



T.C.

KAHRAMANMARAŞ SÜTÇÜ İMAM ÜNİVERSİTESİ

FEN BİLİMLERİ ENSTİTÜSÜ

**SEMİPARAMETRİK REGRESYON MODELLERİ VE  
TARIMDA UYGULANABİLİRLİĞİ**

**ESRA YAVUZ**

**DOKTORA TEZİ  
ZOOTEKNİ ANA BİLİM DALI**

**KAHRAMANMARAŞ 2021**

**T.C.**  
**KAHRAMANMARAŞ SÜTÇÜ İMAM ÜNİVERSİTESİ**  
**FEN BİLİMLERİ ENSTİTÜSÜ**

**SEMİPARAMETRİK REGRESYON MODELLERİ VE**  
**TARIMDA UYGULANABİLİRLİĞİ**

**ESRA YAVUZ**

**Bu tez,**  
**Zootekni Ana Bilim Dalında DOKTORA** derecesi  
**için hazırlanmıştır**

**KAHRAMANMARAŞ 2021**

## TEZ BİLDİRİMİ

Tez içindeki bütün bilgilerin etik davranış ve akademik kurallar çerçevesinde elde edilerek sunulduğunu, ayrıca tez yazım kurallarına uygun olarak hazırlanan bu çalışmada, alıntı yapılan her türlü kaynağa eksiksiz atıf yapıldığını bildiririm.

Esra YAVUZ



Bu çalışma Kahramanmaraş Sütçü İmam Üniversitesi Bilimsel Araştırma Projeleri Yönetim Birimi Başkanlığı tarafından desteklenmiştir.

Proje No:

Not: Bu tezde kullanılan özgün ve başka kaynaktan yapılan bildirişlerin, çizelge, şekil ve fotoğrafların kaynak gösterilmeden kullanımı, 5846 sayılı Fikir ve Sanat Eserleri Kanunundaki hükümlere tabidir.

# SEMİPARAMETRİK REGRESYON MODELLERİ VE TARIMDA UYGULANABİLİRLİĞİ

## DOKTORA TEZİ

ESRA YAVUZ

### ÖZET

Parametrik regresyon modelleri, bağımlı değişkenin bağımsız değişkenlerle doğrusal bir ilişki olduğunu ve ilişkinin şeklinin biliniyor olduğunu varsayar. Bu varsayımlarının sağlanmaması durumunda ise parametre tahminleri güvenilir sonuçlar vermemektedir. İlişki şeklinin bilinmediği veya varsayımların sağlanmadığı durumlarda parametrik olmayan regresyon yöntemleri uygulanmaktadır. Ancak birden fazla bağımsız değişken olduğunda, bağımsız değişkenlerin bir kısmı bağımlı değişkenle doğrusal ilişki içerisinde iken, bir kısmı da doğrusal olmayan ilişki içerisinde olabilir. Bu değişkenlerin modellenebilmesi için, parametrik ve parametrik olmayan regresyon yöntemlerinin toplamsal olarak birleşiminden oluşan semiparametrik regresyon modellerinden faydalanılmaktadır.

Bu çalışmada parametrik, parametrik olmayan ve semiparametrik regresyon modellerinde parametrelerin tahminleri, modellerin uyum istatistikleri değerleri, güven aralıkları ve standart hata değerleri hesaplanmıştır. Çalışmada Kahramanmaraş ilinde bulunan 60 tarımsal hayvancılık işletmesine ait, barınakların fiyat üzerinde etkili olduğu düşünülen, işletmenin fiyatı, sağım ünitesi varlığı, karantina bölgesi varlığı, barınak alanı, havalandırma alanı, havalandırma sayısı, karantina alanı varlığı, revir alanı varlığı, gübre çukuru varlığı, doğum ünitesi varlığı ve şehir merkezine uzaklık değişkenleri ele alınmıştır. Modele bazı değişkenler parametrik formda dâhil edilirken, bazı değişkenler ise parametrik olmayan formda dâhil edilmiştir.

Yapılan analiz sonucunda parametrik değişkenlerden sağım ünitesi ve karantina bölgesi, parametrik olmayan değişkenlerden ise işletme alanı, havalandırma alanı, havalandırma sayısı, karantina alanı, revir alanı, gübre çukuru ve merkeze uzaklık değişkenleri de istatistiksel olarak çok önemli bulunmuştur ( $P < 0.01$ ). Sonuç olarak tarımsal işletmelerin işletme maliyetinin, buna bağlı olarak ta satış fiyatının belirlenmesinde etkili olan değişkenlerin doğru tanımlanmaları (parametrik ve parametrik olmayan) ve uygun modelin seçilmesi konusunun son derece önemli olduğu ve bu alanda semiparametrik modellerinin rahatlıkla kullanılabileceği sonucuna ulaşılmıştır.

**Anahtar Kelimeler:** Semiparametrik, Regresyon, Tarımsal İşletmeler

Kahramanmaraş Sütçü İmam Üniversitesi

Fen Bilimleri Enstitüsü

Zootekni Ana Bilim Dalı

**Danışman:** Prof. Dr. Mustafa ŞAHİN

**Sayfa Sayısı:** 75

# SEMIPARAMETRIC REGRESSION MODELS AND APPLICABILITY IN AGRICULTURE

PhD Thesis

ESRA YAVUZ

ABSTRACT

Parametric regression models assume that the dependent variable is a linear relationship with the independent variables and the form of the relationship is known. If these assumptions are not provided, the parameter estimates do not give reliable results. Nonparametric regression methods are applied in cases where the relationship type is not known or assumptions cannot be provided. However, when there is more than one independent variable, some of the independent variables may be in a linear relationship with the dependent variable, while some may be in a nonlinear relationship. In order to model these variables, semiparametric regression models, which are a combination of parametric and nonparametric regression methods, are used.

In this study parametric, nonparametric and semiparametric regression models, parametric estimates, fit statistical values of the models, confidence intervals and standard error values were calculated. In the study belonging to 60 agricultural livestock enterprises in Kahramanmaraş province thought to have an effect on the price of shelters, the variables of the price of the enterprise, the existence of the milking unit, the presence of the quarantine zone, the shelter area, the ventilation area, the number of ventilation, the quarantine area, the presence of the infirmary area, the presence of a manure pit, the presence of the birthing unit and the distance to the city center were discussed. While some variables were included in the model in parametric form, some variables were included in non-parametric form.

As a result of the analysis, the parameters of the milking unit and the quarantine area among the parametric variables, the operation area, the ventilation area, the number of ventilation, the quarantine area, the infirmary area, the manure pit and the distance to the center among the non-parametric variables were found to be statistically very important ( $P < 0.01$ ). As a result, it was concluded that the correct definition of the variables (parametric and non-parametric) that are effective in determining the operating cost of agricultural enterprises and consequently the sales price, and the selection of the appropriate model are extremely important and that semiparametric models can be used easily in this field.

**Key Words:** Semiparametric, Regression, Agricultural Businesses

University of Kahramanmaraş Sutcu Imam  
Graduate School of Natural and Applied Sciences  
Department of Animal Science

**Supervisor:** Prof. Dr. Mustafa SAHIN

**Page Numbers:** 75

## TEŐEKKÜR

Çalıőmalarım sırasında beni yönlendiren, her türlü bilgi ve tecrübeleriyle bana yardım eden sayın danışmanlarım Prof. Dr. Mustafa ŐAHİN ve Doç. Dr. Fatih ÜÇKARDEŐ' e, hocalarım Prof. Dr. Ercan EFE, Prof. Dr. Hasan ÖNDER, Prof. Dr. Hüseyin OKUT ve Prof. Dr. Ecevit EYDURAN'a en içten teşekkürlerimi sunarım. Aynı zamanda beni bugünlere getiren aileme ve her anımda yanımda olan anneme en derinden teşekkürü borç bilirim.



# İÇİNDEKİLER

	Sayfa No
ÖZET .....	i
ABSTRACT .....	ii
TEŞEKKÜR .....	iii
İÇİNDEKİLER .....	iv
ÇİZELGELER DİZİNİ .....	vi
ŞEKİLLER DİZİNİ .....	vii
SİMGELER VE KISALTMALAR DİZİNİ .....	viii
1. GİRİŞ .....	1
2. ÖNCEKİ ÇALIŞMALAR .....	4
3. MATERYAL ve METOT .....	17
3.1. Materyal .....	17
3.2. Metot .....	18
3.2.1. Regresyon analizi .....	18
3.2.1.1. Parametrik regresyon yöntemi .....	19
3.2.1.1.1. Basit doğrusal regresyon .....	20
3.2.1.1.1.1. En küçük kareler yöntemi .....	21
3.2.1.1.1.2. En çok olabilirlik yöntemi .....	22
3.2.1.1.2. Çoklu doğrusal regresyon .....	22
3.2.1.1.2.1. Çoklu doğrusal regresyon analizinde katsayıların tahmini .....	23
3.2.1.2. Parametrik olmayan regresyon yöntemi .....	24
3.2.1.2.1. Parametrik olmayan regresyon yönteminin özellikleri .....	25
3.2.1.2.2. Parametrik olmayan regresyon yönteminde yoğunluk tahmini .....	26
3.2.1.2.2.1. Histogram yöntemi .....	26
3.2.1.2.2.2. Kernel tahminleyeni yöntemi .....	28
3.2.1.2.3. Parametrik olmayan regresyonda tahmin yöntemleri .....	28
3.2.1.2.3.1. Kernel regresyon .....	29
3.2.1.2.3.1.1. Nadaraya-Watson tahmin yöntemi .....	29
3.2.1.2.3.1.2. Gasser-Müller tahmin yöntemi .....	29
3.2.1.2.3.1.3. En yakın komşuluk yöntemi .....	30
3.2.1.2.3.2. Lokal polynomial regresyon yöntemi .....	30
3.2.1.2.3.2.1. Lokal doğrusal regresyon .....	31
3.2.1.2.4. Düzgünleştirme parametrelerinin seçimi .....	32
3.2.1.2.4.1. Plug- in seçicileri .....	32
3.2.1.2.4.2. Çapraz geçerlilik yöntemi .....	33
3.2.1.2.4.3. Akaike bilgi kriteri .....	34
3.2.1.2.4.4. Bayesian bilgi kriteri .....	34

3.2.1.2.4.5. Pürüzlük ceza yaklaşımı .....	35
3.2.1.2.4.6. Etkili serbestlik derecesi .....	35
3.2.1.2.5. Parametrik olmayan regresyonda hata varyansının tahmini .....	35
3.2.1.2.6. Hata terimleri arasında otokorelasyon .....	36
3.2.1.2.7. Güven aralığı ve güven bantlarının seçimi .....	37
3.2.1.2.7.1. Güven aralıkların yinelemeli örnekleme yöntemi ile seçilmesi. 37	
3.2.1.2.8. Hipotez testleri .....	38
3.2.1.2.9. Parametrik olmayan regresyonda uyum iyiliğinin hesaplanması .....	39
3.2.1.3. Semiparametrik regresyon yöntemi .....	40
3.2.1.3.1. Semiparametrik regresyon modellerin tahmini .....	42
3.2.1.3.1.1. Geriye uyum (Backfitting) algoritması .....	42
3.2.1.3.1.2. Kısmi splayn yaklaşımı .....	44
3.2.1.3.1.3. Speckman yaklaşımı .....	44
4. BULGULAR ve TARTIŞMA .....	46
5. SONUÇ ve ÖNERİLER .....	58
KAYNAKLAR .....	61
EKLER .....	69
ÖZGEÇMİŞ .....	<b>Hata! Yer işareti tanımlanmamış.</b>

## ÇİZELGELER DİZİNİ

	Sayfa No
Çizelge 3.1. Çalışmada kullanılan veri seti .....	18
Çizelge 4.1. Değişkenlere ait tanımlayıcı istatistik değerleri .....	47
Çizelge 4.2. Doğrusal regresyon modeline ait değişkenlerin tahmin sonuçları .....	48
Çizelge 4.3. Semiparametrik regresyon modeli parametrik değişkenlerin tahmin sonuçları .....	49
Çizelge 4.4. Semiparametrik regresyon modeli parametrik olmayan değişkenlerin tahmin sonuçları .....	50
Çizelge 4.5. Semiparametrik regresyon modelde uyum istatistikleri .....	55
Çizelge 4.6. Düzeltme bileşenleri için parametre tahminleri .....	56

## ŞEKİLLER DİZİNİ

	Sayfa No
Şekil 4.1. Fiyat ile işletme alanı değişkeni arasındaki ilişki için A ve S(A) grafiği .....	51
Şekil 4.2. Fiyat ile havalandırma alanı değişkeni arasındaki ilişki için HA ve S(HA) grafiği .....	52
Şekil 4.3. Fiyat ile havalandırma sayısı değişkeni arasındaki ilişki için HS ve S(HS) grafiği .....	52
Şekil 4.4. Fiyat ile karantina alanı değişkeni arasındaki ilişki için KA ve S(KA) grafiği ..	53
Şekil 4.5. Fiyat ile revir alanı değişkeni arasındaki ilişki için RA ve S(RA) grafiği .....	54
Şekil 4.6. Fiyat ile gübre çukuru değişkeni arasındaki ilişki için GÇ ve S(GÇ) grafiği .....	54
Şekil 4.7. Fiyat ile merkeze uzaklık değişkeni arasındaki ilişki için MU ve S(MU) grafiği .....	55

## SİMGELER VE KISALTMALAR DİZİNİ

<b>A</b>	Alan
<b>AIC</b>	Akaike Bilgi Kriteri
<b>BIC</b>	Bayesian Bilgi Kriteri
<b>ÇG</b>	Çapraz Geçerlilik (Cross Validation)
<b>FY</b>	Fiyat
<b>GCV</b>	Genelleştirilmiş Çapraz Geçerlilik (Generalized Cross Validation)
<b>GÇGS</b>	Genelleştirilmiş Çapraz Geçerlilik Skoru
<b>GÇ</b>	Gübre çukuru
<b>HA</b>	Havalandırma alanı
<b>HS</b>	Havalandırma sayısı
<b>HKO</b>	Hata Kareler Ortalaması
<b>HKT</b>	Hata Kareler Toplamı
<b>İMKB</b>	İstanbul Menkul Kıymetler Borsası
<b>KA</b>	Karantina alanı
<b>KB</b>	Karantina bölgesi
<b>MU</b>	Merkeze uzaklık
<b>P</b>	Önemlilik Düzeyi
<b>RA</b>	Revir alanı
<b>R<sup>2</sup></b>	Belirleme Katsayısı
<b>S</b>	Sapma
<b>SÜ</b>	Sağım ünitesi
<b>TL</b>	Türk Lirası

## 1. GİRİŞ

İstatistikte regresyon, verilen ya da elde edilen veri kümesine dayanarak bağımsız değişkenlerin bağımlı değişkene fonksiyonel etkisini istatistiksel olarak analiz edilmektedir. Regresyon analizi değişkenler arasındaki ilişkinin belirlenmesinde yaygın olarak kullanılan önemli bir yöntemdir. Temeli 19. yüzyıla dayanan regresyon analizi, bağımsız değişkenlerin belli değerleri için bağımlı değişkenin koşullu dağılımını incelemektedir. Fen bilimleri, tıp, mühendislik ve sosyal bilimler gibi birçok alanda değişkenler arasındaki ilişkilerin belirlenmesi ve tahmin edilmesi amacıyla kullanılmaktadır (Alpar, 2003; Aytac, 1991).

Doğrusal regresyon analizi en basit regresyon modeli olarak, bazı şartlar sağlandığında, bağımsız değişkenlerin bağımlı değişkenini doğrusal olarak etkilemesi ve bağımlı değişkeninin normal dağılıma sahip olması varsayımları üzerine incelenmektedir. Doğrusal regresyon analizi için birçok teorik ve uygulama çalışmaları yapılmıştır ve bu çalışmalarda sonuçlar daha karmaşık regresyon modellerinin incelenmesi için teorik ve pratik taban ortaya çıkarmaktadır. Bazı koşullar sağlandığında, doğrusal regresyon analizi pratik tahmin problemlerinin çözümünde uygun sonuçlar ortaya çıkarmaktadır. Ancak çoğu tahmin probleminde bağımsız değişkenlerin bir kısmı bağımlı değişkenini doğrusal olarak etkilememektedir. Böylece tam doğrusal olmayan, daha karmaşık bağıntı içeren regresyon modellerinin incelenmesi ihtiyacı ortaya çıkmaktadır. Böylelikle regresyon analizi parametrik ve parametrik olmayan regresyon olarak iki farklı grupta incelenmektedir (Aneiros-Pérez, 2008; Begun ve ark., 1983).

Parametrik regresyon analizinin en önemli özelliği regresyon fonksiyonunun şeklinin önceden biliniyor olmasıdır. Ayrıca bağımsız değişkenin değerlerinin tümü için hata varyanslarının sabit olması, hata terimlerinin normal dağılması, hata terimleri arasında otokorelasyon olmaması ve bağımsız değişkenler arasında çoklu doğrusal bağlantı olmaması gibi varsayımların sağlanması şartı aranmaktadır. Varsayımların sağlanmaması durumunda ise regresyon fonksiyonu için yapılan tahminlerin sonuçları yanlış yorumlara neden olmaktadır. Böylelikle parametrik regresyon analizi ile oluşturulan modelin varsayımlarının sağlanmadığı durumda, bazı düzeltmeler yapılarak varsayımlar sağlanabilmektedir. Böylece gerekli varsayımlar sağlandığı için tahminler yapılabilmektedir (Berry, 1993; Buckley ve ark., 1988; Yatchew, 2003).

Parametrik olmayan regresyon analizinde ise, fonksiyonun şekli önceden bilinmemektedir. Parametrik regresyon analizindeki gibi önemli varsayımlar şartı aranmamaktadır. Aranan tek varsayım, hata terimlerinin ortalamasının sıfır ve varyansının da sonlu bir sayı olmasıdır. Bu nedenle değişkenler arasındaki ilişkinin belirlenmesinde esneklik olmaktadır. Parametrik olmayan regresyon analizinde avantajların yanı sıra bazı dezavantajlar da ortaya çıkmaktadır. En büyük dezavantajlarından biri çok sayıda veriye ihtiyaç vardır. Ayrıca işlemlerin matematik olarak altyapısının karmaşıklığı ve bilgisayar kullanımı gerektirmesi yaşanan diğer dezavantajlarıdır. Parametrik olmayan regresyon, parametrik regresyon analizine göre daha az etkin olmaktadır (Çağlayan, 2012; Robinson, 1988; Severini ve Staniswalis, 1994).

Semiparametrik regresyon yöntemi standart regresyon yöntemlerini genişleten ve her bir değişkenin etkisinin uygun bir şekilde yorumlanmasını sağlayan toplamsal modellerin özel bir durumu olarak incelenmektedir. Semiparametrik regresyon yöntemi, parametrik ve parametrik olmayan regresyon fonksiyonun ve toplamsal olarak birleşiminden dolayı “kısmi doğrusal regresyon modelleri” de denilmektedir. Semiparametrik regresyon modelinde bağımsız değişkenler ilişkisiz ise, modelin parametrik değişkenlerinin katsayıları en küçük kareler yöntemi uygulanarak ve parametrik olmayan değişkenlerinin kısmi regresyon fonksiyonları ise splayn gibi parametrik olmayan yöntemler ile tahmin yapılmaktadır. (Harezlak ve ark., 2018; Mahmoud ve ark., 2016; Newey, 1989).

İstatistiksel açıdan, eğer regresyon fonksiyonu parametrelerin sınırlı sayıları ile tanımlanmış doğrusal bir model sınıfına ait ise regresyon modeli parametrik olarak, buna karşılık regresyon modeli devamlı değişebilir doğrusal olmayan bir fonksiyon sınıfına ait ise parametrik olmayan regresyon olarak tanımlanabilir. Semiparametrik ifadesi ise parametrik ve parametrik olmayan kavramların toplamı biçiminde gösterilebilir (Mahmoud ve Kim, 2019; Manski, 1986; Nash ve Bradford, 2001; Omay ve ark., 2006).

Semiparametrik tahmin yöntemlerinin, parametrik ve parametrik olmayan yöntemlere nazaran daha az sınırlayıcı olması semiparametrik yöntemlerinin daha çok uygulanması ve yayılmasına sebep olmuştur. Bu yöntem ayrıca uygulamada oluşabilecek bazı olumsuzlukları giderme imkânı da sağlamaktadır. Aynı zamanda semiparametrik yöntemler uygulanırken örneklem büyüklüğünün önemi de azalmaktadır. Özellikle küçük örneklerden oluşan uygulamalarda araştırmacı, parametre ve değişken hakkında bilgi sahibi olmasa bile araştırmaya devam etme imkânı elde edebilmektedir. Parametrik ve

parametrik olmayan yöntemlerde bazı varsayımlara ihtiyaç duyulurken, semiparametrik regresyon yönteminde ise varsayımlar yerine gelmese bile araştırma devam etmektedir (Rajarithinan ve Parmar, 2011; Shi, 2009; Toprak, 2015).

Son zamanlarda istatistikte parametrik ve parametrik olmayan regresyon analizinin yanı sıra semiparametrik regresyon analizi alanında yapılan çalışmalara da ihtiyaç artmıştır. Bu alanla ilgili her ne kadar çalışmalar yapılmış ve farklı metotlar tavsiye edilmiş olsa da ülkemizde bu alanla ilgili çok az çalışma yapılmıştır. Bu çalışmada semiparametrik regresyon modelinin teorik kısmına ilaveten uygulamalı bir şekilde incelenmesine de yer verilmiştir. Bu çalışmanın amacı, parametrik regresyon analizleri ile uygun sonuç vermeyen regresyon sorunlarına çözüm bulmak amacıyla yapılan çalışmalarda, öncelikle karşılaşılan bu tür sorunların çözümü için, oldukça geniş bir uygulama alanına sahip semiparametrik regresyon yöntemini araştırmaktır. Aynı zamanda semiparametrik regresyon yönteminin parametrelerine ilişkin tahminler yapılarak yönteminin istatistiksel açıdan yorumlanmasına olanak sağlanmıştır (Demir, 2005; Dhekale ve ark., 2017).

Bu çalışmada parametrik, parametrik olmayan ve semiparametrik regresyon yöntemleri olmak üzere üç farklı regresyon yöntemi, regresyonda düzgünleştirme yöntemi, pürüzlük ceza yaklaşımı ve splayn düzeltme teknikleriyle beraber düzgünleştirme parametresinde kullanılan tahmin yöntemleri açıklanmıştır. Daha sonra çalışmanın esas konusu olan semiparametrik regresyon ele alınmış olup, modelin tahmininde kısmi splayn ve backfitting algoritması olmak üzere iki farklı yaklaşım incelenmiştir. Ayrıca semiparametrik regresyon modeline ilişkin çıkarsamalar, hem parametrik hem de parametrik olmayan regresyon yöntemi için uygulanmıştır.

