ANADOLU ÜNİVERSİTESİ BİLİM VE TEKNOLOJİ DERGİSİ
ANADOLU UNIVERSITY JOURNAL OF SCIENCE AND TECHNOLOGY
Cilt/Vol.:8-Sayı/No: 1 : 41-52 (2007)

ARAŞTIRMA MAKALESİ /**RESEARCH ARTICLE**

YARI PARAMETRİK MODELLERDE SPLAYN DÜZELTME İLE TAHMİN VE ÇIKARSAMALAR

Dursun AYDIN[1], Mammadagha MAMMADOV, Ali Faut YÜZER, Embiya AĞAOĞLU

ÖZ

Yarı parametrik regresyon modelleri için splayn düzeltme metoduna dayalı yeni bir tahmin kavramı Eubank vd. (1998) tarafından tanıtılmıştır. Bu tahmin kavramına göre, splayn düzeltmeyi esas alan farklı iki yaklaşım Schimek (2000) tarafından karşılaştırılmıştır. Bu çalışmada, yarı parametrik regresyon modelinin tahmini Green–Silverman'nın kısmi splayn ve Speckman yaklaşımı olarak adlandırılan farklı iki yaklaşımla yapılmış ve sıradan en küçük kareler tahminleri ile karşılaştırılmıştır. Splayn düzeltmede, *genelleştirilmiş çapraz geçerlilik* metoduyla düzeltme parametresi seçilmiş ve yarı parametrik modelin hem parametrik hem de parametrik olmayan bileşeni için çıkarsamalar tartışılmıştır. Bu amaçla, MATLAB ortamında yazılan bir program kullanılarak bir uygulama yapılmıştır.

Anahtar Kelimeler : Yarı parametrik regresyon, Splayn düzeltme, Kısmi splayn, Düzeltme parametresi

ESTIMATIONS AND INFERENCES IN SEMIPARAMETRIC MODELS WITH SMOOTHING SPLINES

ABSTRACT

A new estimation concept for semiparametric regression models based on smoothing splines has been introduced in Eubank et al. (1998). The two different approaches based on smoothing splines have been compared by Schimek (2000) according to this estimation concept. In this study, the estimation of the semiparametric regression model is performed with two different approaches called *partial splines of Green and Silverman and Speckman's approach* and compared with the estimations of the ordinary least squares. The smoothing parameter is selected via *generalized cross-validation* in smoothing splines and inferences have been discussed for both the parametric and the nonparametric part of the model. For this purpose an application is performed by using a program coded in MATLAB.

Keywords: Semiparametric regression, Smoothing splines, Partial splines, Smoothing parameter

[1,] Anadolu Üniversitesi, Bilecik Meslek Yüksekokulu, Bilecik
E-Posta: duaydin@anadolu.edu.tr

Geliş: 15 Mart 2006; **Düzeltme:** 07 Haziran 2006; **Kabul:** 20 Haziran 2006

1. GİRİŞ

Yarı parametrik (semiparametric) regresyon modelleri, bağımlı değişkenin bazı açıklayıcı değişkenlerle *doğrusal*, diğer açıklayıcı değişkenlerle *doğrusal olmayan* ilişki içerisinde olduğu regresyon modellerdir. Yarı parametrik modeller doğrusal parametrik bileşen ve parametrik olmayan bileşenlerin her ikisini de içerdiğinden bu modellere, "kısmi doğrusal modeller" de denir. Bu modeller, "boyutluluğun verdiği sıkıntı" nedeniyle parametrik olmayan regresyona tercih edilebilen, her bir değişkenin etkisinin daha kolay yorumlanmasına olanak sağlayan ve standart regresyon tekniklerini genelleştiren toplamsal regresyon modellerin özel bir durumu şeklindedir.

Araştırmada dikkate alınan yarı parametrik regresyon modeli

$$y_i = z_i^T\beta + f(x_i) + \varepsilon_i, 1 \le i \le n \qquad \textbf{(1)}$$

biçiminde ifade edilir. Burada,

y_i : y bağımlı değişkeninin i. gözlem değeri;

z_i : Parametrik kısma karşı gelen bağımsız değişkenlerin k boyutlu i. gözlemler vektörü;

β: k boyutlu regresyon katsayıları vektörü;

$f \in C^2[a,b]$: Modelin parametrik olmayan kısmına karşılık gelen ve $[a,b]$ aralığında ikinci mertebeden sürekli türeve sahip olan bilinmeyen bir pürüzsüz fonksiyon;

x_i : Parametrik olmayan kestirici değişkenin i. gözlem değeri;

$\varepsilon_i, i = 1,\ldots,n$, bağımsız sıfır ortalamalı ve σ^2 ortak varyanslı rassal hata terimleridir.

(1) modeli matris-vektör terimi ile,

$$y = Z\beta + f + \varepsilon \qquad \textbf{(2)}$$

şeklinde de ifade edilebilir.

Bu araştırmada, (1) yarı parametrik regresyon modeli esas alınacaktır. Amaç *model uyumunu elde etmektir*. Başka bir anlatımla, β *parametre vektörünü,* $f \in C^2[a,b]$ *fonksiyonunu ve* $u = Z\beta + f$ *ortalama vektörünü etkin olarak tahmin etmek, uyumu sayısal olarak özetlemek ve ayrıca uyumu grafiksel olarak görüntülemektir*. Konu ile ilgili olarak kuramda düzeltme tekniklerine dayalı olarak birkaç yaklaşım mevcuttur: Green vd. (1985), Engle vd (1986), Wahba (1990) ve Green ve Silverman (1994), Eubank vd (1998) ve Schimek (2000) tarafından yapılan çalışmalarda (1) modeline *splayn düzeltme metodunu* uygulanmıştır. Robinson (1988) *kernel (çekirdek) düzeltmesini*, Speckman (1998) *splayn ve kernel düzeltmesini*, Chen (1988) ise parçalı polinom yaklaşımını önermişlerdir. Bu araştırmada Green ve Silverman'nın *kısmi splayn yaklaşımı* ve Speckman *yaklaşımı* ele alınmıştır.

2. SPLAYN DÜZELTMEYE DAYALI TAHMİN

Esas alınan $\{(y_i, z_i, x_i)\}_{i=1}^n$ gözlem ya da ölçüm değerleri, eşitlik (1)'deki modele uygulandığında, giriş kısmında da belirtildiği gibi β ve f 'i tahmin etmek için birkaç yaklaşım vardır. Bunlar içerisinde en çok kullanılan, splayn düzelteme metodudur. Söz konusu metodun esasını, *cezalı en küçük kareler* toplamının minimum yapılması problemi oluşturur. Splayn düzletme metoduna göre (1) modelinin çözümü,

$$S(\beta, f) = \sum_{i=1}^{n}\{y_i - z_i^T\beta - f(x_i)\}^2 + \lambda\int_a^b \{f''(x)\}^2 dx \qquad \textbf{(3)}$$

$S(\beta, f)$ *cezalı en küçük kareler* toplamını minimum yapan f fonksiyonu ve β parametre vektörü olarak tanımlanır. Bu minimumu veren f fonksiyonu $x_1,\ldots,x_n$ düğümlü bir doğal kübik splayndır (Wahba, 1990; Green ve Silverman, 1994). Eşitlik (3)'de birinci terim hata kareler toplamını (RSS) gösterir ve bu ifade uyumun verilere yakınlığını ölçer. İkinci terim pürüzlülük cezasını gösterir ve bu kısım pürüzlülüğe bir ceza yükler. Ayrıca, λ bir skaler olup, $\{(x_i, y_i)\}_{i=1}^n$ gözlem değerlerine dayalı olarak belirlenir ve bilinmeyen regresyon fonksiyonu ile ilgili sabit bir düzeltme parametresidir. λ parametresi 0'dan $+\infty$'a değişirken, çözüm interpolasyondan basit bir doğrusal modele değişir. Eğer λ değeri $+\infty$'a yaklaşırsa, $S(\beta, f)$ sabit eğimli doğrusal regresyon uyumu üretir. Eğer λ değeri 0'a yaklaşırsa, $S(\beta, f)$ tümüyle esnek eğimli bir interpolasyon uyumuna karşı gelir. Yarı parametrik regresyon analizinde, istenen çözüm ne doğrusal ne de bir interpolasyondur. Bu nedenle λ 'nın uygun bir değerinin seçilmesi gerekir. Buda ancak farklı seçim teknikleri kullanılarak yapılabilir (Lee, 2003-2004; Mammadov, Yüzer ve Aydın, 2005; Aydın, 2005).

