

# FUZZY ROBUST REGRESYON'UN DİĞER REGRESYON TEKNİKLERİYLE KARŞILAŞTIRILMASI VE BİR UYGULAMA

*İsmail Hakkı ARMUTLULU<sup>1</sup>, Murat YAZICI*

<sup>1</sup>Marmara Üniversitesi, İ.İ.B.F., Sayısal Yöntemler Anabilim Dalı, Profesör Dr.

## FUZZY ROBUST REGRESYON'UN DİĞER REGRESYON TEKNİKLERİYLE KARŞILAŞTIRILMASI VE BİR UYGULAMA

**Abstract:** In regression analysis, data analysis is very important. Because, even one observation may be large effect over parameters estimates in regression model. Outliers and/or extreme values are removed from the data is recommended at the time of this regression equation can change completely. Robust methods are given better results than classical regression if data set includes extreme and/or outlier values. In this study, based on the classical method of EKK Regression Analysis, Robust Regression Analysis, Fuzzy Regression Analysis and Fuzzy Robust Regression Analysis were examined. After, with technical data of 45 automobile brands, each analysis of above mentioned were applied and results were compared. Based on the results we have obtained, the basic assumptions are not met Robust Regression Assessment of the classical regression Analysis are given beter results. Another result is, for we've got a set of data, Fuzzy Regression Analysis gave beter result than Fuzzy Robust Regression Analysis. In case of the data set have outliers or extreme values, we think that might be the opposite situation. We determined that techniques which are examined not have an advantage over one another and according to the structure of data set, each technique can be beter than the other.

**Keywords:** Fuzzy Robust Regression, Fuzzy Regression, Robust Regression, Multiple Linear Regression, Automobile, Automobile Rating Function

## COMPARISON OF FUZZY ROBUST REGRESSION WITH OTHER REGRESSION TECHNIQUES AND AN APPLICATION

**Özet:** Regresyon analizinde veri setini analiz etmek oldukça önemlidir. Çünkü tek bir gözlem bile regresyon modelindeki parametre kestirimleri üzerinde büyük bir etkiye sahip olabilir. Aykırı ve/veya uç değerlerin veri setinden çıkartılması zamanında önerilse de bu regresyon denklemini tamamen değiştirebilir. Robust yöntemlerin aykırı ve/veya uç değer olması durumunda klasik regresyona göre daha iyi sonuçlar verdiği ileri sürülmektedir. Bu çalışmada, EKK yöntemine dayanan Klasik Regresyon Analizi, Robust Regresyon Analizi, Fuzzy Regresyon Analizi ve Fuzzy Robust Regresyon Analizi incelenmiştir. Sonrasında 45 otomobil markasının teknik verileriyle bu analizler uygulanmış ve sonuçlar karşılaştırılmıştır. Elde ettiğimiz sonuca göre, temel varsayımların sağlanmadığı durumda Robust Regresyon Analizi'nin Klasik Regresyon Analizine göre daha iyi sonuç verdiği görülmektedir. Diğer bir sonuç ise, elimizdeki veri seti için Fuzzy Regresyon Analizi'nin Fuzzy Robust Regresyona Analizi'ne göre daha iyi sonuç verdiği'dir. Veri setinde aykırı ve/veya uç değer veya değerler bulunması durumunda tersi durumun söz konusu olabileceği düşünülmektedir. İncelenen tekniklerin birbiri üzerine bir üstünlüğe sahip olmadığı, veri setinin yapısına göre her bir tekniğin birbiri üzerine üstün olabileceği saptanmıştır.

**Anahtar Kelimeler:** Fuzzy Robust Regresyon, Fuzzy Regresyon, Robust Regresyon, Çoklu Doğrusal Regresyon, Otomobil, Otomobil Beğeni Fonksiyonu

## I. GİRİŞ

Regresyon Analizi en basit anlamda bir veya birden fazla bağımsız değişkenle bir bağımlı değişkeni açıklama esasına dayanmaktadır. Regresyon terimi ilk olarak 19. yüzyılda İngiliz istatistikçi Francis Galton tarafından bir biyolojik inceleme sırasında ortaya atılmıştır. Galton, baba ve annenin boyu ile çocukların boyu arasındaki bağlantıyı araştırmak istemiştir. Sonrasında Udny Yule, Karl Pearson yöntemi çeşitli istatistiksel alanlarda uygulamış ve geliştirmişlerdir. [1]

Veri setinde aykırı ya da uç değer olması En küçük Kareler (EKK) Yöntemi ile kestirilen regresyon parametrelerinin etkinliğini düşürmektedir. Gerçek hayat verilerinde özellikle sosyal verilerde aykırı ya da uç değerlerle karşılaşıldığından regresyon parametrelerini daha iyi kestirebilecek yeni tekniklere ihtiyaç duyulmuş

ve robust kestirim yöntemleri geliştirilmiştir. İlk kez robust kelimesi 1963 yılında Box tarafından kullanılmış ve EKK kestiricisindeki özellikle normallik varsayımına duyarısız yaklaşımlar olarak tanımlanmıştır.

1960'lı yıllarda Zadeh'in geliştirdiği Bulanık (Fuzzy) Küme Teorisi yaklaşımı, nitel değişkenleri sayısal olarak değerlendirme imkanı sunmuştur. Daha sonra 1980'li yıllarda Tanaka, Bulanık Küme Teorisi'nden yararlanarak Bulanık Doğrusal Regresyon Analizi Yöntemi'ni geliştirmiştir. Tanaka'nın geliştirdiği Bulanık Doğrusal Regresyon Analizi yayınlımların toplamını minimum yapma esasına dayanmaktadır. Sonrasında Sakawa ve Yano, Tanka'nın modelini esneterek, gözlem değerlerinin tamamının belirli aralığa düşmesi yerine, sadece bir kısmının (en az  $h$  kadarının) düşmesini yeterli görmüşlerdir. Sonrasında ise Georg Peters, gerçek verilerle Bulanık Doğrusal Regresyon

Analizi üzerinde çalışırken yeni bir model geliştirmiştir.[2]

Fuzzy Regresyon Analizi'nde veri setinde aykırı ya da uç değerler bulunmasında yayılım değerleri büyümektedir. Her ne kadar bağımlı değişken yüksek yayılımla iyi bir şekilde açıklanabilse de yayılımın düşük olması tercih edilmektedir. Yayılımın yüksek olduğu ve düzensiz verilerin veri setinde bulunması durumunda robust yöntemlere ihtiyaç duyulmuştur. Günümüzde Fuzzy Robust Regresyon Analizi üzerinde çalışmalar devam etmekte, veri setini iyi ifade edebilecek çeşitli robust yöntemleri içeren algoritmalar geliştirilmektedir.

Bu çalışma kapsamında iki tip fuzzy sayısı gösterimi kullanılmıştır. Bu gösterimlerden ilki,  $A = (a_1, a_2, a_3)$ 'dir. Burada  $a_2$  merkez değerini,  $a_1$  sol değeri,  $a_3$  ise sağ değeri ifade etmek üzere,  $A = (a_1, a_2, a_3)$  üçgen fuzzy sayısıdır. Gösterimlerden ikincisi,  $A = (m, s)$ 'dir. Burada  $m$  merkez değerini,  $s$  ise yayılım değerini ifade etmek üzere,  $A = (m, s)$  simetrik üçgen fuzzy sayısıdır.

Çoklu Doğrusal Regresyon Analizi'nde  $k$  adet bağımsız değişkenle  $Y$  bağımlı değişkeni arasındaki doğrusal ilişkinin gösterimi,

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \varepsilon \quad (1)$$

şeklindedir.

Regresyon denklemi, bir  $i$ 'inci gözlem için,

$$Y_i = \beta_0 + \beta_1 X_{i1} + \dots + \beta_k X_{ik} + \varepsilon_i \quad (2)$$

şeklindedir. Bu eşitlik örneklemeden elde edilen istatistiklerle,

$$\hat{Y}_i = b_0 + b_1 X_{i1} + \dots + b_k X_{ik} + e_i \quad (3)$$

şeklini alacaktır.

(3) no'lu ifadedeki regresyon denklemin iyi bir kestirimi için şu varsayımlar sağlanmalıdır.[3]

1.  $X$  matrisi  $(k+1)$  ranklı olmalıdır. Yani denklemler doğrusal bağımsız olmalıdır.

$$2. E(\varepsilon_i) = 0, \quad i = 1, 2, \dots, n \text{ için.} \quad (4a)$$

$$3. V(\varepsilon_i) = \sigma^2, \quad i = 1, 2, \dots, n \text{ için.} \quad (4b)$$

$$4. Cov(\varepsilon_i, \varepsilon_j) = 0 \quad (4c)$$

$X_j$ 'lerin raslantı değişkeni oldukları durumlarda,

$$5. Cov(x_i, \varepsilon_i) = 0, \quad i = 1, 2, \dots, n \text{ için.} \quad (4d)$$

İstatistiksel çıkarımlarda,

$$6. \varepsilon \sim N(0, \sigma^2) \quad (4e)$$

Parametre kestirimleri için En Küçük Kareler (EKK) yöntemi kullanılmak istenildiğinde amaç fonksiyonu:

$$\min \sum e_i^2 = \min \sum (y_i - \hat{y}_i)^2 \quad (5)$$

şeklindedir.  $\underline{\beta}$ 'nin EKK ile kestirimi şu şekilde olacaktır (altı çizili ifadeler ifadenin vektör olduğu anlamındadır);

$$\begin{aligned} \sum e_i^2 = \underline{\hat{e}}' \underline{\hat{e}} &= (\underline{Y} - \underline{\hat{Y}})' (\underline{Y} - \underline{\hat{Y}}) = (\underline{Y} - \underline{X} \underline{\hat{\beta}})' (\underline{Y} - \underline{X} \underline{\hat{\beta}}) \\ &= \underline{Y}' \underline{Y} - \underline{Y}' \underline{X} \underline{\hat{\beta}} - \underline{\hat{\beta}}' \underline{X}' \underline{Y} + \underline{\hat{\beta}}' \underline{X}' \underline{X} \underline{\hat{\beta}} \quad (6a) \end{aligned}$$

$$\underline{Y}' \underline{X} \underline{\hat{\beta}} = \underline{\hat{\beta}}' \underline{X}' \underline{Y} \text{ olduğundan,}$$

$$= \underline{Y}' \underline{Y} - 2 \underline{\hat{\beta}}' \underline{X}' \underline{Y} + \underline{\hat{\beta}}' \underline{X}' \underline{X} \underline{\hat{\beta}} \quad (6b)$$

eşitliği  $\underline{\hat{\beta}}$ 'ya göre türevi alınıp sıfıra eşitlenirse,

$$\frac{\partial (\underline{Y}' \underline{Y} - 2 \underline{\hat{\beta}}' \underline{X}' \underline{Y} + \underline{\hat{\beta}}' \underline{X}' \underline{X} \underline{\hat{\beta}})}{\partial \underline{\hat{\beta}}} = -2 \underline{X}' \underline{Y} + 2 \underline{X}' \underline{X} \underline{\hat{\beta}} = 0$$

(7)

olacaktır. Buradan,

$$\underline{X}' \underline{X} \underline{\hat{\beta}} = \underline{X}' \underline{Y} \quad (8a)$$

$$\underline{\hat{\beta}} = (\underline{X}' \underline{X})^{-1} \underline{X}' \underline{Y} \quad (8b)$$

$$\underline{\hat{\beta}} = \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_k \end{bmatrix}$$

$$\underline{X}' \underline{X} = \begin{bmatrix} n & \sum X_1 & \sum X_2 & \dots & \sum X_k \\ \sum X_1 & \sum X_1^2 & \sum X_1 X_2 & \dots & \sum X_1 X_k \\ & \sum X_2 & \sum X_2^2 & \dots & \sum X_2 X_k \\ & & & \ddots & \\ & & & & \sum X_k^2 \end{bmatrix}$$

(Simetrik)

$$\underline{X}' \underline{Y} = \begin{bmatrix} \sum Y \\ \sum X_1 Y \\ \sum X_2 Y \\ \vdots \\ \sum X_k Y \end{bmatrix}$$

Çoklu doğrusal regresyonda, parametrelerin ve regresyon modelinin anlamlılığını sınamak için çoğu kez iki tür test kullanılır.

1. Tümel-F testi (varyans çözümlemesi), Kısmi F testi

2. T testi

Tümel-F testinde, açıklayıcı değişkenlerle açıklanan değişken arasında doğrusal bağıntı şeklinde bir ilişkinin olup olmadığını sınamır.[4]

Sınama,

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_k = 0 \quad (9a)$$

$$H_1 : \beta_j \neq 0, \text{ en az bir } j \text{ için.} \quad (9b)$$

şeklindedir.  $H_0$  hipotezinin kabul edilmesi durumunda, açıklayıcı değişkenlerin hepsinin birden, açıklanan değişken üzerinde anlamlı bir etkiye sahip olmadıkları sonucuna varılır. Bu durumda model bir bütün olarak istatistiksel açıdan anlamsızdır. Yine aynı hipotezin reddedilmesi durumunda ise, en az bir açıklayıcı değişkenin, açıklanan değişken üzerinde anlamlı bir etkiye sahip olduğu sonucuna varılır. Kısmi F testinde  $F$  istatistiği şu şekilde hesaplanır:

$$F = \frac{RKT / k}{AKT / (n - k - 1)} \quad (10a)$$

Burada Regresyon Kareler Toplamını temsil eden  $RKT$ ,

$$RKT = \hat{\beta}' X' Y - n\bar{y}^2 \quad (10b)$$

şeklinde hesaplanır. Artık Kareler Toplamını temsil eden  $AKT$  ise,

$$AKT = Y' Y - \hat{\beta}' X' Y \quad (10c)$$

şeklinde hesaplanır.

