ANADOLU ÜNİVERSİTESİ BİLİM VE TEKNOLOJİ DERGİSİ
ANADOLU UNIVERSITY JOURNAL OF SCIENCE AND TECHNOLOGY
Cilt/Vol.:8-Sayı/No: 1 : 153-159 (2007)

ARAŞTIRMA MAKALESİ /**RESEARCH ARTICLE**

SEMİPARAMETRİK TOPLAMSAL REGRESYON MODELİ İLE TAHMİN: ESKİŞEHİR'DEKİ EVLERİN KİRA FİYATLARI VE ÖZELLİKLERİ ARASINDAKİ İLİŞKİLERİN ANALİZİ

Rabia Ece OMAY[1], Dursun AYDIN[2], Mammadagha MAMMADOV[3]

ÖZ

Çalışmada Eskişehir merkezde yer alan evlerin kira fiyatları ile evlerin özelikleri arasındaki ilişkilerin incelenmesinde farklı regresyon modelleri ele alınmıştır. Yapılan istatistiksel analizler sonucunda, evlerin kira fiyatları üzerinde etkili olan bağımsız değişkenlerin bir kısmının fiyatları doğrusal, bir kısmının da doğrusal olmayarak etkilediği görülmüştür. Böylece model, parametrik doğrusal bileşenlere ilaveten birkaç nonparametrik bileşeni de bulunduran semiparametrik toplamsal regresyon modeli şeklinde oluşturulmuştur. Elde edilen uygun semiparametrik toplamsal regresyon modelin istatistiksel açıdan anlamlı olduğu, hem parametrik doğrusal hem de semiparametrik modellerden daha iyi sonuçlar verdiği gözlenmiştir.

Anahtar Kelimeler : Toplamsal model, Semiparametrik toplamsal model, Semiparametrik model, Splayn düzeltme

ESTIMATION WITH SEMIPARAMETRIC ADITTIVE REGRESSION MODEL: ANALYSIS OF RELATIONSHIPS AMONG HOUSE RENTS AND TRAIT VARIABLES IN ESKİŞEHİR

ABSTRACT

In this paper, different regression models have been discussed for investigation of the relationships among house rents and trait variables in Eskişehir. According to statistical analysis, it is concluded that some of the explanatory variables have had linear and some of them have had nonparametric effect on house rents. Thus, the model has obtained as semiparametric adittive regression model that contain a few nonparametric components in additon to parametric linear components. It is observed that suitable semiparametric adittive regression model is significant, and given better results than parametric linear and semiparmetric models.

Keywords: : Adittive models, Semiparametric adittive model, Semiparametric model, Spline smoothing

[1,] Anadolu Üniversitesi, Fen Fakültesi, İstatistik Bölümü, ESKİŞEHİR
E-posta: reayar@anadolu.edu.tr
[2,] Anadolu Üniversitesi, Bilecik Meslek Yüksekokulu, BİLECİK
E-posta: duaydin@anadolu.edu.tr
[3,] Anadolu Üniversitesi, Fen Fakültesi, İstatistik Bölümü, ESKİŞEHİR
E-posta: mmammadov@anadolu.edu.tr

Geliş: 03 Ağustos 2006; **Düzeltme:** 11 Ekim 2006; **Kabul:** 08 Kasım 2006

1. GİRİŞ

Çalışmada parametrik ve nonparametrik bileşenleri içeren semiparametrik (kısmi parametrik) toplamsal (adittive) regresyon modeli ele alınmıştır. Böyle bir modelin kestirimi için splayn düzeltme yöntemi uygulanmıştır. İncelenen problemde, uygun düzeltme parametrelerinin seçilmesi ve düzeltme matrislerinin yardımıyla splayn düzeltme kestiricilerinin tahmin edilmesi, toplamsal modellerin kestirimi için temel oluşturmaktadır. Verilen düzeltme parametreleri için bu kestiriciler, yaygın kullanılan backfitting algoritması ile elde edilebilir (Green ve Silverman, 1994; Hastie ve Tibshirani, 1999). Düzeltme parametresinin seçimi için ise otomatik bir seçim yöntemi olan *Genelleştirilmiş Çapraz Geçerlilik* (GCV) kullanılır. Fakat toplamsal modellerde, birden çok fonksiyonu eş zamanlı olarak minimum yapmak çok zor olduğundan, düzeltme parametresinin seçimi, serbestlik derecesi miktarı belirlenerek yapılabilir.

