistiği n-k-1 serbestlik derecesi ile t dağılımına sahiptir (Montgomery, Peck and Vinning, 2001).

1.5 Çoklu Doğrusal Regresyon Modeliyle İlgili Ölçüler

Bu kısımda doğrusal regresyon modeli ile ilişkili bazı ölçülere yer verilecektir. Bu ölçüler şapka matrisi, hata kareler ortalaması ve çoklu belirlilik katsayısı olarak sıralanabilir.

1.5.1 Şapka (H) matrisi

H matrisi aşağıdaki gibi ifade edilebilmektedir:

$$\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \tag{1.9}$$

H matrisi *(nxn)* boyutlu, gözlenen değerlerden oluşan $\mathbf{y}$ vektörünü onun EKK kestiricisi olan $\hat{\mathbf{y}} = \mathbf{H}\mathbf{y}$ vektörüne dönüştürdüğü için şapka matrisi veya dönüşüm matrisi olarak isimlendirilmektedir. **H** matrisi, eşgüçlü (idempotent) $(\mathbf{HH} = \mathbf{H})$ ve simetrik $(\mathbf{H}' = \mathbf{H})$ dir.

Matrisin izi (trace) ve rankı ise;

$$iz(\mathbf{H}) = \sum_{i=1}^{n} h_{ii} = k + 1 \tag{1.10}$$

$$rank(\mathbf{H}) = \sum_{i=1}^{n} h_{ii} = k + 1 \tag{1.11}$$

şeklinde elde edilir. **H** matrisi eşgüçlü ve simetrik olma özelliklerinden dolayı h_{ii} aşağıdaki gibi ifade edilir (Rousseeuw and Leroy, 2003).

$$\begin{aligned} h_{ii} &= (\mathbf{HH})_{ii} = \sum_{j=1}^{n} h_{ij} h_{ji} = \sum_{j=1}^{n} h_{ij} h_{ij} \\ h_{ii} &= \sum_{j=1}^{n} h_{ij}^2 \,, \quad \textit{bütün i'ler için} \end{aligned} \tag{1.12}$$

Şapka matrisi $\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ formülünden şapka matrisinin köşegen elemanları olan h_{ii} değeri $h_{ii} = x_i'(\mathbf{X}'\mathbf{X})^{-1}x_i$, ($\mathbf{x}_\mathbf{i}' = (1 \quad x_{i1} \quad x_{i2}.............x_{ik})$ gözlem vektörü iken) formülü ile hesaplanmaktadır. Yani h_{ii} dönüşüm matrisinin *i.* köşegen öğesini ifade etmektedir. Değer aralığı $(1/n \le h_{ii} \le 1)$ olan h_{ii} değerinin aritmetik ortalaması ise (k+1)/n' dir (Myers, 1990).

Şapka matrisinin (gözlem uzaklıkları matrisi) köşegen öğeleri h_{ii} herhangi bir $\mathbf{x}_\mathbf{i}'$ gözlem vektörünün X_j' lerin aritmetik ortalamasından oluşan merkeze standart uzaklığını belirtir (Draper and John, 1981).

1.5.2 Hata kareler ortalaması

$\hat{\boldsymbol{\beta}}$ kestiricisinin hata kareler ortalaması aşağıdaki gibi tanımlanır:

$$HKO(\hat{\boldsymbol{\beta}}) = E(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^2 = Var(\hat{\boldsymbol{\beta}}) + (\boldsymbol{\beta} - E(\hat{\boldsymbol{\beta}}))^2 \qquad (1.13)$$

burada $\hat{\boldsymbol{\beta}}$ örneklemden elde edilen parametre kestirimini, $\boldsymbol{\beta}$ ise evren için parametre değerlerini ifade etmektedir. $Var(\hat{\boldsymbol{\beta}})$ parametre kestirimlerinin varyansını, $E(\hat{\boldsymbol{\beta}})$ parametre kestirimlerinin beklenen değerini vermektedir. $HKO(\hat{\boldsymbol{\beta}})$ değeri, şayet evrenin tamamı bilinmiyorsa gerçek hayatta bulunamaz. Gerçek hayatta evrenin tamamına ulaşmak oldukça zordur. Bu nedenle $HKO(\hat{\boldsymbol{\beta}})$ değeri ancak artık kareler ortalaması (AKO) ile tahmin edilir. AKO değerini ise $AKO = \dfrac{\Sigma e_i^2}{n-k'}$ şeklinde elde edilir. Burada Σe_i^2 artıkların karelerinin toplamını, n gözlem sayısını, k' ise parametre sayısını vermektedir.

1.5.3 Çoklu belirlilik katsayısı (R^2)

Bir regresyon modelinin çoklu belirlilik katsayısı aşağıdaki gibi ifade edilmektedir:

$$R^2 = \frac{\sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \tag{1.14}$$

bu katsayı doğrusal regresyon modeli tarafından açıklanan bağımlı değişkendeki toplam değişim oranı olarak yorumlanır.

Modelin verilere ne kadar iyi uyduğunun bir göstergesi olarak kullanılır. R^2 değeri 0 ile 1 arasında değişir. Belirlilik katsayısı (R^2) bağımlı değişkendeki değişmelerin yüzde kaçının bağımsız değişkenler tarafından açıklanabildiğini ifade eder. R^2' nin 1' e yakın olması modelin verilere iyi uyumunun bir göstergesidir (Draper and Smith, 1981).

BÖLÜM 2

AYKIRI DEĞERLER VE ETKİN GÖZLEMLER

2.1 Aykırı Değerlerin ve Etkin Gözlemlerin Tanımı

2.1.1 Aykırı değer tanımı

Veri setindeki bir gözlem geri kalan diğer gözlemlerden önemli derecede farklı veya alışılmamış ise aykırı değer olarak tanımlanır (Freund and Wilson, 1998). Ayrıca bir gözlem artık ortalamasından 3 standart sapma kadar veya daha büyük ise o gözlem aykırı değer olarak değerlendirilebilir.

Gerçek veri ve model tarafından kestirilmiş değerler arasındaki fark olarak ifade edilebilen artıkların çok büyük değerli olanları ciddi anlamda kestirimi çarpıtabilirler. Bu artıklar aykırı değer olarak isimlendirilirler. Aykırı değerler hata varyansını artırır ve standart hata da artar, katsayı aralığı genişlemiş olur. Dolayısıyla kestirim asimtotik olarak tutarlı olmamaktadır (Yaffee, 2002).

Bu aykırı değerlerin büyük artıklı olanları sıklıkla EKK regresyon fonksiyonunda büyük etkiye sahip olabilirler. Bundan dolayı çalışmalarda aykırı değerlere dikkat edilmeli ve bu aykırı değerlerin silinip silinmemesine karar verilmeli; silinmemesine karar verildi ise bu aykırı değerlerin uygulama sürecinde etkilerinin azaltılıp azaltılmayacağına ve/veya regresyon modelini değiştirip değiştirmediğine dikkat edilmelidir (Kutner, Nachtsheim, Neter and Li, 2005).

Aykırı değerler tek değişkenli veya çok değişkenli durumda, kesikli veya sürekli, bağımlı veya bağımsız değişkenler arasında ve veri seti içinde veya analiz sonuçları içinde yer alabilir (Tabachnick and Fidell, 2001).

Her hangi bir regresyon analizinde temel adım; veri setindeki bir veya birkaç gözlem tarafından etkilenmiş olabileceğini de hesaba katarak regresyon modelini belirlemektir.

Bir veya iki bağımlı değişkenli regresyon modellerinde x veya y-yönündeki aykırı değerler, kutu grafiği, gövde-yaprak grafiği, serpilme grafiği ve artık grafiği ile oldukça basit bir şekilde saptanabilir. Böylece aykırı değerin regresyon fonksiyonunu etkileyip etkilemediği görülebilir (Kutner, Nachtsheim, Neter and Li, 2005).

Bir veri setinde aykırı değer tespit edildiğinde tespit edilen bu aykırı değerin etkisiyle, veri setinden elde edilmiş olan regresyon modelinin veri setinin geri kalanındaki gözlemlerden uzaklaştığı bilinmektedir.

Böyle durumlarda sağlam regresyon yöntemlerinin kullanılması ile aykırı değer haricindeki gözlem artıklarının büyümesinden dolayı meydana gelen sapmalardan etkilenmemektedir.

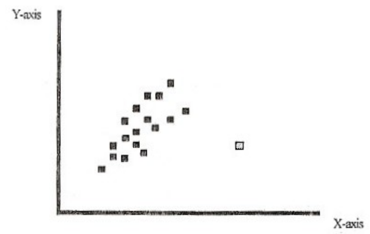
Aykırı değer varlığında regresyon model kestirimi yapılırken sağlam regresyon ile EKK yöntemine kıyasla daha az etkilenmiş parametre kestirimleri elde edilmektedir (Rousseeuw and Leroy, 1987).

Aykırı değerler konum ve etkileri bakımından tanımlanabilir.

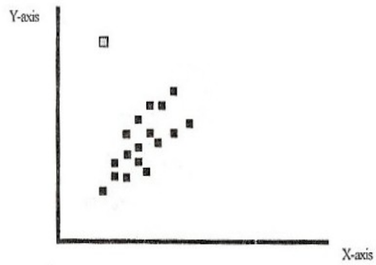
Şekil 2.1 a,b,c de görüldüğü gibi konum bakımından aykırı değer x-yönünde, y-yönünde ve her iki yönde de ortaya çıkabilir. x-yönündeki bağımsız değişken değerlerini, y-yönüdeki bağımlı değişken değerini ifade eder.

Diğer bir anlatım ile y-yönünde aykırı değer var ise bağımlı değişkende aykırı değer, x-yönünde aykırı değer var ise bağımsız değişkende aykırı değer var demektir.

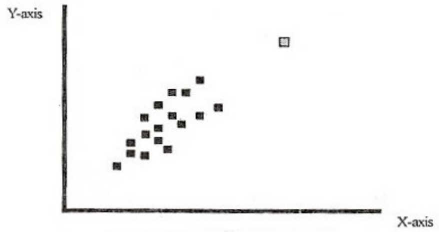
Şekil 2.1a,b,c Çeşitli aykırı değer konumları



Şekil-2.1a   x-yönündeki aykırı değer.



Şekil-2.1b   y-yönündeki aykırı değer



Şekil-2.1c Hem x- yönündeki hem y- yönündeki aykırı değer

Aykırı değerlerin regresyon denklemindeki katsayı kestirimlerini etkileme düzeyi konumlarına göre değişmektedir, y-yönündeki bir aykırı değer regresyon katsayı kestiriminde az oranda (hafif şiddette) bir etkiye sahipken, x-yönündeki bir aykırı değer regresyon katsayılarının kestiriminde daha fazla bir etkiye sahiptir. Aykırı değerler etkilerine göre sınıflandırılırlar. Etkinin ölçüsü ise; kaldıraç noktası, uyuşmazlık ve etki fonksiyonu ile değerlendirilir (Montgomery, Peck and Vinning, 2001).

2.1.2 Kaldıraç noktası

Diğer gözlem değerlerinden çok uzakta yer alan bir gözlem olarak tanımlanır ve kaldıraç noktası gözlem değerlerin geri kalanından uzaklığının yönünü hesaba katmaz. Aykırı değer regresyon doğrusundan uzakta ise doğru aykırı değer yönüne doğru kayar ve kötü kaldıraç noktası (bad leverage point) olarak isimlendirilir.

Kötü kaldıraç noktası parametre kestirimlerini yanlı yapan aykırı değerlerdir. Gözlem değeri kestirimi yapılan doğru boyunca uzanan fakat verilerin çoğunluğundan uzaktaysa iyi kaldıraç noktası (good leverage point) olarak değerlendirilir çünkü iyi kaldıraç noktası standart hatanın büyüklüğünü azaltır (Rouseeuw and Leroy, 1987; Yaffee, 2002).

2.1.3 Uyuşmazlık (discrepancy)

Regresyon doğrusu ile bir gözlem değerinin arasındaki uzaklığın ölçüsüdür. Diğer bir ifadeyle aykırı değerin aykırılığının derecesidir.

Yüksek kaldıraç noktalı gözlem değerleri aynı regresyon doğrusu üzerinde fakat diğer gözlem değerlerinden uzakta iken, uyuşmazlığı yüksek olan gözlem değeri hem regresyon doğrusundan hem de diğer gözlem değerlerinden uzaktadır (Tabachnick and Fidel, 2001).

2.1.4 Etki fonksiyonu

Belirli bir istatistik üzerinde keyfi bir x noktasının etkisini değerlendirmenin en basit yolu x gözleminin olduğu ve olmadığı istatistik değerlerinin arasındaki farkı hesaplamaktır.

Bir F dağılımına sahip evrenden çekilen aynı dağılımlı birbirinden bağımsız (i.i.d) birimlerden oluşan n birimlik örneklemden ($X_1, X_2, \ldots\ldots X_n$) elde edilen T kestiricisinin etki fonksiyonu ;

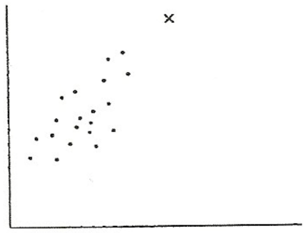
$$IF(x,T,F) = \lim_{\varepsilon \to \infty} \frac{T((1-\varepsilon)F + \varepsilon\Delta_x) - T(F)}{\varepsilon} \tag{2.1}$$

olarak tanımlanır. Burada Δ_x, x noktasında gözlem değerlerini 1, diğer yerlerde 0 veren bir olasılık dağılımıdır. x' in bir fonksiyonu olarak değerlendirilen (2.1) no' lu eşitlikteki nicelik Hampel(1968, 1974) tarafından etkinlik eğrisi veya etki fonksiyonu olarak isimlendirilmiştir.

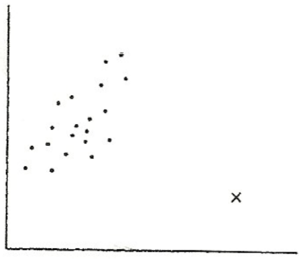
Etki fonksiyonu iki temel kullanıma sahiptir. Bunlardan ilki, bir kestirim değeri veya test istatistiğine karşı her bir gözlemin göreli etkinliğini değerlendirmemizi sağlar. İkinci olarak da bir kestirimin asimptotik özelliklerini basit ve sezgisel olarak değerlendirmemizi sağlar (Huber, 1981).

Etki fonksiyonu istatistiksel sürecin sağlamlılığını sınırlı olarak tanımlar ve etki fonksiyonu, kırılma noktası olarak adlandırılmış geniş çaplı güvenilirlik ölçüsü tarafından tamamlanır **(**Bab-Hadiashar and Suter, 2000).

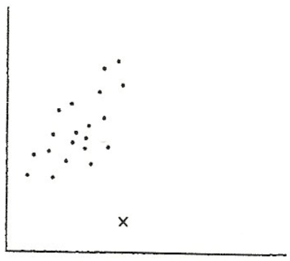
Şekil 2.2 a,b,c Kaldıraç noktası, uyuşmazlık ve etkin gözleme ilişkin örneklerdir (Tabachnick and Fidell, 2001).



Şekil 2.2.a Yüksek kaldıraç, düşük uyuşmazlık, orta şiddette etki



Şekil 2.2.b Yüksek kaldıraç, yüksek uyuşmazlık, yüksek etki



Şekil 2.2.c Düşük kaldıraç, yüksek uyuşmazlık, orta şiddette etki

2.1.5 Etkin gözlem tanımı

Bir gözlem veri setinden silindiğinde regresyon parametre kestiriminde önemli bir farklılığa neden oluyor ise etkin gözlem olarak tanımlanmaktadır (Freund and Wilson, 1998). Etkin gözlemler, veri setindeki diğer gözlemlerle karşılaştırıldığında regresyon denklemini bireysel olarak veya topluca aşırı derecede etkileyen gözlemlerdir.

Bununla birlikte bir gözlem regresyon sonuçlarının tamamında aynı etkiye sahip olmayabilir. Önemli olan soru "ne üzerine etkili" olduğudur. Örneğin bir gözlem $\hat{\beta}$ üzerinde, $\hat{\beta}$' nın kestirilen varyansı üzerinde, önkestirim değerleri üzerinde ve/veya uyum iyiliği istatistikleri üzerinde etkili olabilir.

Analizin temel amacı hangi etkinin değerlendirileceği, inceleneceği sorusuna olan cevabı ortaya koyabilmektedir. Örneğin β' nın kestirimi ile öncelikli olarak ilgileniliyorsa bu durumda $\hat{\beta}$ üzerinde gözlemlerin etkisinin ölçülmesi uygun olurken, önkestirim amacı öncelikli ise önkestirim yapılan değerler üzerindeki etkilerin ölçülmesi $\hat{\beta}$ üzerindeki etkilerin ölçülmesinden daha uygun olur.

Etkin gözlemler, aykırı değerler ve yüksek kaldıraç noktaları arasındaki ilişkiler aşağıda verildiği gibidir:

1- Aykırı değerler etkin gözlem olmak zorunda değildir,

2- Etkin gözlemler aykırı değer olmak zorunda değildir,

3- Büyük artığa sahip gözlemler arzu edilmezken küçük artığa sahip gözlemin de her zaman iyi gözlem olması beklenmemelidir, küçük artıklı bir aykırı değer aynı zamanda yüksek kaldıraç noktası olabilir,

4- Aykırı değerlerde olduğu gibi yüksek kaldıraç noktalı gözlemler etkin gözlem olmak zorunda değildir ve etkin gözlemlerde yüksek kaldıraç noktalı gözlemler olmak zorunda değildir (Chatterjee and Hadi, 1988).

Bir gözlemin y-eksenine göre, x-eksenine göre veya her iki eksene göre dışarıda olabileceği bir durum söz konusu olabilir.

Şekil 2.3' de verilen serpilme grafiğinde;

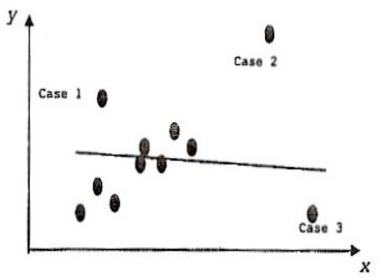
Gözlem 1, y-eksenine göre dışarıda iken x-eksenine göre gözlemlere ilişkin değişim aralığının ortasına yakın bir değerdir.

Gözlem 2, x-eksenine göre oldukça uzakta bir gözlem olduğu için y-eksenine göre değil x-eksenine göre aykırı değerdir. Diğer yandan;

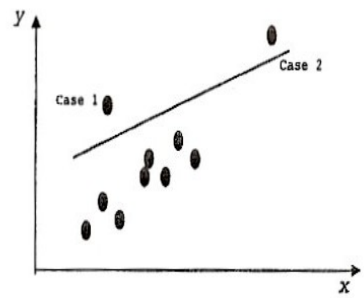
Gözlem 3, hem x-eksenine göre hem y-eksenine göre aykırı değerdir. x ve y eksenlerinin dışında kalan gözlem etkin gözlem olarak isimlendirilir. Çünkü regresyon katsayılarının uyumunda güçlü bir etkiye sahiptir.

Şekil 2.4.a ve Şekil 2.4.b' de gözlem 3' ün varlığının veya yokluğunun kestirimi yapılan regresyon doğrusunun eğimini nasıl etkilediği gösterilir. Diğer taraftan x ve y gözlem değerlerinin ötesinde (dışında) kalan değerler kestirilen regresyon katsayılarının üzerinde güçlü etkiye sahip olamazlar. Kestirimi yapılan regresyon doğrusuna uyumda diğer x gözlem değerlerine benzer olduğu için gözlem 1 etkin gözlem özelliği taşımamaktadır. Aynı zamanda gözlem 2' de etkin gözlem değildir. Çünkü bu ikinci gözlemin y değeri aykırı değer olmayan gözlemlerle ifade edilen regresyon ilişkisi ile tutarlıdır (Kutner, Nachtsheim, Neter and Li, 2005).

Şekil 2.3 Aykırı değerli veri kümesinin serpilme grafiği



Şekil 2.4.a



Şekil 2.4.b

Şekil 2.4.a ve Şekil 2.4.b, Üçüncü gözlemin regresyon doğrusunun eğimi üzerindeki etki derecesinin gösterimi

2.2 Aykırı Değerlerin ve Etkin Gözlemlerin Ortaya Çıkış Nedenleri

Aykırı değerlerin özellikle etkin gözlemlerin ortaya çıkış nedenleri aşağıda verilmiştir:

1- Kayıt hatası veya ölçüm hatası (ölçüm aletindeki bozukluktan kaynaklanan hata) olabilir.

2- Normal hata terimine sahip doğrusal bir terimin veri seti için yetersiz kalması da bir aykırı değerin göstergesi olabilir. Gerçek dağılım çarpıksa aykırı değer yerinde bir aşırı (uç) gözlem olabilir.

3- Normal olmayan koşullar (çalışan bir makinenin aşırı ısınması gibi) altındaki gözlemlerin kayıt edilmesi söz konusu olabilir.

4- Etkileşim teriminin unutulması veya karesel terimin hesaba katılmaması gibi yanlış modelin seçilmesinden kaynaklanabilir (Cohen, Cohen, West and Aiken, 2003).

Aykırı değerlerin regresyon çözümlemesinde yarattığı sorunlar aşağıda gibi sıralanabilir:

1- Yanlı (biased) kestirimlerin elde edilmesi,
2- Kareler toplamlarının ve buna bağlı olarak hesaplanan ölçülerin şişmesi (büyümesi),

3- İstatistiksel anlamlılık için kullanılan p değerlerinin bozulması,
4- Hatalı karar verme veya yanlış sonuç çıkarımına neden olurlar (High, 2004).

Aykırı değerlerle başa çıkmanın bir alternatif yolu onların yerine bir değer ikame etmektir. Yerine koyma işlemi birkaç sağlam regresyon kestirim metotlarından birini kullanarak gerçekleştirilir. Sağlam regresyon aykırı değerlere küçük ağırlık veren ve böylece onların etkilerini azaltan kestirim metotlarını kullanır. Çok sayıda sağlam regresyon kestirim metotları mevcuttur. Bunlardan bazıları: En küçük mutlak sapma (LAD), ağırlıklandırılmış en küçük mutlak sapma (WLAD), en küçük medyan kare (LMS), en küçük kırpılmış kareler toplamı(LTS), M kestirimi (M), MM kestirimi (MM), genelleştirilmiş M kestirimi (GM) sayılabilir (Rousseeuw and Leroy, 2003).

2.3 Aykırı Değerler ve Etkin Gözlemlerin Sonuçları

Modelde aykırı değer mevcutsa yanlış olan nedir? Ya modelin fonksiyonel şekli tamamen yetersizdir veya modelin şekli doğru tahmin edilmiş fakat regresyon katsayılarının kestirimleri etkin gözlemlerden bozulmuştur (etkilenmiştir).

2.4 Aykırı Değerlerin ve Etkin Gözlemlerin Belirlenmesi

2.4.1 Aykırı değer teşhis yöntemleri

Aykırı değerler iki boyutlu uzayda serpilme grafiğinde kolaylıkla belirlenebilirken değişken sayısının artması durumunda belirlenmeleri sorun yaratabilmektedir. Bu tür değerlerin tespiti için geliştirilen istatistiklere “aykırı değer teşhisçileri” denmektedir.

Aykırı değer teşhisçileri EKK kestiricileri üzerinde büyük etkiye sahip gözlemlere dikkati çeken istatistiklerdir.

Bir veya birden fazla aykırı değeri belirlemek için nümerik ve grafiksel araçları birleştiren çok sayıda teşhis ölçüsü geliştirilmiştir. Bu teşhisçilerin çoğu da EKK' den elde edilen artıklara dayanır.

Ancak bir aykırı değer aynı zamanda bir kaldıraç noktası olduğunda küçük artığa sahip olabilir ve EKK' ya dayalı teşhisçiler bu tür gözlemlerle başa çıkmada genellikle başarısızdırlar.

Bir diğer teşhisçi sınıfı veri setinden gözlem silme prensibine dayanır. Bu yaklaşımda *i.* gözlemin olmadığı $\hat{\beta}_i$ tahmincisiyle *i.* gözlemin olduğu durumdaki $\hat{\beta}$ tahmincisi arasındaki farka bakılır. Bu fark *i.* gözlemin varlığının regresyon katsayıları üzerinde nasıl bir etkiye sahip olduğunu gösterir.

Bu tür teşhisçiler "Tek Gözlem Teşhisçileri (Single Case Diagnostics)" olarak isimlendirilir. Birden fazla gözlemin eşanlı olarak etkisini belirlemek için bu teşhisçiler "Çoklu Gözlem Teşhisçileri" ne genelleştirilebilir (Rousseeuw and Leroy, 2003).

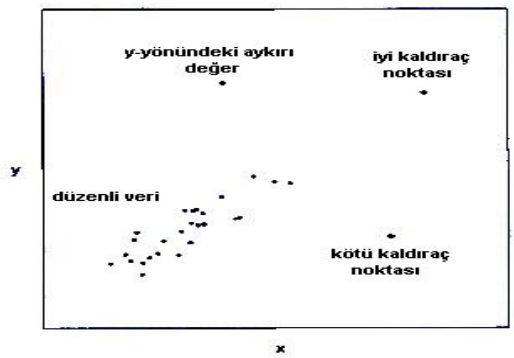
Araştırmacıların varsayımların sağlanıp sağlanmadığını görmek için verileri dikkatli şekilde incelemesi gerekir. Veri setinde aykırı değer olduğunda hataların normal dağılımlılık varsayımı sağlanmaz. Aykırı değerler istatistiksel ve grafiksel yöntemlerin kullanılmasıyla belirlenebilir.

Regresyon verilerindeki gözlemler;

1- $\mathbf{x_i}$**'** nin içinde ve y_i' ye iyi uyum sağlayan (gösteren) düzenli gözlemler
2- $\mathbf{x_i}$**'** nin içinde y_i' ye uyum sağlamayan (göstermeyen) dikey aykırı değerler (y-yönünde aykırı değerler)
3- $\mathbf{x_i}$**'** nin dışında y_i' ye iyi uyum sağlayan iyi kaldıraç noktaları,
4- $\mathbf{x_i}$**'** nin dışında ve y_i' ye uyum sağlamayan kötü kaldıraç noktaları olmak üzere dört farklı şekilde ortaya çıkabilir.

Şekil 2.5 klasik regresyondaki bu dört tip durumu gösterir. Regresyon teşhisçileri bu tip bir veya daha fazla gözlemi belirlemeyi hedeflemektedir. Bu amaçla çalışmada özellikle literatürde yaygın olarak kullanılan aşağıdaki teşhis ölçülerine yer verilecektir.

1- Standartlaştırılmış artıklar,
2- Sağlam standartlaştırılmış artıklar,
3- Student artıklar,
4- Mahalanobis uzaklığı,
5- Sağlam uzaklıklara dayalı teşhis grafiği,
6- Minimum hacimli elipsoid (MVE) metoduna dayalı sağlam uzaklıklar
7- Minimum determinantlı kovaryans (MCD) metoduna dayalı sağlam uzaklıklar.



Şekil 2.5 Regresyon verilerindeki gözlem türleri

2.4.1.1 Standartlaştırılmış artıklar

Regresyon analizi sonucu elde edilen artıklar ölçeklendirilmemiş artıklar olduğundan karşılaştırılmaları mümkün olamamaktadır. Bu nedenle artıkların standartlaştırılmaları istenir. Her bir artığın kendi standart sapmasına bölünmesiyle standartlaştırılmış artıklar elde edilir ve

$$r_i = \frac{e_i}{s} \tag{2.2}$$

biçiminde tanımlanır. Eşitlik (2.2)' deki standart sapma

$$s = \frac{1}{n-p}\sum_{i=1}^{n} e_i^2 \tag{2.3}$$

olarak hesaplanır.

Standart normal dağılım gösteren ε/σ ögelerinin % 95 'inin -1,96 ile +1,96 sınırları arasında değiştiği; buna benzer olarak da, çoğu kez e_i / s ' nin (-2,+2) arasında olabileceği belirtilir.

2.4.1.2 Sağlam standartlaştırılmış artıklar

Sağlam standartlaştırılmış artıklar aşağıdaki gibi tanımlanır:

$$r_i(\theta)/s(\theta) \tag{2.4}$$

burada $s(\theta)$ artıklara dayalı sağlam ölçek kestirimini ifade eder. Sağlam ölçek kestirimi olarak LTS metoduna dayalı ölçek kestiricisi kullanılacaktır.

$$s_{LTS}(\theta) = d_{h,n}\sqrt{\frac{1}{h}\sum_{i=1}^{h} r^2(\theta)_{i:n}} \tag{2.5}$$

Sağlam standartlaştırılmış artıkların mutlak değerleri ile bir ölçütü karşılaştırarak iyi uyum sağlayan gözlemlerle uyum sağlamayan gözlemler arasındaki farkı ayırt etmemize yardımcı olur. Diğer bir deyişle, $|r_i(\theta)|/s(\theta)$ ile 2,5 değeri karşılaştırılır.

Hata terimleri normal dağılıma sahip ise e_i' ler için % 99 güven aralığını kabaca belirlediğimizden 2,5 değerini kullanırız. Sağlam standartlaştırılmış artıklar e_i' ye yakınsadığı için bir gözlemin standartlaştırılmış artığı bu tolerans bölgesinin dışında yer alıyorsa gözlemin y_i' ye iyi uyum göstermediği düşünülür.

2.4.1.3 Student artıklar

Belsley, Kuh,Welsch (1980) her bir artığın, artıklardan bağımsız kendi standart sapmasının kestirimi ile standartlaştırılmasını önermektedir. Bu durum analizden *i.* gözlemin çıkartılarak artık kareler ortalamasının (AKO) kullanımı ile gerçekleştirilir.

Veri setinden *i.* satırın çıkartılması sonucu kalan n-1 gözlem değeri ile elde edilen $\hat{\sigma}_{(i)} = s_{(i)}\sqrt{1-h_{ii}}$ biçimindedir.

Analizden *i.* gözlemin çıkartılması ile elde edilen $\hat{\sigma}_{(i)}$ ' ye bağlı student türü artık

$r_i^* = \dfrac{e_i}{s_{(i)}\sqrt{1-h_{ii}}}$ ile gösterilir. r_i^* artıkları gözlem çıkartılarak elde edildiğinden "dışsal student artıkları" olarak isimlendirilir ayrıca literatürde "jacknife artıkları", "RSTUDENT" v.b isimlendirmeler de söz konusudur. $\varepsilon \sim N(0, \sigma^2 I)$ olduğunda her bir student türü artık (n-k-1) serbestlik derecesi ile t dağılımına sahiptir, buradan elde edilen kritik değer kullanılarak gözlemin aykırı değer olup olmadığı araştırılabilir (Belsley, Kuh and Welsch, 1980).

2.4.1.4 Mahalanobis uzaklığı

Çoklu aykırı değerleri belirlemekte kullanılan bir istatistik olan Mahalanobis uzaklığı bir gözlemin, ilişkisel ve değişimsel yapısını ele alarak diğer gözlemlerin ağırlık merkezinden olan uzaklığını ölçer.

Çoğu veri setinde gözlemler ağırlık merkezi etrafında yığın halinde bulunurlar ve her bir gözlem yığında tek bir nokta ile temsil edilmektedir (Schinka and Velicer, 2003).

Mahalanobis uzaklığında eğer iki değişken birbirine bağımlı ise, istatistik uzaklık için iki değişken arasındaki kovaryans veya korelasyon katsayısının da dikkate alınması gerekir ve Mahalanobis uzaklığı, iki birim değeri veya noktası arasındaki uzaklığı ölçmede iki değişken arasındaki kovaryans veya korelasyon katsayısını da dikkate alan bir uzaklık ölçütüdür. Kısaca, i ve j birim değerleri arasındaki Mahalanobis uzaklığı,

$$MD_{ij}^2 = \left[\frac{1}{1-r^2}\right] \times \left[\frac{(x_{i1}-x_{j1})^2}{s_1^2} + \frac{(x_{i2}-x_{j2})^2}{s_2^2} - \frac{2r \times (x_{i1}x_{j1}) \times (x_{i2}-x_{j2})}{s_1 s_2}\right] \quad (2.6)$$

ile hesaplanmaktadır. Buradaki s_1^2, s_2^2 sırasıyla birinci ve ikinci değişkenin varyanslarını, r ise iki değişken arasındaki korelasyon katsayısını göstermektedir. Yukarıdaki bağıntıdan, iki değişken birbirinden bağımsız ise Mahalanobis uzaklığı istatistik uzaklığına eşit olduğu ve eğer değişkenlerin bağımsızlığının yanında varyanslarının bir olması halinde Mahalanobis uzaklığı öklid uzaklığına dönüştüğü görülmektedir. Yani, öklid ve istatistik uzaklık Mahalanobis uzaklığının özel bir şeklidir. Genel olarak, k değişkenli uzaydaki iki birim değeri arasındaki Mahalanobis uzaklığı aşağıdaki eşitlikle hesaplanmaktadır:

$$MD_{\mathbf{ik}} = \left(\mathbf{x_i} - \mathbf{x}_j\right)' \mathbf{S}^{-1} \left(\mathbf{x_i} - \mathbf{x}_j\right) \quad (2.7)$$

Eşitlik (2.7)' de **x**, $(k \times 1)$ boyutunda bir koordinatlar vektörünü; **S** ise, $(k \times k)$ boyutunda bir kovaryans matrisini göstermektedir. Bağımsız standart değişkenler için **S** matrisi, matris köşegeni üzerinde standart varyansları gösteren bir birim matrisdir (Albayrak, 2006). Mahalanobis uzaklığı regresyonda kaldıraç noktalarının bir ölçüsü olarak kullanılırken aykırı değerlerin belirlenmesinde güvenli bir metot değildir.

Mahalanobis uzaklığının aykırı değerleri bulmada başarısız olduğu durumlar söz konusudur. Örneğin; y-yönündeki aykırı değerler Mahalanobis uzaklığı ile belirlenemez. Ayrıca veri setinde çok sayıda kaldıraç noktası olan aykırı değerler iyi kaldıraç noktalarının etkileri tarafından maskelenebilir.

2.4.1.5 Minimum hacimli elipsoid (MVE) metoduna dayalı sağlam uzaklıklar

Rousseeuw (1983) tarafından tanımlanan minimum hacimli elipsoid (Minimum Volume Elipsoid (MVE)) yöntemi gözlemlerin en azından yarısını içine alacak şekilde belirlenen kestiricilere dayalı olarak tanımlanan %50' lik kırılma noktasına sahip bir yöntemdir. Gözlem sayısının fazla olması durumunda hesaplaması zaman alıcı ve problemlidir. MVE hesaplanması için X_{nxk} olmak üzere rastgele olacak şekilde belirlenen $k+1$ gözleme ait alt küme için ortalama ve varyans-kovaryans matrisi yardımıyla karşılık gelen Mahalanobis uzaklıkları hesaplanır.

Alt kümedeki gözlem sayısı s ise elde edilen Mahalanobis uzaklıklarından minimum $s+1$ tanesini alarak yeni alt küme belirlenir. Alt kümede n-h gözlem olana kadar yukarıdaki işlemler tekrarlanır (h=[(n+k+1)/2]). Son adımda elde edilen alt kümeye ait Mahalanobis uzaklıkları yardımıyla bu alt kümeye karşılık gelen elipsoidin hacmi hesaplanır. Bu işlem $\binom{n}{k+1}$ kadar seçilen tüm alt kümeler için tekrarlanır. İçlerinden minimum hacmi veren alt küme belirlenir. Bu alt kümedeki gözlemler temiz, dışarıda kalanları ise aykırı değer olarak bildirilir (Kıral ve Billor, 2001).

k değişkenli $k \geq 2$ veri setinde aykırı değerleri belirlemenin bir yolu Mahalanobis uzaklıklarını kullanmaktır, ancak bilinen ortalama ve örneklem kovaryans matrisi yerine yüksek kırılma noktalı kestiriciler yer değiştirilebilir.