Elde edilen  $F$  istatistiği,  $F$  tablosundan elde edilen  $F_{\alpha, k, n-k-1}$  değeri ile karşılaştırılır. Elde edilen  $F$  istatistiğinin,  $F$  tablo değerinden büyük olması durumunda  $\% 100(1 - \alpha)$  olasılığı ile  $H_0$  hipotezi reddedilir. [4]

$H_0$  hipotezinin reddedildiği durumda, hangi açıklayıcı değişkenin açıklanan değişken üzerinde istatistiksel açıdan anlamlı bir etkiye sahip olduğu araştırılmak istenildiğinde  $t$  testine başvurulur.

Bir  $\beta_j$  katsayısı için sınama,

$$H_0 : \beta_j = 0 \quad (11a)$$

$$H_1 : \beta_j \neq 0 \quad (11b)$$

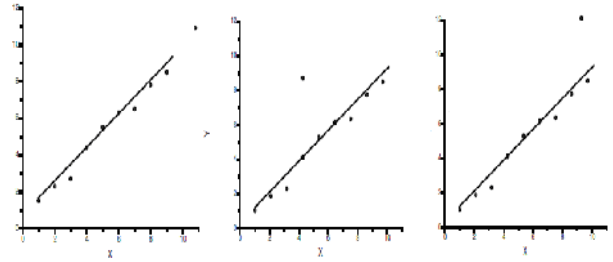
şeklindedir.  $H_0$  hipotezinin kabul edilmesi durumunda,  $j$ . açıklayıcı değişkenin, açıklanan değişken üzerinde anlamlı bir etkiye sahip olmadığı sonucuna varılır. Yine aynı hipotezin reddedilmesi durumunda ise,  $j$ . açıklayıcı değişkenin, açıklanan değişken üzerinde anlamlı bir etkiye sahip olduğu sonucuna varılır.

$t$  istatistiği şu şekilde hesaplanır:

$$t = \frac{b_j}{S_{b_j}} \quad (12)$$

Burada  $b_j$ , kestirilen  $j$ . parametreyi,  $S_{b_j}$  ise, bu parametrenin standart hatasını temsil etmektedir. Elde edilen  $t$  istatistiği,  $t$  tablosundan elde edilen  $t_{\alpha, n-k-1}$  değeri ile karşılaştırılır. Elde edilen  $t$  istatistiğinin,  $t$  tablo değerinden büyük çıkması durumunda  $\% 100(1 - \alpha)$  olasılığı ile  $H_0$  hipotezi reddedilir.

Aşağıdaki üç şekilde görüldüğü üzere veri setinde aykırı ya da uç değer bulunması durumunda EKK Kestiricisi iyi sonuç vermemekte, veri setini iyi ifade edecek kestirimlerde bulunamamaktadır. Bu gibi durumlarda yeni kestirim yöntemlerinde ihtiyaç duyulmuş ve veri setini EKK'dan daha iyi ifade edebilecek robust yöntemler geliştirilmiştir.



Şekil 1. Regresyon Doğrusu ile Aykırı ve/veya Uç değer

## II. ROBUST REGRESYON

EKK kestirim yönteminin uygulanabilmesi için gerekli olan varsayımların karşılanmaması durumunda yeni kestirim yöntemlerine ihtiyaç duyulmuştur. Veri setinde aykırı ve/veya uç değerlerin bulunması EKK kestirim yönteminin etkinliğini düşürmektedir. Diğer bir ifade ile, EKK kestirim yöntemi, veri setini açıklamaya çalışan regresyon denklemi için iyi bir kestirici olmamaktadır. En Küçük Kareler (EKK) Kestirim Yöntemi aykırı ve/veya uç değerlere sahip veri setine uygulandığında elde edilen tahmini  $\gamma$  değeri ( $\hat{Y}$ ) normal dağılıma uysa bile varyansın sıfır olduğunu ileri süren  $H_0 : \sigma = 0$  hipotezi aykırı ve/veya uç değer nedeni ile

kabul edilemeyebilir. Robust kestiriciler bu ihtiyaçları karşılamak için geliştirilmiştir. Robust kestiricilerin bazıları,  $LMS$  (Least Median Square),  $LTS$  (Least Trimmed Square),  $WLS$  (Weight Least Square),  $M$  kestirim yöntemleridir.

Robust kestirim yöntemine değinmeden önce robust regresyon kestiricilerini cazip kılan özelliklerden bahsederek;

- 1) Temel varsayımlar sağlanmadığında EKK'dan daha iyi bir performans göstermesi,
- 2) Temel varsayımların sağlandığı durumlarda ise hemen hemen EKK kadar başarı sağlaması,
- 3) Aşırı derecede hesaplama ve anlaşılma zorluğuna sahip olmamasının hedeflendiği söylenebilir.

Sıralı istatistiklerin doğrusal kombinasyonlarını temel alan robust kestirim yöntemleri  $L$  kestiriciler olarak isimlendirilmektedir.  $L$  kestiriciler matematiksel olarak;  $Z_{(1)}, Z_{(2)}, \dots, Z_{(n)}$   $n$  büyüklüğündeki örneklemin sıralı istatistikleri ve  $a_1 \leq a_2 \leq \dots \leq a_n$  gerçel sayılar ( $0 \leq a_i \leq 1, i = 1, 2, \dots, n$ ) olmak üzere; [5]

$$\sum_{i=1}^n a_i = 1 \quad (13a)$$

koşuluyla  $L$  kestiricisi,

$$T = \sum_{i=1}^n a_i x_i \quad (13b)$$

eşitliği ile tanımlanır.

Örneğin  $n$  büyüklüğündeki bir örneklemin ortalamasının ağırlığının  $1/n$  olduğu durum bir  $L$  kestiricisidir.

$L$  kestiricilerinin hesaplanması kolay olmasına rağmen aykırı değerlere ağırlık vererek sağlamlıklarından ya da aykırı olmayan değerlere daha küçük ağırlık vererek etkinliklerinden ödün vermelerine neden olmaktadır. [6]

Ağırlıklandırılmış hata karelerinin medyanını minimum yapmayı amaçlayan  $L$  kestiricisi:  $LMS$  (Least Median Square) kestirim yöntemidir. Bu yöntemde amaç fonksiyonu,

$$\min(\text{med}(w_i e_i^2)) \quad (14a)$$

biçiminde tanımlanır. Burada  $e_i, i$ 'nci gözlem hatasıdır. Bu yöntemde  $i$ 'inci gözlemin ağırlığı,

$$w_i = \begin{cases} 1 & |e_i / \hat{\sigma}| \leq 2.5 \\ 0 & |e_i / \hat{\sigma}| > 2.5 \end{cases} \quad (14b)$$

biçiminde hesaplanır.

$[-2.5, 2.5]$  aralığının dışına düşen standartlaştırılmış hata değerleri, potansiyel uç ya da aykırı değer olarak düşünülmektedir. Ağırlık fonksiyonunda ( $w_i$ ) bu aralığın dışındaki standartlaştırılmış hata değerlerine 0 değeri, aralık içindeki standartlaştırılmış hata değerlerine ise 1 değeri atanmaktadır.

2.5 değeri keyfi bir değerdir fakat normal durumda çok az artıklar  $2.5 \hat{\sigma}$ 'dan daha büyük olabileceğinden oldukça makul bir değerdir. [7]

14b numaralı ifadedeki  $\hat{\sigma}$ ,

$$\hat{\sigma} = 1.4826 \left( 1 + \frac{5}{n-p} \sqrt{\text{med}(e_i^2)} \right) \quad (14c)$$

olarak tanımlanır. Burada  $n$ , toplam gözlem sayısını;  $p$ , değişken sayısını ifade etmektedir. 1.4826 ve 5 sayısı,  $LMS$  kestirim yöntemi için, hata karenin medyanının dağılımını Normal Dağılım'a yaklaştırmak adına Rousseeuw ve Leroy tarafından önerilmiştir.

Budanmış (trimmed) hata karelerinin toplamını minimum yapmayı amaçlayan  $L$  kestiricisi:  $LTS$  (Least Trimmed Square) kestirim yöntemidir.  $LTS$  kestirim yöntemi 1984 yılında Rousseeuw tarafından geliştirilmiştir. Bu yöntemde amaç fonksiyonu şu şekilde tanımlanmaktadır: [8]

$$\min \sum_{i=1}^q e_{(i)}^2 \quad (15)$$

Burada  $q = [n(1-td) + 1]$ 'dir. Yöntem budanmış her bir gözlem için hesaplanan hata karelerinin toplamı şeklinde işlemektedir. Burada  $td$ , budanma oranını;  $n$  ise, örneklem büyüklüğünü ifade etmektedir.

Ağırlıklandırılmış hata karelerinin toplamını minimum yapmayı amaçlayan  $L$  kestiricisi:  $WLS$  (Weight Least Square) kestirim yöntemidir.  $WLS$  yönteminde amaç fonksiyonu;

$$\min \sum_i w_i e_i^2 \quad (16)$$

biçiminde tanımlanır.  $WLS$  yönteminde ağırlıklar, (14b) numaralı ifadedeki fonksiyonla hesaplanır. (14b) numaralı ifadedeki  $\hat{\sigma}$  değeri, (14c) numaralı eşitlik yardımıyla hesaplanır.

Bir diğer robust yöntem olan  $M$  kestirim yönteminde amaç fonksiyonu şu şekilde tanımlanır: [8]

$$\min \sum_{i=1}^n \rho(w e_i) \quad (17)$$

Burada  $\rho$ , bir fonksiyondur ve literatüre baktığımız zaman,  $M$  kestirim yöntemine ait çeşitli fonksiyonların bulunduğunu görmekteyiz. Bu fonksiyonlardan bazıları *Huber* fonksiyonu ve *Biweight* fonksiyonudur.

*Huber* fonksiyonunu kullanarak oluşturulan amaç fonksiyonu şu şekildedir: [9]

$$\min \sum_{i=1}^n \rho_H(w_H e_i) \quad (19a)$$

Burada  $p_H$ , *Huber* fonksiyonunu ifade etmekte,  $w_H$  ise, hatanın ağırlığını ifade etmektedir.  $p_H$  ve  $w_H$  aşağıdaki gibidir. [9]

$$\rho_H(w_H e) = \begin{cases} \frac{1}{2}(w_H e)^2 & , |w_H e| \leq k \\ k|w_H e| - \frac{1}{2}k^2 & , |w_H e| > k \end{cases} \quad (19b)$$

$$w_H = \begin{cases} 1 & , |e| \leq k \\ k/|e| & , |e| > k \end{cases} \quad (19c)$$

Burada  $k$  ifadesi *tuning* (ayar) sabitidir. Genelde, bu fonksiyon için  $k = 1.345$  değerini alır.

*Biweight* fonksiyonunu kullanarak ise oluşturulan amaç fonksiyonu şu şekildedir: [9]

$$\min \sum_{i=1}^n \rho_B(w_B e_i) \quad (20a)$$

Burada  $p_B$ , *Biweight* fonksiyonunu ifade etmekte,  $w_B$  ise, hatanın ağırlığını ifade etmektedir.  $p_B$  ve  $w_B$  aşağıdaki gibidir. [9]

$$\rho_B(w_B e) = \begin{cases} \frac{k^2}{6} \left\{ 1 - \left[ 1 - \left( \frac{w_B e}{k} \right)^2 \right]^3 \right\} & , |w_B e| \leq k \\ k^2/6 & , |w_B e| > k \end{cases} \quad (20b)$$

$$w_B = \begin{cases} \left[ 1 - \left( \frac{e}{k} \right)^2 \right]^2 & , |e| \leq k \\ 0 & , |e| > k \end{cases} \quad (20c)$$

*Tuning* sabiti  $k$ , genelde bu fonksiyon için  $k = 4.685$  değerini alır.

### III. FUZZY REGRESYON

Bir fuzzy regresyon modeli, sistemin içermesi gereken olanaklı bütün gözlemlenmiş veriyi değerlendirir. Bir başka ifade ile fuzzy regresyon modeli, gözlemlenen tüm verileri içeren bir model oluşturmayı amaçlar.[10]

FLR (Fuzzy Linear Regression) modelinin genel yapısı aşağıdaki gibidir:

$$\hat{Y}_i = A_0 + A_1 X_1 + \dots + A_k X_k, \quad i = 1, 2, \dots, n \quad (21a)$$

$$\hat{Y}_i = (m_0, s_0) + (m_1, s_1)X_1 + \dots + (m_k, s_k)X_k \quad (21b)$$

Bu modelde bağımlı değişken olan gözlem değerleri  $X_j$  kesin sayılar olmasına karşın ilişkinin bulanıklığı nedeni ile bulanık sayı olarak kestirilir. Regresyon denkleminin hata miktarı, sistem parametrelerinin toplam bulanıklığına eşit olacaktır [11]. Bu nedenden dolayı aynı zamanda regresyon denklemi olan bu modelde, epsilon ( $\varepsilon$ ) değeri bulunmamakta ve epsilon, kendini sistem parametrelerinin toplam bulanıklığında göstermektedir. Bulanıklığı minimize ederek en az hataya ulaşılmış olunur. Bu nedenle doğrusal programlama yöntemi kullanılarak tahmin değerlerin gözlem değerlerini belirli bir  $h$  seviyesinde içermesi kısıtı altında bulanık parametrelerin toplam yayılımları minimize edilir.