Eşitlik (3) ile verilen cezalı en küçük kareleri minimum yapan f fonksiyonu doğal kübik splayn ve verilen gözlem değerleri $y = (y_1,\ldots,y_n)^T$ vektörü olduğundan, f fonksiyonunu parametreleştirmek için (3) eşitliğindeki pürüzlülük ceza terimi, $\int_a^b f''(x)^2 dx = \mathbf{f}^T K\mathbf{f}$ olarak yazılabilir (bak. Geren ve Silverman (1994)). Burada $\mathbf{f} = (f(x_1),\ldots,f(x_n))^T$ ve K bir yarı pozitif tanımlı matris olup, $x_1 < \cdots < x_n$ düğüm noktalarından elde edilen R ve Q matrisleri yardımıyla

$$K = QR^{-1}Q^T \qquad \textbf{(4)}$$

şeklinde hesaplanır. (4) eşitliğindeki R, $(n-2)\times(n-2)$ simetrik bant matris; Q, $n\times(n-2)$

bant matristir. $R(r_{ij})$ ve $Q(q_{ij})$ bant matrislerinin elemanları aşağıdaki gibi hesaplanır:

$$r_{ij}=\begin{cases}\frac{1}{6}h_{j-1} & ,i=j-1;\\ \frac{1}{3}(h_{j-1}+h_j) & ,i=j,i=2,3,\ldots,n-1;\\ 0 & ,|i-j|\geq 2,\ j=2,3,\ldots,n-1;\\ \frac{1}{6}h_j & ,i=j+1;\end{cases}$$

ve

$$q_{ij}=\begin{cases}\frac{1}{h_{j-1}} & ,i=j-1;\\ -\left(\frac{1}{h_{j-1}}+\frac{1}{h_j}\right) & ,i=j,i=1,2,\ldots,n;\\ 0 & ,|i-j|\geq 2,\ j=2,3,\ldots,n-1;\\ \frac{1}{h_j} & ,i=j+1;\end{cases}$$

Burada $h_i = x_{i+1} - x_i, i = 1,2,\ldots,n-1$. Söz konusu düğüm noktalarının $x_1 < \cdots < x_n$ koşulunu sağlaması, diğer bir ifadeyle farklı ve sıralı olması gerekir. x_i'ler farklı ve sıralı değilse, onlar belirli bir "*N tekrarlanma matrisi*" (incidence matrix) yardımıyla farklı ve sıralı hale getirilir. $x_1,\ldots,x_n$ düğüm noktalarının farklı ve sıralı değerleri $s_1,\ldots,s_q$ ile gösterilsin. Bu durumda, N tekrarlanma matrisinin n_{ij} elamanları

$$n_{ij}=\begin{cases}1, & x_i = s_j \ \ ise\\ 0, & \text{diğ. dur.}\end{cases}\quad ,\ i=1,\ldots,q;\ j=1,\ldots,n$$

biçiminde ifade edilir. Söz konusu N matrisi yardımıyla, (3)'te belirtilen $S(\beta, f)$ cezalı kareler toplamı

$$S(\beta, f) = (y - Z\beta - N\mathbf{f})^T (y - Z\beta - N\mathbf{f}) + \lambda \mathbf{f}^T K \mathbf{f} \qquad \textbf{(5)}$$

biçiminde yazılabilir. Pürüzlülük ceza yaklaşımı olarak adlandırılan cezalı en küçük karelerin esası geleneksel doğrusal regresyon modelinin çözümünde kullanılan sıradan en küçük karelere benzer olarak, (5) eşitliğini minimum yapan $\mathbf{f}$ ve β vektörlerinin kestirimidir. Eşitlik (5)'in sırasıyla β ve $\mathbf{f}$ 'e göre türevleri alınıp sıfıra eşitlenirse

$$Z^T Z\beta = Z^T (y - N\,\mathbf{f}) \qquad \textbf{(6)}$$

$$(N^T N + \lambda K)\,\mathbf{f} = N^T (y - Z\beta) \qquad \textbf{(7)}$$

denklem sistemleri elde edilir. (7) ile verilen denkleminden $\mathbf{f}$ vektörü

$$\mathbf{f} = N(N^T N + \lambda K)^{-1} N^{-1}(y - Z\beta) = S_\lambda (y - Z\beta) \qquad \textbf{(8)}$$

olarak elde edilir. (8) eşitliğinden, $\mathbf{f}$ uyum değerleri vektörünü elde etmek için $\mathbf{y}$ vektörüne uygulanan ve verilen bir $\lambda > 0$ sabitine bağlı *düzeltme matrisi*

$$S_\lambda = N(N^T N + \lambda K)^{-1} N^T \qquad \textbf{(9)}$$

şeklinde ifade edilir. Ayrıca, önceden x_i düğüm noktaları farklı ve sıralıysa, $N = I$ (birim matris) olması nedeniyle, S_λ düzeltme matrisi aşağıdaki şekle indirgenir:

$$S_\lambda = (I + \lambda K)^{-1}$$

2.1 Kısmi Splayn Yaklaşımı

Kısmi splayn yaklaşımına göre, λ'nın önceden belirlenmiş değeri için (1) modelindeki β ve $\mathbf{f}$ vektörüne karşı gelen kestiriciler, (9) eşitliğinde verilen pozitif tanımlı bir S_λ düzeltme matrisi yardımıyla elde edilir: Parametrik olmayan $f = f(x_i)$ değerlerine karşı gelen (7) denklemdeki $\mathbf{f}$ vektörü

$$\hat{\mathbf{f}} = S_\lambda (y - Z\beta) \qquad \textbf{(10)}$$

biçiminde elde edilir. (10) ifadesi (6)'da yerine yazılarak, β parametrik katsayılar vektörü

$$\hat{\beta} = \left[Z^T (I - S_\lambda) Z\right]^{-1} Z^T (I - S_\lambda)\mathbf{y} \qquad \textbf{(11)}$$

şeklinde elde edilir. Green ve Silvermanın kısmi splayn yaklaşımı (6) ve (7) normal denklemlerin iteratif çözümüne dayalıdır. Diğer taraftan, parametik değişkenler matrisi Z 'nin sütun bileşenleri x'e bağlı olduğunda λ'nın seçimi için kısmi splayn kestiricilerinin genellikle yanlı olduğu Rice (1986) tarafından gösterilmiştir. Speckman'nın geliştirdiği yöntemde ise sözü edilen yan önemli ölçüde indirgenebilmektedir.

2.2 Speckman Yaklaşımı

Bu bölümde, Speckman (1988) tarafından önerilen, kısmi splayn yaklaşımına alternatif bir yaklaşım incelenecektir. Bu metotta (1) modelindeki z_i açıklayıcı değişkenin x_i parametrik olmayan kestirici değişkenine bağımlı olduğu varsayılır:

$$z_i = m(x_i) + \eta_i\ , i = 1,\ldots,n \qquad \textbf{(12)}$$

(12) eşitliğinde, $m(x_i)$, x_i'in pürüzsüz vektör-fonksiyonu ve η_i, hata terimleri vektörüdür. (12) modeli (1)'de yerine yazılarak

$$\begin{aligned} y_i &= \left(m(x_i) + \eta_i\right)^T\beta + f(x_i) + \varepsilon_i \\ &= m(x_i)^T\beta + f(x_i) + \left(\eta_i^T\beta + \varepsilon_i\right) \end{aligned} \quad \textbf{(13)}$$

elde edilir. (13) eşitlindeki $\left(\eta_i^T\beta + \varepsilon_i\right) = hata$ ve $m(x_i)^T\beta + f(x_i) = f_0(x_i)$ olarak ifade edilsin. Bu durumda modelin

$$y_i = f_0(x_i) + hata \quad \textbf{(14)}$$

şeklinde olabilmesi için aşağıdaki gibi bir f_0 fonksiyonu tanımlanabilir:

$$f_0(x_i) = m(x_i)^T\beta + f(x_i) \quad \textbf{(15)}$$

(1) ve (15) eşitliklerinin farkı alınarak, aşağıdaki ifade elde edilir:

$$y_i - f_0(x_i) = \left(\mathbf{z}_i - m(x_i)\right)^T\beta + hata \quad \textbf{(16)}$$

Eşitlik (16) gösterir ki, β parametre vektörü, söz konusu açıklayıcı değişkenlere göre y_i değişkenine ilişkin artıklarının doğrusal regresyonundan elde edilir. S_λ, verilen herhangi bir λ düzeltme parametresi için x_i düğüm noktaları ile belirlenen (9) eşitliğindeki düzeltme matrisi olsun. Ayrıca $\mathbf{y}$, i. gözlem değeri y_i ile gösterilen bağımlı değişken vektörü, Z, i. satırı $\mathbf{z}_i^T$ ile belirtilen bağımsız değişkenlerin gözlem değerlerinin matrisi ve θ, i. satırı $m(x_i)^T$ ile belirtilen matris olarak gösterilirse, 16) eşitliğine göre aşağıdaki adımlar gerçekleştirilir (Speckman, 1988).