Bu tip metotlardan ilki Rousseeuw ve Van Zomeren (1990) tarafından önerilen MVE kestiricileridir.

Mahalanobis uzaklığı bir aykırı değer olduğunda verilerin merkeze olan uzaklığını varyans-kovaryans matrisini de dikkate alarak hesaplayabilmesine rağmen birden fazla aykırı değer olduğunda bu aykırı gözlemlerin dayanıklı olmayan ortalama ve varyans-kovaryansı kendisine doğru çekerek aykırı gözlemlerin teşhis edilmesini engelleyebilmektedir.

Bu nedenle aykırı değerlerin doğru teşhis edilebilmesi için Rousseeuw ve Van Zomeren Mahalanobis uzaklık formülündeki konum ve ölçek için kullanılan ortalama ve varyans-kovaryans matrisi yerine dayanıklı konum ve ölçek ölçülerinin kullanılması gerekliliğini belirtmişlerdir ve bu şekilde hesapladıkları Mahalanobis uzaklıklarını sağlam uzaklıklar (Robust distances) olarak tanımlamışlardır.

((k+1)x1) boyutlu **C** sütun vektörü MVE konum kestirimi ve *(kxk)* boyutlu **M** matrisi de kovaryans matrisi olmak üzere, $\mathbf{x}'_i = (x_{i1},......,x_{ik})$ noktasından **C**’ nin uzaklığı aşağıdaki gibi verilir:

$$MVED_i = \sqrt{(\mathbf{x_i} - \mathbf{C})'\mathbf{M}^{-1}(\mathbf{x_i} - \mathbf{C})} \qquad (2.8)$$

Eşitlik (2.8)’ de $MVED_i > \sqrt{\chi^2_{0,975,k}}$ ise, $\mathbf{x_i}$ aykırı değer olarak belirlenir.

Rousseeuw ve Van Zomeren(1990) bu metodu her değişkende en az 5 gözlem olması durumunda (diğer bir deyişle; $\frac{n}{k} > 5$) önermiştir.

2.4.1.6 <u>Minimum determinantlı kovaryans (MCD) metoduna dayalı sağlam uzaklıklar</u>

MVE’ ye alternatif olarak tanımlanmış, minimum determinantlı kovaryans (Minimum Covariance Determinant (MCD)) (Rousseeuw, 1984) yönteminde, amaç n gözlem üzerinden klasik kovaryans matrisinin determinantı en küçük olan h gözlemi bulmaktır. Bu durumda konum ve ölçek parametrelerinin MCD tahminleri sırasıyla bu h gözlemin ortalama ve kovaryans matrisleri olacaktır.

Yöntemin kırılma noktası MVE yöntemi ile aynıdır. Ancak MCD' nin asimtotik olarak normal olması (Butler, Davies ve Jhun, 1993) nedeniyle MVE ile karşılaştırıldığında avantajlara sahiptir.

Yöntem MVE' ye göre istatistiksel olarak daha etkindir. MCD' ye dayalı sağlam (robust) uzaklıklar MVE' ye dayalı olarak elde edilenlere göre daha kesindir. Bu nedenle de çok değişkenli veri kümeleri içerisinde problemli gözlemleri belirlemede daha uygun bir yöntemdir (Rousseeuw and Van Zomeren, 1990; Kıral ve Billor, 2001).

2.4.1.7 Sağlam uzaklıklara dayalı teşhis grafiği

Regresyon verilerindeki gözlemlerin ortaya çıkarılmasındaki dört tip durum Şekil 2.6' da sınıflandırılmıştır.

Kaldıraç noktaları $\mathbf{x_i}$ bağımsız değişken uzayının dışında kaldığından $\mathbf{x}$ bileşenlerini analiz ederek dikey aykırı (y-yönündeki) değerlerden bu kaldıraç noktaları ayırt edilebilir. Bunun için minimum hacimdeki elipsoid (MVE) konum ve ölçek kestirimine dayalı sağlam uzaklıklar hesaplanır.

Bu uzaklığın karesi, $\mathbf{x_i}$' ler arasında aykırı değer olmadığında ki-kare dağılımına sahip olduğundan, $RD(x_i) > \sqrt{\chi^2_{p,\,0.975}}$ ise gözlem kaldıraç noktası olarak sınıflandırılacaktır. Bu bilgi standartlaştırılmış LTS artığı ile birleştirildiğinde Şekil 2.6' daki grafik elde edilir (Rousseeuw and Hubert, 1997).

Şekil 2.6 Sağlam uzaklıklara karşı sağlam standartlaştırılmış artık grafiği

Diğer istatistiksel aykırı değer teşhis yöntemleri bir gözlem değerin silinmesinden sonra regresyon katsayılarının tekrar hesaplanması prensibine dayanır. Bu istatistikler analizden bir gözlemin çıkarılmasından sonra tüm regresyon katsayılarındaki değişimi ölçer. Bu teşhis istatistikleri yüksek kaldıraç noktalı aykırı değerlerin belirlenmesinde özellikle yararlıdır ve silme (deletion) istatistikleri olarak ifade edilir. Bir kaç silme istatistiği Cook uzaklığı, student artıkları, jacknife artıkları olarak ifade edilebilir. Tek gözlemin silinmesine ilişkin istatistikler iki değişkenli durumda çok faydalıdırlar fakat çok değişkenli aykırı değerlerin bileşik etkilerinin belirlenmesinde daha az kullanışlıdırlar. Artık çizimlerine dayalı aykırı değerlerin belirlenmesine ilişkin grafik metotları vardır. Bazı artık çizim teşhisçileri artıkların histogramı, artıkların uyum kestirim değerlerine karşı çizimi $(e, \hat{y})$, bağımlı değişkenin uyum çizimi değerlerine karşı çizimi $(y, \hat{y})$ veya çok değişkenli durumda kullanılan kısmi artık çizimini içerir. Bu metotlar aykırı değerleri belirlemede kullanışlıdırlar. Fakat ilgili aykırı değerler x-yönünde ve kötü kaldıraç noktasıysa problem oluşturabilirler. Kötü kaldıraç noktası olan x-yönündeki aykırı değerler regresyon doğrusunu küçük artıklara sebep olan bu noktalara doğru çekerler. Sonuç olarak; artıklara dayalı teşhis yöntemleri aykırı değer olarak isimlendirilen bu gözlem değerlerini ortaya çıkarmada başarısız olabilir. Bununla birlikte grafiksel teşhis yöntemleri belirli durumlardaki aykırı değerleri belirlemede yardımcıdır.

Ayrıca çok sayıda bağımsız değişken varsa her bir bağımsız değişkene karşı artıkların çizimi zor ve zaman kaybı olabilir (Rousseeuw and Leroy, 1987).

2.4.2 Etkin gözlemin teşhis ölçüleri

Bir gözlemin aykırı değer olup olmadığı ve aykırı değerse x-yönünde veya y-yönünde olup olmadığı saptandıktan sonra gelecek adım aykırı gözlemlerin etkili olup olmadığının belirlenmesidir (Kutner, Nachtsheim, Neter and Li, 2005). Çok değişkenli doğrusal modeldeki bir gözlemin etkili olup olmadığı nasıl belirlenir? Cevabı öncelikle grafiksel olarak belirlemektir. Bununla birlikte serpilme grafiği aracılığıyla etkin gözlemlerin belirlenmesi sadece bir veya iki bağımsız değişkenli regresyon için söz konusudur.

Bir bağımsız değişkendeki etkin gözlemler çoklu regresyon modelinde etkin olamayabileceğinden ve diğer taraftan çoklu regresyon modelindeki bazı etkin gözlemler de tek değişkenli modellerde belirlenemediğinden çok sayıdaki bağımsız değişkenli çoklu regresyon için etkin değildir (Kutner, Nachtsheim, Neter and Li, 2005). Gözlem değerleri regresyon modelinde farklı etkilere sahiptir. Farklı etkinin anlamı bazı gözlem değerlerinin regresyon katsayılarının kestiriminde diğer gözlem değerlerine göre daha büyük bir etkiye sahip olmasıdır.

2.4.2.1 Welsch-Kuh uzaklığı (DFFITS ölçüsü)

Veri setinden bir gözlem silindiğinde kestirim üzerindeki etkiyi ölçen bir teşhis ölçüsüdür. Veri setinin tamamı için kestirilen $\hat{y}_i$ ile veri setinden *i.* gözlemin çıkarıldığı $\hat{y}_{(i)}$ arasındaki standartlaştırılmış fark eşitlik (2.9)' daki gibi verilir:

$$WK_i = DFFITS_i = \frac{\hat{y}_i - \hat{y}_{(i)}}{s_{(i)}\sqrt{h_i}}, \qquad i = 1, \ldots\ldots, n \tag{2.9}$$

Buradaki $s_{(i)}$, *i.* gözlemin yer almadığı model için örneklem standart sapması, $\hat{y}_{(i)}$ *i.* gözlem olmadığı durumda elde edilen y_i' nin kestirilmiş değeri ve h_i ise *i.* kaldıraç değeridir. Orta büyüklükteki veri setleri için $|DFFITS_i| > 1$ ise *i.* gözlem ($\hat{y}$ uyum kestirimi üzerinde etkili olduğu) etkin gözlem olarak değerlendirilir ve büyük veri setleri için de $|DFFITS_i| > 2\sqrt{\frac{k'}{n}}$ tercih edilmektedir.

DFFITS istatistiği eşitlik (2.10)' daki gibi de elde edilebilir:

$$DFFITS_i = \left[\frac{h_i}{1-h_i}\right]^{\frac{1}{2}} . \frac{e_i}{s_{(i)}\sqrt{1-h_i}}, \qquad i = 1,2,....,n \qquad (2.10)$$

Bu istatistiğin, şapka matrisinin köşegen öğesi olan *i.* kaldıraç ve r-student türü artıklardan etkilendiği görülmektedir (Belsley, Kuh and Welsch, 1980).

2.4.2.2 <u>Cook uzaklığı (CD)</u>

Bir diğer etkin gözlem teşhis ölçütü ise;

$$CD_i = \frac{(\hat{\beta} - \hat{\beta}_{(i)})'(X'X)(\hat{\beta} - \hat{\beta}_{(i)})}{k'(AKO)}, \qquad i = 1,2,....,n \qquad (2.11)$$

biçiminde ifade edilir. Cook uzaklığı regresyon katsayılarındaki değişimi ölçer.

CD_i değeri $F_{\alpha,k'(n-k')}$ kritik değeri ile karşılaştırılır. CD_i' nin kritik değerden büyük olduğu durumlar için *i.* gözlemin etkin gözlem olduğu kabul edilir. $F_{0,05,k'(n-k')} = 1$ olduğunda $CD_i > 1$ değerlerin etkin olduğu ifade edilmiştir (Montgomery, Peck and Vining, 2001).

Cook uzaklığının farklı bir hesaplama şekli

$$CD_i = \left(\frac{r_i^2}{k'}\right)\left(\frac{h_{ii}}{1-h_{ii}}\right), \qquad i = 1,2,...,n \tag{2.12}$$

olarak tanımlanmaktadır. Buradaki r_i standartlaştırılmış artıkları, h_{ii} şapka matrisinin köşegen elemanını, k' ise parametre sayısını vermektedir (Stevens, 2002).

2.4.2.2 DFBETAS ölçüsü

Belsley ve arkadaşlarının 1980' de tanımladığı bu teşhis istatistiği *i.* gözlemin çıkartılmasıyla sadece *j.* kestirilmiş regresyon katsayısı üzerindeki değişime dayanır (Rousseeuw and Leroy, 2003).

Tam model için kestirilen *j.* regresyon katsayısı $\left(\hat{\beta}_j\right)$ ile *i.* gözlemin atıldığı model için kestirilen *j.* regresyon katsayısı $\hat{\beta}_{j(i)}$ arasındaki standartlaştırılmış fark aşağıdaki gibidir:

$$(DFBETAS)_{j(i)} = \frac{\hat{\beta}_j - \hat{\beta}_{j(i)}}{s_{(i)}\sqrt{d_j}}, \quad i = 1,....,n \quad j = 0,....,k \tag{2.13}$$

Burada d_j, $(\mathbf{X'X})^{-1}$ matrisinin *j.* köşegen elemanıdır. Eğer küçük veya orta büyüklükteki veri setlerinde tüm veya neredeyse tüm j' ler için $\left|(DFBETAS)_{j(i)}\right| > 1$ ve büyük veri setlerinde de $\left|(DFBETAS)_{j(i)}\right| > \frac{2}{\sqrt{n}}$ ise *i.* gözlem etkin gözlem olarak değerlendirilir (Montgomery, Peck and Vining 2001).

2.4.2.3 Düzeltilmiş Cook uzaklığı $\left(CD^{*}\right)$

Welsch ve Kuh (1977) tarafından ilk olarak ve ardından da Atkinson (1981a) tarafından önerilen Cook Uzaklığı' nın düzeltilmiş hali olan bu istatistik aşağıda verilmiştir:

$$CD_i^{*} = \left|t_i\right|\left[\left(\frac{n-k'}{k'}\right)\left(\frac{h_{ii}}{1-h_{ii}}\right)\right]^{1/2} = \left|DFFITS_i\right|\left(\frac{n-k'}{k'}\right)^{1/2} \qquad \text{i=1,2,....,n} \qquad (2.14)$$

her bir CD_i^{*} değeri, $2\left[\left(n-k'\right)/n\right]^{1/2}$ değeri ile karşılaştırılarak, bu değerden büyük olan CD_i^{*} değerine sahip gözlemler etkin gözlem olarak değerlendirilir. Düzeltilmiş Cook uzaklığı, aşırı değerlere (extreme values) daha fazla önem vermesi, grafiksel gösterimler için daha uygun olması ve $h_{ii} = k'/n$, $i = 1,....,n$ olması durumunda düzeltilmiş Cook uzaklığının grafiğinin $\left|t_i\right|$ ' nin grafiğiyle aynı olması avantajlarına sahiptir (Ravishanker and Dey, 2001).

2.4.2.5 COVRATIO istatistiği

Belsley, Kuh ve Welsch (1980) tarafından önerilen bu istatistik kovaryans matrisi üzerindeki *i.* gözlemin etkisini ölçer; bu amaçla *i.* gözlemin veri setinden çıkartılmasıyla kestirilmiş katsayıların kovaryans matrisinin determinantını tüm gözlemlerin ele alındığı katsayıların kovaryans matrisinin determinantına oranlayarak karşılaştırır.

COVRATIO istatistiği aşağıda verildiği gibi hesaplanır:

$$COVRATIO_i = \frac{\det\left\{ s_{(i)}^2 \left(X'_{(i)}X_{(i)}\right)^{-1} \right\}}{\det\left\{ AKO\left(X'X\right)^{-1} \right\}}, \qquad i = 1,2,....,n \qquad (2.15)$$

Bu istatistikte $\left|COVRATIO_i - 1\right| > 3k'/n$ değerlerine sahip gözlemler etkin gözlem olarak ifade edilirler (Chatterjee and Hadi, 1986). $COVRATIO_i > 1$ ise *i.* gözlemin kestirimin kesinliğini artırdığı, $COVRATIO_i < 1$ ise kestirimin kesinliğini azalttığı ifade edilmektedir (Montgomery, Peck and Vinning, 2001).

COVRATIO istatistiği aşağıda verildiği gibi de hesaplanabilir:

$$COVRATIO_i = \left(\frac{n - k' - r_i^2}{n - k' - 1} \right)^{k'} \Big/ \left(1 - h_{ii}\right), \quad \text{i=1,2,.....,n} \qquad (2.16)$$

Eşitlik (2.16) incelendiğinde COVRATIO istatistiğinin, yüksek kaldıraç noktası ve student artıkları ile yakından ilişkisi olduğu görülmektedir.

2.4.2.6 DFTSTAT istatistiği

Veri setinden *i.* gözlem satırı silindiğinde regresyon katsayılarına ilişkin duyarlılığı değerlendirmek, diğer bir deyişle *j.* regresyon katsayısının sıfıra eşit olup olmadığını sorgularken elde edilen *t*-istatistik değerindeki değişimi incelemek yararlı olabilir. Bu amaçla geliştirilen DFTSTAT istatistiği;

$$DFTSTAT_{j(i)} = \frac{\hat{\beta}_j}{\sqrt{AKO\left(C_{jj}\right)}} - \frac{\hat{\beta}_{j(i)}}{\sqrt{s_{(i)}^2 \left(\left(X'_{(i)} X_{(i)}\right)^{-1}\right)_{jj}}}, \quad i = 1,2,....,n \qquad (2.17)$$

biçiminde hesaplanır. Büyük $\left|DFTSTAT_{j(i)}\right|$ değerleri için *i.* gözlemin etkin olduğu ifade edilmektedir (Belsley, Kuh and Welsch, 1980).

2.4.2.7 WELSCH uzaklığı (W)

Welsch (1982) tarafından önerilen bu istatistik,

$$W_i = |DFFITS_i| \sqrt{\frac{n-1}{1-h_{ii}}}, \quad i = 1,2,....,n \tag{2.18}$$

biçiminde hesaplanır. n>15 için bu istatistik değerleri, $3\sqrt{k'}$ ile karşılaştırılır. Bu değerden büyük olan W_i değerine sahip gözlemler etkin gözlem olarak değerlendirilmektedir. Ayrıca bu istatistiğin $DFFITS_i$ ve h_{ii}' den daha duyarlı olduğu ifade edilmiştir **(**Chatterjee and Hadi**,** 1986).

2.4.2.8 PRESS istatistiği

Allen (1971-1974) tarafından önerilen bu istatistik regresyon modellerinin karşılaştırılmasında kullanılır ve önkestirim hata kareler toplamı (Prediction Error Sum of Squares, PRESS) olarak isimlendirilir (Montgomery, Peck and Vinning, 2001). Etkin gözlemlere dikkat çekmesi ve veri seti üzerinden farklı kestirim yapılan modellerin durağanlığı konusunda ayrıntılı bilgi sunma avantajına sahiptir (Draper and Smith, 1981).

PRESS İstatistiği,

$$PRESS = \sum_{i=1}^{n} e_{(i)}^2 = \sum_{i=1}^{n} (y_i - \hat{y}_{i(i)})^2 = \sum_{i=1}^{n} \left(\frac{e_i}{1-h_{ii}} \right)^2 \tag{2.19}$$

biçiminde hesaplanır. Küçük değerli PRESS istatistiğine sahip model arzu edilir (Montgomery, Peck and Vinning, 2001).

2.4.2.9 Andrews-Pregibon istatistiği (AP)

Andrews ve Pregibon (1978) tarafından önerilen bu istatistik;

$$AP_i = \frac{\det\{Z'_{(i)} Z_{(i)}\}}{\det\{Z'Z\}}, \qquad i = 1,2,....,n \tag{2.20}$$

biçiminde hesaplanır (Jarrell, 1991). Veri setinden *i.* gözlemin çıkartılması durumunda $\det\{Z'Z\}$ değerlerindeki göreli değişimi değerlendirmek için kullanılır (Chatterjee and Hadi, 1986).

Eşitlik (2.21)' de belirtilen **Z** matrisi,

$$Z = \begin{bmatrix} 1 & x_{11} & x_{12} & . & . & . & x_{1k} & y_1 \\ 1 & x_{21} & x_{22} & . & . & . & x_{2k} & y_2 \\ . & . & . & & & & . & . \\ . & . & . & & & & . & . \\ . & . & . & & & & . & . \\ 1 & x_{n1} & x_{n2} & . & . & . & x_{nk} & y_n \end{bmatrix} \tag{2.21}$$

biçiminde tanımlanmaktadır.

$Z_{(i)}$ ise eşitlik (2.21)' de tanımlanan matrisin *i.* satırının çıkartılmasıyla elde edilen matristir. Diğer gözlemlerle karşılaştırıldığında AP_i' nin küçük değerleri için ilgili gözlemin etkin gözlem olduğu ifade edilmektedir (Rousseeuw and Leroy, 2003).

Andrews-Pregibon İstatistiği,

$$AP_i = 1 - h_{ii} - \frac{e_i^2}{e'e}, \qquad i = 1,2,....,n \tag{2.22}$$

biçiminde de hesaplanabilir.

Eşitlik (2.22)' de AP_i değerlerinin x-yönündeki yüksek kaldıraç noktaları ile y-yönündeki aykırı değerleri ayırt edemediği belirtilmektedir (Chatterjee and Hadi, 1986).

2.4.3 Çok sayıda etkin gözlemin çıkarılmasına dayalı ölçüler

Kısım 2.4.2' de tek bir gözlem satırının çıkartılmasına dayalı etkin gözlemlerin belirlenmesi için geliştirilen teşhis istatistikleri verilmiştir. Birden fazla sayıda etkin gözlemlerin varlığı durumunda bu istatistikler yetersiz kalmaktadır. Bu amaçla geliştirilen çoklu gözlemlerin çıkarılmasına dayalı istatistiklere aşağıda yer verilmiştir.

2.4.3.1 MDFFITS istatistiği

Etkin gözlem olduğundan şüphelenilen m tane gözlem satırının veri setinden çıkartılmasının ardından önkestirimdeki değişimi gösteren çoklu teşhis istatistiği aşağıdaki gibi verilmiştir:

$$MDFFITS_{(D_m)} = \left[\hat{\beta} - \hat{\beta}_{(D_m)}\right]' \mathbf{X}'_{(D_m)} \mathbf{X}_{(D_m)} \left[\hat{\beta} - \hat{\beta}_{(D_m)}\right] \tag{2.23}$$

Burada D_m, m sayıda veri setinden çıkartılan gözlemlerden oluşan kümeyi; $\hat{\beta}_{(D_m)}$ m sayıdaki gözlem satırı çıkartıldıktan sonra elde edilen parametre kestiricilerinin oluşturduğu sütun vektörünü, $\mathbf{X}_{(D_m)}$ ise m sayıda gözlemin çıkartıldıktan sonra kalan $\mathbf{X}$ matrisini göstermektedir.

Burada m sayısının hesaplama zorluğundan dolayı yaklaşık olarak 20 gözlemi aşmaması tavsiye edilmektedir. MDFFITS istatistiğinin büyük değer alması, ilgili gözlemlerin etkin gözlem olduğu hakkında fikir vermektedir (Belsley, Kuh and Welsch, 1980).

2.4.3.2 Çoklu COVRATIO istatistiği

Tek gözlemin silinmesine dayalı COVRATIO istatistiği çok sayıda gözlemin çıkartılmasını sağlayacak biçimde genelleştirilerek çoklu COVRATIO istatistiği elde edilir:

$$COVRATIO_{(D_m)} = \frac{\det\left\{s^2_{(D_m)}\left(\mathbf{X}'_{(\mathbf{D_m})}\mathbf{X}_{(\mathbf{D_m})}\right)^{-1}\right\}}{\det\left\{AKO\left(\mathbf{X}'\mathbf{X}\right)^{-1}\right\}}, \qquad i = 1, 2, ..., n \tag{2.24}$$

Burada $s^2_{(D_m)}$, m sayıda gözlemin veri setinden çıkartılarak elde edilen varyans kestiricisidir. Çoklu COVRATIO istatistiğine göre çıkarılan gözlem gruplarının her biri diğer gözlem grupları ile karşılaştırıldığında çok büyük ya da çok küçük değerler veren gözlem grubunda bulunan gözlemler etkin gözlem olarak ifade edilmektedir.

2.4.3.3 Çoklu Andrews-Pregibon istatistiği (AP)

Çoklu Andrews Pregibon istatistiği (2.25)' deki gibi hesaplanmaktadır:

$$AP_{(D_m)} = \frac{\det\left\{\mathbf{Z}'_{(\mathbf{D_m})}\mathbf{Z}_{(\mathbf{D_m})}\right\}}{\det\left\{\mathbf{Z}'\mathbf{Z}\right\}} \tag{2.25}$$

Andrews ve Pregibon (1978) tarafından önerilen bu istatistik tek gözlemin veri setinden çıkartıldığı Andrews-Pregibon istatistiğinin genelleştirilmiş halidir.

Çoklu gözlem kümelerinden elde edilen her bir değer birbiri ile kıyasladığında küçük çoklu Andrews-Pregibon değerine sahip gözlem kümesinde bulunan gözlemler etkin gözlem olarak tanımlanır (Belsley, Kuh and Welsch, 1980).

BÖLÜM 3

SAĞLAM REGRESYON

Huber (2004) sağlam kelimesini pek çok anlamla (bazen tutarsız olan) yüklü olduğunu ifade etmiştir. Bununla birlikte bir kestiriciyi değerlendirmede sağlamlığın iki türünün değerlendirilmesi gerektiği kabul edilir. Bunlar Mosteller ve Tukey (1977) tarafından özetlenmiştir. Çalışmalarında sağlam bir kestiricinin iki koşulu sağlaması gerektiği ifade edilmiştir. Bu koşullardan ilki gözlem değerlerinde küçük bir değişim söz konusuysa kestirimde önemli bir değişime neden olmaması; ikincisi kestirim çeşitli koşullar altında etkin olmasıdır. Kuşkulu gözlemler için kestiricinin sağlamlılığını gösteren ilk koşul "geçerliliğin sağlamlılığı" (robustness of validity) olarak değerlendirilebilir. Diğer bir deyişle kestirici gözlemlerin büyük bir kısmı için geçerli bir kestirim sunar. Dağılımsal varsayımlarla ilişkili ikinci koşul "etkinliğin sağlamlılığı" (robustness of efficiency) olarak düşünülebilir. Bu koşul, kestiricinin dağılımsal varsayımlarının sağlanmasındaki başarısızlığı durumunda kestirimlerin kesinliği veya kestirimlerin standart hataları üzerinde küçük etkinliğe sahip olduğunu ima eder. Sağlam regresyon metotları hem geçerliliğin sağlamlılığı hem de etkinliğin sağlamlığıyla ilgilidir. Bu metotlar tüm gözlem değerlerindeki bilgiyi kullanır. Fakat kuşkulu olan gözlemlere daha az ağırlık verir. Farklı sağlam regresyon türleri kuşkulu artıklar, kaldıraç noktaları veya her ikisinin birleşimiyle ilgilenir. Bu metotların çoğu hem hata dağılımının uzun kuyruklu olduğu durumda hem de normal dağılımlı olması durumunda oldukça etkin kestirimler sunar.

Sağlam regresyon olarak ifade edilen metotlar etkinlik ile çok az ilgilidir. Bu metotların temel amacı regresyon eğim değerlerini önemli şekilde etkileyen kuşkulu gözlemleri önlemektir. Sağlam metotlar, sadece kuşkulu gözlemlere düşük ağırlık vermekten ziyade çoğu kez bu gözlemleri analizden tamamiyle çıkarılması kriterlerini oluştururlar.

Sağlam regresyon hata teriminin normal dağılmaması durumunda ve/veya aykırı değerlerin regresyon denklemini etkilediği durumda EKK' ya alternatif olarak kullanılan bir metottur ve aykırı değerlerin analizden çıkarılıp çıkarılmayacağı hakkında bize bilgi verir. En küçük kareler yaklaşımı veri setinde aykırı değerlerin yer aldığı durumlarda kararsız kestirimler üretebilmektedir. Sonuç olarak sağlam kestirimler elde etmek için bir regresyon yaklaşımına ihtiyaç duyulur. Sağlam regresyon aykırı değerlerden fazla etkilenmemiş regresyon katsayılarını belirlemeye çalışır (Ryan, 1997). Sağlam regresyon analizinin başlıca amaçları verinin çoğunluğunu bir modele uydurmaya çalışmaktır. Diğer önemli bir amacı ise model dağılımlarından küçük sapmalara karşı yeni istatistiksel yöntemleri oluşturmaktır (Bhar, 2010). Burada istenen, aykırı değerlerin varlığında sağlam sonuçlar elde etmek ve bunu başarmak için aykırı değerlerin etkilerini sınırlamaktır. x-yönündeki, y-yönündeki, hem x hem y yönündeki aykırı değerlerin yarattığı sorunlara karşı duyarsız regresyon teknikleri mevcuttur (Chen, 2010).

Sağlam bir kestiricinin istenilen iki özelliği dayanıklılığı (dirençliliği) ve sağlamlığıdır. Bu kavramlar Huber(1972), Hampel(1974), Mosteller ve Tukey (1977) tarafından ayrıntılı olarak incelenmiştir. Bir kestirici, eğer az sayıdaki büyük hatalardan veya az sayıdaki küçük yuvarlamalardan ve gruplanmış hatalardan bir dereceye kadar etkileniyorsa dayanıklı olarak nitelendirilir. Bir kestirici, eğer örneklemin küçük bir alt kümesi kestirim üzerinde oransız bir etkiye sahip değilse bu kestirici büyük hatalara karşı dayanıklıdır.

Bir kestirici eğer küçük hatalardan etkileniyorsa, ayrıca bu kestirici gözlemlerin küçük bir kısmının yuvarlanması/gruplanmasıyla belirlenmemiş ise bu kestirici grupsal hatalara/yuvarlama hatalarına karşı dayanıklıdır. Bir kestiricinin varyansı veya yanlı kestiriciler için HKO' sı her dağılım için minimuma yakınsa o kestirici sağlamdır. Sağlamlık tam olarak bilinmeyen bir dağılımdan çekilen tekrarlı örneklemler için kestiricinin iyi olduğunu garanti eder. Ayrıca örneklemin bozulması (contaminated) durumunda kestirim çok küçük miktarda değişmelidir. Bozulma durumu ya büyük hatalardan ya da gözlemler arasındaki yuvarlama ve gruplama hatalarından kaynaklanabilir (Hoaglin, Mosteller and Tukey, 1983).

Tahmin edicilerin sağlamlığı (model varsayımlardan küçük sapmaların tahminleri etkilememesi) ve dağılımsal sağlamlık (gerçek dağılım şeklinden varsayılan modelin az miktarda sapma göstermesi) düşünüldüğünde tercih edilen genellikle dağılımsal sağlamlıktır. Sağlam istatistiğe olan ilgi eski yunan tarihine kadar uzanır. Yunanlılar daha sağlam olduğu için ortalama yerine medyanı kullanmayı tercih etmişlerdir. Sağlamlık kavramını kullanan çok sayıda istatistikçi (Newcomb, K.Pearson, Gosset, Jeffreys ve E.S. Pearson) olmasına rağmen sağlam tahmin teorisi özellikle son 20 yılda ciddi bir şekilde gelişmiş ve formüle edilmiştir (Hampel, Ronchetti, Rousseeuw and Stahel, 1986).

Sağlam kestirim yöntemlerinin sahip olması gereken özellikler aşağıdaki gibi sıralanabilir;

1. Yöntem varsayılan model altında oldukça iyi (en iyi veya yaklaşık en iyi) etkinliğe sahip olmalıdır.

2. Model varsayımlarından küçük sapmalar yöntemin performansına çok az zarar vermelidir.

3. Model varsayımlarından daha büyük sapmalar çok kötü sonuçlara neden olmamalıdır.

Gözlemlerin küçük bir kısmında ortaya çıkan büyük hataların varlığı küçük bir sapma olarak değerlendirilir. Fakat bazı klasik yöntemlerin bu sapmalara karşı çok fazla duyarlı olması bu yöntemler açısından dezavantaj olmakta, sağlam yöntemlerin büyük hatalara karşı duyarsız olması ise avantaj sağlamaktadır (Huber, 1981).

Sağlam yöntemlerin amaçları aşağıdaki gibi özetlenebilir;

1. Veri setinin çoğunluğunu kullanarak en iyi uyan model yapısını tanımlamak,

2. Gözlem değerlerinden sapan aykırı değerleri belirlemek,

3. Yüksek etkin gözlem değerleri (kaldıraç noktalarını) belirlemek,

4. Gözlemlerin bağımsız olduğu düşüncesiyle varsayılan ilişki yapılarından sapmaları belirlemek (Hampel, Ronchetti, Rouseeuw and Stahel, 1986).

3.1 Sağlam Regresyon Kestiricisine İlişkin Bazı Temel Kavramlar

Sağlam regresyon kestiricisine ilişkin kullanılan temel kavramlardan bazıları; kırılma noktası, göreli etkinlik ve sınırlı etkidir. Etkinlik, kırılma noktası ve sınırlı etkinin özellikleri teorik anlamda sağlam tekniklerin performansını ölçeklendirerek tanımlamada kullanılır (Bhar, 2010).

3.1.1 Kırılma (bozulma) noktası

İlk olarak Hodges (1967) tarafından kullanılan kırılma noktası kavramı tek boyutlu konum kestirimi ile sınırlıyken, Hampel (1971) daha genel bir tanım vermiştir. Ancak bu kavram Donoho ve Huber (1983) tarafından yaygın bir şekilde tanınır hale getirilmiştir (Rousseuw and Leroy, 2003). Bir kestiricinin kırılma noktası kestiriciyi kullanışsız hale getiren aykırı değer sayısıdır. Genellikle yüzde olarak ifade edilir. Bazı regresyon kestiricileri mümkün en küçük kırılma noktası olan $1/n$ veya % 0' a sahiptir. Aritmetik ortalamanın kırılma noktası sıfır, medyanın kırılma noktası 0,5 dir. EKK kestiricilerinin kırılma noktası ise sıfırdır. Bu nedenden dolayı bir tek aykırı değer EKK sonuçlarını anlamsız yapmaya yetecektir. Başka sağlam regresyon kestiricileri de olası en yüksek kırılma noktası olan $n/2$ veya % 50' ye sahiptir. Bir kestirim yönteminin diğerine göre güvenilirliği kırılma noktası ile ölçülür, kırılma noktası büyük olan yöntem daha güvenilir olarak ifade edilmektedir.

Sağlam kestirim tekniği %50 kırılma noktasına sahipse, bu durumda verilerin % 50' si aykırı değer içerse bile katsayılar hala kullanılabilir durumda olacaktır. Şayet kırılma noktası % 50’ yi aşar ise aykırı değer ile normal gözlemler birbirinden ayırt edilememektedir (Hampel, Ronchetti, Rousseeuw and Stahel, 1986). Sağlam kestiriciler için kırılma noktası % 0’ dan fazla olmalıdır. En yüksek beklenen kırılma noktası değeri % 50’ dir. Yüksek kırılma noktasına sahip bir sağlam kestirici diğer düşük kırılma noktasına sahip kestiricilere göre daha iyi bir kestirici olarak ifade edilmektedir (Heikkila, 2010).

Sonlu bir örneklem için kırılma noktasının açıklanması amacıyla n tane birimden oluşan bir Z örneklemi olduğu ve $T_n(Z_1,Z_2,\ldots,Z_n)$ ise θ parametreli P olasılık dağılımına sahip bir kestirici olduğu varsayılsın.