Bir x açıklayıcı ve bir y bağımlı değişkenin $x_i, y_i \ {}_{i=1}^{n}$ gözlem değerlerinin yer aldığı, *nonparametrik regresyon modeli* aşağıdaki şekilde tanımlanır:

$$y_i = f(x_i) + \varepsilon_i, \quad a < x_1 < ... < x_n < b, \quad \varepsilon_i \sim N(0, \sigma^2) \quad \textbf{(1.1)}$$

Burada, $f \in C^2[a,b]$, bilinmeyen pürüzsüz fonksiyon ve ε_i rassal hata terimleridir. Nonparametrik regresyonda temel amaç (1.1) modelindeki bilinmeyen $f \in C^2[a,b]$ fonksiyonunun tahminidir. Belirli bir $\lambda > 0$ için (1.1) modelinin *splayn düzeltmeye dayalı çözümü*,

$$S(f) = \sum_{i=1}^{n} y_i - f(x_i)^2 + \lambda \int_a^b f''(x)^2 dx \quad \textbf{(1.2)}$$

eşitliği ile belirtilen $S(f)$ *cezalı hata kareler toplamını* minimum yapan bir $\hat{f} \in C^2[a,b]$ fonksiyonu olarak tanımlanır (Wahba, 1990; Greeen ve Silverman, 1994).

Eşitlik (1.2) ile verilen minimum probleminin splayn düzeltmeye dayalı çözümü, $x_1, ..., x_n$ düğümleri ile bir *"doğal kübik splayn"* olarak bilinir (Green ve Silverman, 1994). $\mathbf{y} = (y_1, ..., y_n)^T$ verilen gözlem değerleri vektörü olsun. (1.2) denkleminin çözümünün bir kübik splayn olduğu gerçeğini ve x_i düğüm noktalarında $f(x_i)$ değerler vektörünün $\mathbf{f} = (f_1, ..., f_n)^T = (f(x_1), ..., f(x_n))^T$ olduğunu kullanarak, (1.2) denklemi aşağıdaki şekilde ifade edilebilir:

$$(\mathbf{y} - \mathbf{f})^T (\mathbf{y} - \mathbf{f}) + \lambda \mathbf{f}^T \mathbf{K} \mathbf{f} \quad \textbf{(1.3)}$$

Burada $\mathbf{K}$, belli bir karesel ceza matrisidir. (1.3) denkleminin çözümü,

$$\hat{\mathbf{f}} = (I + \lambda \mathbf{K})^{-1} \mathbf{y} = \mathbf{S}_\lambda \mathbf{y} \quad \textbf{(1.4)}$$

şeklinde ifade edilen vektördür. Burada $\mathbf{S}_\lambda$, verilen bir $\lambda > 0$ düzeltme parametresi ve $x_1, ..., x_n$ düğüm noktaları olarak bilinen nonparametrik bir kestirici değişkenin gözlem değerleri yardımıyla hesaplanan bir *düzeltme matrisidir*. (1.4)'deki $\hat{\mathbf{f}}$ tahmin vektörü, (1.2) eşitliğini minimum yapan $\hat{f} \in C^2[a,b]$ fonksiyonun x_i noktalarında aldığı değerler vektörüdür (Wahba, 1990; Greeen ve Silverman, 1994).

Semiparametrik regresyon modelleri, bağımlı değişkenin bazı açıklayıcı değişkenlerle *doğrusal*, diğer açıklayıcı değişkenlerle ise *doğrusal olmayan* ilişki içerisinde olduğu regresyon modellerdir ve bu modeller genel olarak aşağıdaki gibi ifade edilebilir.

$$y_i = \mathbf{z}_i^T \beta + f(x_i) + \varepsilon_i, \quad i = 1, 2, ..., n \text{ veya}$$
$$\mathbf{y} = \mathbf{Z}\boldsymbol{\beta} + \mathbf{f} + \boldsymbol{\varepsilon}. \quad \textbf{(1.5)}$$

Burada $\mathbf{z}_i$, parametrik kısma karşılık gelen bağımsız değişkenlerin k boyutlu i.gözlemler vektörü; β, k boyutlu regresyon katsayıları vektörüdür.

Eşitlik (1.5) ile verilen semiparametrik modelin uyumunu elde etmek için, β *parametre vektörünü,* $f \in C^2$ *a, b fonksiyonu ve* $\boldsymbol{\mu} = \mathbf{x}\beta + \mathbf{f}$ *ortalama vektörünü tahmin etmek* gerekir. Bunun için farklı düzeltme tekniklerine dayalı birkaç yaklaşım önerilmiştir. Bu yaklaşımlardan biri de *splayn düzeltme yöntemdir* (Engle vd, 1986; Wahba, 1990; Greeen ve Silverman, 1994; Green vd, 1985).

2. TOPLAMSAL REGRESYON MODELLERİNİN TAHMİN DENKLEMLERİ

Bir *toplamsal regresyon modeli*,

$$y_i = \sum_{j=1}^{p} f_j(x_{ji}) + \varepsilon_i, \ i = 1, ..., n, \ \varepsilon_i \sim N(0, \sigma^2) \text{ veya}$$
$$\mathbf{y} = \sum_{j=1}^{p} \mathbf{f}_j + \boldsymbol{\varepsilon} \quad \textbf{(2.1)}$$

biçiminde tanımlanır (Hastie ve Tibshirani, 1999). Eşitlik (2.1)'de f_j 'ler bilinmeyen tek değişkenli fonksiyonlardır, $\mathbf{f}_j = (f_j(x_{j1}), ..., f_j(x_{jn}))^T, \quad j = 1, 2, ..., p$ ise f_j fonksiyonunun düğüm noktalarındaki değerleri vektörüdür.

