ANADOLU ÜNİVERSİTESİ BİLİM VE TEKNOLOJİ DERGİSİ
ANADOLU UNIVERSITY JOURNAL OF SCIENCE AND TECHNOLOGY
Cilt/Vol.:6Sayı/No: 1 : 37-44 (2005)

ARAŞTIRMA MAKALESİ/**RESEARCH ARTICLE**

VERİLERİN BULANIK VE VERİ KÜMESİNDE AYKIRI DEĞER OLMASI DURUMUNDA BULANIK ROBUST REGRESYON ÇÖZÜMLEMESİ

Kamile ŞANLI, Ayşen APAYDIN[1]

ÖZ

Regresyon çözümlemesinde veri analizi oldukça önemlidir. Çünkü, tek bir gözlem bile regresyon modelindeki parametre tahminleri üzerinde büyük bir etkiye sahip olabilir. Bu gözlemin veri kümesinden çıkartılması regresyon denklemini tamamen değiştirebilir. Bu nedenle büyük artık değere sahip gözlemler ya da aykırı değer, regresyon çözümlemesinde oldukça etkilidir. Veri kümesinde aykırı değer olması durumunda, parametre tahminlerinde robust yöntemler olarak bilinen Huber, Hampel, Andrews ve Tukey'in M yöntemleri kullanılmaktadır. Bu çalışmada çoklu regresyon çözümlemesinde x'in kesin Y'nin bulanık sayı ve veri kümesinde aykırı değer olması durumunda, üyelik fonksiyonu yardımıyla ağırlık matrisi tanımlanmıştır. Regresyon model tahmininde ise bulanık regresyon çözümlemesi kullanılmıştır. Klasik en küçük kareler (EKK), Huber, Hampel, Andrews ve Tukey M yöntemleri ve önerilen bulanık robust yöntem ile regresyon model tahminleri elde edilmiş ve sonuçlar karşılaştırılmıştır.

Anahtar Kelimeler: Robust regresyon, Aykırı değer, Bulanık regresyon, Üyelik fonksiyonu

WHEN DATA ARE FUZZY AND DATA SET HAS OUTLIER, FUZY ROBUST REGRESSION ANALYSIS

ABSTRACT

In regression analysis, data analysis is very important. Because, even one data point (observed) may be large effect over parameters estimates in regression model. If the data point (observed) is removed in data set then regression model is completely change. Therefore, data point have large residual or outlier which fairly effect in regression analysis. In case of data set has outlier, robust methods are used in parameters estimates (Huber, Hampel, Andrews and Tukey M methods). In this paper, when x is crisp, Y is fuzzy data and data set has outlier, weighted matrix will be defined with respect to membership function. In regression model estimate, fuzzy regression analysis will be used. Regression model estimates are obtained with least square methods (LSM), Huber, Hampel, Andrews and Tukey M methods and suggested fuzzy robust method and the results are compared.

Key words: Robust regression, Outlier, Fuzzy regression, Membership function

1 Ankara Üniversitesi, Fen Fakültesi, İsatatistik Bölümü,06100 Tandoğan/ANKARA
e-posta: apaydin@science.ankara.edu.tr

Geliş:17 Şubat 2004; **Kabul:**14 Aralık 2004

1. GİRİŞ

Bir doğrusal regresyon modeli matris gösterimi ile,

$$\underline{Y} = X\underline{\beta} + \underline{\varepsilon} \tag{1}$$

olarak tanımlanır. Burada $\underline{Y}$, nx1 boyutlu bağımlı değişken için gözlemlerin vektörü; X , [nx(m+1)] boyutlu bağımsız değişkenlere ilişkin matris; $\underline{\beta}$, [(m+1)x1] boyutlu regresyon katsayılarının vektörü ve $\underline{\varepsilon}$, (nx1) boyutlu hata vektörüdür.

En küçük kareler tahmin edicisi $\underline{\hat{\beta}}$'yi bulmak için

$$\sum_{i=1}^{n} \varepsilon_i^2 = \underline{\varepsilon}'\varepsilon = (\underline{Y} - X\underline{\beta})'(\underline{Y} - X\underline{\beta}) \tag{2}$$

eşitliğinin en küçüklenmesi gerekir. (2) eşitliğinden parametre vektörü

$$\underline{\hat{\beta}} = (X'X)^{-1} X'\underline{Y} \tag{3}$$

olarak bulunur (Apaydın vd, 2002).

Test istatistiklerinin ve katsayıların belirlenmesinde her bir gözlemin rolüne önem verilmeli ve veri ayrıntılı bir şekilde test edilmelidir. Çünkü elde edilen parametre tahminleri sadece bir gözleme bile bağlı olabilir ve bu gözlemin veriden çıkartılması analiz sonucunu ciddi bir şekilde değiştirebilir. Diğerlerine göre büyük artık değere sahip bu gözlemlere aykırı değer (outlier) denir (Alpar, 1997). Aykırı değer olması durumunda regresyon model tahmininde EKK yöntemine göre daha az etkilenen robust yöntemler kullanılır.

