G.Ü. Fen Bilimleri Dergisi ISSN 1303-9709 *G.U. Journal of Science*
17(3):71-84 (2004) *17(3):71-84 (2004)*

THE FUZZY ROBUST REGRESSION ANALYSIS, THE CASE OF FUZZY DATA SET HAS OUTLIER

Kamile ŞANLI, Ayşen APAYDIN*
Ankara University, Faculty of Science, Department of Statistics,06100, Ankara, TÜRKİYE
e.mail:sanli@science.ankara.edu.tr, apaydin@science.ankara.edu.tr

ABSTRACT

In regression analysis, data analysis is very important. Because, even one observation (data point) may be large effect over parameters estimates in regression model. If the observation is removed in data set then regression model is completely change. Therefore, observation has large residual or outlier which fairly effect in regression analysis. In case of data set has outlier, robust methods are used in parameter estimates. In this paper, when input data are fuzzy ($X_i = (x_i, \underline{\xi_i}, \overline{\xi_i})$, $Y_i = (y_i, \underline{\eta_i}, \overline{\eta_i})$) and data set has outlier in multi regression analysis, 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 squares method (LSM), robust methods and suggested fuzzy robust method and the results will be compared.

Key Words: Robust Regression, Outlier, Fuzzy Regression, Membership Function

GİRDİ DEĞİŞKENLERİNİN BULANIK OLMASI DURUMUNDA ROBUST REGRESYON ÇÖZÜMLEMESİ

ÖZET

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 kullanılmaktadır. Bu çalışmada çoklu regresyon çözümlemesinde girdi değişkenlerinin bulanık sayı ($X_i = (x_i, \underline{\xi_i}, \overline{\xi_i})$, $Y_i = (y_i, \underline{\eta_i}, \overline{\eta_i})$) 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), robust yöntemler ve önerilen bulanık robust yöntem ile regresyon model tahminleri elde edilmiş ve sonuçlar karşılaştırılmıştır.

Anahtar Kelimeler: Robust Regresyon, Aykırı Değer, Bulanık Regresyon, Üyelik Fonksiyonu.

1. GİRİŞ

Bir doğrusal regresyon modeli matris gösterimi ile,

$$\underline{Y} = X\underline{\beta} + \underline{\varepsilon} \quad [1]$$

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.

1. INTRODUCTION

A linear regression model is given by matrix notation as,

$$\underline{Y} = X\underline{\beta} + \underline{\varepsilon} \quad [1]$$

where $\underline{Y}$ is a nx1 vector of response variable, X is a [nx(m+1)] matrix of independent variables, $\underline{\beta}$ is a [(m+1)x1] vector of regression coefficient, $\underline{\varepsilon}$ is a (nx1) vector of random error.

The least squares estimates for the regression coefficients $\hat{\underline{\beta}}$ are obtained by minimizing equation

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

eşitliğinin en küçüklenmesi gerekir. Eş. 2'den parametre vektörü

From Eq. 2 parameters vector is obtained as (1)

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

olarak bulunur (1).

....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 (2). 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 (3,4).

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) ileri sürmüştür. Celmins (1987)'in eleştirisine karşılık Tanaka ve Ishibuchi (1991) etkileşimli bulanık katsayıları elde etmek için bir yöntem geliştirmişlerdir (3). 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 (3).

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 (3).

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

In the determining of test statistics and coefficients, the role of each observation must taken into consideration and the data must be tested detaily. Because the results of parameters estimations may even related to an observation and removes of this observation from data may change the result of the analyze. This kind of observations which has a bigger residual value than the others is called outlier (2). In the event that outlier value, robust methods are being used which is less affected than LSM method during the estimation of regression model.

Taking the fuzzy regression analysis and in case of data set has outlier as basis, the work accomplished can be summarized as follows:

Tanaka (1982) and the others have suggested first study in the linear regression analysis that has a fuzzy model. This approach generalized by Bardossy (1990). The approach of Tanaka (1987) is too complicated to solve the optimization problem. The approach of least squares is not clear and the measurement of the best suitability has not completed by residuals. Dimond (1988) has suggested the fuzzy least squares method, to eliminate that disadvantages of model Tanaka (3,4)

Model Tanaka (1987) produced frequently crisp numbers therefore, criticized by Celmins (1987). Later, Jos'zef (1992) has suggested that minimization problem has a scale independent which also suggested by Tanaka (1982) and the others. As a reaction to Celmin's critique, Tanaka and Ishibuchi had improved a method to get interaction fuzzy coefficients (3). Shown by Redden and Woddall (1996), this method had a strong sensitive to outlier, undervalued the knowledge that contains absolute data and produced unbounded solution (3).

Sakawa and Yano (1992) suggested interaction algorithm which is iterative to satisfy the decider and leaning to interaction with decider, to formulize multiobjective programming problems (4). But this algorithm is much sensitive against the outlier too. By the assist of the orthogonal least squares method, Redden and Woddall (1996) had improved the model of Sakawa and Yano and also suggested the linear programming problem (3).

Chang and Lee (1996), for a outlier condition, making weighted with degree of membership and lean on an interaction with the decider, have suggested generalized fuzzy weighted the least squares method (5).

For a simple regression, Yang and Ko (1997) suggested weighted fuzzy the least squares of analyzed iterative sürmüşlerdir (5).

Yang 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 (6)

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 (7).

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 (8).

Bu çalışmada çoklu regresyon çözümlemesinde girdi değişkenlerinin bulanık sayı ($X_i = (x_i, \underline{\xi_i}, \overline{\xi_i})$, $Y_i = (y_i, \underline{\eta_i}, \overline{\eta_i})$) ve veri kümesinde aykırı değer olması durumunda, üyelik fonksiyonu yardımıyla ağırlık matrisi tanımlanmıştır. Regresyon model tahmininde ise bulanık regresyon çözümlemesi kullanılmıştır.

Bu 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 ve 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 EKK, M, En Küçük Medyan Kare (LMS) ve LMS'ye Dayanan Yeniden Ağırlıklandırılmış EKK (RLS) yöntemleri ve önerilen bulanık robust yöntem ile regresyon model tahminleri elde edilecek ve sonuçlar karşılaştırılacaktır.

algorithm. This algorithm has two stages. First of all chosen the fuzzy classification method which gives class membership of the observations, then values of this membership used as weights. In the solution of fuzzy regression, weighted least squares method was thought as an optimization problem (6).

