

## BULANIK YEREL POLİNOMİYAL PARAMETRİK OLMAYAN REGRESYON MODELLERİ

Memmedağa MEMMEDLİ\* Münevvere YILDIZ\*\* Özer ÖZDEMİR\*\*\*

### ÖZET

*Bu çalışmada parametrik olmayan bulanık regresyon modelleri göz önüne alınmıştır. Mevcut bulanık regresyon modelleri, bulanık versiyonu geliştirilen parametrik olmayan yerel kübik regresyon modeli ile karşılaştırılmıştır. Bu yeni yaklaşımda, önerilen modelin performansının diğerlerine göre daha iyi sonuçlara sahip olduğu bulunmuştur.*

**Anahtar Kelimeler:** Bulanık parametrik olmayan regresyon, LR bulanık sayısı, Yerel polinomiyal düzeltme.

### 1. GİRİŞ

Kesin durumlara ait bilginin yer aldığı çalışmalarda klasik regresyon modelleri yaygın olarak çeşitli alanlarda kullanılmaktadır. Ancak bazı veri gruplarını bu modellerle ifade etmek mümkün olamamaktadır. Bununla birlikte, bazen de bulanık yapıya sahip verilerin kullanılmasının gerekliliği, bu klasik regresyon modellerinin kullanımını olanaksız hale getirmektedir. Bu durumda, belirsizliğin olduğu ve klasik regresyon modellerinin yetersiz kaldığı durumlarda yeni arayışlar başlamış ve bulanık regresyon modelleri ileri sürülmüştür.

Bulanık regresyon modelleri konusunda ilk çalışmalar Tanaka vd. (1982) ile başlamıştır ve bu çalışmaların çoğu bulanık parametrik regresyon modelleri üzerine odaklanmıştır. Bulanık regresyon modelleri için farklı yaklaşımlar ileri sürülerek; matematiksel programlama temelli yöntemler (Tanaka vd., 1982; Nasrabadi ve Nasrabadi, 2004), en küçük kareler yöntemi (Diamond, 1988; Hong vd., 2004) ve sinir ağı yaklaşımları (Cheng ve Lee, 2001) gibi çeşitli yöntemler kullanılmıştır.

Cheng ve Lee (2001) bulanık regresyon analizinde, radyal tabanlı fonksiyonları kullanan sinir ağları üzerine çalışma yapmışlardır. Bulanık parametrik regresyon analizi konusunda çeşitli pratik çalışmalar da literatürde yer almaktadır (Hong vd., 2004; Kao ve Chyu, 2002; Sanchez ve Gomez, 2003). Pratik çalışmalarda yanıt değişkeni ile açıklayıcı değişkenler arasındaki ilişkinin belirli bir parametrik modelle ifade edilmesi çok kısıtlayıcı olmaktadır. Ayrıca yanlış sonuçlara da neden olabilmektedir. Bu açıdan, son yıllarda parametrik olmayan regresyon modellerine olan ilgi giderek artmıştır (Fan ve Gijbels, 1996; Loadar, 1999).

Ishibuchi ve Tanaka (1992), geri yayılım sinir ağı algoritmasını kullanarak, çeşitli parametrik olmayan bulanık regresyon modellerini önermişlerdir. Parametrik olmayan regresyon modellerine ilginin giderek artmasıyla bu modellerin bulanık versiyonları da geliştirilmeye başlanmıştır (Fan ve Gijbels, 1996; Wood, 2006). Cheng ve Lee (1999) k-

\*Prof. Dr., Anadolu Üniversitesi, Fen Fakültesi, İstatistik Bölümü, e-posta: [mmammadov@anadolu.edu.tr](mailto:mmammadov@anadolu.edu.tr)

\*\*Anadolu Üniversitesi, Fen Fakültesi, İstatistik Bölümü, e-posta: [munevvere@hotmail.com](mailto:munevvere@hotmail.com)