Bulanık regresyon çözümlemesi ve veri kümesinde aykırı değer olması durumuyla ilgili yapılan çalışmalar ele alındığında aşağıdaki gibi özetlenebilir:

Tanaka vd. (1982) bulanık modele sahip doğrusal regresyon çözümlemesindeki ilk çalışmayı önermişlerdir. Bu yaklaşım Bardossy (1990) tarafından genelleştirilmiştir. Optimizasyon problemini çözmek için Tanaka (1987) yaklaşımı çok karmaşıktır. En küçük kareler yaklaşımı açık değildir ve artıklar tarafından en iyi uygunluğun ölçümü tanımlanmamıştır. Tanaka modelinin bu dezavantajlarını gidermek için Dimond (1988) bulanık en küçük kareler yaklaşımını önermiştir (Redden ve Woddall, 1996; Yang ve Ko, 1997).

Tanaka (1987) modeli sıklıkla kesin sayılar ürettiği için Celmins (1987) tarafından eleştirilmiştir. Daha sonra Tanaka vd. (1982) tarafından ileri sürülen minimizasyon probleminin ölçek bağımlı olduğunu Jo'zsef (1992)'de ileri sürmüştür. Celmins (1987)'in eleştirisine karşılık Tanaka ve Ishibuchi (1991) etkileşimli bulanık katsayıları elde etmek için bir yöntem geliştirmişlerdir (Redden ve Woddall, 1996). Bu yöntemin aykırı değerlere güçlü bir şekilde duyarlı olduğu, verinin içerdiği kesin bilgiyi göz ardı ettiği ve sınırsız çözüm ürettiği Redden ve Woddall (1996) tarafından gösterilmiştir.

Sakawa ve Yano (1992) çok amaçlı programlama problemlerini formüle etmek için karar vericiyi tatmin edecek şekilde iteratif, karar vericiyle etkileşime dayanan, etkileşimli algoritmayı ileri sürmüşlerdir. Fakat bu algoritma da aykırı değerlere karşı çok fazla duyarlıdır. Redden ve Woddall (1996) ortogonal en küçük kareler yardımıyla Sakawa ve Yano'nun modelini geliştirmişler ve doğrusal programlama problemini önermişlerdir.

Chang ve Lee (1996) aykırı değer olması durumu için üyelik dereceleriyle ağırlıklandırma yapan ve karar verici ile etkileşime dayanan genelleştirilmiş bulanık ağırlıklandırılmış en küçük kareler yöntemini ileri sürmüşlerdir.

Yan ve Ko (1997) basit regresyon için ağırlıklandırılmış bulanık en küçük kareler çözümlemesinin iteratif algoritmasını ileri sürmüşlerdir. Bu algoritma iki aşamalıdır. İlk olarak gözlemlerin sınıf üyeliklerini veren bulanık sınıflama yöntemi seçilir, daha sonra üyeliklerin bu değerleri ağırlıklar olarak kullanılır. Bulanık regresyon çözümlemesinde ağırlıklandırılmış bulanık en küçük kareler bir optimizasyon problemi olarak düşünülmüştür.

Yang ve Lin (2002) bulanık girdi ve bulanık çıktı için bulanık en küçük kareler doğrusal regresyon çözümlemesini ileri sürmüşlerdir. Heterojen veri kümesi ve aykırı değerleri belirlemek için kümeleme analizinden yararlanmışlardır.

Yang ve Liu (2003) etkileşimli bulanık doğrusal regresyon modelleri için bulanık en küçük kareler algoritmasını önermişlerdir. Bu algoritma basit regresyon için aykırı değere karşı robusttır. Bu algoritmada ortogonal koşullar optimizasyon problemine kısıt olarak eklenmiştir.

Bu çalışmada çoklu regresyon çözümlemesinde X 'in kesin Y 'nin bulanık sayı ve veri kümesinde aykırı değer olması durumunda, üyelik fonksiyonu yardımıyla ağırlık matrisi tanımlanmıştır. Regresyon model tahmininde ise bulanık regresyon çözümlemesi kullanılmıştır.

Çalışmanın İkinci Bölümünde literatürde sıklıkla kullanılan robust yöntemler için tanımlamalar yapılacak, üçüncü bölümde bulanık en küçük kareler yöntemi tanımlanacaktır. Dördüncü bölümde ise bağımsız değişken(ler)in kesin, bağımlı değişkenin bulanık sayı olması ve veri kümesinde aykırı değer olması durumunda üyelik fonksiyonu yardımıyla ağırlık matrisi tanımlanacaktır. Regresyon modelinin tahmininde ise bulanık regresyon çözümlemesi kullanılacaktır. Son Bölümde ise klasik en küçük kareler (EKK), Huber, Hampel, Andrews ve Tukey M yöntemleri ve önerilen bulanık robust yöntem ile ele alınan veri kümesi için regresyon model tahminleri elde edilecek ve sonuçlar karşılaştırılacaktır.

2. ROBUST YÖNTEMLER

Bu kesimde literatürde sıklıkla kullanılan Huber, Hampel, Andrews ve Tukey M yöntemleri için tanımlar verilecektir.