Teorik olarak incelenen semiparametrik regresyon çalışmasını uygulamayla desteklemek amacıyla, tarımsal işletmelerin satış fiyatı ile işletmelerin karakteristik özelliklerini içeren değişkenler arasındaki ilişkinin incelenmesine ilişkin bir uygulama yapılmıştır. İstatistiksel analizlerde SAS 9.4 programı kullanılmıştır. Söz konusu uygulamada parametrik, parametrik olmayan ve semiparametrik regresyon modellerinin tahmin sonuçları, değişkenlere ait eğrilerin grafiklerle gösterimi, modele uyum istatistik değerleri, düzeltme bileşenleri için parametre tahmin değerleri elde edilmiştir.

Çalışmanın sonuç kısmında ise tezin amacı ve elde edilen sonuçlar özetlenerek, çalışmanın literatüre olan katkısı ve çalışma ile ilgili bazı önerilere yer verilmiştir.

## 2. ÖNCEKİ ÇALIŞMALAR

Akaike (1973), çalışmada, maksimum olabilirlik tahmin yönteminin çok genel bilgi kriterine göre optimum bir tahminin asimptotik olarak gerçekleştirilmesi için bir yöntem olarak kabul edileceğini göstermiştir. Ayrıca bu gözlem, istatistiksel model uydurmanın birçok pratik problemine cevap verme ilkesini de gösterdiğini belirtmiştir.

Clark (1977), çalışmasında, küçük örnekler için bazı alternatif tahmin edicilere göre teorik avantajlara sahip olduğu gösterilen, düzgün bir regresyon fonksiyonunun parametrik olmayan tahmin edicilerinin yeni bir yöntemini sunmuştur. Uygulama olarak simülasyon çalışmasını kullanarak, uygun düzeltme derecesini belirleme aracı olarak çapraz doğrulama tekniğinin olağanüstü başarısını göstermiştir.

Peterson (1977), çalışmasında Kaplan-Meier tahmincisini tanıtmış ve bu tahmin yönteminin hayatta kalma işlevlerinin bir fonksiyonu olduğunu ifade etmiştir. Kaplan-Meier tahmincisini bu şekilde ifade etmenin kullanılabilirliğinin bir örneği olarak Kaplan-Meier tahmincisinin güçlü tutarlılığını kolayca kanıtlandığını da göstermiştir.

Craven ve Wahba (1979), çalışmalarında spline eğrileriyle bazı verileri optimum yumuşatma miktarını tahmin etmek için pratik etkili bir yöntem incelemişler. Ayrıca düzgünleştirici eğrilerin bazı verileri yumuşatan eğriler olduğunu da göstermişler.

Green ve ark. (1985), çalışmalarında bağımsız hata modelini varsayarak, hem işlem etkilerini hem de bilinmeyen eğilimi tahmin etmek için en küçük kareler düzeltme yöntemlerini ele almışlar. Temel yöntemin uygulanmasını üç veri seti üzerinde göstermişler ve sonuçları diğer analizlerle karşılaştırmışlar.

Hurvich ve ark. (1988), çalışmalarında, düzgün bir regresyon fonksiyonunun parametrik olmayan tahminlerini oluşturmak için yerel polinom, çekirdek ve düzgünleştirici spline tahmin edicileri dahil olmak üzere birçok farklı yöntem önermişler. Bu çalışmada, Akaike bilgi kriterine dayalı bir kriterin geliştirilmiş versiyonu olan düzeltilmiş akaike bilgi kriterini incelemişler ve yumuşatma parametresini seçmenin bir yolu olarak göstermişler.

Powell ve ark. (1989), çalışmalarında genel bir fonksiyonunun yoğunluk ağırlıklı ortalama türevinin tahmin edilmesiyle, indeks modellerinin katsayılarının tahmin edilmesi problemine bir çözüm sunmuşlar. Tahmin edicilerin asimptotik varyans-kovaryans

matrislerinin tutarlı tahmin edicilerini vermişler ve prosedürlerin pratik performansını incelemek için Monte Carlo simülasyon çalışmasını uygulamışlar.

Lee (1990), tez çalışmasında, öncelikle semiparametrik regresyon modellerini açıklamış ve daha sonra ekonometri uygulamalarında semiparametrik yöntemlerini uygulayarak verimlilik artışlarını incelemiştir. Ayrıca tek denklem modellerini ve koşullu moment kısıtlamaları ile maksimum olabilirlik yöntemlerini açıklamıştır.

Chamberlain (1992), çalışmasında, semiparametrik regresyon modelleri ve parametrik olmayan bileşenler için verimlilik sınırlarını incelemiştir. Ayrıca verilerin dağılımını F' den rastgele örnek oluşturarak incelemiştir. Verilerin ve parametrelerinin belirli bir işlevi oluşunu açıklamıştır.

Cuzick (1992), çalışmada, hata dağılımı bilinmeyen semiparametrik regresyon modellerinde etkili tahminleri incelemiştir. Semiparametrik regresyon modelinde, hata dağılımının normal dağılım gösterdiğinde parametrelerin bilinmediğini göstermiştir. Bununla birlikte, sonlu Fisher bilgisi ile genel hata dağılımları için hata tahmininin bilinmediği durumlarda bile etkili kestirimin mümkün olabileceğini göstermiştir.

Horowitz (1993), çalışmasında, iş seyahat için otomobil ve transit arasında tercih edilen ikili yanıt modeli, sabit ve rastgele katsayıları semiparametrik regresyon modeli tahmin etmiştir. Tahmin yöntemlerini, probit modelleri için maksimum olasılık, tek indeksli model için ise semiparametrik regresyon modelini tercih etmiştir.

Klein ve Spady (1993), çalışmalarında seçim olasılık fonksiyonunun fonksiyonel formu ile ilgili herhangi bir varsayım yapmayan bu fonksiyonun bir indeks ile karakterize edilebileceği ayrı seçim modelleri için bir tahminci olduğunu önermişlerdir. Tahmin edicinin tutarlı, asimptotik olarak normal dağılmış olduğu ve yarı parametrik modeline bağlı olduğunu göstermişlerdir.

Schick (1996), semiparametrik regresyon ve semiparametrik katkı regresyon modellerini açıklamıştır. Otoregresif hatalar ile semiparametrik katkı regresyon modelinde etkili tahmin yöntemini göstermiştir. Semiparametrik regresyon modellerinde verimli tahminlerin oluşturulmasında önemli bir yöntem olan sonlu boyutlu parametrenin ayrık ve tutarlı tahminlerini uygulamıştır.

Ai (1997), çalışmasında, maksimum olabilirlik tahmin yöntemi için semiparametrik regresyon modelini uygulamıştır. Tahmin edicinin sonlu örnekleme performansını değerlendirmek için Monte Carlo çalışmasını yapmıştır. Tahmincinin, hata yoğunluklarının

bilinmediği çok değişkenli doğrusal olmayan regresyon, eşzamanlı denklemler, kısmen belirtilen regresyon, indeks regresyonu, sentezlenmiş regresyon, anahtarlama regresyonu ve dengesizlik modelleri için karakterize edilen semiparametrik verime bağlı olduğunu göstermiştir.

Chen ve Randall (1997), çalışmalarında, doğal kaynak değerlendirme uygulamasıyla ikili yanıt modellerinin semiparametrik regresyon modeli ile tahminini ele almışlar. Model tahmini için maksimum olabilirlik tahmin yöntemini uygulamışlar. Anket veri çalışmasıyla semiparametrik regresyon tahminin parametrik regresyon tahmine göre daha anlamlı sonuç gösterdiğini belirtmişler.

Liang ve ark. (1997), çalışmalarında kısmen doğrusal modellerde parametrik olmayan regresyon tahmininin asimptotik özelliklerini incelemişler. Ayrıca yarı parametrik regresyon modelinde parametrik olmayan kısımdaki değişken, bilinen bir dağılıma sahip ardışık hatalarla ölçüldüğünde, tutarlı ve asimptotik olarak normal olduğunu gösterilebilen  $\beta'$  nın bir tahminini açıklamışlar.

Liang ve Hardle (1999), makalelerinde, semiparametrik bir model için hata dağılımının kestirimi için büyük örnek teorisini incelemişler. Uygun koşullar altında, tahmin edicilerinin olasılıkta yakınsadıklarını göstermişler. Asimptotik normallik ve tahminin yakınsama oranlarını da araştırmışlar. Son olarak, tahmin ediciler için yinelenen logaritma yasasını da göstermişler.

Fox (2000), kitabında, çekirdek tahmini, parametrik olmayan regresyon yöntemlerini ve verilerin dönüşümlerini seçmek için parametrik olmayan regresyonun uygulanacağı yolları tanıtmış. Ayrıca parametrik olmayan regresyonun logit, probit ve Poisson regresyonuna genelleştirebileceği adımları da incelemiştir.

Van ve ark. (2001), çalışmalarında, anlaşma etki eğrisini tahmin etmek için semiparametrik regresyon modelini uygulamışlar. Parametrik regresyon modellerindeki bpyutsallık probleminden dolayı semiparametrik regresyon modelini tercih etmişler. Hem parametrik hem de parametrik olmayan bileşenlerin tahmini için semiparametrik regresyon modelinin daha anlamlı sonuç olduğunu göstermişler.

Carota ve Parmigiani (2002), çalışmalarında, sayım verileri için semiparametrik regresyon modelini uygulamışlar. Orta veya büyük boyutlu veri kümeleri için geliştirilmiş doğrusal modellerin esnek, uygulanması kolay ve sağlam bir uzantı sağlamışlar. Ayrıca yanıt değişkeninin önemli olduğu regresyon problemleri için bir sınıf

Bayesian semiparametrik modelini arařtırmıřlar. Açıklayıcı deęiřkenlerin yanıt üzerindeki etkilerini toplam kütle parametresi ile modellemiřler. Modelleme seeneklerini ve dięer yaklařımlarla iliřkilerini de incelemiřler.

aęlayan (2002), tez alıřmasında, deęiřkenler arasındaki iliřkiyi matematiksel bir fonksiyon ile belirlemek iin regresyon modellerini kullanmıřtır. Bu iliřkiyi aıklamak iin parametrik regresyon modellerini ve parametrik olmayan regresyon modellerine alternatif olarak yarı parametrik modellerini de ele almıřtır. Öncelikle regresyon modellerini aıklayıcı bir řekilde tanımlamıřtır. Daha sonra uygulama da yařam boyu sürekli gelir hipotezinin Türkiye iin geerli olduęunu regresyon modellerinde uygulayarak sonuları yorumlamıřtır.

Fung ve ark. (2002), alıřmalarında, doęrusal modellerin parametrik olmayan bir bileřenin dahil edilmesini saęlamak iin maksimum cezalandırılmıř olasılık tahmin edici yöntemini ele almıřlar. Ayrıca teřhis ve aykırı deęer testlerini etkileyen yarı parametrik modelleri ele almıřlar. Önerilen etki teřhisini simülasyon bir alıřma ile göstermiřler.

Kim ve ark. (2002), alıřmalarında, yarı parametrik regresyon modellerinde tanıyı etkileme ve yanıt tahmin edicileri iin bazı etki teřhislerini saęlamıřlar. Katsayı vektörünün tahmin edici üzerindeki etkili bir gözlemin, parametrik olmayan bileřen üzerinde etkili olmayabileceęini aıklamıřlar. Böylelikle bir gözlemin etkisi her bir tahminci üzerinde ayrı ayrı deęerlendirilmesi gerektięini göstermiřler. Ayrıca gerek bir veri setine dayalı aılayıcı bir örnek de vermiřler.

Kou ve Efron (2002), alıřmalarında parametrik olmayan regresyon modellerin tahmininde yaygın olarak kullanılan düzgünleřtirme yöntemini incelemiřler. Düzgünleřtirme parametresinin seimi iin genelleřtirilmıř maksimum olabilirlik yöntemini ve seim kriterini tercih etmiřler. Ayrıca seim kriterlerinin grafiklerle gösterimini de vermiřler.

Ruppert ve ark. (2003), “Semiparametric Regression” isimli eserlerinde, semiparametrik regresyonun tekniklerini ve faydalarını kısa ve modüler bir řekilde aıklamıřlardır. Ayrıca örnek ve grafiklerin gösterimi ile ilgili uygulamalara yer vermiřlerdir.

Zhu ve Cui (2003), arařtırmalarında, tüm deęiřkenlerde ölçüm hataları olan kısmi bir doęrusal regresyon modelini incelemiřler. Parametrik olmayan bir çekirdek tahmincisi ile birlikte yeni bir parametrik tahminci sınıfı oluřturmak iin momentler yöntemini

kullanmışlar. Ayrıca semiparametrik regresyon modelinde hem parametrik olmayan kısımdaki değişkenler hem de parametrik kısımdaki değişkenler hatalı ölçüldüğünde  $\beta$ ' nın tutarlı ve asimtotik olarak normal olduğu gösterilebilen bir tahmin problemini ortaya çıkarmışlar. Güçlü yakınsama, zayıf yakınsama oranı ve tahmin edicilerin asimtotik normalliklerini de araştırmışlar. Araştırma sonucunda ölçüm hatalarının dağılımının düzgünlüğü, özellikle parametrik olmayan kısımdaki hatalar, tahmin edicilerin performansı üzerinde önemli bir etkiye sahip olduğunu gözlemlemişler.

Hardle ve ark. (2004), kitaplarında parametrik olmayan modelleri ve yarı parametrik modellerini incelemişler. Parametrik olmayan modellerden, histogram, çekirdek yoğunluğu tahmini ve parametrik olmayan regresyon yöntemlerini, yarı parametrik modellerden ise genelleştirilmiş regresyon, tek endeksli modeller, genelleştirilmiş kısmi doğrusal modeller, katkı ve genelleştirilmiş katkı modellerini ele almışlar. Ayrıca yarı parametrik modelleme tekniğinde, kısmi parametrik bileşenleri tanıtarak istatistiksel prosedürlerin esnekliğini ve basitliğini de ele almışlar. Kitapta, yüksek boyutlu nesnelere yoğunluk fonksiyonları ve regresyon yöntemleri ile göstermişler.

Kaki (2004), tez çalışmasında, parametrik olmayan düzleştirme fonksiyonu ve sürekli bağımsız değişkenler için doğrusal bir fonksiyon kullanarak oluşturulan yarı parametrik regresyon yöntemiyle çeşitli dönem koyun ağırlıklarının süttan kesim ağırlığı üzerine etkisini incelemiştir. Bununla birlikte aynı verilere uygulanan yarı parametrik regresyon bulgularıyla parametrik ve parametrik olmayan regresyon modellerinden elde edilen bulguları karşılaştırmıştır. Ayrıca yarı parametrik ve parametrik olmayan regresyonda parametre tahminlerini backfitting algoritması kullanarak incelemiştir.

Koop ve Poirier (2004), makalelerinde, kısmi doğrusal normal regresyon modelinde semiparametrik çıkarım için yeni Bayesci yöntemler incelemişler. Semiparametrik regresyon alternatiflere karşı parametrik regresyon modellerini test etme yöntemlerini geliştirmişler. Bu yöntemlerin, hesaplama karmaşıklığı açısından diğer modellere nasıl genişletebileceğini tartışmışlar. Ayrıca normal hataların varsayımının nasıl genişletebileceğini göstermişler.

Schennach (2004), parametrik olmayan yoğunluk tahmini probleminin aksine, bağımsız değişkenlerin bilinmeyen bir dağılımdan gelen bir hata ile ölçüldüğünde, koşullu beklentilerin parametrik olmayan bir regresyon tahminçisi olduğunu sunmuştur. Schennach, bilinmeyen bir dağılımın hatası bulaşmış bağımsız değişkenleri göz önüne alan

geleneksel Nadaraya-Watson kernel regresyon tahmin edicisini genişleterek bu boşluğu doldurmuştur. Ayrıca sadece tahmin edicinin yakınsama oranını ortaya koymamış, onun asimtotik dağılımını da göstermiştir.

Aydın (2005), tez çalışmasında, semiparametrik regresyon modelinin kestirimi için kısmi splayn ve Speckman yaklaşımını uygulamıştır. Bu iki yaklaşımı MATLAB programında uygulamış ve modelin hem parametrik hem de nonparametrik bileşeni hakkında bilgi vermiştir. Nonparametrik ve semiparametrik regresyon modellerinde cezalı en küçük kareleri splayn düzeltme yöntemi ile incelemiştir. Bu yöntem doğrultusunda düzeltme parametresinin seçiminde yaygın olarak ele alınan seçim kriterlerini incelemiştir. Bu seçim kriterlerini belirlemek amacıyla, MATLAB programında yazılan bir simülasyon çalışması ile göstermiştir.

Hu (2005), çalışmasında ilk olarak parametrik ve parametrik olmayan regresyon yöntemlerini açıklamış. Daha sonra yöntemlerin geçerliliği ve fizibilitesi bir simülasyon örneği ile göstermiştir.

Aneiros-Pérez ve Vieu (2006), çalışmalarında açıklayıcı değişkenlerden bazılarının işlevsel olduğu durumlarda, bazı gerçek değerli yanıt değişkenlerini tahmin etme problemini ele almışlar. Yani, hem yarı doğrusal modellemenin avantajını hem de fonksiyonel veriler için parametrik olmayan istatistikler için yeni modeller tanıtmışlar. Ayrıca yarı parametrik regresyon modeli kurulduktan sonra, tahmin sonuçlarını yorumlamışlar. Son olarak ta gerçek bir veri örneği ile modelin kullanılabilirliğini göstermişler.

Aydın ve ark. (2007), çalışmalarında yarı parametrik modelini esas alarak Green ve Silverman'ın kısmi splayn yaklaşımını ve Speckman yaklaşımını incelemişler. Yarı parametrik regresyon modelleri için splayn düzeltme metoduna dayalı yeni bir tahmin kavramı Eubank vd. tarafından göstermişler. Bu tahmin kavramına göre, splayn düzeltmeyi esas alan farklı iki yaklaşım Schimek tarafından karşılaştırılmışlar. Ayrıca, yarı parametrik regresyon modelinin tahminini Green-Silverman'ın kısmi splayn ve Speckman yaklaşımı olarak adlandırılan farklı iki yaklaşımla ele almışlar ve sıradan en küçük kareler tahminleri ile karşılaştırmışlar. Kullandıkları veriler, Kanada'nın başkenti olan Ottawa'da, 1987 yılında 92 müstakil eve ait olup MATLAB ortamında yazılan bir program kullanarak uygulamada kullanmışlar. Speckman ve kısmi splayn yaklaşımına göre

elde ettikleri modelin parametrik katsayılarının tahmin değerleri ve standart sapmaları, istatistikleri, parametrik katsayıların güven aralıkları ve  $F$ -test istatistiği sonuçlarını tabloda göstermişler. Ayrıca çalışmada, splayn düzeltmede, genelleştirilmiş çapraz geçerlilik metoduyla düzeltme parametresini elde etmişler ve yarı parametrik modelin hem parametrik hem de parametrik olmayan bileşeni için çıkarsamaları tartışmışlar.

Omay ve ark. (2007), çalışmalarında, Eskişehir' deki evlerin kira fiyatları ve özellikleri arasındaki ilişkilerin analizini semiparametrik toplamsal regresyon modeli ile tahmin etmişlerdir. Yapılan istatistiksel analizler sonucunda, evlerin kira fiyatları üzerinde bağımsız değişkenlerinin bir kısmının fiyatları lineer, bir kısmının da nonlineer olarak etkilediğini göstermişlerdir. Bu amaç doğrultusunda model, parametrik bileşenlerine ilaveten nonparametrik bileşenleri de bulduran semiparametrik toplamsal regresyon modeli şeklinde göstermişlerdir. Sonuç olarak çalışmalarında, semiparametrik toplamsal regresyon modelin istatistiksel olarak önemli olduğunu, ayrıca hem parametrik doğrusal hem de semiparametrik modellerden daha anlamlı sonuçlar verdiğini göstermişlerdir.

Akkuş ve ark. (2008), çalışmalarında, iki düzeyli bağımlı değişken için parametrik, parametrik olmayan ve yarı parametrik regresyon metotları üzerinde araştırma yapmışlar. Öncelikle parametrik ve parametrik olmayan metotları tanıtmışlar. Daha sonra parametrik model varsayımlarının sağlanmaması durumunda parametrik ve parametrik olmayan modellerin en iyi yöntemi olan yarı parametrik yöntemini tanıtmışlar. Bu yöntem doğrultusunda parametreler, Klein ve Spady' nin yarı parametrik en çok olabilirlik tahmin edicisi ile parametrik olmayan Nadaraya-Watson çekirdek kestirim yöntemi ile iki aşamada tahmin etmişler. Ayrıca modellerin uygulamasını da sayısal bir örnekle göstermişler. Çalışmanın sonucunda, parametrik model varsayımlarının doğruluğu kontrol edilmeden elde edilen sonuçlar yanıltıcı olabileceğinden dolayı, doğrusal indeks varsayımı dışında başka ek varsayımlara ihtiyaç duymayan yarı parametrik yöntemlerin istatistiksel çalışmalarda uygulanabilirliğini göstermişler.

Ata ve ark. (2008), bu çalışmada, orantılı hazardlar varsayımının sağlanmadığı durumlarda parametrik yaşam modelleri olan üstel, weibull, log-lojistik, log-normal, Gompertz, Gamma regresyon modelleri ve yarı parametrik yaşam modelleri olan Cox regresyon modeli ve uzanımlarını incelenmişler. Orantılı hazardlar varsayımını sağlamayan gerçek yaşam verilerini kullanarak, parametrik ve yarı parametrik yaşam modellerinin uygulamasını göstermişler.

Harezlak ve ark. (2008), kitaplarında, semiparametrik regresyon modellerini tanımlayarak bu modellerin kodlarını R programında göstermişler. Semiparametrik regresyonun astronomi, biyoloji, tıp, ekonomi ve finans gibi farklı alanlarda birçok uygulamada büyük değeri olduğu kanıtlamışlardır. R programını kullanarak semiparametrik regresyon yöntemlerine ilişkin kolay takip edilen kitaplarında, mevcut metodoloji ve pratikte kullanımı arasındaki boşluğu kapatmayı amaçlamışlardır. Yazarlar, çok çeşitli semiparametrik regresyon modellerini ve R analizlerini bir araya getirerek bunları uygulamalı araştırmacılar için yararlı bir forma getirmişlerdir. Akdeniz ve Akdeniz (2009), çalışmalarında, semiparametrik regresyon modelinde parametrelerin vektörü için Liu tahmin yöntemini incelemişler. Ayrıca her bir parametrik bileşen için Liu tahmin edici yöntemini kullanarak semiparametrik regresyon modelini elde etmişler.

Keele (2008), kitapta, parametrik, nonparametrik ve semiparametrik regresyon modellerini açıklamıştır. Semiparametrik regresyon modellerinde backfitting algoritması tahmin yöntemini ele almıştır. Ayrıca her örnek için tüm data setlerine ve R kodlarına kitapta yer vermiştir.