\*\*\*Öğr. Gör., Anadolu Üniversitesi, Fen Fakültesi, İstatistik Bölümü, e-posta: [ozerozdemir@anadolu.edu.tr](mailto:ozerozdemir@anadolu.edu.tr)

en yakın komşu ve çekirdek tahminleme yöntemlerinin bulanık versiyonlarını incelemiştir. Wang vd. (2007) bulanık yerel doğrusal regresyon modelini ele almışlar ve deneysel çalışmalar ile bu yaklaşımın (Cheng ve Lee, 1999)'deki bulanık modellere göre avantajlarını göstermişlerdir. Bu çalışmalarda düzeltme parametresinin seçimi için çapraz-geçerlilik kriteri kullanılmıştır. Ancak, Fan ve Gijbels (1996) yerel regresyon uyumlaması (fitting) için daha etkin seçim yöntemleri önermişlerdir.

Klasik teoride parametrik olmayan regresyon modelleri üzerine yapılmış pek çok çalışmanın henüz bulanık versiyonları geliştirilmemiştir, bu açıdan parametrik olmayan regresyon modelleri teorisinde kullanılan yöntemlerin bulanık versiyonlarının geliştirilmesi önemli bir konudur.

Çalışmada parametrik olmayan bulanık yerel polinomial regresyon modelleri, bulanık yerel doğrusal regresyon modeli ve literatürde bulanık versiyonu üzerine çalışma yapılmamış olan bulanık yerel kübik regresyon modelleri incelenmiştir. Gerçekleştirilen simülasyon çalışması ile farklı çekirdek fonksiyonları için uygun düzeltme parametresi belirlenerek, bu modellerin performansları ASE (ortalama hata kareler) performans kriteri yardımıyla değerlendirilmiştir. Bulanık yerel kübik regresyon modelinin, bulanık yerel doğrusal regresyon modeline göre daha iyi performans gösterdiği gözlemlenmiştir.

## 2. BULANIK PARAMETRİK OLMAYAN REGRESYON MODELİ İÇİN YEREL POLİNOMİYAL YÖNTEM

Tek değişkenli bulanık parametrik olmayan regresyon modeli eşitlik 1'deki gibi ele alınır:

$$y = F(x) \{+\} \varepsilon \quad (1)$$

Burada,

$$F(x) = (m(x), \alpha(x), \beta(x)) \quad (2)$$

kesin  $\mathcal{D}$  kümesinde tanımlanmış  $LR$  bulanık değerler alan bulanık fonksiyon,  $\varepsilon$  bulanık hata terimi,  $\{+\}$  - kullanılan bulanık sıralama yöntemine bağlı olarak tanımlanan operatör'dür (Cheng ve Lee, 1999; Wang vd., 2007).

$(x_i, y_i)$ 'lerin ( $i = 1, 2, \dots, n$ ), kesin girdilere ve  $LR$  bulanık çıktılara sahip eşitlik 1 modeli için gözlem verileri oldukları varsayımı altında, her bir  $y$  çıktısı,  $c_y$ - merkez,  $\alpha_y$  ve  $\beta_y$  sol ve sağ yayılım bileşenleri ile  $(c_y, \alpha_y, \beta_y)$  şeklinde bir  $LR$  bulanık sayıdır. Bu sayı  $l_y = c_y - \alpha_y$  ve  $u_y = c_y + \beta_y$  alınarak, sol uç, merkez ve sağ uç noktaları yardımıyla  $(l_y, c_y, u_y)$  şeklinde yazılabilir.  $l(x), c(x), u(x)$  fonksiyonlarının ise sırasıyla  $LR$  - fonksiyonun sol ucu, merkezi ve sağ ucu olarak,  $\mathcal{D}$  tanım kümesinde  $p$ . mertebeye kadar sürekli türeve sahip olmasıyla, bu fonksiyonlar verilen bir  $x_0 \in \mathcal{D}$  noktasının komşuluğunda yaklaşık olarak  $p$ . mertebeden Taylor polinomu olarak ele alınırlar ve aşağıdaki gibi yazılırlar;