p tane nonparametrik bileşene sahip olan, semiparametrik toplamsal regresyon modeli ise aşağıdaki gibi tanımlanır:

$$y_i = \mathbf{z}_i^T\boldsymbol{\beta} + f_1(x_{1i}) + ... + f_p(x_{pi}) + \varepsilon_i,\ i = 1,2,...,n$$

$$\text{veya } \mathbf{y} = \mathbf{Z}\boldsymbol{\beta} + \sum_{j=1}^{p}\mathbf{f}_j + \boldsymbol{\varepsilon} \qquad \textbf{(2.2)}$$

Eşitlik (2.1) toplamsal regresyon modelinin tahmini için splayn düzeltme yaklaşımı uygulandığında, ikinci mertebeden sürekli türevi olan tüm f_j, $j = 1,2,...,p$ fonksiyonlar uzayında, aşağıdaki *genelleştirilmiş cezalı hata kareler toplamının* minimizasyonu problemi ele alınır:

$$\sum_{i=1}^{n}\left\{ y_i - \sum_{j=1}^{p} f_j(x_{ij}) \right\}^2 + \sum_{j=1}^{p}\lambda_j \int \left[f_j''(x) \right]^2 dx \quad \textbf{(2.3)}$$

(2.3) ifadesinin ceza kısmındaki her fonksiyon, seçilen bir λ_j *düzeltme parametresine* bağlıdır. Bu parametre, uygun fonksiyonun çözümdeki düzeltme katkısını belirtir.

Eşitlik (1.3) ile verilen tek bir kestirici f fonksiyonunun olduğu duruma benzer olarak, (2.3) problemi aşağıdaki şekilde yazılabilir:

$$\left(\mathrm{y} - \sum_{j=1}^{p}\mathrm{f}_j \right)^T \left(\mathrm{y} - \sum_{j=1}^{p}\mathrm{f}_j \right) + \sum_{j=1}^{p}\lambda_j \mathrm{f}_j^T \mathrm{K}_j \mathrm{f}_j \qquad \textbf{(2.4)}$$

Burada K_j, *uygun kestiricinin ceza matrisidir* ve tek bir kestirici durumundaki K matrisine benzer olarak tanımlanır. (2.4) ifadesi, f_j, $j = 1,2,...,p$ vektörlerine göre bir kare formdur. Bu ifadenin f_j'lere göre türevini sıfıra eşitleyerek,

$$2\lambda_j \mathbf{K}_j \mathbf{f}_j - 2(\mathbf{y} - \sum_j \mathbf{f}_j) = \mathbf{0} \qquad \textbf{(2.5)}$$

denklemler sistemi elde edilir. (2.5) eşitliği, *tahmin denklemleri* olarak adlandırılan aşağıdaki gibi bir $np \times np$ sistemi şeklinde yazılabilir:

$$\begin{pmatrix} \mathrm{I} & \mathrm{S}_1 & \mathrm{S}_1 & \cdots & \mathrm{S}_1 \\ \mathrm{S}_2 & \mathrm{I} & \mathrm{S}_2 & \cdots & \mathrm{S}_2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathrm{S}_p & \mathrm{S}_p & \mathrm{S}_p & \cdots & \mathrm{I} \end{pmatrix} \begin{pmatrix} \mathrm{f}_1 \\ \mathrm{f}_2 \\ \vdots \\ \mathrm{f}_p \end{pmatrix} = \begin{pmatrix} \mathrm{S}_1\mathrm{y} \\ \mathrm{S}_2\mathrm{y} \\ \vdots \\ \mathrm{S}_p\mathrm{y} \end{pmatrix} \qquad \textbf{(2.6)}$$

Burada $\mathrm{S}_j = \mathrm{S}_{\lambda_j} = \left(\mathrm{I} + \lambda_j \mathrm{K}_j \right)^{-1}$ uygun düzeltme matrisidir. (2.6) sistemi kısaca $\hat{\mathbf{P}}\mathbf{f} = \hat{\mathbf{Q}}\mathbf{y}$ olarak yazılabilir.

Eşitlik (2.5)'ten görülüyor ki, (2.3)'ün çözümünden elde edilen her bir $\hat{f}_k$ tahmin fonksiyonu,

$$\hat{\mathrm{f}}_k = \mathrm{S}_k \left(\mathrm{y} - \sum_{j \neq k} \hat{\mathrm{f}}_j \right),\ k = 1,2,...,p \qquad \textbf{(2.7)}$$

doğrusal düzeltici (splayn düzeltme) yardımıyla hesaplanan bir kübik splayndır. (2.7) ve (2.6) sistemleri denk sistemlerdir. (2.7) formülü, *backfitting algoritması*nın uygulanması için (2.6) sisteminin uygun bir şeklidir.