Tanaka'nın önerdiği doğrusal programlama modeli aşağıdaki gibidir: [12]

$$\text{Minimize } ns_0 + \sum_i \sum_j s_j |X_{ij}| \quad (22a)$$

$$\forall i \text{ için, } (m_0 + (1-h)s_0) + \sum_j (m_j + (1-h)s_j) |X_{ij}| \geq y_i + (1-h)e_i \quad (22b)$$

$$\forall i \text{ için, } (m_0 - (1-h)s_0) + \sum_j (m_j - (1-h)s_j) |X_{ij}| \leq y_i - (1-h)e_i \quad (22c)$$

$$s_0 \geq 0 \quad (22d)$$

$$s_j \geq 0 \quad (22e)$$

$$s_j |X_{ij}| \geq 0 \quad (22f)$$

$$i = 1, 2, \dots, n \quad (22g)$$

$$j = 1, 2, \dots, k \quad (22h)$$

Doğrusal programlama modelinde ifade edilen amaç fonksiyonunda toplam yayılım minimize edilmeye çalışılmaktadır. Kısıtların olduğu bölümde ise, gözlem değerlerinin belirlenen  $h$  seviyesinde tahmin edilen bulanık küme tarafından içerilmesi gerektiği formüle edilmiştir. Doğrusal programlama modelinin çözülmesiyle bulanık sistem parametrelerinin diğer bir

ifade ile  $\tilde{A}_j$ 'lerin merkez ve yayılım değerleri hesaplanılmaktadır.

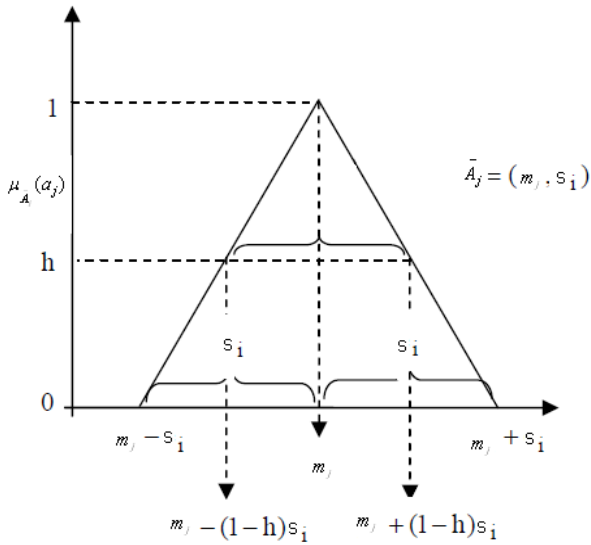
Tanaka'nın önerdiği doğrusal programlama probleminin en önemli varsayımı  $Y_i$  gözlem değerlerinin tamamının, tahmin edilen alt ve üst sınır tarafından kapsanması gerektiği varsayımdır. Bu varsayım aşağıdaki gibi gösterilebilir.

$$y_i \in \left[ Y_i \right]_h, \quad i = 1, 2, \dots, n \quad (23)$$

$a_j$  katsayılarının üyelik fonksiyonu aşağıdaki gibidir.

$$\mu_{\tilde{A}_j}(a_j) = \begin{cases} 1 - \frac{|m_j - a_j|}{s_j}, & m_j - s_j < a_j < m_j + s_j, \quad \forall j = 1, 2, \dots, k \\ 0, & \text{ö.d.} \end{cases} \quad (23)$$

Şekil 2.  $a_j$  Katsayılarının Üyelik Fonksiyonu [12]

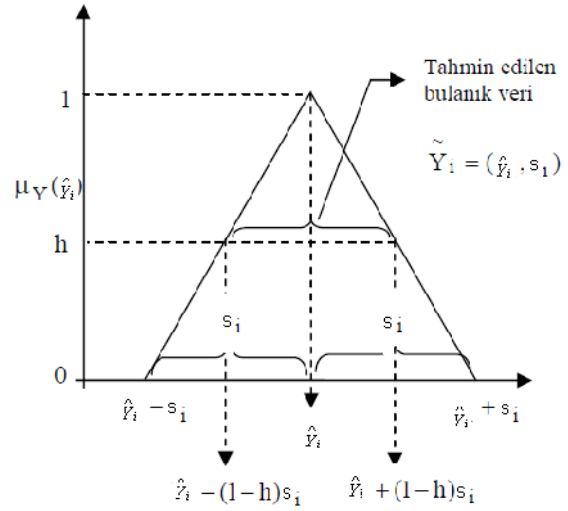


$Y_i$  değerlerinin üyelik fonksiyonu ise,

$$\mu_{\tilde{Y}_i}(\hat{Y}_i) = \begin{cases} 1 - \frac{|\hat{Y}_i - Y_i|}{s_i}, & \hat{Y}_i - s_i < Y_i < \hat{Y}_i + s_i \\ 0, & \text{ö.d.} \end{cases} \quad (24)$$

şeklindedir.

Elde edilen modelin uygunluğunu ölçmek için aşağıdaki iki istatistik kullanılır. Bu iki istatistik değerinin mümkün olduğunca küçük olması, modelin uygunluğu açısından önemlidir. Tanaka uygun bir model için  $FR1$  istatistiğinin %5'den,  $FR2$  istatistiğinin %10'dan küçük olması gerektiğini öne sürmüştür.

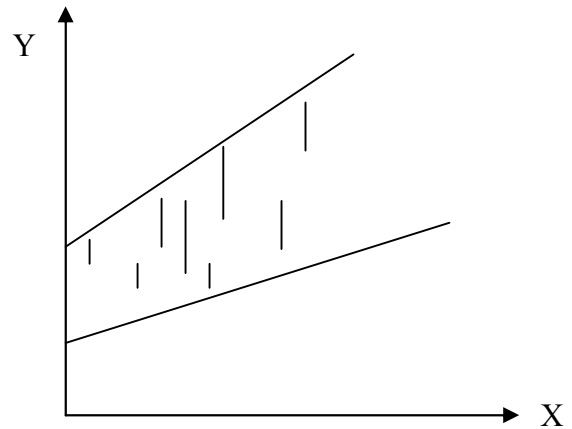


Şekil 3.  $Y_i$  Değerlerinin Üyelik Fonksiyonu [12]

$$FR1 = \frac{|\hat{Y}_i - Y_i|}{Y_i} \quad (25a)$$

$$FR2 = \frac{\sum s_j |X_{ij}|}{Y_i} \quad (25b)$$

Bu doğrusal programlama probleminin çözülmesiyle elde edilen tahminlerin görünümü aşağıdaki şekilde görüldüğü gibi olacaktır,



Şekil 4. Tanaka Modelinde Tahminlerin Görünümü

Fuzzy Regresyon Analizi'nde diğer bir yöntem "Regresyon Katsayılarının Bulandırılması" yöntemidir. Bu yöntemde Klasik Regresyon Analizi ile elde edilen iyi modelin katsayıları, seçilen bir  $h$  seviyesinde bulandırılır. Yöntemin adımları aşağıdaki gibi sıralanabilir: [13]

**1. adım.** Klasik regresyon yöntemi ile en uygun model tespit edilir.

**2. adım.** Açıklayıcı değişkenlerin gözlem değerleri ortalama değerlerinden saptırılarak bulandırma işlemine başlanır ve klasik regresyon modeli şu hale dönüşür:

$$\hat{Y} = m_0 + \sum_i m_i (X_i - X_{i \text{ ortalama}}) \quad (27)$$

Örneğin iki adet bağımsız değişken olduğunda,

$$\hat{Y} = m_0 + m_1 (X_1 - X_{1 \text{ ortalama}}) + m_2 (X_2 - X_{2 \text{ ortalama}}) \quad (28)$$

$X_{\text{ortalama}}$  değeri tüm gözlem değerlerin aritmetik ortalaması olabileceği gibi, sadece en büyük ve en küçük gözlem değerlerinin ortalaması alınarak da bulunabilir.

$$X_{\text{ortalama}} = 0.5 (\max X_j + \min X_j) \quad (29)$$

**3. adım.** Bağımsız değişkenler ortalamalarından saptırıldıktan sonra bu adımda parametreler bulandırılır. Bu işlem alttaki formül kullanılarak gerçekleştirilir.

$$\hat{\beta}_j - \delta_j L_j^{-1}(h) \leq \beta_j \leq \hat{\beta}_j + \eta_j R_j^{-1}(h) \quad (30)$$

$L(X) = R(X) = 1 - X$  ve  $\delta_j = \eta_j$  olduğunu varsayarak bulandırma yapılır.  $\delta_j$  ve  $\eta_j$ 'nin değerleri analist tarafından belirlenir.  $h$  seviyesi bir sabit girdi olarak  $[0,1]$  aralığından seçilir ve analistin çalışacağı bulanıklık seviyesini ifade eder.  $h = 0.5$  alınırsa,  $L(h) = R(h) = 1 - 0.5 = 0.5$  bulunur. Bulandırma işleminde kullanılmak üzere  $L_j^{-1}(h)$  ve  $R_j^{-1}(h)$  değerleri aşağıdaki gibi bulunur:

$$L_j^{-1}(h) = R_j^{-1}(h) = 1/0.5 = 2$$

elde edilir.

Bulandırma işlemi sonucunda  $\beta_j$  parametresi için bir alt ve bir üst sınır hesaplanarak  $\beta_j$  aralığı bulunur. Aslında yapılan işlem aralık regresyonu gibi görünse de  $h \neq 0$  olduğunda analiz, bulanık regresyon analizidir.

**4.adım.**  $\hat{Y}$  bulanık değerler tahmin edilir. Diğer bir ifade ile  $\hat{Y} = \hat{\beta}_0 + \sum_i \hat{\beta}_i (X_i - X_{i \text{ ortalama}})$  denkleminde hesaplanır.

**5.adım.**  $\hat{Y}_i$  tahmin değerleri,  $y_i$  gözlem değerleri ve klasik regresyondan tahmin edilen  $\hat{Y}_i$ 'lerle kıyaslanarak hangisinin daha iyi olduğuna ilişkin yorumlar yapılır.

Başka bir yöntem, Celmins'in geliştirdiği veri seti ve regresyon denklemi arasındaki uyum ölçüsüne dayanan *Maksimum Uyum Kriteri* yöntemidir.  $\mu_A(x)$  ve  $\mu_B(x)$

şırasıyla  $\hat{A}$  ve  $\hat{B}$  bulanık sayılarının üyelik fonksiyonları olmak üzere,  $\hat{A}$  ve  $\hat{B}$  için uyum ölçüsü  $\gamma(\hat{A}, \hat{B})$  biçiminde gösterilsin. Eğer  $\mu_A(x)$  ve  $\mu_B(x)$  normalleştirilmiş üçgen üyelik fonksiyonları ise Şekil 2.10'da gösterildiği gibi  $\gamma(\hat{A}, \hat{B})$ ,

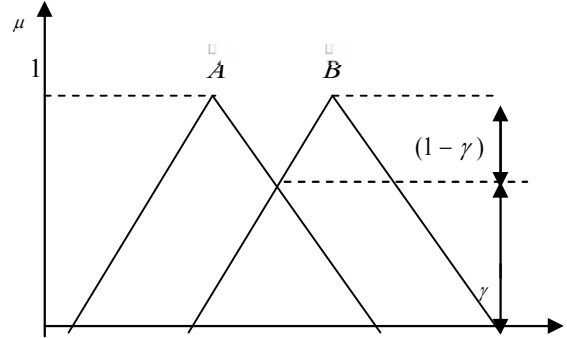
$$\gamma(\hat{A}, \hat{B}) = \max_x \min \{ \mu_A(x), \mu_B(x) \} \quad (36)$$

biçiminde açıklanır. Burada  $\gamma \in [0,1]$  olacaktır ve Şekil 2.11'de gösterildiği gibi eğer iki bulanık sayı örtüşmüyorsa  $\gamma = 0$ , eğer iki bulanık sayısının merkezleri çakışıyorsa  $\gamma = 1$  değerini alacaktır. [14]

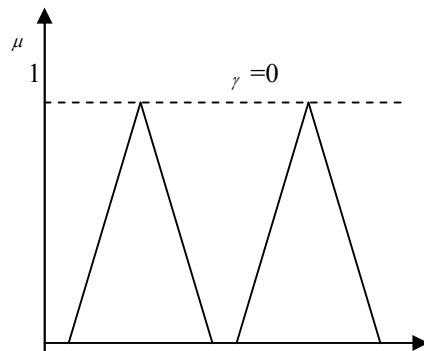
Bu yaklaşıma göre amaç, veri ve model arasındaki maksimum genel uyum derecesine sahip bir modelin belirlenmesidir.  $\gamma_i$ , veri seti ve tahmin denklemi arasındaki uyum derecesini göstermek üzere; genel uyum ölçüsü  $\gamma_i$ 'nin 1'den sapmalarının kareleri toplamına eşit olacaktır. Burada amaç, [15]

$$\min W = \sum_{i=1}^m (1 - \gamma_i)^2 \quad (37)$$

biçiminde verilen  $w$  fonksiyonunun en küçüklenmesidir.