$$l(x) \approx l(x_0) + l'(x_0)(x - x_0) + \dots + \frac{l^{(p)}(x_0)}{p!} (x - x_0)^p \quad (3)$$

$$c(x) \approx c(x_0) + c'(x_0)(x - x_0) + \dots + \frac{c^{(p)}(x_0)}{p!} (x - x_0)^p \quad (4)$$

$$u(x) \approx u(x_0) + u'(x_0)(x - x_0) + \dots + \frac{u^{(p)}(x_0)}{p!}(x - x_0)^p \quad (5)$$

Diamond uzaklığı (Diamond, 1988) kullanılarak, belirli bir  $K(\cdot)$  çekirdek fonksiyonu yardımıyla ağırlıklı yerel en küçük kareler yöntemi yardımıyla eşitlik 3-4-5'teki yerel tahminler elde edilir:

$$(\hat{l}(x_0), \hat{l}'(x_0), \dots, \hat{l}^{(p)}(x_0))^T = \left( X^T(x_0)W(x_0; h)X(x_0) \right)^{-1} X^T(x_0)W(x_0; h)l_y \quad (6)$$

$$(\hat{c}(x_0), \hat{c}'(x_0), \dots, \hat{c}^{(p)}(x_0))^T = \left( X^T(x_0)W(x_0; h)X(x_0) \right)^{-1} X^T(x_0)W(x_0; h)c_y \quad (7)$$

$$(\hat{u}(x_0), \hat{u}'(x_0), \dots, \hat{u}^{(p)}(x_0))^T = \left( X^T(x_0)W(x_0; h)X(x_0) \right)^{-1} X^T(x_0)W(x_0; h)u_y \quad (8)$$

Burada;

$$X(x_0) = \begin{pmatrix} \mathbf{1} & (x_1 - x_0) & (x_1 - x_0)^2 & \dots & (x_1 - x_0)^p \\ \dots & \dots & \dots & \dots & \dots \\ \mathbf{1} & (x_n - x_0) & (x_n - x_0)^2 & \dots & (x_n - x_0)^p \end{pmatrix}, \quad (9)$$

$$l_y = \begin{pmatrix} l_{y_1} \\ \vdots \\ l_{y_n} \end{pmatrix}, c_y = \begin{pmatrix} c_{y_1} \\ \vdots \\ c_{y_n} \end{pmatrix}, u_y = \begin{pmatrix} u_{y_1} \\ \vdots \\ u_{y_n} \end{pmatrix} \quad (10)$$

$$W(x_0; h) = \text{Diag}(K_h(|x_1 - x_0|), \dots, K_h(|x_n - x_0|)) \quad (11)$$

$$K_h(z) = \frac{1}{h} K\left(\frac{z}{h}\right) \quad (12)$$

olmak üzere  $n \times n$  boyutunda köşegen ağırlık matrisidir. Gauss çekirdek fonksiyonu  $K(\cdot)$  aşağıdaki gibi yazılır;

$$K(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \quad (13)$$

Epanechnikov çekirdek fonksiyonu kullanılması durumunda,

$$K(x) = \begin{cases} 0.75(1 - x^2), & |x| < 1 \\ \mathbf{0} & , \quad d.d \end{cases} \quad (14)$$

şeklinde ifade edilir.

Çalışmada eşitlik 1'deki model için, polinomun mertebeleri  $p = 1$  ve  $p = 3$  alınarak, eşitlik 3-4-5'teki uygun formüller yardımıyla bulanık yerel doğrusal ve bulanık yerel kübik regresyon modelleri incelenmiştir. Bu amaçla bir sonraki bölümde yer alan simülasyon çalışması yapılmıştır.