İlgilenilen örnek uygulama probleminde sadece iki nonparametrik açıklayıcı değişkenin kullanılması nedeniyle, bu çalışmada özel olarak iki lineer düzeltici içeren toplamsal model için backfitting algoritması incelenmiştir. Bu durumda (2.6) veya (2.7) sistemi aşağıdaki gibi yazılabilir:

$$\begin{aligned} \mathrm{f}_1 &= \mathrm{S}_1(\mathrm{y} - \mathrm{f}_2) \\ \mathrm{f}_2 &= \mathrm{S}_2(\mathrm{y} - \mathrm{f}_1) \end{aligned} \qquad \textbf{(2.8)}$$

Backfitting algoritmasının *m.* adımındaki tahminleri $\mathrm{f}_1^{\,m}$ ve $\mathrm{f}_2^{\,m}$ olsun. Başlangıç adım için $\mathrm{f}_1^{\,0}$ ve $\mathrm{f}_2^{\,0}$ tanımlanır. Backfitting, (2.6) sisteminin çözümünü bulmak için *Gaus-Seidel* prosedürü ile aşağıdaki tekrarlama (recursion) şeklinde gerçekleştirilir:

$$\begin{aligned} \mathrm{f}_1^{\,m} &= \mathrm{S}_1(\mathrm{y} - \mathrm{f}_2^{\,m-1}) \\ \mathrm{f}_2^{\,m} &= \mathrm{S}_2(\mathrm{y} - \mathrm{f}_1^{\,m}) \end{aligned} \qquad \textbf{(2.9)}$$

$\mathrm{f}_1^{\,m}$ ve $\mathrm{f}_2^{\,m}$'nin yakınsaması için $\left\| \mathrm{S}_1\mathrm{S}_2 \right\|_2$ normu 1'den küçük olmalıdır: $\left\| \mathrm{S}_1\mathrm{S}_2 \right\|_2 < 1$ [5]. Bu durumda (2.9)'un çözümü

$$\begin{aligned} \mathrm{f}_1^{\,\infty} &= \left[\mathrm{I} - (\mathrm{I} - \mathrm{S}_1\mathrm{S}_2)^{-1}(\mathrm{I} - \mathrm{S}_1) \right] \mathrm{y} \\ \mathrm{f}_2^{\,\infty} &= \mathrm{S}_2(\mathrm{I} - \mathrm{S}_1\mathrm{S}_2)^{-1}\mathrm{y} = \left[\mathrm{I} - (\mathrm{I} - \mathrm{S}_2\mathrm{S}_1)^{-1}(\mathrm{I} - \mathrm{S}_2) \right] \mathrm{y} \end{aligned} \qquad \textbf{(2.10)}$$

olur ve

$$\hat{\mathrm{y}} = \mathrm{f}_1^{\,\infty} + \mathrm{f}_2^{\,\infty} = \left[\mathrm{I} - (\mathrm{I} - \mathrm{S}_2)(\mathrm{I} - \mathrm{S}_1\mathrm{S}_2)^{-1}(\mathrm{I} - \mathrm{S}_1) \right] \mathrm{y} \qquad \textbf{(2.11)}$$

elde edilir. (2.11) eşitliği S_1 ve S_2'ye göre simetrik olup kolaylıkla hesaplanabilir. Eğer $\left\| \mathrm{S}_1\mathrm{S}_2 \right\|_2 < 1$ ise (2.6) tahmin denklemi tutarlıdır, çözüm tektir ve (2.9) backfitting algoritması çözüme yakınsamaktadır (Hastie ve Tibshirani, 1999). S_1 ve S_2 simetrik matrislerinin öz değerleri (-1,1] aralığına ise, (2.6) tahmin denklemi en az bir çözüme sahip olur ve (2.9) backfitting algoritması bu çözümlerden birine yakınsar. Bu durumda çözüm, başlangıç $\mathrm{f}_2^{\,0}$ durumuna bağımlı olur.

İki nonparametrik bileşene sahip *semiparametrik* bir model ele alındığında, bu model aşağıdaki gibi ifade edilecektir:

$$y_i = \mathbf{z}_i^T\beta + f_1(x_{1i}) + f_2(x_{2i}) + \varepsilon_i \ , \ \text{i=1,2,...,n}$$
$$\text{veya} \quad \mathbf{y} = \mathbf{Z}\boldsymbol{\beta} + \mathbf{f}_1 + \mathbf{f}_2 + \boldsymbol{\varepsilon} \qquad \textbf{(2.12)}$$

Burada $\mathrm{y}, \mathrm{f}_1, \mathrm{f}_2$ ve ε n-boyutlu sütun vektörleri, $\mathbf{Z}$ $n \times k$ boyutlu matris ve β k-boyutlu katsayı vektörüdür. (2.1) regresyon problemine splayn düzeltme yöntemi uygulandığında, (2.7) denklemler sistemi aşağıdaki şekilde yazılabilir:

$$\begin{aligned} \mathrm{f}_0 &= \mathrm{S}_0(\mathrm{y} - \mathrm{f}_1 - \mathrm{f}_2) \\ \mathrm{f}_1 &= \mathrm{S}_1(\mathrm{y} - \mathrm{f}_0 - \mathrm{f}_2) \\ \mathrm{f}_2 &= \mathrm{S}_2(\mathrm{y} - \mathrm{f}_1 - \mathrm{f}_0) \end{aligned} \qquad \textbf{(2.13)}$$