Şekil 5.  $\gamma$  Uyum Ölçüsü [15]



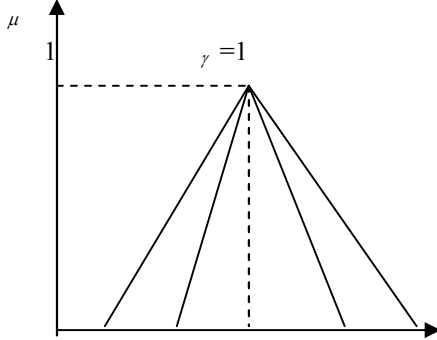
Şekil 6.  $\gamma = 0$  Uyum Ölçüsü [15]

Bu yaklaşıma göre maksimum uyum kriterini kullanarak bulanık en küçük kareler regresyonu için model,

$$\hat{Y} = \hat{A}_0 + \hat{A}_1 X$$

$$= m_0 + m_1 X \pm \sqrt{s_0^2 + 2s_{01}X + s_1^2 X^2} \quad (38)$$

biçiminde oluşturulur. [15]



Şekil 7.  $\gamma =1$  Uyum Ölçüsü [15]

(38) numaralı denklemin ilk kısmı olan  $m_0 + m_1 X$ , bulanık regresyon modelinin merkez doğrusunu göstermektedir.  $m_0$  ve  $m_1$  katsayıları ağırlıklı en küçük kareler regresyonu yardımıyla elde edilir. (38) numaralı denklemin ikinci kısmı  $\pm \sqrt{s_0^2 + 2s_{01}X + s_1^2 X^2}$ , bulanık regresyon modelinin alt ve üst sınırlarını göstermektedir.  $s_0$  ve  $s_1$  ise  $\hat{A}_0$  ve  $\hat{A}_1$  bulanık katsayılarının genişliklerini ifade etmektedir. Celmins'e göre  $s_{01}$ ,  $\hat{A}_0$  ve  $\hat{A}_1$  arasındaki bulanık uyumluluk olarak tanımlanmıştır ve klasik parametrelerdeki kovaryansa karşılık gelmektedir. İteratif hesaplamalar sonucunda  $s_0$ ,  $s_1$  ve  $s_{01}$ ,  $[0,1]$  aralığı içerisinde elde edilir. [16]

Başka bir Fuzzy Regresyon Analizi Yöntemi de *Fuzzy Linear Regresyon Analizine EKK Yaklaşımıdır*.  $Y_i$  simetrik üçgen fuzzy sayısı,  $Y_i \equiv (m_i, s_i)$   $i \in \{1, 2, \dots, n\}$  biçiminde tanımlanabilir. Lineer regresyon analizini ise matris formatında şu şekilde tanımlayabiliriz: [17]

$$\mathbf{m} = \mathbf{m}^* + \varepsilon_m \text{ burada } \mathbf{m}^* = \mathbf{X} \mathbf{a}, \quad (39a)$$

$$\mathbf{s} = \mathbf{s}^* + \varepsilon_s \text{ burada } \mathbf{s}^* = \mathbf{m}^* \mathbf{b} + \mathbf{1} d \quad (39b)$$

Burada  $\mathbf{X}$ ;  $n \times (k+1)$  boyutlu girdi verilerinden oluşan matrisi,  $\mathbf{a}$ ; regresyon parametrelerini içeren  $1 \times (k+1)$  boyutlu vektörü,  $\mathbf{m}$  ve  $\mathbf{m}^*$ ; sırasıyla gözlenen ve tahmin edilen çıktı verilerinden oluşan, her ikisi de  $n \times 1$  boyutlu matrisi,  $\mathbf{s}$  ve  $\mathbf{s}^*$ ; sırasıyla gözlenen ve tahmin edilen yayılımlardan oluşan, her ikisi de  $n \times 1$  boyutlu matrisi,  $\mathbf{1}$ ; bütün değerleri 1'den oluşan  $n \times 1$  boyutlu vektörü,  $b$  ve  $d$ ; yayılım regresyon modeli olarak adlandırılan ikinci regresyon modelinin parametrelerini ifade etmektedir.

$Y_i \equiv (m_i, s_i)$  ve  $Y_i^* \equiv (m_i^*, s_i^*)$  arasındaki öklid uzaklık şu şekilde tanımlanabilir:

$$\delta_i \equiv \delta(Y_i, Y_i^*) = \sqrt{(m_i - m_i^*)^2 + (s_i - s_i^*)^2} \quad (40)$$

Fuzzy linear regresyon analizine EKK yaklaşımı öklid uzaklıklarının toplamını minimize etme esasına dayanmaktadır. [17]

$$\varphi(\mathbf{a}, \mathbf{b}, d) \equiv \sum_{i=1}^n \delta_i^2 = (\mathbf{m} - \mathbf{m}^*)'(\mathbf{m} - \mathbf{m}^*) + (\mathbf{s} - \mathbf{s}^*)'(\mathbf{s} - \mathbf{s}^*) \quad (41)$$

EKK'nın iteratif çözümü sonrasında parametre değerlerin kestirimi şu şekilde olacaktır. [17]

$$\mathbf{a} = \frac{1}{(1 + b^2)} ((\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'(\mathbf{m} + \mathbf{s}b - \mathbf{1}d)), \quad (42a)$$

$$b = (\mathbf{a}'\mathbf{X}'\mathbf{X}\mathbf{a})^{-1} (\mathbf{s}'\mathbf{X}\mathbf{a} - \mathbf{a}'\mathbf{X}'\mathbf{1}d), \quad (42b)$$

$$d = \frac{1}{n} (\mathbf{s}'\mathbf{1} - \mathbf{a}'\mathbf{X}'\mathbf{1}b). \quad (42c)$$

#### IV. FUZZY ROBUST REGRESYON

Fuzzy Robust Regresyon yaklaşımı, veri seti üzerindeki düzensiz verilerin model üzerindeki etkisini azaltmak amacıyla kullanılan bir yöntemdir. Ishibuchi, bazı marjinal etkiye sahip verileri silmeye yönelik bir yaklaşım önermiştir. Watana ve Yabuuchi, genetik algoritma yöntemini kullanarak düzensiz verileri içerisinde barındıran bir yaklaşım önermiştir. Ayrıca Watana ve Yabuuchi, hipereliptik fonksiyona ( $y^2 = f(x)$ ) dayanan bir robust FLR yaklaşımı önermiştir. Chen, önceden belirlenmiş  $k$  sınırlayıcı (pre-assigned  $k$ -limiting) yöntemini kullanarak modifiye güven aralığı yöntemini önermiştir. Yang ve Liu, gürültü küme (noise cluster) yöntemini kullanarak yeni bir fuzzy en küçük kareler algoritmasını önermişler ve bu algoritma ile düzensiz verilerin model üzerinde olumsuz etkilerini gidermeye çalışmışlardır.

Fuzzy robust regresyon modeli, modeldeki tüm verilerle inşa edilmeye, oluşturulmaya çalışılır. Gerçek dünyada veriler arasında düzensiz verilerin bulunması kaçınılmazdır. Bu düzensiz veriler, gözlenen değerlerdeki hatalardan, gözlem metotlarındaki eksiklik, hata veya yanlışlıklardan, gözlenen değerlerin kaybolmasından, verinin aykırı ya da uç değer olmasından ya da diğer nedenlerden kaynaklanmaktadır. Bunun sonucu olarak, bu tip verileri klasik regresyon yaklaşımı ile kurmaya çalıştığımızda, kurulan model gerçekliğini yitirmektedir. Bu yüzden bu yaklaşımda düzensiz verilerin model üzerindeki etkisi, modelin kurulması aşamasında ortadan kaldırılması amaçlanmaktadır. [18]



Üçgen fuzzy sayısı,  $a_{1j}$  ve  $a_{3j}$  iki uç noktayı,  $a_{2j}$  merkez noktayı göstermek üzere,  $A_j = (a_{1j}, a_{2j}, a_{3j})$  bulanık aralığı olarak tarif edilebilir.

$$Y_i = A_0 + A_1x_{ij} + \dots + A_jx_{ij}, \quad i = 1, \dots, n \quad (43)$$

olmak üzere,  $A_j$ 'nin üyelik fonksiyonu: [18]

$$\mu_{A_j}(x) = \begin{cases} 0 & , x < a_{1j} \\ 1 - \frac{a_{2j} - x}{a_{2j} - a_{1j}} & , a_{1j} \leq a_{2j} \\ 1 - \frac{x - a_{3j}}{a_{3j} - a_{2j}} & , a_{2j} \leq a_{3j} \\ 0 & , x > a_{3j} \end{cases} \quad (44)$$

ve  $A_{jh} = (a_{1j}^{(h)}, a_{3j}^{(h)})$ ,  $h$  -kesit (cut) aralığıdır. Burada  $a_{1j}^{(h)} = a_{1j} + (a_{2j} - a_{1j})h$  ve  $a_{3j}^{(h)} = a_{3j} - (a_{3j} - a_{2j})h$  'dir.

Fuzzy üçgen sayılarla oluşturulan FLR model:

$$Y_j^* = (a_{10}, a_{20}, a_{30})x_{i0} + (a_{11}, a_{21}, a_{31})x_{i1} + \dots + (a_{1j}, a_{2j}, a_{3j})x_{ij} \quad (45a)$$

$$= \left( \sum_j a_{1j}x_{ij}, \sum_j a_{2j}x_{ij}, \sum_j a_{3j}x_{ij} \right) \quad (45b)$$

Bu modele ait üyelik fonksiyonu ise; [18]

$$\mu_{Y_j}(y) = \begin{cases} 0 & , y < \sum_j a_{1j}x_{ij} \\ 1 - \frac{\sum_j a_{2j} - y}{\sum_j (a_{2j} - a_{1j})x_{ij}} & , \sum_j a_{1j}x_{ij} \leq y \leq \sum_j a_{2j}x_{ij} \\ 1 - \frac{y - \sum_j a_{3j}}{\sum_j (a_{3j} - a_{2j})x_{ij}} & , \sum_j a_{2j}x_{ij} \leq y \leq \sum_j a_{3j}x_{ij} \\ 0 & , y > \sum_j a_{3j}x_{ij} \end{cases} \quad (46)$$

Aşağıdaki fuzzy lineer programlama probleminin çözümü ile  $a_{1j}$ ,  $a_{2j}$  ve  $a_{3j}$  tahmin edilir. [18]

$$\min \sum_{i=0}^{n-1} \sum_{j=0}^p (a_{2j} - a_{1j})x_{ij} + \sum_{i=0}^{n-1} \sum_{j=0}^p (a_{3j} - a_{2j})x_{ij} \quad (47a)$$

$$\forall i \text{ için, } \sum_j a_{2j} - (1-h) \sum_j (a_{2j} - a_{1j})x_{ij} \leq y \quad (47b)$$

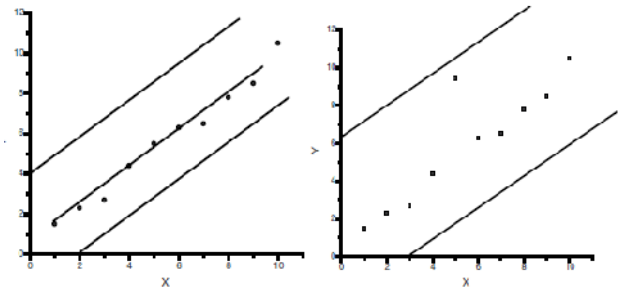
$$\forall i \text{ için, } \sum_j a_{2j} - (1-h) \sum_j (a_{2j} - a_{1j})x_{ij} \geq y \quad (47c)$$

$$(a_{2j} - a_{1j}) \geq 0 \quad (47d)$$

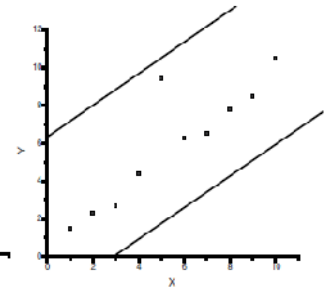
$$(a_{3j} - a_{2j}) \geq 0 \quad (47e)$$

Şekil 5'de uç noktalar olan  $\sum_j a_{1j}x_{ij}$  ve

$\sum_j a_{3j}x_{ij}$  fuzzy üçgen sayıları sırasıyla alt limit çizgisi ve üst limit çizgisidir. Aykırı değer çıktısına sahipsek Şekil 6'da görüldüğü gibi üst ve alt limit sınırları aykırı değeri kapsamaktadır.