.

Yang and Lin (2002) had suggested the fuzzy least squares linear regression analysis for fuzzy input and fuzzy output. They had benefited from cluster analyzing to determine the heterogeneous data group and outlier (7).

Yang and Liu (2003) had suggested the fuzzy least squares algorithm for the models of fuzzy interaction linear regression. This algorithm is robust against the outlier for simple regression. In this algorithm, orthogonal conditions have added constraint to optimization problem (8).

In this study when input data are fuzzy number ($X_i = (x_i, \underline{\xi_i}, \overline{\xi_i})$, $Y_i = (y_i, \underline{\eta_i}, \overline{\eta_i})$) in multi regression analysis and data set has outlier, weighted matrix will be defined with respect to membership function. In regression model estimates are used fuzzy regression analysis.

Section 2 of this paper there will be definitions robust methods which are used usually in literature. Section 3 will be given the definitions fuzzy least squares method. Section 4 when independent variable(s) and response variable is fuzzy number 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. Finally regression model estimates are obtained with least squares method, robust methods which M, Least Median of Squares (LMS) and Reweighted Least Squares Based on The LMS (RLS) and suggested fuzzy robust method and the results will be compared.

2. ROBUST YÖNTEMLER

Bu kesimde literatürde sıklıkla kullanılan Robust yöntemlerden M, LMS ve RLS yöntemleri için tanımlar verilecektir.

2.1. M Yöntemi

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

2. ROBUST METHODS

In this section, there will be definitions robust methods which M, (LMS) and RLS are used usually in literature.

2.1. M Methods

M method is minimizing of residual function much than minimizing the sum of squared residuals. Regression coefficients are obtained by the minimizing the sum

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

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

By taking the first partial derivative of the sum in Eq. 4 with respect to each $\hat{\beta}_j$ and setting it a zero it may be found regression coefficient that p equations.

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

p denklem sistemi için regresyon katsayıları elde edilir. Huber'in ρ fonksiyonu

Huber's ρ function is defined as

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

$$z = r_i / d$$

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

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. Eş. 5'in türevi alınırsa

where k is called tuning constant and k is set at 1.5. Sometimes the numerator of d is called the median of the absolute deviations (MAD). By taking derivative of Eq. 5

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

fonksiyonu elde edilir. ψ fonksiyonu ρ 'nun türevidir. Eş. 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

it may be function. The function ψ is the derivative of ρ . They are typically set up such that large residuals will be given only marginal or zero ψ weights in Eq. 5. So ψ is often labeled as "redescending to zero".
Hampel ψ function is defined as

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

biçiminde tanımlanır. Genellikle sabitlerin değerleri a=1.7, b=3.4 ve c=8.5 olarak seçilir. Andrews (sinüs tahmini) ise ψ fonksiyonunu

where k is taken 1.5 or k=2.1.

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

olarak tanımlamıştır, burada k=1.5 ya da k=2.1 alınır.
Tukey'in iki ağırlıklı tahmini için ψ fonksiyonu ise

In Tukey's biweight estimate, ψ function is defined as

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

biçiminde tanımlanır. *k*, 5 ya da 6 olarak seçilir (9-12). ölçekleme bağıntısı aşağıda verilmektedir:

where *k* is selected as 5 or 6 (7- 10).

2.2. En Küçük Medyan Kare (LMS)

Hata karelerinin medyanını en küçük yapmayı amaçlayan bu yöntemde amaç fonksiyonu,

$$\min_{\hat{\theta}} \ \underset{i}{med}\ r_i^2$$

biçiminde tanımlanır. Burada r_i i'inci gözlem hatası, $\hat{\theta}$ tahmin biçiminde tanımlanır. En küçük medyan kare yönteminde i'inci gözlemin ağırlığı

2.2. Least Median of Squares (LMS)

Minimizing the median of the squared residual is purposed in this methods, the objective function is defined as

$$\min_{\hat{\theta}} \ \underset{i}{med}\ r_i^2$$

where r_i is *i*th observed error, $\hat{\theta}$ is defined as estimated. In least median of squares method, a weight for *i*th observation is defined as

$$\min_{\hat{\theta}} \sum_i w_i r_i^2$$

biçiminde hesaplanır. Burada n toplam gözlem sayısı, *p* değişken sayısı olmak üzere $s_0 = 1,4826\left(1 + \frac{5}{n-p}\sqrt{\underset{i}{med}\ r_i^2}\right)$ olarak tanımlanır. LMS yöntemi, Rousseuw ve Leroy (1987) tarafından verilen yeniden örnekleme algoritması ile hesaplanır (13-15).

where n is the sum number of observation, *p* is the number of variable. Then $s_0 = 1,4826\left(1 + \frac{5}{n-p}\sqrt{\underset{i}{med}\ r_i^2}\right)$. LMS method is calculated by Rousseuw and Leroy (1987) were taken with resampling algorithm (13-15).

2.3. LMS'ye Dayanan Yeniden Ağırlıklandırılmış EKK (RLS)

Ağırlıklandırılmış hata kare toplamını minimum yapmaya çalışan RLS yönteminde amaç fonksiyonu

2.3. Reweighted Least Squares Based on The LMS (RLS)

Minimizing the weighted median of the squared residual is purposed in RLS, the objective function is defined as

$$\min_{\hat{\theta}} \sum_i w_i r_i^2$$

biçiminde tanımlanır. RLS yönteminde ağırlıklar

In RLS method, weights are defined as

$$w_i = \begin{cases} 1 & |r_i / \hat{\sigma}| \le 2,5 \\ 0 & |r_i / \hat{\sigma}| > 2,5 \end{cases}$$

şeklinde hesaplanır. Burada

where $\hat{\sigma}$ is defined as

$$\hat{\sigma} = \sqrt{\frac{\sum w_i r_i^2}{\sum w_i - p}}$$

olarak tanımlanır (13-15).

(13-15).

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

$$Y=a+bX$$

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

3. FUZZY LEAST SQUARES METHOD

Triangular fuzzy numbers is defined as $X=(m,\underline{m},\overline{m})$, where *m* is modal value of *X*, $\underline{m}$ is left spreads and $\overline{m}$ is right spreads. For fuzzy least squares method, when $X_i=(x_i,\underline{\xi_i},\overline{\xi_i})$, $Y_i=(y_i,\underline{\eta_i},\overline{\eta_i})$ triangular fuzzy number is taken, following model will be considered

$$Y=a+bX$$

where $a,\, b$ are crisp numbers. When parameters are crisp, least squares optimization problem is defined as

$$Minimum \quad r(a,b)=\sum d(a+bX_i,Y_i)^2 \qquad [7]$$

olarak tanımlanır. Bu durumda

.In this case $d\left(a+bX_i,Y_i\right)$ is defined as

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

şeklinde tanımlanır (16).