Burada $\mathbf{S}_0 = \mathbf{Z}(\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{Z}^T$ matrisi doğrusal parametrik kısmın düzeltici matrisidir ve $\mathbf{f}_0 = \mathbf{Z}\boldsymbol{\beta}$, parametrik terimin kestiricisidir. (2.13) denklemlerine uygun backfitting algoritması aşağıdaki gibi elde edilir:

$$\begin{aligned} \mathrm{f}_0^{\,m} &= \mathrm{S}_0(\mathrm{y} - \mathrm{f}_1^{\,m-1} - \mathrm{f}_2^{\,m-1}) \\ \mathrm{f}_1^{\,m} &= \mathrm{S}_1(\mathrm{y} - \mathrm{f}_0^{\,m} - \mathrm{f}_2^{\,m}) \\ \mathrm{f}_2^{\,m} &= \mathrm{S}_2(\mathrm{y} - \mathrm{f}_1^{\,m} - \mathrm{f}_0^{\,m}) \end{aligned} \qquad \textbf{(2.14)}$$

3. TOPLAMSAL REGRESYON MODELLERİ İÇİN ÇIKARSAMALAR

Eşitlik (1.7) ile verilen modeli değerlendirebilmek için hem parametrik bileşenler hem de nonparametrik bileşenler üzerinde testler yapmak gerekir. Bu amaçla, izleyen alt başlıklarda semiparametrik toplamsal modelin değerlendirmesinde kullanılan bazı temel kavramların tanıtılmasına yer verilmiştir.

3.1 Sapma

İlgilenilen modelin uyum iyiliğinin (goodness of fit) testi ve modelleri karşılaştırmanın bir yolu, tahmin edilebilecek maksimum parametreyi içeren *doymuş (saturated) modelle*, ilgilenilen modeli karşılaştırmaktadır. İlgilenilen model ile doymuş modelin maksimize edilmiş log-olabilirlik değerlerinin oranına dayanan sapma (deviance) değeri,

$$D(\mathbf{y};\mathbf{b}) = 2\ l(\mathbf{b}_{\max};\mathbf{y}) - l(\mathbf{b};\mathbf{y}) \qquad \textbf{(3.1)}$$

olarak tanımlanır. Burada, b_{max} doymuş model için parametre vektörü β_{max}'ın maksimum olabilirlik tahmincisi, $l(\mathbf{b}_{\max}, \mathbf{y})$ doymuş modelin olabilirlik fonksiyonu ve $l(\mathbf{b}, \mathbf{y})$ ilgilenilen model için olabilirlik fonksiyonunun maksimum değerini gösterir.

Eşitlik (3.1) ile verilen sapma değeri yaklaşık bir χ^2 dağılımı gösterir. Sapma değeri en küçük olan model, verileri en iyi açıklayan model olarak seçilmektedir. Nonparametrik ve toplamsal modeller için sapma, modelleri ve bu modellerin farklarını değerlendirmek için kullanılır. Fakat farkların dağılım teorisi geliştirilmemiş olmasına karşın, χ^2 dağılımı, modelleri karşılaştırmak için bir referans dağılım olarak kullanılır (Hastie ve Tibshirani, 1999).

3.2 Serbestlik derecesi

Farklı düzelticileri veya modelleri karşılaştırabilmek için *etkin parametre sayısı* ya da *serbestlik derecesi* (*degrees of freedom– df*) kullanılabilir. Gerçekte, bir düzeltici için serbestlik derecesini (*df*) belirleyerek basit olarak düzeltme parametresinin değerini seçmek mümkündür (Hastie ve Tibshirani, 1999). Bir değişkenli nonparametrik bileşen için serbestlik derecesi, λ düzeltme parametresine bağlı olarak hesaplanan bir $\mathbf{S}_\lambda$ düzeltme matrisinin izidir: $df = tr(\mathbf{S}_\lambda)$. Birden çok düzeltme gerektiren nonparametrik kestirici değişken olması durumunda, *model için toplam serbestlik derecesi*,

$$df = tr(\mathbf{R}_\lambda)$$

biçiminde tanımlanır. Burada yer alan $\mathbf{R}_\lambda$, $\hat{\mathbf{f}}_+ = \mathbf{R}_\lambda\mathbf{y}$ tahmin vektörleri toplamını ($\hat{\mathbf{f}}_+ = \sum_{j=1}^{p}\hat{\mathbf{f}}_{\lambda_j}$) üreten bir düzeltme matrisidir.

3.3 Düzeltme parametresinin seçimi

Teoride, bir değişkenli fonksiyon için geçerli olan seçim teknikleri, (genelleştirilmiş çapraz geçerlilik (GCV), Akaike bilgi kriteri (AIC) gibi) düzeltme parametresinin seçimi için toplamsal modeller ortamına genişletilir. Özellikle GCV ve AIC gibi klasik model seçme kriterleri $\lambda_j, j = 1,..., p$ düzeltme parametrelerinin seçimi için tasarlanabilir (Wood, 2000). $p-$terimli bir toplamsal modelde, p tane λ_j düzeltme parametresi GCV gibi bir kriteri optimum yapan değer olarak düşünülebilir.