Şekil 8. FLR Model



Şekil 9. FLR Model ile Aykırı Değer

Merkez nokta olan  $a_{2j}$ 'nin ve artık olarak  $r_{mj}$ 'nin  $M$ -Kestirici yöntemiyle tanımlanması aşağıdaki gibidir [18].

$$r_{mj} = y_i - \sum_j a_{mj}x_{ij} \quad (48a)$$

$$IQR(r_{mj}) = Q_3 - Q_1 \quad (48b)$$

Burada  $Q_1$  ve  $Q_3$ , sırasıyla birinci ve üçüncü kartillerdir. Düzensiz çıktı verisi, sınırı aşan arama puanlama yöntemiyle bulunabilir.

$$Q_1 - 1.5 \times IQR(r_{mj}) \text{ veya } Q_3 + 1.5 \times IQR(r_{mj}) \quad (49)$$

Şimdi  $h_m$ 'yi tanımlayalım. [18]

$$h_m = \frac{|y_i - y_{mi}|}{y_i} = \frac{|r_{mi}|}{y_{mi}} \quad (50)$$

Burada  $y_{mi} = \sum_j a_{mj}x_{ij}$  'dir.  $i$  . çıktı verisi düzensiz ise (47b) ve (47c) numaralı kısıtlar altındaki  $i$  . kısıtın  $h$  ve  $y_i$ 'si,  $M$  - Kestirim Yöntemine dayanan  $h_m$  ve yudurulmuş değer olan  $y_{mi}$  tarafından değiştirilir.

Büyük değer  $y_i$ , büyük yayılımın ve düzensiz çıktı verisinin etkisini azaltmak için,  $h_m$  ve  $y_{mi}$  tarafından küçültülür. [18]

**V. UYGULAMA**

Çoklu Doğrusal Regresyon analizi ile bulunan modelin, diğer üç teknikle (Robust Regresyon Analizi, Fuzzy Regresyon Analizi, Fuzzy Robust Regresyon Analizi) nasıl sonuç verdiği otomobil beğeni değerleri için gözlenen 45 otomobil markası verileriyle denenmiştir. Çoklu doğrusal regresyon çalışmasında Otomobil Beğeni Değeri'ni açıklamada sekiz açıklayıcı değişken olarak otomobilin fiyatı, silindir hacmi, beygir gücü, maksimum hız, saniye olarak 0'dan 100 km/saat'e çıkma süresi, 100 kilometrede şehir dışı kaç litre yaktığı, 100 kilometrede şehir içinde kaç litre yaktığı ve bir kilometre gitmenin maliyeti belirlenmiştir. Aynı değişkenlerle diğer üç tekniğin nasıl sonuç vereceği araştırılmıştır.

$X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8$ , ve  $Y$  değerleri;

$X_1$ : Otomobilin Fiyatını (TL),

$X_2$ : Otomobilin Silindir Hacmi (cc),

$X_3$ : Otomobilin Beygir Gücü (HP),

$X_4$ : Otomobilin Maksimum Hızı (km/saat),

$X_5$ : 0'dan 100 km/saat'e Çıkma Süresi (saniye),

$X_6$ : Otomobilin 100 Kilometrede Şehir Dışında Kaç Litre Yaktığı'nı (litre),

$X_7$ : Otomobilin 100 Kilometrede Şehir İçinde Kaç Litre Yaktığını (litre)

$X_8$ : Otomobilin Bir Kilometre Gitmesinin Maliyeti

$Y$ : Otomobil Beğeni Değeri'ni ([0-10] aralığı üzerinden) ifade etmektedir.

Fuzzy hedef çıktısı ( $y_i$ ) aşağıdaki şekilde oluşturulmuştur:

Otomobilin bir kilometre gitme maliyeti,

$$\frac{((X_6 + X_7) / 2) \cdot Y a k}{100} + \frac{B a k}{10000 \text{ veya } 15000} + \frac{C a k}{15000 \text{ veya } 25000 \text{ veya } 30000}$$

Çıktıya baktığımızda oluşturulan regresyon denkleminin,

$$\hat{Y} = 11.645 + 2.348 E - 6 X_1 + 0.009 X_3 - 0.019 X_4 - 0.101 X_5 + 0.233 X_6 + 0.042 X_7 - 4.589 X_8$$

olduğunu ve  $R^2$ 'nin 0.268 olduğunu görmekteyiz. Bu sonuca göre, bağımlı değişkendeki %26.8'lik değişim modele dahil ettiğimiz bağımsız değişkenler tarafından

$$y_i = f_j(x_{i1}, x_{i2}, \dots, x_{i8}), \quad i = 1, \dots, 45 \quad j = 1, \dots, 8. \quad (51)$$

Tablo 1'deki  $s$  (spread) yayılım değerleri ise, gözlenen otomobil beğeni değerlerinin minimum veya maksimum değerleri yardımıyla aşağıdaki gibi hesaplanmıştır.

$$s_i = Y_i - Y_{i \min} = Y_{i \max} - Y_i \quad (52)$$

Ayrıca  $APE'$  adında bir hata ölçütü tanımlanmıştır.  $APE'$  ne kadar küçükse tahmin değerleri o denli başarılı olduğunu söyleyebiliriz. [19]

$$APE' = \frac{1}{N} \sum_{i=1}^N \left| \frac{Y_i - \hat{Y}_i}{Y_i} \right| \quad (53)$$

Gözlem değerleri ile Bölüm 2.1'de açıklanan Çoklu Doğrusal Regresyon Analizi çalışması SPSS programıyla yapılmış ve aşağıdaki sonuçlar bulunmuştur.

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.517 <sup>a</sup>	.268	.105	.33483

a. Predictors: (Constant), X2, X3, X4, X5, X6, X7, X8

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	.474	8	.059	1.873	.147 <sup>a</sup>
	Residual	4.056	36	.112		
	Total	5.510	44			

a. Predictors: (Constant), X2, X3, X4, X5, X6, X7, X8

b. Dependent Variable: Y

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error			
1	(Constant)	11.645	2.323		5.012	0.000
	X1	-2.348E-06	.000	-.111	-.294	.770
	X2	.000	.000	-.115	-.623	.537
	X3	.000	.000	.043	1.402	0.167
	X4	-.019	.010	-.362	-1.824	0.078
	X5	-.101	.066	-.565	-1.808	0.079
	X6	.233	.206	.523	1.122	0.264
	X7	.042	.124	.183	.338	0.737
	X8	-4.589	2.682	-.313	-1.778	0.084

a. Dependent Variable: Y

açıklanmaktadır. ANOVA tablosuna baktığımızda, tablodaki 1.643 F değeri, modelimizin bir bütün olarak anlamlı olmadığını göstermektedir (Sig.=0.147>0.05).

Tablo 1: Girdi Değişken Verileri ve Hedef Çıktılar

n	X1	X2	X3	X4	X5	X6	X7	X8	Ymin	Y	Ymax	s
1	57997,768	1598	122	203	7,6	4,8	7,6	0,40	7,0	7,35	7,7	0,4
2	44600	1290	124	185	11,9	5,9	9	0,43	7,9	8,1	8,3	0,2
3	96173,64	1480	156	230	9,5	5,7	9,9	0,53	4,6	6,45	8,3	1,9
4	60642,512	1390	122	203	8,9	4,4	6,8	0,43	7,3	7,55	7,8	0,3
5	28170	1390	85	174	12,3	5,3	8,6	0,42	7,2	7,45	7,7	0,3
6	46600	1598	120	190	10,9	5,3	9,6	0,41	7,5	7,85	8,2	0,4
7	55980	1368	120	195	9,4	5,4	8,6	0,41	8,1	8,2	8,3	0,1
8	93169	1956	160	218	9,5	4,7	7,7	0,41	7,6	7,75	7,9	0,2
9	33908	1598	124	190	10,9	5,1	8,8	0,39	7,7	7,95	8,2	0,3
10	23900	1086	69	156	15	4,1	5,8	0,28	7,5	7,75	8,0	0,3
11	27516	1206	81	164	12,1	4,2	6,6	0,31	7,3	7,65	8,0	0,4
12	37175	1368	120	197	9,6	5,5	8,6	0,40	7,2	7,35	7,5	0,2
13	57730	1598	156	204	10,9	5,4	8,8	0,42	7,4	7,45	7,5	0,0
14	33900	1396	108	185	11,5	4,8	7,8	0,36	7,6	7,8	8,0	0,2
15	48163	1398	140	202	9,9	4,7	7,8	0,37	7,3	7,45	7,6	0,2
16	55000	1598	120	203	11,5	4,6	9	0,39	7,1	7,4	7,7	0,3
17	39800	1197	105	187	10,9	4,9	7,1	0,34	8,0	8,1	8,2	0,1
18	29282	1240	80	167	13,5	5,1	7,4	0,35	7,0	7,35	7,7	0,4
19	61700	1390	122	205	10,3	5,2	7,3	0,38	7,5	7,65	7,8	0,1
20	29904	1368	77	165	13,2	5	7,5	0,36	7,5	7,6	7,7	0,1
21	39100	1582	90	172	13,8	4,2	5,7	0,29	7,7	7,9	8,1	0,2
22	23990	1248	74	170	16	4	6,4	0,26	7,1	7,4	7,7	0,3
23	48900	1364	90	175	11,9	4,0	5,6	0,29	7,9	8,1	8,3	0,2
24	42600	1598	90	180	11,5	3,7	5,2	0,26	7,3	7,5	7,7	0,2
25	59700	1560	112	190	12,8	4,5	6,4	0,32	7,5	7,55	7,6	0,0
26	44170	1598	105	190	11,8	3,9	5,7	0,29	7,8	7,9	8,0	0,1
27	56850	1461	110	182	12,5	4,4	6,1	0,31	7,1	7,2	7,3	0,1
28	61578	1598	105	185	11,3	3,7	5,5	0,28	8,1	8,2	8,3	0,1
29	55300	1598	105	190	11,7	4,2	5,6	0,29	7,8	7,95	8,1	0,2
30	44600	1595	102	188	11,3	5,5	8,8	0,37	7,5	7,75	8,0	0,3
31	37200	1598	90	178	11,8	3,6	5,1	0,26	7,2	7,3	7,4	0,1
32	39430	1560	90	173	13,4	4,1	6,1	0,28	7,0	7,25	7,5	0,3
33	39300	1598	110	183	11,7	5,3	8,8	0,36	7,6	7,8	8,0	0,2
34	56019	1248	95	170	14,7	3,6	5,1	0,25	7,3	7,45	7,6	0,2
35	37050	1598	110	185	11,7	5,3	8,8	0,36	7,1	7,2	7,3	0,1
36	48700	1461	105	185	10,9	4	5,5	0,28	7,1	7,2	7,3	0,1
37	80600	1598	200	237	7,5	5,6	9,1	0,43	8,0	8,25	8,5	0,3
38	72649,6	1390	160	218	8	5,4	8,7	0,43	7,8	8,15	8,5	0,4
39	49600	1560	156	214	7,3	5,1	9,4	0,41	7,5	7,85	8,2	0,4
40	39970	1197	105	175	11,8	5,9	7,9	0,39	7,7	7,85	8,0	0,2
41	47750	1598	120	195	11,6	5,6	9,8	0,44	7,4	7,65	7,9	0,3
42	34400	1598	105	160	12,8	7	10,4	0,48	7,7	8,1	8,5	0,4
43	75900	1987	158	185	11	6,5	9,6	0,53	7,4	7,7	8,0	0,3
44	28370	1198	70	158	15,9	5,5	8,7	0,39	7,5	7,7	7,9	0,2
45	42750	1591	125	178	12,2	5,8	9	0,41	7,7	7,9	8,1	0,2

Kaynak: Otohaber Test Yılığ 2011, Auto Show Dergisi Sayı:26, car.tr.diploptop.com ( $s_i = Y_i - Y_{i\min} = Y_{i\max} - Y_i$ )

Katsayılar tablosunda  $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \beta_8$  için bulunan  $b_0, b_1, b_2, b_3, b_4, b_5, b_6, b_7, b_8$  istatistikleri kullanılarak  $H_0 : \beta_0 = 0, H_0 : \beta_1 = 0, H_0 : \beta_2 = 0, H_0 : \beta_3 = 0, H_0 : \beta_4 = 0, H_0 : \beta_5 = 0, H_0 : \beta_6 = 0, H_0 : \beta_7 = 0, H_0 : \beta_8 = 0$  hipotezleri için sınama sonuçları görülmektedir.  $H_0 : \beta_0 = 0$  hipotezi için elde edilen sig. değeri dışında diğer hipotezler için elde edilen sig. değeri %5'den büyük çıkmıştır. Sonuç olarak  $H_0 : \beta_0 = 0$  hipotezi %95 güven düzeyinde reddedilmiş, hesaplanan istatistiğin sıfır olamayacağı belirlenmiştir. Diğer hipotezler ise %95 güven düzeyinde reddedilememiş, hesaplanan istatistiklerin sıfır olduğu belirlenmiştir. Modelin bir bütün olarak anlamlı bulunmamasından ve en az bir bağımsız değişkenin anlamlı bulunmamasından dolayı elimizdeki veri seti için çoklu doğrusal regresyon analizinin uygun olmadığına karar verilmiştir.

Veri seti üzerine, 3. Bölüm'de açıklanan LMS Kestirme Yöntemi ve  $LMS$ 'ye dayalı yeniden ağırlıklandırılmış  $EKK$  Yöntemi ( $WLS$  or  $RLS$ ) SAS/IML programı ile yapılmış ve aşağıdaki sonuçlar bulunmuştur.