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

(16).

When x is crisp and $Y=(y,\underline{\eta},\overline{\eta})$ is triangular fuzzy number, fuzzy regression model is defined as

$$Y=A+xB \qquad [8]$$

olarak tanımlanır. Burada $A=(a,\underline{\alpha},\overline{\alpha})$ ve $B=(b,\underline{\beta},\overline{\beta})$ bulanık parametrelerdir. Eş.8'deki parametre tahminlerinin yapabilmesi için en küçük kareler optimizasyon problemi

where $A=(a,\underline{\alpha},\overline{\alpha})$ and $B=(b,\underline{\beta},\overline{\beta})$ are fuzzy parameters. For parameters can be estimated in Eq.8, least squares optimization problem is

$$Minimum \quad r(A,B)=\sum d(A+x_iB,Y_i)^2$$

biçimindedir. Burada

where $d(A+x_iB,Y_i)$ is defined as

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

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 (16,17).

solving $\dfrac{\partial d}{\partial a}=0$ and $\dfrac{\partial d}{\partial b}=0$, a and b parameters and solving $\dfrac{\partial d}{\partial \alpha}=0$, $\dfrac{\partial d}{\partial \beta}=0$, α and β parameters are obtained (16,17).

4.GİRDİ DEĞİŞKENLERİNİN BULANIK OLMASI DURUMUNDA ROBUST REGRESYON ÇÖZÜMLEMESİ

Bulanık girdi değişkenleri ile ilgili yapılan çalışmalarda genellikle basit doğrusal regresyon modeli ele alınmıştır.

4. THE FUZZY ROBUST REGRESSION ANALYSIS, THE CASE OF INPUT VARIABLES IS FUZZY

In fuzzy input variables are taken the work accomplished, simple linear regression model has been taken in general. In this study, form of

Bu çalışmada $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_p X_p + \varepsilon$ biçimindeki çoklu regresyon modeli ele alınarak, Eş. 7 genelleştirilmiş ve bir lemma olarak verilmiştir.

$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_p X_p + \varepsilon$ multi linear regression model has been taken, Eq. 7 has been generalized and lemma will be given.

Lemma: Çoklu doğrusal regresyon modeli durumunda Eş.7 ile tanımlanan optimizasyon problemi

Lemma: The case of multi linear regression model, optimization problem which defined in Eq.7 has been generalized as

$$\text{P:} \qquad Min\ r(a, b_1, b_2, ..., b_p) = \sum d\left(a + b_1 X_{1i} + ... + b_p X_{pi}, Y_i\right)^2$$

olarak genelleştirildi. Burada $a, b_1, ..., b_p$ kesin sayılar, $X = (x, \underline{\xi}, \overline{\xi})$, ve $Y_i = (y_i, \underline{\eta}_i, \overline{\eta}_i)$ üçgensel bulanık sayılar olmak üzere

where $a, b_1, ..., b_p$ are crisp number, $X = (x, \underline{\xi}, \overline{\xi})$ and $Y_i = (y_i, \underline{\eta}_i, \overline{\eta}_i)$ are triangular fuzzy number. Then

$$\begin{aligned} d\left(a + b_1 X_{1i} + ... + b_p X_{pi}, Y_i\right) = & \left[a + b_1 x_{1i} + b_2 x_{2i} + ... + b_p x_{pi} - y_i - \left(b_1 \underline{\xi}_1 + b_2 \underline{\xi}_2 + ... + b_p \underline{\xi}_p - \underline{\eta}_i\right)\right]^2 \\ & + \left[a + b_1 x_{1i} + b_2 x_{2i} + ... + b_p x_{pi} - y_i + \left(b_1 \overline{\xi}_1 + b_2 \overline{\xi}_2 + ... + b_p \overline{\xi}_p - \overline{\eta}_i\right)\right]^2 \\ & + \left(a + b_1 x_{1i} + b_2 x_{2i} + ... + b_p x_{pi} - y_i\right)^2 \end{aligned}$$

dir. P probleminin en küçüklenmesi sonucunda

When P problem is minimized, $\hat{\beta}$ is obtained as

$$\hat{\beta} = (X'X + A'A + B'B)^{-1}(X'Y + A'C + B'D) \qquad [9]$$

olarak bulunur. Burada

where

$$X = \begin{bmatrix} 1 & x_{11} & . & . & x_{p1} \\ 1 & x_{12} & . & . & x_{p2} \\ . & & . & & . \\ . & & & . & . \\ 1 & x_{1n} & & & x_{pn} \end{bmatrix} \quad Y = \begin{bmatrix} y_1 \\ . \\ . \\ y_n \end{bmatrix} \quad C = \begin{bmatrix} y_1 - \underline{\eta}_1 \\ . \\ . \\ y_n - \underline{\eta}_n \end{bmatrix} \quad D = \begin{bmatrix} y_1 + \overline{\eta}_1 \\ . \\ . \\ y_n + \overline{\eta}_n \end{bmatrix}$$

$$B = \begin{bmatrix} 1 & \left(x_{11} + \overline{\xi}_{11}\right) & . & . & \left(x_{p1} + \overline{\xi}_{p1}\right) \\ 1 & \left(x_{12} + \overline{\xi}_{12}\right) & . & . & \left(x_{p2} + \overline{\xi}_{p2}\right) \\ . & & . & & . \\ . & & & . & . \\ 1 & \left(x_{1n} + \overline{\xi}_{1n}\right) & . & . & \left(x_{pn} + \overline{\xi}_{pn}\right) \end{bmatrix} \quad A = \begin{bmatrix} 1 & \left(x_{11} - \underline{\xi}_{11}\right) & . & . & \left(x_{p1} - \underline{\xi}_{p1}\right) \\ 1 & \left(x_{12} - \underline{\xi}_{12}\right) & . & . & \left(x_{p2} - \underline{\xi}_{p2}\right) \\ . & & . & & . \\ . & & & . & . \\ 1 & \left(x_{1n} - \underline{\xi}_{1n}\right) & & & \left(x_{pn} - \underline{\xi}_{pn}\right) \end{bmatrix}$$

olarak tanımlanır. Gözlem sayısı $i = 1, ..., n$ ve değişken sayısı $j = 1, ..., p$ 'dir. Parametre tahminlerinin elde edilebilmesi için $(X'X + A'A + B'B)$ 'nin tersinin alınabilmesi gereklidir. Klasik regresyonda X matrisi için gerekli olan koşulların sağlanması durumunda $(X'X + A'A + B'B)$ 'nin tersi alınabilir.