M yöntemi artıkların kareleri toplamını minimum yapmaktan çok artıkların fonksiyonunu minimum yapar. Regresyon katsayıları

$$\sum_{i=1}^{n} \rho\left[(y_i - \sum_{j=1}^{p} x_{ij}\hat{\beta}_j)/d \right] \quad \textbf{(4)}$$

toplamı minimum yapılarak elde edilir. Eşitlik (4)'ün $\hat{\beta}_j$'ya göre türevi alınıp sıfıra eşitlenirse

$$\sum_{i=1}^{n} x_{ij}\psi\left[(y_i - \sum_{j=1}^{p} x_{ij}\hat{\beta}_j)/d \right] = 0 \qquad j = 1,...,p$$

p denklem sistemi için regresyon katsayıları elde edilir.

Huber'in ρ fonksiyonu

$$\rho(z) = \begin{cases} \dfrac{z^2}{2} & |z| \le k \\ k|z| - \dfrac{k^2}{2} & |z| > k \end{cases} \quad \textbf{(5)}$$

$$z = r_i / d$$

$$d = median\left| r_i - median(r_i) \right| / 0.6745$$

biçiminde tanımlanır. Burada k ifadesi tuning sabiti (ayar sabiti) olarak ifade edilir ve k=1.5 değerini alır. d'nin payı genellikle mutlak sapmaların medyanı (MAD) olarak tanımlanır. Eşitlik (5)'in türevi alınırsa

$$\psi(z) = \begin{cases} -k & z < -k \\ z & |z| \le k \\ k & z > k \end{cases} \quad \textbf{(6)}$$

fonksiyonu elde edilir. ψ fonksiyonu ρ'nun türevidir. Eşitlik (5)'de aykırı değere genellikle sıfır ya da sıfıra çok yakın ψ ağırlıkları verilir. Bu nedenle ψ "sıfıra geri azalan" (redescending to zero) olarak nitelendirilir.

Hampel ψ fonksiyonu

$$\psi(z) = (signz)\begin{cases} |z| & 0 \le |z| \le a \\ a & a \le |z| \le b \\ a\left(\dfrac{c-|z|}{c-b}\right) & b \le |z| \le c \\ 0 & c \le |z| \end{cases}$$

biçiminde tanımlanır. Genellikle sabitlerin değerleri a=1.7, b=3.4 ve c=8.5 olarak seçilir.

Andrews (sinüs tahmini) ise ψ fonksiyonunu

$$\psi(z) = \begin{cases} \sin(z/k) & |z| \le k\pi \\ 0 & |z| > k\pi \end{cases}$$

olarak tanımlamıştır, burada k=1.5 ya da k=2.1 alınır.

Tukey'in iki ağırlıklı tahmini için ise ψ fonksiyonu

$$\psi(z) = \begin{cases} z(1-(z/k)^2)^2 & |z| \le k \\ 0 & |z| > k \end{cases}$$

biçiminde tanımlanır. k, 5 ya da 6 olarak seçilir (Hampel vd, 1986; Hogg, 1979; Huber, 1981; Huynh, 1982).

3. BULANIK EN KÜÇÜK KARELER YÖNTEMİ

$X = (m, \underline{m}, \overline{m})$ şeklinde tanımlanan üçgensel bulanık sayıda, m X'in merkezi (modal value), $\underline{m}$ sol yayılma (left spread) ve $\overline{m}$ sağ yayılma (right spead) olarak tanımlanır. Bulanık en küçük kareler yaklaşımı için $X_i = (x_i, \underline{\xi_i}, \overline{\xi_i})$, $Y_i = (y_i, \underline{\eta_i}, \overline{\eta_i})$ üçgensel bulanık sayıları ele alındığında

$$Y = a + bX \quad \textbf{(7)}$$

modeli düşünülebilir. Burada a, b kesin sayılardır. Parametrelerin kesin olduğu model ele alındığında en küçük kareler optimizasyon problemi

$$Minimum \quad r(a,b) = \sum d(a + bX_i, Y_i)^2$$

olarak tanımlanır. $b \ge 0$ ya da $b < 0$ olmasına göre iki durum ortaya çıkar. $b \ge 0$ olması durumunda

$$d(a + bX_i, Y_i) = \left[a + bx_i - y_i - (b\underline{\xi_i} - \underline{\eta_i}) \right]^2 + \left[a + bx_i - y_i + (b\overline{\xi_i} - \overline{\eta_i}) \right]^2 + (a + bx_i - y_i)^2 \quad \textbf{(8)}$$

şeklinde tanımlanır. (8) eşitliğinde $\dfrac{\partial d}{\partial a}=0$ ve $\dfrac{\partial d}{\partial b}=0$ çözümlemesiyle a,b parametreleri elde edilir.