Çıktıya baktığımızda oluşturulan regresyon denkleminin,

$$\hat{Y} = 6.45 - 1.52E - 5X_1 - 8.65E - 04X_2 + 0.015X_3 + 0.00295X_4 + 0.021X_5 + 0.413X_6 - 0.074X_7 - 2.069X_8$$

olduğunu ve Robut  $R^2$ 'nin 0.52 olduğunu görmekteyiz. Bu sonuca göre, bağımlı değişkendeki %52'lik değişim, modele dahil ettiğimiz bağımsız değişkenler tarafından açıklanmaktadır.

$$APE' = \frac{1}{N} \sum_{i=1}^N \frac{|Y_i - \hat{Y}_i|}{Y_i} \cong 0.038 \quad (54)$$

olarak hesaplanmıştır.

LMS Parameter Estimates Based on LMS

Variable	Estimate	Approx Std Err	t Value	Pr >  t	Lower OCI	Upper OCI
VAR1	1.73887E-6	7.47245E-5	0.24	0.8123	-0.3003125	0.30031643
VAR2	-0.3005654	0.30053067	-1.15	0.2589	-0.3003347	0.30026352
VAR3	0.03406322	0.30457855	0.32	0.4196	-0.3053882	0.31532625
VAR4	-0.3020572	0.30925053	-0.22	0.8247	-0.3202173	0.31593033
VAR5	-0.016322	0.35025535	-0.32	0.7474	-0.1143211	0.38217705
VAR6	0.24586545	0.17565317	1.40	0.1710	-0.0964276	0.59016075
VAR7	-0.3047332	0.10565363	-0.34	0.5546	-0.2252222	0.20275585
VAR8	-1.5452326	2.24603775	-0.73	0.4590	-6.3474455	2.75636055
Intercept	7.63944255	2.1557173	3.55	0.0012	3.41824325	11.8636613

Bu tabloda,  $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \beta_8$  için bulunan  $b_0, b_1, b_2, b_3, b_4, b_5, b_6, b_7, b_8$  istatistikleri kullanılarak  $H_0 : \beta_0 = 0, H_0 : \beta_1 = 0, H_0 : \beta_2 = 0, H_0 : \beta_3 = 0, H_0 : \beta_4 = 0, H_0 : \beta_5 = 0, H_0 : \beta_6 = 0,$

Least Squares (LS) Method  
Minimizing 27th Ordered Squared Residual.  
Highest Possible Breakdown Value = 42.22 %

Estimated Coefficients				
VAR1	VAR2	VAR3	VAR4	VAR5
-0.000013172	-0.00055769	0.0145241195	0.0029542571	0.0209746301
Estimated Coefficients				
VAR6	VAR7	VAR8	Intercept	
0.4179147057	-0.74156355	-0.068717759	6.453497507	
LMS Objective Function = 0.190132302				
Preliminary LMS Scale = -0.2510012071				
Robust R Squared = 0.5219454454				
Final LMS Scale = 0.3201629512				

$H_0 : \beta_7 = 0, H_0 : \beta_8 = 0$  hipotezleri için sınama sonuçları görülmektedir.  $H_0 : \beta_0 = 0$  hipotezi için elde edilen  $|t|$  değeri dışında diğer hipotezler için elde edilen  $|t|$  değeri %5'den büyük çıkmıştır.

Least Trimmed Squares (LTS) Method  
Minimizing Sum of 27 Smallest Squared Residuals.  
Highest Possible Breakdown Value = 42.22 %

The best half of the entire data set obtained after full iteration consists of the cases:

2	4	5	6	7	9	13	14	16	17	18	20	22	24
25	30	31	32	35	36	37	38	41	42	43	44	45	

Estimated Coefficients				
VAR1	VAR2	VAR3	VAR4	VAR5
-0.000013172	-0.00055769	0.0145241195	0.0029542571	0.0209746301
Estimated Coefficients				
VAR6	VAR7	VAR8	Intercept	
0.4179147057	-0.74156355	-0.068717759	6.453497507	
LTS Objective Function = 0.190132302				
Preliminary LTS Scale = -0.2510012071				
Robust R Squared = 0.5219454454				
Final LTS Scale = 0.3201629512				

Sonuç olarak  $H_0 : \beta_0 = 0$  hipotezi %95 güven düzeyinde reddedilmiş, hesaplanan istatistiğin sıfır olamayacağı belirlenmiştir. Diğer hipotezler ise %95 güven düzeyinde reddedilememiş, hesaplanan istatistiklerin sıfır olduğu belirlenmiştir.  $LMS$  'ye dayalı yeniden ağırlıklandırılmış EKK Yöntemi ( $WLS$  or  $RLS$ ) ile anlamlı bir model kurulamamıştır.

Çıktıya baktığımızda oluşturulan regresyon denkleminin,

$$\hat{Y} = 11.93 - 3.885E - 7X_1 - 9.845E - 04X_2 - 0.01X_3 - 0.0154X_4 - 0.124X_5 + 0.218X_6 + 0.107X_7 - 3.98X_8$$

olduğunu ve Robut  $R^2$ 'nin 0.79 olduğunu görmekteyiz. Bu sonuca göre, bağımlı değişkendeki %79'lık değişim, modele dahil ettiğimiz bağımsız değişkenler tarafından açıklanmaktadır.

$$APE' = \frac{1}{N} \sum_{i=1}^N \frac{|Y_i - \hat{Y}_i|}{Y_i} \cong 0.033 \quad (55)$$

olarak hesaplanmıştır.

LMS Parameter Estimates Based on LTS

Variable	Estimate	Approx Std Err	t Value	Pr >  t	Lower CI	Upper CI
VAR1	1.091773-E	5.996158-E	0.18	0.8570	-0.0001197	0.0002284
VAR2	-0.0000000	0.00022715	-0.39	0.0023	-0.00012597	-0.00033391
VAR3	0.0070000	0.00000000	1.34	0.0000	-0.00000000	0.01666663
VAR4	-0.0009701	0.00000000	-1.42	0.0000	-0.00000000	-0.00000000
VAR5	-0.0770711	0.00000000	-2.13	0.0000	-0.00000000	-0.00000000
VAR6	0.2043151	0.00000000	1.55	0.0000	-0.00000000	0.00000000
VAR7	0.0000000	0.00000000	1.45	0.0000	-0.00000000	0.00000000
VAR8	-3.480452E	0.00000000	-2.14	0.0000	-0.00000000	-0.00000000
Intercept	10.0684114	0.00000000	6.30	<.0001	6.96000000	13.13682280

Bu tabloda,  $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \beta_8$  için bulunan  $b_0, b_1, b_2, b_3, b_4, b_5, b_6, b_7, b_8$  istatistikleri kullanılarak  $H_0 : \beta_0 = 0, H_0 : \beta_1 = 0, H_0 : \beta_2 = 0, H_0 : \beta_3 = 0, H_0 : \beta_4 = 0, H_0 : \beta_5 = 0, H_0 : \beta_6 = 0, H_0 : \beta_7 = 0, H_0 : \beta_8 = 0$  hipotezleri için sına sonuçları görülmektedir.  $H_0 : \beta_0 = 0, H_0 : \beta_2 = 0, H_0 : \beta_5 = 0$  ve  $H_0 : \beta_8 = 0$  hipotezleri için elde edilen  $|t|$  değerleri dışında diğer hipotezler için elde edilen  $|t|$  değerleri %5'den büyük çıkmıştır. Sonuç olarak  $H_0 : \beta_0 = 0, H_0 : \beta_2 = 0, H_0 : \beta_5 = 0$  ve  $H_0 : \beta_8 = 0$  hipotezleri %95 güven düzeyinde reddedilmiş, hesaplanan istatistiklerin sıfır olamayacağı belirlenmiştir. Diğer hipotezler ise %95 güven düzeyinde reddedilememiş, hesaplanan istatistiklerin sıfır olduğu belirlenmiştir.

Bağımlı değişken üzerinde anlamlı bir etkisi bulunmayan  $X_1, X_3, X_4, X_6, X_7$  değişkenleri modelden çıkarıldığında elde edilen sonuç aşağıdaki gibidir.

Veri seti üzerine, 3. Bölüm'de açıklanan  $LTS$  Kestirim Yöntemi ve  $LTS$  'ye dayalı yeniden ağırlıklandırılmış EKK Yöntemi ( $WLS$  or  $RLS$ ) SAS/IML programı ile yapılmış ve aşağıdaki sonuçlar bulunmuştur.

LMS Parameter Estimates Based on LTS

Variable	Estimate	Approx Std Err	t Value	Pr >  t	Lower CI	Upper CI
VAR2	-0.0002000	0.00000000	-1.05	0.0000	-0.00000000	0.00000000
VAR5	-0.0482147	0.00000000	-1.75	0.0000	-0.00000000	0.00000000
VAR8	0.97547500	0.00000000	1.29	0.0000	-0.00000000	0.00000000
Intercept	3.3021707	0.00000000	12.72	<.0001	0.00000000	0.00000000

Bu tabloda,  $\beta_0, \beta_2, \beta_5, \beta_8$  için bulunan  $b_0, b_2, b_5, b_8$  istatistikleri kullanılarak  $H_0 : \beta_0 = 0, H_0 : \beta_2 = 0, H_0 : \beta_5 = 0, H_0 : \beta_8 = 0$  hipotezleri için sına sonuçları görülmektedir.  $H_0 : \beta_0 = 0$  hipotezi için elde edilen  $|t|$  değeri dışında diğer hipotezler için elde edilen  $|t|$  değerleri %5'den büyük çıkmıştır. Sonuç olarak  $H_0 : \beta_0 = 0, H_0 : \beta_2 = 0, H_0 : \beta_5 = 0, H_0 : \beta_8 = 0$  hipotezleri için sına sonuçları görülmektedir.  $H_0 : \beta_0 = 0, H_0 : \beta_2 = 0, H_0 : \beta_5 = 0$  ve  $H_0 : \beta_8 = 0$  hipotezleri için elde edilen  $|t|$  değerleri dışında diğer hipotezler için elde edilen  $|t|$  değerleri %5'den büyük çıkmıştır. Sonuç olarak  $H_0 : \beta_0 = 0, H_0 : \beta_2 = 0, H_0 : \beta_5 = 0, H_0 : \beta_8 = 0$  hipotezleri %95 güven düzeyinde reddedilmiş, hesaplanan istatistiğin sıfır olamayacağı belirlenmiştir. Diğer hipotezler ise %95 güven düzeyinde reddedilememiş, hesaplanan istatistiklerin sıfır olduğu belirlenmiştir.  $LTS$  'ye dayalı yeniden ağırlıklandırılmış EKK Yöntemi ( $WLS$  or  $RLS$ ) ile anlamlı bir model kurulamamıştır.

Veri seti üzerine, 4. Bölüm'de açıklanan Tanaka Yöntemi hem Kesin  $Y$  değerleri için hem de Fuzzy (Bulanık) ( $Y, S$ ) değerleri için uygulanmış, çalışmada Lingo programı kullanılmış ve aşağıdaki sonuçlar bulunmuştur.

$$\hat{Y}_i = A_0 + A_1 X_1 + \dots + A_8 X_8, \quad i=1,2,\dots,45 \quad (56a)$$

$$\hat{Y}_i = (m_0, s_0) + (m_1, s_1)X_1 + \dots + (m_8, s_8)X_8, \quad i=1,2,\dots,45 \quad (56b)$$

Tabloya baktığımızda Kesin  $Y$  değerleri ile kurulan modelde  $X_2, X_3, X_4, X_7, X_8$  değişkenlerinin modelde olmadığı görülmektedir. Fuzzy  $Y$  değerleri ile kurulan modelde ise  $X_3, X_4, X_7, X_8$  değişkenlerinin modelde olmadığı görülmektedir.  $Y$  değerlerini fuzzy olarak ele aldığımızda  $X_2$  değişkeninin modele dahil olduğunu görmekteyiz.

**Tablo 4: Kesin ve Fuzzy Y için  $(m_j, s_j)$  'ler**

$h = 0.5$  için

Kesin Y için $(m_j, s_j)$	Fuzzy Y için $(m_j, s_j)$
$(7.36, 0) \Rightarrow A_0$	$(6.342, 0) \Rightarrow A_0$
$(0, 0.24E-04) \Rightarrow A_1$	$(0, 0.3801E-04) \Rightarrow A_1$
$(0.0165, 0) \Rightarrow A_5$	$(0.6985E-04, 0) \Rightarrow A_2$
$(0.01614, 0) \Rightarrow A_6$	$(0.0685, 0) \Rightarrow A_5$
	$(0.04496, 0) \Rightarrow A_6$

Kesin Y için üyelik fonksiyonları:

$$\mu_{A_0}(a_0) = \begin{cases} 1, & a_0 = 7,36 \\ 0, & \text{ö.d.} \end{cases} \quad (59a)$$

$$\mu_{A_1}(a_1) = \begin{cases} 1 - \frac{|a_1|}{0.24E-04}, & -0.24E-04 < a_1 < 0.24E-04 \\ 0, & \text{ö.d.} \end{cases} \quad (59b)$$

$$\mu_{A_5}(a_5) = \begin{cases} 1, & a_5 = 0.0165 \\ 0, & \text{ö.d.} \end{cases} \quad (59c)$$

$$\mu_{A_6}(a_6) = \begin{cases} 1, & a_6 = 0.01614 \\ 0, & \text{ö.d.} \end{cases} \quad (59d)$$

Kurulan modeller:

$$\text{Kesin Y için: } \tilde{Y}_i = (7.36, 0) + (0, 0.24E-04)X_1 + (0.0165, 0)X_5 + (0.01614, 0)X_6$$

$$\text{Fuzzy Y için: } \tilde{Y}_i = (6.342, 0) + (0, 0.3801E-04)X_1 + (0.6985E-04, 0)X_2 + (0.0685, 0)X_5 + (0.04496, 0)X_6$$

$$\text{Kesin Y için: } APE' = \frac{1}{N} \sum_{i=1}^N \frac{|Y_i - \hat{Y}_i|}{Y_i} \cong 0.037, \quad (57)$$

$$\text{Fuzzy Y için: } APE' = \frac{1}{N} \sum_{i=1}^N \frac{|Y_i - \hat{Y}_i|}{Y_i} \cong 0.045 \quad (58)$$

olarak hesaplanmıştır.