Adım 1: (12) ve (14) modellerinde splayn düzeltme yöntemi kullanılarak sırasıyla,
$S_\lambda Z = \theta = \left(m(x_1),...,m(x_n)\right)^T$ matrisi ve
$S_\lambda \mathbf{y} = \mathbf{f}_0 = \left(f_0(x_1),...,f_0(x_n)\right)^T$ vektörü hesaplanır

Adım 2: (16) eşitliğinden

$$\mathbf{y} - \mathbf{f}_0 = \mathbf{y} - S_\lambda\mathbf{y} = (I - S_\lambda)\mathbf{y}$$

ve

$$Z - \theta = Z - S_\lambda Z = (I - S_\lambda)Z$$

olarak yazılabilir. Buradan, $\tilde{\mathbf{y}} = (I - S_\lambda)\mathbf{y}$ ve $\tilde{Z} = (I - S_\lambda)Z$ dönüştürülmüş değişkenleri elde edilir.

Adım 3: (16)denklemi, $\tilde{\mathbf{y}}$ 'nin $\tilde{Z}$ 'ye göre regresyon denklemi olarak

$$\tilde{\mathbf{y}} = \tilde{Z}\beta + hata$$

şeklinde ifade edilir. Bu modele *sıradan en küçük kareler (SEK) yöntemi* uygulanarak, (1) modelinin parametrik bileşeni için katsayılar kestirimi aşağıdaki gibi bulunur:

$$\hat{\beta} = \left(\tilde{Z}^T\tilde{Z}\right)^{-1}\tilde{Z}^T\tilde{\mathbf{y}} \quad \textbf{(17)}$$

Adım 4: (17) eşitliğindeki $\hat{\beta}$ tahmini (1) modelinde yerine yazılarak (1) yarı parametrik regresyon modeli,

$$\mathbf{y}_i^* = f(x_i) + hata$$

şeklindeki parametrik olmayan bir modele dönüşür. Burada $\mathbf{y}_i^* = y_i - \mathbf{z}_i^T\hat{\beta}$. Bağımlı $\mathbf{y}_i^*$ değişkenine uygulanan splayn düzeltme yönteminde, (1) modelinin parametrik olmayan bileşeni için aşağıdaki gibi bir $\hat{\mathbf{f}}$ tahmini elde edilir:

$$\mathbf{f} = S_\lambda\mathbf{y}_i^* = (I + \lambda K)^{-1}\mathbf{y}_i^* \quad \textbf{(18)}$$

3. DÜZELTME PARAMETRESİNİN SEÇİMİ

Splayn düzeltme metoduna dayalı yarı parametrik regresyonda düzeltme parametresinin seçimi önemli bir problemdir. Bu konuda, Aydın (2005) tarafından yapılan simülasyon çalışmasında λ düzeltme parametresinin seçimi için kullanılan diğer metotlara göre genelleştirilmiş çapraz geçerlilik (GCV), daha iyi bir performans göstermiştir. Bu nedenle çalışmada λ düzeltme parametresinin seçimi için GCV metodu kullanılmıştır. λ parametresinin GCV tahmini, aşağıdaki GCV(λ) fonksiyonunu minimum yapan değer olarak belirlenir (Wahba, 1990):

$$GCV(\lambda) = \frac{1}{n}\frac{\sum_{i=1}^{n}\left(y_i - f_\lambda(x_i)\right)^2}{\left(1 - n^{-1}tr(S_\lambda)\right)^2} = \frac{n^{-1}\left\|(I - S_\lambda)\mathbf{y}\right\|^2}{\left[n^{-1}tr(I - S_\lambda)\right]^2} \quad \textbf{(19)}$$

Yarı parametrik regresyon modelinin geçerliliğinin değerlendirilebilmesi için, modelin hata terimlerinin varyansına ve modeli oluşturan hem parametrik hem de parametrik olmayan bileşenin varyanslarına ihtiyaç vardır. İzleyen bölümde varyansların hesaplanmasına ve model hakkındaki çıkarsamalara yer verilmiştir.

4. VARYANS-KOVARYANS TAHMİNİ VE ÇIKARSAMALAR

Yarı parametrik regresyon modelinin varyans ve kovaryansların tahminleri aşağıda belirtilen amaçlar için gereklidir:

i. Modelin parametrik bileşeni hakkındaki çıkarsamalar;

ii. Modelin parametrik olmayan fonksiyonu hakkındaki çıkarsamalar.

Yarı parametrik modelin parametrik bileşeni için geliştirilen varyans kestiricisi, aşağıdaki gibi elde edilir (Eubank vd. 1998; Schimek 2000):

$$\hat{\sigma}^2 = \frac{\mathbf{y}^T \mathbf{A}^T (\mathbf{I}-\mathbf{P})\mathbf{A}\mathbf{y}}{tr\left(\mathbf{A}^T (\mathbf{I}-\mathbf{P})\mathbf{A}\right)} \qquad \textbf{(20)}$$

Burada $\mathbf{P} = \mathbf{A}Z\left(Z^T \mathbf{A}^T \mathbf{A}Z\right)^{-1} Z^T \mathbf{A}^T$. Eşitlik (20)'de $(n-2)\times n$ boyutlu A matrisinin herhangi bir i. satırının i. elemanı $a_i c_i$, $(i+1)$. elemanı $-c_i$ ve $(i+2)$. elemanı $b_i c_i$ olmak üzere, bu satırın diğer tüm elemanları sıfırdır. Burada,

$$\mathbf{a}_i = \frac{x_{i+1}-x_i}{x_{i+1}-x_{i-1}}, \quad \mathbf{b}_i = \frac{x_i - x_{i-1}}{x_{i+1}-x_{i-1}} \quad ve$$

$$\mathbf{c}_i = \left(\mathbf{a}_i^2 + \mathbf{b}_i^2 + 1\right)^{-\frac{1}{2}}, \; i = 1,\ldots,n-1.$$

Schimek (2000) tarafından yapılan çalışmada yarı parametrik regresyonun parametrik katsayılarının varyans-kovaryans kestiricileri hem kısmi splayn (β_p) hem de Speckman yaklaşımı (β_s) için hesaplanmıştır. Bu kestiriciler için sırasıyla varyans-kovaryans matrisleri aşağıdaki gibi verilir:

$$\mathbf{Var}_p = \sigma^2 \left(Z^T Z\right)^{-1} \tilde{Z}^T \tilde{Z} \left(Z^T \tilde{Z}\right)^{-1} \qquad \textbf{(21)}$$

$$\mathbf{Var}_S = \sigma^2 \left(Z^T \tilde{Z}\right)^{-1} \tilde{Z}^T \left(I - S_\lambda\right)^2 \tilde{Z} \left(Z^T \tilde{Z}\right)^{-1} \qquad \textbf{(22)}$$

(21) ve (22)'de $\tilde{Z} = (I - S_\lambda)Z$, S_λ ise (9) ile tanımlanan düzeltici matristir. Varyans-kovaryans matrislerinin asal köşegeni üzerindeki elemanlar varyansları, diğerleri ise kovaryansları gösterir.