### 3. SİMÜLASYON ÇALIŞMASI

Simülasyon çalışması; bulanık yerel doğrusal ve bulanık yerel kübik modelleri için uygun düzeltme parametresinin seçiminde çapraz geçerlilik (CV) kriterini kullanarak, bu modellerin performanslarını karşılaştırma amacıyla gerçekleştirilmiştir. Cheng ve Lee (1999) çalışmasındaki fonksiyonlar kullanılarak, boyutları 100 olan veri setleri türetilmiştir ve simülasyon denemelerinde 100 tekrarlıma gerçekleştirilmiştir. Bulanık yanıt çıktıları, merkez ve sağ-sol saçılımları Örnek-1 ve Örnek-2'deki gibi türetilen, simetrik üçgen bulanık sayılardır. Simülasyon çalışmasında gürültü, merkez için  $[-0.5, 0.5]$ , simetrik saçılımlar için ise  $[-0.25, 0.25]$  aralığında düzgün dağılımdan elde edilmektedir.

**Örnek-1.**  $g_1(x)$ ,  $[0,10]$  aralığında aşağıdaki gibi tanımlanmış bir fonksiyon olsun:

$$g_1(x) = \frac{1}{5}x^2 + 2\exp\left(\frac{x}{10}\right) \quad (15)$$

$[0,10]$  aralığındaki  $x_i = 0.1i$  ( $i = 1,2, \dots, 100$ ) noktaları için,

$$y_i = g_1(x_i) + \mathit{rand}[-0.5, 0.5] \quad (16)$$

$$\sigma_i = \frac{1}{4}g_1(x_i) + \mathit{rand}[-0.25, 0.25] \quad (17)$$

olarak hesaplanırlar. Burada  $\mathit{rand}[a_1, a_2]$ , her bir  $i$  için  $[a_1, a_2]$  aralığı üzerinde düzgün dağılımdan bağımsız olarak türetilen bir rassal sayıyı göstermektedir. Gözlenen bulanık çıktıların simetrik üçgen bulanık sayılar olduğu varsayılarak, bu çıktılara ait gösterim aşağıdaki gibi tanımlanır.

$$Y_i = (l_{y_i}, c_{y_i}, u_{y_i})_T = (y_i - \sigma_i, y_i, y_i + \sigma_i)_T \quad i = 1, 2, \dots, 100 \quad (18)$$

**Örnek-2.**  $[0,10]$  aralığında ikinci fonksiyon aşağıdaki gibi tanımlanmıştır:

$$g_2(x_i) = 10 + 5\sin(0.025\pi(1 - x)^2) \quad (19)$$

$[0,10]$  aralığındaki aynı  $x_i = 0.1i$  ( $i = 1,2, \dots, 100$ ) noktaları için

$$y_i = g_2(x_i) + \mathit{rand}[-0.5, 0.5] \quad (20)$$

$$\sigma_i = \frac{1}{3}g_2(x_i) + \mathit{rand}[-0.25, 0.25] \quad (21)$$

olarak hesaplanırlar. Bulanık çıktılar;

$$Y_i = (l_{y_i}, c_{y_i}, u_{y_i})_T = (y_i - \sigma_i, y_i, y_i + \sigma_i)_T \quad i = 1, 2, \dots, 100 \quad (22)$$

olarak belirlenirler. Çalışmada bu fonksiyonların yardımıyla türetilen iki veri seti için bulanık yerel doğrusal ve bulanık yerel kübik regresyon modelleri incelenmiştir. Her iki modele uygun düzeltme parametresinin (h) optimum değerinin seçimi için çapraz geçerlilik (CV) kriteri kullanılmıştır.