The observed number is $i = 1, ..., n$ and the variable number is $j = 1, ..., p$. To parameter estimates is obtained, $(X'X + A'A + B'B)$ must be inverse. In ordinary regression when X matrix is provided necessary condition, inverse of $(X'X + A'A + B'B)$ will be taken.

4.1. Önerilen Algoritma

Regresyon çözümlemesinde aykırı değer olması durumunda parametre tahminleri ile ilgili oldukça çok çalışma yapılmış ve Robust tahmin ediciler

4.1. Suggested Algorithm

In regression solution, many studies were done about parameter estimation in the event that outlier and were defined the Robust estimators. Studies were made as for tanımlanmıştır. Son yıllarda yapılan çalışmalarda ise, bulanık tahminler de oldukça etkindir. Ancak bu çalışmalarda genellikle basit doğrusal regresyon modeli ele alınmış ve tanımlanan iteratif süreçlerde model parametrelerinin kesin sayı, merkezlerin sıfır olması gibi bazı dezavantajlara rastlanmıştır. Bu sakıncaları gidermek ve süreci çoklu doğrusal regresyon için genelleştirmek bu çalışmanın amacını oluşturmuştur. Bu nedenle aşağıda önerilen algoritmada tanımlı $\hat{\beta}$'ların tahmininde ağırlık matrisi üyelik fonksiyonları yardımıyla tanımlanmış, üyelik matrisinin tanımlanmasında uzaklık kullanılmış ve üyelikler kullanılarak ağırlıklandırılmış bulanık en küçük kareler tahmin edicisi elde edilmiştir. Önerilen algoritmanın adımları aşağıda verilmiştir.

Adım 1: $X_i = (x_i, \underline{\xi_i}, \overline{\xi_i})$, $Y_i = (y_i, \underline{\eta_i}, \overline{\eta_i})$ üçgensel bulanık sayıları için parametre tahmini Eş. 9'dan elde edilir.

Adım 2: $\hat{y}_j$ tahmin değerleri 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

the last years, fuzzy estimations were quite active. But in this operations have generally taken up simple linear regression model and in defined iterative processes, were met disadvantages like crisp number and zero centered of model parameters. The aim of this studies that to remove those objections and to generalize the processes for the multi linear regression. For this reason, estimation of defined $\hat{\beta}$ which was suggested algorithm in below, was defined help of the functions of weights matrix membership, distance was used defining of matrix membership and was found estimator of the weighted fuzzy least squares by using memberships. Suggested algorithm steps are given in below.

Step 1: For $X_i = (x_i, \underline{\xi_i}, \overline{\xi_i})$, $Y_i = (y_i, \underline{\eta_i}, \overline{\eta_i})$ triangular fuzzy numbers, estimation of regression parameter is obtained from Eq.9.

Step 2: $\hat{y}_j$ are estimated and residuals (e_i) are determined.

Step 3: According to absolute residual value, median is determined and distances are calculated

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

where $\|.\|$ is euclide distance.

Step 4: According to distance, membership function has been defined

$$\mu(x) = \begin{cases} 1 & x \le a \\ \dfrac{b - |e|}{b - a} & a < x < b \\ 0 & d.d \end{cases} \qquad [10]$$

biçiminde tanımlanmıştır. Burada;

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

dır.

Adım 5: Eş.10'da tanımlanan üyelik fonksiyonundan üyelik dereceleri belirlenir ve ağırlık matrisi (W) oluşturulur. Ağırlık matrisi, köşegen elemanları üyeliklerden oluşan diagonal matristir. Eş. 9'dan ağırlıklandırılmış bulanık en küçük kareler tahmin edicisi

where

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

Step 5: From Eq. 10 has been defined membership function, membership values are determined and weighted matrix (*W*) is constituted. Weighted matrix is diagonal matrix which diagonal elements are consist of membership value. From Eq.9, weighted fuzzy least squares parameters coefficient is defined as

$$\hat{\beta} = (X'WX + A'WA + B'WB)^{-1}(X'WY + A'WC + B'WD) \qquad [11]$$

biçiminde tanımlanır. Bu tahmin edici kullanılarak parametre tahmini yapılır.

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.

As this parameters coefficient is used, regression parameters will be estimated.

Step 6: If $\left|\hat{\beta}^{k+1} - \hat{\beta}^{k}\right| < \varepsilon$ then stop. Otherwise is go to Step 2. Where $\hat{\beta}$ is estimates of regression model coefficients, k is iteration number and $\varepsilon > 0$ is a small number.

4.2. Uygulama

Önerilen algoritmayı irdeleyebilmek ve İkinci Bölümde verilen klasik yöntemlerle karşılaştırabilmek için Gujarati (2001)'den 31 gözlem ve üç bağımsız değişkeni içeren bir örnek ele alınmıştır (18). Aykırı gözlem durumu için 24'üncü gözlemin değeri 30.1 olarak değiştirilmiştir. Veri kümesi Çizelge 1'de verilmiştir. Bu veri kümesi için Bölüm 1'de tanımlanan EKK, Bölüm 2'de tanımlanan M, LMS, RLS ve önerilen yöntemden elde edilen mutlak artıklar Çizelge 2'de verilmiştir. Önerilen yöntem için bağımsız değişken değerleri; merkez (x_i), sol yayılma $\underline{\xi_i} = x_i / 7$, sağ yayılma $\overline{\xi_i} = x_i / 6$ ve bağımlı değişken değerleri; merkez (y_i), sol yayılma $\underline{\eta_i} = y_i / 8$ ve sağ yayılma $\overline{\eta_i} = y_i / 7$ alınarak çözümleme yapılmıştır. Sol ve sağ yayılmalar literatürde belirtildiği gibi belirlenmiştir. M yöntemi ve tanımlanan yöntem için Matlab paket programında program yazılmıştır. Elde edilen regresyon model tahmini Çizelge 3'de, mutlak hatalara ilişkin indeks plot ise Şekil 1, 2, 3, 4, 5, 6, 7 ve 8'de verilmiştir..