4.1 Parametrik Bileşen İçin Çıkarsama

Geleneksel regresyon analizine benzer olarak, yarı parametrik regresyon analizi de örneklem verileri üzerinden yapıldığından, elde edilen $\hat{\beta}$ tahmin vektörü, β parametrelerinin anlamlığının test edilmesinde kullanılır. Dolayısıyla, parametrik katsayıların testi, modelin anlamlılığını da test eder. Buna göre, her bir $\hat{\beta}$ parametrik katsayının istatistiksel açıdan anlamlı olup olmadığı test edilmek istendiğinde, β parametresiyle ilgili hipotezler;

$$H_0 : \beta_j = 0$$
$$H_1 : \beta_j \neq 0$$

şeklinde formüle edilir. $\beta_j = 0$ hipotezi x_j bağımsız değişkeninin bağımlı değişken üzerinde etkili olmadığını gösterirken, $\beta_j \neq 0$ hipotezi, x_j bağımsız değişkeninin bağımlı değişken üzerinde etkili olduğunu göstermektedir. Her bir parametrik katsayıyı test etmek için aşağıdaki t istatistiği kullanılır:

$$t_{df} = \frac{\hat{\beta}_j}{SE(\hat{\beta}_j)}, \quad j = 1,\ldots,k \qquad \textbf{(23)}$$

Burada $SE(\hat{\beta}_j)$, $\hat{\beta}_j$ katsayılarının standart hataları olup (21) yada (22) varyans-kovaryans matrisinin köşegen elemanları yardımıyla hesaplanır $\left(var(\hat{\beta}_j) = SE(\hat{\beta}_j)\right)$. Eşitlik (23)'de verilen t istatistiği $df = n - tr(S_\lambda) - k$ serbestlik derecesine göre t-tablosundan bulunan değerden büyükse sıfır hipotezi reddedilir, diğer bir deyişle, β_j katsayısının anlamlı olduğu sonucuna varılır.

Speckman (1988), varyansın (σ^2) bir kestiricisi olarak

$$\hat{\sigma}^2 = MSE = \frac{Hata\,KarelerToplamı\,(RSS)}{Serbestlik\,Derecesi} = \frac{n^{-1}\left\|(I-\mathbf{H}_s)\mathbf{y}\right\|^2}{tr\left[(I-\mathbf{H}_s)^T (I-\mathbf{H}_s)\right]} \qquad \textbf{(24)}$$

şeklinde ifade edilen hata kareler ortalamasını kullanmayı önermiştir. (24) eşitliğindeki H_s, şapka (hat) matrisi olup aşağıdaki şekilde hesaplanır:

$$H_s = S_\lambda + \tilde{Z}\left(Z^T \tilde{Z}\right)^{-1} \tilde{Z}^T (I - S_\lambda) \qquad \textbf{(25)}$$

Eşitlik (25)'de $\tilde{Z} = (I - S_\lambda)Z$ olarak ifade edilir.

Parametrik katsayıların toplu olarak istatistiksel açıdan anlamlı olup olmadığını test edebilmek için hipotezler

$$H_0 : \hat{\beta} = 0$$
$$H_1 : \hat{\beta} \neq 0$$

veya

$$H_0 : \beta_1 = \cdots = \beta_k = 0$$
$$H_1 : \beta_1 \neq \cdots \neq \beta_k \neq 0 \text{ (en az bir } \beta_j \neq 0)$$

şeklinde formüle edilir. Söz konusu bu hipotezler, $df_1 = k$ ve $df_2 = tr\left[(I - H_s)^T (I - H_s)\right]$ serbestlik derecelerine sahip bir F istatistiği yardımıyla test edilir (Schimek, 2000):

$$F_{df_1, df_2} = \frac{MSS_{par}}{MSE} = \frac{\hat{\beta}_s^T \left(q^T q\right)^{-1} \hat{\beta}_s / n}{RSS / tr\left[(I - H_s^T)(I - H_s)\right]} \qquad \textbf{(26)}$$

Eşitlik (26)'da, $q^T = \left(Z^T \tilde{Z}\right)^{-1} \tilde{Z}^T (I - S_\lambda)$.

Diğer taraftan, istatistiksel çıkarsamalarda yapılan kestirimlerin, gerçek değerlerle genellemesi aralık kestirimiyle yapılır. (6) ve (7) denklemlerinden tahmin edilen β parametrik katsayılar vektörü için *100(*$1-\alpha$*)% güven aralığı*

$$P\left(\hat{\beta}_i - t_\alpha SE(\hat{\beta}_i) \le \beta \le \hat{\beta}_i + t_\alpha SE(\hat{\beta}_i)\right) = 1-\alpha \qquad \textbf{(27)}$$

biçiminde verilir. (27) ifadesinde $SE(\hat{\beta}_i)$ parametrik katsayıların standart hataları, t_α, α anlam düzeyi ve $\mathrm{df} = n - tr(\mathbf{H}_\lambda) - k$ serbestlik derecesindeki t-tablo değerdir.

4.2 Parametrik Olmayan Bileşen İçin Çıkarsama

Burada amaç f eğrisinin biçimsel olarak şeklini değerlendirmektir. Test edilmek istenen sıfır ve alternatif hipotezler aşağıdaki şekilde ifade edilir:

$$H_0 : E(y_i) = \mu \;(\text{doğrusal fonksiyon})$$
$$H_1 : E(y_i) = f(x_i) \;(\text{pürüzsüz fonksiyon})$$

Böyle bir test sade parametrik bir modelle karşılaştırılan yarı parametrik modelin bir anlam ifade edip etmeyeceğine karar verilmesine olanak sağladığı için ilgi odağıdır. Hastie and Tibshirani (1990), yarı parametrik ortama uygulanabilen, bir parametrik olmayan uyum $\hat{f}_1$'e karşın bir doğru denklem uyumu $\hat{f}_0$'ı için bir F testi önermiştir. Söz konusu F-test istatistiği

$$F_{\mathrm{df}_1-\mathrm{df}_0,\mathrm{n}-\mathrm{df}_1} = \frac{\left(\sum_{i=1}^{n}\hat{\varepsilon}_i^2 - \sum_{i=1}^{n}\hat{v}_i^2\right)/(\mathrm{df}_1-\mathrm{df}_0)}{\sum_{i=1}^{n}\hat{v}_i^2/(n-\mathrm{df}_1)} \qquad \textbf{(28)}$$

formülü ile verilir. Burada $\hat{\varepsilon}_i = (y_i - z_i^T\hat{\beta}_{OLS})$ ve $\hat{v}_i = z_i^T\hat{\beta}_s + \hat{f}(x_i) - z_i^T\beta_{OLS}$. Serbestlik dereceleri df_0, sıradan en küçük karelerde olduğu gibi, modeldeki parametre sayısına (k tane) ve $\mathrm{df}_1 = tr(\mathrm{H}_s - \mathrm{H}_s\mathrm{H}_s^T)$ ifadesine eşittir. Eşitlik (28)'de verilen bu F-test istatistiği σ^2'nin tahminini gerektirmeyen bir avantaja sahiptir.

Amaç, eşitlik (1) ile verilen modeldeki parametrik olmayan bileşene karşı gelen f eğrisinin şekli hakkında sağlam bir karar verebilmektir. Açıktır ki, örnek değişkenliğinden dolayı parametrik olmayan tahminlerde her zaman bazı eğriliklerin beklenmesi gerekir. Güven aralıkları önceden belirlenen bir olasılık düzeyine ve örneklem sonuçlarına göre şekillenirler. $\hat{f}(x)$ bilinmeyen pürüzsüz fonksiyonun splayn düzeltme tahmini olmak üzere, *parametrik olmayan bileşenin 100(*$1-\alpha$*)% güven aralığı*

$$\hat{f}(x) - 2SE(\hat{f}(x)) \le f(x) \le \hat{f}(x) + 2SE(\hat{f}(x)) \qquad \textbf{(29)}$$

olarak ifade edilir. Eşitlik (29)'da $SE(\hat{f}(x))$, f eğrisinin tahmini standart hatasıdır ($\hat{\mathrm{f}}$ vektörünün standart hatasıdır).