Bir toplamsal model için GCV kriteri,

$$GCV(\lambda_1,...,\lambda_p) = \frac{\sum_{i=1}^{n}\left\{y_i - \sum_{j=1}^{p}\hat{f}_{\lambda_j}(x_{ij})\right\}^2}{n\ 1 - tr\mathbf{R}(\lambda_1,...,\lambda_p)/n\ ^2} \qquad \textbf{(3.3)}$$

olarak tanımlanır. Burada $\mathbf{R}(\lambda_1,...,\lambda_p)$, verilen düzeltme parametrelerinin değerleri için toplamsal uyum

operatörünü ve $\hat{f}_{\lambda_j}$ terimleri uyum fonksiyonlarını gösterir.

Toplamsal modellerde birden çok düzeltme parametresinin optimum seçim probleminin zor olması ve bu parametrelerin serbestlik derecesiyle doğrudan ilişkili olması nedeniyle, uygulamada serbestlik derecesi değiştirilerek uygun bir model seçilebilir. S-Plus gam() toplamsal model fonksiyonu her bir toplamsal bileşen için başlangıçta serbestlik derecesinin değerini 3 olarak kabul eder. Bu bir makul başlangıç noktasıdır ancak, bu serbestlik derecesi her zaman kullanılan düzeltme miktarı olmayabilir (Ruppert, 2003)

4. UYGULAMA

Uygulamada, Eskişehir merkezde bulunan evlerin kira fiyatlarını etkileyen ve bağımsız değişkenler olarak dikkate alınan evlerin özelliklerinin fiyatlar üzerindeki etkileri incelenmiştir. Değişkenlerden bazılarının fiyat ile (doğrusal veya doğrusal olmayan) ilişkisinin şekli önceden bilinmeyebilir. Çalışmanın bu bölümünde, böyle bir ilişki şekli incelenerek ve diğer regresyon modelleri ile karşılaştırılarak, *uygun bir semiparametrik toplamsal regresyon modeli* belirlenmiştir. Yapılan istatistiksel analizlerle onun anlamlılığı değerlendirilmiştir.

Çalışmada kullanılan veriler, Eskişehir merkezde ikiden çok katlı binalarda yer alan kiralık evlerden tarafımızca elde edilmiştir. Ele alınan veriler, 2006 mayıs ayı içersinde 108 kiralık evin kira fiyatları ve karakteristiklerini gösteren değişkenlere ilişkin gözlem değerlerinden oluşmaktadır. Söz konusu değişkenler aşağıdaki gibi tanımlanır:

Fiyat: Evlerin kira fiyatları (YTL)
Odas: Evlerde bulanan oda sayıları
Dakat: Bina içerisinde evlerin kaçıncı katta yer aldığı
Katsay: Evlerin bulunduğu binadaki kat sayısı
Kombi: Evlerde kombi sistemi olup olmadığını götseren dummy değişken
Depozito: Evlerin kiralanması durumunda kiracıdan alınan depozito (YTL)
Yas: Evlerinin yaşı

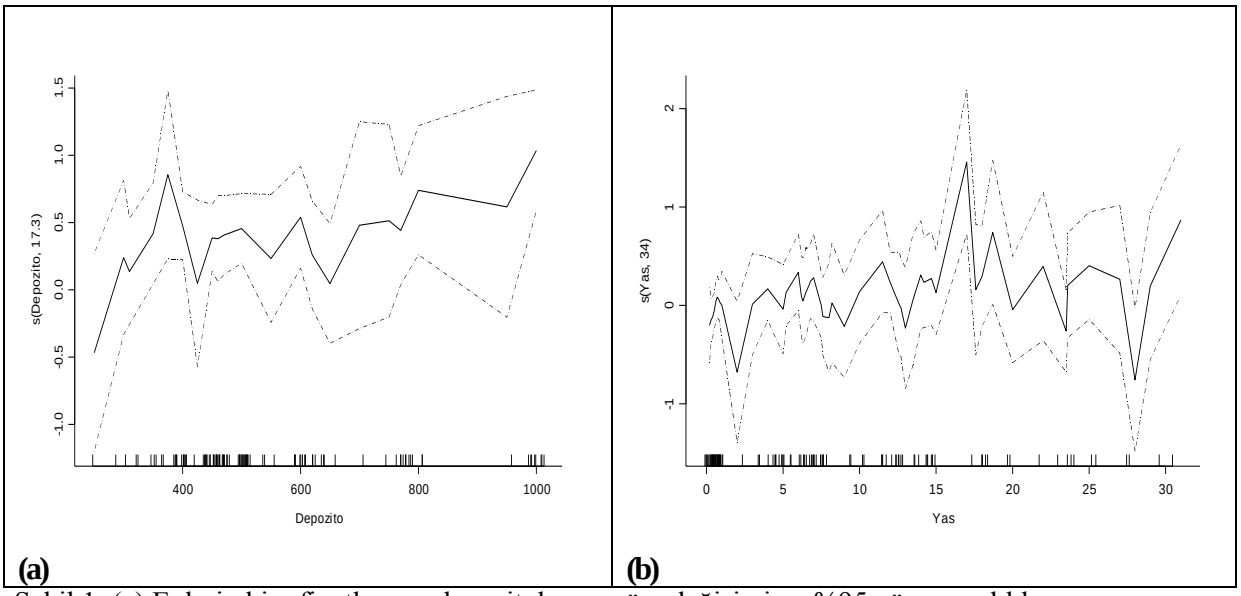
İfade edilen bu değişkenlerden, Fiyat, Depozito ve Yas değişkenleri sürekli değişkenler, Odas, Dkat ve Katsay değişkenleri kesikli değişkenlerdir. Kombi değişkeni ise evlerde kombi sistemi olup olmadığını gösteren dummy değişkendir.