Omay ve Memmedli (2008), çalışmalarında hava kirliliğinin insan sağlığı üzerindeki etkisini incelemek için geliştirilmiş toplamsal modellerini uygulamışlar. Ayrıca çalışmalarında, doğrusal model, toplamsal model, ince tabakalı splayn ve geliştirilmiş toplamsal model gibi regresyon modellerini de incelemişler. Uygulamada, S-Plus programında gam paketinin gelişmiş versiyonu olan gam.exact paketinde incelemişler.

Akdeniz ve Tabakan (2009), çalışmalarında, semiparametrik regresyon modelinde her bir parametrik bileşen için sınırlı sayıda tahmin ediciler elde etmişler. Uygulama olarak ise konut fiyatlarından oluşan bir veri setiyle ve önerilen ilgili tahmincilerin performanslarını Monte Carlo simülasyonu aracılığıyla karşılaştırmışlar.

Akkuş ve ark. (2009), çalışmalarında, iki düzeyli, bağımlı değişken modelinde yarı parametrik tahminin ilk aşamadaki parametreler vektörü Klein ve Spady tahmin edici yöntemini kullanarak elde etmişler. Nadaraya-Watson ve uyarlanabilir Nadaraya-Watson tahmin edicilerinin yarı parametrik model tahminindeki performanslarını ise farklı örneklem büyüklükleri için yapılan bir benzetim çalışması ile Ortalama Hata Kareler Ortalaması kriterlerine göre karşılaştırmışlar.

Linton (2009), makalede, semiparametrik ve parametrik olmayan regresyon yöntemini ARCH modelleme ile tahminini incelemiştir. Ayrıca modellerin fonksiyonel formu, hata şeklini, ortalama ve varyans arasındaki ilişkiyi incelemiştir.

Tabakan (2009), tez çalışmasında, parametrik regresyon model varsayımlarının sağlanmaması durumunda parametrik ve parametrik olmayan yaklaşımların en uygun yöntemi olan semiparametrik regresyon modelini tanıtmış ve semiparametrik regresyon modelinde parametre tahmini için üç farklı yaklaşım yöntemlerini göstermiştir. Bu yöntemlerden cezalı en küçük kareler temelinde olan eğrisel çizgi düzeltme yöntemi, semiparametrik regresyon modelinde parametre tahminine dayanan dalgacık tabanlı tahmin yöntemi ve semiparametrik modelde parametre tahmini için fark alma fikri yöntemlerini incelemiştir. Ayrıca semiparametrik modelde çoklu iç bağlantının olması durumunda farka dayalı ridge tahmin edici yönteminden bahsetmiştir. Böylelikle farka dayalı ridge tahmin edici ile farka dayalı tahmin edici yöntemlerini hata kareler ölçütüne göre kıyaslama yapmıştır. Çalışma sonucunda, yanlış tahmin edici olan farka dayalı ridge tahmin edicinin farka dayalı tahmin ediciden hata kareler ortalamasına göre daha önemli olmadığı, yalnızca belirli şartlar altında önemli olduğunu göstermiştir.

Bağdatlı (2010), tez çalışmasında, parametrik olmayan regresyon modelinin tahmini çok boyutluluğun yarattığı sıkıntı nedeniyle model tahmini için semiparametrik regresyon modelini önermiştir. Böylelikle bir semiparametrik regresyon modellerinin temelini oluşturan toplamsal modelleri açıklamış ve bu modelleri ayrıntılı bir şekilde ele almıştır.

Sinan (2010), tez çalışmasında, panel veri analizleri için parametrik, parametrik olmayan ve yarı parametrik modellerin kullanımını araştırmıştır. Bu modeller için 3 farklı bağımsız değişkenlerini incelemiş ve modellerin tahminini yapmıştır. Yapılan tahminler sonucunda, yarı parametrik regresyon modelinin parametrik modellere göre daha anlamlı olduğunu ve parametrik olmayan regresyon modele göre yorumlamasının daha kolay olduğunu göstermiştir. Tez uygulamasında, 2009 Avrupa ekonomik krizini yarı parametrik regresyon modelini kullanarak araştırmıştır. Ayrıca medyana dayalı parametrik olmayan tahmin yöntemlerini de yorumlamıştır ve bu yöntemlerini simülasyon yardımıyla karşılaştırmıştır.

Akkuş (2011), çalışmasında, parametrik ve semiparametrik regresyon modellerini Xplore komutlarında uygulanabilirliğini göstermiştir. Semiparametrik kestiriminin tüm aşamalarında kestirim için gerekli komutlarını vermiştir. Ayrıca parametrik regresyon için

lojistik, probit ve tamamlayıcı log log regresyon modellerini de ayrıntılı biçimde göstermiştir. Uygulamada, komutların uygulanabilirliğini göstermek için yapay veri setini kullanmıştır. Çalışmanın amacı, araştırmacılara özellikle semiparametrik modeller için standart istatistiksel paketlerde bulunması kolay olmayan ek çıktıları açıklamaktır.

Çağlayan ve Dayıoğlu (2011), çalışmalarında, bir hanenin yoksulluk ve yaşam standartlarını etkileyebilecek faktörleri belirlemişler. Ayrıca parametrik ve semiparametrik logit modellerini kullanarak Türkiye’deki hanelerde yoksulluk olasılıklarını göstermişler. Parametrik ve parametrik olmayan yaklaşımların en iyi özelliklerini birleştiren semiparametrik yöntemin model varsayımlarında daha etkin olduğunu göstermişler.

Duran ve ark. (2011), çalışmalarında öncelikle semiparametrik regresyon modelinin tanımını ve tahmin edicilerin performansını göstermek için sayısal bir örnek vermişler. Ayrıca semiparametrik regresyon yöntemini ve Liu tahminci yöntemini ile iki aşamalı olarak karşılaştırmışlar.

Evren ve Tuna (2011), çalışmalarında, parametrik, parametrik olmayan ve yarı parametrik panel veri modellerini tahmin etmişler ve tahmin edilen sonuçları karşılaştırmışlar. Modellerin tahminleri sonucunda, parametrik olmayan modellerin verilerini daha iyi temsil ettiğini, yarı parametrik modellerin parametrik ve parametrik olmayan modellerin dezavantajlarını ortadan kaldıran model olmadığını göstermişler. Uygulamada, göç araştırmalarının uluslar arası göç üzerinde etkisi bulunan birçok değişkenleri ele almışlar. Bu amaç doğrultusunda rassal ve sabit etkili kısmi doğrusal panel veri modelleri ile ele alınan temel makro ekonomik değişkenlerinin uluslararası göç trendi üzerindeki etkisini incelemişler.

Turanlı ve Bağdatlı (2011), çalışmalarında semiparametrik regresyon modellerinin tanımını ve parametrik ve parametrik olmayan regresyon modellerin farklılıklarını da açıklamışlardır. Ayrıca semiparametrik regresyon modellerinin tahmini (backfitting algoritması), güven bantları, standart hataların hesaplanması ve hipotez testlerini göstermişlerdir.

Tezcan (2011), çalışmasında parametrik ve parametrik olmayan regresyon tekniklerini ve farklılıklarını incelemiştir. Bununla birlikte Kernel tahmini, düzgünleştirme parametresi ve LOWESS yöntemini parametrik olmayan regresyon analizini uygulayarak İstanbul Menkul Kıymetler Borsası (İMKB) İmalat sanayisinde işlem gören işletmelerin borsa performansı ile uygun finansal oranlar arasındaki ilişkiyi araştırmıştır. Bu amaçla,

İMKB’de işlem gören şirketlerin piyasa değeri/defter değeri oranı çeşitli finansal oranlar ile göstermiştir.

Turanlı ve Bağdatlı Kalkan (2012), çalışmalarında, öncelikle semiparametrik regresyon modelinin tanıtımını yaparak daha sonra inşaat sektöründe çok önemli yeri olan site içerisindeki daire fiyatlarını etkileyen özellikleri semiparametrik regresyon analizi ile belirlenmesini incelemişler. Uygulamada, ilk önce modele düzgünleştirerek ilave edilmesi gereken değişkenleri oluşturup sonra anlamsız olan değişkenleri belirlemişler. Semiparametrik regresyon modelleri ile logaritmik, lineer ve karesel dönüşümle elde edilen modelleri araştırmışlar. Ayrıca parametrik kısımdaki değişkenleri t testi ile, parametrik olmayan kısımdaki değişkenleri ise F testi ve güven bantlarıyla test etmişler.

Engle ve ark. (2012), çalışmalarında, hava ve elektrik satışı arasındaki fonksiyonel ilişkiyi semiparametrik tahmin yöntemi ile incelemişlerdir. Uygulamada, tekniğin mantıklı ve sezgisel olarak çekici fonksiyonel ilişkiler üretme vaadini göstermişlerdir. Tahminler, dört elektrikli tesisin konut satışları için sunulmuştur. Ayrıca sadece ısıtma ve soğutma derece günlerini kullanan hava parçalarını temsil eden modellerle ve parçalı doğrusal spline’ larla karşılaştırmışlar.

Fit (2012), tez çalışmasında, parametrik ve parametrik olmayan regresyon yöntemlerini açıklamış. Parametrik olmayan regresyonda parametre tahmini için düzeltme parametrelerinin seçimini incelemiştir.

Türkan (2012), tez çalışmasında, doğrusal regresyon modelinde etkili gözlemleri belirlemek için önerilen ölçütlerin yarı parametrik regresyon modeli için geliştirilmesi ve geliştirilen ölçütlerin etkili gözlemleri belirlemede başarılı olup olmadıklarının gerçek bir veri kümesi kullanarak ve simülasyon çalışması yaparak araştırmıştır. Ayrıca gerçek bir veri kümesi ve yapay bir veri kümesi kullanarak önerilen ölçütlerin etkili gözlemleri ortaya çıkarmadaki başarıları da incelemiştir. Simülasyon çalışması yaparak farklı örneklem büyüklüklerinde önerilen ölçütlerin etkili gözlemleri ortaya çıkarmadaki başarıları incelenmiş ve karşılaştırmıştır.

Omay ve Canpolat (2013), çalışmalarında Türkiye’ deki gelir ve çevre kirliliği arasındaki bağlantıyı parametrik olmayan sabit etkili panel veri modelleri ile incelemişler. Panel veri için İstatistiksel Bölge Birimleri Sınıflandırılmasında yer alan 12 farklı bölgeyi içeren verileri kullanmışlar. Uygulamada Türkiye’ de kişi başına düşen kükürt dioksit ve partiküler madde ile kişi başına düşen gayri safi yurtiçi hasıla arasındaki bağlantıyı

incelemişler. Uygulama sonucunda, parametrik olmayan modellerin iyi sonuçlar verdiğini göstermişler.

Aydın ve Alma (2015), çalışmalarında, semiparametrik regresyon modellerindeki etkili gözlemlerin belirlenmesini incelemişler. Bu etkili gözlemlerin belirlenmesinde Cook's uzaklık ölçüsünü ele almışlar. Ayrıca çalışmalarında semiparametrik regresyon modellerinde etkili gözlemlerin belirlenebilmesi için Dffits, Dfbetas ve CovRatio istatistiklerini incelemişler. Etkili gözlemlerin belirlenmesinde her bir ölçüm için gerçek ve benzetim çalışması veri kümelerini kullanmışlar. Uygulama için ayrıca semiparametrik regresyon için Monte Carlo simülasyon çalışmasını da göstermişler.

Toprak (2015), tez çalışmasında, yarı parametrik regresyon modelinde kısmi doğrusal modelde ölçüm hatasından dolayı hatalı ölçülmüş verilerin olmasında ve ölçüm hatasının yoğunluk fonksiyonunun açıklanmamasında parametrelerinin tahmininde uygulanan metodu incelemiştir. İlk kısımda parametrik, nonparametrik ve yarı parametrik regresyon modelleri, regresyonda düzeltme kavramı, pürüzlülük ceza yaklaşımı ve parametrik olmayan regresyonda düzeltme yöntemlerini açıklamıştır. Sonraki kısımda, değişkenleri hatalı modeller üzerinde göstererek değişkenler "bilinen" bir dağılımdan gelen hata ile ölçüldüğünde parametrik olmayan ve yarı parametrik regresyon yöntemlerini göstermiştir. Son olarak, değişkenlerin "bilinmeyen" bir dağılımdan gelen hata ile ölçüldüğünde parametrik olmayan regresyon yöntemleri içerisinde incelenen değişkenlerin hatalı parametrik olmayan regresyon yöntemlerini incelemiştir. Bu yöntem doğrultusunda, parametrik olmayan bölümde ölçüm hataları olması durumunda yarı parametrik regresyon ve hem parametrik hem de parametrik olmayan bölümde ölçüm hataları olması durumunda yarı parametrik regresyon modelleri için kernel dekonvolüsyon tekniğine alternatif olarak görülen yöntemleri açıklamıştır. Ayrıca tahmin edicilerin sonlu örneklem özelliklerini Monte Carlo simülasyon yöntemi ile incelemiştir.

Qingguo ve Kong (2017), makalede, skaler yanıtların koşullu kantilitesinin hem skaler hem de fonksiyonel eş değişkenler ve ek bilinmeyen parametrik olmayan bir fonksiyon terimi ile modellenmiş olduğu fonksiyonel lineer semiparametrik kantil regresyon için parametrik olmayan tahmin yöntemlerini önermişler. Önerilen tahmin edicilerin sonlu örneklem performansını araştırmak için simülasyon çalışması yapmışlar. Ayrıca dikkat eksikliği hiperaktivite bozukluğu özelliklerine sahip 200 örnekten elde edilen gerçek verilerin analizini de göstermişler.

Liu ve ark. (2018), çalışmalarında, süreksiz bilinmeyen tek indeksli modelin tahminini ve minimum ortalama varyans tahmini yöntemini incelemişler. Ayrıca koşullu varyans, minimum ağırlıklı hata kareler ortalamaya sahip olan, merkezlenmiş, sol taraftaki ve sağ taraftaki çekirdek fonksiyonları olan yerel doğrusal uyumlar arasından elde edildiğini göstermişler. Gerçek veri analizi ve simülasyon çalışmasını yöntemle uygulayarak sonuçların anlamlı olduğunu göstermişler.

Shen ve Chen (2018), çalışmalarında, küme içi alt örnekleme değişkenliği ve bilgilendirici küme boyutu için marjinal ortalama için semiparametrik regresyon model seçimini incelemişler. Ayrıca model seçimi için yeniden örnekleme küme bilgi kriteri örnekleme yönteminde uygulamışlar. Yeniden örnekleme küme bilgi kriteri yöntemi, küme içi alt örnekleme göre modelin değişkenliğini açıklayan ek bileşen içerdiğini göstermişler.

Mahmoud (2019), çalışmasında öncelikle parametrik, parametrik olmayan ve semiparametrik regresyon modellerini tanıtmıştır. Bu modellerin bağımlı ve bağımsız değişkenler arasındaki ilişkinin şekli ve rastgele dağılımı hakkında bilgi vermiştir. Ayrıca modeller arasındaki farklılıklar, tahmin yöntemleri ve uygulamalara yer verilmiştir. Uygulamadaki tüm grafikler ve analizler R programında analiz edilmiştir ve üç farklı modelin sonuçları tablo halinde verilmiş ve sonuçlar yorumlanmıştır.

### 3. MATERYAL ve METOT

#### 3.1. Materyal

Bu tezde kullanılan veriler Kahramanmaraş ilinde bulunan 60 tarımsal hayvancılık işletmeye aittir. Barınakların fiyatı üzerinde etkili olduğu düşünülen, işletmenin fiyatı, sağım ünitesi varlığı, karantina bölgesi varlığı, barınak alanı, havalandırma alanı, havalandırma sayısı, karantina alanı varlığı, revir alanı varlığı, gübre çukuru varlığı, doğum ünitesi varlığı ve şehir merkezine uzaklık değişkenleri ele alınmıştır. Modele bazı değişkenler parametrik formda dahil edilirken, bazı değişkenler ise parametrik olmayan formda dahil edilmiştir. Modele parametrik formda dâhil edilen değişkenler,

- Fiyat ( $y$ ) (İşletmenin Fiyatı, TL)
- Sağım ünitesi ( $x_8$ ) (varsa 1, yoksa 0)
- Karantina bölgesi ( $x_4$ ) (varsa 1, yoksa 0)

Modele parametrik olmayan formda dahil edilen değişkenler,

- Alan ( $x_1$ ) ( $m^2$ ),
- Havalandırma alanı ( $x_2$ ) ( $m^2$ )
- Havalandırma sayısı ( $x_3$ )
- Karantina alanı ( $x_5$ ) (varsa  $m^2$ , yoksa 0)
- Revir alanı ( $x_6$ ) (varsa  $m^2$ , yoksa 0)
- Gübre çukur ( $x_7$ )
- Merkeze uzaklık ( $x_9$ ) (km)

özelliklerini kapsamaktadır.

Çizelge 3.1. Çalışmada kullanılan veri seti

İşletme	Fiyat	A	HA	HS	KB	KA	RA	GÇ	SÜ	MU
1	2.679.14	2160	96	10	1	36	142	291	1	155
2	2.774.05	2604	100.8	28	1	25	100	200	1	145
3	1.155.45	780	86.35	12	1	80	0	217	0	145
4	1.738.59	2520	99	9	1	53	100	238	1	30
5	1.307.94	1612	75.6	12	0	0	0	435	1	145
6	1.131.52	1061	37	5	0	0	0	320	0	50
.	-	-	-	-	-	-	-	-	-	-
.	-	-	-	-	-	-	-	-	-	-
.	-	-	-	-	-	-	-	-	-	-
59	1.345.52	1540	45	11	1	40	45	355	0	50
60	1.340.45	1525	42	10	0	0	40	200	1	45

Kullanılan değişkenlerden karantina bölgesi (KB), sağım ünitesi (SÜ) ve doğum ünitesi (DÜ) kesikli değişken, alan (A), havalandırma alanı (HA), havalandırma sayısı (HS), karantina alanı (KA), revir alanı (RA), gübre çukuru (GÇ) ve merkeze uzaklık (MU) değişkenleri ise sürekli değişkendir. Uygulamada parametrik ve parametrik olmayan değişkenleri değerlendirmek için toplamsal model olan semiparametrik regresyon modeli ele alınmıştır. Yapılan istatistiksel değerlendirmelerde SAS 9.4 paket programından yararlanılmıştır.

## 3.2. Metot

### 3.2.1. Regresyon analizi

Regresyon analizi, aralarında neden-sonuç ilişkisi bulunan iki veya daha fazla değişken arasındaki ilişkiyi ölçmek için kullanılan analiz yöntemidir. Bu analiz yönteminde tek bir değişken kullanılarak analiz yapılıyorsa tek değişkenli regresyon, birden fazla değişken kullanılıyorsa çok değişkenli regresyon analizi olarak adlandırılmaktadır. Regresyon analizi, değişkenler arasındaki ilişkinin varlığını, eğer ilişki var ise ilişkinin gücünü ve bu ilişkiyi kullanarak konu ile ilgili tahminler veya kestirimler uygulamak amacıyla kullanılmaktadır. Regresyonda, değişkenlerden biri bağımlı diğerleri bağımsız değişken olarak ele alınmaktadır (Hurvich ve Tsai, 1989; Omay, 2007).

Regresyon yöntemini ilk olarak Adrien Marie Legendre tarafından 1805 yılında ileri sürülmüştür (en küçük kareler yöntemi). Daha sonra 1809 yılında C. F. Gauss aynı yöntemi daha ayrıntılı ele almıştır. Legendre ve Gauss regresyon yöntemini astronomi alanında, uyduların güneş etrafındaki yörüngelerini belirlemek amacıyla kullanmışlardır.

Regresyon ifadesi ilk defa İngiliz istatistikçisi Francis Galton tarafından 19. yüzyılında biyolojik bir çalışmada kullanılmıştır. Galton çalışmasında anne ve babaların boyu ile çocuklarının boyları arasındaki ilişkiyi incelemiştir. Böylelikle Galton'un çalışmaları günümüzde değişkenler arasındaki istatistik ilişkileri inceleyen regresyon analizinin başlangıcı olmuştur.

$Y$ , bağımlı değişken ve  $X$ , bağımsız değişken arasındaki ilişki  $Y = f(X)$  şeklinde tanımlanıyorsa bu fonksiyonel bir ilişki olarak açıklanmaktadır. Regresyon analizi, korelasyon analizi ile birlikte istatistik ilişkinin incelenmesinde uygulanan önemli bir yöntemdir.

Regresyon analizi, parametrik ve parametrik olmayan regresyon olmak üzere iki grupta incelenmektedir. Regresyon analizinin amacı, bağımsız değişkenlerinin bağımlı değişken üzerindeki etkisinin olup olmadığını araştırmak, bağımlı değişkenin alabileceği değerleri tahmin etmek ve bağımsız değişkenlerin bağımlı değişken üzerindeki uyguladığı etkisini analiz etmektir (Green ve Silverman, 1994; Eubank, 2011).

### 3.2.1.1. Parametrik regresyon yöntemi

Parametre, anakütleyle ait olan ortalama, oran, varyans vb. ölçülerdir. Parametrik regresyon, bağımlı ve bağımsız değişkenler arasındaki ortalama ilişkinin matematiksel bir fonksiyonla gösterilmesi ve bu fonksiyonda yer alan parametrelerin açık bir şekilde ifade edilmesidir. Parametrik regresyon, regresyon fonksiyonunun  $x_1, x_2, \dots, x_q$  bağımsız değişkenlerinin bir doğrusal fonksiyonu olarak gösterildiğini varsaymaktadır.  $E(y | X)$  koşullu beklenen değeri  $X$  biliniyorken  $y'$  nin ortalama dağılımının  $X$  ile fonksiyonel ilişkisini açıklamaktadır. Başka bir deyişle  $X = (x_1, x_2, \dots, x_q)$  bağımsız değişkenlerindeki değişime karşılık  $y$  bağımlı değişkenin ortalamasını açıklamaktadır (Speckman, 1988; Schimek ve Michael, 2000).

$$E(y | X) = X\beta \quad (1)$$

veya

$$y = X\beta + \varepsilon_i \quad (2)$$

şeklinde gösterilmektedir.