5. ÖRNEK UYGULAMA

Bu bölümde, yukarıda verilen teorik bilgilerin uygulanabilirliği ve önemini göstermek amacıyla, müstakil evlerin satış fiyatları ile söz konusu evlerin özellikleri arasındaki ilişkiler araştırılmış, müstakil evlerin satış fiyatları üzerinde etkili olan bağımsız değişkenlerin etkilerinin tahmin edilmesinde hem parametrik hem de yarı parametrik regresyon modeli kullanılmıştır. Söz konusu evlerin satış fiyatları ile fiyatı etkileyen açıklayıcı değişkenlerin doğası gereğince hem parametrik hem de parametrik olmayan bir ilişki içerisinde olduğu göz ardı edilemeyecek bir gerçektir. Örneğin, evlerin satış fiyatları ile evlerin bahçe dahil brüt kullanım alanları arasında kesin bir doğrusal ilişki vardır denilemez. Bu nedenle, böyle bir ilişkiyi analiz etmede hem parametrik ve hem de parametrik olmayan kısmın modele dahil edildiği, yarı parametrik regresyon modelinin kullanılması daha uygun olacaktır. İzleyen alt bölümde, adı geçen ilişkiyi analiz etmek için yarı parametrik regresyon modelinin kestiriminde iki farklı yaklaşım kullanılarak, elde edilen sonuçlar tartışılmıştır.

5.1 Veri ve Değişken Tanımları

Veriler Kanada'nın başkenti olan Ottawa'da, 1987 yılında satılan 92 müstakil eve ait olup, Ho (1995) tarafından yapılan "essay on the housing market" adlı doktora tez çalışmasından alınmıştır. Söz konusu verilere ilişkin bazı özet bilgiler Tablo 1'de verilmiştir. Örnek uygulamada kullanılan değişkenler aşağıdaki gibi tanımlanır:

SF: Evin satış fiyatı (dolar);

AU: Evin anayola uzaklığı;

FD: Şömine için yapay (dummy) değişken;

GD: Garaj için yapay değişken;

KG: Söz konusu semtte yaşayan insanların ortalama geliri(dolar);

NKA: Evin net kullanım alanı (square feet);

BKA: Evin bahçe dahil brüt kullanım alanını (square feet) gösterir.

Tablo 1. Değişkenlere ilişkin gözlem değerlerinin özeti

Değişkenler	Ortalama	St Sap.	Max	Min
SF	146.0049	33.4323	240	75
AU	0.8399	0.4477	1.8643	0.1032
FD	0.6957	0.4627	1	0
GD	0.6413	0.4822	1	0
KG	49.2817	11.6262	79.4580	25.9540
NKA	1.1419	0.2862	1.9997	0.6247
BKA	5.2823	1.1551	9.9000	1.8910

5.2. Deneysel Değerlendirmeler

Bu bölümde, yarı parametrik regresyon modelinin kestirimi, hem SEK ve hem de Speckman ve kısmi splayn yaklaşımı ile gerçekleştirilmiştir. Ayrıca splayn düzeltme metodunu esas alan yarı parametrik regresyon modelinin parametreleri hakkında tahmin ve çıkarsamalar yapılmıştır.

5.2.1 Parametrik Regresyon Modelinin Ayrıntıları

Örnek uygulamada kullanılan değişkenlere göre, parametrik doğrusal regresyon modeli

$$SF_i = \beta_0 + \beta_1 AU_i + \beta_2 FD_i + \beta_3 GD_i + \beta_4 KG_i + \beta_5 NKA_i + \beta_6 BKA_i + \varepsilon_i \quad, i = 1,2,\ldots,n \tag{30}$$

şeklinde yazılabilir. (30) modelindeki $\beta = (\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6)$ regresyon katsayıları vektörüdür. Söz konusu bu katsayılar sıradan en küçük kareler yöntemine göre elde edilmiş ve onların standart hataları, $t-test$ istatistikleri ve %95 güven aralıkları hesaplanarak, sonuçlar tablo 2'de verilmiştir.

Tablo 2 incelendiğinde, evlerin ana yola olan uzaklığında bir birimlik artış, evlerin satış fiyatlarında 8.2151 birimlik bir azalışa neden olmakta veya bunun tersi geçerli olmaktadır. Buna karşılık, diğer tüm değişkenler ile evlerin satış fiyatları arasında aynı yönlü bir ilişki gözlenmektedir. Ancak (23) formülü ile belirlenen ve Tablo 2'de verilen $t-test$ istatistikleri incelendiğinde, AU, FD ve BKA değişkenlerine ilişkin katsayılar anlamsızdır. Diğer taraftan, katsayıların %95 güven aralıkları oldukça geniş bir aralığı kapsadığı görülmektedir. Ayrıca, modelin R^2 belirlilik katsayısına göre, evlerin satış fiyatlarındaki değişmelerin ancak %33.287'si söz konusu açıklayıcı değişkenler tarafından açıklanabilmektedir. Bunun yanı sıra, evlerin satış fiyatlarına ilişkin tahminlerde, yapılan hataların oldukça yüksek olduğu görülmektedir. Diğer bir deyişle, tablo 2'de verilen ve $RSS = 67855$ olarak hesaplanan hata kareler toplamı oldukça yüksek bir değerdir. Bu durum, parametrik regresyon modeli ile yapılan tahminlerin yeterince tutarlı olmadığı biçiminde yorumlanabilir.

5.2.2 Yarı Parametrik Regresyon Modelinin Ayrıntıları

Bu uygulama için (1) ile tanımlanan yarı parametrik regresyon modeli

$$SF_i = \beta_1 AU_i + \beta_2 FD_i + \beta_3 GD_i + \beta_4 KG_i + \beta_5 NKA_i + f(BKA_i) + \varepsilon_i \ , i = 1,2,\ldots,n \tag{31}$$

biçiminde ifade edilebilir. Yarı parametrik regresyonda yapay değişkenler her zaman parametrik kısma dahil edilirler. Bu uygulamada yer alan değişkenlerden BKA değişkeni parametrik olmayan değişken olarak ele alınırken, diğer tüm değişkenler parametrik değişken olarak ele alınmıştır. Splayn düzeltmenin esasını oluşturan (3) eşitliğindeki, λ düzeltme parametresinin seçiminde, GCV kriteri kullanılmıştır. Kısmi splayn yakla-

Tablo 2. Parametrik Regresyon Sonuçlar

Değişkenler	Katsayılar	Standart hat.	t- ist.	%95 Güven aralıkları
Sabit	62.67	20.451	3.0644	[(22.586)-(102.75)]
AU	-8.2151	6.6935	-1.2273	[(-21.334)-(4.9042)]
FD	6.0548	7.5919	0.79754	[(-8.8253)-(20.935)]
GD	14.468	6.3102	2.2929	[(2.1005)-(26.836)]
KG	0.56001	0.27959	2.003	[(0.012023)-(1.108)]
NKA	39.27	11.855	3.3124	[(16.033)-(62.506)]
BKA	0.81454	2.6181	0.31112	[(-4.3169)-(5.946)]
$R^2 = 0.33287$, $RSS = 67855$, $df = 85$ ve $t_{0.05}(85) = 1.99$				

[2] Doğrusal regresyon için H=Z (Z' Z)$^{-1}$ Z' şapka matrisi yardımı ile, RSS= **y**' (I-H) **y** formülü ile verilen hata kareler toplamı hesaplanır. Diğer taraftan, **y** ve **ŷ** değerleri arasındaki korelasyon katsayısının karesi R^2 değerini diğer bir deyişle, modelin belirlilik katsayısını vermektedir.

Tablo 3. Kısmi Splayn Yaklaşımı İle Elde Edilen Regresyon Sonuçları

Değişkenler	Katsayılar	Standart hat.	t- istatist.	%95 Güven Aralıkları
Sabit	-	-	-	-
AU	-8.688	0.70338	-12.352	[(-10.067)-(-7.3094)]
FD	5.6203	0.79823	7.041	[(4.0558)-(7.1848)]
GD	14.122	0.66292	21.303	[(12.823)-(15.422)]
KG	0.58242	0.029402	19.809	[(0.52479)-(0.64004)]
NKA	38.043	1.2526	30.372	[(35.588)-(40.498)]
BKA	-	-	-	-
$R^2 = 0.96546$, $RSS = 718.33$, $F = 53.3967$, $df_1 = 5$ *ve* $df_2 = 83.4778$, $F_\alpha(df_1, df_2) = 2.29$ $df = 84.9972$ ve $t_\alpha(df) = 1.99$				

şımına göre, model (31)'de, β parametrik katsayıları ve $f(BKA)$ bilinmeyen fonksiyonunun aldığı f değerler vektörü elde edilmiş olup, sonuçlar Tablo 3'de verilmiştir.