x'in kesin sayı ve $Y=(y,\underline{\eta},\overline{\eta})$ üçgensel bulanık sayı olması durumda bulanık regresyon modeli

$$Y = A + xB \quad \textbf{(9)}$$

olarak tanımlanır. Burada $A=(a,\underline{\alpha},\overline{\alpha})$ ve $B=(b,\underline{\beta},\overline{\beta})$ bulanık parametrelerdir. (9) eşitliğindeki parametre tahminlerini yapabilmek için en küçük kareler optimizasyon problemi

$$Minimum \qquad r(A,B)=\sum d(A+x_iB,Y_i)^2 \quad \textbf{(10)}$$

biçimindedir. Burada

$$d(A+x_iB,Y_i)=(a+bx_i-y_i)^2$$
$$+(a+bx_i-\underline{\alpha}-\underline{\beta}x_i-y_i+\underline{\eta}_i)^2$$
$$+(a+bx_i+\overline{\alpha}+x_i\overline{\beta}-y_i-\overline{\eta}_i)^2$$

olarak tanımlanır. $\dfrac{\partial d}{\partial a}=0$ ve $\dfrac{\partial d}{\partial b}=0$ çözümlemesiyle a ve b parametreleri ve $\dfrac{\partial d}{\partial \alpha}=0$ ve $\dfrac{\partial d}{\partial \beta}=0$ çözümlemesiyle α ve β parametreleri elde edilir [4, 11, 15].

4. VERİLERİN BULANIK VE VERİ KÜMESİNDE AYKIRI DEĞER OLMASI DURUMUNDA BULANIK ROBUST REGRESYON ÇÖZÜMLEMESİ

Verilerin bulanık olması durumunda yapılan çalışmalar genellikle tek bağımsız değişkenin içerildiği basit doğrusal regresyon model tahminleri üzerinedir. Bu çalışmada birden fazla bağımsız değişkeni içeren çoklu doğrusal regresyon modeli ele alınarak eşitlik (10) ile verilen optimizasyon problemi çok değişkenli için genelleştirilmiştir. x'in kesin sayı ve $Y=(y,\underline{\eta},\overline{\eta})$ üçgensel bulanık sayı olması durumda bulanık çok değişkenli regresyon modeli $Y=A+B_1x_{1i}+\ldots+B_px_{pi}$'dir. Bu durumda optimizasyon problemi

$$P: Min\ r(A,B_1,B_2,\ldots,B_p)=$$
$$\sum d\left(A+B_1x_{1i}+\ldots+B_px_{pi},Y_i\right)^2$$

olarak tanımlanır. Burada

$$A=(a,\underline{\alpha},\overline{\alpha}),\ B_1=\left(b_1,\underline{\beta}_1,\overline{\beta}_1\right),\ldots,\ B_p=\left(b_p,\underline{\beta}_p,\overline{\beta}_p\right)$$

biçiminde tanımlanan üçgensel bulanık parametreler olmak üzere

$$d\left(A+B_1x_{1i}+\ldots+B_px_{pi},Y_i\right)=\left(a+b_1x_{1i}+b_2x_{2i}+\ldots+b_px_{pi}-y_i\right)^2$$
$$+\left(a+b_1x_{1i}+b_2x_{2i}+\ldots+b_px_{pi}-\underline{\alpha}-\underline{\beta}_1x_{1i}-\underline{\beta}_2x_{2i}-\ldots-\underline{\beta}_px_{pi}-y_i+\underline{\eta}_i\right)^2$$
$$\left(a+b_1x_{1i}+b_2x_{2i}+\ldots+b_px_{pi}+\overline{\alpha}+\overline{\beta}_1x_{1i}+\overline{\beta}_2x_{2i}+\ldots+\overline{\beta}_px_{pi}-y_i-\overline{\eta}_i\right)^2$$

dir. P probleminin en küçüklenmesi sonucunda $a,\ b_i\ \ (i=1,\ldots,p)$, $\underline{\alpha},\ \overline{\alpha},\ \underline{\beta},\ \overline{\beta}$ değerleri bulunarak bulanık çok değişkenli regresyon modeli tahmin edilir.

4.1. Önerilen Algoritma

Çoklu regresyon çözümlemesinde x'in kesin sayı ve $Y=(y,\underline{\eta},\overline{\eta})$ simetrik üçgensel bulanık sayı olması ve veri kümesinde aykırı değerin olması durumunda regresyon modelinin tahmin edilmesinde önerilen bulanık robust yöntem için algoritmanın adımları aşağıda verilmiştir:

Adım 1: x'in kesin sayı ve $Y=(y,\underline{\eta},\overline{\eta})$ simetrik üçgensel bulanık sayı olması durumda P probleminden başlangıç bulanık tahmini regresyon modeli bulunur.

Adım 2: Merkezler ve yayılmalar için tahmin değerleri ($\hat{y}_j$) ve artıklar (e_i) hesaplanır,

Adım 3: Artıkların mutlak değerlerine göre medyanı belirlenir ve

$$d(i)=\left\|abs(e(i))-medyan(|e_i|)\right\|$$

uzaklığı hesaplanır. Burada $\|.\|$ öklid uzaklığıdır.