Burada,  $\varepsilon_i = y - E(y | \mathbf{X})$  ifadesi  $E(y | \mathbf{X})$  koşullu beklenen değerinin  $y'$ den sapması olarak açıklanmaktadır. Parametrik regresyon yönteminde hedef, model uygunluğunu ve modelin oluşturulmasını incelemektir. Doğrusal regresyon modelini ispatlamak için bilinmeyen  $\beta$  parametrelerini açıklamak gerekir. Eşitlik 2'de  $(y_i, x_i)$  gözlem değerlerine karşılık gelen gözlenen nokta ile bu noktaların en küçük kareler metodu ile belirtilmiş olan doğru üzerindeki izdüşümlerinin toplamı,  $\varepsilon_i = y_i - \hat{y}_i$  sıfıra eşit olmalıdır. Aynı zamanda farkın kareler toplamı minimum değere sahip olmalıdır. Böylece eşitlik 2'deki parametre tahminleri eşitlik 3'teki gibi gösterilmektedir.

$$\begin{aligned} \sum_{i=1}^n \varepsilon_i^2 &= \sum_{i=1}^n (y_i - \hat{y}_i)^2 \\ &= \sum_{i=1}^n (y_i - x_i' \hat{\beta})^2 \end{aligned} \quad (3)$$

Parametrik regresyon, varsayımlara dayanmaktadır. Eşitlik 2' de bağımsız değişkenler arasındaki fonksiyonel yapının  $X\beta$  şeklinde doğrusal olduğu ve  $\beta$  parametrelerinin sonlu olduğu varsayılmaktadır. Fakat bağımsız değişkenler arasında doğrusal olmayan ilişkiler var ise parametrik regresyon yöntemi yerine parametrik olmayan regresyon yöntemlerinin uygulanması gerekmektedir (Stefanski, 1985).

### 3.2.1.1.1. Basit doğrusal regresyon

Basit doğrusal regresyon, değişkenlerden birisi bağımlı diğeri ise bağımsız değişken olarak kabul edilen birbiriyle ilişki olan iki değişkenin olduğu yöntemdir.

$$Y_i = \beta_0 + \mathbf{X}_i \beta_1 + \varepsilon_i \quad , \quad i = 1, 2, \dots, n \quad (4)$$

Eşitlik (4)'te  $y$  bağımlı (açıklanan),  $x$  bağımsız (açıklayıcı) değişkeni,  $\beta_0$  ve  $\beta_1$  modelin bilinmeyen parametrelerini ve  $\varepsilon$  hata terimlerini ifade etmektedir.

Basit doğrusal regresyon yöntemini uygulamak için bazı varsayımlarının sağlanması gerekmektedir. Varsayımlardan herhangi birinin sağlanmaması durumunda doğru sonuca ulaşmak mümkün olmayabilir. Berry (1993), parametrik regresyon modelinin varsayımlarını aşağıdaki gibi sıralamaktadır:

1. Bağımsız değişkenlerin varyanslar sıfırdan farklı olmalıdır.

2. Hata terimleri ortalaması sıfıra eşit olmalıdır. Yani,  $\sum \varepsilon_i = 0$  olmalıdır.
3. Hata terimi ve bağımsız değişkenler arasında ilişki yoktur.
4. Hata terimleri arasında korelasyon yoktur.
5. Hata terimleri normal dağılıma sahip olmalıdır.
6. Ayrıca bağımsız değişkenlerin ( $X$ ) nicel ya da nitel, bağımlı değişkenin ( $Y$ ) ise nicel ve sürekli olarak ölçülmüş olması gerekmektedir.

Parametrik yöntemlerle oluşturulan regresyon denklemlerinde, ilişkilerin doğrusal (linear) varsayılması, bağımsız değişkenlerin katsayılarının doğrusal olması anlamını taşımaktadır. Parametrik regresyon modellerin bağımlı değişken ve bağımsız değişkenlerinin doğrusal olma şartı aranmamaktadır. Yani katsayıları itibariyle doğrusaldır. Ayrıca bağımsız değişkenlerin üssel, parabolik veya kareköklü olması parametrik regresyon modellerin doğrusallık özelliğini etkilememektedir (Benedetti, 1977; Chen, 1988).

Parametrik regresyon modellerinde tahmin yapılırken sıklıkla uygulanan yöntemler, En küçük kareler yöntemi ve En çok olabilirlik yöntemidir.

### 3.2.1.1.1.1. En küçük kareler yöntemi

En küçük kareler yöntemi, birbirine bağlı olarak değişen iki büyüklük arasındaki bağlantıyı, gerçeğe en yakın bir denklem şeklinde yazmak için uygulanan regresyon tahmin yöntemi olarak gösterilmektedir. Ayrıca doğrusal varsayımını dikkate alarak parametreleri tahmin etmektedir. Kolay hesaplanabildiğinden dolayı en çok tercih edilen yöntemdir. En küçük kareler yönteminin amacı, gözlemler içerisinde doğrusal bir model oluşturarak, gözlemlerin bu doğruya olan uzaklık arasındaki farkın karesi toplamını minimum yapmaktır. Yani hata kareler toplamını minimize etmeye çalışmaktır. Başka bir ifade ile  $\min \sum e_i^2$  veya  $\min \sum (y_i - \hat{y}_i)^2$  şeklinde gösterilmektedir.

$\beta_0$  ve  $\beta_1$  parametrelerinin tahmini için uygulanan en yaygın yöntemlerden birisi olan en küçük kareler yöntemidir. Tek değişkenli regresyon doğrusunun denklemi,

$$Y_i = \beta_0 + X_{1i}\beta_1 + \varepsilon_i \quad i = 1, 2, \dots, n \quad (5)$$

şeklinde ifade edilmektedir. Denklem 5'te gösterilen  $\beta_0$  ve  $\beta_1$  parametrelerinin değerlerini bulmak için uygulanan en küçük kareler yönteminin, toplam sapmaların karelerinin toplamını minimum yapacak değerlerin bulunmasıdır. Hata terimlerini, gözlenen değerleri ile beklenen değerlerin farkı olarak gösterilebilir. Yani,

$$\hat{\varepsilon}_i = Y_i - \hat{Y}_i \quad (6)$$

şeklinde gösterilmektedir. Denklemden gösterilen ifade ile ölçülen hata terimleri pozitif, negatif ya da sıfıra eşit olur iken bu farkların toplam,

$$\sum_{i=1}^n \hat{\varepsilon}_i = \sum_{i=1}^n (Y_i - \hat{Y}_i) = 0 \quad (7)$$

şeklindedir. En küçük kareler yöntemi,  $\beta_0$  ve  $\beta_1$  parametrelerinin kestirimleri olan ve bu parametrelerinin farkını minimum yapacak biçimde şekil 8'de gösterilmektedir.

$$\text{en küçük } \sum_{i=1}^n \hat{\varepsilon}_i^2 = \text{en küçük } \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 \quad (8)$$

ve  $\hat{\beta}_0$  ile  $\hat{\beta}_1$  tahminleri aşağıda gibidir:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (X_{1i} - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_{1i} - \bar{X})^2} \quad (9)$$

$$\hat{\beta}_0 = \bar{Y} - \bar{X}\hat{\beta}_1 \quad (10)$$

En küçük kareler yöntemi normal denklemler yöntemi ile ortalamalardan farklar yöntemleriyle parametre tahminini gerçekleştirir (Green ve ark., 1985).

### 3.2.1.1.1.2. En çok olabilirlik yöntemi

En çok olabilirlik yöntemi, anakütlenin olasılık dağılımını belirlemek veya bu yönde varsayım yapmaktadır. Ayrıca bu yöntemde, eldeki örneklem verilerinin olma olasılığının en çok olduğu anakütlenin hangi katsayılara sahip olduğunu incelemektedir.

En çok olabilirlik yönteminde hem  $X$  hem de hata teriminin bir dağılıma sahip olduğu varsayılırsa iki değişkenli olabilirlik fonksiyonu için,

$$L = L_x L_e \quad (11)$$

$$\ln L = \ln L_x + \ln L_e \quad (12)$$

$L_x$  marjinal ve  $e$  hata teriminin olabilirliği şeklinde gösterilmektedir. Eşitlik 12'de  $\ln L$ 'yi maksimize eden tahmin ediciler  $\ln L_x$  ile  $\ln L_e$ 'yi maksimize eder. Verilerin iki değişkenli normal dağılım gösterdiği varsayıldığında  $\ln L_x$  ile  $\ln L_e$ 'yi maksimize eden en çok olabilirlik tahmin edicilerine eşit olmaktadır (Cula, 1988).

### 3.2.1.1.2. Çoklu doğrusal regresyon

Çoklu doğrusal regresyon, bir bağımlı (açıklanan) ve birden fazla bağımsız (açıklayıcı) değişken arasındaki fonksiyonel ilişkiyi açıklamaktadır. Çoklu doğrusal

regresyon yönteminin amacı, bağımlı değişkeni etkilediği belirlenen değişkenler aracılığıyla bağımlı değişkenin değerini tahmin etmek, bağımlı değişkeni etkilediği bağımsız değişkenlerden hangisinin ya da hangilerinin bağımlı değişkeni en fazla etkilediğini belirlemek ve aralarındaki ilişkiyi incelemektedir. Böylelikle çalışmalarda daha iyi sonuç oluşturmak için ele alınan tüm etkenleri gözlemlemek ve bağımsız değişkenlerin bağımlı değişkeni hangi yönde ve nasıl etkilediğini tahmin etmek için çoklu regresyon analizi uygulanmaktadır (Fan ve Yao, 2003; Wood, 2000). Çoklu doğrusal regresyon modelinin açık gösterimi;

$$Y = \beta_0 + X_{i1}\beta_1 + X_{i2}\beta_2 + \dots + X_{im}\beta_m + e_i \quad (i = 1, 2, \dots, n) \quad (13)$$

şeklindedir ve

$$Y = \beta_0 + \sum_{k=1}^m X_{ik}\beta_k + e_i \quad (14)$$

şeklinde toplamı olarak ifade edilebilir. Eşitlik 13 ve Eşitlik 14'te  $Y$  bağımlı değişken,  $m$  sayıda bağımsız değişken arasındaki doğrusal ilişkiyi  $Y$  ve  $X$ 'lere ait  $n$  tane gözlem değerinin çoklu doğrusal regresyon modelini ifade etmektedir.  $\beta_0, \beta_1, \dots, \beta_m$  parametreleri regresyon katsayılarını ya da kısmi regresyon katsayılarını göstermektedir.

### 3.2.1.1.2.1. Çoklu doğrusal regresyon analizinde katsayıların tahmini

Çoklu doğrusal regresyon analizinde değişkenler arasında çoklu bağımlılık olmaması şartı sağlanmalıdır. Hatalar rasgele değişkenler olduğu için hatalara ilişkin bazı varsayımların sağlanması gerekmektedir. Bu varsayımlar, aşağıdaki gibi ifade edilmektedir.

$$Var(\varepsilon_i) = \sigma^2, Cov(\varepsilon_i, \varepsilon_j) = 0 \quad i \neq j \text{ ve } E(\varepsilon_i) = 0$$

Böylelikle  $\varepsilon$  hata terimlerinin normal dağılım varsayımlarını gerçekleştirmektedir. Çoklu doğrusal regresyon modelinde bulunan  $\beta$  parametrelerinin tahmin yöntemi olan  $k + 1$  tane  $\beta$  tahmini en küçük kareler yöntemi ile aşağıdaki gibi gösterilmektedir.

$$\hat{\beta} = (X'X)^{-1}(X'Y) \quad (15)$$

$\beta$  parametrelerine ilişkin güven aralıklarının belirlenmesinde, yokluk ve alternatif hipotez testlerinde serbestlik derecesi  $[n - (k + 1)]$  olan  $t$  dağılımından yararlanmaktadır.

Çoklu doğrusal regresyon analizinde, bağımlı ve bağımsız değişkenler arasındaki bağlantının gücünü göstermek için çoklu korelasyon katsayısı ( $R$ ) incelenmektedir. Fakat çoklu belirleme katsayısına ( $R^2$ ) bakılarak yorum yapmak daha uygundur.

$$R^2 = 1 - \frac{\sum \hat{u}_i^2}{\sum y_i^2} \quad (16)$$

Çoklu doğrusal regresyonda bağımlı değişkeni açıklayabilen birden fazla bağımsız değişken bulunmaktadır.  $R^2$  değerinin 0.75'ten büyük olması iyi sonuç vermektedir (Aydın, 2005).

### 3.2.1.2. Parametrik olmayan regresyon yöntemi

Parametrik olmayan regresyon, biri bağımlı değişken ( $y$ ) ve bağımlı değişkenle ne tür bir ilişkide olduğu bilinmeyen bağımsız değişkenin ( $x$ ) olduğu basit parametrik olmayan regresyon modeli,

$$y_i = f(x_i) + \varepsilon_i, \quad i = 1, 2, \dots, n \quad (17)$$

şeklinde ifade edilmektedir. Eşitlik (17)'de  $f \in C^2[a, b]$  olan düzgün dağılım fonksiyonu,  $x_i$  parametrik olmayan bağımsız değişkenlere ait gözlem değerini,  $y_i$  bağımlı değişkene ait gözlem değerini ve  $\varepsilon_i$  bağımsız özdeş olarak dağılan rasgele hata terimlerini belirtmektedir.

Parametrik olmayan regresyon yönteminin asıl amacı, parametreleri tahmin etmekten çok bilinmeyen ortalama fonksiyonu olan  $f(x_i)$ 'i tahmin etmektir.  $y$  bağımlı değişkenin ve  $x = (x_1, x_2, \dots, x_n)$  bağımsız değişkenlerinin ölçümleri arasındaki ilişkiyi gösteren ve en çok kullanılan yöntem,  $E(y|x) = f(x)$  koşullu beklenen fonksiyonu tahmin eden modeldir. Bu ilişkiyi gösteren genel parametrik olmayan regresyon modeli,

$$y = E(y|x) + \varepsilon = f(x) + \varepsilon \quad (18)$$

şeklinde gösterilmektedir. Burada  $y = (y_1, \dots, y_n)'$ ,  $f = (f_1, \dots, f_n)' = (f(X_1), \dots, f(X_n))'$  ve  $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)'$  şeklinde olan  $x$  bağımsız değişkenlerinin bir fonksiyonudur.

Parametrik olmayan regresyon yönteminde tahmin yaparken sınırlayıcı varsayımlar olmamasına rağmen bazı özellikleri olabilmektedir. Bağımsız değişken sayısı çok olduğunda tahmin yapmak zor olmaktadır. Ayrıca ortaya çıkan grafikler karmaşık bir yapıda gösterilmektedir. Bu durumlar sonucunda “boyutluluk sorunu” ortaya çıkmaktadır. Aynı zamanda parametrik olmayan regresyon yöntemle kesikli bağımsız değişkenleri ele almak ve bağımsız değişken sayısındaki artışla beraber  $y$  değişkenine ait etkileri yorumlamak güç olmaktadır. Bu zorluklar yarı parametrik regresyon yöntemi uygulanarak ortadan kaldırılabilir (Wand ve Gutierrez, 1997; Slama, 2001; Tezcan, 2011).

İstatistiksel arařtırmalarda aykırı gözlemlerin etkilerini farklı yollarda ele alan güçlü parametrik yöntemler vardır. Fakat aykırı gözlemlerden dolayı varsayımlar bozulduğundan bu yöntemler bile veri setlerine uygun olmayabilir ve verinin gerçek yapısını bozabilmektedir. Bu durumda parametrik olmayan regresyon yöntemi, aykırı gözlemlerin oluşturduğu veri setleri için uygun bir analiz yöntemi olmaktadır. Parametrik regresyon yönteminde birçok varsayım gerektiğinden dolayı sonuçların güvenilirliğı azalmaya başlamaktadır. Parametrik olmayan regresyon yönteminde ise herhangi bir varsayıma ihtiyaç duyulmamakta ancak bağımsız değışken sayısının çok olmasından dolayı model tahminin ortaya çıkması zorlaşmaktadır. Yarı parametrik regresyon yöntemi, parametrik ve parametrik olmayan regresyon yöntemleri arasında orta yol oluşturmayı hedeflemektedir (Carota ve Parmigiani, 2002; Chandran ve Prajneshu, 2004; Ata ve ark., 2008).

### 3.2.1.2.1. Parametrik olmayan regresyon yönteminin özellikleri

Parametrik olmayan regresyon yönteminin en önemli özelliğı, herhangi bir fonksiyonel formu dikkate almadan bağımlı ve bağımsız değışkenler arasındaki ilişkiyi incelemeye çalışan bir yöntem olarak ifade edilmektedir. Böylelikle parametrik olmama ifadesi, regresyon doğrusunun esnek fonksiyonel formunu göstermektedir. Parametrik olmayan regresyon modeli,

$$y_i = f(x_i) + \varepsilon_i \quad (19)$$

şeklinde ifade edilmektedir.

Parametrik olmayan regresyon yönteminin geliştirilmesindeki amaç, parametrik olmayan tahminci  $f(x_i)$ 'in 'düzgün' tahminlerini oluşturmak amacıyla uygulanmaktadır. Parametrik olmayan yöntemde regresyon ilişkisini incelemek için en önemli avantajları aşağıdaki gibi ifade edilmektedir.

- Parametrik olmayan tahmin yöntemi bağımlı ve bağımsız değışken arasındaki genel ilişkinin incelenmesinde çok yönlü ve esnek kolaylık sağlamaktadır.
- Verinin elde edildiğı anakütlenin hangi dağılımından geldiğı hakkında güçlü varsayımlara ihtiyaç duymamaktadır.
- Parametrik olmayan yöntemler aykırı gözlemlerden çok fazla etkilenmezler.
- Parametrik olmayan yöntemleri yorumlamak daha kolaydır.

- Parametrik olmayan yöntemler, birimlerin büyüklüklerinden daha çok onların sıralarıyla ilgilendikleri için normal dağılım varsayımı gerektiren analizlerin uygulanmadığı birçok durumda uygulanabilmektedir.

Parametrik olmayan regresyon yöntemler, örneğin çekildiği ana kütle belli bir dağılıma dayanmadığı ve örnekleme dağılımından oluşan test istatistiğine dayalı çıkarımlar olduğu için genellikle dağılımdan bağımsız yöntemler olarak ifade edilmektedir. Dağılımından bağımsız olma özelliği, çok sayıda parametrik olmayan yöntemin kilit noktası olarak gösterilmektedir. Parametrik olmayan testte, istatistiksel bir yoğunluk fonksiyonunda, parametrenin değeri hakkında varsayımda bulunulmazken, dağılımdan bağımsız testte örnek çekilen ana kütle hakkında bir varsayımda bulunmaz. Teste dayalı istatistik bir testte istatistiğin örnekleme dağılımı, değişkenin ana kütle dağılımından bağımsız ise dağılımdan da bağımsız olmaktadır. Ayrıca bir yöntemi, dağılımdan bağımsız veya parametrik olmayan şeklinde ifade etmek, varsayımdan bağımsız anlamını taşımamaktadır. Örneğin, ana kütle hakkında sürekli ve simetrik varsayımına bakılabilmektedir (Clark, 1977; Altman, 1992).

### **3.2.1.2.2. Parametrik olmayan regresyon yönteminde yoğunluk tahmini**

Olasılık yoğunluk fonksiyonu, sürekli tesadüfi değişkenin nasıl dağıldığını incelemektedir. Böylelikle sürekli tesadüfi bir değişkenin olasılık yoğunluk fonksiyonu incelendiğinde, sadece ortalama ve varyans gibi ifadeleri değil, ayrıca değişkenin belli bir olasılıkla hangi değerler alacağı da açıklanmaktadır. Yoğunluk tahmin yönteminin uygulanmasının amacı, yoğunluk tahminleri ile verilere uygun model tahminini yapmaktır. Ayrıca birçok sayıda değişkenin olduğu veri setindeki değişkenler arasındaki ilişkiler hakkında bilgi vermektedir.

Parametrik olmayan yoğunluk tahminleri, tahminleyenin dağılımına ilişkin geniş bilgi vermektedir. Parametrik olmayan yoğunluk tahmini 1951 yılında Fix ve Hodges tarafından dağılımsal varsayımlara ilaveten farklı bir yöntem olarak ortaya çıkarılmıştır. Böylelikle parametrik olmayan yoğunluk tahmininde uygulanan birçok yöntem geliştirilmiştir.

#### **3.2.1.2.2.1. Histogram yöntemi**

Parametrik olmayan yoğunluk tahmininde en çok kullanılan ve bilinen yaklaşımlardan histogram yaklaşımı, kutulama olarak da isimlendirilmektedir. Kullanışlı

olmasına rağmen gerçeği çok iyi yansıtmaz. İki veya daha fazla değişken kullanımı için oldukça zordur. Bu özelliklerinden dolayı,  $f(x)$ 'in düzgün tahminin yapılması amacıyla histogram yöntemi dışında birçok parametrik olmayan tahminleyen ortaya çıkmıştır.

Yoğunluk tahmini için uygulanan histogramın şekli, başlangıç noktasına ve kutu genişliğinin seçimine bağlı olmaktadır. Histogramların şekil ve görünüşüne etki eden tek parametre kutu genişliği değildir. Aynı kutu genişliği ve benzer veriler kullanılsa bile histogramlar verinin özellikleri hakkında farklı bilgiler oluşturmaktadır. Çok fazla sayıda gözlemin olduğu durumlarda, histogramların uygulanmasında herhangi bir sakınca yoktur, ama küçük örneklerde dar aralıklı kutu oluşturulması pek dengeli olmamaktadır. Dengeli bir ortalama hesaplamak için geniş aralıklı az sayıda kutu daha uygun olmaktadır.

Olasılık yoğunluk fonksiyonu olan  $f(x)$ 'in tahminleyeni olarak uygulanan histogram, yanlı bir tahminleyendir.  $x_0 = 0$  olmak üzere, histogramın yanlılığı,

$$\text{Yanlılık}\{\hat{f}_h(x)\} = E\{\hat{f}_h(x) - f(x)\} \approx f'\{m_j\}\{m_j - x\} \quad (20)$$

şeklindedir.  $m_j$ , orta noktasında gerçek yoğunluğunun eğiminde yaklaşık yanlılığın mutlak değeri artar ve  $x = m_j$  eşitliği sağlandığında yaklaşık yanlılık sıfıra eşit olur.

Histogramın varyansı,

$$\text{Var}\{\hat{f}_h(x)\} \approx \frac{1}{nh} f(x) \quad (21)$$

Burada,

$n$ : Örneklem büyüklüğünü,

$h$ : düzgünleştirme parametresini

şeklinde ifade etmektedir.

Histogramın varyansı  $f(x)$  fonksiyonu ile orantılıdır.  $nh$  artarken varyans azalır. Böylelikle  $h$  değerinin artmasıyla varyansı azalmakta ve yanlılığı artmaktadır. Yanlılık ve varyansın  $h$  değeri ile ters orantılı olması, yanlılık ve varyans arasında optimal bir denge kuracak  $h$  değerlerinin sonuçlandırılmasını zorunlu kılar. Böylelikle hata kareler ortalaması da bu amaçla uygulanan kriterlerden biridir. Histogram için hata kareler ortalaması (HKO),

$$\text{HKO}\{\hat{f}_h(x)\} = \text{Var}\{\hat{f}_h(x)\} + [\text{Yanlılık}\{\hat{f}_h(x)\}] \quad (22)$$

şeklindedir. Histogramlar, dar kutularda büyük varyans küçük yanlılık, geniş kutularda küçük varyans büyük yanlılık göstermektedir (Bowman ve Adelchi, 1997; Craven ve Wahba, 1979; Hastie ve Tibshirani, 1999).