Diğer bir deyişle aşağıdaki gibi ifade edilsin.

$$T(Z)=\hat{\theta}\,. \tag{3.1}$$

Z örneklemindeki m tane gözlem değeri çıkarılıp, yerine m tane keyfi bozulmuş (yani, veri setinin geneline uymayan) değerler konularak Z' örneklemi elde edilsin. Böyle bir bozulumun neden olduğu en büyük etki (yan) aşağıdaki gibi ifade edilebilir:

$$etki(m;T,Z)=\sup_{Z'}\left\|T(Z')-T(Z)\right\| \tag{3.2}$$

Etki (m;T,Z) sonsuz ise, bu m aykırı değerin T kestiricisi üzerinde çok büyük etkisi olduğunu gösterir, diğer bir deyişle kestiricinin bozulduğu, verinin geri kalanını yeterince temsil edemediği söylenir. Daha genel bir ifadeyle, sonlu bir Z örneklemi için T kestiricisinin kırılma noktası aşağıdaki gibi tanımlanır:

$$\varepsilon_n^*(T,Z)=\min\left\{\frac{m}{n}:etki(m;T,Z)\,sonsuz\,iken\right\} \tag{3.3}$$

Hampel vd.(1986) veri setinin büyük çoğunluğunu temsil eden genel yapıdan sapan yaklaşık %10 kadar gözlemin veri setinde bulunabileceğini ve sağlam bir tahmin edicinin en azından %10 kırılma noktasına sahip olması gerektiğini belirtmiştir (Andersen, 2008).

3.1.2 Göreli etkinlik

Sağlam regresyonda bir başka önemli kavram göreli etkinliktir. Bir sağlam kestiricinin etkinliği EKK normal dağılımlı hatalara sahip olduğunda bir kestiricinin EKK’ ya benzer sonuçlar üretme derecesi olarak tanımlanır. Örneğin; sağlam regresyon analizi normal dağılımlı bir veri setine uygulanırsa bu veriler için EKK uygun kestirici olduğundan EKK’ ya mümkün olduğunca benzer olan sonuçları üreten bir kestiricinin olması istenir.

Etkinlik genellikle % olarak ifade edilir. Teorik olarak etkinlik 0 ile 1 değişim aralığına sahiptir. Fakat sağlam hata kareler ortalamasının EKK hata kareler ortalamasından (HKO) daha küçük olması durumunda değişim aralığı 1' i aşabilir. Yüksek yüzdelikler arzu edilir; diğer bir ifade ile etkinliğin yüksek yüzdeliğe sahip olması istenir. Göreli etkinlik terimi sağlam tekniğindeki hata kareler ortalamasının EKK hata kareler ortalamasına bölünmesiyle hesaplanır. EKK tekniğinden elde edilen HKO’ sı aykırı değer içermeyen veri setinden tahmin edilir (Ryan, 1997).

Göreli etkinlik aşağıdaki gibi ifade edilir:

$$Göreli\ \ Etkinlik = \frac{HKO_{Robust}}{HKO_{EKK}} \tag{3.4}$$

Daha genel bir ifadeyle, evren parametresi θ için T_1 ve T_2 olmak üzere iki tahmin edici olsun. T_2 tahmin edicisinin göreli etkinliği, onun HKO’ nın T_1 tahmin edicisinin HKO’ ya oranlanmasıyla belirlenir:

$$Etkinlik(T_1, T_2) = \frac{E\left[(T_2 - \theta)^2\right]}{E\left[(T_1 - \theta)^2\right]} \tag{3.5}$$

Doğrusallık, sabit varyans ve hata terimlerinin ilişkisiz olması varsayımları sağlanıyorsa, EKK tahminleri yansız doğrusal kestiriciler arasında en etkin olanıdır. Sonuç olarak, sağlam kestiricilerin göreli etkinliği bu koşullar altında EKK kestiricileriyle karşılaştırılarak değerlendirilmektedir (Andersen, 2008).

3.1.3 Sınırlı etki

Sınırlı etki kavramı kestiricinin, belirli bir artığın aldığı ağırlığı sınırlandırabilme yeteneğini (kabiliyetini) ifade eder. Böylece gözlem değerlerinin regresyon denklemi üzerindeki etki miktarının kontrolü söz konusudur. Özellikle x-ekseni üzerinde yer alan bazı aykırı değerler regresyon denklemindeki katsayı kestirimi üzerinde yüksek derecede etkiye sahip olabilirler. Bir kestiricinin sınırlı etki fonksiyonuna sahip olması istenir, böylece gözlem değerlerindeki aykırı değerlerin etkisi sınırlandırılır (Birkes and Dodge, 1993).

3.2 Sağlam Regresyonun İstenen Özellikleri

Regresyon analizinde, bilinmeyen β parametrelerinin kestirimleri standart hatalarının kestirimleri ve parametre vektörüne ait hipotez testlerinin yapılması istenebilir. Hataların bağımsızlığı normal dağılıma sahip olması $\varepsilon_i \sim N(0,\sigma^2)$ bağımsız değişkenlerin doğrusal bağımsızlığı gibi varsayımlardan bir veya daha fazlasının ihlali EKK kestiricisi $\hat{\beta}$ ' da veya standart hata kestiriminde büyük değişime neden olabilir.

Bu nedenlerden dolayı β ' nın sağlam kestiricilerinin istenen özellikleri:

1- Model varsayımlarının ihlali söz konusu değil ise kestiricinin tutarlı, asimtotik normal, yüksek etkinliğe sahip olmalı ve neredeyse EKK kadar iyi sonuçlar vermeli,

2- Bilinmeyen parametreler için güven aralıklarının ve hipotez testlerinin oluşturulmasını sağlayacak yöntemleri içermeli,

3- Model varsayımlarından hafif sapmalar sözkonusu olduğunda 1' deki özelliklerden ve 2' deki sonuçlardan göreli duyarsız olmalı (EKK' dan daha iyi sonuç vermeli),

4-Teorisi basit (kolay anlaşılabilir) olmalı,

5-Hesaplanması kolay olmalıdır (Staudte and Sheather, 1990).

3.3 Bazı Sağlam Regresyon Metotları

Hataların normal dağılım göstermesi durumunda, EKK tekniğini kullanmak gerekir. Ancak hataların normal dağılmadığı durumlarda diğer metotlar dikkate alınmalıdır. Aykırı değer olarak nitelendirilebilen büyük artığa sahip bir gözlemi veri setinden çıkararak, EKK tekniği hala kullanılır kılınabilir. Ancak birden fazla büyük artığa sahip gözlemin var olması durumunda parametre katsayılarının kestiriminde hatalı sonuçlar elde edilebileceğinden bu gibi durumlarda sağlam regresyon alternatif bir metot olarak karşımıza çıkmaktadır (Faraway, 2009).

Bir veya daha fazla aykırı değer belirlendiyse onu derinlemesine inceleme yapmadan doğrudan veri setinden atmak büyük bir hata olabilir. Bir veri setinde sık sık etkin veya aykırı gözlemlere rastlamak mümkündür. İyi bir kestirim yapabilmek için bazen bu gözlemleri yenileri ile değiştirmek gerekebilir. Ayrıca aykırı değerler bazı olağandışılığı yansıtarak modelin belirlenmesinde bir gelişmeye neden olabilir. Etkin veya aykırı gözlem belirlendiğinde yapılacak ilk şey bu gözlemlerin dikkatlice incelenmesidir. Daha önce de belirtildiği gibi aykırı değerlerin var olmasına neden olan belli nedenler vardır, ölçeklendirme hatası, verilerin sınıflandırılmasındaki hata veya verilerin bilgisayara hatalı kayıt edilmesi gibi. Şayet bu hataları düzeltmede geçerli bir çare bulunamıyorsa haklı olarak o gözlemi veri setinden atmak gerekir. Olağandışı bir koşul veya durum ile bir gözlem bağlantılı olabilir.

Bir araştırmacı yaptığı araştırmalar sonunda gözlemin gerçek olduğundan eminse EKK tekniğinden vazgeçmesine gerek kalmaz, şayet araştırmacı bazı şüphelere sahip ise veriden kaynaklanan hatalı durumlar elde edilmesi sonucunda, diğer bir ifadeyle etkin veya aykırı bir gözlem belirlenmiş ise ya da hataların normal dağılmaması dışında bu gözlemlerin var olmasına neden olan diğer koşulların tamamı ortadan kaldırılmış ise uygun olan sağlam regresyon metodu seçilerek çalışma gerçekleştirilebilir (Kennedy, 2003). Aşağıda bazı sağlam regresyon metotlarına yer verilmiştir.

3.3.1 En küçük mutlak sapma (LAD)

EKK kestiricisi üzerinde çok fazla etkiye sahip olan gözlem veya gözlemler belirlenmiş ise bu değerlere karşı koruma sağlayabilecek farklı bir yaklaşım olan ve değerlere karşı EKK' ya göre daha az duyarlı olan bir metot kullanılabilir. İstatistikçilerin arasında kullanımı yaygın olan böyle bir metot en küçük mutlak sapma olarak isimlendirilir (Wooldridge, 2009).

Çoğu sağlam regresyon metodu son 20-30 yıl boyunca geliştirilmişken bir tanesi EKK' dan neredeyse 50 yıl önce geliştirilmiştir. 1805'lerde Fransa'da Legendre tarafından geliştirilen EKK metodundan neredeyse 50 yıl önce Boscovich 1757'de LAD metodunu önermiştir. 30 yıl sonra metodun teorik yapısını tamamlayan Laplace LAD' ı kullanmıştır. EKK artık karelerinin aykırı değerlere olan etkisi LAD' ın mutlak değerlerinin etkisinden daha büyük bir etkiye sahip olduğu için EKK' ya göre LAD' ın daha üstün olduğunu ileri sürmüştür. Bununla birlikte EKK geliştirilince LAD' a olan ilgi azalmıştır (Birkes and Dodge, 1993).

Aşağıda amaç fonkiyonu verilen LAD metodunda artıkların mutlak değerlerinin toplamını minimize eden katsayılar seçilir ve şu şekilde ifade edilir;

$$\min \sum_{i=1}^{n} |e_i| \tag{3.6}$$

LAD kestiricisi; y-yönündeki aykırı değerlere sağlam fakat x-yönündeki aykırı değerlere sağlam değildir. LAD' ın kırılma noktası 0' dır.

Bu nedenle x-yönündeki bir aykırı değerin etkisi regresyon doğrusunu kendisine doğru çekerek değiştirir. Bununla birlikte LAD sınırlı etkiye sahip bir kestirici değildir (Mosteller and Tukey, 1977).

LAD $L_1 - norm$ olarak bilinirken, EKK $L_2 - norm$ olarak bilinir. Her ikisi de $L_p - norm$ regresyonunun özel durumlarıdır,

$$\min \sum_{i=1}^{n} |\varepsilon_i|^p \tag{3.7}$$

ve burada p, 1 veya 2' dir.

3.3.2 Ağırlıklandırılmış en küçük mutlak sapma (WLAD)

Ağırlıklandırılmış En Küçük Mutlak Sapma (WLAD) regresyonu özellikle x-yönündeki aykırı değerlerin varlığında, yüksek kırılma noktalı kestiricilere rakip performansa sahip ve aynı zamanda hesaplaması kolay olan kestiricinin kırılma noktasını arttırmak için geliştirilmiştir.

Bu regresyon kestiricisi aşağıdaki amaç fonksiyonunu en küçüklemek için tanımlanır:

$$\min \sum_{i=1}^{n} w_i \left| y_i - x_i' \beta \right| \tag{3.8}$$

Bu kestirim problemini Giloni v.d (2006) doğrusal programlama olarak formüle etmişlerdir.

Hubert ve Rousseeuw (1997) ise hem sürekli hem de kesikli bağımsız değişkenleri içeren regresyon modelleri için LAD regresyonun ağırlıklandırılmış versiyonunu önermiştir. Fakat önerdikleri bu metot aynı zamanda sürekli bağımsız değişkenlere dayalı modeller için de kullanılabilmektedir. Ağırlıklar yüksek kırılma noktasına sahip minimum hacimli elipsoid (MVE) kestiricisine dayanmaktadır (Rousseeuw, 1984).

X matrisinin yarısını içeren minimum elipsoidin merkezi C ve kovaryans matrisi de **M** olmak üzere sağlam uzaklıklara (RD_i) dayalı ağırlıklar,

$$w_i = \min(1, k / RD(x_i)^2) \tag{3.9}$$

biçimde hesaplanır. Burada

$$RD(x_i) = \sqrt{(\mathbf{x_i} - \mathbf{C})\mathbf{M}^{-1}(\mathbf{x_i} - \mathbf{C})'} \tag{3.10}$$

biçiminde tanımlanır.

3.3.3 En küçük medyan kare (LMS)

Sağlam regresyon yöntemleri içerisinde en çok kullanılan LMS kestiricisinin amaç fonksiyonu artık karelerin toplamını en küçüklemek yerine kareleri alınmış artıkların medyanını enküçüklemeyi hedeflemektedir. LMS h. sıra artık kareleri minimize ederek hesaplanır. h = [n / 2] + [(k' + 1) / 2] iken burada n ve k' sırasıyla örneklem hacmi ve parametre sayısıdır. LMS ortalama değer yerine medyan değerinin kullanılması hariç EKK' ya benzer olarak düşünülebilir.

LMS' nin amaç fonksiyonu;

$$\min Med(e_i^2) \tag{3.11}$$

olarak ifade edilir.

Bu metod Rousseeuw (1984) tarafından önerilmesine rağmen daha sonra kendisi tarafından onaylanmamıştır (Ryan, 1997).

LMS aykırı değerlere karşı % 50' ye varan yüksek kırılma noktasına sahip bir metottur. LMS yüksek kırılma noktasına sahip metot olarak geliştirilmiş ve bu yöntem y-yönündeki aykırı değerlerin belirlenmesinde kullanılmıştır. Fakat parametre sayısının artışına bağlı olarak hesaplama güçlüğü ile karşılaşıldığı ifade edilmiştir. LMS yüksek göreli etkinliğe sahip değildir. Gerçekte LMS sıfır asimptotik göreli etkinliğe sahiptir. LMS artıkların medyan karelerini minimize eder ve geri kalan n -1 gözlem değerini yok sayar.

Bununla birlikte örneklem hacmi arttıkça LMS gözlem değerlerin geri kalanını yok saydığından regresyon katsayılarını EKK kadar iyi tahmin edemez. LMS genellikle tek başına kullanılan bir süreç olarak önerilmez. Aynı zamanda artıkların en küçük medyan karelerini üreten regresyon katsayılarını veya eğimi bulmada verilerin tüm alt kümelerini ele aldığından fazla sayıda işlem (hesap) gerektirir. Literatürde sıklıkla tartışılmasına rağmen sadece tarihsel bir öneme sahip olduğu düşünülmektedir (Rousseeuw and Leroy, 2003).

LMS tekniğinde ağırlıklar;

$$w_i = \begin{cases} 1 & ,|\varepsilon_i / s_0| \le 2,5 \\ 0 & ,|\varepsilon_i / s_0| > 2,5 \end{cases} \tag{3.12}$$

ve

$$s_0 = 1,4826\left[1 + \frac{5}{n-p}\right]\left[\sqrt{med\left(\overset{2}{\varepsilon_i}\right)}\right] \tag{3.13}$$

olarak hesaplanır (Rousseeuw and Leroy, 1987).

3.3.4 En küçük kırpılmış kareler toplamı (LTS)

Rousseeuw (1984) tarafından geliştirilen bir başka metot En Küçük Kırpılmış Kareler (LTS) regresyonudur. LTS kestiricisi aşağıdaki gibi bulunur:

$$\min \sum_{i=1}^{h} e_{(i)}^2 \tag{3.14}$$

Kırpılmış ortalamadan genişletilerek, LTS regresyonu artıkların kırpılmış kareler toplamını en küçükler. Burada $e_{(1)}^2, e_{(2)}^2, ..., e_{(h)}^2$ küçükten büyüğe doğru sıralı artık karelerdir. h değeri tahmin edicinin hesabında kullanılan gözlem sayısı olup, LMS' de belirlendiği gibidir. Bu metot en büyük artık karelerin toplamının dışında tutulması hariç EKK' ya benzerdir. En büyük artık karelerin toplamın dışında tutulması aykırı değerlerin tamamen dışarıda tutulması anlamına gelir. Aykırı değerlerin yapısına ve h değerine bağlı olarak LTS çok etkin olabilir. Gerçekte belirli sayıdaki aykırı değerler kırpılırsa bu metot hesapsal olarak EKK' ya eşdeğerdir. Bununla birlikte aykırı değerden daha fazla sayıda gözlem kırpılırsa bu metot etkin olmaz. Tersine aykırı değer sayısından daha fazla gözlemin kırpılması iyi gözlem değerlerinin hesap dışı bırakılması demektir. Kırılma noktası bakımından LTS, %50 kırılma noktasına sahip yüksek kırılma noktalı tekniklerden biri olarak değerlendirilir. En küçük kareler kestiriminde artık karelerin tamamının toplamını minimize etmek yerine kareler toplamının kırpılarak minimize edilmesi söz konusudur. Bu durumda n tane artık yerine h tane artığın toplamı söz konusudur ve burada h = [n / 2] + [(k' + 1) / 2] için, LTS $\left(\left[(n-k')/2\right]+1\right)/n$ en büyük kırılma noktasına ulaşır (Rousseeuw and Leroy, 1987).

3.3.5 M kestirimi

M kestirimi sağlam konum ve ölçek ölçülerine en çok olabilirlik fikrini genelleştiren tahmin edici sınıfını içermektedir.

M kestirimi aynı zamanda GM kestirimleri, S kestirimleri ve MM kestirimleri olarak sınıflandırılan pek çok sağlam regresyon kestirimlerinin temelini oluşturmaktadır (Andersen, 2008). M regresyon sağlam yöntemler içerisinde en popüler olarak kullanılan yöntemlerden biri olup, temel mantığı artıkların simetrik bir fonksiyonunu en küçüklemektir. Bu amaçla EKK amaç fonksiyonunun yerine aykırı değerlere daha az ağırlık veren bir başka fonksiyon kullanır. En küçüklenen amaç fonksiyonunun çözümü sonucunda elde edilen M kestiricileri y-yönünde aykırı değerlere karşı sağlam kestiricilerdir. Önerilen çok sayıda M kestiricisi söz konusudur. Bu çalışmada literatürde en yaygın kullanılan M kestiricilerine yer verilecektir.

3.3.5.1 Huber' in M-kestirimi

Huber (1973-1981) artıkların kareleri yerine artıkların farklı bir fonksiyonuna dayalı bir grup kestirici geliştirmiştir. M kestiricilerine ilişkin amaç fonksiyonu;

$$\min \sum_{i=1}^{h} \rho(e_i^2) \tag{3.15}$$

dir. Burada ρ sıfır noktasında tek bir minimuma sahip simetrik bir fonksiyonudur. M kestirimleri iteratif olarak yeniden ağırlıklandırılmış (IRLS) EKK' yı kullanılarak hesaplanır. İteratif olarak yeniden ağırlıklandırılmış EKK' da başlangıç değeri hesaplanır ve sonra yeni ağırlıklar başlangıç değerine dayalı olarak hesaplanır. İterasyonlar belirlenen sayıdaki iterasyon sonuçlanana kadar ya da yakınsama kriteri sağlanana kadar devam eder. Huber M kestirim metodu istatistiksel olarak LAD' dan daha etkin ve y-yönündeki aykırı değerlere karşı sağlamdır. Bununla birlikte M kestiricileri x-yönündeki aykırı değerlere karşı sağlam değildir ve kırılma noktaları 1/n' dir (Rouseeuw and Leroy, 2003). Huber' ın amaç fonksiyonunda, e_i artık terimi, sıfıra uzak değer aldığında ρ fonksiyonunun değeri $|e_i|$ ' ye, e_i artık terimi sıfıra yakın değer aldığında ise e_i^2 ' ye eşitlenmesine dayalı bir fonksiyon ileri sürülmüştür.

Buna göre Huber (1964)' ın önerdiği ρ fonksiyonu artıklar için tanımlandığında;

$$\rho(e) = \begin{cases} e^2 & ,-k \le e \le k \ ise \\ 2k\left|e\right| - k^2, & e < -k \ veya \ k < e \ ise \end{cases} \tag{3.16}$$

biçiminde ifade edilir. Burada $k = 1{,}5.\hat{\sigma}$, en çok kullanılan ölçek kestirimi olan $\hat{\sigma} = 1{,}483.MAD$ ve mutlak sapmaların medyanı (MAD) olan $MAD = med\left\{\left|e_i - med(e_i)\right|\right\}$ biçiminde hesaplanır (Birkes and Dodge, 1993). M kestiricisinin ölçek değişmez (scale invariant) özelliğine sahip olması gerekli değildir. Bu kestiricinin ölçek değişmez özelliğine sahip olması istenirse artıkların sağlam ölçek kestirimine bölünmesi $\left(r = \dfrac{e_i}{\hat{\sigma}}\right)$ gerekir. Bu durumda Huber' ın amaç fonksiyonu aşağıdaki gibi yazılabilir:

$$\rho(r) = \begin{cases} \dfrac{r^2}{2} & , \quad -k \le r \le k \\ k\left|r\right| - \dfrac{k^2}{2}, & r < -k \ veya \ k < r \end{cases} \tag{3.17}$$

ve ağırlık fonksiyonu ise aşağıdaki gibi verilmiştir (Rousseeuw and Leroy, 2003).

$$w_i = \begin{cases} 1, & \left|r_i\right| \le k \\ \dfrac{k}{\left|r_i\right|}, & \left|r_i\right| > k \end{cases} \tag{3.18}$$

3.3.5.2 <u>Hampel' in M kestirimi</u>

Hampel (1971) tarafından öne sürülen amaç fonksiyonu:

$$\rho(r)=\begin{cases}\dfrac{r^2}{2} & , \quad 0<|r|\le a\\ a|r|-\dfrac{a^2}{2} & , \quad a<|r|\le b\\ -\dfrac{a}{2(c-b)}(c-r^2)+\dfrac{a}{2}(b+c-a) & , \quad b<|r|\le c\\ \dfrac{a}{2}(b+c-a) & , \quad c<|r|\end{cases} \tag{3.19}$$

Hampel'in ağırlık fonksiyonu ise,

$$w=\begin{cases}1 & , \quad 0<|r|\le a\\ \dfrac{a}{2}\operatorname{sgn}(r) & , \quad a<|r|\le b\\ \dfrac{a}{r}\left[\dfrac{c-|r|}{c-b}\right]\operatorname{sgn}(r) & , \quad b<|r|\le c\\ 0 & , \quad \varepsilon<|r|\end{cases} \tag{3.20}$$

biçimindedir. Genellikle sabitlerin değerleri a=1,7 ; b=3,4 ; c=8,5 olarak alınır.

3.3.5.3 <u>Tukey' nin M kestirimi</u>

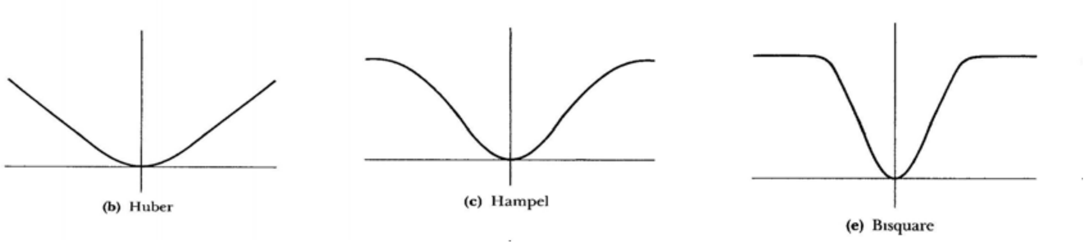
Tukey (1970) tarafından öne sürülen amaç fonksiyonu aşağıdaki biçimdedir:

$$\rho(r)=\begin{cases}\dfrac{1}{6}\left(1-\left(1-\left(\dfrac{r}{k}\right)^2\right)^3\right) & , \quad |r|\le k\\ \dfrac{1}{6} & , \quad |r|>k\end{cases} \tag{3.21}$$

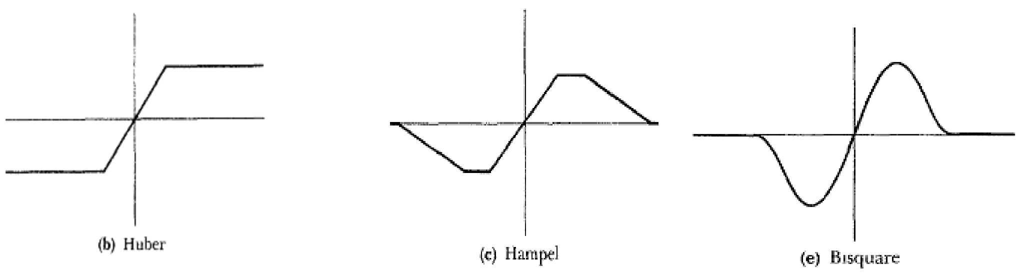
Tukey' in çift kare ağırlık fonksiyonu ise (k = 4,685) aşağıda verilmiştir (Tukey,1977).

$$w_i = \begin{cases} \left[1-\left(\dfrac{r}{k}\right)^2\right]^2 & , \quad |r| \le k \\ 0 & , \quad \textit{diğer durumlarda} \end{cases} \tag{3.22}$$

Şekil 3.1 ve Şekil 3.2' de Huber, Hampel ve Tukey tarafından ileri sürülen amaç fonksiyonlarının grafikleri ve yine Huber, Hampel ve Tukey tarafından ileri sürülen ağırlık fonksiyonlarının grafikleri yer almaktadır (Wu, 1985).



Şekil 3.1 Huber, Hampel ve Tukey' e ait amaç fonksiyonlarının grafikleri



Şekil 3.2 Huber, Hampel ve Tukey' e ait ağırlık fonksiyonlarının grafikleri

3.3.6 MM kestirimi

MM kestirimi Yohai tarafından 1987 yılında önerilmiş, son yıllarda artan bir popülariteye sahip olup, sağlam regresyon teknikleri arasında en yaygın olarak kullanılan, M kestirimin özel bir halidir. "MM" ismi final kestirimlerini hesaplamak için M kestirim sürecinin birden fazla sayıda kullanılmasından kaynaklanmaktadır.

MM kestiriminde sağlam M kestirimi aşağıdaki gibi verilir:

$$\sum_{i=1}^{n} \rho\left(\frac{y_i - x_i^T}{\hat{s}}; c\right) \tag{3.23}$$

Burada $\hat{s}$: artıkların sağlam ölçek kestirimi, $\rho(\bullet; c)$ ise c ayar sabitine sahip artıkların konveks ağırlık fonksiyonudur. Alternatif olarak, $\psi = \rho'$ 'nun azalan bir fonksiyon olduğu durumda $\hat{\beta}$, kestirimi yapılan (3.24) eşitliğinin çözümüdür.

$$\sum_{i=1}^{n} x_i \, \psi\left(\frac{y_i - x_i^T \hat{\beta}}{\hat{s}}\right) = 0 \tag{3.24}$$

Burada ψ monoton (sürekli artan veya sürekli azalan) olmayan bir fonksiyondur. MM kestirimi üç aşamalı bir süreçtir. İlk aşamada sağlam, yüksek kırılma noktalı kestirici hesaplanır. S kestirimi ilk aşamada kullanılır. 2. aşamada sağlam M kestirimi başlangıç adımındaki $\hat{s}$ artık kestirimlerini kullanarak hesaplanır. 3. aşamada final M kestirimi regresyon parametreleri hesaplanır (Yohai, 1987). Süreç ayrıntılı olarak aşağıdaki gibidir:

1. $\hat{\beta}^{(1)}$ katsayılarının başlangıç kestirimleri ve ilgili $e_i^{(1)}$ artıkları yüksek kırılma noktalı (%50 kırılma noktasına sahip) bir regresyondan elde edilir. Kestiricinin tutarlı olmasına rağmen etkin olmasına gerek yoktur. Sonuç olarak, Huber ve Tukey ağırlıklarıyla S kestirimi bu aşamada kullanılabilir.

2. Birinci aşamadaki başlangıç kestiriminden elde edilen $e_i^{(1)}$ artıkları artıkların ölçek tahmincisi $\hat{\sigma}_e$' nin hesaplanması için kullanılır.

3. Birinci aşamadan elde edilen $e_i^{(1)}$ artıkları ve ikinci aşamadan elde edilen artıkların ölçek kestiricisi M regresyon katsayılarının kestirimini elde etmek için ağırlıklı EKK' in ilk iterasyonunda kullanılır.

$$\sum_{i=1}^{n} w_i \left(e_i^{(1)} / \hat{\sigma}_e\right)\mathbf{x_i} = 0 \tag{3.25}$$

burada w_i Huber veya Tukey ağırlıklarıdır.

4. Üçüncü aşamadaki başlangıç ağırlıklı EKK' den artıklar kullanılarak yeni $w_i^{(2)}$ ağırlıkları hesaplanır.

5. İkinci, üçüncü ve dördüncü aşamadaki artıkların ölçek kestiricisi sabit tutularak yakınsama sağlanıncaya kadar iterasyona devam edilir (Andersen, 2008).

3.3.7 Genelleştirilmiş M –kestirimi

M kestiricisi kaldıraç noktalarını hesaba katmadığı için sınırlı olmayan etkiye sahiptir, diğer bir deyişle x yönündeki aykırı değerleri sınırlandırmaz (Hampel vd. 1986). Bu problemle baş edebilmek için sınırlı etki genelleştirilmiş M (GM) kestiricileri önerilmiştir. Amacı hem y hemde x yönündeki aykırı değerleri hesaba katan ağırlıkları oluşturmaktır (Andersen, 2008).

Bu kestiricilerin dezavantajı kırılma noktalarının $1/k'$ ' den (k' parametre sayısı olmak üzere) büyük olmamalarıdır. Dolayısıyla, çok parametreye sahip bir model için kırılma noktası 1/n' den daha iyi olmayacaktır. Ancak parametre sayısı küçük ise bu kestiriciler çok iyi sonuçlar verecektir (Montgomery, Peck and Vining, 2001).

GM sınıfı kestiricilerin genel hali aşağıdaki gibi tanımlanır;

$$\sum_{i=1}^{n} \pi_i \psi\left(\frac{y_i - x\hat{\beta}}{s\pi_i}\right) x_i = 0 \tag{3.26}$$

burada ψ, ρ' nun türevi, ve π_i ağırlıkları ifade etmektedir. Bu Schweppe tipi GM kestiricisi olarak bilinir. Mallows' da bir başka GM kestiricisini önermiştir. Bu iki tip kestirici arasındaki temel fark Mallows hem küçük hem de büyük artıklara düşük ağırlık verirken; Schweppe sadece küçük artıklara düşük ağırlık vermektedir. Literatürde yaygın bir şekilde yer alan iki genelleştirilmiş M-kestirim türü üzerinde durulacaktır. Bunlar Coakley-Hettmansperger GM kestirimi ve Schweppe ağırlıklarıyla sınırlı etkili M regresyon kestirimidir.

3.3.7.1 Coakley-Hettmansperger GM kestirimi

Coakley ve Hettmansperger (1993) tarafından ileri sürülen Schweppe bir adım kestiricisi (S1S) Schweppe kestiricisinden türetilmiştir. Bu kestiricinin orjinaline göre avantajı gözlemlerin iyi ve kötü kaldıraç noktası olup olmadığını dikkate alarak kötü kaldıraç noktalarına daha düşük ağırlık vermesidir. Gauss-Markov varsayımları altında EKK kestiricilerine göre sonuçlar %95 etkindir. S1S kestiricisi EKK tekniğinden ziyade yüksek kırılma noktalı bir regresyondan elde edilen artıkların ölçek ve başlangıç kestirimlerini kullanır. Başlangıç kestirimleri olarak Rousseeuw' un LTS kestiricisinin kullanımı %50 kırılma noktasını verir. Bu metot final M kestirimlerini tek adımda hesaplamasında ve başlangıç kestirimlerini LTS' den elde etmesinden ötürü Mallows ve Schweppe kestirimlerinden ayrılır (Andersen, 2008).

Prosedür LTS kestiricisiyle ($\hat{\beta}_0$) başlar ve Coakley-Hettmansperger kestiricileri aşağıdaki gibi elde edilir;

$$\hat{\beta}_{ch} = \hat{\beta}_0 + (\mathbf{X'BX})^{-1}\mathbf{X'W}\psi(r_i/(w_i s))s \tag{3.27}$$

Burada **W**=diag(w_i), $\mathbf{B} = diag(\psi'(r_i / sw_i))$ ve **X** nx$(k+1)$ boyutlu model matrisi, ve $\psi'(x)$ Huber'ın ψ fonksiyonunun türevidir. Ölçek kestirimi olarak $s = 1.4826(1+5/(n-k')) \times med\{|r_i|\}$ önerilmiştir. $\mathbf{x_i}$**'** ye verilen ağırlıklar aşağıdaki gibi hesaplanır;

$$w_i = \min\left\{1, \left[b/(\mathbf{x_i} - \mathbf{m_x})'\mathbf{C}^{-1}(\mathbf{x_i} - \mathbf{m_x})\right]^{a/2}\right\} \tag{3.28}$$

burada $\mathbf{m_x}$ ve **C** nicelikleri minimum hacimli elipsoid kestiricilerinin konum ve ölçek kestiricisini ifade etmekte olup, b değeri için k' serbestlik dereceli ki-kare dağılımının 0.95 kantil değeri ve a=2 değerinin kullanımı önerilmiştir. Coakley ve Hettmansperger regresyon metodu hata teriminin normal dağılıma sahip olması durumunda yüksek asimptotik etkinliğe sahiptir.

Diğer bir deyişle, örneklem büyüklüğü arttıkça EKK kestiricisinin standart hatası, $\hat{\beta}_{ch}$ 'ın standart hatasından daha küçük olmayacaktır. Sonuç olarak, normal varsayım altında yüksek asimptotik etki, sınırlı etki fonsiyonu ve yüksek kırılma noktasını aynı anda vermesi anlamında Coakley ve Hettmansperger metodu iyi sonuçlar vermektedir (Coakley and Hettmansperger, 1993).

3.3.7.2 Schweppe ağırlıklarıyla sınırlı etkili M regresyon kestirimi

Kaldıraç noktalarına karşı regresyon kestiricilerini bulma problemiyle baş edebilmenin bir yolu bağımsız değişkenlerin aşırı veya kuşkulu olarak değerlendirilen değerlerinin yerine düşük ağırlık verilmesidir. Diğer bir deyişle, yüksek kaldıraç noktalarına düşük ağırlık verilmesidir. Bunun için Mallows (1975) tarafından önerilen Mallows ağırlıkları kullanılabilir. Daha sonraki yıllarda Schweppe daha etkin bir kestirici elde etme amacıyla bir adım daha ileriye gitmiş ve eğer $\mathbf{x}_i$ küçük ağırlığa sahipse, r_i artıklarına daha büyük ağırlık verilmesini önermiştir.

Mallows ağırlıkları kuşkulu x değerleri olduğunda etkinlik kaybına neden olabilirken, Schweppe' nin yaklaşımı r_i artıklarını w_i ağırlıklarına bölerek bu problemle başa çıkabilir (Krasker and Welsch, 1982). Bu durumda M kestiricileri $p+1$ denklemin çözümü ile bulunur:

$$\sum_{i=1}^{n} w_i \psi(r_i/(w_i s))x_{ij} = 0 \qquad j = 0,...,p. \tag{3.29}$$

bu eşitlik basit iteratif sürece dayanır ve $w_i=u(\mathbf{x}_i)$ ağırlıkları yüksek etkinlik gibi istenen özelliklere sahip, $\mathbf{x}_i$' nin bir fonksiyonu olarak değerlendirildiğinde, Markatou ve Hettmansperger (1990) tarafından önerilen iki $u(\mathbf{x}_i)$ tercihi söz konusudur; bunlardan biri $\sqrt{1-h_i}$ ve diğeri de $\frac{(1-h_i)}{\sqrt{h_i}}$ ' dir. İlk tercih Schweppe için söz konusu olup, Handschin vd.(1975) tarafından önerilmiştir.