Adım 4: Uzaklıklara bağlı olarak üyelik fonksiyonu

$$\mu(x)=\begin{cases} 1 & x\le a \\ \dfrac{b-|e|}{b-a} & a<x<b \\ 0 & d.d \end{cases} \quad \textbf{(11)}$$

biçiminde tanımlanır. Burada;

$$a = medyan(d(i))$$

$$b = \max(d(i))$$

dır.

Adım 5: Eşitlik (11)'de tanımlanan üyelik fonksiyonu ile üyelik dereceleri belirlenir ve ağırlık matrisi oluşturulur. Ağırlık matrisi, köşegen elemanları üyeliklerden oluşan diagonal matristir. Elde edilen ağırlık matrisi yardımıyla ağırlıklandırılmış bulanık en küçük kareler tahminleri bulunur.

Adım 6: Eğer $\left|\hat{\beta}^{k+1}-\hat{\beta}^{k}\right|<\varepsilon$ ise durulur. Aksi durumda Adım 2' ye gidilir. Burada $\hat{\beta}$ tahmin edilen regresyon model parametreleri, k iterasyon sayısı ve $\varepsilon > 0$ olmak üzere çok küçük bir sayıdır.

4.2. Uygulama

Önerilen algoritmayı irdeleyebilmek ve ikinci bölümde verilen klasik yöntemlerle karşılaştırabilmek için Huynh 1982'den 20 gözlem ve 5 bağımsız değişkeni içeren bir örnek ele alınmıştır. Veri kümesi Tablo 1'de verilmiştir. Bu veri kümesi için Bölüm 1'de tanımlanan EKK, Bölüm 2'de tanımlanan Huber, Hampel, Andrews, Tukey M yöntemlerinden ve önerilen yöntemden elde edilen mutlak artıklar Tablo 2'de verilmiştir. Önerilen yöntem için bağımlı değişken değerleri, merkez (y_i), yayılma ise literatürde tanımlandığı gibi $\underline{\eta}_i = y_i/6$ alınarak çözümleme yapılmıştır. M yöntemi ve tanımlanan yöntemin çözümü için Matlab paket programında yazılan programdan yararlanılmıştır. Elde edilen regresyon model tahmini Tablo 3'de, mutlak hatalara ilişkin indeks plot ise Şekil 1-8,'de verilmiştir.

Tablo 1. Veri Kümesi

Gözlem No	x_1	x_2	x_3	x_4	x_5	y	Gözlem No	x_1	x_2	x_3	x_4	x_5	y
1	3.83	28.87	7.20	26.60	6.19	37.01	11	2.09	12.20	-12.86	23.51	5.62	23.30
2	2.89	20.10	-11.71	24.40	5.17	26.51	12	2.52	22.55	0.92	23.60	5.34	35.20
3	2.86	69.05	12.32	25.70	7.04	36.51	13	2.22	14.30	4.77	24.51	5.80	34.90
4	2.92	65.40	14.28	25.70	7.10	40.70	14	2.67	31.79	-0.96	25.80	6.19	33.10
5	3.06	29.59	6.31	25.40	6.15	37.10	15	2.71	11.60	-16.04	25.20	5.62	22.70
6	2.07	44.82	6.16	21.60	6.41	33.90	16	3.14	68.47	10.62	25.01	6.94	39.70
7	2.52	77.37	12.70	24.90	6.86	41.80	17	3.54	42.64	2.66	25.01	6.33	31.80
8	2.45	24.67	-0.17	25.01	5.78	33.40	18	2.52	16.70	-10.99	24.80	6.01	31.70
9	3.13	65.01	9.85	26.60	6.51	41.01	19	2.68	86.27	15.03	25.51	7.51	43.10
10	2.44	9.99	-0.05	28.01	5.57	37.20	20	2.37	76.73	12.77	24.51	6.96	41.01

Tablo 2. Mutlak Hatalar ve Bunlara Karşılık Gelen Ağırlık Matrisinin Köşegen Elemanları