### 3.2.1.2.2. Kernel tahminleyeni yöntemi

Parametrik olmayan yoğunluk fonksiyonu tahmin yöntemi olan kernel tahmini ilk uygulayan Rosenblatt'dır.  $f(x)$ 'in tahminleyeni olan  $\hat{f}_h(x)$ ,

$$\hat{f}_h(x) = \frac{F_n(x+h) - F_n(x-h)}{2h} \quad (23)$$

Burada,

$F_n$ : n tane gözleme dayalı örnek dağılım fonksiyonu,

$n$ : Örneklem büyüklüğünü,

$h$ : düzgünleştirme parametresini

şeklinde ifade etmektedir.  $x_1, x_2, \dots, x_n$  sürekli olasılık fonksiyonu  $f(x)$ 'e sahip bağımsız ve düzgün dağılmış rasgele değişkenlerdir.  $h = h_n$ , örnek büyüklüğü  $n$ 'in fonksiyonudur ve  $n$  sonsuza giderken sıfıra yaklaşır.

Kernel fonksiyonu pozitif, sürekli, sınırlı ve sıfıra yaklaşan simetrik fonksiyondur. Kernel tahmin yöntemi üzerinde Rosanblatt'tan sonra birçok araştırmacı çalışmalar geliştirmiştir. Araştırmacı Parzen, çalışmasında  $f(x)$  fonksiyonun tahminleyenlerinden oluşan sınıfın tutarlılık ve asimptotik yansızlığını açıklamıştır (Box ve Jenkins, 1970; Altman, 1990).

Kernel fonksiyonları olasılık yoğunluk fonksiyonları olarak incelenmektedir. Çünkü  $K'$ 'nin tanım kümesindeki tüm  $z$ 'ler için  $K(z) \geq 1$ 'dir ve  $\int K(z)dz = 1$  denklemi  $\int \hat{f}_h(x)dx = 1$  şeklindedir. Böylece kernel yoğunluk tahminleyeni olasılık yoğunluk fonksiyonudur. Aynı zamanda  $\hat{f}_h$ ,  $K'$ 'nin türevlenebilme ve süreklilik özelliklerini göstermektedir (Sheather ve Jones, 1991; Brockmann ve ark., 1993).

### 3.2.1.2.3. Parametrik olmayan regresyonda tahmin yöntemleri

Parametrik olmayan regresyonda  $f(x)$  anakütle regresyon eğrisinin belirlenmesiyle bağımlı ve bağımsız değişkenler arasında ilişkinin boyutunu ve yönünü açıklamak için en küçük kareler tahmincileriyle benzerlik göstermektedir. Fakat bu tahmin yöntemini belirlemek için genellikle parametrik olmayan yöntemlere dayanan tahminciler

uygulanmaktadır. Parametrik olmayan regresyon yöntemi ile tahmin yapılmasının avantajları; iki değişken arasında var olan ilişkinin belirlenmesinde esneklik olması, sabit bir parametrik modele bağlı olmaksızın tahmin yapılması ve ardışık  $X$  değerleri arasında kayıp gözlemlerin yerine konulması için esnek yöntemler oluşturmaktır.

### 3.2.1.2.3.1. Kernel regresyon

Parametrik olmayan regresyon analizinde kernel regresyon yöntemi, tahmin yöntemi olarak en çok kullanılan bir yöntem olarak bilinmektedir. Kernel regresyon yönteminde, Nadaraya-Watson Tahminleyeni ve Gasser- Müller Tahminleyeni yöntemleri bulunmaktadır.

#### 3.2.1.2.3.1.1. Nadaraya-Watson tahmin yöntemi

Nadaraya- Watson tahmin yöntemi, herhangi bir  $x$  değerinin komşuluğunda bağımsız değişken değerlerine karşılık gelen bağımlı değişken değerlerinin ağırlıklı ortalaması olarak gösterilmektedir. Nadaraya- Watson tahminci ile oluşturulan regresyon eğrisinin tahmini:

$$\hat{m}(x) = \sum_{i=1}^n w_{ni}(x)y_i \quad (24)$$

Burada,

$y_i$  :  $x_i$ 'lere karşılık gelen ağırlıklı ortalaması,

$n$ : örneklem büyüklüğü,

$w(x)$  : ağırlık fonksiyonu

şeklinde ifade edilmektedir.

Nadaraya- Watson tahmincisine ait en küçük kareler tahmini,

$$\sum_{i=1}^n w_{ni}[y_i - \hat{m}(x)]^2 \quad (25)$$

şeklindedir.

#### 3.2.1.2.3.1.2. Gasser-Müller tahmin yöntemi

Gasser-Müller tahminleyeni, integral kernel tahminleyeni olarak da adlandırılmaktadır. Gasser-Müller tahminleyeni eşit aralıklı dağılan noktalar için uygulansa da eşit aralıklarla dağılmayan noktalar için de uygulanmaktadır. Ayrıca Gasser-Müller tahminleyeni aykırı gözlemlere karşı güçlü bir tahmin yöntemidir.

Nadaraya-Watson tahminleyeni, sınır bölgelerinde önemli ölçüde bir yanlılığa sahiptir. Gasser-Müller tahminleyeni ise bu yanlılığı azaltmakta ve buna karşılık varyansı arttırmaktadır. Böylelikle hem varyansı hem de yanlığı azaltan yöntem olan Gasser-Müller tahminleyeni yöntemi uygulanmaktadır.

### 3.2.1.2.3.1.3. En yakın komşuluk yöntemi

Parametrik olmayan regresyon yönteminde sıklıkla kullanılan analiz yöntemlerinden, en yakın komşuluk tahmin yöntemi tutarlı tahminler ortaya çıkardığından dolayı parametrik regresyon analiz yöntemine bir alternatif olarak ortaya çıkmıştır. Bu yöntemde yakın komşuluk yöntemi isminin verilmesinin sebebi, yoğunluk tahmini belli bir noktanın en yakın komşularından çok yakın komşu noktalarına gitmesidir.

Kernel regresyon yöntemiyle en yakın komşuluk yönteminde, tahminde  $x'$ in komşuluğundaki  $y$  değişkenlerinin ağırlıklı bir ortalaması alınmasıdır. Ancak bu komşuluk mesafesinin sabit olmayıp değişken olmasıdır.

En yakın komşuluk tahminin dezavantajlarından biri, en yakın komşuluk tahmininin düzgün bir eğri tahmini olmamasıdır. Normalde sürekli olmasına rağmen, o nokta için türevleri sürekli değildir. Ayrıca Kernel tahminin aksine en yakın komşuluk tahmininin integrali 1'e eşit olmayacağı için olasılık yoğunluk fonksiyonu özelliğini oluşturmamaktadır.

### 3.2.1.2.3.2. Lokal polynomial regresyon yöntemi

Lokal polynomial regresyon yönteminin amacı, Kernel yönteminde ortaya çıkan eksiklikleri ortadan kaldırmaya çalışmaktadır. Lokal polynomial regresyon; LOcally WEighted Smoothing Scatterplot ifadesindeki baş harflerin birleştirilmesiyle oluşan LOWESS adı ile bilinmektedir. LOWESS yöntemi Cleveland tarafından 1979 yılında ilk olarak uygulanmıştır. Lokal polynomial regresyon, genel lineer regresyon modelinin özel halidir. Ayrıca zaman serileri analizinde uygulanan ağırlıklı bir hareketli ortalamanın bir uzantısı şeklindedir. Regresyon doğrusunu eğrisel duruma getirmek için, bağımsız değişkenlerin karesini ve yüksek dereceden değerlerini almak gerekmektedir. Bağımsız değişkenlerinde  $p$ . dereceden polynomial regresyon modeli,

$$y = \alpha + x\beta_1 + x^2\beta_2 + \dots + x^p\beta_p + \varepsilon \quad (26)$$

Burada,

$y$ : bağımlı değişken,

$x$ : bağımsız değişken,

$\alpha, \beta$ : parametreler,

$\varepsilon$  : hata terimi

şeklinde ifade edilmektedir. Bu model en küçük kareler yöntemi ile çözülebilir.  $p = 1$  ise doğrusal,  $p = 2$  ise kuadratik bir fonksiyon oluşturulur.

Lokal polynomial regresyon,  $w = K \left[ \frac{(x_i - x_0)}{h} \right]$  lokal Kernel ağırlıkları uygulanarak,  $x_0$  odak noktasında Kernel tahmininin polynomial tahmin uzantısı olarak gösterilmektedir.

Burada,

$h$ : düzgünleştirme parametresi,

$K$ : Kernel tahmini,

$w$ : ağırlık fonksiyonunu ifade etmektedir.

Lokal polynomial regresyonda serbestlik derecesi, parametrik regresyonda tam sayı olmak zorunda iken parametrik olmayan regresyonda tam sayı olmak zorunda değildir. Ancak pozitif bir sayı olmak zorundadır.

### 3.2.1.2.3.2.1. Lokal doğrusal regresyon

Regresyon modeline uygulanan lokal polynomial yaklaşımının uzantısı olan Lokal doğrusal regresyon, bir nokta tahmininin oluşmasını sağlayan bir yöntem olarak bilinmektedir. Lokal doğrusal regresyon tahminleyenin diğer tahminleyenlere göre avantajları, hem sabit hem de tesadüfi duruma kolaylıkla uygulanabilir olmasıdır. Ayrıca Nadaraya-Watson ve Gasser-Müller tahminleyenleri sınır etkilerine sahip iken lokal doğrusal regresyonunda buna ihtiyaç yoktur. Böylelikle lokal doğrusal regresyonun bu esnekliklerinden dolayı “Düzene Uyarlanır (Design– Adaptive) Regresyon Tahminleyeni” olarak da isimlendirilmektedir. Lokal doğrusal regresyonun, Nadaraya-Watson ve Gasser-Müller tahminleyenlerine göre asimptotik göreceli etkinliği daha iyi olduğu araştırmalarda kanıtlanmıştır. Ayrıca tüm doğrusal düzgünleştiriciler arasında lokal doğrusal regresyon en iyisi olarak gösterilmektedir.

Lokal doğrusal regresyondaki bazı varsayımlar aşağıdaki gibidir.

1. Regresyon modeli sınırlıdır ve ikinci dereceden türevelere sahiptir.

2. Koşullu varyans sürekli ve sınırlıdır.
3. Bağımsız değişken  $X$ 'in marjinal yoğunluğu olan  $f_x(x)$  sürekli ve  $(a_0, b_0)$  aralığında 0'dan uzakta sınırlıdır.
4. Kernel fonksiyonu sınırlı bir yoğunluk fonksiyonu olarak kabul edilmektedir.

#### 3.2.1.2.4. Düzgünleştirme parametrelerinin seçimi

Parametrik olmayan regresyon yönteminde, tahminin düzgünlük derecesini kontrol eden düzgünleştirme parametresinin belirlenmesi çok önemlidir.  $f(x)$  fonksiyonunun istatistik özellikleri düzgünleştirme parametresi ile ilişkilidir. Düzgünleştirme parametresinin belirlenmesi, tahminin yanlılığı ve varyansı arasında bir denge göstermektedir. Daha geniş bir düzgünleştirme parametresi uygulandığında tahminde daha küçük bir varyans ortaya çıkarken gerçek yoğunluk aşırı düzgünleştirilmiş olabileceği için yanlı bir tahmin ortaya çıkmaktadır. Aynı zamanda daha dar bir düzgünleştirme parametresi seçilirse yanlılık miktarı azalırken tahminin varyansı artmaktadır.

Bir regresyon fonksiyonunun tepe noktası etrafındaki tahmininin yanlılığı önemlidir. Bu amaçla yanlılığın etkilerini azaltmak için düzgünleştirme parametresini düşük almak gerekmektedir. Aynı şekilde regresyon fonksiyonu düz olduğunda yanlılığı fazla arttırmadan varyansın etkilerini azaltmak için daha büyük bir düzgünleştirme parametresi seçilebilir. Ayrıca düzgünleştirme parametre genişliğinin belirlenmesinde bağımsız değişkenin yoğunluğu da önemlidir. Şayet bağımsız değişkenin yoğunluğu düşük ise, yeterli sayıda gözlemi kapsaması için düzgünleştirme parametresi büyük alınabilir.

Parametrik olmayan regresyonda düzgünleştirme parametresinin belirlenmesi için birçok metot ortaya çıkmıştır.

##### 3.2.1.2.4.1. Plug-in seçicileri

Plug-in yöntemi düzeltme parametresi seçim yöntemlerinin tamamen farklı bir sınıfı, yanlılık ve varyansın tahmin edilmesi ile bir risk ölçümünün tahmin edilmesini sağlamaktadır. Bu yöntem ilk kez Scott vd. tarafından 1977 yılında önerilmiştir. Bu yaklaşımda optimal düzgünleştirme parametresi formülünde uygulanan yoğunluk fonksiyonu tahmincisi için yinelemeli süreçler oluşturmuştur. Plug-in yönteminde amaç, eğer içinde bilinmeyen bir parametre içeren değer var ise, bilinmeyen parametre yerine tahmini değeri koyarak sonuç elde edilebilir. Bilinmeyen tek terim yoğunluk fonksiyonunun ikinci dereceden türevi;

$$h_{opt} = \mu_2^{-2/5} \{ \int K^2(\theta) d\theta \}^{1/5} \{ \int f^2(x) dx \}^{-1/5} n^{-1/5} \quad (27)$$

Burada,

$K(\theta)$ : Kernel fonksiyonu,

$f^2(x)$  : yoğunluk fonksiyonun 2. Dereceden türevi,

n: örneklem büyüklüğü,

h: düzgünleştirme parametresi

şeklinde ifade edilmektedir. Eşitlik 27'de  $\int f^2(x)^2 dx$ 'in tahminini elde etmek için yoğunluk fonksiyonunun bilinen bir dağılıma uyma varsayımını uygulamak yerine parametrik olmayan tahmin seçilebilir. Bu fonksiyonun parametrik olmayan regresyon tahmini için de uygun bir düzgünleştirme parametresi ifadesi gösterilmelidir. Böylece gösterilen düzgünleştirme parametresi başlangıç düzgünleştirme parametresi olarak ifade edilmektedir.  $\int f^2(x)^2 dx$ 'in parametrik olmayan tahmini, başlangıç tahmin olarak gösterilmektedir. Başlangıç tahmininde en çok tercih edilen yaklaşım, Sheather ve Jones (1991) tarafından ortaya çıkan denklem çözümü olarak adlandırılan plug-in yaklaşımıdır.

Plug-in yönteminin en önemli özelliği, bant genişliğinin belirlenmesinde çapraz geçerlilik yöntemine göre daha düşük değişkenlik göstermesi ve iyi bir yakınsama oranına sahip olmasıdır. Ayrıca plug-in yönteminde, küçük örneklerde bile anlamlı sonuçlar elde edilmektedir.

### 3.2.1.2.4.2. Çapraz geçerlilik yöntemi

Çapraz geçerlilik- ÇG, düzgünleştirme parametresinin seçiminde kullanılan ve en çok tercih edilen yöntemlerden biridir. Özellikle polinom için derecesinin belirlenmesi problemi veya çok değişkenli parametrik regresyondaki değişken seçimi problemi ile bağlantılıdır. Çapraz geçerlilik yöntemindeki amaç, bağımlı değişkenin her bir  $Y_i$  değeri, geri kalan verilerden tahmin edilmektedir.  $Y_i$ 'nin değerinin tahmini  $g_a^{-i}$  şeklinde gösterilmektedir. Böylece  $\alpha'$ 'nın çapraz geçerlilik skoru,

$$\text{ÇGS}(\alpha) = n^{-1} \sum_{i=1}^n (Y_i - g_a^{-i}(t_i))^2 \quad (28)$$

şeklindedir. Burada  $\alpha'$ 'nın çapraz geçerlilik değeri, çapraz geçerlilik skorunu minimize eden değerdir. Çapraz geçerlilik değerinin minimize edilmesi için bu yöntem değiştirilerek "Genelleştirilmiş Çapraz Geçerlilik –GÇG" hesaplama yöntemi uygulanabilir. Genelleştirilmiş çapraz geçerlilik skoru,

$$G\check{C}GS(\alpha) = n^{-1} \sum_{i=1}^n (g_{\alpha}(t_i) - Y_i)^2 / (1 - T_n(\alpha))^2 \quad (29)$$

şeklindedir. Genelleştirilmiş çapraz geçerlilik yöntemi, büyük örnekler ve hataların bağımsız olduğu durumlarda iyi istatistik sonuçlar göstermektedir. Ayrıca genelleştirilmiş çapraz geçerlilik değeri 0 ile 4 aralığının da olmaktadır.

### 3.2.1.2.4.3. Akaike bilgi kriteri

Parametrik olmayan tahmin yöntemlerinin sıklıkla kullanılan düzgünleştirme kriterlerinden biri olan Akaike Bilgi Kriteri kısaca AIC şeklinde ifade edilmektedir. AIC, Plug-in yönteminden farklı olarak doğrusal tahminci için düzgünleştirme parametresinin belirlenmesinde uygulanabilir. Ayrıca AIC büyük ölçüde değişkenlikten kaçındığı için eksik düzgünleştirme yapmaz.

AIC yöntemi öncelikle parametrik modeller için kullanılmıştır. Daha sonraki yıllarda Hurvich ve Tsai (1989), doğrusal regresyon regresyon ve zaman serisi yöntemleri için küçük örneklerde AIC değerinin oldukça sapmalı olduğunu belirtmiştir ve bu sorunun çözümü için Akaike Bilgi Kriterinin düzeltilmiş versiyonunu geliştirmişlerdir. Böylece Akaike Bilgi Kriterinin düzeltilmiş versiyonun daha az sapmalı olduğu gösterilmiştir. Akaike Bilgi Kriteri' nin düzeltilmiş versiyonu,

$$AIC_c = \log(\hat{\sigma}^2) + \frac{1+tr(H)/n}{1-\{tr(H)+2\}/n} \quad (30)$$

şeklinde gösterilmektedir. Burada varyansın tahmini,  $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n \{y_i - \hat{m}(x_i)\}^2 = y'(I - H)'(I - H)\frac{y}{n}$ , H ise  $n \times n$  boyutlu ağırlık matrisi olarak tanımlanmaktadır. Minimum  $AIC_c$  değerini veren  $h$  değeri optimum düzgünleştirme parametresi değerini gösterir.  $AIC_c$  değeri, küçük örneklerde en küçük kareler çapraz geçerlilik yöntemine göre daha iyi sonuçlar gösterirken büyük örneklerde iki yaklaşım arasında önemli bir farklılık çıkmamaktadır.

### 3.2.1.2.4.4. Bayesian bilgi kriteri

Parametrik olmayan tahmin yöntemlerinin sıklıkla kullanılan diğer düzgünleştirme kriterlerinden biri olan Bayesian Bilgi Kriteri kısaca BIC şeklinde ifade edilmektedir. Bayesian bilgi kriteri için bir ölçüt olan model seçimi modelleri sonlu veri seti arasında; en düşük BIC değerine sahip model tercih edilmektedir. Modelleri yerleştirirken, parametre ekleyerek olasılığı artırmak mümkündür, ancak bunu yapmak aşırı uyuma neden

olabilmektedir. Hem BIC hem de AIC, modeldeki parametrelerin sayısı için bir ceza terimi getirerek bu sorunu çözmeye çalışmaktadır.

#### **3.2.1.2.4.5. Pürüzlük ceza yaklaşımı**

Fonksiyonel veri analizinde verilere bir eğrinin uyumunu sağlarken tek amaç yalnızca iyi bir uyum yapmak değil, aynı zamanda bu amaçla aslında çatışan diğer bir amaç da çok fazla iniş çıkış göstermeyen bir eğri tahmini elde etmektir. Fonksiyonel veri analizinde splaynları düzgünleştirirken yaygın olarak kullanılan Pürüzlü Ceza Yaklaşımının temel amacı eğrinin pürüzlülüğünü ölçmek ve verilerin eğriye uyumu ve eğrinin pürüzlülüğü arasında bir uzlaşma sağlamaktır. Yani iyi modelin elde edilmesi için optimum düzeltme parametresi veya en uygun serbestlik derecenin seçimidir.

#### **3.2.1.2.4.6. Etkili serbestlik derecesi**

Belirsizlik analizi yaparken, çeşitli olasılık dağılımları ile karakterize edilen çoklu varyanslar değerlendirilir ve birleştirilir. Artan karmaşıklık nedeniyle, çoğu zaman serbestlik dereceleri uygunsuz veya tanımsızdır. Bu nedenle, gerçek serbestlik derecelerini yaklaşık olarak tahmin etmek için, çıkarım amacıyla etkili veya eşdeğer serbestlik dereceleri hesaplanır.

#### **3.2.1.2.5. Parametrik olmayan regresyonda hata varyansının tahmini**

Parametrik olmayan regresyonda çıkarımda bulunmak için eğrilere ait varyansın bulunması gerekir. Bağımlı ve bağımsız değişkenlerin sürekli olduğu durumlarda, hata varyansı önemli bir parametreye sahiptir. Parametrik ve parametrik olmayan regresyon yöntemlerindeki ortak varsayım, hata varyansının homojen çıkmasıdır. Bu varsayımının sağlanmaması durumunda tahminleyenlerin uyumu konusunda problem olmaktadır.

Varyansın tahmin edilmesi için ilk olarak güven aralıklarının hesaplanması gerekmektedir. Şayet varyans sabit tutulursa toplu tahminleyen uygulanabilir. Varyansın sabit alındığı zaman ve veri en çok düzen noktalarında tekrarlar içerdiğinde, anakütle varyansını tahmin etmek için hata varyansı uygulanabilir. Genellikle veride az sayıda tekrarlanan noktalar olmaktadır. Varyansın tahmin edilmesinde ayrıca hata kereler toplamından da elde edilebilir.

Varyansın tahmininde kullanılan diğer yöntem, değişken bant genişliğine dayalı yaklaşımdır. Değişken bant genişliği yaklaşımında, varyansı tahmin edilecek sabit bir

noktanın komşuluğundaki  $m$  tane gözlemin kareler toplamı ile başlangıç varyansının tahmininde bulunmasıdır. Tahmininde bulunan başlangıç varyansı tutarlı olmadığı için tahminler Kernel tahmini ile düzgünleştirilerek elde edilmektedir.

Değişen varyanslığın elde edilmesinde uygulanan en kolay yöntem olarak hata terimlerine ilişkin grafikler yorumlanabilir ve hata varyansının yaklaşık eğilimin ne olduğu inceleyebilir. Ancak bu yöntem subjektif olduğundan yorumlama konusunda sorun çıkarmaktadır.