Veri seti üzerine, 4. Bölüm'de açıklanan Fuzzy Lineer Regresyon Analizine EKK Yaklaşımı Yöntemi uygulanmıştır. Çalışmada SAS/IML programı kullanılmış ve aşağıdaki sonuçlar bulunmuştur.

$$Y = (m, s), Y^* = (m^*, s^*) \quad (61a)$$

$$m = m^* + \varepsilon_c, m^* = Xa \quad (61b)$$

$$s = s^* + \varepsilon_r, s^* = m^*b + 1d \quad (61c)$$

$$\delta_i \cong \delta(Y_i, Y_i^*) = \sqrt{(c_i - c_i^*)^2 + (r_i - r_i^*)^2} \quad (61d)$$

$$\mu_{Y_i}(\hat{Y}_i) = \begin{cases} 1 - \frac{|\hat{Y}_i - Y_i|}{s_i}, & \hat{Y}_i - s_i < Y_i < \hat{Y}_i + s_i \\ 0, & \text{ö.d.} \end{cases} \quad (59e)$$

Fuzzy Y için üyelik fonksiyonları:

$$\mu_{A_0}(a_0) = \begin{cases} 1, & a_2 = 6.342 \\ 0, & \text{ö.d.} \end{cases} \quad (60a)$$

$$\mu_{A_1}(a_1) = \begin{cases} 1 - \frac{|a_1|}{0.3801E-04}, & -0.3801E-04 < a_1 < 0.3801E-04 \\ 0, & \text{ö.d.} \end{cases} \quad (60b)$$

$$\mu_{A_2}(a_2) = \begin{cases} 1, & a_2 = 0.6985E-04 \\ 0, & \text{ö.d.} \end{cases} \quad (60c)$$

$$\mu_{A_5}(a_5) = \begin{cases} 1, & a_5 = 0.0685 \\ 0, & \text{ö.d.} \end{cases} \quad (60d)$$

$$\mu_{A_6}(a_6) = \begin{cases} 1, & a_6 = 0.04496 \\ 0, & \text{ö.d.} \end{cases} \quad (60e)$$

$$\mu_{Y_i}(\hat{Y}_i) = \begin{cases} 1 - \frac{|\hat{Y}_i - Y_i|}{s_i}, & \hat{Y}_i - s_i < Y_i < \hat{Y}_i + s_i \\ 0, & \text{ö.d.} \end{cases} \quad (60f)$$

Amaç fonksiyonu şu şekilde tanımlanır:

$$\min \sum_{i=1}^n \delta_i^2 = (c - c^*)(c - c^*) + (r - r^*)(r - r^*) \quad (62)$$

Analizin sonucunda kestirilen katsayılar şunlardır:

**Tablo 6. Kestirilen  $a_j, b$  ve  $d$**

$a_j$
$10.34168 \Rightarrow a_0$
$-3.354E-6 \Rightarrow a_1$
$0.0000374 \Rightarrow a_2$
$0.0083946 \Rightarrow a_3$
$-0.012231 \Rightarrow a_4$ $b = -0.990758$ $d = 7.8380299$
$-0.066096 \Rightarrow a_5$
$0.2118132 \Rightarrow a_6$
$-0.0222 \Rightarrow a_7$
$-3.653391 \Rightarrow a_8$
İterasyon sayısı: 65086

$$APE' = \frac{1}{N} \sum_{i=1}^N \left| \frac{Y_i - \hat{Y}_i}{Y_i} \right| \cong 0.0335 \quad (63)$$

olarak hesaplanmıştır.

Veri seti üzerine 5. Bölüm'de açıklanan  $M$  - Kestiricisi Yöntemi'ne Dayanan Fuzzy Robust Regresyon Analizi Yöntemi ile hem Kesin  $Y$  değerleri için hem de Fuzzy  $Y$  değerleri için Fuzzy Robust Regresyon Analizi Lingo programı yardımıyla uygulanmıştır. Fuzzy  $Y$  değerleri ile yapılan analizde uygun bir sonuç bulunamamış, Kesin  $Y$  değerleri ile yapılan analizde ise, aşağıdaki sonuçlar bulunmuştur.

$$\hat{Y}_i = A_0 + A_1 X_1 + \dots + A_8 X_8 \quad (64a)$$

$$\hat{Y}_i = (a_{01}, a_{02}, a_{03}) + (a_{11}, a_{12}, a_{13})X_1 + \dots + (a_{81}, a_{82}, a_{83})X_8 \quad (64b)$$

Kesin  $Y$  için oluşturulan model:

$$\begin{aligned} \hat{Y}_i = & (0.928E-04, 0.943E-04, 0.943E-04)X_1 + (0.0127E-03, 0.127E-03)X_2 \\ & + (0.0406E-02, 0.406E-02)X_4 + (0.234, 0.234, 0.234)X_7 \\ & + (7.972, 7.972, 7.972)X_8 \end{aligned} \quad (65)$$

**Tablo 9:**  $A_2, A_3, A_5, A_6$  Analizden Çıkartıldığında Kesin  $Y$  için  $(a_{j1}, a_{j2}, a_{j3})$ 'ler

$h = 0$  için

Kesin  $Y$  için  $(a_{j1}, a_{j2}, a_{j3})$ :

$$(0.928E-04, 0.943E-04, 0.943E-04) \Rightarrow A_1$$

$$(0, 0.127E-03, 0.127E-03) \Rightarrow A_2$$

$$(0, 0.406E-02, 0.406E-02) \Rightarrow A_4$$

$$(0.234, 0.234, 0.234) \Rightarrow A_6$$

$$(7.972, 7.972, 7.972) \Rightarrow A_8$$

**Tablo 8:** Kesin  $Y$  için  $(a_{j1}, a_{j2}, a_{j3})$ 'ler

$h = 0$  için

Kesin  $Y$  için  $(a_{j1}, a_{j2}, a_{j3})$ :

$$(0, 0, 0) \Rightarrow A_0$$

$$(0.928E-04, 0.943E-04, 0.943E-04) \Rightarrow A_1$$

$$(0, 0.127E-03, 0.127E-03) \Rightarrow A_2$$

$$(0, 0, 0) \Rightarrow A_3$$

$$(0, 0.406E-02, 0.406E-02) \Rightarrow A_4$$

$$(0, 0, 0) \Rightarrow A_5$$

$$(0.234, 0.234, 0.234) \Rightarrow A_6$$

$$(0, 0, 0) \Rightarrow A_7$$

$$(7.972, 7.972, 7.972) \Rightarrow A_8$$

Üyelik fonksiyonları:

$$\mu_{A_1}(a) = \begin{cases} 1 & , a = 0.928E-04 \\ 0 & , \text{ö.d} \end{cases} \quad (66a)$$

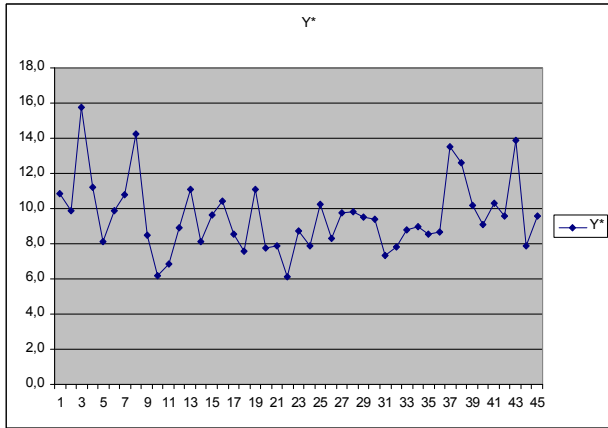
$$\mu_{A_2}(a) = \begin{cases} \frac{a}{0.127E-03} & , 0 \leq a \leq 0.127E-03 \\ 0 & , \text{ö.d} \end{cases} \quad (66b)$$

$$\mu_{A_4}(a) = \begin{cases} \frac{a}{0.406E-02} & , 0 \leq a \leq 0.406E-02 \\ 0 & , \text{ö.d} \end{cases} \quad (66c)$$

$$\mu_{A_6}(a) = \begin{cases} 1 & , a = 0.234 \\ 0 & , \text{ö.d} \end{cases} \quad (66d)$$

$$\mu_{A_8}(a) = \begin{cases} 1 & , a = 7.972 \\ 0 & , \text{ö.d} \end{cases} \quad (66e) \text{ ve}$$

$$\mu_{Y_i}(y) = \begin{cases} 0 & , y < 0.928E-04x_{i1} + 0.234x_{i6} + 7.972x_{i8} \\ 1 - \frac{8.21 - y}{0.15E-05x_{i1} + 0.127E-03x_{i2} + 0.406E-02x_{i4}} & , 0 \leq y \leq 0.127E-03x_{i2} + 0.406E-02x_{i4} \\ 0 & , y > 0.943E-04x_{i1} + 0.127E-03x_{i2} \\ & + 0.406E-02x_{i4} + 0.234x_{i6} + 7.972x_{i8} \end{cases} \quad (66f)$$



Şekil 10. Tahmin Edilen  $Y^*$ 'ların Görünümü

$Y^*$  değerlerini küçükten büyüğe doğru sıraladıktan sonra 1. ve 3. kartil değerleri:

$$Q_1 = \frac{n+1}{4} = \frac{45+1}{4} = 11.5'uncu\ de = 8.1 \quad (67a)$$

$$Q_3 = \frac{3n+1}{4} = \frac{3 \times 45 + 1}{4} = 34. de = 10.3 \quad (67b)$$

olarak hesaplanır. Bu değerlerden hareketle,

$$IQR(r_{mj}) = Q_3 - Q_1 = 10.3 - 8.1 = 2.2 \quad (67c)$$

olarak bulunur. Sınırı aşan değerler için puanlama yöntemi,

$$Q_1 - 1.5 \times IQR(r_{mj}) = 4.8\ veya\ Q_3 + 1.5 \times IQR(r_{mj}) = 13.6 \quad (67d)$$

şeklinde olacaktır. 4.8 değerinden daha düşük veya 13.6 değerinden daha büyük değerler düzensiz çıktı verisi olarak adlandırılır. Şekil 3.1'e veya Tablo 10'a baktığımızda, 15.7 değeriyle 3. tahmin değerinin, 14.3 değeriyle 8. tahmin değerinin ve 13.9 değeriyle 43. tahmin değerinin 13.6 değerini aştığını görmekteyiz. Bu tahmin değerlerinin düzensiz çıktı verisi olduğunu söyleyebiliriz. Bu değerler,

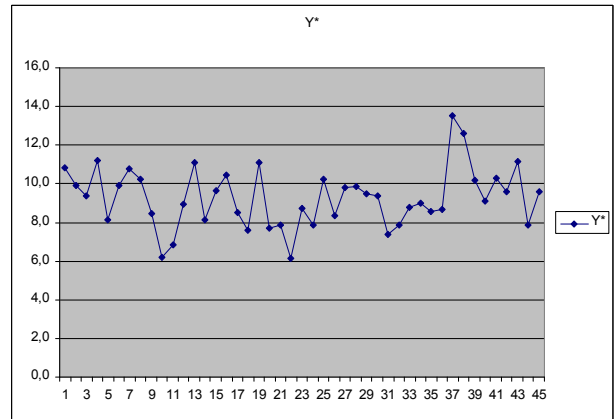
$$h_m = \frac{|y_i - y_{mi}|}{y_i} = \frac{|r_{mi}|}{y_{mi}}$$

işlemi ile elde edilen  $y_{mi}$  değeriyle değiştirilir.

Keyfi bir değer olan  $h_m = 0.45$  için  $y_{m3} = 9.35$ ,  $y_{m8} = 10.24$  ve  $y_{m43} = 11.17$  olarak hesaplanmıştır. Elde edilen son grafik Şekil 11'deki gibidir.

$$APE' = \frac{1}{N} \sum_{i=1}^N \frac{|Y_i - \hat{Y}_i|}{Y_i} \cong 0.23 \quad (68)$$

olarak hesaplanmıştır.



Şekil 11.  $y_{m3}$ ,  $y_{m8}$ ,  $y_{m43}$  ile Birlikte Tahmin Edilen  $Y^*$ 'ların Görünümü

## VI. SONUÇ

Bu verilerle, Klasik Regresyon Analizi'nde  $EKK$  yöntemiyle kestirilen parametreler ve  $LMS$ 'ye dayalı yeniden ağırlıklandırılmış  $EKK$  yöntemi ( $WLS$  or  $RLS$ ) 'nde kestirilen parametreler  $t$ -sınavasından geçememiştir.

Robust Regresyon Analizi'nde  $LMS$  kestirim yöntemiyle kurulan model testlerden geçmiş ancak belirginlik katsayısı  $R^2 = 0.52$  olarak düşük kalmıştır.