Söz konusu yaklaşıma göre tahmin edilen modelin parametrik katsayıları (1) eşitliğinde yerine yazılarak, yarı parametrik regresyon modeli aşağıdaki gibi ifade edilebilir:

$$SF = -8.688AU + 5.6203FD + 14.122GD + 0.58242KG + 38.043NKA + f(BKA) + \varepsilon \quad \textbf{(32)}$$

Tablo 3 veya (32) eşitliği incelendiğinde, evlerin ana yola olan uzaklığında bir birimlik bir artış, evlerin satış fiyatlarında 8.688 birimlik bir azalışa neden olur veya bunun tersi geçerlidir. Buna karşılık, diğer tüm değişkenler ile evlerin satış fiyatları arasında aynı yönlü bir ilişki olduğu görülmektedir.Parametrik olmayan bileşen ise, *brüt kullanım alanlarına göre, evlerin tahmini satış fiyatlarını gösterir:*

$$\hat{f}(BKA) = S_\lambda\left(SF - (AU, FD, GD, KG, NKA)\hat{\beta}\right)$$

Bu değer 92 müstakil eve ilişkin fiyatlarını verdiğinden o ancak grafiksel olarak Şekiller1-2'de gösterilmiştir.

Bölüm 2.1'de incelenen Speckman yaklaşımına göre elde edilen sonuçlar ise, özetle Tablo 4'de verilmiştir. Speckman yaklaşımına göre tahmin edilen parametrik katsayılar (1)'de yerine yazılarak, yarı parametrik regresyon modeli aşağıdaki gibi ifade edilebilir:

$$SF = -8.947AU + 5.2692FD + 13.933GD + 0.59775KG + 37.657NKA + f(BKA) + \varepsilon \quad \textbf{(33)}$$

Tablo 4 veya (33) eşitliği incelendiğinde, (1) modelinin sonuçlarına benzer olarak, evlerin ana yola olan uzaklığındaki bir birimlik bir artışın, evlerin satış fiyatlarında 8.974 birimlik bir azalışa neden olduğu görülür. Buna karşılık, diğer tüm değişkenler ile evlerin satış fiyatları arasında aynı yönlü bir ilişki olduğu görülmektedir.

5.3 Yarı Parametrik Regresyon Modeline İlişkin Çıkarsamalar

Speckman ve kısmi splayn yaklaşımına göre elde edilen (1) modelinin parametrik katsayılarının tahmin değerleri ve standart sapmaları, $t-test$ istatistikleri, parametrik katsayıların güven aralıkları ve *F*-test istatistiği sonuçları bir önceki bölümde verilmiştir. Bir sonraki paragraflarda bu sonuçlara ilişkin yorum ve çıkarsamalar yapılmıştır.

5.3.1 Parametrik Bileşen Hakkında Çıkarsamalar

Kısmi splayn yaklaşımına göre kestirilen (1) modelinin, parametrik katsayılarının anlamlı olup olmadığını belirlemek amacıyla, Tablo 3'de verilen t-test istatistikleri incelendiğinde, örneğin, $\hat{\beta}_s$ katsayısına ilişkin hesaplanan, $t-test$ istatistiği = 30.372 olup, bu değer, $t_{0.05}(84.9972) = 1.99$ değerinden büyük olduğundan sıfır hipotez reddedilir. Buna göre,

Tablo 4. Speckman Yaklaşımı İle Elde Elden Regresyon Sonuçları

Değişkenler	Katsayılar	Standart hat.	t- istatist.	%95 Güven Aralıkları
Sabit	-	-	-	-
AU	-8.947	0.7044	-12.702	[(-10.328)-(-7.5664)]
FD	5.2692	0.80091	6.579	[(3.6994)-(6.839)]
GD	13.933	0.66395	20.984	[(12.631)-(15.234)]
KG	0.59775	0.029568	20.216	[(0.5398)-(0.65571)]
NKA	37.657	1.2581	29.931	[(35.191)-(40.123)]
BKA	-	-	-	-
$R^2 = 0.96555$, $RSS = 717.53$ $F = 39.2996$, $df_1 = 5$ *ve* $df_2 = 84.4432$, $F_\alpha(df_1, df_2) = 2.29$ $df = 84.9972$ ve $t_\alpha(df) = 1.99$				

$\hat{\beta}_s$ katsayısı istatistiksel açıdan anlamlıdır. Benzer şekilde, aynı değişken için Tablo 4'de verilen Speckman yaklaşımına ait $t-test$ istatistiği $t-tablo$ değeri ile $\hat{\beta}_s$ katsayısının da istatistiksel açıdan anlamlı olduğu görülmektedir. Bu bilgiler doğrultusunda, her iki yaklaşımdan elde edilen parametrik katsayıların istatistiksel açıdan anlamlı oldukları görülmektedir (bak, Tablo 3 ve 4). Diğer taraftan, Tablo 3 ve 4'te verilen *F*-istatistikleri, %5 anlam düzeyindeki *F*-tablo değerinden büyük olduğundan sıfır hipotezi reddedilir. Dolayısıyla her iki yönteme göre de kestirilen (1) modelinin, t-testi sonucunda olduğu gibi, *F*- test istatistiğine göre de istatistiksel açıdan anlamlı olduğu söylenebilir. Eşitlik (1) ile verilen yarı parametrik regresyon modelinin parametrik katsayılarının (27) güven aralıkları da incelenmiş olup, bu sonuçlar da yine Tablo 3 ve 4'de verilmiştir. Örneğin, Speckman yaklaşıma göre, kestirimi yapılan ve *NKA* parametrik açıklayıcı değişkeninin etkisini gösteren β_s katsayısı için %95 güven aralığı,

$$P\left(35.191 \leq \beta_s \leq 40.123\right) = 0.95$$

biçimde elde edilmiştir. Buna göre, ele alınan örneklemin alındığı ana kütlenin β_s katsayısının 35.191 ile 40.123 aralığını kapsama olasılığı %95 ve bu aralığın dışında olma olasılığı %5 tir. Benzer olarak, kısmi splayn yaklaşımına göre kestirimi yapılan β_s katsayısının %95 güven aralığı,

$$P\left(35.588 \leq \beta_s \leq 40.498\right) = 0.95$$

olarak hesaplanmıştır. Buna göre, incelenen örneklemin çekildiği ana kütlenin β_s regresyon kat sayısının, 35.588 ile 40.498 aralığını kapsama olasılığı %95 ve bu aralığın dışında olma olasılığı ise %5 tir. Kalan diğer parametrik katsayıların %95 güven aralıklarına ilişkin benzer yorumlar yapılabilir.

5.3.2 Parametrik Olmayan Bileşen Hakkında Çıkarsamalar

Eşitlik (1) ile verilen modelin parametrik olmayan bileşeni sayısal olarak özetlenemediğinden, o ancak grafiksel olarak görüntülenebilir. Söz konusu grafikler Şekil 1 ve 2'de görülmektedir. Az sayıda parametre ile özetlenemeyen parametrik olmayan bileşeni, biçimsel olarak değerlendirmek amacıyla *F*-testi yapılmıştır. Bu test yarı parametrik modelin parametrik olmayan bileşeninin, doğrusal veya doğrusal olmayan pürüzsüz bir fonksiyonla temsil edilip edilmemesi konusunda bilgi vermesi bakımından analizin önemli bir kısmını oluşturmaktadır.