İkinci tercih ise Welsch(1980) için söz konusudur. Schweppe ağırlıklarıyla M regresyon kestiricisini hesaplamak için sabit terim ve eğim kestiricileri EKK tekniği ile hesaplanır.

İzleyen aşamalar aşağıda verilmiştir:

1- $r_{ik} = y_i - \hat{\beta}_{0k} - \beta_{1k}x_{i1} - .... - \hat{\beta}_{pk}x_{ip}$ artıkları hesaplanır, $|r_{i,k}|$ ' ların en büyük n-p tanesinin medyanı M_k ' ya eşitlenir, $s = 1{,}48 \times M_k$ ve $e_{ik} = \frac{r_{ik}}{s}$ olur.

2- Ağırlıklar,

$$w_{i,k} = \frac{\sqrt{1-h_{ii}}}{e_{i,k}} \; \Psi\left(\frac{e_{i,k}}{\sqrt{1-h_{ii}}}\right) \tag{3.30}$$

eşitliği ile hesaplanır. Burada Huber' ın ψ fonksiyonu, $\Psi = \max[-K, \min(K,x)]$ ve

$K = 2\sqrt{(p+1)/n}$ olarak alınır.

3- Bu ağırlıklar, ağırlıklı en küçük kareler kestirimlerini $\hat{\beta}_{0,k+1}, \ldots\ldots\ldots, \hat{\beta}_{p,k+1}$ elde etmek için kullanılır ve k değeri 1 artırılır.

4- 1., 2. ve 3. adımlar yakınsama sağlanıncaya kadar tekrarlanır. Böylelikle kestirimi yapılan parametrelerdeki değişim çok küçük oluncaya kadar iterasyon tekrarlanır.

Hipotez testleri sözkonusu olduğunda ε hata teriminin simetrik dağılıma sahip olduğu varsayılarak, Markatou ve Hettmansperger (1990) $w_i = (1-h_{ii})/\sqrt{h_i}$ kullanımını önermiştir. Bununla birlikte, ε asimetrik dağılıma sahip ise Carrol ve Welsh (1988)' deki sonuçlar Mallows ağırlıklarını veya $w_i = \sqrt{1-h_i}$ ' nın kullanımını göstermektedir.

BÖLÜM 4

UYGULAMA

Bu bölümde iki ayrı veri seti üzerinde çalışma yapılacaktır. Birinci çalışma ikinci el otomobil satış fiyatları veri seti üzerinde, ikinci çalışma ise Solow büyüme modeli veri seti üzerinde gerçekleştirilecektir. Bu amaçla aşağıda ilk olarak ikinci el otomobil satış fiyatları veri seti üzerinde, sonrasında ise ikinci veri seti olan Solow büyüme modeli üzerinde aykırı değer, etkin gözlem ve sağlam regresyon metotları incelenerek ayrıntılı bir şekilde aktarılmıştır.

4.1 İkinci El Otomobil Satış Fiyatları Veri Seti için Amaç ve Kapsam

İkinci el otomobil satışında otomobiller bazen çeşitli sebeplerden dolayı normal piyasa fiyatının altındaki fiyatlara satılmak zorunda kalır. Bu fiyat düşüşüne sebep olan bazı durum ve özelliklerin göz önüne alınması gerekir. İkinci el otomobil alıcılarının ilk olarak dikkat ettiği durum otomobilin kazalı olup olmadığıdır. Bunun anlamı daha sonra otomobilin değerinde düşüşe sebep olabilecek bir kazanın olup olmaması durumudur. Buna bağlı olarak otomobil dış yüzeyinde tamamen veya kısmi boya olup olmamasına dikkat edilir. Bir diğer dikkat edilen ve otomobilin fiyatını olumsuz yönde etkileyen faktör ise, değişmiş parçanın olması eksik ya da çalışmayan parça ve aksesuarın bulunmasıdır. Tüm bu etkenler sözel nitelikli olduğu için çalışma kapsamına alınmamıştır. Otomobilin kilometresinin yüksek olması da yukarıda sayılan faktörler kadar önemlidir. Otomobilin kilometresinin yüksek olması demek aracın uzun yollar katettiği anlamına gelir ki bu da otomobilde bir yıpranmaya neden olacaktır. Bu yıpranmaya bağlı olarak ileride doğabilecek tamir masrafları alıcıya yükleneceğinden fiyatta düşüşlere neden olur. Otomobilin yaşının otomobilin fiyatı üzerindeki etkisi büyüktür.

Otomobilin yaşı demek imal edildiği tarihten sonra geçen süre demektir ve hızla gelişen teknoloji dünyasında son model otomobiller tercih edilir çünkü insanlar son teknolojinin konforundan ve rahatlığından yararlanmak ister, dolayısıyla yaşı büyük otomobiller çoğunlukla tercih edilmemesinden dolayı ikinci el otomobil satış fiyatlarının düşük olmasında büyük etkiye sahip olmaktadırlar. İkinci el otomobil satış fiyatlarının düşük olmasındaki bir diğer etken ise motor gücüdür. Motor gücü otomobilin gücüdür yani otomobilin yapacağı hız kadar, sert hava koşullarına dayanıklılığı, yokuş tırmanışındaki zorlanmayışına kadar otomobilin motorunun gücü önemlidir. Dolayısıyla motor gücü düşük otomobiller ikinci el otomobil satış fiyatlarındaki düşüşte etkilidir. Tüm bu bilgiler ışığında ikinci el otomobil satış fiyatlarında etkili olabilecek sayısal değişkenler olarak otomobilin kilometresi, yaşı ve motor gücü değişkenleri çalışmaya alınmıştır. İkinci el otomobil satış fiyatlarıyla buna ilişkin bağımsız değişkenlerin değerleri EK-1' de verilmiştir.

Uygulama çalışmasında oluşturulan çoklu doğrusal regresyon modeli;

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon$$

olmak üzere;

Y : İkinci el otomobil satış fiyatını (TL)

X_1: Otomobilin katettiği mesafeyi (km)

X_2: Otomobilin yaşını (yıl)

X_3: Otomobilin motor gücünü (BG) ifade etmektedir.

4.1.1 İkinci el otomobil satış fiyatları veri seti için modelin EKK kestirim tekniğiyle çözümünden elde edilen sonuçlar

İkinci el otomobil satış fiyatları veri setinin E-Views 4.1 paket programında EKK yöntemiyle çözümünden elde edilen sonuçlar aşağıdaki gibi ifade edilebilmektedir.

Tablo 4.1 İkinci el otomobil satış fiyatları veri setinde EKK sonuçları

	katsayılar	Std. hata	t değerleri	P
sabit	-72.6996	6.8850	-10.5591	0.0000
X1	0.1947	0.1041	1.8700	0.0669
X2	-4.1865	1.1711	-3.5749	0.0007
X3	1.0509	0.0333	31.5432	0.0000

$R^2 = 0{,}949$ $\overline{R}^2 = 0{,}946$ F=336.8 $\hat{\sigma}_e$ =25,82

İkinci el otomobil satış fiyatları veri setindeki çalışmalar sonucu ortaya çıkan farklı varyanslılık ve otokorelasyon sorunlarının ortadan kaldırılması için R 2.10.1 programı ile bağımlı ve bağımsız değişkenlerimizin logaritmaları alınmış ve EKK yöntemiyle çözümünden elde edilmiş sonuçlar Tablo 4.2' de verilmektedir.

Tablo 4.2 İkinci el otomobil satış fiyatları veri setinde bağımlı ve bağımsız değişkenlerin logaritmasının alınması sonucu elde edilen EKK sonuçları

	katsayılar	Std. hata	t değerleri	P
sabit	-5.4907	0.5419	-10.13	4.29e-14
X1	0.1441	0.0668	2.16	0.0353
X2	-0.4086	0.0714	-5.72	4.78e-07
X3	1.8550	0.1057	17.55	2e-16

$R^2 = 0.8551$ $\overline{R}^2 = 0{,}847$ F=106.2 $\hat{\sigma}_e$ =0,3774

İkinci el otomobil satış fiyatları veri setinin bağımlı ve bağımsız değişkenlerinin logaritması alınarak değişen varyanslılık, otokorelasyon ve normal dağılım testlerinin R 2.10.1 programı yardımıyla elde edilen sonuçlarını ise şu şekilde özetleyebilmek mümkündür:

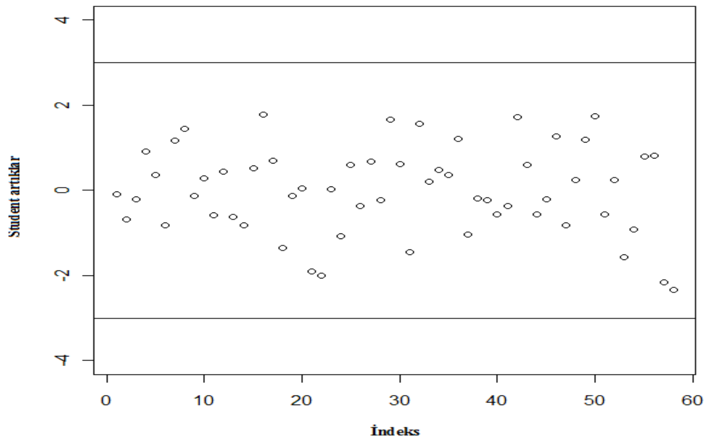
Breusch-Pagan farklı varyanslılık test istatistiği değeri 2,5147 olarak elde edilmiş ve p değeri ise 0,4726 olarak bulunmuştur. $\alpha = 0{,}05$ ile p değeri karşılaştırılarak $p < \alpha$ ise Ho RED edilir genel kuralına göre burada $p > \alpha$ olduğundan Ho Kabul edilir; diğer bir deyişle değişen varyanslılık yoktur sonucu elde edilmiştir.

Breusch-Godfrey otokorelasyon test istatistiği değeri 0,5581 ve p değeride 0,455 elde edilmiş $\alpha = 0,05$ ile p değeri karşılaştırılarak $p > \alpha$ olduğundan Ho Kabul edilir; diğer bir deyişle otokorelasyon yoktur sonucu elde edilmiştir.

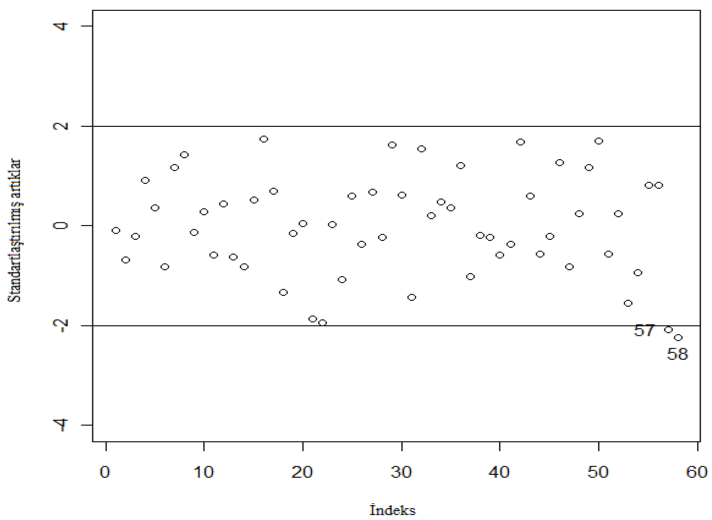
Shapiro-Wilk normallik test istatistiği değeri 0,9774 ve p değeri de 0,3518 elde edilmiş $\alpha = 0,05$ ile p değeri karşılaştırılarak $p > \alpha$ olduğundan Ho Kabul edilir; dolayısıyla veri seti normal dağılıma sahiptir sonucu elde edilmiştir.

4.1.1.1 Aykırı değerlerin istatistiklerle ve grafiklerle incelenmesi

Aykırı değerlerin istatistiklerle incelenmesi aşamasında Ek-3' e göre Mahalanobis uzaklığı *i.* gözlemin bağımsız değişken değerlerinin merkezinden ne kadar uzakta olduğunu göstermektedir. Fazla uzaktaki gözlem bağımsız değişkenler üzerinde aykırı değer olarak düşünülür. Bu ölçüt için Barnett ve Lewis (1978-1984) kritik değer tablosu düzenlemiş ve bu tabloya göre 58 gözlemli, 3 bağımsız değişken için tablo değeri 14.18 olarak bulunmuştur. Bu değerle yapılan karşılaştırma sonucu 6. ve 54. gözlemlerin en uzak gözlem değerleri olduğu görülmektedir. Mahalanobis uzaklığı kriterlerine göre aykırı değer olarak tanımlanan gözlemler 1., 4., 5., 6., 11., 12., 14., 16., 17., 18., 19., 46., 54 ve 55. gözlemlerdir. MCD sağlam uzaklıkları istatistiğine göre 1., 4., 6., 16., 17 ve 54. gözlemler aykırı değer olarak ifade edilmektedir. MVE sağlam uzaklıkları MCD sağlam uzaklıkları ile aynı aykırı değer gözlemlerine sahiptir. Student artıklar da [-3,+3] değer aralığı dışındaki gözlemler aykırı değer olarak kabul edildiğinden, student artıklarda aykırı değerin mevcut olmadığı Şekil 4.1 ile de gözlenebilmektedir. Standartlaştırılmış artıklarda ise [-2,+2] değer aralığı dışındaki gözlemler aykırı değer olarak kabul edildiğinden standartlaştırılmış artıklara göre ise 57. ve 58. gözlemler aykırı değer olarak kabul edildiği Şekil 4.2 ile de ifade edilebilmektedir. Sağlam standartlaştırılmış artıklarda ise belirlenen aykırı değerler, 21, 54, 57. ve 58. gözlemlerdir. Aykırı değerleri incelemek için kullandığımız grafikler ise sırasıyla Şekil 4.1-Şekil 4.7 olmak üzere sıralanmıştır.

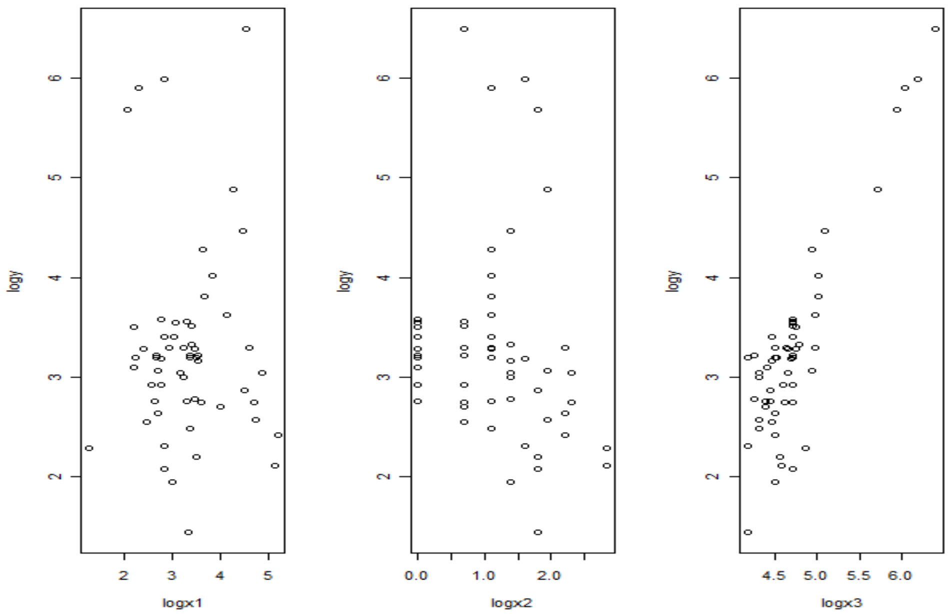


Şekil 4.1 Student artıklarının indeks grafiği.

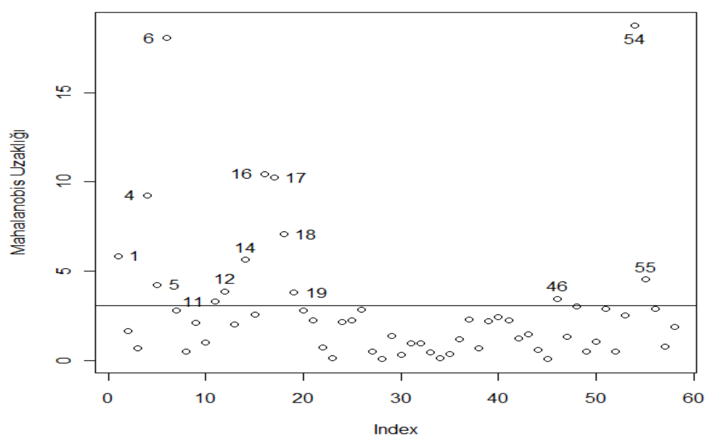


Şekil 4.2 Standartlaştırılmış artık grafiği.

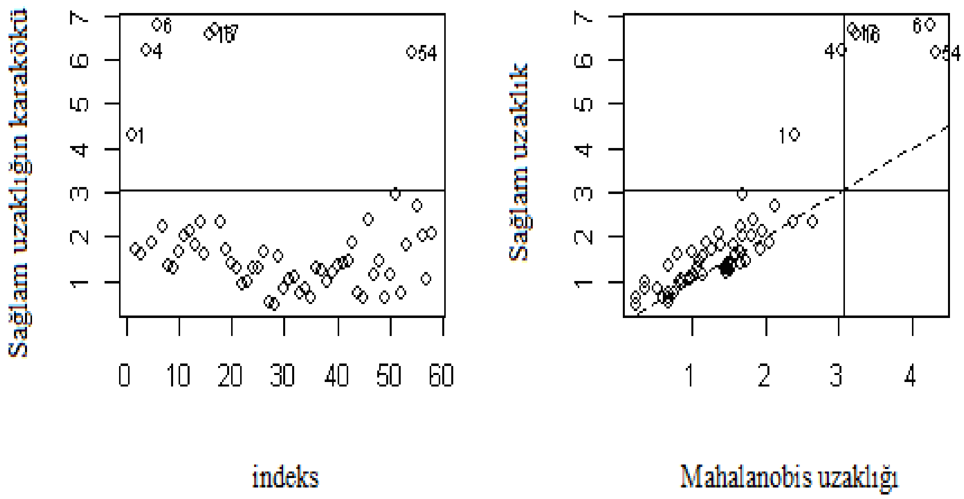
Ön inceleme yapmak amacı ile ikinci el otomobil satış fiyatları verisine ait X_1, X_2 ve X_3 bağımsız değişkenlerine karşı serpilme grafikleri Şekil 4.3' de verilmiştir. Bu grafiklerde gözlem değerlerinin çoğunluğundan uzakta olan gözlemler dikkati çekmektedir.



Şekil 4.3 Bağımsız değişkenlerin bağımlı değişken ile serpilme grafikleri.



Şekil 4.4 Mahalanobis uzaklığı grafiği.

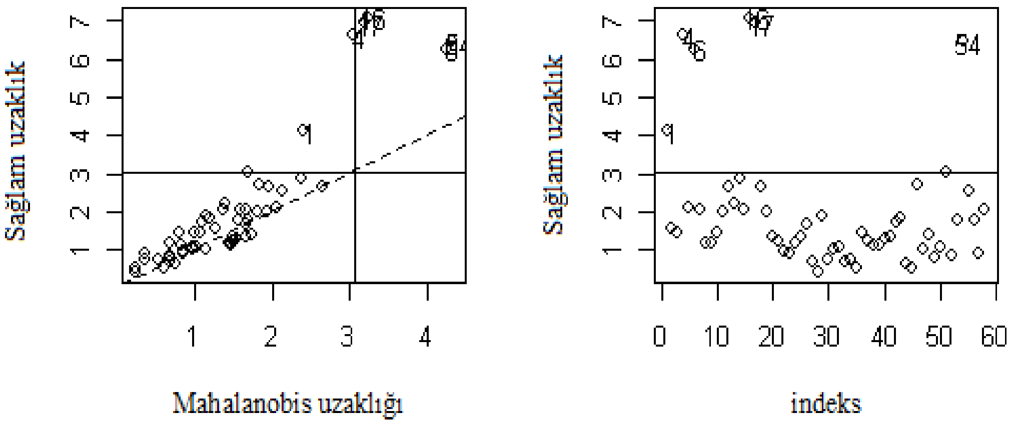


Şekil 4.5 MCD’ ye göre çizilen grafikler.

Mahalanobis uzaklıklarının sağlam uzaklıklardan farklı sonuçlar verdiği hem Ek 3’ den hem de Şekil 4.5 ve Şekil 4.6’ dan görülmektedir. Sağlam uzaklıkların kareköküne ilişkin indeks grafiğinde numaralandırılmış gözlemlerin veri setinin geri kalanından uzakta kaldığını göstermektedir.

Mahalanobis uzaklıklarına karşı çizilen sağlam uzaklıkların grafiğinde eğer kirlenmiş değerler olmasaydı tüm gözlem değerleri kesikli doğru etrafında yer alacaklardı. Uygulama verilerimizde noktaların çoğu bu doğru etrafında yer alırken, 6 noktanın doğrudan uzaklaştığı görülmektedir.

MVE’ ye ve MCD’ ye göre çizilen grafiklerin benzer sonuçlar verdiği görülmektedir. Bu durumda bu grafiklerden bir tanesinin seçimi gerekli açıklamaların yapılabilmesi için yeterli olabilecektir.

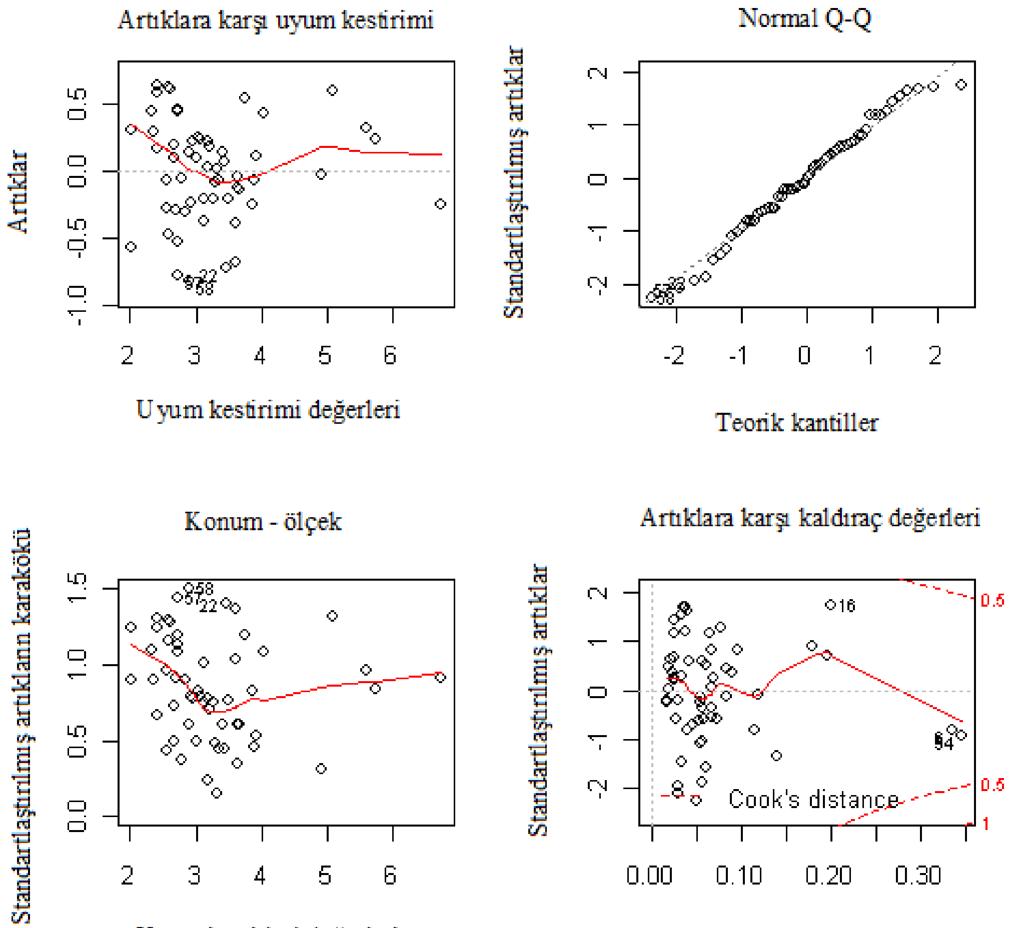


Şekil 4.6 MVE' ye göre çizilen grafikler

Şekil 4.7.a' da çizilen artıklara karşı uyum kestirimi grafiğinde noktaların rassal dağılmasına karşın, 22., 57. ve 58. gözlemlerin verilerin geri kalanından uzakta yer aldıkları gözlenmektedir.

Şekil 4.7.b' deki normal Q-Q grafiğinde ise değerlerin bir doğru etrafında yer aldığı 22., 57. ve 58. gözlemlerin aykırı değer olarak nitelendirildiği görülmektedir. Benzer yorumlar Şekil 4.7.c' için de geçerlidir.

Şekil 4.7.d' deki kaldıraç noktalarına karşı standart artık grafiğinde ise 6., 16. ve 54. gözlemlerin etkin gözlem olduğu görülmektedir.



Şekil 4.7 a,b,c,d EKK’ ya dayalı grafikler;

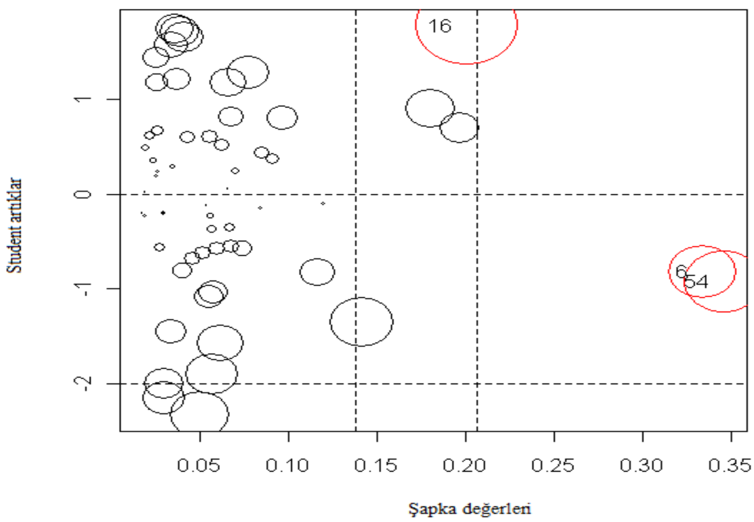
a) Artıklara (r_i) karşı $\hat{y}$ uyum kestirimi yapılan değerlerin grafiği
b) Normal Q-Q grafiği
c) Uyum kestirimine karşı standartlaştırılmış artıkların karekök grafiği
d) Kaldıraç değerlerine karşı standart artık grafiği

4.1.1.2 Etkin gözlemlerin istatistiklerle ve grafiklerle incelenmesi

Etkin gözlemlerin istatistiklerle incelenmesine başlandığında Ek-4’ de Cook uzaklığı 4/58-4 = 0.0741 kritik değerinden büyük olan 6., 16. ve 54. gözlemler regresyon doğrusu üzerinde etkilidir. Bu durumu Şekil 4.7.d’ de Artıklara karşı çizilen kaldıraç noktası çiziminden de görmek mümkündür.

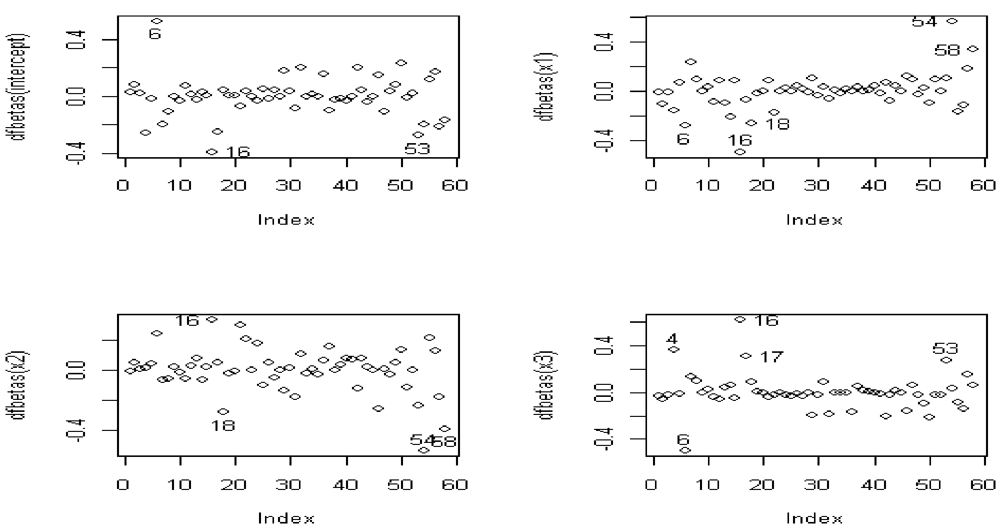
COVRATIO istatistiği için hesaplanan değerler kritik değer $1 \pm 3(4/58) = 1 \pm 0.2069$ aralıklarının dışına çıkan değerlerle karşılaştırıldığında 1., 4., 6., 17., ve 54. gözlemlerin etkin olduğu söylenir.

Düzeltilmiş Cook uzaklığı istatistiği için ise hesaplanan değerlerden 6., 16., 21., 54. ve 58. gözlemler kritik değer $2\sqrt{54/58} = 1.9298$ 'den büyük olduklarından etkin noktalar olarak görülmektedir. Welsch uzaklığı ölçüsü için de $3\sqrt{4} = 6$ kritik değerinden büyük olan 16. ve 54. gözlemler etkin gözlem olarak nitelendirilir. Güven elipsoidlerine bağlı etkinlik ölçülerinden olan Andrews-Pregibon istatistiği için kritik değer 1-2(3+1)/58=0.8621 ile karşılaştırıldığında bu değerden küçük gözlem değerleri etkin gözlem olarak nitelendirilir. Diğer bir deyişle, 4., 6., 16., 17., 18, 54., ve 58 gözlemler bu ölçüte göre etkin gözlem kabul edilirler. EKK' ya dayalı PRESS istatistiği 8,895762 olarak bulunmuştur. Ek 5' de DFFITS değerleri için $|DFFITS_i| > 2\sqrt{\frac{k'}{n}}$ kritik değeri ile hesaplanan $2\sqrt{\frac{4}{58}} = 0,5252$ ' yi aşan 6., 16., 18., 54., ve 58. gözlemler etkin gözlem olarak tanımlanmaktadır. DFBETAS değerleri *i.* gözlemin silinmesinin her bir katsayı üzerindeki etkisini göstermektedir. Ek 5' den bu değerlere bakıldığında 6., 16., 18., 54. ve 58. gözlemlerin $2/\sqrt{58} = 0.263$ kritik değerini aştığını ve bu nedenle de X_1(otomobilin kilometresi) değişkenine ait katsayı üzerinde etkili olduğunu söyleyebiliriz. Bu gözlemlerden 6., 16. ve 18. gözlemin etkisi negatif yöndedir. X_2(otomobilin yaşı) değişkenine ait katsayı üzerinde etkili olan 16., 18, 54, ve 58. gözlemlerden 18., 54. ve 58. gözlemler negatif yönde etkilidir. 4., 6., 16., 17. ve 53. gözlemin X_3(otomobilin motor gücü) değişkenine ilişkin katsayı üzerinde etkili ve bunlardan 6. gözlem negatif yönde etkilidir. Tüm bu bilgiler Şekil 4.9' dan da görülebilmektedir. DFTSTAT değerleri *i.* gözlem silindiğinde j. regresyon katsayısının sıfıra eşit olup olmadığını test ederken t istatistik değerindeki değişimi ölçer. Gözlem değerleri birbiri ile kıyaslandığında büyük $|DFTSTAT_{j(i)}|$ değerine sahip gözlemler etkin gözlem olarak nitelendirilirler. DFTSTAT B_0' a göre 1., 4., 17., 54., gözlemler, DFTSTAT B_1' e göre 16. ve 54. gözlemler DFTSTAT B_2' e göre 16., 18., 21., 22., 46., 54., gözlemler, DFTSTAT B_3' e göre 1., 4., 6., 16., 17., ve 58. gözlemler etkin gözlem olarak ifade edilmektedir.

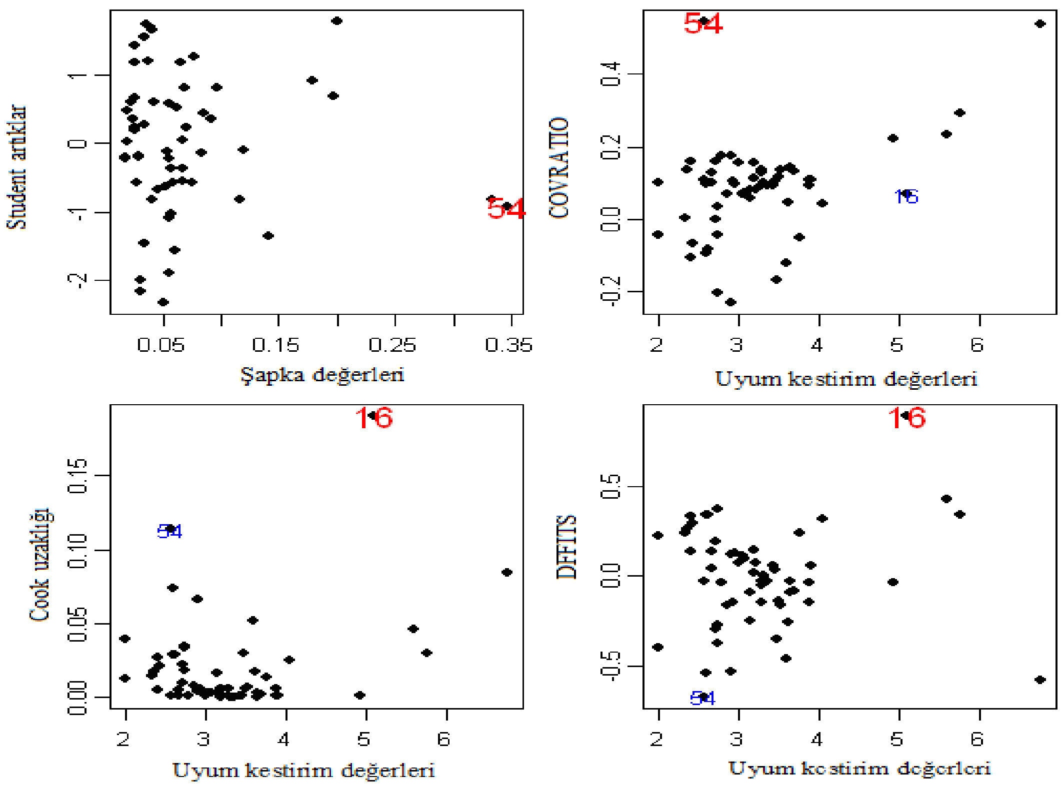


Şekil 4.8 Etki grafiği

Şapka değerlerine karşı Student artıklarının çizildiği etki grafiğine göre 16. gözlem en büyük etkiye sahip gözlemdir. Bu gözlemi sırasıyla 54. ve 6. gözlemler izlemektedir.



Şekil 4.9 DFBETAS grafikleri



Şekil 4.10 Bazı teşhis ölçülerinin uyum kestirimi değerlerine karşı grafiği

Şekil 4.10' a göre 16 ve 54 nolu gözlemlerin Student artıkları, COVRATIO, Cook uzaklığı ve DFFITS ölçüleri bakımından etkin gözlem oldukları dikkati çekmektedir.