4.2. Numerical Example

To investigate suggested algorithm and compare classical method that given in second part, an example was taken which includes 31 observations and 3 independent variables from Gujarati (2001) (18). For a outlier observation, the value of the 24th observation was changed into 30.1. Data set was given in Table 1. For this data set, LSM that defined in Section 1, M, LMS, RLS that defined in Section 2 and absolute residuals was taken from the suggested method, was shown in Table 2. For the suggested method, was taken the values of independent variables; center (x_i), left spread $\underline{\xi_i} = x_i / 7$, right spread $\overline{\xi_i} = x_i / 6$ and dependent variables; center (y_i), left spread $\underline{\eta_i} = y_i / 8$, right spread $\overline{\eta_i} = y_i / 7$ analyze has done. Left and right spreads were determined as shown in literature. For the M method and suggested method was written a program in Matlab software. The gotten estimation of regression model in Table 3, as for the index plot belonging to absolute errors was given in Figures 1,2,3,4,5,6,7 and 8.

Table 1. Data set
Çizelge 1. Veri kümesi

Observed Number /Gözlem No	x_1	x_2	x_3	y	Observed Number Gözlem No	x_1	x_2	x_3	y
1	4,89	5,52	487,67	8,01	**17**	3,96	7,80	874,29	10,75
2	4,83	5,05	490,59	9,06	**18**	3,85	8,30	925,86	9,47
3	4,68	5,41	533,55	10,31	**19**	3,75	8,81	980,98	10,31
4	4,42	6,16	576,57	11,76	**20**	3,69	8,66	1007,72	8,88
5	4,36	6,26	598,62	12,43	**21**	3,56	8,78	1051,83	8,88
6	4,55	6,34	621,77	13,31	**22**	3,56	9,18	1078,76	9,70
7	4,66	6,81	613,67	13,1	**23**	3,48	9,03	1075,31	7,69
8	5,54	7,15	654,80	14,94	**24**	3,53	9,00	1107,48	**30,1**
9	4,44	7,17	668,84	16,17	**25**	3,39	8,78	1171,10	7,54
10	4,75	6,71	681,02	14,71	**26**	3,68	8,38	1234,97	7,47
11	4,56	7,05	679,53	13,2	**27**	5,92	8,01	1217,81	8,63
12	4,29	7,04	720,53	13,19	**28**	6,03	7,78	1202,36	9,21
13	4,19	7,18	736,86	11,7	**29**	6,12	7,88	1271,01	9,23
14	4,17	7,33	755,34	10,99	**30**	6,05	7,88	1332,67	9,96
15	4,11	7,54	799,15	10,80	**31**	5,89	8,67	1385,10	10,78
16	4,04	7,61	830,7	10,66					

***R, residual and W, weights**

Table 2. Absolute Residuals and Diagonal Elements of Weighted Matrix
Çizelge 2.Mutlak Hatalar ve Bunlara Karşılık Gelen Ağırlık Matrisinin Köşegen Elemanları

Observed Number /Gözlem No	LSM R /EKK Hata	Huber R /Hata	Huber W Ağırlık	Hampel R /Hata	Hampel W Ağırlık	Tukey R /Hata	Tukey W Ağırlık	Andrews R /Hata	Andrews W Ağırlık	LMS R /Hata	LMS W Ağırlık	RLS R /Hata	RLS W Ağırlık	Suggested R /Hata	Suggested W Ağırlık	Estimated /Tahmin
1	3.3093	4.3406	0.6045	4.2117	0.7108	4.6024	0.3753	4.3690	0.3559	6.3138	0	6.2188	0	4.5623	0.7980	12,5723
2	0.6494	1.7388	1	1.5388	1	2.0622	0.8505	1.7431	0.4557	5.4822	0	4.9001	0	2.5225	0.9020	11,5825
3	0.3012	0.4924	1	0.2991	1	0.7461	0.9797	0.4689	0.4747	3.6916	0	3.2250	0	1.1545	0.9717	11,4645
4	0.3926	0.0877	1	0.2386	1	0.0275	1	0.1147	0.4760	1.4229	1	1.3701	1	0.1089	1	11,8689
5	1.1499	0.9609	1	1.1171	1	0.8673	0.9727	1.0320	0.4689	0.5322	1	0.4696	1	0.7470	0.9924	11,6830
6	1.8176	1.5217	1	1.6519	1	1.4192	0.9277	1.5693	0.4595	0.4359	1	0.3763	1	1.4418	0.9570	11,8682
7	0.1509	0.4417	1	0.3920	1	0.4883	0.9913	0.4279	0.4749	0.3874	1	0.2058	1	0.0952	1	13,0048
8	1.3814	1.3519	1	1.3926	1	1.3687	0.9326	1.3888	0.4631	2.7303	0	2.0163	0	2.0169	0.9277	12,9231
9	2.8977	2.9856	0.8789	3.0439	0.9835	3.0166	0.6948	3.0424	0.4153	4.1077	0	3.4790	0	3.5486	0.8497	12,6214
10	2.5874	2.3105	1	2.3918	1	2.2407	0.8248	2.3371	0.4397	2.3016	1	1.9507	1	2.5494	0.9006	12,1606
11	0.2744	0.2459	1	0.3082	1	0.2464	0.9978	0.2914	0.4756	1.0728	1	0.5152	1	0.7238	0.9936	12,4762
12	1.2824	1.5603	1	1.6801	1	1.5901	0.9097	1.6625	0.4575	1.4429	1	1.1853	1	1.6819	0.9448	11,5081
13	0.2664	0.1626	1	0.2853	1	0.2240	0.9982	0.2819	0.4756	0.1822	1	0.0958	1	0.3034	1	11,3966
14	1.1572	0.6414	1	0.5270	1	0.5556	0.9887	0.5168	0.4744	0.3207	1	0.6714	1	0.4220	1	11,4120
15	1.2989	0.6020	1	0.4867	1	0.4783	0.9917	0.4588	0.4748	0.0961	1	0.4744	1	0.3628	1	11,1628
16	1.1083	0.2774	1	0.1471	1	0.1351	0.9993	0.1143	0.4761	0.0300	1	0.2752	1	0.1307	1	10,7907
17	0.8800	0.1443	1	0.2807	1	0.3238	0.9962	0.3296	0.4754	0.5322	1	0.2298	1	0.2758	1	10,4742
18	2.8386	1.5042	1	1.3943	1	1.2389	0.9446	1.2983	0.4647	0.0961	1	0.6442	1	1.1037	0.9743	10,5737
19	2.6737	1.0294	1	0.9471	1	0.6779	0.9833	0.8034	0.4718	1.4176	1	0.6191	1	0.3553	1	10,6653
20	3.1642	1.4573	1	1.3348	1	1.1205	0.9546	1.2083	0.4663	0.0947	1	0.4450	1	1.0887	0.9750	9,9687
21	2.7229	0.7950	1	0.6500	1	0.4256	0.9934	0.5138	0.4744	0.4958	1	0.0571	1	0.5547	1	9,4347
22	2.7809	0.7073	1	0.6023	1	0.2793	0.9971	0.4283	0.4749	1.7056	1	0.9728	1	0.1355	1	9,8355
23	4.2469	2.1460	1	2.0119	1	1.7310	0.8934	1.8514	0.4531	0.3695	1	0.9225	1	1.7802	0.9398	9,4702
24	18.6356	20.7413	0.1265	20.8872	0	21.1471	0	21.0404	0	22.1923	0	21.7340	0	20.9890	0.0005	9,1110
25	2.1237	0.1621	1	0.3853	1	0.5526	0.9889	0.5118	0.4744	0.0366	1	0.0093	1	0.1669	1	7,7069
26	0.4735	1.5703	1	1.8351	1	1.8719	0.8759	1.9094	0.4516	0.0961	1	0.3215	1	0.7735	0.9911	6,6965
27	1.7785	1.8254	1	1.8295	1	1.8254	0.8818	1.8286	0.4536	0.3959	1	1.0632	1	1.3721	0.9606	10,0021
28	0.8595	1.0868	1	1.0831	1	1.1326	0.9536	1.1057	0.4679	0.0961	1	0.6619	1	0.7463	0.9925	9,9563
29	0.3218	0.4836	1	0.4739	1	0.5235	0.9900	0.4951	0.4745	0.3494	1	0.1579	1	0.2296	1	9,4596
30	1.3837	1.3912	1	1.4388	1	1.3602	0.9335	1.4123	0.4627	1.4877	1	1.2103	1	1.3635	0.9610	8,5965
31	0.7009	1.1355	1	1.1259	1	1.2379	0.9447	1.1767	0.4668	3.1548	0	2.3928	0	1.6520	0.9463	9,1280