Uygulamada ele alınan modeli değerlendirmek için, oluşturulan *uygun semiparametrik toplamsal modelle* yapılan tahmin sonuçları, değişkenlerin tamamının doğrusal olarak yer aldığı çok değişkenli *parametrik doğrusal regresyon modeli* ve hem parametrik (düzeltme gerektirmeyen dummy değişken parametrik kısımda yer alır (Bkz. Omay, Aydın, ve Mammadov, 2006) hem de nonparametrik değişkenleri içeren *semiparametrik regresyon modeli* ile yapılan tahmin sonuçlarıyla karşılaştırılmıştır. Yapılan istatistiksel değerlendirmelerde ***R*** ve ***S-Plus*** paket programlarından yararlanılmış ve *semiparametrik toplamsal* modelin kestirimlerine ilişkin bazı çıkarsamalara yer verilmiştir.

Semiparametrik toplamsal regresyon modelinin ayrıntıları: Uygun semiparametrik toplamsal modeli belirlemek amacıyla, dummy ve kesikli değişkenlerin dışında kalan mevcut bağımsız değişkenlerden bazıları nonparametrik, bazıları ise parametrik olarak ele alınarak, yukarıda adı geçen üç regresyon modeli oluşturulmuştur. Elde edilen bu modeller arasından seçilen en iyi modelle diğer bir ifadeyle uygun modelle yapılan tahmin sonuçları Tablo1'de verilmiştir.

Tablo1. Semiparametrik toplamsal regresyon sonuçları

	Nonparametrik Kısım				Parametrik Kısım			
	Sd Npar	Sd Npar	F	Pr (F)	Katsayılar	St. Hata	t-ist.	Pr ($>\|t\|$)
Odas	1				0.2602	0.0291	8.948067	1.26e-12
Dakat	1				-0.0288	0.0166	-1.730918	5.86e-02
Katsay	1				0.0772	0.0162	4.742307	1.35e-05
Kombi	1				0.1032	0.0505	2.046119	4.52e-02
S(Depozito)	1	17.3	327.12	2.2e-16	-	-	-	-
S(Yas)	1	34	216.51	2.2e-16	-	-	-	-
	Bağımlı değişken: log (Fiyat)				R^2 = 0.7718312		Deviance (sapma) = 4.0313	



Şekil 1: (a) Evlerin kira fiyatlarının depozitolarına göre değişimi ve %95 güven aralıkları
(b) Evlerin kira fiyatlarının yaşlarına göre değişimi ve %95 güven aralıkları

Uygun modelde (s(Depozito) , s(Yas) değişkenlerine bağlı) iki nonparametrik bileşen bulunmaktadır. Tablo1 incelendiğinde, hem parametrik ve hem de nonparametrik değişkenlerin istatistiksel olarak anlamlı oldukları görülmektedir. Bu modelde evlerin bulunduğu katlardaki bir birimlik bir artış, evlerin kira fiyatlarında 0.0288 birimlik bir azalmaya sebep olmaktadır. Diğer bir ifadeyle, binaların üst katlarında kira fiyatların azda olsa düştüğü söylenebilir. Buna karşılık diğer değişkenlerin tümü ile evlerin kira fiyatları arasında aynı yönlü ilişki mevcuttur. Diğer bir deyişle, Odas, Katsay ve Kombi değişkenlerindeki bir birimlik artış kira fiyatlarına artmasına yol açmaktadır. Ancak Katsay değişkenin kira fiyatlarının üzerinde etkisinin çok düşük olduğu görülmektedir (bak. Tablo 1). Ayrıca, uygun model tarafından evlerin kira fiyatlarındaki değişmelerin %77'sini açıklanabildiği gözlemlenmiştir.