Ele alınan bulanık regresyon problemleri için CV kriteri;

$$\begin{aligned}
 CV(\mathbf{h}) &= \frac{1}{n} \sum_{i=1}^n d^2(y_i, \hat{F}(x_i, \mathbf{h})) \\
 &= \frac{1}{n} \sum_{i=1}^n [(l_{y_i} - \hat{l}_{(i)}(x_i, \mathbf{h}))^2 + (c_{y_i} - \hat{c}_{(i)}(x_i, \mathbf{h}))^2 + (u_{y_i} - \hat{u}_{(i)}(x_i, \mathbf{h}))^2] \quad (23)
 \end{aligned}$$

formülüyle hesaplanır. Burada  $(i)$  indisi tahminleme yapılırken  $i$ . gözlem değerinin çıkarıldığını göstermektedir. Optimum  $h_0$  düzeltme parametresi,  $CV(h_0) = \min_h CV(h)$  minimum probleminin çözümü olarak elde edilir.

Simülasyon hesaplamalarında rassallığın etkisini azaltmak için 100 tekrarlar yapılmıştır. Her iki veri seti ile yapılan uyumlamalarda ağırlık dizilerinin oluşturulmasında Gauss ve Epanechnikov çekirdekleri (kernel) kullanılmıştır. Modellerin performansları onların ASE değerleri yardımıyla karşılaştırılmıştır:

$$\begin{aligned}
 ASE(\mathbf{h}) &= \frac{1}{n} \sum_{i=1}^n d^2(F(x_i), \hat{F}(x_i)) \\
 &= \frac{1}{n} \sum_{i=1}^n [(l(x_i) - \hat{l}(x_i))^2 + (c(x_i) - \hat{c}(x_i))^2 + (u(x_i) - \hat{u}(x_i))^2] \quad (24)
 \end{aligned}$$

Tablo 1’de Örnek-1 için bulanık yerel doğrusal (LLS) ve kübik regresyon modellerinin (LCS), CV kriteri yardımıyla optimum düzeltme parametreleri seçilerek, Gauss ve Epanechnikov çekirdeklerine uygun ASE değerleri verilmiştir.

**Tablo 1. Örnek-1 için simülasyon sonuçları**

Yöntem	Çekirdek	Düzeltilme Parametresi (h)	ASE
LLS	Gauss	0,51	0,2642107
	Epanechnikov	1,20	0,2739353
LCS	Gauss	1,10	0,2616963
	Epanechnikov	1,25	0,2496074

Benzer şekilde Tablo 2’de Örnek-2 için bulanık yerel doğrusal ve kübik regresyon modellerinin, CV kriteri yardımıyla optimum düzeltme parametreleri seçilerek, Gauss ve Epanechnikov çekirdeklerine uygun ASE değerleri verilmiştir.