*R, residual and W, weights

Çizelge 2 incelendiğinde, aykırı değer olarak alınan 24. gözlemin ağırlığı Huber'de "0.1265", Hampel, Tukey, Andrews, LMS ve RLS'de "0" önerilen yöntemde ise "0.0005" olarak bulunmuştur. Önerilen yöntem sonucunda bulunan ağırlıklar aslında her bir gözlemin üyelik dereceleridir. Bu üyelik dereceleri gözlemlerin modele etkisini göstermektedir. Dolayısıyla Çizelge 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.

When examined in Table 2, the weight of 24*th* observation, which was taken as outlier, in Huber "0.1265", in Hampel, Tukey, Andrews, LMS and RLS "0" and in the suggested method was found as "0.0005". Weights that were found in the end of the suggested method, actually were degrees of membership of each observations. These memberships show effects of observations to the model. Also as seen in Table 3, outlier effective the model by very small membership degree, the degrees of membership of the other observations values are 1 or near 1 and the effects of those to the estimate of regression model take importance.

Table 3. Estimated of Regression Coefficients
Çizelge 3.Regresyon model tahminleri

Method /Yöntem	**Constant /Sabit**	**Regression Coefficient /Regrasyon Katsayıları** β_1	β_2	β_3	**Sum of Square Residual /Hata Kare Toplamı**
LSM	-6.5535	1.5118	3.1446	-0.0141	456.9177
Huber	-7.6080	2.3910	2.8980	-0.0159	496.8698
Hampel	-8.3784	2.5659	2.9423	-0.0163	502.2505
Tukey	-7.0755	2.5015	2.7527	-0.0159	514.0758
Andrews	-8.3271	2.5460	2.9273	-0.0162	508.7170
LMS	17.9016	0.5117	-0.5676	-0.0060	626.9510
RLS	11.0860	1.1279	0.3715	-0.0091	583.3430
Suggested /Önerilen	0.2126	1.6831	1.8160	-0.0121	512.4447

Çizelge 3'de EKK, Huber, Hampel, Tukey, Andrews, LMS, RLS ve önerilen yöntemle elde edilen regresyon model tahminleri verilmiştir. Önerilen yöntemde bulanık girdi verileri ele alındığı için girdi verileri bir aralıkta ele alınmıştır. Önerilen yöntemden elde edilen parametre tahmin değerleri diğer Robust yöntemlerle işaretçe aynı ve büyüklükçe çok yakındır.

In Table 3, LSM, Huber, Tukey, Andrews, LMS, RLS and the estimations of regression models that gotten by suggested methods was given. In suggested method fuzzy input data handled therefore input data have handled in one interval. Estimation values by suggested method are close to other Robust methods as size and equal as mark.

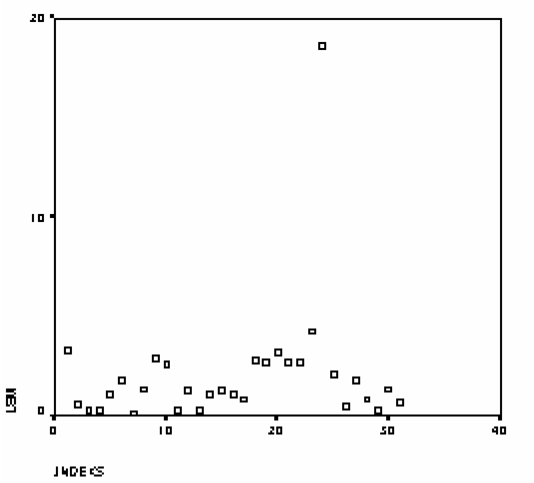

Figure 1. LSM Absolute Residual Index Plot
Şekil 1. EKK Mutlak Hta İçin Indeks Plot