Tablo1'de nonparamtrik kısımda yer alan değişkenlere ilişkin katsayılar, parametrik olarak ifade edilemediğinden onlar ancak grafiksel olarak görüntülenmiştir. Söz konusu eğriler (2.7) formülü kullanılarak, sırasıyla depozito ve yas değişkenlerinin gözlem değerlerinden oluşan düğüm noktalarındaki kira fiyatlarını veren tahmin vektörlerinden elde edilmişlerdir. Adı geçen bu nonparametrik değişkenlerin evlerin kira fiyatları üzerindeki etkileri Şekil 1'de görülen eğriler şeklinde ortaya çıkmıştır. Tablo 1'de görüldüğü gibi eğrilerin fiyatlar üzerinde etkileri istatistiksel açıdan anlamlıdır.

Tablo2'de uygun semi parametrik toplamsal modelin elde edilmesi için incelenen diğer modellerin de performansları verilmiştir. Buna göre doğrusal regresyon modeli, evlerin kira fiyatlarındaki değişmelerin %69'unu açıklarken, uygun modelle kıyaslanmayacak ölçüde sapma içermektedir. Semiparametrik model fiyatlardaki değişimlerin %62 gibi önemli bir kısmını açıklamasına rağmen, uygun modelle kıyaslandığında, içerdiği hatanın oldukça yüksek olduğu görülmektedir. Oluşturulan uygun semiparametrik toplamsal model fiyatlardaki değişmelerin %77'sini açıklarken, içerdiği hata diğer modellerle kıyaslanmayacak ölçüde düşüktür. Modeller için hesaplanan AIC değerleri incelendiğinde en küçük AIC değerine sahip olan model uygun semiparametrik toplamsal modeldir. Bu durumda, uygun model en iyi performans göstergeleriyle diğer modellerden çok daha iyi olduğu söylenebilir.

Tablo2. Modellerin Belirlilik Katsayıları ve Sapmaları

Modeller	R^2	Sapma	AIC
Parametrik Doğrusal Model	0.6921826	29.0474	178.6649
Semiparametrik Model	0.6207759	17.3936	174.3281
Semiparametrik Toplamsal Model	0.7718312	4.0313	49.98315

5. SONUÇ

Çalışmada, Eskişehir merkezde bulunan 105 evin kira fiyatları ile evlerin özellikleri arasındaki ilişkiler, parametrik doğrusal, semiparametrik ve semiparametrik toplamsal regresyon modelleri ile analiz edilmiştir. Bilinen geleneksel doğrusal regresyondan farklı olarak, bazı değişkenlerin kira fiyatını doğrusal etkilemediği gözlenmiştir. Bu durum, örneğin, splayn düzeltme yöntemine dayalı olan tahminlerin geleneksel (parametrik regresyon) yöntemlerden daha iyi olduğunun bir göstergesidir. Yapılan analizde, hem parametrik doğrusal bileşenleri hem de nonparametrik bileşenleri bulunduran uygun bir semiparametrik toplamsal regresyon modeli ile evlerin kira fiyatlarına ilişkin yapılan tahmin sonuçlarının diğer modellerden çok daha üstün olduğu görülmüştür. Böylelikle regresyon modellerinde bağımlı değişkeni etkileyen açıklayıcı değişkenlerin doğasını (doğrusal olduğunu veya doğrusal olmadığını) belirleyerek uygun bir semiparametrik toplamsal modelin bulunması çok önemlidir ve bu durumlarda splayın düzeltme yaklaşımı çok iyi sonuçlar verir.

KAYNAKLAR

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

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

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

Hastie, T.J., Tibshirani, R.J, (1999). Generalized Additive Models, Chapman&Hall/CRC, NewYork.

Omay, E.R., Aydın, D. ve Mammadov, M., (2006). *Splayn Düzeltme Yöntemi ile Semiparametrik Adittive Modellerin Kestirimi*, 5. İstatistik Günleri Sempozyumu Bildiriler Kitabı, Konyaaltı-Antalya.

Ruppert, D., Wand, M.P., Carroll, R.J., (2003). *Semiparametric Regression*, Cambridge University Pres.

Wood, S. N, (2000). Modeling and Smothing Parametr Estimation with Multiple Quadratic Penalties. *J. R. Statist. Soc.* B 62, Part 2, 413-428.

Wahba, G., (1990). Spline Models of Observational Data, University of Wisconsin at Madison, Pensilvanya.

Rabia Ece OMAY, 1976 İskenderun doğumlu olup, ilk, orta ve lise öğrenimini İskenderun'da tamamlamıştır. 1998 yılında Ege Üniversitesi, Fen Fakültesi, İstatistik Bölümü'nden mezun oldu. 2002 yılında Dokuz Eylül Üniversitesi, Sosyal Bilimler Enstitüsü, Ekonometri Anabilim dalında yüksek lisansını tamamladı. 2003 yılında Anadolu Üniversitesi, Fen Bilimleri Enstitüsü, İstatistik Anabilim dalında doktora öğrenimine başladı ve doktora öğrenimi hala devam etmektedir. 2001 yılında Anadolu Üniversitesi Fen Fakültesi İstatistik Bölümü'nde Araştırma Görevlisi olarak işe başlamıştır ve hala aynı birimde görev yapmaktadır.

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.