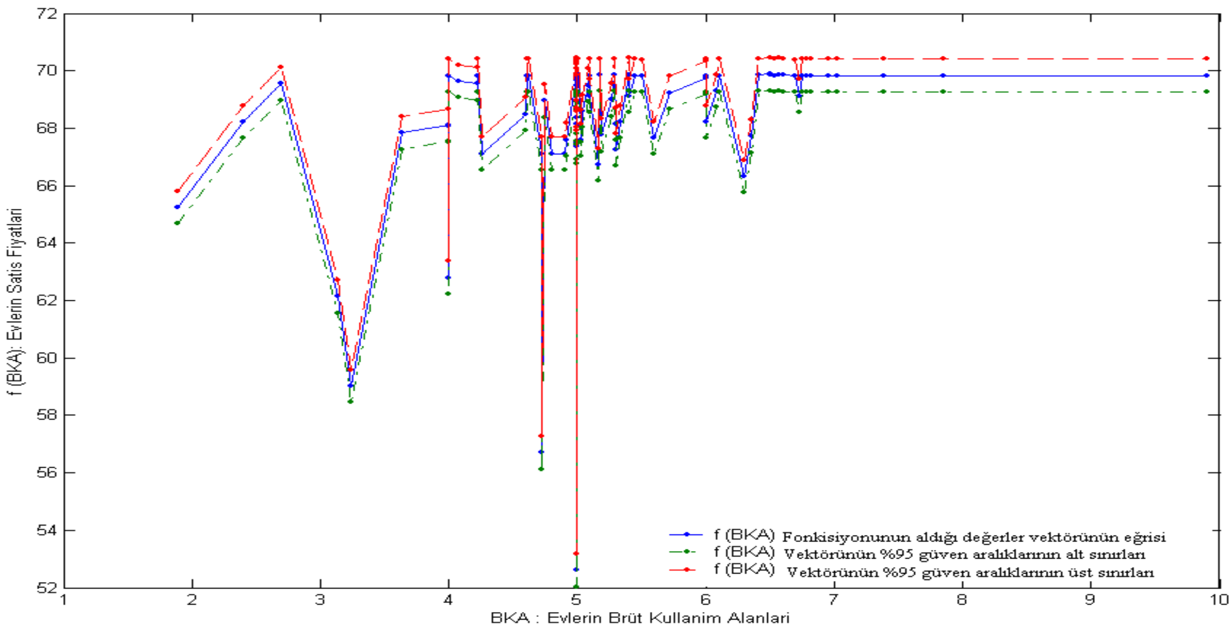
Bölüm 4.2'de incelen F-testi için hipotezler

$$H_0: E(SF) = \beta_0 + \beta_1 AU + \beta_2 FD + \beta_3 GD + \beta_4 KG + \beta_5 NKA$$
$$+ \beta_6 BKA \quad \text{(doğrusal fonksiyon)}$$
$$H_1: E(SF) = f(BKA) \ \text{diğer bir ifadeyle, pürüzsüz fonksiyon}$$

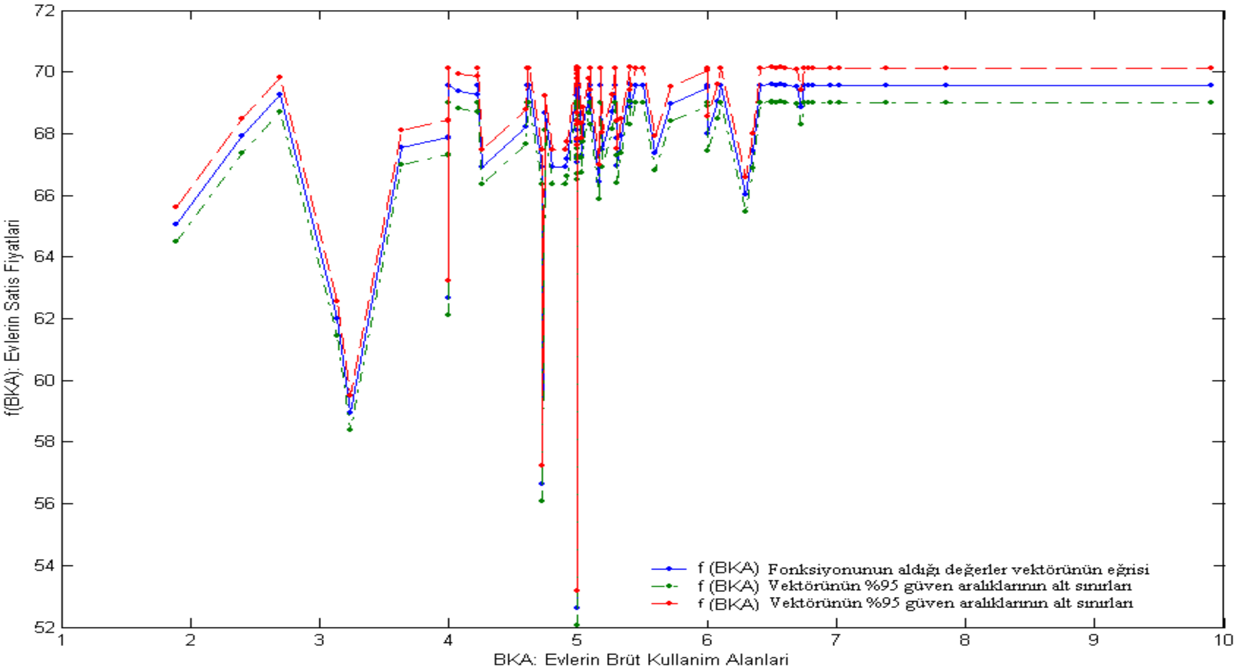
biçiminde ifade edilerek, *F*-test istatistiği ve *F*-tablo değeri aşağıdaki gibi elde edilmiştir:

$$F_{hes}(\mathrm{df}_1 - \mathrm{df}_0 = 3.5568,\ n - \mathrm{df}_1 = 83.4432) = 2870.2$$
$$F_{tab(0.05)} = 2.45$$

Hesaplanan *F*-test istatistiği, *F*-tablo değerinden büyük olduğundan sıfır hipotezi reddedilir. Böylece, yarı parametrik modelin parametrik olmayan bileşe



Şekil 1: Speckman yaklaşımı için splayn düzeltme kestiricisi ve %95 güven aralıkları grafiği



Şekil 2: Kısmi splayn yaklaşımı için splayn düzeltme kestiricisi ve %95 güven aralıkları grafiği

ninin pürüzsüz bir fonksiyon olduğuna karar verilir. Söz konusu parametrik olmayan bileşen istatistiksel açıdan anlamlıdır. Bu durumda, parametrik katsayıların anlamlı olmasın yanı sıra, parametrik olmayan bileşenin de istatistiksel olarak anlamlı bir pürüzsüz fonksiyon olması nedeniyle *yarı parametrik regresyon modeli, istatistiksel açıdan anlamlıdır*

Speckman ve kısmi splayn yaklaşımlarına göre f fonksiyonunun düğüm noktalarındaki $\hat{f}$ tahmin değerleri ve %95 güven aralıklarının grafikleri sırasıyla Şekil 1 ve Şekil 2' de verilmiştir. Şekil 1 Speckman, Şekil 2 ise kısmi splayn yaklaşımına göre evlerin satış fiyatları kestirimlerinin BKA' ya göre değişimleri ve %95 güven aralıkları grafiklerini vermektedir. Her iki grafikte, %95 güven aralıklarının $f(BKA)$ vektörünü çok yakından izlediği görülmektedir. Bu durum yapılan tahminlerin iyi olduğunun bir göstergesidir.

5.4 Parametrik ve Yarı Parametrik Modellerin Karşılaştırılması

Yarı parametrik ve parametrik regresyon modelleri karşılaştırıldığında, parametrik modelde bağımlı değişkendeki değişimlerin %33.287'sinin, yarı parametrik modelde ise %96.546'sının açıklandığı görülmektedir.[3] Diğer taraftan, parametrik modelin hata kareler toplamı 67855 olurken, yarı parametrik modelin, 718.33 olmaktadır. Bunun yanı sıra, parametrik modele ilişkin katsayıların %95 güven aralıklarının, yarı parametrik modelin katsayılarına göre oldukça geniş bir aralığı kapsadıkları görülmek tedir (bak. Tablo 3). Ayrıca parametrik modelde bazı açıklayıcı değişkenlere ilişkin regresyon katsayıları anlamsızken, yarı parametrik modelde regresyon katsayıları anlamlıdır. Bu durum yarı parametrik modelin parametrik modelden çok daha üstün olduğunun bir göstergesidir.