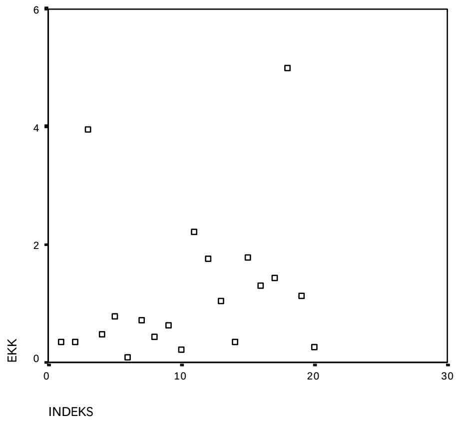
Gözlem No	EKK	Huber		Hampel		Tukey		Andrews		Önerilen Merkez Yayılma			
1	0.35	0.16	1	0.22	1	0.08	1	0.17	0.48	0.2563	0.9928	0.0581	1
2	0.35	0.20	1	0.16	1	0.10	1	0.13	0.48	0.2534	0.9932	0.0583	1
3	**3.95**	**4.19**	**0.26**	**4.17**	**0.15**	**4.27**	**0**	**4.22**	**0.03**	**3.9372**	**0.4452**	**0.6583**	**0.0937**
4	0.47	0.66	1	0.63	1	0.75	0.88	0.68	0.46	0.3996	0.9715	0.0789	0.9928
5	0.78	0.74	1	0.78	1	0.71	0.90	0.75	0.45	0.8164	0.9095	0.1302	0.9133
6	0.09	0.17	1	0.25	1	0.24	0.99	0.24	0.47	0.2771	0.9897	0.0143	1
7	0.72	0.04	1	0.23	1	0.24	0.99	0.24	0.47	0.0236	1	0.1198	0.9293
8	0.43	0.30	1	0.28	1	0.28	0.99	0.28	0.47	0.2759	0.9899	0.0724	1
9	0.62	0.28	1	0.48	1	0.51	0.95	0.50	0.47	0.3166	0.9838	0.1041	0.9537
10	0.21	0.30	1	0.28	1	0.28	0.99	0.27	0.47	0.3256	0.9825	0.0351	1
11	2.21	0.76	1	0.36	1	0.37	0.97	0.37	0.47	0.3792	0.9745	0.3680	0.5442
12	**1.75**	**1.13**	**0.95**	**0.91**	**1**	**1.00**	**0.80**	**0.94**	**0.44**	**0.7898**	**0.9134**	**0.2911**	**0.6635**
13	1.05	0.92	1	0.92	1	0.92	0.83	0.92	0.44	0.9450	0.8904	0.1748	0.8440
14	0.35	0.27	1	0.49	1	0.42	0.96	0.46	0.47	0.5914	0.9430	0.0576	1
15	1.78	0.21	1	0.27	1	0.22	0.99	0.26	0.47	0.2958	0.9869	0.2965	0.6552
16	1.30	1.04	1	1.06	1	0.96	0.81	1.02	0.44	1.2574	0.8439	0.2162	0.7797
17	1.44	1.32	1	1.18	1	1.29	0.68	1.22	0.42	1.1144	0.8651	0.2400	0.7428
18	**5.00**	**6.70**	**0.16**	**7.24**	**0**	**7.14**	**0**	**7.20**	**0**	**7.3087**	**0.0015**	**0.8336**	**0.0155**
19	1.12	0.96	1	1.01	1	0.89	0.84	0.96	0.44	1.3444	0.8309	0.1871	0.8249
20	0.26	0.27	1	0.39	1	0.42	0.96	0.41	0.47	0.1754	1	0.0438	1

Tablo 2 incelendiğinde, aykırı değer olan 18. gözlemin ağırlığı Huber'de "0.16", Hampel, Tukey ve Andrews'te "0", önerilen yöntemde ise merkez için "0.0015", yayılma için "0.0155" olarak bulunmuştur. Önerilen yöntem sonucunda bulunan ağırlıklar aslında her bir gözlemin üyelik dereceleridir. Bu üyelik dereceleri gözlemlerin modele etkisini göstermektedir. Dolayısıyla Tablo 3'ten de görüldüğü gibi aykırı değerler çok küçük üyelik derecesi ile modeli etkilerken diğer gözlemlerin üyelik dereceleri 1 ya da 1'e yakın değerdedir ve bunların tahmini regresyon modeline etkileri önemlidir.

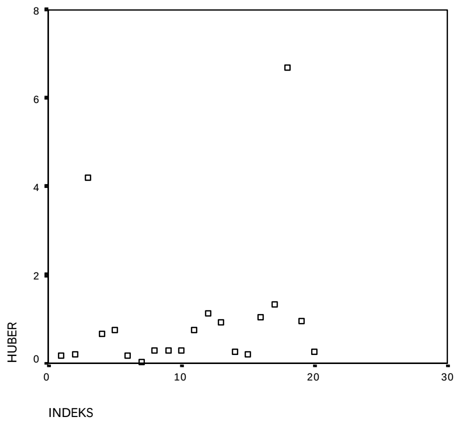
Tablo 3. Regresyon Model Tahmini

Yöntem	Sabit	Regresyon Katsayıları					HKT*
		$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$	
EKK	19.9	-1.79	0.044	0.556	1.11	-1.81	60.2379
Huber	27.7	-1.65	0.075	0.641	1.15	-3.55	70.5059
Hampel	30.5	-1.68	0.085	0.667	1.17	-4.15	77.2119
Tukey	29.3	-1.62	0.082	0.666	1.17	-3.96	76.8528
Andews	30.0	-1.67	0.084	0.667	1.17	-4.08	77.004
Önerilen	(32.212,4.814)	(-1.641,-0.279)	(0.084,0.013)	(0.672,0.109)	(1.141,0.195)	(-4.327,0.642)	(76.9928, 2.0697)