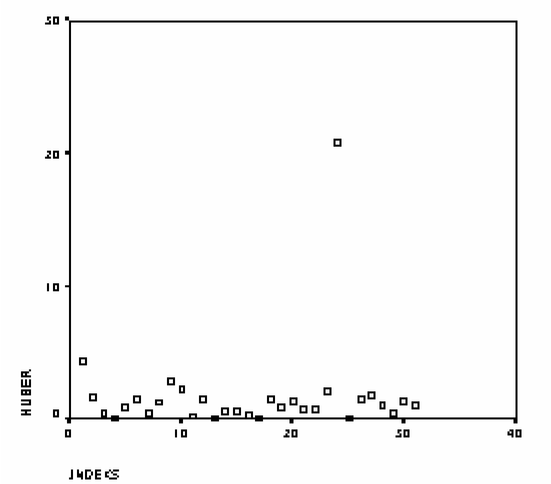

Figure 2. HUBBER absolute Residual Index Plot
Şekil 2.HUBER Mutlak Hata İçin Indeks Plot

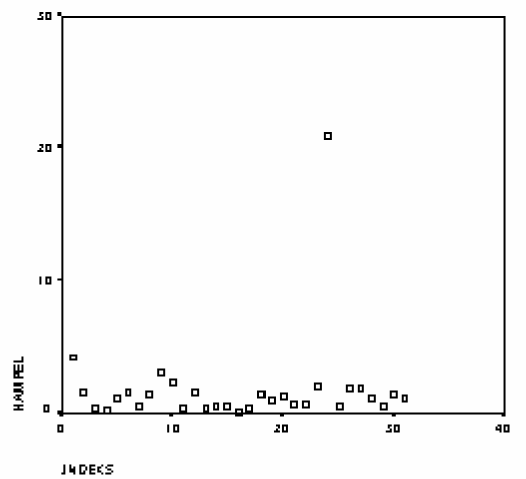

Figure3. HAMPEL Absolute Residual Index Plot
Şekil 3. HAMPEL Mutlak Hata İçin Indeks Plot

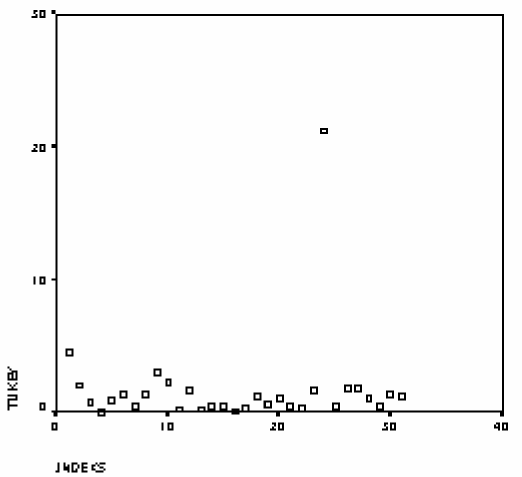

Figure 4. TUKEY Absolute Residual Index Plot
Şekil 4. TUKEY Mutlak Hata İçin Indeks Plot

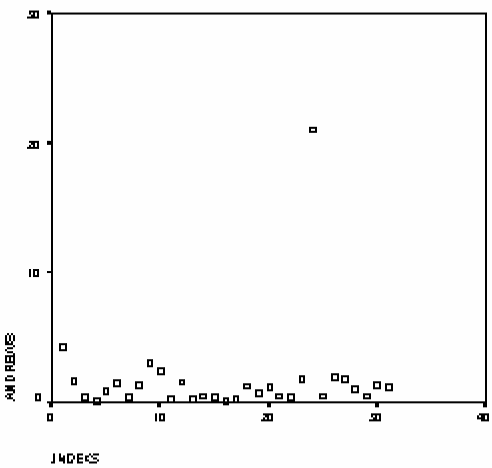

Figure 5. ANDREWS Absolute Residual Index
Şekil 5. ANDREWS Mutlak HATA İçin Indeks Plot

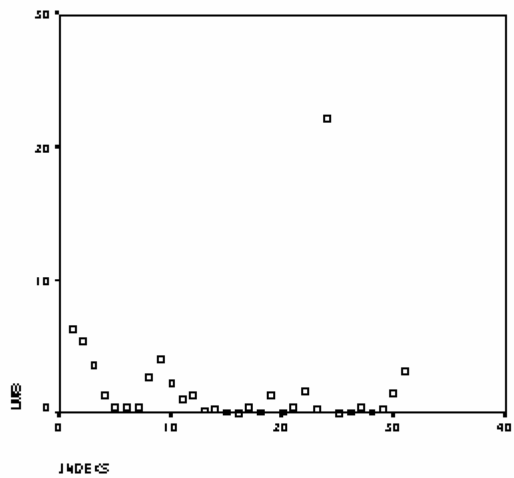

Figure 6. LMS Absolute Residual Index Plot
Şekil 6. LMS Mutlak Hata İçin Indeks Plot

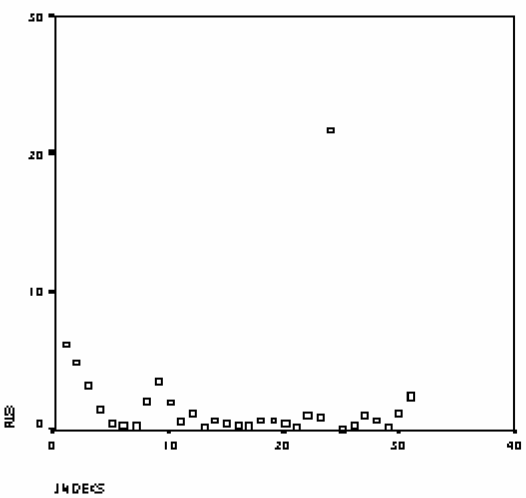

Figure 7.RLS Absolute Residual Index Plot
Şekil 7. RLS Mutlak Hata İçin Indeks Plot

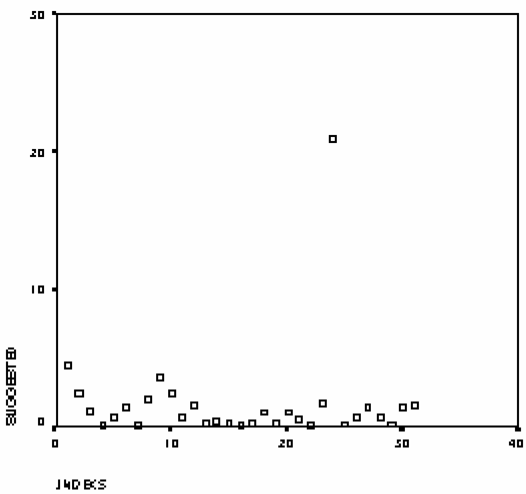

Figure 8 Suggested Absolute Residual Index Plot
Şekil 8. Önerilen Mutlak Hata İçin Indeks Plot