Speckman yaklaşımı, kısmi splayn yaklaşımı ile kıyaslandığında, her iki yöntemin de benzer sonuçlar verdiği görülmüştür. Örneğin, Speckman yaklaşımına göre kestirimi yapılan yarı parametrik modelde, bağımlı değişkendeki değişimlerin %96.555'i açıklanırken, kısmi splayn yaklaşımını esas alan yarı parametrik modelde %96.546'sı açıklamaktadır. Ayrıca Speckman yöntemi için hata kareler toplamı 717.53 olurken, kısmi splayn yaklaşımı için 718.33 olmuştur. Bu durumda, Speckman yaklaşımının az da olsa daha başarılı olduğu söylenebilir.

6. SONUÇ VE TARTIŞMA

Parametrik ve yarı parametrik regresyon modellerinin bazı performans göstergeleri aşağıda verilmiştir:

Parametrik	Kısmi splayn	Speckman
$R^2 = 0.33287$	$R^2 = 0.96546$	$R^2 = 0.96550$
$RSS = 67855$	$RSS = 718.33$	$RSS = 717.53$

Bu sonuçlara göre, yarı parametrik regresyon modellerinin doğrusal regresyon modellerine kıyasla tahminlerde büyük bir iyileşme sağladıkları görülmüştür. Diğer taraftan, yarı parametrik regresyonda bağımlı değişken, açıklayıcı değişkenlerden bazıları ile doğrusal kalan diğer açıklayıcı değişkenlerle doğrusal olmayan bir ilişkiye sahip olduğundan yarı pa-

---

[3] Semiparametrik modelin R^2 belirlilik katsayısı, $R^2 = \hat{\mathbf{y}}^T\hat{\mathbf{y}} / \mathbf{y}^T\mathbf{y}$ formülü ile hesaplanır.

rametrik regresyon modeli, çok daha karmaşık bir ilişki yapısı içerisinde yer alan değişkenler arası ilişkiyi açıklamada hem parametrik hem de parametrik olmayan regresyon modellerinden çok daha üstün olduğu söylenebilir.

KAYNAKLAR

Aydın, D. (2005). *Semiparametrik Regresyon Modellemede Splayn Düzeltme Yaklaşımıyla Tahmin ve Çıkarsamalar*, Doktora Tezi, FBE-Eskişehir.

Chen, H. (1988). Covergance Rates for Parametric Components in a Partially Linear Models. *Annals of Statistics* 16, 136-146.

Engle, R.F., Granger, C.W.J., Rice, C.A. ve Weiss A. (1986). Semiparametric Estimates of the Relation Between Weahter and Electricity Sales. *Journal of Amer. Statist. Assoc.* 81, 310-320.

Eubank, R. L, Kambour, E. L., Kim, T.C., Kipple, K, Reese, S.C. ve Schimek, M. (1998). Estimation in Partiallly Linear Models. *Computational Statistics &Data Analysis* 29, 27-34.

Green, P. J., Jennison, C. ve Seheult, A. (1985). Analysis of Field Experiments By Least Square Smoothing. *J. Roy. Statis. Soc.* 47(B), 299-315.

Green, P. J. ve Silverman, B.W. (1994). *Nonparametric Regression and Generalized Linear Models.* Chapman &Hall, London.

Hastie, T. ve Tibshirani, R. J. (1990). *Generalized Additive Models*. Chapman & Hall, London.

Ho, M. (1995). Essay On The Houshing Market, Unpublished Ph.D. Dissertation University Of Toronto.

Lee, T. C. M. (2003). Smoothing Parameter Selection For Smoothing Splines: A Simulation Study. *Comput. Statistisc & Data Analysis* 42, 139-148.

Lee, T. C. M. (2004). Improved Smoothing Spline Regression By Combining Estimates of Different Smoothness. *Statistiscs & Probability Letters* 67, 133-140.

Mammadov, M. Yüzer, A. F. ve Aydın, D. (2005). Splayn Düzeltme Regresyonu ve Düzeltme Parametresinin Seçimi, 4. İstatistik kongresi bildiri ve poster özetleri kitabı, Belek- Antalya, 148-149.

Rice, J. (1986). Covergance Rates For Partially Spline Models. *Statis. Prob. Lett.* 4, 203-208.

Robinson, M. P. (1988). Root-N-Consistent Semiparametric Regression. *Econometrica* 56, 4, 931-954.

Schimek, G. Michael. (2000). Estimation and İnference in Partially Linear Models with Smoothing Splines. *Journal of Statistical Planning and Inference* 91, 525-540.

Speckman, P. (1988). Kernel Smoothing in Partially Linear Model, *J. Royal Statist. Soc.* 50(B), 413-436.

Wahba, G. (1990). *Spline Models for Observational Data.* SIAM, Philadelphia, PA.

Dursun AYDIN, 16.02.1969 Koyunpınarı-Hanak / ARDAHAN doğumludur. İlk ve orta öğrenimini Koyunpınarı köyünde, lise öğrenimini ise Hanak'ta tamamladı. Eskişehir Anadolu Üniversitesi, Fen Edebiyat Fakültesi, İstatistik Bölümü'nden 1992 yılında mezun oldu ve 1994 yılında Anadolu Üniversitesi Açık Öğretim Fakültesi'nde Öğretim Görevlisi olarak işe başladı ve 1994-1999 yılları arasında Edirne'de görev yaptı. 1999 yılında Marmara Üniversitesi, Sosyal Bilimler Enstitüsü, Ekonometri Anabilim Dalı, İstatistik Bilim Dalında yüksek lisansını ve Kasım 2005'de Anadolu Üniversitesi, Fen bilimleri Enstitüsü, İstatistik Bilim Dalında Doktora öğrenimini tamamladı ve halen Anadolu Üniversitesi Bilecik MYO'de Öğr. Gör. Dr. olarak görev yapmaktadır.

Mammadagha MAMMADOV, Azerbaycan 1947 doğumlu olup Bakü Devlet Üniversitesi'nden 1971 yılında mezun oldu. 1977 yılında Rusya Bilim Akademisi'nin Doktora ünvanını, 1985'de ise Baş Bilim Adamı ünvanını (diplomasını) kazandı. 1977-1990 yılları arasında Azerbaycan Bilim Akademisi Sibernetik Enstitüsü'nde Baş Bilim Adamı olarak, 1991-1998 yılları arasında ise Bakü Devlet Üniversitesi'nde Doçent olarak görev yapmıştır. 1999-2002 yıllarında Çanakkale 18 Mart Üniversitesi Bilgisayar Bölümü'nde Doçent olarak çalışmıştır. 2003 yılından itibaren ise Anadolu Üniversitesi İstatistik Bölümü'nde Doçent olarak çalışmaktadır. Diferansiyel Oyun Teorisi, Yapay Sinir Ağları, Kontrol Teori, Nonparametrik ve Semiparametrik Regresyon Analizi alanlarında, yurtiçi ve yurtdışında çeşitli bilimsel dergilerde yayımlanmış elliden fazla makalesi bulunmaktadır.

Ali Fuat YÜZER, 1944 yılında Eskişehir'de doğdu. 1969 yılında Eskişehir İktisade ve Ticari İlimler Akademisi'ni bitirdi. 1970 yılında aynı akademinin Uygulamalı Matematik ve İstatistik Bölümünde asistan olarak göreve başladı. 1978 yılında doktor, 1985 yılında doçent oldu. 1996 yılında Eskişehir Anadolu üniversitesinde profesör oldu. Halen Anadolu Üniversitesi İstatistik Bölümü'nde Teorik İstatistik Ana Bilim Dalı Başkanı olarak görev yapmaktadır.

Embiya AĞAOĞLU, Şumnu'da 1949 yılında doğdu. EİTİA İstatistik ve Uygulamalı Matematik kürsüsünde doktora derecesini aldı. 1993 yılında Anadolu Üniversitesi Fen Fakültesi İstatistik Bölümü'nde Doçent ve aynı bölümde 1998 yılında Profesör olan E. Ağaoğlu halen Anadolu Üniversitesi, İstatistik Bölümü'nde Bölüm Başkanı olarak çalışmalarına devam etmektedir. Evli ve bir çocuğu olan E. Ağaoğlu'nun ulusal ve uluslar arası çalışmaları bulunmaktadır.