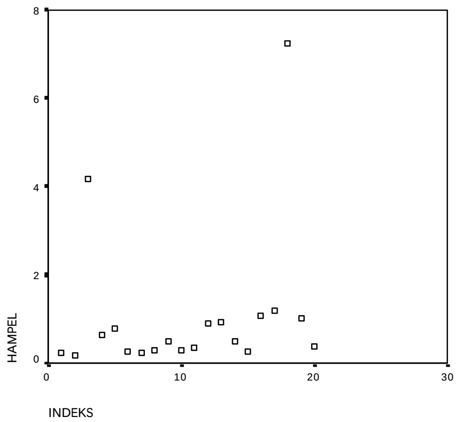
* HKT, Hata kare toplamı



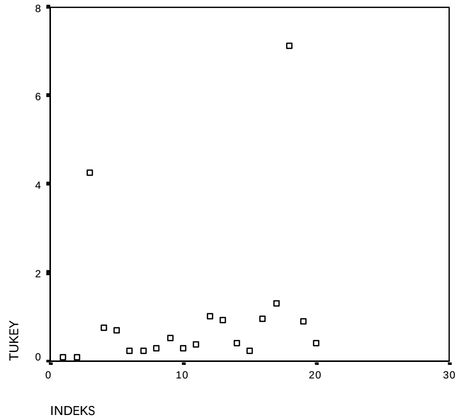
Şekil 1. EKK Mutlak Hata İçin Indeks Plot



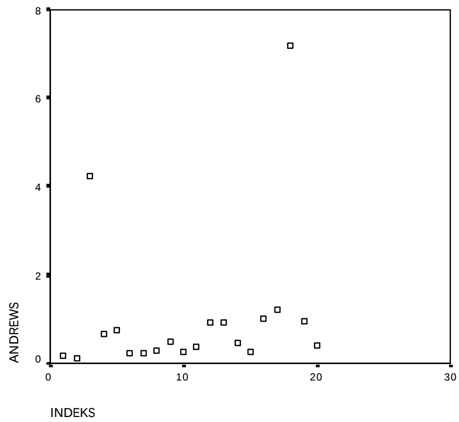
Şekil 2. HUBER Mutlak Hata İçin Indeks Plot



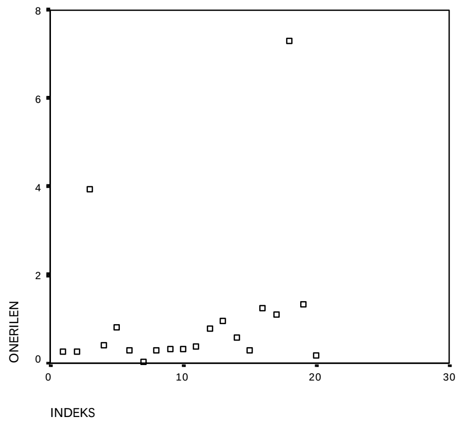
Şekil 3. HAMPEL Mutlak Hata İçin Indeks Plot



Şekil 4. TUKEY Mutlak Hata İçin Indeks Plot



Şekil 5. ANDREWS Mutlak Hata İçin Indeks Pl



Şekil 6. ÖNERİLEN Mutlak Hata İçin Indeks Pl

Tablo 3'de EKK, Huber, Hampel, Tukey, Andrews ve önerilen yöntemle elde edilen regresyon model tahminleri verilmiştir. Önerilen yöntemde bulanık girdi verileri düşünülmüş ve bulanık regresyon parametreleri elde edilmiştir. Böylece parametreler için bir yayılma esnekliği verilmiştir. Önerilen yöntemden elde edilen parametre tahmin değerleri diğer Robust yöntemlerle işaretçe aynı ve $\hat{\beta}$ değerleri çok yakındır.

Robust yöntemler ve önerilen yöntem için çizilen mutlak hata saçılım grafiklerinde, aykırı değere ilişkin hata oldukça büyük değer alırken diğer gözlemlere ilişkin hatalar sıfır etrafında değerler almıştır.

5. SONUÇ ve TARTIŞMA

Bu çalışmada verilerin bulanık olması ve veri kümesinde aykırı değer olması durumda çoklu doğrusal regresyon model tahmininin elde edilmesinde iteratif bir yaklaşım ileri sürülmüştür. Regresyon model tahminlerinin elde edilmesinde bulanık regresyon çözümlemesi kullanılmıştır. Elde edilen parametre tahminleri bulanık sayılar olduğu için tek bir nokta değil belirli bir aralıkta değer almaktadır. Tablo 3 incelenirse $\hat{\beta}_2$ parametresinin merkez değeri 0.084 iken yayılması 0.013'tür. $\hat{\beta}_2$ parametresi merkez değerden $\pm 0,013$ aralığında değerler alabilmektedir. Böylece parametreler için bir yayılma esnekliği verilmiştir Ağırlık matrisi ise uzaklıklara bağlı olarak tanımlanan üyelik fonksiyonundan elde edilmiştir. Her bir gözlem üyelik derecesine göre regresyon model tahminine katılmıştır. Regresyon model tahmininde aykırı değerden EKK yöntemine göre daha az etkilenen model elde edilmiştir. Tablo 2'de verilen mutlak hatalar incelendiğinde önerilen yöntemden elde edilen hataların aykırı değer için (3.9372, 7.3087) gibi büyük değerler, diğer gözlemler için ise (0.0236, 01754) gibi küçük değerler aldığı görülmektedir. Ayrıca aykırı değer için üyelik fonksiyonundan elde edilen üyelik (ağırlık) derecesi diğer gözlemlere göre daha küçüktür. Böylece üyelikler yardımıyla aykırı değerler belirlenebilir. Tablo 3 incelenirse önerilen yöntemden elde edilen regresyon model parametre tahminlerinin literatürde yer alan Robust yöntemlerle işaretçe aynı ve büyüklükçe çok yakın olduğu görülür ve hata kare toplamı klasik robust yöntemlere çok yakın elde edilmiştir. Şekil 6'dan görüldüğü gibi önerilen yöntemde aykırı değer büyük artığa sahipken diğer gözlemler sıfıra yakın artığa sahiptir. Böylece elde edilen regresyon model tahmininin uygunluğunun, aykırı değer dışında kalan gözlemler için daha iyi olduğu söylenebilir. Çoklu regresyon çözümlemesinde gözlemlerin bulanık ve veri kümesinde aykırı değer olması durumunda regresyon model tahmininde önerilen yöntem kullanılabilir.