Spline düzgünleştirme yönteminde, bant genişliğinin belirlenmesi hatalara dayalı olarak hesaplanan varyans tahminleyeni ile yakın ilişkilidir. Şayet düzgünleştirme parametresinin değeri uygun değil ise varyansın tahminleyeni büyük bir yanlılığa sahip olacaktır. Genelleştirilmiş Çapraz Geçerlilik yönteminde veriden oluşan düzgünleştirme parametresinin değeri çok küçük seçildiğinde varyans normalden daha küçük bir değer olacaktır.

#### **3.2.1.2.6. Hata terimleri arasında otokorelasyon**

Parametrik olmayan regresyonda uygulanan varsayımdan biri hata terimlerinin bağımsız ve sonlu varyansa sahip hata terimleridir. Bu varsayımlar parametrik regresyon yönteminde sağlanmadığı zaman regresyon eğrisinin tahminleyenlerine ait istatistik kurallar geçersiz olmaktadır. Homojenlik varsayımı altında uygun düzgünleştirme parametresinin seçiminde yeni yöntemler geliştirilmiştir. Bu yöntemlerden Çapraz Geçerlilik ve Genelleştirilmiş Çapraz Geçerlilik, beklenen hata kareler tahmininin tahminleyeni ve gözlenen hata karelerinde yapılan düzeltmeyi incelemektedir. Böylelikle gözlemlerdeki hatalar, yeni gözlemlerdeki hatalardan bağımsız olur. Ayrıca düzeltmenin oluşturulabilmesi için korelasyon fonksiyonunun bilinmesi gerekir. Ancak genel olarak bu fonksiyon bilinmediği için veriden tahmini yapılır. Bu tahmin için uygulanan yöntemlerden biri de standart momentler yöntemi olmaktadır.

Bazı araştırmalara göre, korelasyon var olduğu sürece birbiri ile ilişkili hatalara sahip verilere kernel regresyon uygulandığında anlamlı sonuçlar oluşmaktadır. Şayet veriden uygun bir bant genişliği oluşturulursa, bu bant genişliğini açıklayan yöntem korelasyon fonksiyonu ile düzeltilmelidir. Korelasyon tahminindeki yanlılık tahminler ve gözlemler arasındaki korelasyondan kaynaklanır. Böylelikle bant genişliği seçimi geniş olmalıdır. Korelasyon tahmini yaklaşık olarak elde edilirse, bant genişliği seçimi de otomatik olarak elde edilebilir. Şayet korelasyon tahmini yapılamazsa bant genişliği

subjektif olarak diğer bant genişliklerine bakılarak tahmini yapılabilir. Ayrıca hata terimleri arasındaki bağımlılığın seçilmesinde farklı yöntemlerin uygulanması önerilmektedir.

### **3.2.1.2.7. Güven aralığı ve güven bantlarının seçimi**

Güven aralığı, istatistik biliminde bir anakütle parametresi için bir çeşit aralık kestirimi olup bir çıkarımsal istatistik çözüm aracıdır. Bir anakütle parametre değerinin tek bir sayı ile nokta kestirimi yapılacağına, bu parametre değerini kapsayabilecek iki (alt ve üst sınır) sayıdan oluşan bir aralık bulunur. Böylece güven aralıkları bir kestirimin ne kadar güvenilir olduğunu göstermektedir. Güven aralığı tarif edilirken genelde %95 değeri kullanılır, bu değer elde ettiğimiz sonuçları söylerken ne kadar emin olduğumuzu gösterir. Bir başka deyişle %95 emin olduğumuz durumda %5 hatalı da olacağımızı kabul ettiğimizi gösterir.

Güven aralıklarının elde edilmesi için bir  $f$  fonksiyonun örnekleme dağılımının bilinmesi gerekir. Yanlılık terimi, regresyon fonksiyonunun yüksek dereceli türevlerine bağlı olduğu için uygulamalarda daha karmaşık bir yapıda olabilir ve bu yüzden genellikle ihmal edilir. Nadaraya-Watson regresyon tahminleyeni için güven aralığı elde edildiği durumda, limit dağılımlarındaki yanlılık teriminin ortadan kalkması gerekir.

Regresyon fonksiyonunun tahmininde parametrik olmayan regresyon tahminleyenler için oluşan problemlerden birisi yanlı olmalarıdır. Böylece güven bantları tahminleyenin yanlılığını ele alarak regresyon eğrisini düzgün biçimde ortalayarak oluşturulmalıdır.

#### **3.2.1.2.7.1. Güven aralıklarının yinelemeli örnekleme yöntemi ile seçilmesi**

Bir tahminleyenin dağılımı hakkında bilgi almak için kullanılan bir yeniden örnekleme yöntemine yineleme örnekleme yöntemi yani Bootstrap yöntemi denilmektedir. Yinelemeli örnekleme yöntemi, parametrik olmayan regresyon yöntemindeki kullanılma amaçlardan biri veriye uygulanan düzgünleştirmenin değerini belirleyen düzgünleştirme parametresinin belirlenmesinde kullanılmasıdır. Bir diğer amaç ise regresyon eğrisi için güven aralıklarının elde edilmesidir. Ortalama ve varyansın tahmininde kullanılan direkt yöntemi, düzgünleştirme parametresi ve güven bantları hesaplamalarında kullanılabilir. Yinelemeli örnekleme yönteminde, güven bantların olasılıklarının elde edilmesinde direkt

yöntemine göre daha az uygulanmasına rağmen, eğiklik vb. standart olmayan özelliklerin elde edilmesinde yinelemeli örnekleme yönteminin avantajlarından birisidir.

Parametrik olmayan regresyon eğrisinin tahmininde güven bandı elde etmek için uygulanan yöntemler, yanlılık düzelten ve klasik Bonferroni bant yöntemi veri odaklı yöntemlerdir. Bonferroni bant yöntemi kolay hesaplama yapıldığından dolayı avantaj elde etmektedir. Ayrıca standart kernel tahminleyeni etrafında ortalanarak asimptotik olarak yanlılık düzeltilmesine gerek yoktur. Yanlılık düzelten yöntemler küçük örnekler için çoğu durumda korumacı olmaktadır. Örnek büyük olduğunda her bir noktada Bonferroni bantlarını yerleştirmek gerekli değildir. Yani düzenin her bir alt kümesinin kullanılması bantların çizimi konusunda daha iyi sonuçlar oluşmaktadır. Böylece Bonferroni bantlarının boyutu azalarak korumacı hale gelecektir ve yanlılık düzelten yöntemlerde rekabet gücünün oluşmasına sebep olacaktır.

### 3.2.1.2.8. Hipotez testleri

Parametrik regresyonda, regresyon katsayılarını tahmin etmek ve buna bağlı olarak güven aralıklarının hesaplanması ve hipotez testleri gibi istatistik çıkarımlar ele alınmaktadır. Parametrik olmayan regresyonda, regresyon katsayısı hesaplanmadığı için regresyon fonksiyonu oluşturulmaktadır. Parametrik olmayan regresyonda hipotez testinin yetersizliği durumunda çeşitli yaklaşımlar gösterilmektedir. Bu yaklaşımlardan biri parametrik regresyonda en çok tercih edilen en küçük kareler yöntemindeki gibi, alternatif iç içe modellerin hata kareleri toplamını incelemektedir. Doğrusal olmayan bir modelin istatistiği, parametrik olmayan regresyon ile doğrusal basit regresyon modeli karşılaştırılabilir. Doğrusal ilişki doğrusal olmayan modelin özel hali olduğu için birbiri ile iç içe olmaktadır. Ayrıca doğrusal regresyon modellerindeki yaklaşım, modellerin hata kareleri toplamını incelemektir. Parametrik olmayan regresyon modellerinde de aynı şekilde hata kareler toplamı incelenebilir.

Hipotezler aşağıdaki gibi gösterilmektedir.

$$H_0 = \text{Değişkenler arasındaki ilişki doğrusaldır.}$$

$$H_1 = \text{Değişkenler arasındaki ilişki doğrusal değildir.}$$

Test istatistiği,

$$F = \frac{(RSS_0 - RSS_1) / (df_{mood} - 2)}{RSS_1 / df_{res}} \quad (31)$$

Burada,

$RSS_0$  = Doğrusal modelin hata kareler toplamı,

$RSS_1$  = Parametrik olmayan regresyon modelinin hata kareler toplamı,

$df_{mood}$  = Parametrik olmayan regresyon modelinin serbestlik derecesi,

$df_{res} = n - df_{mood}$

olarak gösterilmektedir. Parametrik olmayan tahminleyenler, parametrik tahminleyenlere göre daha az varsayım gerektirdiğinden dolayı daha çok tercih edilmektedir. Böylelikle koşullu ortalama veya veri oluşturan süreçlere ilişkin bilgi noksanlığında parametrik olmayan yöntemler kullanılmaktadır (Tezcan, 2009).

Nadaraya-Watson tahminleyeni uygulandığında İç İçe Pivotal Yinelemeli Örnekleme yöntemi, parametrik olmayan regresyonda anlamlılık testi için bilinen bir yöntemdir. Bu yöntemde uygulanan test istatistiği, bilinmeyen bir koşullu ortalamanın kısmi türevlerinin parametrik olmayan tahminleri uygulanılmaktadır. Test istatistiğinin yokluk hipotezinin oluşturulmasında asimptotik teori yerine yeniden örnekleme yöntemi kullanılmaktadır. Bu yönteminin kullanılma amacı, asimptotik yaklaşımların oluşturulması için gerekli koşulların yetersiz olduğu durumlarda kullanılmasıdır.

Test istatistiğinin hipotez testlerinden dolayı oluşturduğu avantajlar, ilk olarak önemlilik testi güvenilir büyüklüktedir ve sürekli olarak iki kere türevlenebilen güçlüktedir. İkinci olarak, test istatistiği bant genişliğine bağlıdır ve bant genişliği belirlenmesinde önemli derecede duyarlıdır. Son olarak, test istatistiği parametrik modellere dayalı test istatistikleri ile aynı yakınsama oranında olmaktadır.

### 3.2.1.2.9. Parametrik olmayan regresyonda uyum iyiliğinin hesaplanması

Parametrik regresyon yönteminde, belirleme katsayısı bağımlı değişkenin azaltılmasında bağımsız değişken veya değişkenlerin hangi ölçüde etkili olduğunu incelemektedir. Parametrik olmayan regresyon yönteminde belirleme katsayısını bulmak mümkündür.

$$R^2 = \frac{[\sum_{i=1}^n (y_i - \bar{y})(\hat{y}_i - \bar{y})]^2}{\sum_{i=1}^n (y_i - \bar{y})^2 \sum_{i=1}^n (\hat{y}_i - \bar{y})^2} \quad (32)$$

Eşitlik 32' deki denklemde,

$y_i$  = Bağımlı değişkenin gözlenen değerini,

$\bar{y}$  = Bağımlı değişkenin ortalamasını,

$\hat{y}_i$  = Bağımlı değişkenin parametrik olmayan regresyon ile tahmin değerini,

$i = 1, 2, \dots, n$ ; Gözlem sayısını ifade etmektedir.

Elde edilen  $R^2$  değeri  $[0,1]$  aralığında bulunmaktadır. 0 değeri tahminin gücü olmadığını, 1 değeri ise örnek verisine çok iyi bir uyum sağladığını belirtmektedir (Tezcan, 2009).

### 3.2.1.3..Semiparametrik regresyon yöntemi

Regresyon modellerinde bağımlı değişken ile bağımsız değişkenler arasındaki ilişkinin fonksiyonel şekli ile ilgili herhangi bir varsayımının olmaması parametrik olmayan regresyon modellerinin en önemli avantajı olmaktadır. Parametrik olmayan regresyon modellerinin sağladığı esneklik, bu modeli uygulanabilir hale getirmektedir. Fakat bu modelde düzgünleştirme işlemini hesaplamak oldukça zor olmaktadır. Ayrıca bağımsız değişken sayısı arttıkça oluşan boyut sorunundan dolayı parametrik olmayan tahminlerin güvenilirliği giderek azalacaktır. Parametrik olmayan regresyon yaklaşımlarda değişkenler arasındaki ilişkinin fonksiyonel şekli ile ilgili yorumu veriye bağlı olarak fonksiyonel şekli gösterir. Ayrıca parametrik regresyon modellerinin tahmininde fonksiyonel şeklin bilindiği varsayımı, istatistiksel sonuçların elde edilmesi ve tahmin sonuçlarının yorumlanması parametrik olmayan regresyon modeline göre daha kolaydır. Fakat parametrik regresyon modelinde fonksiyonel şeklin bulunması varsayımı daha sınırlayıcıdır. Aynı zamanda veri setinde parametrik olmayan bağımsız değişkenler olabilir. Verinin bir kısmı için model hatalı kurulmuş olur ve tahmin sonuçlarıyla ilgili yapılan yorumlar güvenilir olmaz. Böylelikle, regresyon modelinde ilişkinin fonksiyonel şekli bilinmediği zaman parametrik regresyon model veriyi daha iyi açıklayamazken, parametrik olmayan regresyon analiz de veri hakkındaki önemli bir yorumun yapılmamasına sebep olabilir. Bu sorunların üstesinden gelebilmek için bir kısmı parametrik, diğer kısmı parametrik olmayan değişkenlerden oluşan regresyon modelleri olan semiparametrik regresyon modeli (yarı parametrik regresyon modeli) uygulanmaktadır (Rahman ve Ullah, 2002).

Semiparametrik regresyon modelleri standart regresyon yöntemlerini genelleştiren ve her bir değişkenin etkisinin uygun bir şekilde yorumlanmasını sağlayan toplamsal modellerin özel bir durumu olarak incelenmektedir. Değişkenlerin bir kısmının parametrik

bir kısmının ise parametrik olmayan deęişken olarak oluşturulan semiparametrik regresyon modeli;

$$y_i = \alpha + f_1(x_1) + \dots + f_j(x_j) + x_{j+1}\beta_1 + \dots + x_k\beta_k + \varepsilon \quad (33)$$

şeklinde gösterilmektedir. Semiparametrik regresyon modelindeki  $j$  tane deęişkenin  $y$  bağımlı deęişkeni üzerinde doğrusal olmayan etkisi bulunmakta ve modelin parametrik olmayan kısmını göstermektedir. Diğer deęişkenlerin ise  $y$  bağımlı deęişken üzerinde doğrusal etkisi bulunmakta ve modelin parametrik kısmını göstermektedir. Ayrıca modelin parametrik kısmında kukla deęişkenler gibi kesikli deęişkenler de olabilmektedir. Modelin parametrik olmayan kısmında ise birden fazla deęişken olduğunda parametrik olmayan model için oluşan sakıncalar bu modeller için de geçerli olacaktır. Bu sorunları ortadan kaldırmak için modelin parametrik olmayan kısmındaki deęişkenler modele toplamsal olarak dâhil edilerek yeni model oluşturulmaktadır (Zhongyi ve Baocheng, 2001; Aydın, 2005).

Bir model oluştururken ilk önce deęişkenler belirlenmektedir. Deęişkenler belirlendikten sonra modelin fonksiyonel şeklinin ya da matematiksel şeklinin belirlenmesi gerekmektedir. Matematiksel şekli belirlerken öncelikle grafiklerin incelenmesi gerekmektedir. Bağımlı deęişken ile bağımsız deęişkenlerin ayrı ayrı oluşturulan grafikler incelenerek, bağımlı deęişken ile bağımsız deęişkenler arasındaki ilişkinin yapısı hakkında bilgi sahibi olunmaktadır. Her bir deęişkenin ilişkisi incelendikten sonra bağımsız deęişkenlerin bir kısmının parametrik olmayan bir kısmının ise parametrik ilişki için uygun ise semiparametrik regresyon modeli en uygun model olarak tercih edilmektedir (Çağlayan, 2002). Semiparametrik regresyon modelinin parametrik kısmı doğrusal (linear) ilişkiyi, parametrik olmayan kısmı ise doğrusal olmayan (nonlinear) ilişkiyi göstermektedir. Böylelikle semiparametrik regresyon modellerine yarı doğrusal modeller de denilmektedir (Lee, 1990; Ichimura, 1993; Hong, 1998).

Semiparametrik regresyon modellerindeki parametrik kısmında dönüşüm yöntemleri uygulanarak doğrusallaştırılabilen yapıda da incelenebilir. Semiparametrik modelin parametrik kısmı şeklinde tahmin yapılabilir. Deęişkenler arasındaki ilişkiyi en iyi şekilde açıklayacak model çeşitli denemeler sonucunda da elde edilebilir. Ayrıca deęişkenler arasında şekli açıklanamayan ilişkiler olduğunda farklı matematiksel kalıplar ya da farklı deęişkenler için parametrik olmayan ilişkileri oluşturan modellerin denenmesi ve en uygun modelin belirlenmesi gerekmektedir (Çağlayan, 2002).

### 3.2.1.3.1. Semiparametrik regresyon modellerin tahmini

Semiparametrik regresyon modellerin ve toplamsal modellerinin tahmininde tekrarlı(iteratif) algoritmalar kullanılmaktadır. Bu modellerin tahmini için geliştirilen birçok algoritma olup bu algoritmalar farklı bilgisayar programında uygulanmaktadır. Bu algoritmaların analiz edilmesinde R programı ve SAS programı en çok tercih edilen programlardır (Zeger ve Diggle, 1994).

Semiparametrik regresyon modelinde bağımsız değişkenler ilişkisiz olduğunda, birçok sayıda parametrik olmayan değişkenin olduğu semiparametrik regresyon modellerinin tahminini yapmak oldukça kolay olmaktadır. Yani bağımsız değişkenler ilişkisiz (ortogonal) ise modelin parametrik değişkenlerinin katsayıları en küçük kareler yöntemi uygulanarak, parametrik olmayan değişkenlerinin kısmi regresyon fonksiyonlarını ise splayn gibi parametrik olmayan yöntemler ile tahmini yapılabilmektedir. Fakat semiparametrik regresyon modellerinde modelin parametrik ve parametrik olmayan değişkenleri birbiriyle ilişkili olabilmektedir. Böylece değişkenler arasındaki ilişkileri dikkate alındığında farklı algoritmalara ihtiyaç duyulmaktadır. Bu algoritmalarından en çok tercih edilen Newton-Raphson algoritması ve geriye uyum (backfitting) algoritmasıdır (Mammadov, 2005; Liu ve ark., 2013).

#### 3.2.1.3.1.1. Geriye uyum (Backfitting) algoritması

Hastie ve Tibshirani tarafından tanıtılan backfitting algoritması 1990 yılında ortaya çıkmaktadır. Geriye uyum algoritması parametrik ve parametrik olmayan değişkenleri tahmin edebilen en kolay yöntemlerden biridir. Modelin tahmini yapılırken, bağımsız değişkenler birbirine dik ise modelin parametrik kısmı iki değişkenli modeller olarak en küçük kareler yöntemi uygulanarak tahmin yapılabilir. Parametrik olmayan değişkenlerin tahmininde ise splaynlar uygulanabilir. Şayet bağımsız değişkenler arasında korelasyon yok ise toplamsal modeller veya semiparametrik regresyon modelleri tahmin edilirken bağımsız değişkenler arasında ilişki varsa ilişkiyi dikkate alacak yöntemlerden geriye uyum algoritmasına ihtiyaç duyulmaktadır. Geriye uyum algoritması parametrik ve parametrik olmayan değişkenleri tahmin ederken bağımsız değişkenler arasındaki korelasyonu da dikkate almaktadır. Geriye uyum algoritması ayrıca kısmi regresyon fonksiyonlarını da önermektedir.

$$y = \alpha + f_1(x_1) + f_2(x_2) + \varepsilon \quad (34)$$

Eşitlik (28)'de iki bağımsız değişkenli toplamsal modelinde,  $f_2$ 'nin gerçek fonksiyonel şeklinin bilindiği fakat  $f_1$ 'in bilinmediğini varsayalım. Böylece  $f_1$ 'in tahmini için modeli kısmi regresyon fonksiyonu olarak aşağıdaki gibi gösterilmektedir.

$$y - \alpha - f_2(x_2) = f_1(x_1) + \varepsilon \quad (35)$$

denkleminde  $x_1$ 'e karşı  $y - \alpha - f_2(x_2)$ 'nin düzgünleştirilmesi  $f_1(x_1)$ 'in tahminini oluşturmayı sağlamaktadır. Yani, bir kısmi regresyon fonksiyonunu bilmek diğer kısmi regresyon fonksiyonunu tahmin etmeyi amaçlamaktadır. Normalde hiçbir regresyon fonksiyonunu öğrenmek mümkün değildir. Fakat  $f$ 'lerden herhangi biri için bir başlangıç değeri belirlenirse toplamsal modellerin tahmini için kısmi regresyon fonksiyonları tekrarlı metotlar ile açıklanır.

$$y_i = \alpha + f_1(x_1) + \dots + f_k(x_k) + \varepsilon \quad (36)$$

Eşitlik 36'daki modelin tahmini için,  $S_j$ , sütunları  $X$  tahmin tahminlerinden oluşan matrisi göstermektedir.  $X$ , kolonları  $x$  değişkenlerinden oluşan model matrisini göstermektedir. Toplamsal modellerin tahmini için backfitting algoritması için başlangıç değerleri  $\alpha = \bar{y}$  ve  $S_j = X$  ( $j = 1, \dots, m$ ) olarak alınır. Her  $x$  değişkeni için kısmi hatalar hesaplanır.  $x_i$  değişkeni için tahmin edilen kısmi hatalar,

$$\hat{e}_p^j = y_i - \sum_{i=2}^k S_j - \alpha \quad (37)$$

$x_1$  bağımsız değişkeni etrafında  $e_p^j$  düzgünleştirilir. Bu durumda parametrik olmayan regresyon modeli belirlenmelidir.  $S_j$ 'deki  $x_1$  değişkeni,  $x_i$ 'nin düzgünleştirilmiş tahminleri ile değiştirilir. 2'den  $k$ 'ya kadar olan her  $x$  değişkeni için aşamalar tekrarlanır. Eşitlik (32)'deki model ile hata kareler toplamı hesaplanır.

$$HKT = \sum_{i=1}^n [(y_i - \sum_{i=1}^k S_j)^2] \quad (38)$$

Hata kareler toplamındaki değişim belirli bir tolerans seviyesinde ise model yakınsar ve algoritma durur. Belli bir tolerans seviyesinde değil ise bu işlem hata kareler toplamındaki değişim belirli bir tolerans seviyesine kadar devam eder. Geriye uyum algoritması durduğunda  $S_j$ 'nin her sütunu için  $x$  değişkeninin parametrik olmayan tahmin kısmını gösterir. Bu tahminler  $x$  değişkenleri arasındaki ilişkiyi dikkate alır. Böylelikle, üç  $x$  değişkenine ait bir toplamsal model tahmin edildiğinde  $\hat{f}_1$ 'in grafiği,  $x_2$  ve  $x_3$  bağımsız değişkenlerinin etkisi sabit edildiğinde  $x_1$ 'in  $y_i$  üzerindeki etkisi şeklinde ifade edilmektedir. Geriye uyum algoritmasının çok fazla varyasyonu vardır. Varyasyonlardan

en çok tercih edilenlerden biri başlangıç değeri olarak sıradan en küçük kareler tahmincilerini uygulamaktır. Geriye uyum algoritması semiparametrik regresyon modelleri için benzer adımları göstermektedir. İlk olarak modeldeki her bir bağımsız değişken için kısmi hatalar seçilir. Seçilen değişkenin doğrusal olmayan uyumu var ise bu değişken için kısmi hatalar, aynı bağımsız değişkene karşı düzgünleştirme metodu ile düzeltilir. Eğer seçilen değişkenin doğrusal uyumu var ise düzgünleştirme metodu yerine sıradan en küçük kareler metodu uygulanmaktadır (Wahba, 1990; Wang, 2011).