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

5. SONUÇ ve TARTIŞMA

Bu çalışmada $X_i = (x_i, \underline{\xi_i}, \overline{\xi_i})$, $Y_i = (y_i, \underline{\eta_i}, \overline{\eta_i})$ üçgensel bulanık sayılarının ele alınması ve veri kümesinde aykırı değer olması durumda regresyon model tahmini için iteratif bir yaklaşım ileri sürülmüştür. Ağırlık matrisinin elde edilmesinde uzaklıklara bağlı olarak tanımlanan üyelik fonksiyonundan yararlanılmıştır. Her bir gözlem üyelik derecesine göre parametre tahminine katılmıştır. Elde edilen ağırlıklar ile ağırlıklandırılmış bulanık en küçük kareler analizi yapılmıştır. Parametre tahmininde aykırı değerden EKK yöntemine göre daha az etkilenen model tahmini elde edilmiştir. Çizelge 2'de verilen mutlak hatalar incelendiğinde önerilen yöntemle elde edilen hataların aykırı değer için (20.9890) gibi büyük değerler, diğer gözlemlerde ise (0.3553,0.1669) gibi küçük değerler aldığı ve tahmin değerlerinin gözlemlenen değerlere yakın olduğu görülmektedir. Aykırı değer için üyelik fonksiyonundan elde edilen üyelik (ağırlık) "0.0005" iken diğer gözlemler için "1 ya da 1'e yakın" değerler almaktadır. Üyelikler yardımıyla aykırı değerler belirlendiği gibi diğergözlemlerin modele etki dereceleri de ayrı ayrı incelenebilmektedir.

Çizelge 3 incelendiğinde önerilen yöntemden elde edilen parametre tahminleri literatürde yer alan klasik yöntemlere yakın elde edildiği görülmektedir ve hata kare toplamı Huber ve Hampel'e çok yakın diğer yöntemlerden ise daha küçüktür. Saçılım grafiğinden, robust yöntemlerde olduğu gibi önerilen yöntemden de aykırı değerler kolaylıkla belirlenebilmektedir. Şekil 8'den görüldüğü gibi 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.

In absolute error graphics that drawn for Robust and suggested method, error that related to outlier gained big value, errors that related to the other observations gained close to zero.

5. CONCLUSION and DISCUSSION

In this study was suggested an iterative approach for estimation of regression model in the event that $X_i = (x_i, \underline{\xi_i}, \overline{\xi_i})$, $Y_i = (y_i, \underline{\eta_i}, \overline{\eta_i})$ handling of triangular fuzzy numbers and outlier in data set. To gain weight matrix connected to distance, was benefited from identified membership function. Each of observations according to its degree of membership, joined the parametric estimation. With the weights that have gained, weighted the fuzzy least squares have analyzed. In parameter estimation was taken model estimation that effects less than SM method outlier. When was examined absolute errors in Table 2, the errors were gained by suggested estimations for the outlier like (20.9890) bigger values, as for in the other observations like (0.3553, 0.1669) smaller values and the estimation values are closer to the values of observed values can be seen. While the outlier gained from membership function is "0.0005", for the other observations, it becomes "1 or closer to1". By the help of the membership outlier can be determined, by the way, degrees of the effection to the other observations can be analyzed.

When Table 3 has analyzed, can be seen that the estimation of parameter is closer to the classical methods that placed in literature and the sum of error squares is closer to Huber and Hampel, smaller than the other methods. From scatter graphic, like in the robustmethods, from suggested method outliers are can be easily determined. As seen from the Figure 8, when the outlier has a big residual, other observations has a residual that closer to zero. In this way it can be say that the suitability of gained regression model, except the outlier, better for remain observations. In the multi regression analyzing, in the event that observations are fuzzy and outlier in data group, the method can be used which suggested in regression model estimation.

KAYNAKLAR/ REFERENCES

1. Apaydın, A., Kutsal, and A., Atakan, C., *Uygulamalı İstatistik, 3.Baskı*, ***Klavuz Yayınevi***, Ankara, 343-357 (2002).

2. Alpar R., "Uygulamalı Çok Değişkenli İstatistiksel Yöntemlere Giriş I", ***Bağırgan Yayınevi,*** Ankara, 198-238 (1997).

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

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

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

6. Yang, M.S., and Ko, C.H., "On Cluster-Wise Fuzzy Regression Analysis", ***IEE Transaction on Systems, Man and Cybernetics Part B: Cybernetics***, Vol. 27, No:1: 1-13 (1997).

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

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

9. Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J., ve Stahel W.A., "Robust Statistics", ***John-Willey & Sons***, New-York, 307-338 (1986).

10. Hogg, R.V., "Statistican Robustness: One View of Its Use in Applications Today", ***The American Statistican***, Vol.33, No.3: 108-115 (1979).

11. Huber, P.J., Robust Statistics, ***John Willey & Son,*** USA, 153-195 (1981).

12. Huynh, H., "A Comparision of For Approaches to Robust Regression", ***Psychological Bulletin,*** Vol. 92, No.2: 505-512 (1982).

13. Candan, M., "Doğrusal Regresyon Çözümlemesinde Sağlam Kestiriciler", BilimUzmanlığı Tezi, ***Hacettepe Üniversitesi Fen Bilimleri Enstitüsü,*** Ankara, 37-39 (1995).

14. Rousseeuw, P.J., and Leroy, A.M., *Robust Regression and Outlier Detection,* USA, 21-154 (1987).

15. Şanlı, K., and Apaydın, A., "Bulanık Robust Regresyon", ***3. İstatistik Kongresi, Antalya***, 112-116 (2003).

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

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

18. Gujarati, D.N, *Temel Ekonometri*, Çeviri: Ümit Şenesen, Gülay Günlük Şenesen, ***Literatür Yayıncılık,*** İstanbul, 226 (2001).

Received/ Geliş Tarihi:26.01.2004 Accepted/Kabul Tarihi: 04.05.2004