KAYNAKLAR

Alpar R., 1997. *Uygulamalı Çok Değişkenli İstatistiksel Yöntemlere Giriş I*. Bağırgan Yayınevi.

Apaydın, A. Kutsal, A., ve Atakan, C. (2002). *Uygulamalı İstatistik*. Klavuz Yayınevi, 3. Baskı.

Chang P.T. ve Lee, E.S. (1996). A Generalized Fuzzy Weighted Least-Squares Regression. *Fuzzy Sets and Systems*, 82, 289-298.

Diamond, P. (1988). Fuzzy least Squares. *Information Science*, 46, 141-157.

Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J. ve Stahel W.A. (1986). *Robust Statistics*. John-Willey & Sons, New-York.

Hogg, R.V. (1979). Statistical Robustness: One View of Its Use in Applications Today. *The American Statistican* 33 (3) 108-115.

Huber, P.J. (1981). *Robust Statistics.* John Willey & Son.

Huynh, H. (1982). A Comparison of Four Approaches to Robust Regression. *Psychological Bulletin,* 92 (2) 505-512.

Redden, D.T. ve Woddall, W.H. (1996). Futher Examination of Fuzzy Linear Regression. *Fuzzy Sets and Systems* 79, 203-211.

Sakawa, M. ve Yano, H. (1992). Multiobjective Fuzzy Linear Regression Analysis for Fuzzy Input-Output Data. *Fuzzy Sets and Systems* 47, 173-181.

Xu, R., ve Li, C. (2001). Multidimensional Least-Squares Fitting with a Fuzzy Model. *Fuzzy Sets and Systems* 119, 215-223.

Yang, M.S. ve Ko, C.H. (1997). On Cluster-Wise Fuzzy Regression Analysis, IEE Transaction on Systems, *Man and Cybernetics Part B: Cybernetics* 27 (1) 1-13.

Yang, M.S. ve Lin, T.S. (2002). Fuzzy Least-Squares Linear Regression Analysis for Fuzzy Input-Output Data. *Fuzzy Sets and Systems* 126, 389-399.

Yang, M.S. ve Liu, H.H. (2003). Fuzzy Least Squares Algorithms for Interactive Fuzzy Linear Regression Models. *Fuzzy Sets and Systems* 135, 305-316.

Yun-Hsi O.C. (2001). Hybrid Fuzzy Least-Squares Regression Analysis and Its Reliability Measures. *Fuzzy Sets and Systems* 119, 225-246.

Ayşen Apaydın, Lisans öğrenimini Hacettepe Üniversitesi İstatistik Bölümü'nde 1979 yılında, Yüksek Lisans öğrenimini 1981, Doktorasını 1987 yılında Hacettepe Üniversitesi İstatistik Bölümü'nde tamamladı. 1996 yılında Doçent, 2002 yılında "Yöneylem Araştırması" Anabilim Dalında Profesör oldu. 1988 yılından bu yana Ankara Üniversitesi Fen Fakültesi İstatistik Bölümü'nde öğretim üyesi olarak görev yapmaktadır. Evli ve iki çocuk annesidir.

Kamile Şanlı, Lisans öğrenimini Osmangazi Üniversitesi Fen-Edebiyat Fakültesi İstatistik Bölümü'nden 1997 yılında, Yüksek Lisans Öğrenimini 1999 yılında Ankara Üniversitesi Fen Bilimleri Enstitüsü İstatistik Anabilim Dalında tamamladı. Aynı yıl Ankara Üniversitesi Fen Bilimleri Enstitüsü İstatistik Anabilim Dalında Doktora öğrenimine başladı. 1997-1999 yılları arasında Osmangazi Üniversitesi Fen-Edebiyat Fakültesi İstatistik Bölümü'nde Araştırma Görevlisi olarak görev yaptı. 1999 yılından bu yana Ankara Üniversitesi Fen Fakültesi İstatistik Bölümü'nde Araştırma görevlisi olarak görev yapmaktadır.