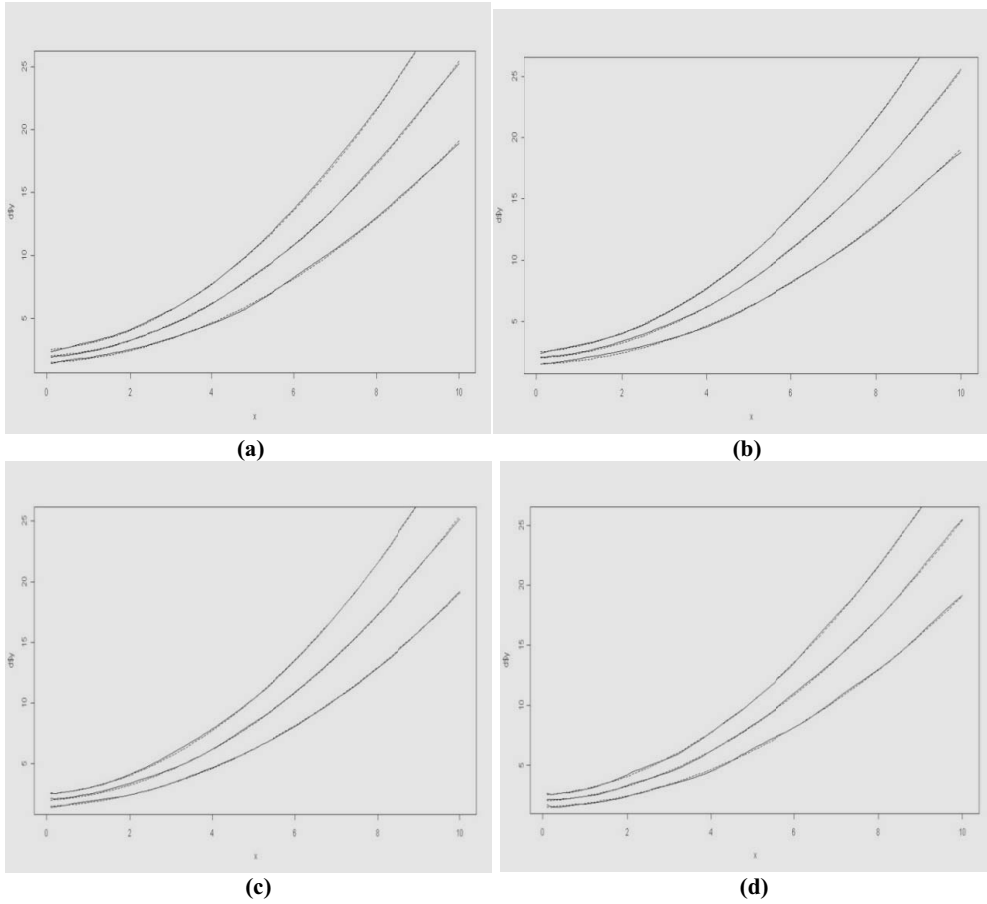
**Tablo 2. Örnek-2 için simülasyon sonuçları**

Yöntem	Çekirdek	Düzeltilme Parametresi (h)	ASE
LLS	Gauss	0,21	0,2301000
	Epanechnikov	0,52	0,2573107
LCS	Gauss	0,40	0,2281975
	Epanechnikov	1,30	0,2483770

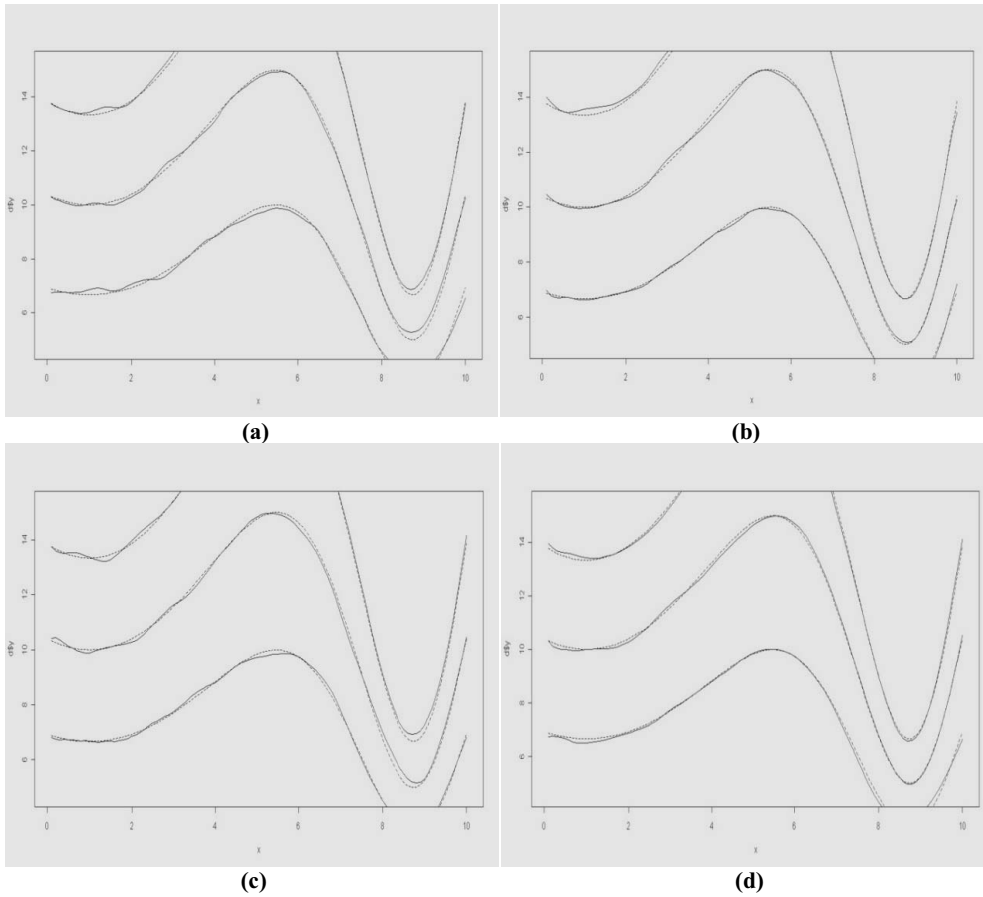
Tablo 1 ve Tablo 2’de LLS ve LCS modelleri için uygun çekirdek fonksiyonu ve düzeltme parametresine karşılık gelen performans kriteri olan ASE değerleri görülmektedir. Tablolar incelendiğinde bulanık yerel kübik modelin performansı