4.1.1.3 <u>Çok sayıda etkin gözlemin istatistiklerle incelenmesi</u>

Tek sayıdaki gözlem satırının çıkarılması ile belirlenen potansiyel etkin gözlemlerden yararlanarak çok sayıda gözlemler daha önce belirtilen ölçüler kullanılarak Tablo 4.3' de incelemeye alınmıştır.

Tablo 4.3’ den elde edilen sonuçlara dayandırılarak, çoklu COVRATIO istatistiğine göre çıkarılan gözlem gruplarının her biri diğer gözlem grupları ile karşılaştırıldığında çok büyük ya da çok küçük değerler veren gözlem grubundaki gözlem değerleri etkin gözlem olarak ifade edilmektedir; bu sebeple çoklu COVRATIO istatistiğine göre 16,2888 değerini elde ettiğimiz gözlem grubuna ait 1., 4., 6., 16., 17., 51., 54., ve 55. gözlemler etkin gözlemdir.

MDFFITS istatistiğine göre ise diğer gözlem kümeleri ile karşılaştırdığımızda büyük değerler elde ettiğimiz gözlem kümesi etkin gözlemlere sahiptir ifadesinden yola çıkılarak Tablo 4.3’ de MDFFITS istatistiğine ait en büyük değer olan 0,59872’ ye karşılık gelen gözlem kümesindeki 1., 4., 6., 16., 17., 51., 54., ve 55. gözlemler etkin gözlemlerdir. Gözlem kümelerini Çoklu Andrews-Pregibon istatistiği için birbiri ile karşılaştırdığımızda en küçük Çoklu Andrews-Pregibon değeri olan 0,3363’ e karşılık gelen gözlem kümesindeki 1., 4., 6., 16., 17., 51., 54., ve 55. gözlemler etkin gözlemlerdir. Dolayısıyla üç ölçüt aynı gözlem değerlerini etkin gözlem olarak göstermektedir.

Tablo 4.3 Çok sayıda etkin gözlemi incelemek için kullanılan istatistikler

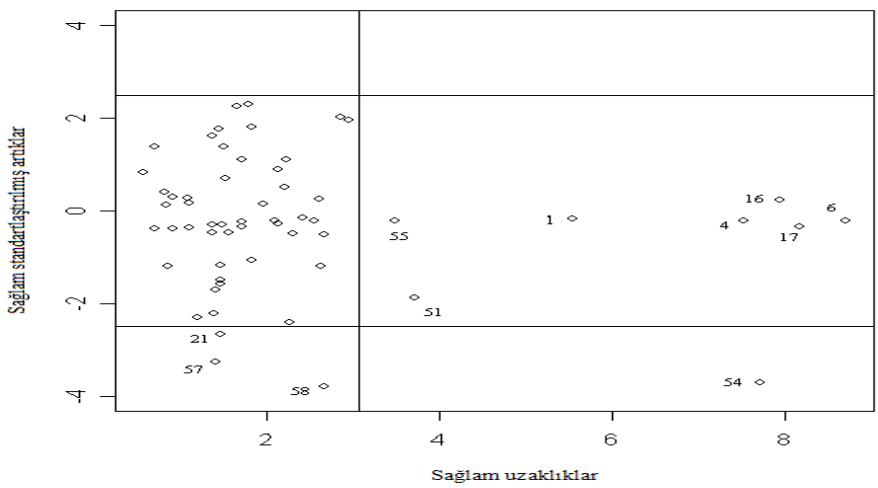
Çıkarılan Gözlem grubu	**Çoklu COVRATIO**	**MDFFITS**	**Çoklu AP**
1-6	1,9667	0,0424	4,2389
6-16	1,7131	0,0738	3,7723
6-54	2,5468	0,0363	3,1217
16- 54	1,8340	0,0452	3,5525
6-16-54	2,3259	0,1708	2,2303
1-4-6-17-54-58	8,5570	0,3594	1.0207
1-4-6-16-17-51-54-55	16,2888	0,5987	0,3363

4.1.2 Sağlam regresyon metotları ile elde edilen sonuçlar

Literatürde aykırı değerlerin belirlenmesinde EKK’ ya göre çizilen Normal Q-Q grafiklerinin çizimi yerine daha sağlam sonuçlar veren grafikler geliştirilmiştir.

Bunlardan biri sağlam uzaklıklara karşı çizilen sağlam standartlaştırılmış artıklardır. Şekil 4.11 ilgilenilen veri seti için bu grafiği göstermektedir. Grafiğe göre 21., 54., 57., ve 58. gözlemler regresyon aykırı değerleri olup 21., 57., ve 58 nolu gözlemler ise y yönündeki aykırı değer olarak karşımıza çıkmaktadır. Yani kestirimi yapılan diğer otomobillere göre fiyat bakımından daha pahalı olduğu söylenebilir.

Ayrıca 54. gözlem aynı zamanda kötü kaldıraç noktası olarak da ifade edilmektedir. 3.057 kritik değerinin sağında yer alan ve -2.5 ile +2.5 değerleri arasındaki gözlemlerde iyi kaldıraç noktalarıdır. Yani diğer otomobillerin özelliklerinden (fiyat hariç) farklıdırlar.



Şekil 4.11 İkinci el otomobil satış fiyatları verisi için sağlam uzaklıklara karşı sağlam standartlaştırılmış artık grafiği.

4.1.2.1 Sağlam regresyon metotları ile elde edilen katsayı kestirimleri ve ölçek kestirim değerleri

Bu bilgiler ışığında 3. bölümde anlatılan sağlam regresyon metotları ikinci el otomobil satış fiyatları verisine uygulandığında elde edilen katsayı kestirim teknikleri ve artık standart hata değerleri (residual standard error) Tablo 4.4’ de verilmiştir.

Tablo 4.4 İkinci el otomobil satış fiyatları veri seti için farklı tekniklerde elde edilen katsayı kestirimleri ve ölçek kestirimi değerleri

Metot	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\sigma}_e$	R^2
LAD	-5.1091	-0.0135	-0.2690	1.8529	0.4004	0,5304
M Huber	-5.4162	0.1345	-0.4000	1.8467	0.3451	0,7630
M Hampel	-5.4625	0.1396	-0.4032	1.8521	0.3422	0,9498
M Tukey	-5.4108	0.1321	-0.3948	1.8457	0.3401	0.7075
MM	-5.4109	0.1321	-0.3948	1.8457	0.3402	0.4326
LTS	-4.8396	0.0894	-0.3686	1.7632	0.3722	0,6917
LMS	-5.4305	0.0052	-0.2035	1.9037	0.2914	0,7466
GM_{BM}	-5.1828	0.0721	-0.3130	1.8239	0.3592	0,9005
GM_{CH}	-5.1616	0.0069	-0.2646	1.8537	0.3344	0,7776
WLAD	-3.3781	0.1465	-0.3308	1.3890	0.3271	0.7297

Regresyon denkleminin kestiriminde kullanılan M kestiriciler için teorik kısımda belirtildiği gibi birçok alternatif söz konusudur. Kullanılan amaç fonksiyonuna göre, artıklar farklı ağırlıklar kullanılarak minimize edilmektedir. Bu nedenle ortaya çıkan tahminler değişkenlik göstermekle birlikte elde edilen parametre tahminleri, artık standart hata değerleri Ek 5' de özetlenmiştir. Artık standart hata değerlerine göre en küçük değeri veren sağlam teknikler LMS ve WLAD dır. Bu tekniklerden sonra küçükten büyüğe doğru sıralama yapıldığında artık standart hata değerleri $GM_{CH,}$ M Tukey, MM, M Hampel, M Huber, GM_{BM}, LTS ve LAD' dır.

Belirlilik katsayısı (R^2) hesaplamaları sonuçlarına göre modele en iyi uyum sağlayan metot M Hampel olmak üzere diğer metotlar sırası ile GM_{BM}, $GM_{CH,}$ M Huber, LMS, WLAD, Tukey, LTS, LAD, MM' dir.

4.1.2.2 <u>Sağlam regresyon metotları ile elde edilen ağırlıkların incelenmesi</u>

Önceki kısımda yer verilen tekniklerden elde edilen artıklar ve aldıkları ağırlıklar da Ek 6 ve Ek 7' de verilmiştir.

Ek 6’ da farklı tekniklere göre elde edilmiş yüksek artık değerlerine her tekniğin farklı ağırlıklar verdiği Ek 7’ den görülmektedir.

Ek 7’ de Huber’ a göre; 8., 16., 18., 21., 22., 29., 31., 32., 42., 50., 53., 57 ve 58. gözlemler aykırı değer, Hampel’ a göre 22., 57 ve 58. gözlem aykırı değer olarak bulunmuştur. Tukey’ e göre en küçük ağırlıklar verilen 57. ve 58. gözlemler aykırı değer olarak tanımlanabilir, diğer gözlemlere ise [0,5098 ile 0,9999] değerleri arasında ağırlık verilmiştir.

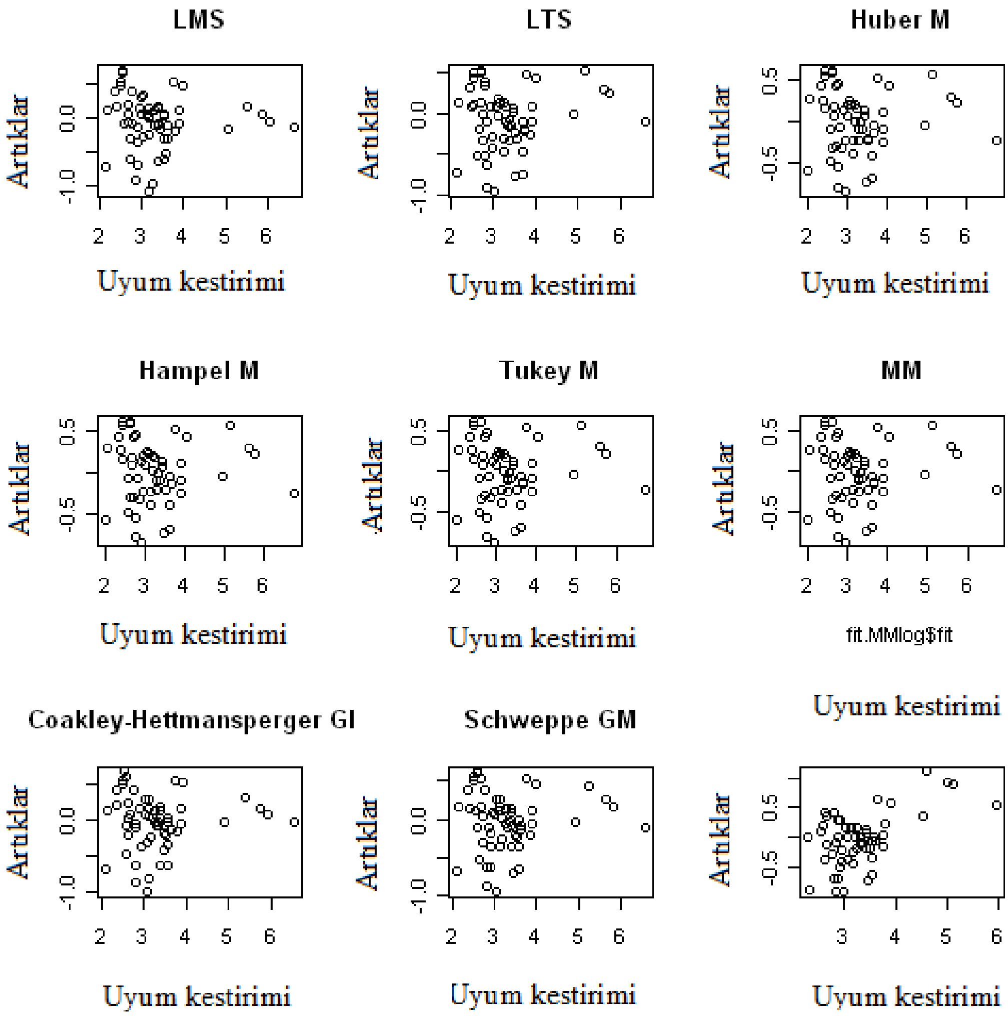
LMS’ ye göre 53., 54., 57., ve 58. gözlemler aykırı değer, LTS’ ye göre 58. gözlem aykırı değer olarak bulunmuştur.

MM’ e göre en küçük ağırlık verilmiş iki gözlem olan 57. ve 58. gözlemler aykırı değer olarak tanımlanabilir. Bununla birlikte, diğer gözlemlere [0,5101 ile 0,9999] aralığında ağırlıklar verilmiştir.

GM_{BM}’ e göre 7., 8., 16., 18., 21., 22., 29., 31., 32., 36., 42., 46., 49., 50., 53., 54., 57., 58 nolu gözlemlere 0,50’ den düşük ağırlık verilmiştir. Bu veri seti için gözlemlerin büyük kısmına düşük ağırlık verilmesi söz konusudur.

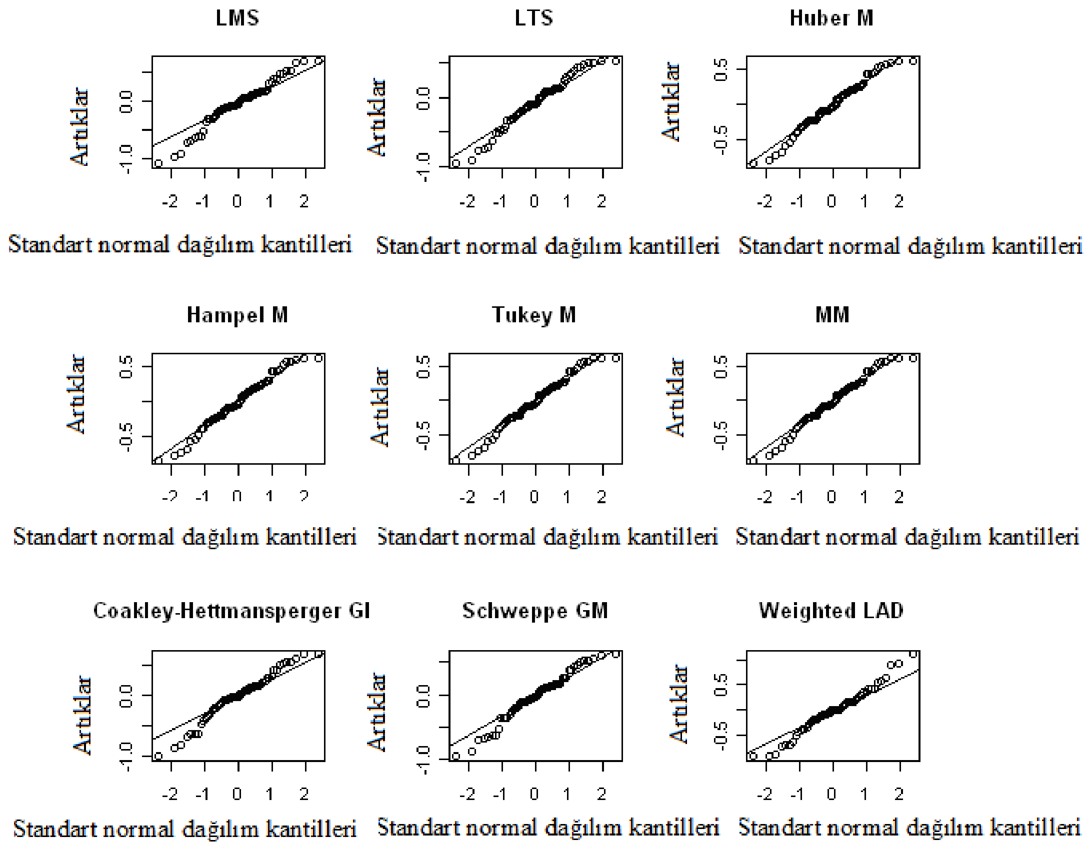
GM_{CH}’ e göre; 10 adet gözleme sıfıra yakın ağırlık verilmesi söz konusudur. WLAD’ a göre; 4., 6., 16., 17. ve 54. gözlemlere sıfıra yakın ağırlık verilmiş olup 1’ den düşük ağırlık verilenler ise [0,1259 ile 0,8830] arasında değer almıştır.

4.1.2.3 Sağlam regresyon metotları ile elde edilen sonuçların grafiksel incelenmesi



Şekil 4.12 Artıklara (r_i) karşı $\hat{y}$ uyum kestirimi yapılan değerlerin grafikleri

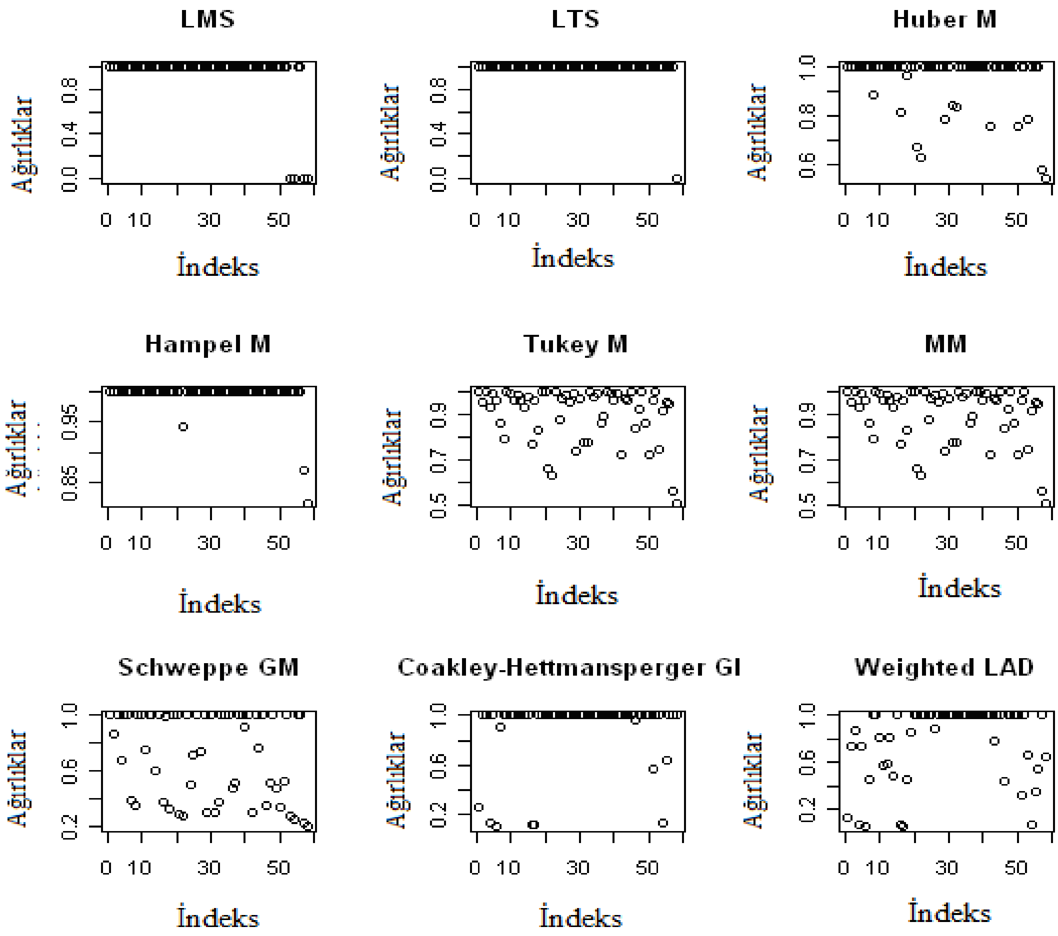
Şekil 4.12' ye göre tüm metotlarda veri setinin çoğunluğundan uzakta yer alan aykırı değerlerin varlığı gözlenmektedir. Ayrıca düzenli gözlemlerin rassal bir serpilme sergilediği söylenebilir.



Şekil 4.13 Artıkların normal Q-Q grafikleri

Şekil 4.13' den normal Q-Q grafiklerine göre verilerin tüm metotlar için normallik varsayımını sağladığı, Shapiro-Wilk testi ile de bu sonuçların doğrulandığı görülmüştür.

Ek 7' de verilen ve yukarıda yorumu yapılan farklı metotlara göre hesaplanmış ağırlıkların grafikleri Şekil 4.14' de verilmektedir.



Şekil 4.14 Farklı sağlam metotlar için ağırlıkların grafikleri

Sonuç olarak EKK kestirim tekniği her gözleme eşit ağırlık verirken, sağlam regresyon metotlarının aykırı gözlemlere farklı ağırlıklar verdiği, aykırı değer olduğu şüphelenilen gözlemlerin denklemde daha az ağırlıkla yer aldığı Şekil 4.14' den de görülebilmektedir.

4.2 Solow Büyüme Modeli Veri Seti için Amaç ve Kapsam

Yukarıda ikinci el otomobil satış fiyatları veri seti için gerçekleştirilen çalışmanın teorik kısımlarında aktarılanların uygulamasından sonra aşağıda Solow büyüme modeli veri seti için uygulamaya geçilmiştir. Solow büyüme modelinin ikinci el otomobil satış fiyatları veri setinden farkı ve çalışmaya alınma sebebi, normallik varsayımını sağlamaması ve sadece y-yönündeki aykırı değerlere sahip olması şeklinde ifade edilebilir.

Nonneman ve Vanhoudt (1996) tarafından ele alınan OECD ülkeleri için Solow büyüme modeli verileri 22 OECD ülkesinden oluşmaktadır ve Ek-2' de veri seti sunulmuştur. Veriler yatay kesit verileri niteliğindedir. Literatürde ele alınan regresyon modeli OECD ülkeleri için Solow büyüme modeli olarak anılmaktadır.

Verilerin alındığı Nonneman ve Vanhoudt (1996) makalesinde Norveç' e ilişkin okul (school) değişkeninin almış olduğu değer Mankiv, Romer ve Weil (1992)' den düzeltilmiştir. Diğer bir deyişle 0.01 yerine doğru olan değer 0,100 alınarak hesaplamalar gerçekleştirilmiştir.

Bu ülkelerin gözlem birimi olarak alındığı çoklu doğrusal regresyon modelindeki değişkenler ise aşağıda verilmiştir:

Y : Çalışan kesimin (15-65 yaş) kişi başına düşen reel gayrisafi yurtiçi hasılası (1985 ve 1960 yıllarındaki) (InY_t)

X_1: 1985 yılı uluslararası fiyatlar (InY_0)

X_2:Yurtiçi yatırımın reel gayrisafi yurtiçi hasılaya oranının yıllık ortalaması (1960-1985) (S_k)

X_3: Yıllık nüfus artışı +%5 (1960-1985 yıllarındaki) (n)

X_4: Ortaokula giden çalışan nüfusun yüzdesi (S_h)

X_5: Gayrisafi yurtiçi harcamaların yıllık oran ortalaması (1975-1985) (S_T)

4.2.1 Solow büyüme modelinin EKK kestirim tekniğiyle çözümünden elde edilen sonuçlar

R 2.10.1 programı yardımı ile bağımlı ve bağımsız değişkenlerin EKK tekniği ile hesaplamalarına bakılarak anlamlılıkları sınanmış ve aşağıda elde edilmiş sonuçlara göre bağımlı değişkeni etkileyen bağımsız değişkenler olarak ve EKK metoduna göre anlamlı sonuçlar veren X_1, X_2 ve X_3 değişkenleri seçilmiştir.

Tablo 4.5 Solow büyüme modeli veri seti için EKK sonuçları

	katsayılar	**Std. hata**	**t değerleri**	**P**
sabit	2.9759	1.0216	2.913	0.0093
X1	-0.3429	0.0565	-6.070	0.0000
X2	0.6501	0.2020	3.218	0.0048
X3	-0.5730	0.2904	-1.973	0.0640
$R^2 = 0{,}7464$ $\overline{R}^2 = 0.7041$ F= 17,66 $\hat{\sigma}_e$ =0,1329				

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Solow büyüme modeli veri setinin bağımlı ve bağımsız değişkenlerinin logaritması alınmış olarak değişen varyanslılık, otokorelasyon ve normal dağılım testlerinin R 2.10.1 programı yardımıyla elde edilen sonuçlarını ise şu şekilde özetleyebilmek mümkündür:

Breusch-Pagan farklı varyanslılık test istatistiği değeri 6,4269 elde edilmiş ve p değeri ise 0,0926 olarak bulunmuştur. $\alpha = 0{,}05$ ile p değeri karşılaştırılarak $p < \alpha$ ise Ho RED edilir genel kuralına göre burada $p > \alpha$ olduğundan Ho Kabul edilmiş; diğer bir deyişle değişen varyanslılık yoktur sonucu elde edilmiştir.

Breusch-Godfrey otokorelasyon test istatistiği değeri 1,814 ve p değeride 0,1780 elde edilmiş $\alpha = 0{,}05$ ile p değeri karşılaştırılarak $p > \alpha$ olduğundan Ho Kabul edilmiş; diğer bir deyişle otokorelasyon yoktur sonucu elde edilmiştir.

Shapiro-Wilk normallik test istatistiği değeri 0,9259 ve p değeri de 0.001651 elde edilmiş $\alpha = 0{,}05$ ile p değeri karşılaştırılarak $p < \alpha$ olduğundan Ho Red edilir. Dolayısıyla veri seti normal dağılıma sahip değildir sonucu elde edilmiştir.

4.2.1.1 Solow büyüme modeli için aykırı değerlerin istatistiklerle ve grafiklerle incelenmesi

Tablo 4.6' da Mahalanobis uzaklığı kriterlerine göre aykırı değer olarak tanımlanan gözlemler 1., 2., 3., 7., 10., 15., 19., 20. ve 21. gözlemlerdir. Tablo değeriyle yapılan karşılaştırma sonucu 19. gözlemin en uzak gözlem değerleri olduğu Şekil 4.18' den de görülmektedir.

MCD sağlam uzaklıkları istatistiğine göre 1., 2., 19., 21. ve 22. gözlemler aykırı değer olarak ifade edildiği Şekil 4.19' dan da gözlenebilir. Şekil 4.20' den de gözlenebilen MVE sağlam uzaklıkları MCD sağlam uzaklıkları ile aynı aykırı değer gözlemlerine sahiptir.

Student artıklar da [-3,+3] değer aralığı dışındaki gözlemler aykırı değer olarak kabul edildiğinden 3. gözlem aykırı değer olarak, Şekil 4.15 ile de gözlenebilmektedir.

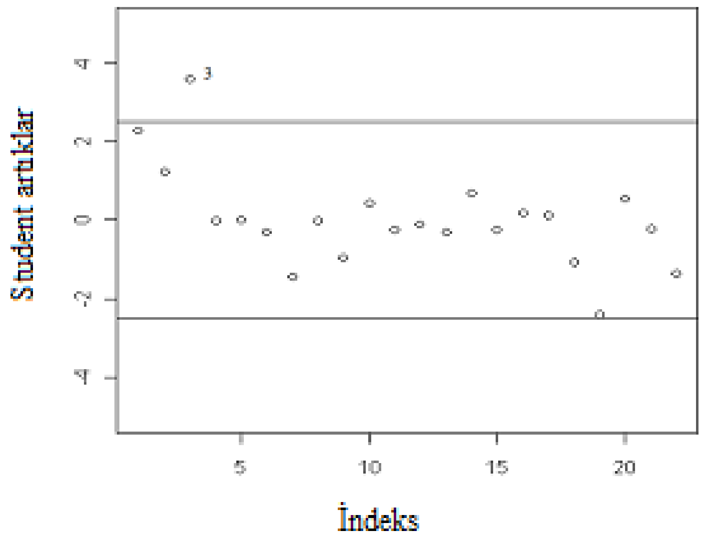
Standartlaştırılmış artıklarda ise [-2,+2] değer aralığı dışındaki gözlemler aykırı değer olarak kabul edildiğinden standartlaştırılmış artıklara göre ise 1., 3., ve 19. gözlemler aykırı değer olarak kabul edildiği Şekil 4.16 ile de ifade edilebilmektedir.

Sağlam standartlaştırılmış artıklarda ise belirlenen aykırı değerler 1., 2., 3. ve 21. gözlemler aykırı değerlerdir.

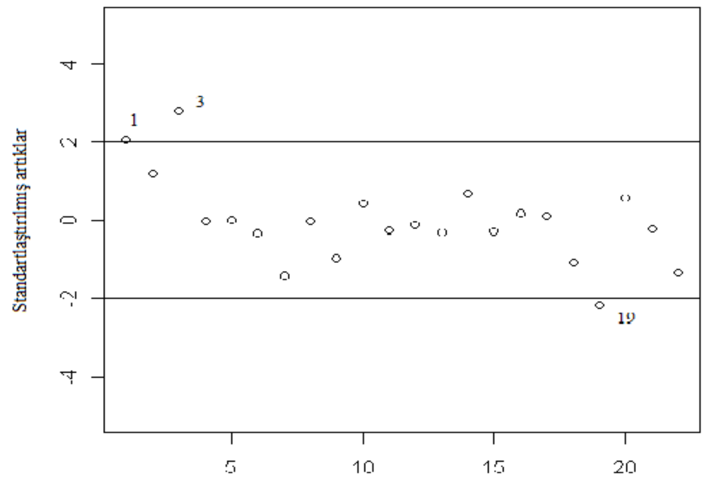
Tablo 4.6- Solow büyüme modeli için aykırı değerlerin belirlenmesinde kullanılan istatistikler

i	Mahalanobis Uzaklığı	MVE Sağlam Uzaklıkları	MCD Sağlam Uzaklıkları	Standartlaştırılmış Artıklar	Student Artıklar	Sağlam standartlaştırılmış artıklar
1	**3.3894**	**17.0520**	**15.3546**	**2.0469**	2.2710	**5.5320**
2	**3.9457**	**13.0673**	**11.7665**	1.2115	1.2285	**3.7898**
3	**4.4700**	2.7085	2.4389	**2.7845**	**3.5866**	**5.6487**
4	1.2333	0.9927	0.8939	-0.0248	-0.0241	-0.2449
5	1.3208	0.6237	0.5616	0.0084	0.0081	-0.1551
6	0.7469	0.4423	0.3982	-0.3167	-0.3087	0.0576
7	**4.2554**	2.6805	2.4137	-1.3991	-1.4403	-0.7052
8	0.2119	0.9474	0.8531	-0.0096	-0.0093	1.1134
9	1.0234	0.8100	0.7294	-0.9447	-0.9417	-1.0573
10	**3.5044**	2.4853	2.2379	0.4460	0.4359	-0.1551
11	0.5691	1.1463	1.0322	-0.2367	-0.2304	-0.1144
12	0.7163	0.3315	0.2985	-0.0985	-0.0958	0.2449
13	0.6137	4.8283	4.3477	-0.3009	-0.2931	0.9156
14	2.0598	1.3184	1.1871	0.7012	0.6909	2.3861
15	**4.7220**	3.3343	3.0024	-0.2610	-0.2541	-1.8891
16	0.9308	0.8656	0.7794	0.1765	0.1717	0.2142
17	2.3814	1.5550	1.4003	0.1330	0.1293	0.2242
18	1.7945	1.7458	1.5720	-1.0563	-1.0599	-0.1551
19	**11.2384**	**33.4409**	**30.1120**	**-2.1374**	-2.4046	-1.9882
20	**7.3728**	4.1222	3.7118	0.5735	0.5625	-0.1551
21	**4.0681**	**13.0725**	**11.7712**	-0.2261	-0.2200	**2.5530**
22	2.4318	**11.4354**	**10.2971**	-1.3326	-1.3642	-0.1604

Aykırı değerleri incelemek için kullandığımız grafikler ise sırasıyla Şekil 4.15-Şekil 4.20 olmak üzere sıralanmıştır.

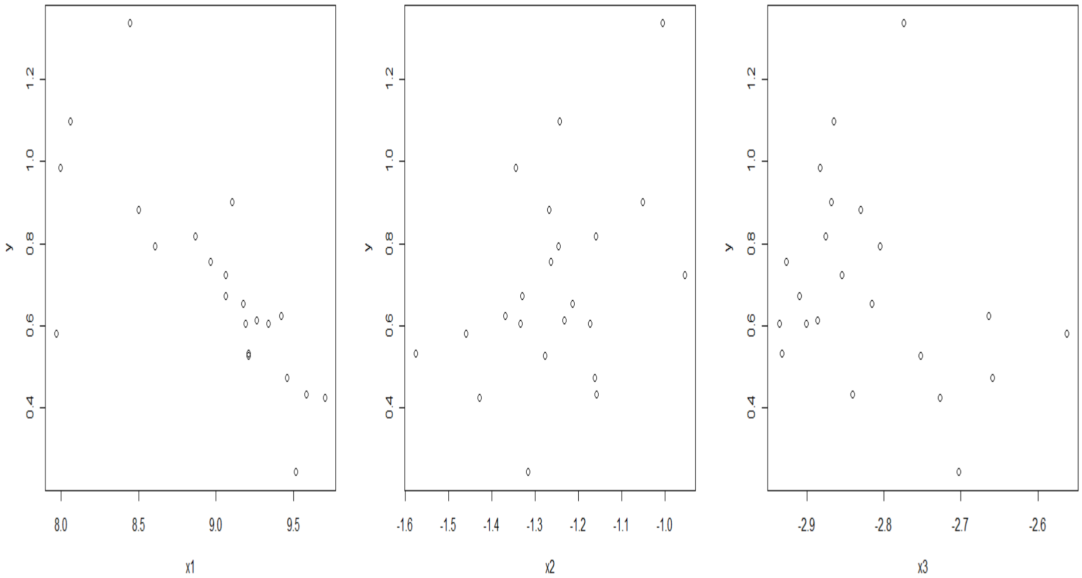


Şekil 4.15 Solow büyüme modeli için student artıklarının indeks grafiği

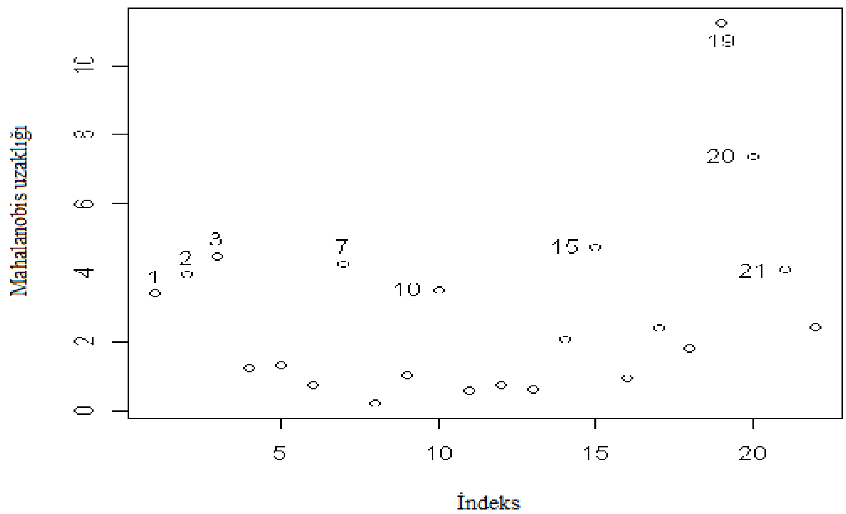


Şekil 4.16 Solow büyüme modeli için standartlaştırılmış artık grafiği

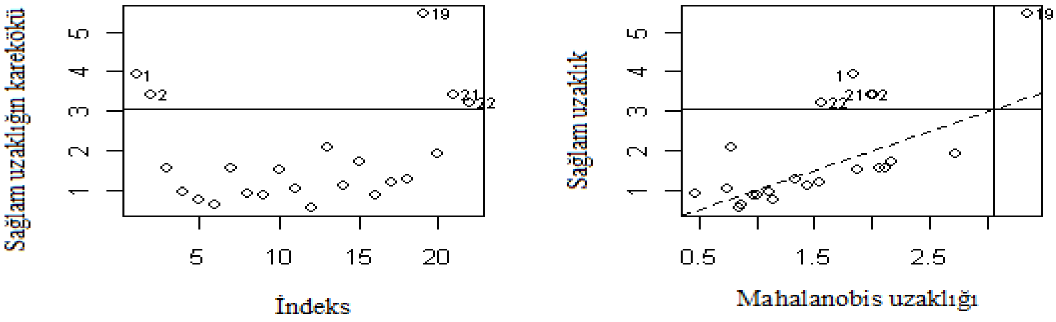
Ön inceleme yapmak amacı ile Solow büyüme modeli verisine ait X_1, X_2, X_3 bağımsız değişkenlerine karşı serpilme grafikleri Şekil 4.17' de verilmiştir.