### 3.2.1.3.1.2. Kısmi splayn yaklaşımı

Bir regresyon yönteminde doğrusal olmayan şekli dikkate almanın bir yolu da örnekleme alt gruplara ayırıp her bir alt grup için doğrusal bir ilişki tahmininde bulunmaktır. En küçük kareler metodu ile tahmin edilebilen parçalı doğrusal regresyon modeli doğrusal olmayan şekli dikkate almasına rağmen regresyon fonksiyonu tüm noktalarda türevlenebilir olmayabilir. Böylelikle parçalı doğrusal regresyona benzer bir yaklaşım olan splayn regresyon uygulanmaktadır. Splayn, her bir örneklem parçası için doğru parçalarını tahmin etmek yerine polinomları tahmin etmektedir. Splaynlar her zaman düzgün olan tahmin edilmiş fonksiyonlar olarak gösterilmektedir. Splaynlar nümerik analizde veri enterpolasyonu amacı ile uygulanırken, istatistikte veri düzgünleştirme amacı ile uygulanmaktadır. Splayn tahmini, düğüm noktalarının belirlenmesine ve konumuna duyarlıdır (Raz, 1990; Thomas, 1991).

Kısmi splayn yaklaşımında, ele alınacak  $(y_i, z_i, x_i)_{i=1}^n$  gözlem veya ölçüm değerleri için semiparametrik regresyon denkleminde uygulandığında,

$$y_i = z_i' \beta + f(x_i) + \varepsilon_i \quad (39)$$

hata kareler toplamını minimum yapan ve  $f$  fonksiyonunu ve  $\beta$  parametreleri en küçük kareler yöntemine göre tahmini için,

$$HKT = \sum_{i=1}^n \{y_i - z_i' \beta - f(x_i)\}^2 \quad (40)$$

eşitlik 40' taki denklem kullanılmaktadır.

### 3.2.1.3.1.3. Speckman yaklaşımı

Speckman yaklaşımı, Speckman (1988) tarafından ele alınan ve kısmi splayn yaklaşımına alternatif olarak ortaya çıkmıştır. Speckman yaklaşımını açıklamak için eşitlik

(33)' teki semiparametrik regresyon modelindeki  $z_i$  bağımsız değişkenlerinin  $x_i$  parametrik olmayan bağımsız değişkenine göre bir regresyon olarak ele alındığında,

$$z_i = m(x_i) + \varepsilon_i \quad (41)$$

şeklinde model ifade edilmektedir. Bu denklemde  $m = (m(x_1), \dots, m(x_n))$ ,  $x$ 'in düzgün fonksiyonu ve  $\varepsilon_i$ , hata terimlerinin vektörleri olarak tanımlanmaktadır (Speckman, 1988).

Bu çalışmada yapılan istatistik analizler için SAS 9.4 programı kullanıldı. Semiparametrik regresyon modelinin tahminlenmesinde geriye uyum(backfitting) algoritması ve kısmi splayn yaklaşımından yorumlandı.



#### 4. BULGULAR ve TARTIŞMA

Tarımsal hayvancılık işletme satış fiyatlarının belirlenmesinde, işletme özelliklerinin fiyat üzerinde etkili olduğu bilinen bir gerçektir. Bu özelliklerin işletmelerin satış fiyatlarını nasıl etkilediğinin bilinmesi, üretim maliyetinin belirlenmesi (işçi, elektrik, su, paketlenme, ulaşım, ... vb), ürünlerin arz-talep zincirinde etkin ve karlı bir şekilde sunulabilmesi, işletme sahibine üretimin sürdürülebilirliği ve işletmenin geleceği ile ilgili önemli bilgiler verecektir. Çünkü sağım ünitesi varlığı, karantina bölgesi varlığı, barınak alanı, havalandırma alanı, havalandırma sayısı, karantina alanı varlığı, revir alanı varlığı, gübre çukuru varlığı, doğum ünitesi varlığı ve şehir merkezine uzaklık durumu doğrudan ya da dolaylı olarak üretimin etkinliği ve ek yatırım gerekip gerekmediği ile alakalıdır.

Bu çalışmada, işletme fiyatı üzerinde etkili değişkenlerin, bir kısmının doğrusal olarak yer aldığı çok değişkenli parametrik regresyon modeli, bir kısmının da doğrusal olmayan değişkenlerin yer aldığı parametrik olmayan regresyon modelini de içeren semiparametrik regresyon modeli sonuçları elde edilmiştir. Çalışmada kullanılan veri seti Kahramanmaraş ili ve çevresinde bulunan 60 tarımsal hayvancılık işletmeye aittir. Parametrik regresyon, parametrik olmayan regresyon ve semiparametrik regresyon modelleri için yapılan analizlerde SAS 9.4 paket programı kullanılmıştır.

Ayrıca çalışmada parametrik ve parametrik olmayan bileşenleri içeren semiparametrik regresyon modelinin tahmini için splayn düzeltme yöntemi uygulanmıştır. Uygun düzeltme parametrelerinin seçilmesi ve düzeltme matrislerinin yardımıyla splayn düzeltmeler tahmin edilmiştir. Düzeltme parametreleri için bu tahminler, en çok kullanılan geriye uyum (bacfitting) algoritması ile elde edilmiştir.

Öncelikle tüm bağımsız değişkenlerin, tarımsal işletmenin satış fiyatı üzerinde doğrusal etkiye sahip olduğunu varsayılarak aşağıda ifade edilen doğrusal regresyon modeli tanımlanmıştır (Fiyat(y), Alan ( $x_1$ ), Havalandırma alanı ( $x_2$ ), Havalandırma sayısı ( $x_3$ ), Karantina bölgesi ( $x_4$ ), Karantina alanı ( $x_5$ ), Revir alanı ( $x_6$ ), Gübre çukuru ( $x_7$ ), Sağım ünitesi ( $x_8$ ), Merkeze uzaklık ( $x_9$ )).

$$y = \beta_0 + x_1\beta_1 + x_2\beta_2 + x_3\beta_3 + x_4\beta_4 + x_5\beta_5 + x_6\beta_6 + x_7\beta_7 + x_8\beta_8 + x_9\beta_9 + \varepsilon \quad (42)$$

Bu modele ait fiyat, alan, havalandırma alanı, havalandırma sayısı, karantina alanı, karantina bölgesi, revir alanı, gübre çukuru, sağım ünitesi, merkeze uzaklık değişkenleri için veri sayısı (N), aritmetik ortalama ( $\bar{x}$ ), standart sapma (S), medyan ve minimum-maksimum değerleri Çizelge 4.1.' de verilmiştir.

Çizelge 4.1. Değişkenlere ait tanımlayıcı istatistik değerleri

Değişkenler	N	$\bar{x}$	S	Medyan	Min-Max
Fiyat (y)	60	1913524	706285	1982230	955230-3150120
Alan ( $x_1$ )	60	1627.07	616.77	1600	780-2790
Havalandırma alanı( $x_2$ )	60	283.40	1115.51	80	37-8635
Havalandırma sayısı ( $x_3$ )	60	13.72	4.77	12.50	5-28
Karantina bölgesi ( $x_4$ )	60	0.58	0.49	1.00	0-1
Karantina alanı ( $x_5$ )	60	26.82	23.82	40	0-80
Revir alanı ( $x_6$ )	60	47.58	24.71	45	0-142
Gübre çukuru ( $x_7$ )	60	311.4	97.56	320	100-500
Sağım ünitesi ( $x_8$ )	60	0.68	0.46	1.00	0-1
Merkeze uzaklık ( $x_9$ )	60	69.85	38.10	56	20-160

Bu modele ait alan, havalandırma alanı, havalandırma sayısı, karantina alanı, karantina bölgesi, revir alanı, gübre çukuru, sağım ünitesi, merkeze uzaklık değişkenleri için katsayı tahminleri, standart hata değerleri, t-hesap değeri, önemlilik düzeyleri (P), belirleme katsayısı, düzeltilmiş belirleme katsayısı, hata kareler toplamı ve sapma değerleri Çizelge 4.2.' de verilmiştir.

Çizelge 4.2. Doğrusal regresyon modeline ait değişkenlerin tahmin sonuçları

Değişkenler	Katsayılar	$S_{\bar{x}}$	t	P
Sabit	-30858.28	22117.62	-1.40	0.890
Alan ( $x_1$ )	827.109	134.654	6.142	0.000**
Havalandırma alanı( $x_2$ )	-52.074	67.963	-0.766	0.447
Havalandırma sayısı ( $x_3$ )	23792.965	12433.654	1.914	0.062
Karantina bölgesi ( $x_4$ )	12812.820	11259.863	1.138	0.261
Karantina alanı ( $x_5$ )	-694084.866	510136.495	-1.361	0.180
Revir alanı ( $x_6$ )	3513.317	2719.605	1.292	0.202
Gübre çukuru ( $x_7$ )	516.844	626.670	0.825	0.414
Sağım ünitesi ( $x_8$ )	-235931.494	122664.109	-1.923	0.060
Merkeze uzaklık ( $x_9$ )	1171.255	1388.689	0.843	0.403
$R^2 = 0.820$ $\bar{R}^2 = 0.783$ F= 22.274 HKT=530700 S= 63944.903				

\*\* : Parametrelerin 0.01 önem düzeyinde istatistiksel olarak çok önemli.

Çizelge 4.2.' de sonuçları gösterilen doğrusal regresyon modeli eşitlik 43' te gösterilmiştir.

$$y = -30858.28 + x_1(827.109) + x_2(-52.074) + x_3(23792.965) + x_4(12812.820) + x_5(-694084.866) + x_6(3513.317) + x_7(516.844) + x_8(-235931.49) + x_9(1171.255) + \varepsilon \quad (43)$$

Çizelge 4.2.'de görüldüğü gibi havalandırma alanı, havalandırma sayısı, karantina alanı, karantina bölgesi, revir alanı, gübre çukur, sağım ünitesi ve merkeze uzaklık değişkenlerin tarımsal işletmelerin satış fiyatı üzerinde etkisi istatistiki olarak önemsiz bulunmuşken ( $P>0.05$ ), alan değişkeni üzerindeki etkisi çok önemli bulunmuştur ( $P<0.01$ ). Çoklu doğrusal regresyon uygulandığında, havalandırma alanı, karantina alanı ve sağım ünitesi değişkenlerinin satış fiyatı üzerinde negatif yönde, buna karşılık diğer değişkenlerinin ise pozitif yönde bir etkisinin bulunduğu görülmektedir. Ayrıca hata kareler toplamı ve sapma değeri oldukça yüksek elde edilmiştir. Böylelikle doğrusal parametrik modelin tarımsal işletmenin satış fiyatları üzerinde etkili olan değişken etkilerini belirlemek için yeterli olmadığı görülmüştür. Bu amaç doğrultusunda değişkenler için semiparametrik regresyon modeli oluşturulmuştur.

Çalışmada yer verilen kesikli değişkenler fonksiyondaki eğriliği etkilemediklerinden diğer bir ifadeyle, düzeltmeye ihtiyaç duyulmayan değişkenler olduklarından modele parametrik olarak dâhil edilmiştir. Diğer yandan, bağımlı değişkenle ilişkinin türü kesin olarak bilinmeyen diğer değişkenler ise modele parametrik olmayan kısım olarak dâhil edilmiştir. Uygun semiparametrik regresyon modelini belirleyebilmek için hem parametrik değişkenlerin hem de parametrik olmayan değişkenlerin yer aldığı ve modele ait tahminlerin nasıl yorumlandığını görmek amacıyla tarımsal işletmelerinin satış fiyatları ile tarımsal işletmenin özellikleri arasındaki ilişkilerin incelendiği semiparametrik regresyon modeli aşağıdaki eşitlikle tanımlanmıştır.

$$y = \beta_0 + x_4 \beta_1 + x_8 \beta_2 + s(x_1) + s(x_2) + s(x_3) + s(x_5) + s(x_6) + s(x_7) + s(x_9) + \varepsilon \quad (44)$$

Bu modele ait karantina bölgesi ve sağım ünitesi değişkenleri için parametre tahminleri, standart hata değerleri, Ki-kare hesap değeri ve önemlilik düzeyleri (P) Çizelge 4.3.'te verilmiştir.

Çizelge 4.3. Semiparametrik regresyon modeli parametrik değişkenlerin tahmin sonuçları

Değişkenler	Katsayılar	$S_{\bar{x}}$	$\chi^2$	P
Sabit	13.569	0.048	77645.992	<.001**
Sağım ünitesi ( $x_8$ )	-0.016	0.005	9.854	<.001**
Karantina bölgesi ( $x_4$ )	-0.812	0.113	51.02	<.001**

\*\* : Parametrelerin 0.01 önem düzeyinde istatistiksel olarak çok önemli.

Çizelge 4.3.'e göre uygulamadaki parametrik değişkenler incelendiğinde, elde edilen sonuçlara göre modelde yer alan parametrik değişkenlerin tümünün istatistiksel olarak çok önemli bulunduğu görülmektedir ( $P < 0.01$ ). Bu değişkenlerden sağım ünitesi ve karantina bölgesi fiyat değişkenini negatif yönde etkilemektedir. Parametrik değişkenler arasında fiyatları negatif yönde en fazla etkileyen değişken ise karantina bölgesi değişkenidir.

Bu modele ait alan, havalandırma alanı, havalandırma sayısı, karantina alanı, revir alanı, gübre çukuru ve merkeze uzaklık değişkenleri için etkili serbestlik derecesi, F hesap değeri ve önemlilik düzeyleri (P) Çizelge 4.4.'te verilmiştir.

Çizelge 4.4. Semiparametrik regresyon modeli parametrik olmayan değişkenlerin tahmin sonuçları

Bileşen	Etkili Serbestlik Derecesi	F	P
Alan ( $x_1$ )	6.8554.04	188.17	<.001**
Havalandırma alanı ( $x_2$ )	6.05	68.22	<.001**
Havalandırma sayısı ( $x_3$ )	5.01	110.35	<.001**
Karantina alanı ( $x_5$ )	7.07	1162.24	<.001**
Revir alanı ( $x_6$ )	7.48	144.04	<.001**
Gübre çukuru ( $x_7$ )	6.61	260.36	<.001**
Merkeze uzaklık ( $x_9$ )	1.00	25.96	<.0002**

\*\* : Parametrelerin 0.01 önem düzeyinde istatistiksel olarak çok önemli.

Çizelge 4.3. ve Çizelge 4.4'te sonuçları gösterilen parametrik ve parametrik olmayan regresyon modellerin toplamsal gösterimi eşitlik 45'te gösterilmiştir.

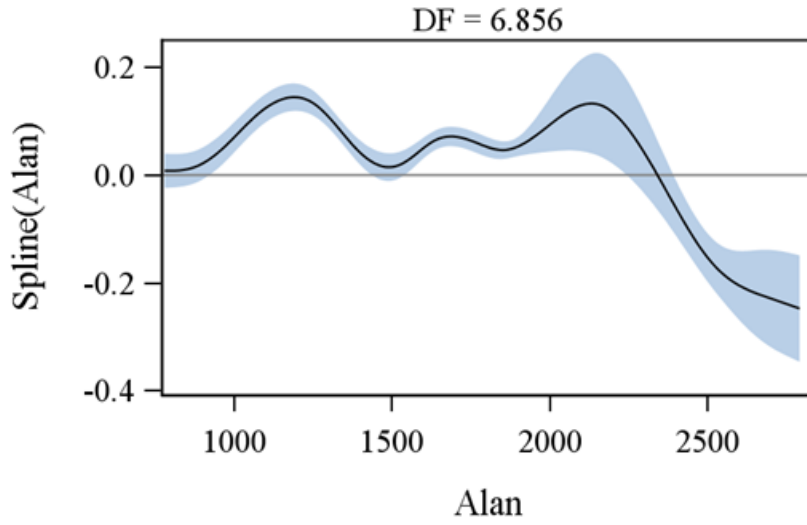
$$y = 13.569 + x_4(-0.812) + x_8(-0.016) + s(x_1) + s(x_2) + s(x_3) + s(x_5) + s(x_6) + s(x_7) + s(x_9) + \varepsilon \quad (45)$$

Eşitlik 45'te, parametrik ve parametrik olmayan regresyon olarak iki bölümden oluşmaktadır. Bu iki bölüm için katsayı yorumları ve çıkarımlar ayrı yöntemlerle incelenmektedir. Semiparametrik regresyon modelinin parametrik regresyon kısmı için yorumlar ve çıkarımlar doğrusal regresyon modelleri ile benzer şekildedir. Parametrik olmayan regresyon kısmı için yorumlar grafik yardımıyla incelenirken, çıkarımlar ise F testi yardımıyla incelenmektedir.

Çizelge 4.4.'te modelin parametrik olmayan kısmına ait tahmin sonuçları her bir parametrik olmayan değişkene ait kısmi regresyon fonksiyonlarının ne kadar düzleştirildiklerini gösteren etkili serbestlik dereceleri ve her bir parametrik olmayan değişkenin istatistiksel olarak önemli olup olmadığını gösteren F hesap değerleri ile verilmiştir. Elde edilen uygun semiparametrik regresyon modelinin parametrik olmayan kısmına ait 7 bileşen bulunmaktadır. Böylece Çizelge 4.4.'e göre modele dâhil edilen

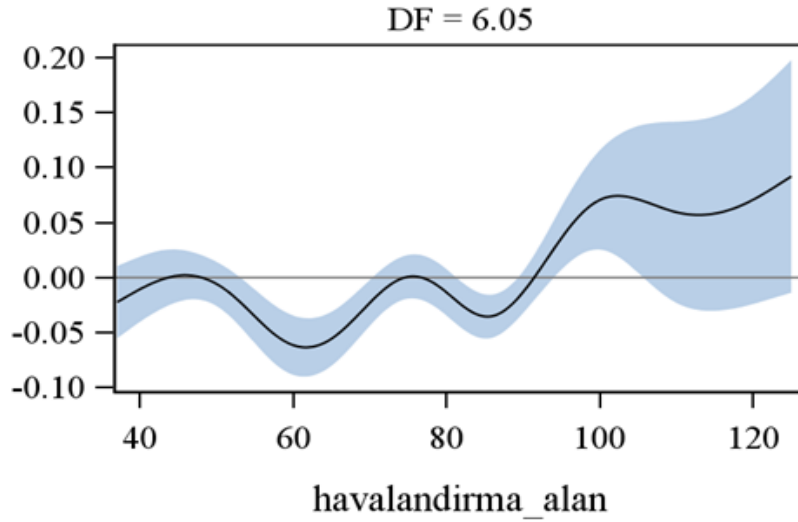
parametrik olmayan deęişkenlerinden alan, havalandırma alanı, havalandırma sayısı, karantina alanı, revir alanı, gübre çukuru ve merkeze uzaklık deęişkenlerin tümü istatistiki olarak çok önemli bulunmuştur ( $P < 0.01$ ).

Elde edilen parametrik olmayan kısım çok sayıda katsayı içerdiğinden başka bir deyişle vektör olarak elde edildiğinden parametrik olarak ifade etmek mümkün değildir ve böylece parametrik olmayan bileşenler ancak grafiklerle gösterilebilir (Turanlı ve Kalkan, 2012). Dolayısıyla fiyat ile modele parametrik olmayan formda dâhil edilen deęişkenler arasındaki ilişki grafiklerle verilmiştir.



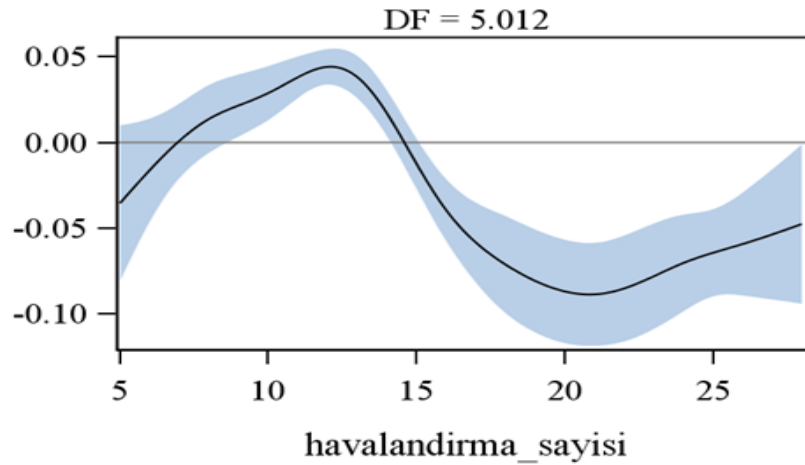
Şekil 4.1. Fiyat ile işletme alanı deęişkeni arasındaki ilişki için A ve S(A) grafięi

Şekil 4.1.'deki grafik incelendiğine fiyat ile modele parametrik olmayan formda dâhil edilen alan deęişkeni arasında doğrusal olmayan ilişkinin olduęu gözlemlenebilir. Semiparametrik modelin tahmini ile ilgili olarak, dikey ekseninde spline(Alan) deęerlerinin, yatay ekseninde ise parametrik olmayan alan deęişkenine ait deęerler elde edilmiştir. Yani, her bir parametrik olmayan deęişkenin deęerindeki deęişmeye karşılık katsayı tahminlerinin nasıl deęiştięi gösterilmiştir. Şekildeki gölgeli alanlar ise %95 güven aralığı bandında olduęunu ifade etmektedir. Şekil 4.1.'de görüldüğü üzere 2000 m<sup>2</sup>'ye kadar çok fazla işletme sayısı ve 2000 m<sup>2</sup>'den sonrasında az sayıda işletme olduęu güven bandının genişlemesinden anlaşılmaktadır. Yani en çok tercih edilen alan 1000-2000 m<sup>2</sup> aralığıdır.