çekirdek fonksiyonları açısından daha iyi sonuçlar vermektedir. Tablolarda yer alan bilgilerin grafiksel gösterimleri Şekil 1 ve Şekil 2’de verilmiştir.

Şekil 1 ve Şekil 2’de yer alan grafikler incelendiğinde, bulanık yerel kübik regresyon modeline ait grafiklerde, merkez, sağ – sol saçılım eğrilerinin gerçek bulanık regresyon fonksiyonu eğrisi ve onun tahminlerini gösteren eğri arasındaki sapmanın, bulanık yerel doğrusal regresyon modeline göre daha az olduğu görülmektedir. Şekil 1 ve Şekil 2’de, Örnek-1 ve Örnek-2 için ilgili grafikler verilmiştir.



Şekil 1. Örnek-1 için Gauss çekirdeği kullanıldığında: (a)-  $h=0.51$  ve  $p=1$ , (b)-  $h=1.1$  ve  $p=3$   
Epanechnikov çekirdeği kullanıldığında: (c)-  $h=1.20$  ve  $p=1$ , (d)-  $h=1.25$  ve  $p=3$



Şekil 2. Örnek-2 için Gauss çekirdeği kullanıldığında: (a)-  $h=0.21$  ve  $p=1$ , (b)-  $h=0.4$  ve  $p=3$   
Epanechnikov çekirdeği kullanıldığında: (c)-  $h=0.52$  ve  $p=1$ , (d)-  $h=1.3$  ve  $p=3$

#### 4. TARTIŞMA VE SONUÇ

Parametrik olmayan yerel polinomial regresyon problemlerinde yerel polinomun mertebesi ile ilişkili olarak düzeltme parametresinin ( $h$ ) seçimi oldukça önemli bir rol oynar. Bu modellerde polinomun mertebesi arttıkça yerel düzeltme parametresinin genişliği de artar.

Bulanık yerel polinomial regresyon modelleri incelenirken, genel teknik olarak ağırlıklı en küçük kareler yöntemi, bulanık uzaklık olarak ise Diamond uzaklığı kullanılmıştır. Yapılan simülasyon çalışması ile öncelikle bulanık yerel doğrusal ve bulanık yerel kübik regresyon modelleri için çapraz geçerlilik kriteri yardımıyla uygun düzeltme parametresi seçilmiştir. Gauss ve Epanechnikov çekirdek fonksiyonları modellerin uyumlanmasında kullanılmıştır. Tüm bu kriterler göz önüne alınarak bulanık yerel doğrusal ve bulanık yerel kübik regresyon modelleri karşılaştırılmıştır. Bulanık yerel kübik regresyonun, bulanık yerel doğrusal regresyon modeline göre her iki çekirdek fonksiyonu içinde  $ASE$  değerleri bakımından daha iyi performans gösterdiği gözlenmiştir.