Şekil 4.17 Solow büyüme modeli için bağımsız değişkenlerin bağımlı değişken ile serpilme grafikleri

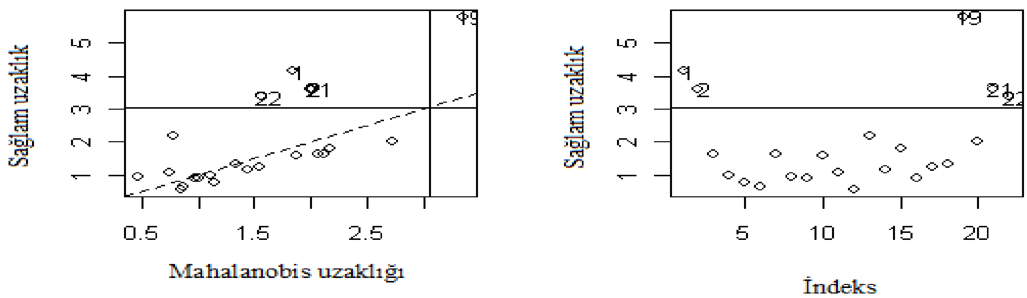


Şekil 4.18 Solow büyüme modeli için Mahalanobis uzaklığı grafiği



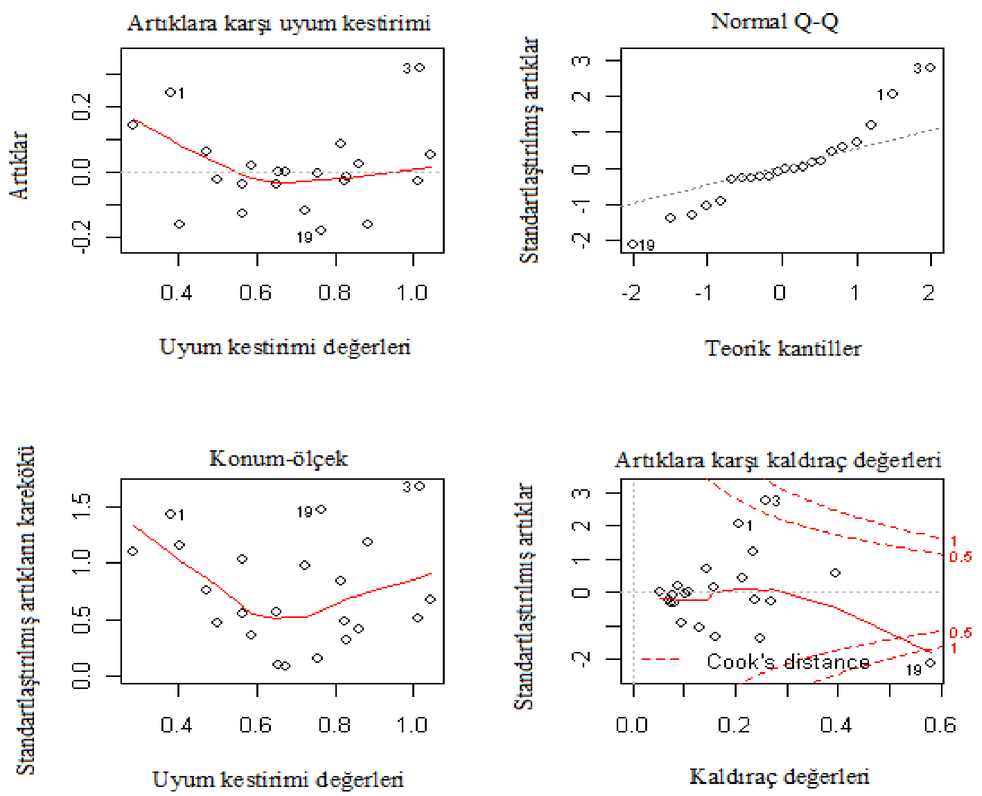
Şekil 4.19 Solow büyüme modeli için MCD' ye göre çizilen grafikler

Mahalanobis uzaklıklarının sağlam uzaklıklardan farklı sonuçlar verdiği hem Tablo 4.6' dan hem de Şekil 4.19 ve Şekil 4.20' den görülmektedir. Sağlam uzaklıkların kareköküne ilişkin indeks grafiğinde numaralandırılmış gözlemlerin veri setinin geri kalanından uzakta kaldığını göstermektedir. Mahalanobis uzaklıklarına karşı çizilen sağlam uzaklıkların grafiğinde eğer kirlenmiş değerler olmasaydı tüm gözlem değerleri kesikli doğru etrafında yer alması beklenirdi. Uygulama verilerimizde noktaların çoğu bu doğru etrafında yer alırken, 5 noktanın doğrudan uzaklaştığı görülmektedir. MVE' ye ve MCD' ye göre çizilen grafiklerin benzer sonuçlar verdiği görülmektedir. Bu durumda bu grafiklerden bir tanesinin seçimi gerekli açıklamaların yapılabilmesi için yeterli olabilecektir.



Şekil 4.20 Solow büyüme modeli için MVE' ye göre çizilen grafikler

Şekil 4.21.a’ da çizilen artıklara karşı uyum kestirimi grafiğinde 1., 3. ve 19. değerlerin verilerin geri kalanından uzakta yer aldıkları gözlenmektedir. Normal Q-Q grafiğinde ise değerlerin bir doğru etrafında yer almadığı, diğer bir değişle veri setinin normal dağılmadığı ve 1., 3. ve 19. gözlemlerin aykırı değer olduğu görülmektedir. Şekil 4.21.c’ de uyum kestirimine karşı standart artıkların karekökü grafiğinde 1., 3. ve 19. gözlemlerin aykırı değer olduğu görülmektedir. Şekil 4.21.d’ de kaldıraç noktalarına karşı standart artık grafiğinde ise 19. gözlemin en etkin gözlem olduğu ve 3. ve 1. gözlemlerin de sırasıyla etkili olduğu görülmektedir.



Şekil 4.21 a,b,c,d Solow büyüme modeli için EKK’ ya dayalı grafikler

a) Artıklara (r_i) karşı $\hat{y}$ uyum kestirimi yapılan değerlerin grafiği
b) Normal Q-Q grafiği
c) Uyum kestirimine karşı standartlaştırılmış artıkların karekök grafiği
d) Kaldıraç değerlerine karşı standart artık grafiği

4.2.1.2 Solow büyüme modeli için etkin gözlemlerin istatistiklerle ve grafiklerle incelenmesi

Tablo 4.7' de Cook uzaklığı 4/22-4 = 0.2222 kritik değerinden büyük olan 1., 3., ve 19. gözlemler regresyon doğrusu üzerinde etkilidir. Bu durumu Şekil 4.21.d' de artıklara karşı çizilen kaldıraç noktası çiziminden de görmek mümkündür. COVRATIO istatistiği için hesaplanan değerler kritik değer $1 \pm 3(4/22) = 1 \pm 0.5455$ aralıklarının dışına çıkan değerlerle karşılaştırıldığında 3., 15., 20. ve 21. gözlemlerin etkin olduğu söylenir. Düzeltilmiş Cook uzaklığı istatistiği için hesaplanan değerlerden 1., 3. ve 19. gözlemler kritik değer $2\sqrt{18/22} = 1.809$' dan büyük olduklarından etkin noktalar olarak, Welsch uzaklığı ölçüsü için de $3\sqrt{4} = 6$ kritik değerinden büyük olan 3. ve 19. gözlemler etkin gözlem olarak nitelendirilir. Güven elipsoidlerine bağlı etkinlik ölçülerinden olan Andrews-Pregibon istatistiği için kritik değer 1-2(3+1)/22=0.6364 ile karşılaştırıldığında bu değerden küçük gözlem değerleri etkin gözlem, diğer bir deyişle 1., 3., 19. ve 20. gözlemler bu ölçüte göre etkin gözlem kabul edilirler.

Tablo 4.7 Solow büyüme modeli için etkin gözlemlerin belirlenmesinde kullanılan istatistikler

i	CD	COVRATIO	CD*	W	AP
1	**0.273**	0.549	**2.4602**	5.9677	**0.6085**
2	0.112	1.167	1.4377	3.5470	0.7041
3	**0.675**	**0.178**	**4.4900**	**11.2627**	**0.4222**
4	1.79e-05	1.403	0.0175	0.0399	0.8958
5	2.13e-06	1.410	0.0060	0.0138	0.8916
6	2.21e-03	1.337	0.1944	0.4381	0.9139
7	0.161	1.055	1.7550	4.3723	0.6701
8	1.36e-06	1.331	0.0048	0.0107	0.9445
9	0.023	1.132	0.6442	1.4621	0.8609
10	0.013	1.526	0.4801	1.1685	0.7790
11	1.10e-03	1.338	0.1367	0.3066	0.9246
12	2.10e-04	1.363	0.0597	0.1345	0.9199
13	1.83e-03	1.331	0.1767	0.3967	0.9207
14	0.021	1.314	0.6000	1.4006	0.8331
15	0.006	**1.697**	0.3281	0.8297	0.7269
16	7.68e-04	1.371	0.1144	0.2590	0.9086
17	8.35e-04	1.488	0.1192	0.2808	0.8403
18	0.042	1.120	0.8726	2.0221	0.8152
19	**1.580**	0.929	**6.0019**	**20.0211**	**0.3129**
20	0.054	**1.935**	0.9673	2.6899	**0.5924**
21	4.02e-03	**1.633**	0.2617	0.6482	0.7587
22	0.085	0.989	1.2688	2.9929	0.7560

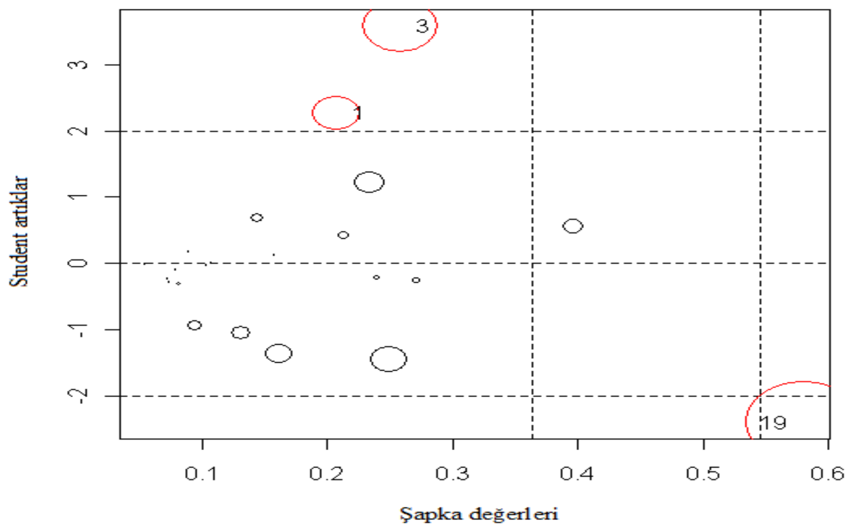
EKK' ya dayalı PRESS istatistiği 0,6609218 olarak bulunmuştur. Tablo 4.8' de DFFITS değerleri için $|DFFITS_i| > 2\sqrt{\frac{k'}{n}}$ kritik değeri ile hesaplanan $2\sqrt{\frac{4}{22}} = 0,8528$' i aşan 1., 3., ve 19. gözlem etkin gözlem olarak tanımlanmaktadır.

DFBETAS değerleri için Tablo 4.8' e bakıldığında 1., 2., 3., ve 19. gözlemlerin $2/\sqrt{22} = 0.4264$ kritik değerini aştığını ve bu nedenle de X_1 değişkenine ait katsayı üzerinde etkili olduğunu söyleyebiliriz. Bu gözlemlerden 3. gözlemin etkisi negatif yöndedir. X_2 değişkenine ait katsayı üzerinde etkili olan 3., 7., ve 19. gözlemlerden 7. gözlem negatif yönde etkilidir. 1., 3., 19., gözlemin X_3 değişkenine ilişkin katsayı üzerinde etkili ve bunlardan 19. gözlem negatif yönde etkilidir. Tüm bu bilgiler Şekil 4.23' den de görülebilmektedir.

Gözlem değerleri birbiri ile kıyaslandığında büyük $|DFTSTAT_{j(i)}|$ değerine sahip gözlemler etkin gözlem olarak nitelendirilirler. DFTSTAT β_0' a göre 3. gözlem, DFTSTAT β_1' a göre 1., 3., 10., 15. ve 19. gözlemler DFTSTAT β_2' e göre 1., 3. ve 19. gözlemler, DFTSTAT β_3' e göre 1., 3. ve 19. gözlemler etkin gözlem olarak ifade edilmektedir.

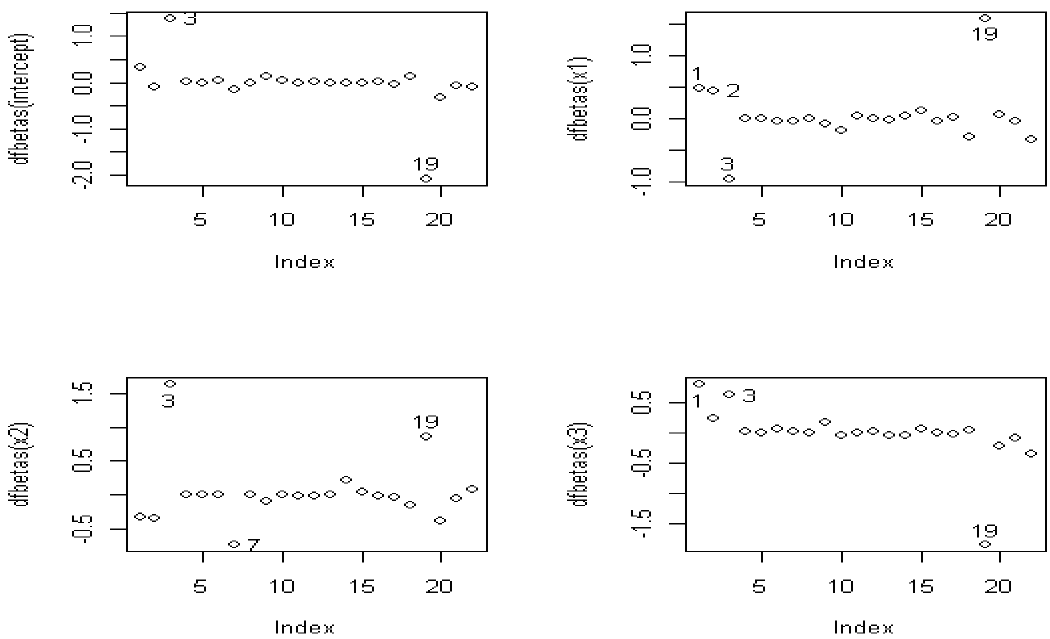
Tablo 4.8- Solow büyüme modeli için etkin gözlemlerin belirlenmesinde kullanılan diğer istatistikler

i	DFFITS	DFBETAS-0	DFBETAS-1	DFBETAS-2	DFBETAS-3	DFTSTAT B0	DFTSTAT B1	DFTSTAT B2	DFTSTAT B3
1	**1.1598**	0.3280	**0.4838**	-0.3236	**0.7907**	0.0384	**0.9900**	**-0.6371**	**0.8410**
2	0.6777	-0.1100	**0.4310**	-0.3329	0.2231	-0.1385	0.1449	-0.2529	0.2150
3	**2.1166**	**1.3900**	**-0.9552**	**1.6233**	**0.6111**	**0.7112**	**0.5621**	**0.9206**	**1.1347**
4	-0.0082	0.0050	0.0003	0.0013	0.0062	0.1448	-0.1711	0.0966	-0.1092
5	0.0028	-0.0019	0.0003	-0.0013	-0.0019	0.1550	-0.1735	0.1271	-0.1075
6	-0.0917	0.0533	-0.0369	-0.0024	0.0468	0.1680	-0.2329	0.0796	-0.0257
7	-0.8273	-0.1490	-0.0519	**-0.7336**	0.0213	-0.2177	0.1226	-0.3875	0.0791
8	-0.0023	7.65e-05	-0.0008	-0.0006	-0.0001	0.0822	-0.1917	0.0961	-0.0559
9	-0.3037	0.1420	-0.0841	-0.0929	0.1616	0.1825	-0.1269	-0.0668	0.1247
10	0.2263	0.0569	-0.1941	-0.0004	-0.0529	0.1466	**-0.8282**	0.0729	-0.1115
11	-0.0644	-0.0263	0.0388	-0.0035	-0.0059	0.0698	-0.2055	0.0827	-0.0592
12	-0.0282	2.28e-03	0.0053	-0.0124	0.0107	0.0844	-0.1731	0.1033	-0.0563
13	-0.0833	-0.019	-0.0297	0.0046	-0.0418	0.0618	-0.2157	0.0878	-0.1113
14	0.2828	9.19e-04	0.0356	0.2156	-0.0490	0.0435	-0.0611	0.3969	-0.0826
15	-0.1546	-0.0120	0.1259	0.0476	0.0506	0.0679	**-0.6612**	0.1851	-0.0393
16	0.0539	0.0133	-0.0372	-0.0046	-0.0064	0.1014	-0.3364	0.0845	-0.0616
17	0.0562	-0.0463	0.0206	-0.0228	-0.0390	0.2026	-0.2214	0.1142	-0.1735
18	-0.4114	0.122	-0.2879	-0.1552	0.0304	0.1307	-0.4701	-0.1305	0.0363
19	**-2.8293**	**-2.090**	**1.5808**	**0.8542**	**-1.8630**	-1.1381	**0.9572**	**0.6115**	**-1.6911**
20	0.4560	-0.312	0.0669	-0.3837	-0.2351	0.1420	-0.0916	0.2937	-0.4045
21	-0.1234	-0.0618	-0.0518	-0.0479	-0.0927	0.1242	-0.3701	0.1110	-0.2888
22	-0.5981	-0.0999	-0.3399	0.0889	-0.3522	-0.1605	-0.3706	0.0196	-0.3585

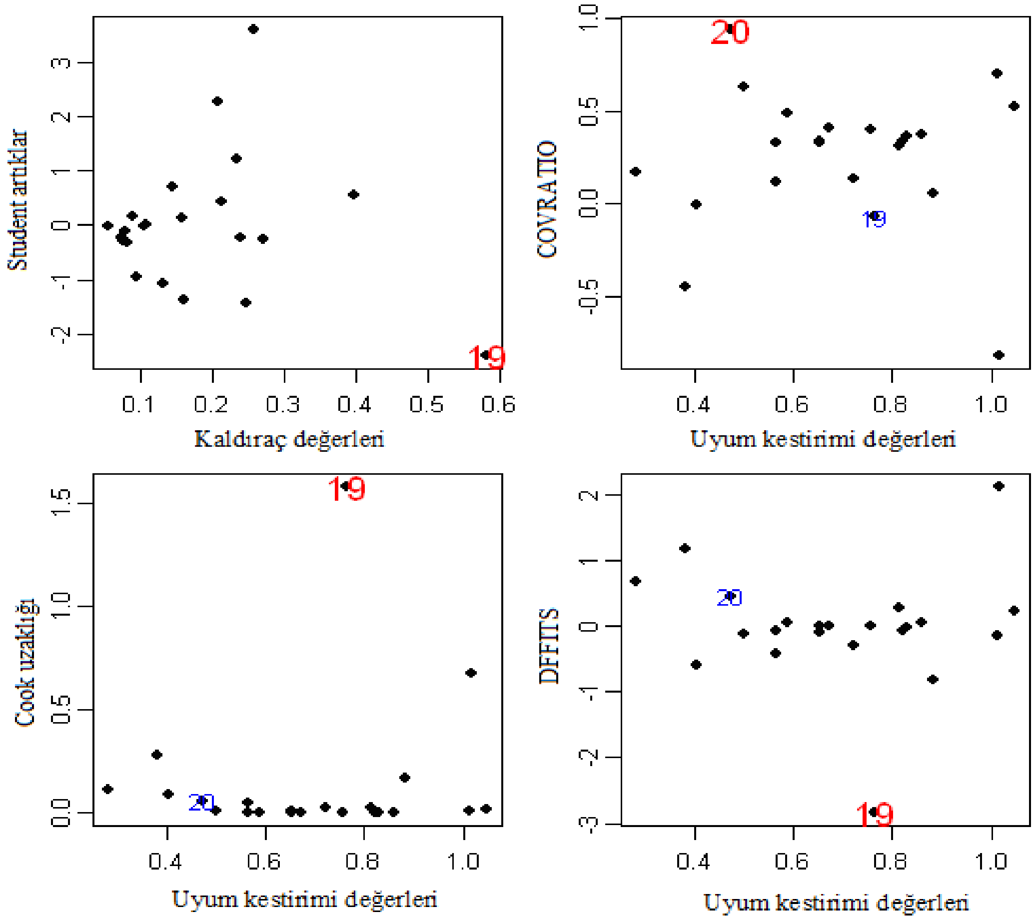


Şekil 4.22 Solow büyüme modeli için etki grafiği

Etki grafiğine göre 19. gözlem en büyük etkiye sahip gözlemdir. Bu gözlemi sırasıyla 3. ve 1. gözlemler izlemektedir.



Şekil 4.23 Solow büyüme modeli için DFBETAS grafikleri



Şekil 4.24 Solow büyüme modeli için bazı teşhis ölçülerinin uyum kestirimi değerlerine karşı grafiği

Şekil 4.24' e göre 19. gözlemin Student artıkları bakımından, 19. ve 20. gözlemlerin COVRATIO, Cook uzaklığı ve DFFITS ölçüleri bakımından etkin gözlem oldukları dikkati çekmektedir.

4.2.1.3 Solow büyüme modeli için çok sayıda etkin gözlemlerin istatistiklerle incelenmesi

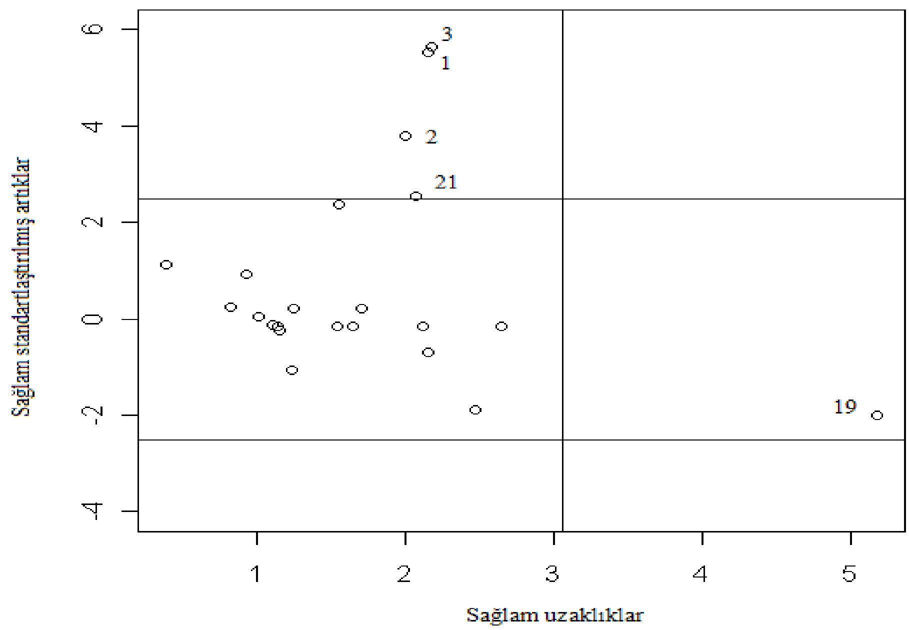
Tablo 4.9 Solow büyüme modeli için çok sayıda etkin gözlemi inceleme de kullanılan istatistikler

Çıkarılan Gözlem grubu	Çoklu COVRATIO	MDFFITS	Çoklu AP
2,1	0.3548	0.0520	0.3354
3,1	0.0428	0.0463	0.2060
19,1	0.8540	0.0262	0.1892
21,1	0.9938	0.0156	0.4383
2,3	0.2344	0.0283	0.2956
19,2	1.2528	0.0425	0.2203
21,2	2.0994	0.0054	0.5204
21,3	0.3140	0.0352	0.3143
2,1,3	0.0172	0.0617	0.0989
2,1,21	0.5629	0.0595	0.2187
2,3,21	0.4015	0.0291	0.2134
1,3,21	0.0610	0.0523	0.1375
1,2,3,21	**0.0130**	**0.0771**	0.0526
1,2,3,21,19	0.0416	0.0479	**0.0148**

Tek sayıdaki gözlem satırının çıkarılması ile belirlenen potansiyel etkin gözlemlerden yararlanarak çok sayıda gözlemler daha önce belirtilen ölçüler kullanılarak Tablo 4.9’ de incelemeye alınmıştır. Tablo 4.9’ den elde edilen sonuçlara göre, çoklu COVRATIO istatistiğine göre çıkarılan gözlem gruplarının her biri diğer gözlem grupları ile karşılaştırıldığında çok büyük ya da çok küçük değerler veren gözlem grubundaki gözlem değerleri etkin gözlem olarak ifade edilmektedir. Bu sebeple çoklu COVRATIO istatistiğine göre 0,0130 değerini elde ettiğimiz gözlem grubuna ait 1., 2., 3. ve 21. gözlemler etkin gözlemdir. MDFFITS istatistiğine göre ise diğer gözlem kümeleri ile karşılaştırdığımızda büyük değerler elde ettiğimiz gözlem kümesi etkin gözlemlere sahiptir ifadesinden yola çıkılarak Tablo 4.9’ da MDFFITS istatistiğine ait enbüyük değer olan 0,0771’ ye karşılık gelen gözlem kümesindeki 1., 2., 3. ve 21. gözlemler etkin gözlemlerdir. Gözlem kümelerini Çoklu Andrews-Pregibon istatistiği için birbiri ile karşılaştırdığımızda en küçük Çoklu Andrews- Pregibon değeri olan 0,0148’ e karşılık gelen gözlem kümesindeki 1., 2., 3., 19. ve 21. gözlemler etkin gözlemlerdir.

4.2.2 Solow büyüme modeli için sağlam regresyon metotları ile elde edilen sonuçlar

Literatürde aykırı değerlerin belirlenmesinde EKK' ya göre çizilen Normal Q-Q grafiklerinin çizimi yerine sağlam uzaklıklara karşı çizilen sağlam standartlaştırılmış artık grafiğine göre 1., 2., 3., 19. ve 21. gözlemler regresyon aykırı değerleri olup 1., 2., 3. ve 21. gözlemler ise y-yönündeki aykırı değer olarak karşımıza çıkmaktadır. Ayrıca 19. gözlem aynı zamanda iyi kaldıraç noktası olarak da ifade edilmektedir (x yönündeki aykırı değerdir). 3.057 kritik değerinin sağında yer alan ve -2.5 ile +2.5 değerleri arasındaki gözlemlerde iyi kaldıraç noktalarıdır. Sağlam standartlaştırılmış artıklar sütununda 19. gözlem yer almamaktadır. Çünkü 19. gözlem yukarıda da belirtildiği gibi iyi kaldıraç noktasıdır.



Şekil 4.25 Solow büyüme modeli için sağlam uzaklıklara karşı sağlam standartlaştırılmış artık grafiği

4.2.2.1 Solow büyüme modeli için sağlam regresyon metotları ile elde edilen katsayı kestirimleri ve ölçek kestirim değerleri

Bu bilgiler ışığında 3. bölümde anlatılan sağlam regresyon değerleri Solow büyüme modeli veri setine uygulandığında elde edilen katsayı kestirim teknikleri, artık standart hata değerleri ve ayrıca R^2 değerleri Tablo 4.10' da verilmiştir.

Tablo 4.10 Solow büyüme modeli verisi için farklı metotlarda elde edilen katsayı kestirimleri, ölçek kestirimi değerleri ve R^2 değerleri.

Metot	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\sigma}_e$	R^2
LAD	2.2628	-0.3122	0.5520	-0.6793	0,1362	0.7987
M Huber	2.5438	-0.3466	0.4850	-0.6595	0.1366	0.6937
M Hampel	1.9301	-0.3179	0.4613	-0.7727	0,1401	0.9530
M Tukey	1.7491	-0.3214	0.4380	-0.8351	0.1433	0.6680
MM	3.9793	-0.3643	0.1341	-0.0568	0.0820	0.5453
LTS	1.4248	-0.3523	0.5026	-1.0702	0.0887	0.8281
LMS	3.0751	-0.3823	0.2767	-0.4941	0.0524	0.9470
GM_{BM}	2.9271	-0.3588	0.4111	-0.5302	0.0818	0.7882
GM_{CH}	2.1291	-0.3424	0.5517	-0.8197	0.1111	0.7619
WLAD	3.0108	-0.4171	0.3949	-0.6774	0.0650	0.7797

Regresyon denkleminin kestiriminde kullanılan M kestiriciler için teorik kısımda belirtildiği gibi birçok alternatif söz konusudur. Kullanılan amaç fonksiyonuna göre, artıklar farklı ağırlıklar kullanılarak minimize edilmektedir. Bu nedenle ortaya çıkan tahminler değişkenlik göstermekle birlikte elde edilen parametre tahminleri, artık standart hata değerleri ve R^2 değerleri Tablo 4.10' da özetlenmiştir. Artık standart hata değerlerine göre en küçük değeri veren sağlam metotlar LMS ve WLAD' dır. Bu metotlardan sonra küçükten büyüğe doğru sıralama yapıldığında artık standart hata değerleri GM_{BM} , MM, LTS, GM_{CH}, LAD, M Huber, M Hampel ve M Tukey' dir. Belirlilik katsayısı (R^2) hesaplamaları sonuçlarına göre modele en iyi uyum sağlayan metot M Hampel olmak üzere diğer metotlar sırası ile, LMS, LTS, LAD, GM_{BM}, WLAD, GM_{CH}, M Huber, Tukey, MM' dir. Dolayısıyla Solow büyüme modelinin sadece y-yönündeki aykırı değerlere sahip olmasından dolayı beklenen sonuçun elde edildiği görülmektedir. M kestirimleri y-yönündeki aykırı değerlere karşı sağlam olduğu için R^2 değeri en yüksek değer olan M Hampel modele en iyi uyan kestirim olarak elde edilmiştir.

4.2.2.2 Solow büyüme modeli için sağlam regresyon metotları ile elde edilen ağırlıkların incelenmesi

Önceki kısımda yer verilen metotlardan elde edilen artıklar ve aldıkları ağırlıklar da Ek 8 ve Ek 9' da verilmiştir. Ek 8' de farklı metotlara göre elde edilmiş yüksek artık değerlerine her metodun farklı ağırlıklar verdiği Ek 9' dan görülmektedir.

Ek 8' de Huber' a göre; 1., 2., 3., 7., 9., 14., 18., 19. ve 22. gözlemler aykırı değer, Hampel' a göre 1., 2., 3., 14. ve 22. gözlemler aykırı değer olarak bulunmuştur. Tukey' e göre en küçük ağırlıklar verilen 1. ve 3. gözlemler aykırı değer olarak tanımlanabilir, diğer gözlemlere ise [0,7401 ile 0,9998] değerleri arasında ağırlık verilmiştir.

LMS' ye göre 1., 3., 14., 19. ve 22. gözlemler aykırı değer, LTS' ye göre 1. ve 3. gözlem aykırı değer olarak bulunmuştur.

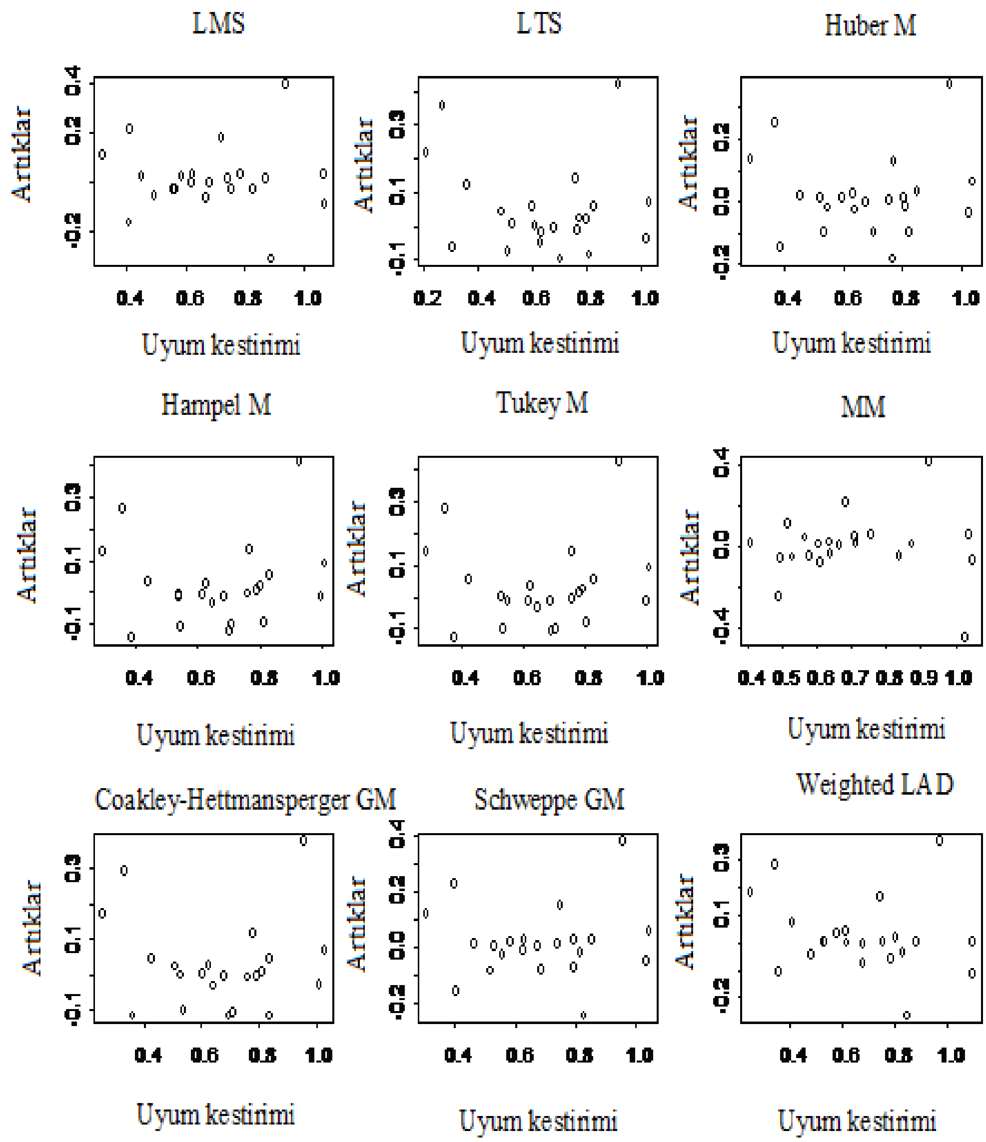
MM' e göre en küçük ağırlık verilmiş iki gözlem olan 1. ve 3. gözlemler aykırı değer olarak tanımlanabilir. Bununla birlikte, diğer gözlemlere [0,7644 ile 0,9999] aralığında ağırlıklar verilmiştir.