Fuzzy Regresyon modellerinden Tanaka'nın modeli hem kesin  $Y$  hem de fuzzy  $Y$  değerleri için uygun görülürken *İteratif EKK* yöntemiyle yapılan 65086 iterasyon sonucu bulunan model de uygun görülmektedir.

Fuzzy Robust Regresyon Analizi'nde uygun model bulunmakla birlikte Fuzzy Regresyon Analizi'nde bulunan sonuçların daha iyi olduğu gözlenmiştir.

Bu çalışma kapsamında Çoklu Regresyon Analizi'ne karşı diğer üç tekniğin (Robust Regresyon Analizi, Fuzzy Regresyon Analizi, Fuzzy Robust Regresyon Analizi) nasıl sonuç verdiği; tekniklerin birbirlerine karşı bir üstünlüklerinin olup olmadığını incelemiştir. Elde edilen sonuçlar, elimizdeki veri setinden yola çıkılarak elde edilmiştir. Başka bir veri setinde (örneğin aykırı ve/veya uç değerlerin olduğu bir veri seti) başka sonuçlar elde edilebilecek, yaptığımız analizde üstün gelmeyen bir teknik üstün gelebilecektir. Sonuç olarak bu tekniklerin birbirlerine karşı bir üstünlüklerinin olmadığına, elimizdeki veri setinin yapısına göre her bir tekniğin birbirine üstün gelebileceği saptanmıştır.



**Tablo 2: Dört Tekniğin Karşılaştırması**

	Regresyon		Robust Regresyon		Fuzzy Regresyon		Fuzzy Robust Regresyon
			LMS - RLS	LTS - RLS	Tanaka	İteratif EKK	M Yöntemi
APE'	----		0,038 - ----	0,033 - ----	Kesin Y 0,037 Fuzzy Y 0.045	0,0335	0,23
R <sup>2</sup>	----		0,52 - ----	0,79 - ----	----	----	----

**YARARLANILAN KAYNAKLAR**

- [1] Pearson K., Yule U., Blanchard N., Lee A. (1903) "The Law of Ancestral Heredity", Biometrika
- [2] Peters G., (1994) "Fuzzy Linear Regression with Fuzzy Intervals", Fuzzy Sets and Systems
- [3] Erar A., (2006) "Bağlanım (Regresyon) Çözümlemesi, Ders Notları...", MSGSÜ, İstatistik, s:73
- [4] Armutlulu İ. H., (2008) "İşletmelerde Uygulamalı İstatistik", Alfa Yayınları, 2.Baskı, s:224
- [5] Vural A., (2007) "Aykırı Değerlerin Regresyon Modellerine Etkileri ve Sağlam Kestiriciler", Yüksek Lisans Tezi, Marmara Üniversitesi, S.B.E., Ekonometri Ana Bilim Dalı, s:29
- [6] Hoaglin D. C., Mosteller F., Tukey J. W., (1983) "Understanding Robust and Explonatory Data Analysis", New York: John Wiley & Sons, s:339
- [7] Rousseeuw P. J., Leroy A. M., (2003) "Robust Regression and Outlier Detection", Jonh Wiley & Sons, s:17
- [8] Andersen R., (2008) "Modern Methods for Robust Regression", SAGE Publications, s:49, s:17
- [9] Fox J., (2002) "Robust Regression", Appendix to An R and S-PLUS Companion to Applied Regression, s:3
- [10] Watada J., Yabuuchi Y., (1994) "Fuzzy Robust Regression Analysis", Department of Industrial Management, Osaka Institute of Technology, 5-16-1 Omiya, Asahi, Osaka 535 Japan, 0-7803-1896-X/94 IEEE, s:1370-1371
- [11] Yücel L. İ., "Bulanık Regresyon : Türkiye'de 1980-2004 Döneminde Kayıt Dışı Ekonominin Bulanık Yöntemlerle Tahminine İlişkin Bir Uygulama", Yüksek Lisans Tezi, İstanbul Üniversitesi, S.B.E., Ekonometri Ana Bilim Dalı, 2005, s:48
- [12] Shapiro A. F., (2005) "Fuzzy Regression Models", Smeal College of Business, University Park, PA 16802, USA s:7-8
- [13] Uras Y., (1998) "Bulanık Mantığın Doğrusal Regresyon Analizinde Kullanılmasına İlişkin Bir Uygulama", Dokuz Eylül Üniversitesi, s:101-106
- [14] Başer F., (2007) "Aktüeryal Modellemede Melez Bulanık Regresyon Analizi", Yüksek Lisans Tezi, Ankara Üniversitesi, F.B.E., İstatistik Ana Bilim Dalı, s: 21
- [15] Shapiro A. F., (2005) "Fuzzy Regression Models", Smeal College of Business, University Park, PA 16802, USA s:14
- [16] Başer F., (2007) "Aktüeryal Modellemede Melez Bulanık Regresyon Analizi", Yüksek Lisans Tezi, Ankara Üniversitesi, F.B.E., İstatistik Ana Bilim Dalı, s: 23
- [17] D'Urso P., Gastaldi T., (2000) "A Least-Squares Approach to Fuzzy Linear Regression Analysis", Computational Statistics & Data Analysis 34, s:430
- [18] Sohn B. Y., (2005) "Robust Fuzzy Linear Regression Based on M-estimators", J. App1. Math. & Computing Vol:18, No:1-2, s:596-597
- [19] Cheng C., (2001) Lee E. S., "Fuzzy Regression with Radial Basis Function Network", Fuzzy Sets and Syatems 119, s:29

**EK 1: İteratif EKK Yaklaşımı için SAS/IML Yazılımı****proc iml;**

start line(coeff, intercept, minX, maxX, step, color);

x = do(minX, MaxX, step);

y = coeff # x + intercept;

call gdraw(x, y, , color);

finish;

do;

c = {7.35, 8.1, 6.45, 7.55, 7.45, 7.85, 8.2, 7.75, 7.95, 7.75, 7.65,  
7.35, 7.45, 7.8, 7.45, 7.4, 8.1, 7.35, 7.65, 7.6, 7.9, 7.4, 8.1, 7.5,  
7.55, 7.9, 7.2, 8.2, 7.95, 7.75, 7.3, 7.25, 7.8, 7.45, 7.2, 7.2, 8.25,  
8.15, 7.85, 7.85, 7.65, 8.1, 7.7, 7.7, 7.9};r = {0.35, 0.2, 1.85, 0.25, 0.25, 0.35, 0.1, 0.15, 0.25, 0.25, 0.35,  
0.15, 0.05, 0.2, 0.15, 0.3, 0.1, 0.35, 0.15, 0.1, 0.2, 0.3, 0.2, 0.2,  
0.05, 0.1, 0.1, 0.1, 0.15, 0.25, 0.1, 0.25, 0.2, 0.15, 0.1, 0.1, 0.25,  
0.35, 0.35, 0.15, 0.25, 0.4, 0.3, 0.2, 0.2};

Data = { 57997.768 1598 122 203 7.6 4.8 7.6 0.40,

44600 1290 124 185 11.9 5.9 9 0.43,

96173.64 1480 156 230 9.5 5.7 9.9 0.53,

60642.512 1390 122 203 8.9 4.4 6.8 0.43,

28170 1390 85 174 12.3 5.3 8.6 0.42,

.

.

.

75900 1987 158 185 11 6.5 9.6 0.53,

28370 1198 70 158 15.9 5.5 8.7 0.39,

42750 1591 125 178 12.2 5.8 9 0.41};

ObjNum = nrow(Data);

IndVarNum = ncol(Data);

One = J(ObjNum, 1);

X = One || Data;

N = X` \* X;

M = inv(N);

P = X` \* r;

threshold = 1e-8;

IterNum = 0;

a=(1: (IndVarNum+1))`;

b = 1;

d = 1;

do until (abs((a-a01d) // (b-b01d) // (d-d01d))

&lt; threshold);

IterNum = IterNum + 1;

a01d = a;

b01d = b;

d01d = d;

a = M \* (X` \* c + P # b - X` [, +] # b # d) / (1 + b ## 2);

b = (P` \* a - a` \* X` [, +] # d) / (a` \* N \* a);

d = (r[+] - a` \* X` [, +] # b) / ObjNum;

end;

cStar = X \* a;

rStar = cStar # b + One # d;

Ec = (c - cStar);

Er = (r - rStar);

Of = Ec[##] + Er[##];

rStarVar = (rStar - rStar[:])[##] / ObjNum;

reset noname;

print

"Doubly linear adaptive fuzzy regression model: Iterative solution",,

"Data", Data,, "a:" a,, "b:" b, "d:" d, "c:" c, "cStar:" cStar,,

"Sum Ec" (Ec[+]), "r:" r "Sum Er" (Er[+]), "rStar:" rStar,,

"rStar variance:" rStarVar,, "Objective function:" Of,,

"Number of iterations:" IterNum;

/\*if IndVarNum = 1 then do;

call gstart;

minX = min(Data);

maxX = max(Data);

minY = min(C - r, cStar - rStar);

maxY = max(C + r, cStar + rStar);

rangeX = maxX - minX;

rangeY = maxY - minY;

XMarginFactor = 1 / 3;

YMarginFactor = 1 / 3;

marginX = rangeX # XMarginFactor;

marginY = rangeY # YMarginFactor;

WminX = (minX - marginX);

WmaxX = (maxX + marginX);

WminY = (minY - marginY);

WmaxY = (maxY + marginY);

call gwindow(WminX, WminY, WmaxX, WmaxY);\*/

call gpoint(Data, C, "black");

step = 1E-2;

**run** line(a[2], a[1], minX-marginX, maxX+marginX, step, "green");**run** line(a[2] # (1 + b), a[1] # (1 + b) + d, minX-marginX, maxX+marginX, step, "Blue");**run** line(a[2] # (1 - b), a[1] # (1 - b) - d, minX-marginX, maxX+marginX, step, "Blue");

call gdrawl(Data || (C - r), Data || (C + r), );

call gXaxis(0 || 0, rangeX + marginX, 5);

call gYaxis(0 || 0, rangeY + marginY, 5);

call gshow;

call gstop;

end;

end;

**quit;**

**EK 2: Analizde Kullanılan Otomobil Markaları**

n	Marka
1	MINI Cooper
2	Toyota Corolla 1.6 A/T
3	Mercedes-Benz C 180 BlueEfficiency
4	Audi A1 1.4 TFSI S tronic
5	Skodia Fabia 1.4
6	Citroen DS3 1.6 Vti 120 HP
7	Alfa Romeo Giulietta 1.4 TB
8	Opel Insignia 2.0 CDTi
9	Chevrolet Cruze 1.6
10	Hyundai i10 1.1 Team A/T
11	Chevrolet Spark 1.2 LT
12	Fiat Bravo 1.4 T-Jet
13	Citroen C4 1.6 THP MCP
14	Hyundai Accent Blue 1.4
15	Opel Astra 1.4 HB Turbo
16	Peugeot 508 1.6 Vti Auto 6R
17	Seat Leon 1.2 TSI
18	Nissan Micra 1.2 CVT
19	Volkswagen Passat 1.4 TSI DGS
20	Fiat Punto Evo 1.4 Fire
21	Kia cee'd 1.6 CRDi
22	Tata Vista 1.3 CRDi Quadrojet
23	Toyota Corolla 1.4 D-4D
24	Volkswagen Polo 1.6 TDI
25	Citroen C5 1.6 Hdi
26	Skoda Octavia 1.6 TDI
27	Renault Latitude 1.5 dCi
28	Alfa Romeo Giulietta 1.6 JTD-2
29	Volkswagen Jetta 1.6 TDI DSG
30	Volkswagen Golf 1.6 TDI DSG
31	Seat Ibiza Sportourer 1.6 TDI
32	Citroen C3 1.6 Hdi
33	Renault Fluence 1.5 dCi EDC
34	Opel Astra HB 1.3 CDTi
35	Renault Megane HB dCi EDC
36	Renault Megane HB 1.5 dCi
37	Peugeot RCZ 1.6 HP
38	Volkswagen Scirocco 1.4 TSI DGS
39	Citroen DS3 1.6 THP
40	Skoda Yeti 1.2 TSI
41	Peugeot 3008 1.6 VTI
42	Dacia Duster 1.6 4x4
43	Toyota RAV4 2.0 Valvematic CVT
44	Skoda Roomster 1.2
45	Kia Venga 1.6 CVVT A/T



**İsmail Hakkı ARMUTLULU**

[ismailhakki@marmara.edu.tr](mailto:ismailhakki@marmara.edu.tr)

He is a professor of Faculty of Economic and Administrative Sciences at Marmara University. He is also head of Department of Quantitative Methods. His research areas include Statistics, Quantitative Methods, Fuzzy Mathematics, Forecasting and History of Science.



**Murat YAZICI**

[asiscrus.muratyazici@gmail.com](mailto:asiscrus.muratyazici@gmail.com)

He graduated from Marmara University's Institute Social Sciences, M.Sc. in Business Administration / Quantitative Methods, in 2011. He graduated from Mimar Sinan Fine Arts University's Faculty of Arts and Sciences, B.Sc. in Statistics, in 2007. He graduated from Haydar Paşa Anatolian High School's Science Programme, in 2003. His interest areas include Applied Mathematics, Fuzzy Probability and Statistics, Robust Statistics, Mathematical and Statistical Modelling, Computational Statistics.