Sonuçta, polinomun mertebesi artarken düzeltme parametresinin genişliği de artmaktadır. Bu durum, işlemsel olarak yerel uyumlama noktalarının azalmasından dolayı işlem basamaklarını azaltmakta ve zamandan kazanç sağlanmaktadır. Özellikle daha eğrisel olan veri yapılarında yerel doğrusal modellerin kullanılması çok uygun olmamaktadır. Çünkü bu tür durumlarda düzeltme parametresinin değeri oldukça küçülmekte ve işlem karmaşasına yol açmaktadır.

Sonraki çalışmalarda düzeltme parametresinin seçimi için farklı kriterlerin bulanık versiyonlarının belirlenmesi, modelin bu parametreye bağlı olarak esnekliğinin incelenmesi merak uyandırmaktadır. Ayrıca, diğer parametrik olmayan regresyon modellerinin bulanık versiyonları üzerine çalışmaların yapılması oldukça önemlidir.

## 5. KAYNAKLAR

Cheng, C. B., Lee E. S., 1999. Nonparametric Fuzzy Regression – k-NN and Kernel Smoothing Techniques, *Computers and Mathematical with Applications*, 38, 239-251.

Cheng, C. B., Lee E. S., 2001. Fuzzy Regression with Radial Basis Function Networks, *Fuzzy Sets and Systems*, 119, 291-301.

Diamond, P., 1988. Fuzzy Least Squares, *Information Sciences* 46, 141-157.

Fan, J., Gijbels, I., 1996. *Local Polynomial Modelling and its Applications*, Chapman & Hall/CRC.

Hong, H., Song, J. K., Do, H., 2001. Fuzzy Least Squares Linear Regression Analysis Using Shape Preserving Operations, *Information Sciences* 138, 185-193.

Hong, Y. Y., Chao, Z. T., Yang, M. S. (2004). A Fuzzy Multiple Linear Regression Based Loss Formula in Electric Distribution Systems, *Fuzzy Sets and Systems* 142, 293-306.

Ishibuchi, H., Tanaka, H. (1992). Fuzzy Regression Analysis Using Neural Networks, *Fuzzy Sets and Systems* 50, 257-265.

Kao, C., Chyu, C. L., 2002. A Fuzzy Linear Regression Model with Better Explanatory Power, *Fuzzy Sets and Systems* 126, 401-409.

Loader, C., 1999. *Local Regression and Likelihood*, Springer.

Nasrabadi, M., Nasrabadi, E., 2004. A Mathematical-Programming Approach to Fuzzy Linear Regression Analysis, *Applied Mathem. and Computation* 155, 673-688.

Sanchez, J. D., Gomez, A. T., 2003. Applications of Fuzzy Regression in Actuarial Analysis, *Journal of Risk Insurance* 70, 665-699.

Tanaka, H., Uejima, S., Asai, K., 1982. Linear Regression Analysis with Fuzzy Model, *IEEE Transactions on Systems, Man. and Cybernetics* 12, 903-907.



Wang, N., Zhang, W. X., Mei, C. L., 2007. Fuzzy Nonparametric Regression Based on Local Linear Smoothing Technique, An International Journal Information Sciences, 177, 3882-3900.

Wood, S. N., 2006. Generalized Additive Models (An Introduction with R), Chapman & Hall/CRC.

## FUZZY LOCAL POLYNOMIAL NONPARAMETRIC REGRESSION MODELS

### ABSTRACT

*In this study, nonparametric fuzzy regression models are considered. Existing fuzzy regression models are compared with improved fuzzy version of local nonparametric cubic regression model. In this new approach, it is found that performance of this model has better results according to others.*

**Keywords:** Fuzzy nonparametric regression, LR Fuzzy number, Local polynomial smoothing.