GM_{BM}' e göre 1., 2., 3., 7., 9., 14., 18., 19. ve 22. gözlemlere olduğundan düşük ağırlık verilmiştir. GM_{CH}' e göre 19. gözleme sıfıra yakın ağırlık verilmesi sözkonusudur.

WLAD' a göre 19. gözleme sıfıra yakın ağırlık verilmiş olup 1' den düşük ağırlık verilenler ise [0,4518 ile 0,7919] arasında değer almıştır.

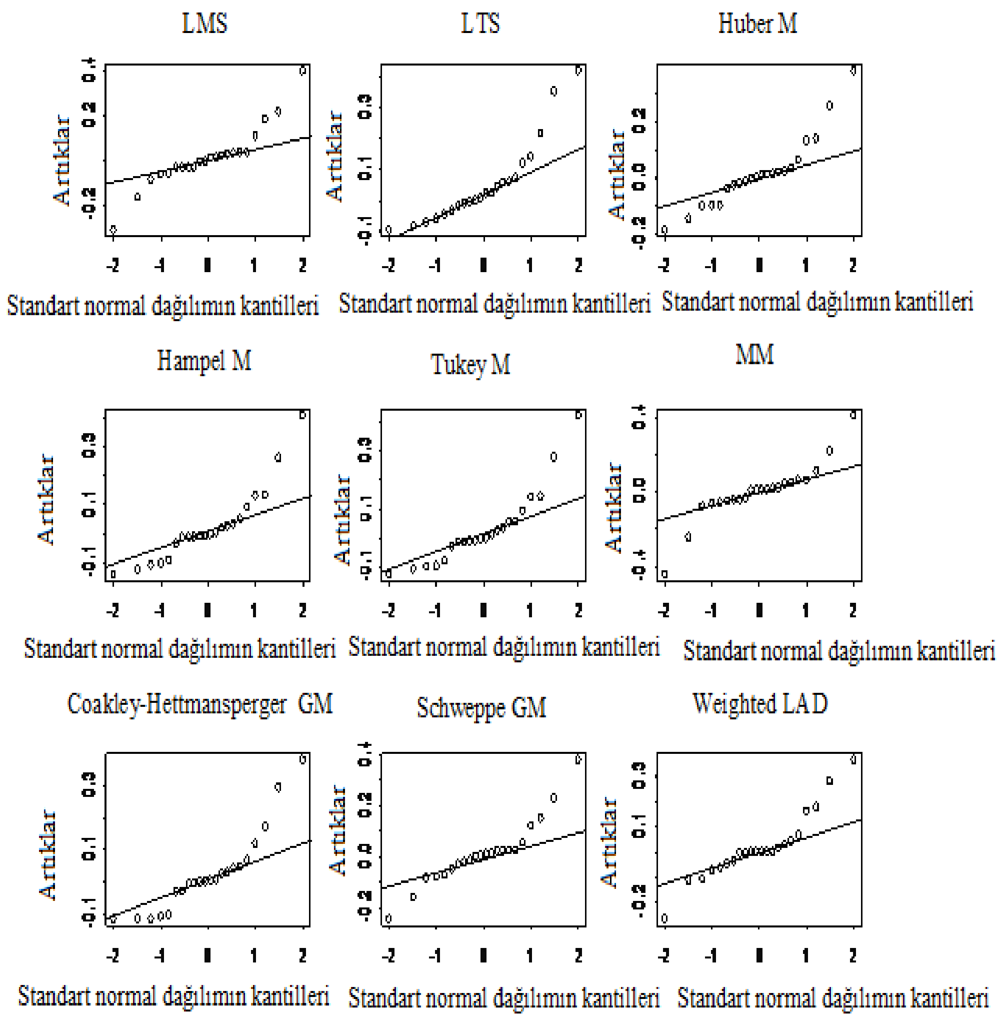
4.2.2.3 Solow büyüme modeli için sağlam regresyon metotları ile elde edilen sonuçların grafiksel incelenmesi

Ele alınan sağlam regresyon metotlarından hesaplanan artıklara karşı uyum kestirimi grafikleri Şekil 4.26' da verilmiştir.



Şekil 4.26 Solow büyüme modeli için artıklara (r_i) karşı $\hat{y}$ uyum kestirimi yapılan değerlerin grafikleri

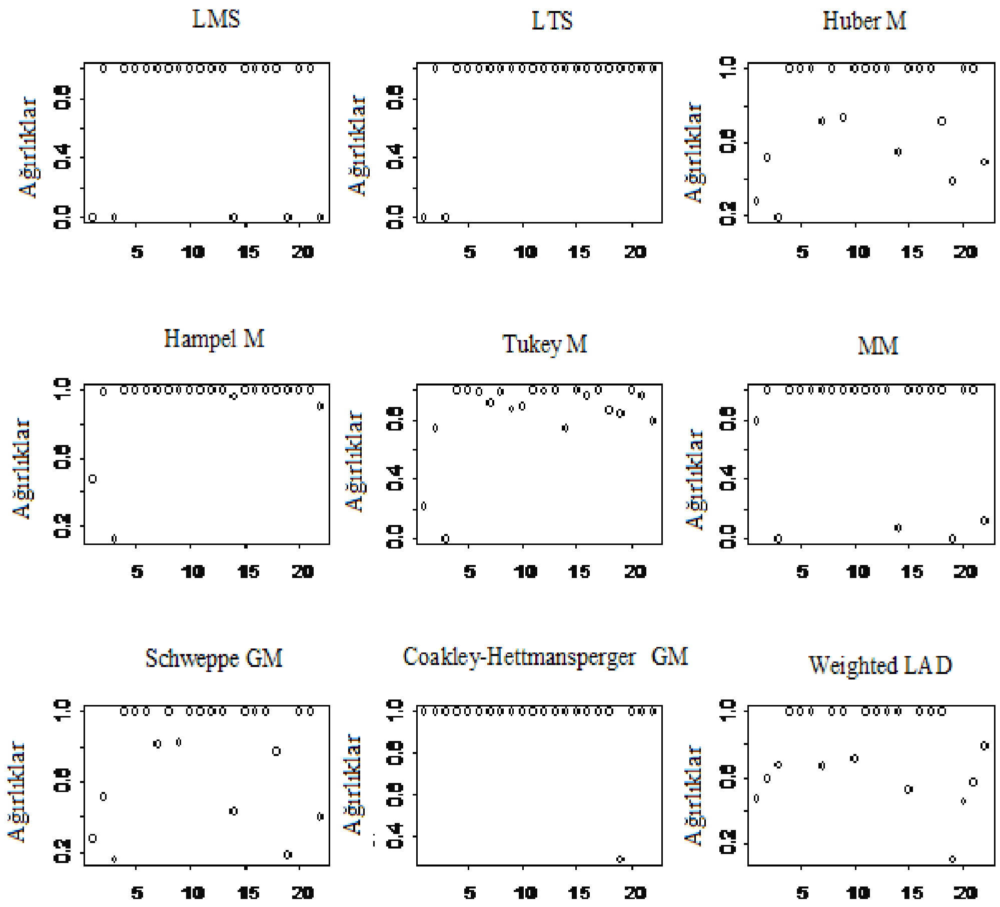
Şekil 4.26' ya göre tüm metotlarda veri setinin çoğunluğundan uzakta yer alan aykırı değerlerin varlığı gözlenmektedir. Ayrıca düzenli gözlemlerin rassal bir serpilme sergilediği söylenebilir.



Şekil 4.27 Solow büyüme modeli için artıkların normal Q-Q grafikleri

Normal Q-Q grafiklerine göre verilerin tüm metotlar için normallik varsayımını sağlamadığı, diğer bir deyişle doğru üzerinde olması beklenen noktaların aykırı değerlerden dolayı doğrudan uzaklaştığı Şekil 4.27' den görülmektedir.

Ek 9' da verilen ve yukarıda yorumu yapılan farklı tekniklere göre hesaplanmış ağırlıkların grafikleri Şekil 4.28' de verilmektedir.



Şekil 4.28 Solow büyüme modeli için farklı sağlam metotların ağırlıklarının grafikleri

BÖLÜM 5

SONUÇ VE ÖNERİLER

Sağlam yöntemlerin uygulama esaslarına dayandırılan bu tezin ilk bölümünde doğrusal regresyon kavramı hakkında genel bilgiler verilerek çoklu doğrusal regresyon modeli ve EKK varsayımları konuları irdelenmiştir. Ardından çoklu doğrusal regresyon hakkında parametre kestirimleri, katsayı çıkarsamaları, ölçü ve kavramlar gibi genel bilgiler verilmiştir.

İkinci bölümde, EKK metodunun varsayımlarının bozulumuna sebep olan ve EKK metodunun aşırı duyarlı olduğu aykırı ve etkin gözlemler tanımlanmıştır. Aykırı ve etkin gözlemlerin ortaya çıkış nedenleri, belirlenmesi, sonuçları, aykırı/etkin gözlemlerin teşhis ölçüleri üzerinde durulmuş ve çok sayıda etkin gözlemin çıkarılmasına dayalı ölçüler irdelenmiştir.

Aykırı gözlemlerin varlığında EKK metodunun uygulanamaması durumunda alternatif regresyon metotlarının geliştirilmesine ihtiyaç duyulmuştur. Bu nedenle üçüncü bölümde aykırı ve etkin gözlemlere dayanıklılığından dolayı EKK metoduna alternatif olarak geliştirilmiş sağlam regresyon metotları üzerinde durulmuştur. Öncelikle üçüncü bölüme sağlam regresyonda temel kavramlar ile başlanmış, sağlam regresyonun istenen özellikleri ile devam edilerek bazı sağlam regresyon metotları üzerinde durulmuştur.

Çalışmanın dördüncü bölümü olan uygulama bölümüne geçildiğinde ise, hem x hem de y-yönünde aykırı değerlerin olduğu ikinci el otomobil satış fiyatları veri seti üzerinde ve ayrıca literatürde yer alan sadece y-yönünde aykırı değerin olduğu Solow büyüme modeli olarak adlandırılan veri seti üzerinde önceki bölümlerde anlatılan bilgilerin ışığında çalışma yürütülmüştür.

5.1 İkinci El Otomobil Fiyatları Veri Seti İçin Elde Edilen Sonuçlar ve Öneriler

İkinci el otomobil satış fiyatları veri seti için ham verilerde ortaya çıkan değişen varyanslılık, otokorelasyon ve normallik testi varsayım bozulumlarının verilerin doğal logaritmaları alındığında ortadan kalktığı görülmüştür. Varsayım bozulumları ortadan kaldırıldığında çoğu araştırmacı EKK ile parametre kestirimlerini elde ederek yorumlama yoluna gidebilmektedir. Ancak varsayım bozulumlarının olmaması doğru kestirim değerlerinin elde edileceği anlamına gelmemektedir. Ayrıca veri setinde aykırı değer ve etkin gözlem araştırılması yapılmalıdır. Bu amaçla veri seti için hesaplanan aykırı değer teşhis ölçülerinden sağlam olan ölçülerin tercih edilmesi önerilmektedir. Sağlam aykırı değer teşhis ölçülerine göre veri setinde hem x hem y-yönünde aykırı değer olduğu belirlenmiştir.

Araştırmacı yapacağı çalışmada etkin gözlemleri belirlemek için analizin farklı özelliklerine dair etkileri hasaplayan ölçüleri tercih etmelidir. Bu amaçla uyum kestirimi değerleri üzerindeki değişim araştırılmak istenirse DFFITS ölçüsü, etkin gözlemin regresyon katsayıları üzerindeki etkisi istenirse DFBETAS, Cook uzaklığı ve düzeltilmiş Cook uzaklığı ölçüsü, kovaryans matrisi kestirimindeki değişme incelenmek istenirse COVRATIO ölçüsü, t istatistik değeri üzerindeki değişim incelenmek istenirse DFTSTAT ölçüsü tercih edilmelidir.

Veri setinde tek aykırı değer ve etkin gözlem varolabilmesinden ziyade birden fazla aykırı ve etkin gözlem de söz konusu olabilir. Çalışmada çok sayıda etkin gözlemin incelenmesinde kullanılan çoklu COVRATIO, MDFFITS ve çoklu AP ölçülerinin birbirleriyle tutarlı sonuçlar verdiği görülmüştür.

Veri setinde aykırıdeğer ve etkin gözlemlerin olduğu ortaya çıkarıldıktan sonra, sağlam metot tercihine geçilmiştir. EKK tekniği için R^2 değeri GM_{BM} ve M Hampel değerlerinden sonra üçüncü sırada büyük değer olarak elde edilmiştir. Sonuç itibariyle aykırı değer varlığında sağlam metotların bağımsız değişkenlerin bağımlı değişkeni açıklamadaki oranının yüksek olduğu çalışmamızca da doğrulanmıştır.

Sağlam metotlar arasında belirlilik katsayısı sonuçları karşılaştırıldığında ise modele en iyi uyum sağlayan metot M Hampel olmak üzere diğer metotlar sırası ile GM_{BM}, $GM_{CH,}$ M Huber' dir. Veri setindeki x ve y yönündeki aykırı değerler düşünüldüğünde her iki yönde de aykırı değerlere daha düşük ağırlık veren Schweppe ağırlıklarıyla Genelleştirilmiş M metodunun tercih edilmesi gerektiği düşünülmektedir.

5.2 Solow Büyüme Modeli Veri Seti İçin Elde Edilen Sonuçlar ve Öneriler

Solow büyüme modeli veri setinde değişen varyanslılık, otokorelasyon varsayımlarının sağlandığı ancak normal dağılım varsayımının sağlanmadığı görülmüştür.

Veri seti üzerinde hesaplanan aykırı değer teşhis ölçülerine bakıldığında ise y yönünde aykırı değerlerin varlığı tespit edilmiştir. Çok sayıda etkin gözlem ölçülerinin sonuçlarında Çoklu COVRATIO ve MDFFITS ölçülerinin aynı gözlemleri etkin gözlem olarak belirlediği, Çoklu AP'nin ise bu etkin gözlemlere bir gözlem eklendiğinde daha küçük değer elde edildiği gözlenmiştir.

Verilerin doğal logaritmaları ile yapılan işlemlerde, sağlam olmayan EKK tekniğine göre elde edilen $\hat{\sigma}_e$ değeri 0.1329 olarak elde edilmiş, sağlam metotlarla artık standart hata değerlerine göre en küçük değeri veren metotlar ise LMS ve WLAD, GM_{BM} olarak bulunmuştur. R^2 değeri ise EKK metodu için 0.7464 bulunmuş, sağlam metotlara göre modeli en iyi açıklayan metot M Hampel için 0.9530 olarak bulunmuştur. Bu sonuca göre bu veri seti için sadece y yönündeki aykırı değerlere düşük ağırlık veren M metodunun tercih edilmesi önerilebilir.

Bu sonuçlara dayandırılarak yapılabilecek öneriler düşünüldüğünde varsayım bozulumlarının olmamasının doğru kestirim değerlerinin elde edileceği anlamına gelmeyeceği, veri setinde aykırı değer ve etkin gözlem araştırılmasının yapılması gerekliliği unutulmamalıdır.

Veri setinde olması muhtemel aykırı değer ve etkin gözlemlerin tek tek değil, aynı zamanda çok sayıda olabileceği durumların da araştırılması gerekmektedir.

Veri setinde kötü kaldıraç noktalarının yanında y-yönlü aykırı değerlerin olması araştırmacının kullanacağı metot seçiminde özenli olmasını gerektirmektedir.

Diğer bir deyişle sadece $\hat{\sigma}_e$ değeri düşük veya R^2' si yüksek metodun seçimi yerine hangi metodun hangi durumda iyi sonuç verdiği göz önüne alınarak metot seçimine gidilmesi iyi bir tercih olacaktır**.**

Elde edilen sonuçlar doğrultusunda sağlam yöntemlerin EKK' ya göre üstünlüğü görülmektedir. Çoğu sağlam regresyon metodu belirli aykırı değerler varlığında en iyi sonucu vermek için tasarlanmıştır. Ancak sağlam teknikler kendi içinde değerlendirildiğinde hangi yöntemin kullanılacağına veri kümesindeki aykırı değer türü ve miktarının dikkate alınması gerektiği unutulmamalıdır.

KAYNAKLAR DİZİNİ

Acock, A. C., 2006, A gentle introduction to Stata, StataCorp LP, 289 p.

Albayrak, A. S., 2006, Uygulamalı çok değişkenli istatistik teknikleri, Asil Yayın, 500 s

Allen, D.M., 1971, In the prediction sum of squares as a criterion for selecting predictor variables, Technical Report no:23, Department of Statistics, University of Kentucky.

Allen, D.M., 1974, The relationship between variable selection and data augmentation and a method for prediction, Technometrics, v16, 125-127p.

Andersen, R., 2008, Modern methods for robust regression, Universiy of Toronto, Sage Publications, Inc., p.2-4, 107 p.

Andrews, D.F. and Pregibon, D., 1978, Finding the outliers that matters, Journal Royal Statistical Society, Series B, v40, 85-93p.

Atkinson, A.C., 1981a, Two graphical displays for outlying and influential observations in regression, Biometrika, v68, 13-20p

Bab-Hadiashar, A., and Suter, D., 2000, Data segmentation and model selection for computer vision, A statistical approach, Springer-Verlag New York,Inc., 208p.

Belsley, D.A., Kuh, E., and Welsch, R.E., 1980, Regression diagnostics: Identifying influential data and sources of collinearity, John Wiley. Sons p.13-24, 292 p.

Bhar, L., (Erişim tarihi: 03.02.2010) Robust regression, I.A.S.R.I., Library Avenue, New Delhi-110 012, Web adresi: Imbhar@iasri.res.in

KAYNAKLAR DİZİNİ (devam ediyor)

Birkes, D., and Dodge, Y., 1993, Alternative methods of regression, New York, John Wiley and Sons, Inc, p.206, 228 p.

Butler, R.W., Davies, P.L. and Jhun, M.,1993, Asymptotics for the minimum covariance determinant estimator, The Annals of Statistics, v21, 1385-1400p.

Carroll, R.J. and Welsh, A.H.. 1988, A note on asymmetry and robustness in linear regression, Amer. Stat. v42, 285-287p.

Chatterjee, S., and Hadi, A.S., 1986, Influential observations, high leverage points and outliers in linear regression, vol.1, no:3, p.379-393

Chatterjee, S., and Hadi, A.S., 1988, Sensitivity analysis in linear regression, John Wiley & Sons, Inc., 317 p.

Chen ,C., (Erişim tarihi 27.01.2010), Robust regression and outlier detection with the robustreg procedure, SAS Institute Inc., Cary, NC, p. 265-27

Coakley C.W., and Hettmansperger T,P., 1993, A bounded influence high breakdown efficient regression estimator, Journal of the American Statistical Association, Sep, 88,423 ABI/INFORM Global p.872 -880

Cohen, J., Cohen, P., West, S, G. and Aiken, L,S., 2003, Applied Multiple Regression / Correlation Analysis for the Behavioral Sciences, Third Edition, Lawrence Erlbaum Associates Inc., New Jersey, 703p.

KAYNAKLAR DİZİNİ (devam ediyor)

Donoho, D.L. and Huber, P.J., 1983, The notion of breakdown point, in A Festschrift for Erich Lehmann P.J. Bickel, K. Doksum, and J.L. Hodges, Jr.(Eds), Wadsworth.

Draper, N,R., and Smith, H., 1981, Applied regression analysis, John Wiley & Sons, 701 p.

Draper, N, R., and John, J. A., 1981, Influential observations and outliers in regression, Technometrics, American Statistical Association and American Society for Quality, Vol. 23, p.21-26

Faraway, J. J., 2009, Texts in statistical science, Linear models with R, Taylor & Francis e- Library, 242 p.

Freund, R. J., and Wilson.W. J., 1998, Regression analysis, statistical modeling of a response variable, Academic Press, 444 p.

Graybill, F.A., 1961, An introduction to linear statistical models, v1, McGraw-Hill, New York.

GroB, J., 2003, Linear regression, Springer-Verlag, Berlin Heidelberg, 394 p.

Gujarati, D.N., 1999, Temel Ekonometri, (Çev. Ü. Şenesen, G. Şenesen), Literatür Yayıncılık, İstanbul, 849 s.

Hampel, F.R., 1971, A generalized qualitative definition of robustness, Ann.Math. Statist., v42, 1887-1896.

KAYNAKLAR DİZİNİ (devam ediyor)

Hampel, F.R., 1974, The influence curve and its role in robust estimation, J. Amer. Statist. Assoc. v69, 383-393p.

Hampel, F.R., Ronchetti, E.M., Rousseeuw P.J., and Stahel W.A., 1986, Robust statistics: the approach based on influence functions, John Wiley & Sons.Inc 502 p.

Handschin, E., Schweppe, F.C., Kohlas, J. and Fiechter, A., 1975, Bad data analysis for power system state estimation, IEEE Trans. Power App. Sys., PAS-94, 329-337p.

Heikkila, J., (Erişim tarihi 05.03.2010) Robust regression, graduate course on advanced statistical signal processing, http://www.ee.oulu.fi/~jth/robust.pdf 38 p.

High, R., 2004, Outlier...s, University of Oregon http://darkwing.uoregon.edu/~robinh/outl.txt (16.02.2010)

Hoaglin D.C., Mosteller F., and Tukey J.W., 1983, Understanding robust and exploratory data analysis, John Wiley & Sons.Inc New York, 447 p.

Hocking, R.R., 2003, Methods and applications of linear models, regression and the analysis of variance, John Wiley & Sons .Inc, Second edition, 741 p.

Hodges, J.L.Jr., 1967, Efficiency in normal samples and tolerance of extreme values for some estimates of location, Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability, University of California, Berkeley, CA.

KAYNAKLAR DİZİNİ (devam ediyor)

Huber, P.J., 1964, Robust estimation of a location parameter, Ann.Math.Statist., v35, 73-101p.

Huber, P.J., 1972, Robust statistics: A review, Ann. Math. Stat., v43, 1041-1067p.

Huber, P.J., 1981, Robust statistics, New York, NY, John Wiley and Sons Inc, 313 p.

Huber, P.J., 2004, Robust Statistics, Hoboken, NJ:Wiley.

Jarrell M. G., 1991 Generating an empirical probability distribution for the Andrews-Pregibon statistic, The University of Alabama, 12 p.

Kardaun, O. J. W.F., 2005, Classical methods of statistics, Springer-Verlag, Berlin Heidelberg, 388 p.

Kennedy, P., 2003, A guide to econometrics, MPG Books, Bodmin, Cornwall, 623 p.

Kıral, G., ve Billor, N., 2001, Bacon temel bileşenler analizi ile sapan değerlerin belirlenmesi, Çukurova Üniversitesi, Adana (http://idari.cu.edu.tr/sempozyum/bil27.htm) (02.01.2010)

Koutsoyiannis, A., 1989, Ekonometri kuramı (Çev: Ü. Şenesen,G. Şensen), Verso Yayıncılık, 688 p.

Kutner, M, H., Nachtsheim C, J., Neter, J. and Li ,W., 2005, Applied linear statistical models, 5. edition, McGraw-Hill Companies, Inc., 1396 p.

Mallows, C.L., 1975, On some topics in robustness, Technical Memorandum, Bell Telephone Laboratories, Murray Hill, NJ.

KAYNAKLAR DİZİNİ (devam ediyor)

Markatou, M. And Hettmansperger, T.P., 1990, Robust bounded influence tests in linear models, J. Amer. Statist. Assoc. 85, 187-190p.

Montgomery D.C., Peck A.E., and Vining G.G., 2001, Introduction to linear regression analysis, third edition, John Wiley & Sons.Inc., 641 p.

Mosteller, F., and Tukey, J.W., 1977, Data analysis and regression, Philippines: Addison-Wesley Publishing Company, 588 p.

Myers H.R., 1990, Classical and modern regression with applications, Duxbury Press Second Edition, 488 p.

Nonneman, W., Vanhoudt, P., 1996, A further Augmentation of the Solow model and the empirics of Economic Growth for OECD Countries, The MIT Press.

R Development Core Team, 2006, R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna.

Ravishanker, N., and Dey, D.K., 2001, A first course in linear model theory, Chapman & Hall/CRC.

Rousseeuw,P.J., and Hubert.M., 1997, Recent developments in PROGRESS in L1-Statistical procedures and related topics,Y, Dodge (ed) The IMS lecture notes –monograph series Vol.31

Rousseeuw P.J., Leroy A.M., 1987, Robust regression and outlier detection, John Wiley & Sons.Inc 329 p.

KAYNAKLAR DİZİNİ (devam ediyor)

Rousseeuw P.J., Leroy A.M., 2003, Robust regression and outlier detection, John Wiley& Sons,Inc., 329 p.

Rousseeuw P.J., Van Zomeren B.C., 1990 Unmasking multivariate outliers and leverage Points, Journal of the American statistical Association Vol: 85, 633-639 p.

Rousseeuw, P.J., 1984, Least median of squares regression, Journal of the American Statistical Association, 79, 871-880

Ryan, T.P., 1997, Modern regression methods, New York, NY, John Wiley and Sons, Inc, 515 p.

Scanlon, S.E., 1996, Residuals and influence in regression, Elsevier Science.

Scheffe, H., 1959, The analysis of variance, Wiley, New York.

Schinka, J.A., and Velicer,W.F., 2003, Handbook of psychology, Volume 2, Research methods in psychology, John Wiley &Sons, Inc, 711p.

Staudle, R, G., and Sheather, S, J., 1990, Robust estimation and testing, A Wiley – Interscience Publication John Wiley & Sons , Inc., New York, 351 p.

Stevens, J., 2002, Applied multivariate statistics for the social sciences, Lawrence Erlbaum Associates, Inc, 4. edition, p. 699

Tabachnick, B.G., and Fidel, L.S., 2001, Using Multivariate Statistics, California State University, Northridge Boston, 966 p.

Tukey, J.W., 1977, Exploratory Data Analysis, Reading, MA: Addison-Wesley.

KAYNAKLAR DİZİNİ (devam ediyor)

Welsch, R.E and Kuh, E.,1977, Linear regression diagnostics, Technical Report 923-77, MIT, Cambridge:MA.

Welsch, R.E., 1982, Influence functions and regression diagnostics, 149-169p. of Launer, R.L. and Siegel, A.F. (Eds), 1982, Modern Data Analysis, Academic Press, New York, 328p.

Wooldridge J. M., 2009, Introductory econometrics, A modern approach, Cengage Learning, Inc, 865p.

Wu, L. L., 1985, Robust M-Estimation of location and regression, Sociological Methology 1985, San Francisco: Jossy-Bass, 316-388 p

Yaffee, R.A., 2002, Robust regression analysis**:** **s**ome popular statistical package options, (http://www.nyu.edu/its/socsci/Docs/RobustReg2.pdf) (08.03.2010)

Yohai , V.J., 1987, High breakdown-point and high efficiency robust estimates for regression, The Annals of Statistics,Vol.15, no:2 p.644

EKLER

Ek 1 İkinci el otomobil satış fiyatları veri seti

i	Marka	(y) (x1000TL)	(x1) (x1000km)	(x2) (yaş)	(x3) (BG)
1	Bentley Continental	131.400	73	7	300
2	Volkwagen Passat TDİ Comfortline	37.500	63	3	144
3	Alfa Romeo 159 JTD Distinctive	45.000	39	3	150
4	Aston Martin Vantage V8	368.000	9.9	3	420
5	Audi A4	20.850	130	10	105
6	Bentley Continenal GT Speed	662.531	95	2	600
7	BMW 520 D Premium Paket	87.000	88	4	163
8	Cadillac BLS 1.9 D	72.563	38	3	140
9	Chery Tiggo 1.6	36.000	16	1	110
10	Volkswagen Passat 2.0 TDİ Comfortline	55.500	47	3	150
11	Chrysler PT Cruiser 2.0	27.000	100	9	145
12	Citroen C3	13.000	114	7	75
13	Dacia Logan 1.5 dcı	14.900	55	2	80
14	Daewoo Lanos 1.5	11.200	180	9	90
15	Daihatsu Terios SX MT AC	17.500	92	6	85
16	Ferrari 360 Modena F1	294.459	8	6	380
17	Ferrari 430 F1	397.519	16.876	5	490
18	Fiat Tempra SXA	8.250	170	17	97
19	Ford Focus	15.500	112	10	100
20	Mazda 3	24.500	9.3	1	90
21	Geely Familia	18.500	13	1	110
22	Samand LX	15.500	37	2	110
23	Alfa Romeo 156 T.Stark Distinctive	27.750	29.988	4	120
24	Chery Alia	25.000	14.5	1	110
25	Hyundai I20 1.4 CRDİ	30.000	17	1	86
26	Opel Astra 1.6 Njoy	33.000	9	1	115
27	Peugeot 207 Outdoor SW	27.000	26	2	90
28	Peugeot 307	24.500	33	3	110
29	Peugeot 207 1.4 HDİ Comfort	25.000	34	2	70
30	Seat Leon 1.6 Stylance	27.000	19	3	102

Ek 1 (devam)

i	Marka	(y) (x1000TL)	(x1) (x1000km)	(x2) (yaş)	(x3) (BG)
31	Lada 112 1.5	9.000	33.5	6	95
32	Seat Ibiza	20.000	26	4	75
33	Toyota Corolla Verso	33.500	30.3	2	110
34	Toyota Corolla SD 1.6 Sol	30.000	21	3	110
35	Toyota Avensis 1.6 Elegant	35.000	27	2	110
36	Tata Marina LX	16.000	32.7	4	70
37	Tata Indigo DLX	15.750	14.18	1	85
38	Tata Indigo SW 1.4	12.000	29	3	75
39	Suzuki Swift 1.3 4x4	24.500	14.5	1	92
40	Suzuki Sx4 1.6 GL	26.500	11	1	105
41	Subaru Impreza Elegance	34.500	22	1	110
42	Subaru Impreza 1.5 Comfort	25.000	29.6	2	70
43	Rover 45	24.000	16	5	109
44	Renault Clio 1.4 Auht	18.500	16	2	98
45	Renault Master Panelvan Uzun Şase	26.750	32	3	115
46	Renault Symbol 1.5 DCİ Expression	24.500	29	1	65
47	Proton Arena 1.5	12.750	12	2	86
48	Proton Waja	22.000	9	1	82
49	Honda city 1.4 Elite	23.500	35	4	86
50	Honda City	21.000	24	4	75
51	Nissan Skystar	21.500	15	7	140
52	Nissan Micra 1.2 Passion	15.750	27	3	80
53	Skoda Favorit 136 L	4.200	28	6	65
54	Nissan Altima	9.800	3.6	17	130
55	Nissan Primera	14.000	15	9	90
56	Nissan Micra	10.000	17	5	65
57	Lada Samara	7.000	20	4	90
58	Lada Vega 1.5	8.000	17	6	110

Kaynak: www.araba.com
Erişim tarihi: 10 /01/ 2010 01:45

Ek 2 Solow büyüme modeli veri seti

Ülkeler	Gsyih85	Gsyih60	Yatırım	Okul	Harcama	Nüfus artışı
Kanada	23060	12361	0.2542	0.106	0.0125	0.0197
ABD	25014	16364	0.2397	0.119	0.0255	0.0154
Japonya	17669	4648	0.3658	0.109	0.024	0.0124
Avusturya	16646	7827	0.2828	0.08	0.011	0.0036
Belçika	16876	8609	0.2645	0.093	0.014	0.0045
Danimarka	19406	10515	0.2915	0.107	0.011	0.0058
Finlandiya	17776	8630	0.3852	0.115	0.012	0.0076
Fransa	18546	9650	0.2972	0.089	0.0205	0.0099
Almanya	17969	9819	0.3095	0.084	0.0245	0.005
Yunanistan	9492	3164	0.2885	0.079	0.002	0.007
İrlanda	12054	5454	0.2877	0.114	0.008	0.0105
İtalya	16055	7086	0.3139	0.071	0.0095	0.0064
Hollanda	16937	10008	0.2789	0.107	0.0205	0.0138
Norveç	22107	8977	0.3494	0.1	0.0145	0.0068
Portekiz	7925	2965	0.2608	0.058	0.0035	0.006
İspanya	11876	4916	0.2817	0.08	0.0045	0.009
İsveç	20826	11364	0.2636	0.079	0.0225	0.0031
İsviçre	22428	14532	0.3142	0.048	0.023	0.0084
Türkiye	5150	2884	0.2323	0.055	0.002	0.0271
İngiltere	17034	10004	0.2067	0.089	0.0225	0.0033
avustra	20617	12824	0.3128	0.098	0.0105	0.02
Yeni Zellanda	17319	13569	0.268	0.119	0.0095	0.017

Source: A further Augmentation of the Solow model and the empirics of Economic Growth for OECD Countries.
Accessed: 09/04/2010 04:03

Ek 3 Aykırı değerlerin belirlenmesinde kullanılan istatistikler

i	Mahalanobis Uzaklığı	MVE Sağlam Uzaklıkları	MCD Sağlam Uzaklıkları	Standartlaştırılmış Artıklar	Student Artıklar	Sağlam standartlaştırılmış artıklar
1	**5.8148**	**16.8331**	**18.4079**	-0.0994	-0.0984	-0.1608
2	1.6295	2.5160	2.9715	-0.6850	-0.6816	-0.4879
3	0.6714	2.0902	2.5993	-0.2066	-0.2047	-0.2750
4	**9.2484**	**43.6682**	**38.8622**	0.9134	0.9120	-0.2222
5	**4.2149**	4.5408	3.3840	0.3723	0.3693	0.9045
6	**18.0481**	**39.4279**	**46.1762**	-0.8216	-0.8191	-0.2222
7	2.7727	4.1815	4.9221	1.1747	1.1789	1.9681
8	0.4641	1.4569	1.8481	1.4297	1.4440	1.8182
9	2.1154	1.3913	1.6132	-0.1229	-0.1217	-0.2969
10	0.9914	2.1763	2.7455	0.2859	0.2834	0.5054
11	**3.2690**	3.9638	4.1421	-0.5827	-0.5791	-0.5054
12	**3.8587**	7.0165	4.4417	0.4401	0.4368	1.1175
13	1.9829	4.9780	3.3030	-0.6293	-0.6258	-0.2222
14	**5.6296**	8.1040	5.4232	-0.8215	-0.8190	-0.2222
15	2.5597	4.2441	2.5795	0.5250	0.5214	1.1104
16	**10.4441**	**50.1995**	**43.0212**	1.7451	1.7799	0.2447
17	**10.2214**	**48.2591**	**44.3746**	0.6956	0.6922	-0.3325
18	**7.0465**	6.9233	5.3320	-1.3380	-1.3480	-1.1966
19	**3.7784**	3.9821	2.9159	-0.1400	-0.1387	0.1572
20	2.7994	1.7474	1.8972	0.0563	0.0558	-0.4707
21	2.2212	1.5228	1.6994	-1.8571	-1.9016	**-2.6561**
22	0.7304	0.9048	0.8546	-1.9516	-2.0054	-2.3022
23	0.1280	0.8617	1.0137	0.0244	0.0242	-0.2882
24	2.1464	1.4175	1.6238	-1.0781	-1.0797	-1.5844
25	2.2114	1.7622	1.6652	0.6012	0.5976	0.7119
26	2.8154	2.7711	2.7079	-0.3610	-0.3581	-1.0660
27	0.4700	0.4693	0.2883	0.6783	0.6749	0.8228
28	0.0648	0.1805	0.2267	-0.2283	-0.2263	-0.3828
29	1.3507	3.5774	2.4495	1.6323	1.6585	2.2927
30	0.2791	0.5350	0.6805	0.6201	0.6166	0.2793
31	0.9613	0.9871	1.0816	-1.4391	-1.4539	-2.2241

Ek 3 (devam)

i	Mahalanobis Uzaklığı	MVE Sağlam Uzaklıkları	MCD Sağlam Uzaklıkları	Standartlaştırılmış Artıklar	Student Artıklar	Sağlam standart laştırılmış artıklar
32	0.9665	1.1816	1.1961	1.5478	1.5686	1.6300
33	0.4553	0.4416	0.4992	0.1989	0.1971	0.2938
34	0.1253	0.5794	0.7069	0.4878	0.4843	0.1767
35	0.3612	0.2964	0.3963	0.3608	0.3579	0.4094
36	1.1568	2.0794	1.6569	1.2042	1.2093	1.3847
37	2.2914	1.5588	1.5758	-1.0264	-1.0269	-1.4886
38	0.6869	1.3032	0.9384	-0.1884	-0.1867	-0.3646
39	2.1741	1.3186	1.3923	-0.2297	-0.2277	-0.4788
40	2.4023	1.6417	1.7595	-0.5769	-0.5733	-1.1624
41	2.2520	1.7525	1.9522	-0.3644	-0.3614	-0.3459
42	1.2095	2.9176	2.0653	1.6842	1.7141	2.2502
43	1.4534	3.4611	3.5234	0.6067	0.6031	-0.1424
44	0.5813	0.3808	0.5150	-0.5731	-0.5695	-1.1899
45	0.0535	0.2873	0.4026	-0.2020	-0.2002	-0.3816
46	**3.4273**	7.3191	5.6936	1.2696	1.2770	2.0189
47	1.3177	1.0315	1.2887	-0.8163	-0.8137	-1.7026
48	3.0338	1.9395	2.1494	0.2482	0.2460	-0.2396
49	0.4733	0.6267	0.3817	1.1773	1.1816	1.3772
50	1.0405	1.1802	1.3013	1.7116	1.7436	1.7736
51	2.8703	9.3030	8.6087	-0.5583	-0.5548	-1.8757
52	0.4830	0.6952	0.5132	0.2482	0.2460	0.1266
53	2.4975	3.0732	3.2894	-1.5480	-1.5688	-2.4110
54	**18.7368**	**41.5826**	**37.7719**	-0.9284	-0.9272	**-3.7127**
55	**4.5184**	6.5140	7.1474	0.8079	0.8053	-0.2159
56	2.8950	3.1985	4.0214	0.8201	0.8176	0.2576
57	0.7369	0.8463	1.1026	**-2.0882**	-2.1578	**-3.2695**
58	1.8609	4.1685	4.2420	**-2.2447**	-2.3354	**-3.8028**

Ek 4 Etkin gözlemlerin belirlenmesinde kullanılan istatistikler

i	CD	COVRATIO	CD*	W	AP
1	3.34e-04	**1.223**	0.1551	0.2914	0.8806
2	5.63e-03	1.091	0.6396	1.1546	0.9459
3	3.19e-04	1.106	0.1515	0.2712	0.9702
4	4.56e-02	**1.234**	1.8264	3.5552	**0.8078**
5	3.48e-03	1.174	0.5008	0.9263	0.9065
6	**8.46e-02**	**1.538**	**2.4828**	5.3640	**0.6578**
7	2.43e-02	1.040	1.3406	2.4458	0.9102
8	1.33e-02	0.948	0.9979	1.7822	0.9377
9	2.17e-04	1.138	0.1250	0.2266	0.9454
10	7.33e-04	1.110	0.2299	0.4125	0.9639
11	6.84e-03	1.135	0.7040	1.2905	0.9196
12	4.49e-03	1.161	0.5698	1.0503	0.9118
13	5.43e-03	1.104	0.6277	1.1368	0.9410
14	2.21e-02	1.159	1.2704	2.3824	0.8729
15	4.57e-03	1.126	0.5747	1.0464	0.9331
16	**1.91e-01**	1.069	**3.8160**	**7.5250**	**0.7544**
17	2.96e-02	**1.294**	1.4660	2.8839	**0.7962**
18	7.34e-02	1.096	2.3372	4.4461	**0.8307**
19	4.47e-04	1.174	0.1793	0.3303	0.9161
20	5.63e-05	1.154	0.0637	0.1162	0.9336
21	5.14e-02	0.877	**1.9870**	3.6064	0.8835
22	2.95e-02	0.829	1.5115	2.7062	0.9015
23	2.97e-06	1.099	0.0146	0.0260	0.9805
24	1.69e-02	1.045	1.1142	2.0209	0.9248
25	5.36e-03	1.111	0.6234	1.1314	0.9376
26	2.33e-03	1.143	0.4096	0.7476	0.9311
27	3.01e-03	1.069	0.4673	0.8347	0.9662

Ek 4 (devam)

i	**CD**	COVRATIO	CD*	W	AP
28	2.44e-04	1.094	0.1326	0.2360	0.9807
29	2.84e-02	0.918	1.4672	2.6417	0.9117
30	2.18e-03	1.071	0.3972	0.7083	0.9709
31	1.83e-02	0.954	1.1698	2.0987	0.9288
32	2.12e-02	0.931	1.2638	2.2675	0.9230
33	2.56e-04	1.102	0.1358	0.2425	0.9741
34	1.18e-03	1.080	0.2920	0.5199	0.9762
35	7.86e-04	1.093	0.2381	0.4249	0.9741
36	1.41e-02	1.004	1.0226	1.8379	0.9366
37	1.61e-02	1.057	1.0855	1.9714	0.9242
38	2.68e-04	1.107	0.1389	0.2485	0.9701
39	7.73e-04	1.136	0.2361	0.4283	0.9437
40	5.25e-03	1.118	0.6168	1.1213	0.9348
41	2.00e-03	1.131	0.3796	0.6891	0.9409
42	2.84e-02	0.903	1.4679	2.6395	0.9110
43	4.11e-03	1.095	0.5457	0.9834	0.9507
44	2.32e-03	1.081	0.4096	0.7324	0.9666
45	1.89e-04	1.094	0.1166	0.2076	0.9811
46	3.38e-02	1.035	1.5834	2.9067	0.8951
47	7.01e-03	1.069	0.7145	1.2861	0.9478
48	1.17e-03	1.154	0.2901	0.5305	0.9285
49	9.08e-03	0.997	0.8192	1.4632	0.9494
50	2.70e-02	0.894	1.4322	2.5715	0.9122
51	5.65e-03	1.129	0.6396	1.1679	0.9270
52	4.07e-04	1.101	0.1712	0.3057	0.9732
53	3.90e-02	0.957	1.7130	3.1170	0.8973
54	**1.14e-01**	**1.545**	**2.8875**	**6.2955**	**0.6436**
55	1.74e-02	1.136	1.1270	2.0906	0.8926
56	1.23e-02	1.100	0.9458	1.7276	0.9204
57	3.39e-02	0.793	1.6296	2.9177	0.8915
58	6.61e-02	0.767	**2.2914**	4.1450	**0.8615**

Ek 5 Etkin gözlemlerin belirlenmesinde kullanılan diğer istatistikler

i	**DFFITS**	**DFBETAS-0**	**DFBETAS-1**	**DFBETAS-2**	**DFBETAS-3**	**DFTSTAT B0**	**DFTSTAT B1**	**DFTSTAT B2**	**DFTSTAT B3**
1	-0.0362	0.0304	-0.0130	-0.0045	-0.0279	**-0.5116**	0.0255	-0.0630	**0.7959**
2	-0.1494	0.0800	-0.1065	0.0537	-0.0575	-0.0397	-0.0687	0.0075	0.0918
3	-0.0354	0.0189	-0.0142	0.0089	-0.0183	-0.1139	0.0102	-0.0474	0.2076
4	0.4266	-0.2605	-0.1562	0.0180	**0.3632**	**-0.6553**	-0.1196	0.0080	**1.6090**
5	0.1170	-0.0190	0.0692	0.0425	-0.0119	-0.1137	0.1221	-0.0409	0.1381
6	**-0.5799**	**0.5274**	**-0.2847**	0.2430	**-0.4996**	-1.1957	-0.1430	-0.0200	**2.1873**
7	0.3131	-0.1979	0.2330	-0.0639	0.1379	-0.2989	0.2619	-0.0517	0.1931
8	0.2331	-0.1040	0.0917	-0.0556	0.1005	-0.0289	0.0743	-0.0026	-0.0325
9	-0.0292	-0.0026	0.0009	0.0215	-0.0027	-0.0977	0.0207	-0.1175	0.1624
10	0.0537	-0.0301	0.0293	-0.0163	0.0261	-0.1726	0.0590	-0.0746	0.2494
11	-0.1644	0.0704	-0.0883	-0.0580	-0.0359	-0.0663	-0.0494	-0.1212	0.1059
12	0.1331	0.0168	0.0812	0.0300	-0.0532	-0.0668	0.1322	-0.0264	0.2072
13	-0.1466	-0.0225	-0.0966	0.0769	0.0467	-0.0863	-0.0581	0.0015	0.1941
14	-0.2967	0.0312	-0.2107	-0.0625	0.0585	-0.0071	-0.1296	-0.0962	0.1565
15	0.1342	0.0081	0.0832	0.0255	-0.0430	-0.0616	0.1236	-0.0200	0.1348
16	**0.8912**	**-0.3919**	**-0.4978**	**0.3366**	**0.6197**	-0.4231	**-0.4409**	**0.3428**	**1.2319**
17	0.3424	-0.2484	-0.0763	0.0529	**0.3059**	**-0.8755**	-0.0525	0.0086	**1.8554**
18	**-0.5459**	0.0467	**-0.2652**	**-0.2812**	0.0935	0.1169	-0.2356	**-0.3535**	0.0033
19	-0.0419	0.0037	-0.0220	-0.0178	0.0067	-0.0923	0.0243	-0.1159	0.1869
20	0.0149	0.0064	-0.0062	-0.0071	-0.0024	-0.1538	0.0268	-0.1054	0.1762
21	-0.4641	-0.0697	0.0854	0.3055	-0.0370	0.1659	0.0358	**0.3644**	-0.4538
22	-0.3530	0.0384	-0.1751	0.2097	-0.0244	0.3161	-0.2256	**0.3346**	-0.5075
23	0.0034	-0.0001	-0.0001	0.0010	0.0004	-0.0945	0.0200	-0.0569	0.1656
24	-0.2602	-0.0308	0.0270	0.1824	-0.0224	-0.0194	0.0244	0.1086	-0.0455
25	0.1456	0.0492	-0.0004	-0.1022	-0.0289	-0.0454	0.0125	-0.2162	0.0963
26	-0.0957	-0.0187	0.0400	0.0483	-0.0100	-0.1145	0.0705	-0.0495	0.1391
27	0.1092	0.0399	0.0136	-0.0468	-0.0330	-0.0289	0.0250	-0.0893	0.0766
28	-0.0310	-0.0003	-0.0076	0.0040	0.0001	-0.0897	0.0127	-0.0473	0.1550
29	0.3427	0.1763	0.1063	-0.1327	-0.1957	0.2806	0.0759	-0.0588	-0.3538
30	0.0928	0.0378	-0.0410	0.0166	-0.0174	-0.0395	-0.0237	-0.0184	0.0906
31	-0.2732	-0.0871	0.0326	-0.1750	0.0892	-0.0014	0.0110	-0.1565	-0.0578

Ek 5 (devam)

i	DFFITS	DFBETAS-0	DFBETAS-1	DFBETAS-2	DFBETAS-3	DFTSTAT B0	DFTSTAT B1	DFTSTAT B2	DFTSTAT B3
32	0.2952	0.2014	-0.0619	0.1119	-0.1888	0.2523	-0.0892	0.1737	-0.2957
33	0.0317	-0.0009	0.0101	-0.0176	0.0019	-0.0916	0.0322	-0.0911	0.1597
34	0.0682	0.0166	-0.0228	0.0079	-0.0028	-0.0614	-0.0050	-0.0336	0.1224
35	0.0556	0.0014	0.0110	-0.0288	0.0030	-0.0808	0.0295	-0.0933	0.1459
36	0.2388	0.1573	0.0006	0.0687	-0.1664	0.1143	-0.0086	0.0838	-0.0753
37	-0.2535	-0.1016	0.0349	0.1609	0.0553	-0.1452	0.0350	0.0929	0.0718
38	-0.0324	-0.0208	-0.0007	-0.0014	0.0207	-0.1734	0.0187	-0.0529	0.2840
39	-0.0551	-0.0173	0.0065	0.0368	0.0070	-0.1356	0.0264	-0.0867	0.1700
40	-0.1441	-0.0352	0.0438	0.0838	-0.0023	-0.1173	0.0633	-0.0125	0.1073
41	-0.0887	0.0008	-0.0184	0.0727	-0.0093	-0.0811	0.0019	-0.0864	0.1384
42	0.3428	0.1994	0.0669	-0.1189	-0.2048	0.3092	0.0301	-0.0308	-0.3894
43	0.1275	0.0438	-0.0836	0.0823	-0.0156	-0.0426	-0.0497	-0.0044	0.0940
44	-0.0957	-0.0411	0.0372	0.0207	0.0173	-0.1308	0.0552	-0.0190	0.1356
45	-0.0272	0.0016	-0.0058	0.0035	-0.0024	-0.0892	0.0144	-0.0484	0.1556
46	0.3698	0.1517	0.1179	-0.2588	-0.1584	0.1385	0.1140	**-0.3350**	-0.1251
47	-0.1669	-0.1044	0.0977	0.0074	0.0581	-0.2174	0.1192	-0.0107	0.1573
48	0.0677	0.0352	-0.0291	-0.0298	-0.0172	-0.1550	0.0048	-0.1206	0.1788
49	0.1913	0.0831	0.0195	0.0507	-0.0916	0.0949	0.0119	0.0663	-0.1029
50	0.3345	0.2342	-0.0940	0.1350	-0.2117	0.3299	-0.1311	0.2243	-0.4083
51	-0.1494	-0.0120	0.0973	-0.1102	-0.0200	-0.0795	0.1420	-0.2538	0.1042
52	0.0400	0.0245	-0.0020	0.0025	-0.0229	-0.1135	0.0170	-0.0477	0.2055
53	-0.4001	**-0.2752**	0.0983	-0.2382	**0.2730**	-0.2893	0.0734	-0.2242	0.2965
54	**-0.6744**	-0.1937	**0.5665**	**-0.5381**	0.0306	**-0.4167**	**0.8025**	**-1.2445**	0.0630
55	0.2632	0.1209	-0.1674	0.2180	-0.0812	-0.0254	-0.1118	-0.0063	0.0647
56	0.2209	0.1697	-0.1138	0.1269	-0.1435	-0.0758	-0.0855	0.0403	0.1760
57	-0.3806	-0.2144	0.1794	-0.1803	0.1540	0.0727	0.1146	-0.0097	-0.3849
58	**-0.5352**	-0.1670	**0.3339**	**-0.3906**	0.0649	0.2162	0.2659	-0.2356	**-0.6377**

Ek 6 Farklı metotlara göre hesaplanmış artıklar

i	e-EKK	e-LAD	e-M_{Huber}	e-M_{Hampel}	e-M_{Tukey}	e-LMS	e-LTS	e-MM	e-GM_{BM}	e-GM_{CH}	e-WLAD
1	-0.0352	-8.8818e-16	-0.0378	-0.0373	-0.0367	-0.1761	-0.0050	-0.0367	-0.0419	-0.0482	0.3492
2	-0.2525	-0.1238	-0.2554	-0.2530	-0.2510	-0.2044	-0.2639	-0.2510	-0.2118	-0.1647	-0.1439
3	-0.0768	-0.0236	-0.0839	-0.0793	-0.0807	-0.0973	-0.1107	-0.0806	-0.0693	-0.0547	0.0519
4	0.3123	0.1514	0.3005	0.3066	0.3015	0.0511	0.2978	0.3015	0.2530	0.1475	0.9241
5	0.1339	0.2081	0.1250	0.1293	0.1246	0.0511	0.0850	0.1247	0.1020	0.1474	0
6	-0.2531	-8.8818e-16	-0.2366	-0.2451	-0.2276	-0.1341	-0.0946	-0.2276	-0.0994	-0.0486	0.5513
7	0.4285	0.5700	0.4274	0.4284	0.4313	0.4580	0.4353	0.4313	0.4697	0.5210	0.5717
8	0.5327	0.5817	0.5248	0.5299	0.5279	0.5120	0.4910	0.5279	0.5362	0.5511	0.6294
9	-0.0451	0.0203	-0.0538	-0.0466	-0.0474	0.0511	-0.1123	-0.0474	-0.0065	0.0125	0.0267
10	0.1060	0.1886	0.1007	0.1044	0.1044	0.1115	0.0823	0.1044	0.1269	0.1537	0.2343
11	-0.2116	-0.1634	-0.2194	-0.2158	-0.2195	-0.3249	-0.2410	-0.2195	-0.2423	-0.2185	-0.1863
12	0.1589	0.2615	0.1490	0.1546	0.1498	0.1474	0.0861	0.1498	0.1411	0.2053	-0.1038
13	-0.2313	-0.0686	-0.2368	-0.2318	-0.2312	-0.0902	-0.2878	-0.2312	-0.1798	-0.1044	-0.3647
14	-0.2915	-0.1516	-0.2976	-0.2945	-0.2969	-0.3000	-0.3326	-0.2969	-0.2947	-0.2184	-0.4899
15	0.1919	0.2824	0.1823	0.1878	0.1835	0.1761	0.1251	0.1835	0.1773	0.2312	0
16	0.5889	0.2975	0.5683	0.5782	0.5651	0.1609	0.5259	0.5651	0.4449	0.2949	1.1006
17	0.2353	0.0876	0.2256	0.2298	0.2254	-0.0640	0.2438	0.2254	0.1705	0.0704	0.8780
18	-0.4680	-0.4258	-0.4796	-0.4745	-0.4822	-0.6186	-0.5308	-0.4822	-0.5338	-0.4942	-0.6808
19	-0.0506	-4.4409e-16	-0.0613	-0.0560	-0.0621	-0.1517	-0.1122	-0.0621	-0.0948	-0.0576	-0.2069
20	0.0205	-8.8818e-16	0.0049	0.0160	0.0098	0.0511	-0.0949	0.0098	0.0137	0.0034	0
21	-0.6809	-0.6482	-0.6916	-0.6834	-0.6857	-0.6135	-0.7595	-0.6857	-0.6573	-0.6518	-0.6087
22	-0.7254	-0.6245	-0.7320	-0.7268	-0.7272	-0.6549	-0.7744	-0.7272	-0.6926	-0.6525	-0.7095
23	0.0091	-0.0197	-0.0048	0.0032	-0.0039	-0.0960	-0.0712	-0.0039	-0.0368	-0.0466	0.0121
24	-0.3955	-0.3456	-0.4052	-0.3975	-0.3990	-0.3130	-0.4682	-0.3990	-0.3640	-0.3515	-0.3236
25	0.2204	0.2949	0.2102	0.2185	0.2165	0.3371	0.1339	0.2165	0.2557	0.2860	0.1773
26	-0.1316	-0.1568	-0.1455	-0.1356	-0.1405	-0.1175	-0.2263	-0.1405	-0.1331	-0.1530	-0.0378
27	0.2527	0.2975	0.2410	0.2491	0.2448	0.2840	0.1659	0.2448	0.2538	0.2769	0.1759
28	-0.0854	-0.05917	-0.0967	-0.0895	-0.0941	-0.1139	-0.1569	-0.0941	-0.0996	-0.0866	-0.1008
29	0.6033	0.6899	0.5921	0.6001	0.5963	0.6841	0.5081	0.5963	0.6159	0.6639	0.4087
30	0.2314	0.1704	0.2142	0.2245	0.2153	0.1298	0.1227	0.2153	0.1751	0.1543	0.1821
31	-0.5338	-0.6023	-0.5522	-0.5421	-0.5533	-0.6953	-0.6458	-0.5533	-0.6178	-0.6330	-0.6715

Ek 6 (devam)

i	e-EKK	e-LAD	e-M$_{Huber}$	e-M$_{Hampel}$	e-M$_{Tukey}$	e-LMS	e-LTS	e-MM	e-GM$_{BM}$	e-GM$_{CH}$	e-WLAD
32	0.5741	0.5217	0.5548	0.5661	0.5549	0.4720	0.4428	0.5549	0.5032	0.4982	0.3583
33	0.0741	0.1435	0.0655	0.0718	0.0699	0.11688	0.0141	0.0699	0.0925	0.1195	0.0905
34	0.1823	0.1372	0.1667	0.1761	0.1681	0.09093	0.0860	0.1681	0.1355	0.1190	0.1679
35	0.1345	0.1857	0.1248	0.1317	0.1290	0.16128	0.0682	0.1290	0.1446	0.1641	0.1511
36	0.4459	0.4295	0.4282	0.4387	0.4288	0.37902	0.3208	0.4288	0.3894	0.4013	0.1974
37	-0.3761	-0.3302	-0.3881	-0.3789	-0.3823	-0.2841	-0.4736	-0.3823	-0.3542	-0.3354	-0.4242
38	-0.0700	-0.0650	-0.0858	-0.0760	-0.0840	-0.0979	-0.1839	-0.0839	-0.1055	-0.0895	-0.2637
39	-0.0843	-0.0347	-0.0954	-0.0867	-0.0894	0.00699	-0.1733	-0.0894	-0.0583	-0.0404	-0.0956
40	-0.2112	-0.2049	-0.2239	-0.2145	-0.2184	-0.1647	-0.3032	-0.2184	-0.2010	-0.2051	-0.1602
41	-0.1336	-0.0179	-0.1392	-0.1336	-0.1320	0.00693	-0.1833	-0.1320	-0.0720	-0.0322	-0.0626
42	0.6233	0.6880	0.6107	0.6195	0.6146	0.68479	0.5205	0.6146	0.6259	0.6649	0.4290
43	0.2240	0.0648	0.2013	0.2138	0.1994	-0.0095	0.0915	0.1994	0.1085	0.0498	0.1663
44	-0.2133	-0.2449	-0.2290	-0.2189	-0.2263	-0.2537	-0.3189	-0.2263	-0.2446	-0.2557	-0.2494
45	-0.0755	-0.0541	-0.0867	-0.0797	-0.0842	-0.1105	-0.1447	-0.0842	-0.0906	-0.0810	-0.0702
46	0.4603	0.6184	0.4529	0.4599	0.4602	0.66475	0.3773	0.4602	0.5253	0.5988	0.2854
47	-0.3018	-0.3790	-0.3213	-0.3091	-0.3194	-0.3757	-0.4351	-0.3194	-0.3579	-0.3838	-0.3980
48	0.0903	0.0644	0.0736	0.0853	0.0783	0.1209	-0.0354	0.0783	0.0783	0.0685	0.0265
49	0.4386	0.4334	0.4234	0.4324	0.4243	0.37119	0.3362	0.4243	0.3934	0.4037	0.2860
50	0.6344	0.5694	0.6144	0.6261	0.6142	0.52122	0.4987	0.6143	0.5578	0.5475	0.4188
51	-0.2035	-0.4194	-0.2277	-0.2151	-0.2312	-0.5272	-0.3300	-0.2312	-0.3480	-0.4347	-0.1706
52	0.0925	0.0864	0.0766	0.0864	0.0783	0.05152	-0.0193	0.0783	0.0538	0.0633	-0.0709
53	-0.5661	-0.6637	-0.5894	-0.5764	-0.5914	-0.7341	-0.7228	-0.5913	-0.6749	-0.6904	-0.8803
54	-0.2834	-0.8483	-0.3296	-0.3066	-0.3412	-0.9838	-0.5304	-0.3412	-0.6180	-0.8384	-0.3508
55	0.2898	0.0379	0.2598	0.2755	0.2545	-0.0639	0.1127	0.2545	0.1075	-0.0218	0.0973
56	0.2988	0.1480	0.2724	0.2873	0.2701	0.0989	0.1221	0.2701	0.1715	0.1323	0
57	-0.7761	-0.8695	-0.7964	-0.7848	-0.7968	-0.9235	-0.9051	-0.7968	-0.8602	-0.8878	-0.9063
58	-0.8258	-1.0009	-0.8494	-0.8367	-0.8521	-1.0887	-0.9614	-0.8521	-0.9541	-1.0179	-0.8936

Ek 7 Farklı metotlara göre hesaplanmış ağırlıklar

i	w- EKK	w-M huber	w-MHampel	w-MTukey	w-LMS	w-LTS	w-MM	w-GM_{BM}	w-GM_{CH}	w-WLAD
1	1	1	1	0.9989	1	1	0.9989	1	**0.2552**	**0.1259**
2	1	1	1	0.9510	1	1	0.9510	**0.8702**	1	**0.7392**
3	1	1	1	0.9949	1	1	0.9949	1	1	**0.8704**
4	1	1	1	0.9296	1	1	0.9297	**0.6751**	**0.1379**	**0.0648**
5	1	1	1	0.9878	1	1	0.9878	1	1	**0.7349**
6	1	1	1	0.9596	1	1	0.9596	1	**0.1031**	**0.0473**
7	1	1	1	0.8588	1	1	0.8589	**0.3881**	**0.9039**	**0.4495**
8	1	**0.8845**	1	0.7925	1	1	0.7927	**0.3473**	1	1
9	1	1	1	0.9982	1	1	0.9982	1	1	1
10	1	1	1	0.9914	1	1	0.9914	1	1	**0.8043**
11	1	1	1	0.9624	1	1	0.9624	**0.7491**	1	**0.5689**
12	1	1	1	0.9824	1	1	0.9824	1	1	**0.5880**
13	1	1	1	0.9583	1	1	0.9584	1	1	**0.8084**
14	1	1	1	0.9318	1	1	0.9318	**0.6018**	1	**0.4804**
15	1	1	1	0.9736	1	1	0.9737	1	1	1
16	1	**0.8168**	1	0.7642	1	1	0.7644	**0.3790**	**0.1236**	**0.0612**
17	1	1	1	0.9604	1	1	0.9604	**0.9910**	**0.1169**	**0.0560**
18	1	**0.9679**	1	0.8252	1	1	0.8254	**0.3276**	1	**0.4584**
19	1	1	1	0.9970	1	1	0.9970	1	1	**0.8531**
20	1	1	1	0.9999	1	1	0.9999	1	1	1
21	1	**0.6712**	1	0.6638	1	1	0.6641	**0.2788**	1	1
22	1	**0.6341**	**0.9416**	0.6268	1	1	0.6271	**0.2682**	1	1
23	1	1	1	0.9999	1	1	0.9999	1	1	1
24	1	1	1	0.8785	1	1	0.8786	**0.5038**	1	1
25	1	1	1	0.9634	1	1	0.9634	**0.7168**	1	1
26	1	1	1	0.9845	1	1	0.9845	1	1	**0.8830**
27	1	1	1	0.9533	1	1	0.9534	**0.7338**	1	1
28	1	1	1	0.9930	1	1	0.9930	1	1	1
29	1	**0.7840**	1	0.7395	1	1	0.7397	**0.2999**	1	1
30	1	1	1	0.9638	1	1	0.9638	1	1	1
31	1	**0.8407**	1	0.7733	1	1	0.7735	**0.3001**	1	1

Ek 7 (devam)

i	**w-EKK**	**w-M huber**	**w-MHampel**	**w-MTukey**	**w-LMS**	**w-LTS**	**w-MM**	**w-GM$_{BM}$**	**w-GM$_{CH}$**	**w-WLAD**
32	1	**0.8366**	1	0.7721	1	1	0.7723	**0.3684**	1	1
33	1	1	1	0.9961	1	1	0.9962	1	1	1
34	1	1	1	0.9779	1	1	0.9779	1	1	1
35	1	1	1	0.9869	1	1	0.9869	1	1	1
36	1	1	1	0.8604	1	1	0.8605	**0.4753**	1	1
37	1	1	1	0.8882	1	1	0.8883	**0.5170**	1	1
38	1	1	1	0.9945	1	1	0.9945	1	1	1
39	1	1	1	0.9937	1	1	0.9937	1	1	1
40	1	1	1	0.9628	1	1	0.9628	**0.9102**	1	1
41	1	1	1	0.9863	1	1	0.9863	1	1	1
42	1	**0.7601**	1	0.7245	1	1	0.7248	**0.2956**	1	1
43	1	1	1	0.9689	1	1	0.9689	1	1	**0.7885**
44	1	1	1	0.9601	1	1	0.9601	**0.7607**	1	1
45	1	1	1	0.9944	1	1	0.9944	1	1	1
46	1	1	1	0.8401	1	1	0.8402	**0.3450**	**0.9610**	**0.4414**
47	1	1	1	0.9212	1	1	0.9213	**0.5164**	1	1
48	1	1	1	0.9952	1	1	0.9952	1	1	1
49	1	1	1	0.8632	1	1	0.8633	**0.4733**	1	1
50	1	**0.7555**	1	0.7248	1	1	0.7251	**0.3322**	1	1
51	1	1	1	0.9583	1	1	0.9584	**0.5235**	**0.5682**	**0.3224**
52	1	1	1	0.9952	1	1	0.9952	1	1	1
53	1	**0.7877**	1	0.7434	**0**	1	0.7437	**0.2709**	1	**0.6559**
54	1	1	1	0.9104	**0**	1	0.9104	**0.2470**	**0.1312**	**0.0716**
55	1	1	1	0.9496	1	1	0.9497	1	**0.6428**	**0.3554**
56	1	1	1	0.9433	1	1	0.9434	1	1	**0.5409**
57	1	**0.5829**	**0.8721**	**0.5623**	**0**	1	**0.5627**	**0.2160**	1	1
58	1	**0.5465**	**0.8179**	**0.5098**	**0**	**0**	**0.5101**	**0.1927**	1	**0.6485**

Ek 8 Solow büyüme modeli için farklı metotlara göre hesaplanmış artıklar

i	e-EKK	e-LAD	e-M_{Huber}	e-M_{Hampel}	e-M_{Tukey}	e-LMS	e-LTS	e-MM	e-GM_{BM}	e-GM_{CH}	e-WLAD
1	0.2423	0.2489	0.2536	0.2621	0.2779	0.2139	0.3559	0.2681	0.2277	0.3306	0.2796
2	0.1410	0.1264	0.1382	0.1299	0.1414	0.1068	0.2169	0.1317	0.1195	0.2150	0.1775
3	0.3188	0.3793	0.3769	0.4096	0.4236	0.3964	0.4218	0.4253	0.3803	0.3921	0.3647
4	-0.0031	0	0.0013	-0.0043	-0.0039	0.0110	-0.0087	-0.0029	0.0117	-0.0018	0
5	0.0011	-0.0035	-0.0037	-0.0118	-0.0116	-0.0074	-0.0052	-0.0122	0.0007	-0.0018	-0.0041
6	-0.0404	-0.0391	-0.0263	-0.0351	-0.0305	-0.0064	-0.0187	-0.0315	-0.0153	0.0008	-0.0034
7	-0.1613	-0.1232	-0.0992	-0.0921	-0.0798	-0.0336	-0.0846	-0.0777	-0.0741	-0.0579	-0.0645
8	-0.0012	0.0121	0.0218	0.0240	0.0331	0.0309	0.0577	0.0306	0.0240	0.0604	0.0417
9	-0.1195	-0.1118	-0.0971	-0.1041	-0.0993	-0.0648	-0.0969	-0.0989	-0.0806	-0.0733	-0.0739
10	0.0526	0.0920	0.0623	0.0902	0.0917	0.0336	0.0720	0.0948	0.0552	0.0155	0
11	-0.0303	-0.0015	-0.0139	0.0050	0.0121	-0.0336	0.0234	0.0112	-0.0223	-0.0124	-0.0370
12	-0.0126	0.0093	0.0131	0.0187	0.0243	0.0326	0.0216	0.0254	0.0234	0.0223	0.0151
13	-0.0385	-0.0258	-0.0203	-0.0136	-0.0019	-0.0336	0.0428	-0.0071	-0.0305	0.0301	-0.0025
14	0.0863	0.1121	0.1312	0.1332	0.1426	0.1802	0.1420	0.1438	0.1513	0.1649	0.1596
15	-0.0296	0	-0.0384	-0.0131	-0.0153	-0.0875	-0.0346	-0.0124	-0.0515	-0.1005	-0.1147
16	0.0224	0.0496	0.0327	0.0513	0.0559	0.0091	0.0596	0.0560	0.0248	0.0205	0
17	0.0162	0	0.0097	-0.0094	-0.0099	0.0195	-0.0009	-0.0110	0.0206	0.0236	0.0281
18	-0.1309	-0.1273	-0.0993	-0.1105	-0.1002	-0.0598	-0.0725	-0.1030	-0.0847	-0.0358	-0.0460
19	-0.1840	-0.1309	-0.1844	-0.1247	-0.1098	-0.3115	-0.0473	-0.1176	-0.2477	-0.1858	-0.2673
20	0.0592	0.0235	0.0124	-0.0083	-0.0148	-0.0336	0.0069	-0.0182	0.0033	0.0015	0
21	-0.0262	0	0.0198	0.0327	0.0537	0.0240	0.1204	0.0460	0.0091	0.1150	0.0672
22	-0.1622	-0.1576	-0.1453	-0.1427	-0.1278	-0.1641	-0.0597	-0.1362	-0.1611	-0.0672	-0.1087

Ek 9 Solow büyüme modeli için farklı metotlara göre hesaplanmış ağırlıklar

i	w- EKK	w-M huber	w-MHampel	w-MTukey	w-LMS	w-LTS	w-MM	w-GM_{BM}	w-GM_{CH}	w-WLAD
1	1	0.2789	0.4784	0.2186	0	0	0.3168	0.2728	1	0.4657
2	1	0.5113	0.9873	0.7429	1	1	0.8003	0.5109	1	0.5853
3	1	0.1882	0.1255	0	0	0	0	0.1580	1	0.6722
4	1	1	1	0.9998	1	1	0.9999	1	1	1
5	1	1	1	0.9981	1	1	0.9982	1	1	1
6	1	1	1	0.9872	1	1	0.9880	1	1	1
7	1	0.7094	1	0.9141	1	1	0.9279	0.8156	1	0.6640
8	1	1	1	0.9849	1	1	0.9887	1	1	1
9	1	0.7271	1	0.8688	1	1	0.8846	0.8231	1	1
10	1	1	1	0.8880	1	1	0.8936	1	1	0.7140
11	1	1	1	0.9980	1	1	0.9985	1	1	1
12	1	1	1	0.9920	1	1	0.9922	1	1	1
13	1	1	1	0.9999	1	1	0.9994	1	1	1
14	1	0.5426	0.9611	0.7401	0	1	0.7644	0.4267	1	1
15	1	1	1	0.9968	1	1	0.9981	1	1	0.5229
16	1	1	1	0.9575	1	1	0.9622	1	1	1
17	1	1	1	0.9986	1	1	0.9985	1	1	1
18	1	0.7113	1	0.8667	1	1	0.8751	0.7671	1	1
19	1	0.3867	1	0.8415	0	1	0.8387	0.1825	0.2923	0.1001
20	1	1	1	0.9970	1	1	0.9960	1	1	0.4518
21	1	1	1	0.9605	1	1	0.9745	1	1	0.5690
22	1	0.4889	0.8978	0.7884	0	1	0.7870	0.3967	1	0.7919