Şekil 4.2. Fiyat ile havalandırma alanı değişkeni arasındaki ilişki için HA ve S(HA) grafiği

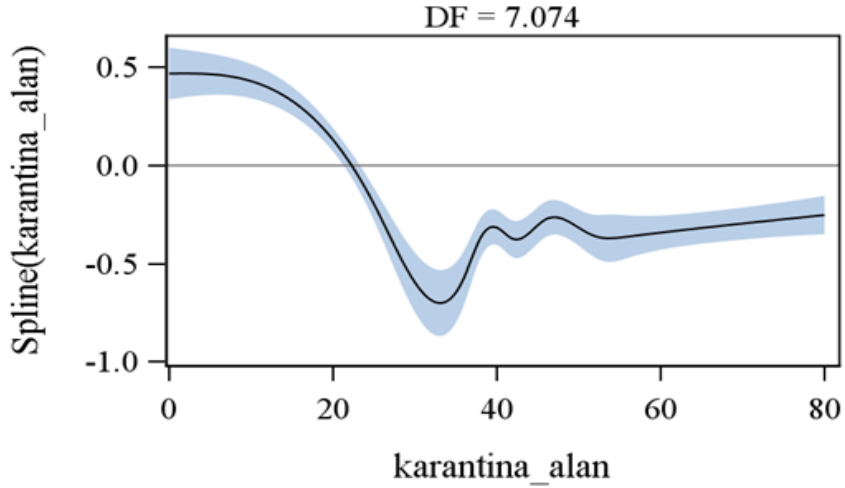
Şekil 4.2.'deki grafik incelendiğine fiyat ile modele parametrik olmayan formda dâhil edilen havalandırma alanı değişkeni arasında doğrusal olmayan ilişkinin olduğu gözlenebilir. Semiparametrik modelin tahmini ile ilgili olarak, dikey ekseninde spline (havalandırma alanı) değerlerinin, yatay ekseninde ise parametrik olmayan havalandırma alanı değişkenine ait değerler elde edilmiştir. Şekildeki gölgeli alanlar ise %95 güven aralığı bandında olduğunu ifade etmektedir. Şekil 4.2.'de görüldüğü üzere 100 m<sup>2</sup>'ye kadar çok fazla işletme sayısı ve 100 m<sup>2</sup>'den sonrası az sayıda işletme olduğu güven bandının genişlemesinden anlaşılmaktadır. Yani en çok tercih edilen alan 40-100 m<sup>2</sup> aralığıdır.



Şekil 4.3. Fiyat ile havalandırma sayısı değişkeni arasındaki ilişki için HS ve S(HS) grafiği

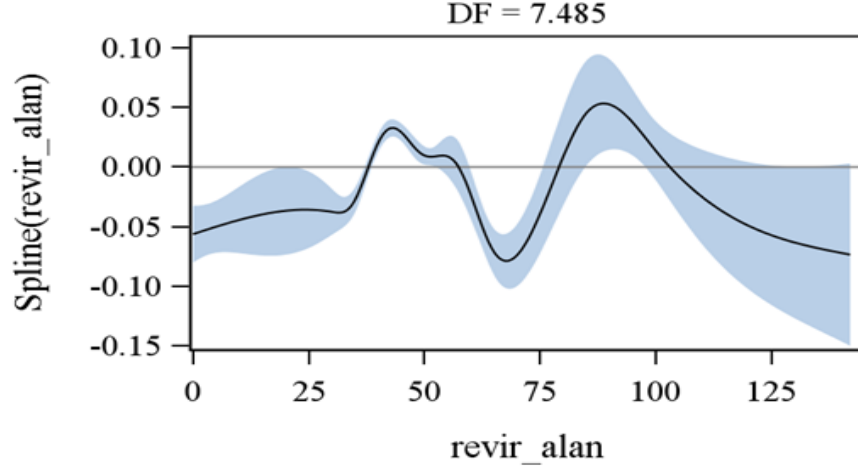
Şekil 4.3.'teki grafik incelendiğine fiyat ile modele parametrik olmayan formda dâhil edilen havalandırma sayısı değişkeni arasında doğrusal olmayan ilişkinin olduğu

gözlenebilir. Semiparametrik modelin tahmini ile ilgili olarak, dikey ekseninde spline(havalandırma sayısı) değerlerinin, yatay ekseninde ise parametrik olmayan havalandırma sayısı değişkenine ait değerler elde edilmiştir. Şekildeki gölgeli alanlar ise %95 güven aralığı bandında olduğunu ifade etmektedir. Şekil 4.3.'te görüldüğü üzere 10 ile 15 m<sup>2</sup>'ye kadar çok fazla işletme sayısı olduğu ve 10 ile 15 m<sup>2</sup> dışındaki alanlarda ise az sayıda işletme olduğu güven bandının genişlemesinden anlaşılmaktadır. Yani en çok tercih edilen alan 10-15 m<sup>2</sup> aralığıdır.



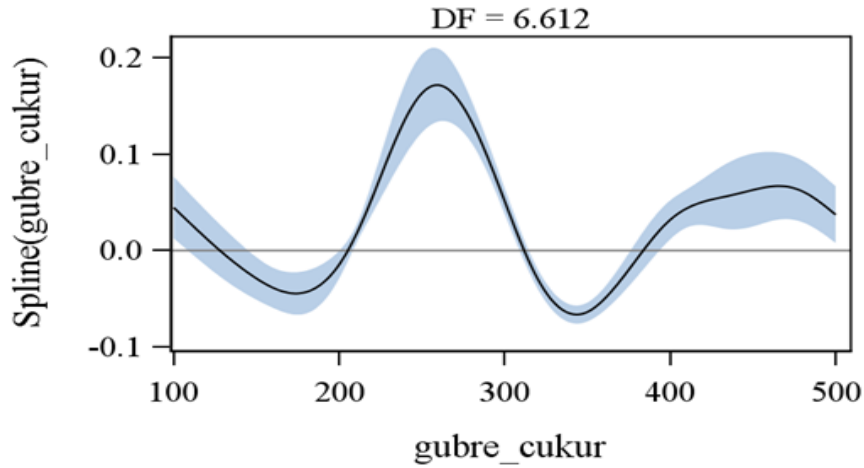
Şekil 4.4. Fiyat ile karantina alanı değişkeni arasındaki ilişki için KA ve S(KA) grafiği

Şekil 4.4.'teki grafik incelendiğine fiyat ile modele parametrik olmayan formda dâhil edilen karantina alanı değişkeni arasında doğrusal olmayan ilişkinin olduğu gözlenebilir. Semiparametrik modelin tahmini ile ilgili olarak, dikey ekseninde spline(karantina alanı) değerlerinin, yatay ekseninde ise parametrik olmayan karantina alanı değişkenine ait değerler elde edilmiştir. Şekildeki gölgeli alanlar ise %95 güven aralığı bandında olduğunu ifade etmektedir. Şekil 4.4.'te görüldüğü üzere baştan sona doğru tüm alanlarda çok fazla işletme olduğu güven bandının daralmasından anlaşılmaktadır.



Şekil 4.5. Fiyat ile revir alanı değişkeni arasındaki ilişki için RA ve S(RA) grafiği

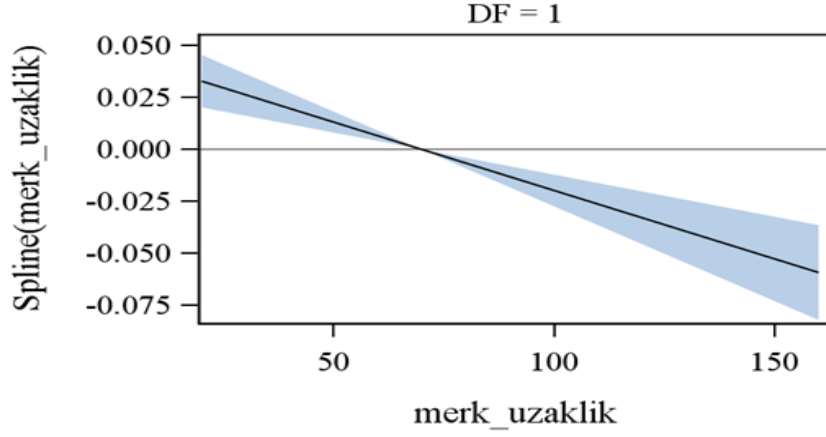
Şekil 4.5.'teki grafik incelendiğine fiyat ile modele parametrik olmayan formda dâhil edilen revir alanı değişkeni arasında doğrusal olmayan ilişkinin olduğu gözlenebilir. Semiparametrik modelin tahmini ile ilgili olarak, dikey ekseninde spline (revir alanı) değerlerinin, yatay ekseninde ise parametrik olmayan revir alanı değişkenine ait değerler elde edilmiştir. Şekildeki gölgeli alanlar ise %95 güven aralığı bandında olduğunu ifade etmektedir. Şekil 4.5.'te görüldüğü üzere 25 ile 75 m<sup>2</sup>'ye kadar çok fazla işletme sayısı olduğu güven bandının daralmasından ve 25 ile 75 m<sup>2</sup> dışındaki alanlarda ise az sayıda işletme olduğu güven bandının genişlemesinden anlaşılmaktadır. Yani en çok tercih edilen alan 25 ile 75 m<sup>2</sup> aralığıdır.



Şekil 4.6. Fiyat ile gübre çukuru değişkeni arasındaki ilişki için GÇ ve S(GÇ) grafiği

Şekil 4.6.'daki grafik incelendiğine fiyat ile modele parametrik olmayan formda dâhil edilen gübre çukuru değişkeni arasında doğrusal olmayan ilişkinin olduğu gözlenebilir. Semiparametrik modelin tahmini ile ilgili olarak, dikey ekseninde spline (gübre çukuru)

değerlerinin, yatay ekseninde ise parametrik olmayan gübre çukuru değişkenine ait değerler elde edilmiştir. Şekildeki gölgeli alanlar ise %95 güven aralığı bandında olduğunu ifade etmektedir. Şekil 4.6.'da görüldüğü üzere baştan sona doğru tüm alanlarda çok fazla işletme olduğu güven bandının daralmasından anlaşılmaktadır.



Şekil 4.7. Fiyat ile merkeze uzaklık değişkeni arasındaki ilişki için MU ve S(MU) grafiği

Şekil 4.7.'deki grafik incelendiğine fiyat ile modele parametrik olmayan formda dâhil edilen merkeze uzaklık değişkeni arasında doğrusal olmayan ilişkinin olduğu gözlenebilir. Semiparametrik modelin tahmini ile ilgili olarak, dikey ekseninde spline(merkeze uzaklık) değerlerinin, yatay ekseninde ise parametrik olmayan merkeze uzaklık değişkenine ait değerler elde edilmiştir. Şekildeki gölgeli alanlar ise %95 güven aralığı bandında olduğunu ifade etmektedir. Şekil 4.7.'de görüldüğü üzere 50 ile 100 km'ye kadar çok fazla işletme sayısı olduğu güven bandının daralmasından ve 50 ile 100 km dışındaki alanlarda ise az sayıda işletme olduğu güven bandının genişlemesinden anlaşılmaktadır. Yani en çok tercih edilen alan 50 ile 100 km aralığıdır.

Çizelge 4.5. Semiparametrik regresyon modelde uyum istatistikleri

Uyum İstatistikleri	
Cezalı maksimum olabilirlik	203.422
Pürüzlük ceza yaklaşımı	0.00265
Etkili serbestlik derecesi	45.08
Hatanın etkili serbestlik derecesi	12.68
Akaike bilgi kriteri	-316.66
Düzeltilmiş akaike bilgi kriteri	-17.876
Bayesian bilgi kriteri	-222.235
Genelleştirilmiş çapraz geçerlilik	0.00094

Çizelge 4.5.'te cezalı maksimum olabilirlik tahminin değeri 203.422 elde edilmiştir. Tahminin Akaike bilgi kriteri değeri -316.66 olup yoğunluk fonksiyonu tahminine ait Akaike bilgi kriteri değerini göstermektedir. Verilerin tahmin edilmesinde modelin verimliliğini ölçmek için kullanılan Bayesian bilgi kriteri değeri -222.235 olarak elde edilmiştir. Pürüzlük ceza yaklaşımı, eğrinin pürüzlüğünü ölçmek ve verilerin uyumu için eğrinin pürüzlülüğü arasında dengeyi sağlar. Burada modelin pürüzlük ceza yaklaşım değeri, 0.00265 olup etkili serbestlik derecesi ise 45.08 olarak elde edilmiştir. Ayrıca modelde genelleştirilmiş çapraz geçerlilik değeri 0.00094 olup bu değer 0 ile 4 aralığının da olduğu için semiparametrik regresyon modeli için uygundur.

Çizelge 4.6. Düzeltme bileşenleri için parametre tahminleri

Bileşen	Etkili Serbestlik Derecesi	Düzeltilme Parametresi	Pürüzlük Ceza Yaklaşımı	Parametrelerin Sayısı	Boğum Sayısı
Alan ( $x_1$ )	6.85	10883.6	0.000356	9	46
Havalandırma alanı ( $x_2$ )	6.05	9.87	0.000909	9	34
Havalandırma sayısı ( $x_3$ )	5.01	0.59	0.000464	9	17
Karantina alanı ( $x_5$ )	7.07	0.004	0.000078	9	14
Revir alanı ( $x_6$ )	7.48	3.06	0.000401	9	18
Gübre çukuru ( $x_7$ )	6.61	162.6	0.000443	9	23
Merkeze uzaklık ( $x_9$ )	1.00	$3.407.10^{-15}$	$1.24.10^{-16}$	9	26

Doğrusal regresyon modelinde modele dâhil edilen tüm parametrik değişkenler için tek bir serbestlik derecesi hesaplanır. Modelde artan karmaşıklık nedeniyle serbestlik derecesi tanımsız çıkmaktadır. Semiparametrik regresyon modelinde artan karmaşık nedeniyle etkili serbestlik derecelerini elde etmek için genelleştirilmiş çapraz geçerlilik kriterlerinden yararlanılmaktadır. Böylelikle her bir değişken için etkili serbestlik derecesinin belirlenmesi, optimal düzeltme parametresinin seçimine karşılık gelmektedir ve her bir değişken için ne kadar düzgünleştirileceği bilgisini vermektedir. Bu nedenle serbestlik derecelerini yaklaşık olarak tahmin etmek için Çizelge 4.6.'da görüldüğü gibi parametrik olmayan değişkenler için etkili serbestlik dereceleri hesaplanmıştır. En düşük etkili serbestlik derecesini merkeze uzaklık değişkeni 1.00 ve en yüksek etkili serbestlik derecesini ise revir alanı değişkeni 7.48 olarak belirlemiştir.

Boğum sayısının tahmin üzerindeki etkilerini azaltmak için düzeltme parametresi uygulanmaktadır. Düzeltme parametresinin değeri yükseldikçe fonksiyonun düzgünlüğü de artmaktadır. Yani alan değişkenin düzgünlüğü diğer değişkenlere göre daha çok artmıştır. Düzeltme parametresi tahmin edilen fonksiyondaki dalgalanmaları göstermektedir. Düzeltme parametresi aynı zamanda cezalandırma parametresi olarak adlandırılmaktadır. Çizelge 4.6. incelendiğinde boğum sayısı karantina alanı değişkeni için 14 düzeltme parametresi değeri 0.004 iken diğer değişkenler için bu değerler artmıştır. En fazla boğum sayısına sahip alan değişkenin boğum sayısı 46 ve optimal düzeltme parametresi değeri ise 10883.6 olarak hesaplanmıştır.

Semiparametrik regresyon modelinin incelendiği bu çalışmadan elde edilen sonuçlar, semiparametrik modellerin yeterliliği bağlamında Omay ve ark. (2007), Engle ve ark. (1986) Hurvich ve ark. (1988), Green ve Silverman (1994) ve Başer ve ark. (2019)'nın elde ettiği sonuçlarla ile uyum içerisindedir.

## 5. SONUÇ ve ÖNERİLER

Bu çalışmada Kahramanmaraş ilinde bulunan 60 tarımsal hayvancılık işletmesine ait, barınakların fiyatı üzerinde etkili olduğu düşünülen, işletmenin fiyatı, sağım ünitesi varlığı, karantina bölgesi varlığı, barınak alanı, havalandırma alanı, havalandırma sayısı, karantina alanı varlığı, revir alanı varlığı, gübre çukuru varlığı ve şehir merkezine uzaklık değişkenleri ele alınmıştır. Modele bazı değişkenler parametrik formda dâhil edilirken, bazı değişkenler ise parametrik olmayan formda dâhil edilmiştir.

Uygulamada, modele düzgünleştirerek dâhil edilmesi gereken değişkenler incelenmiş ve modelde hem parametrik hem de parametrik olmayan değişkenler olduğundan semiparametrik regresyon analizinin uygulanması uygun bulunmuştur. Semiparametrik regresyon modelinin en önemli özelliği, bağımlı değişken ile bağımsız değişkenler arasındaki ilişkilerinin şeklini istatistiksel testlerle inceleyebilmesidir. Başka bir deyişle, bir değişkenin düzgünleştirerek, doğrusal veya dönüştürme metotlarıyla doğrusallaştırarak modele dâhil edileceğine karar verir. Ayrıca modellerin karşılaştırılmasını yaparak hangi modelin uygun olduğunu gösterir. Semiparametrik regresyon yöntemi ile modelleme yapılmasının yanı sıra, bu yöntemi kullanarak değişkenlerin yapılarının belirlenmesi en iyi tahminlere ulaşılmasını da sağlamaktadır.

Çalışma da öncelikle tüm değişkenlerin doğrusal olduğunu varsayarak parametrik regresyon modeli elde edilmiştir. Yapılan analiz sonucunda aşağıdaki denklem elde edilmiştir.

$$y = -30858.28 + x_1(827.109) + x_2(-52.074) + x_3(23792.965) + x_4(12812.820) + x_5(-694084.866) + x_6(3513.317) + x_7(516.844) + x_8(-235931.49) + x_9(1171.255) + \varepsilon$$

Değişkenlerin doğrusal olduğu varsayımı altında elde edilen parametrik regresyon modelinde havalandırma alanı, havalandırma sayısı, karantina alanı, karantina bölgesi, revir alanı, gübre çukur, sağım ünitesi ve merkeze uzaklık değişkenlerinin tarımsal işletmelerin satış fiyatı üzerinde etkisi istatistiki olarak önemsiz bulunmuşken ( $P>0.05$ ), alan değişkeni üzerindeki etkisi çok önemli bulunmuştur ( $P<0.01$ ). Belirleme katsayısı değeri, 0.820 ve düzeltilmiş belirleme katsayısı değeri ise 0.783 elde edilmiştir. Ayrıca HKT ve sapma değerleri ise oldukça yüksek elde edilmiştir. İşletme fiyatının belirlenmesinde yatırım maliyetlerinin etkili olduğu bilinen bir gerçektir. Ancak parametrik regresyon modelinde havalandırma alanı, havalandırma sayısı, karantina alanı, karantina

bölgesi, revir alanı, gübre çukuru, sağım ünitesi ve merkeze uzaklık gibi yatırım maliyeti ve satış fiyatı üzerinde oldukça belirleyici olan değişkenler önemsiz bulunmuştur ( $P>0.05$ ). Bu durum değişkenlerin doğrusal olduğu varsayımı altında elde edilen parametrik regresyon modelinin oldukça yetersiz kaldığının açık göstergesidir.

Değişkenlerin bir kısmının parametrik bir kısmının ise parametrik olmayan kısım olarak belirlendiği semiparametrik regresyon modelinden elde edilen denklem ise,

$$y = 13.569 + x_4(-0.812) + x_8(-0.016) + s(x_1) + s(x_2) + s(x_3) + s(x_5) + s(x_6) + s(x_7) + s(x_9) + \varepsilon$$

şeklindedir. Bu analiz sonucunda parametrik değişkenler sağım ünitesi ve karantina bölgesi istatistiksel olarak çok önemli bulunduğu görülmüştür ( $P<0.01$ ). Modele dâhil edilen parametrik olmayan değişkenlerinden alan, havalandırma alanı, havalandırma sayısı, karantina alanı, revir alanı, gübre çukuru ve merkeze uzaklık değişkenleri de istatistiksel olarak çok önemli bulunmuştur ( $P<0.01$ ).

Diğer yandan parametrik olmayan kısım çok sayıda katsayı içerdiğinden başka bir deyişle vektör olarak elde edildiğinden parametrik olarak ifade etmek mümkün değildir ve parametrik olmayan bileşenlerin gösterimi grafiklerle incelenmiştir. İncelemeler sonucunda parametrik olmayan formda dahil edilen tüm değişkenlerin işletme satış fiyatı ile doğrusal olmayan bir ilişki içinde oldukları sonucuna varılmıştır ( F testleri ve Şeki 4.1, 4.2, 4.3, 4.4, 4.5, 4.6 ve 4.7 deki gölgeli alanlar parametrik olmayan değişkenlerin % 95 güven aralıkları).

Tarımsal işletmelerin maliyetleri ve dolayısıyla satış fiyatlarını belirleyen birçok yatırım ve çevresel faktörlerin oluşu doğru modellemenin ne kadar önemli olduğunu apaçık ortaya koymaktadır. Zira burada incelenen değişkenler modele parametrik olarak alındığında maliyet ve satış fiyatı üzerinde çok etkili olduğu bilinen birçok değişkenin önemsiz bir etkiye sahip olduğu gibi hatalı bir sonuç ortaya çıkmaktadır. Semiparametrik modelde ise değişkenlerin bir kısmının parametrik bir kısmının ise parametrik olmayan formda modele dahil edilmesinin ise pratikte de önemli olduğu bilinen değişkenlerin istatistiksel olarak çok önemli olduğu ortaya çıkmıştır.

Sonuç olarak tarımsal işletmelerin işletme fiyatının, buna bağlı olarakta satış fiyatının belirlenmesinde etkili olan değişkenlerin doğru tanımlanmaları (parametrik ve parametrik olmayan) ve uygun modelin seçilmesi konusunun son derece önemli olduğu ve bu alanda semi parametrik modellerinin rahatlıkla kullanılabileceği söylenebilir.

Bu alıřmada 60 adet tarımsal iřletmeye ait veri seti modellemeye alınmıřtır. Diđer yandan iřletme sayısının arttırılması, yani rnek byklğnn daha yksek tutulması durumunda semiparametrik regresyon yntemiyle deđiřkenlerin maliyet ve satıř fiyatları zerindeki etkilerinin daha net ortaya ıkacađını sylemek mmkndr.



## KAYNAKLAR

- Ai, C. (1997). A semiparametric maximum likelihood estimator. *Econometrica. Journal of the Econometric Society*, 933-963.
- Akaike, H. (1973). Information Theory and an Extension of the Maximum Likelihood Principle, in Petrov and Csaki, eds., *Proceedings of the Second International Symposium on Information Theory*, 267–281.
- Akdeniz, F., Akdeniz D. E. (2009). Liu-type Estimator in Semiparametric Regression Model. *Journal of Statistical Computation Simulation*.
- Akdeniz, F., Tabakan, G. (2009). Restricted Ridge Estimators of the Parameters in Semiparametric Regression Model, *Communications in Statistics-Theory and Methods*, 38: 1852-1869.
- Akkuş, Ö., Demir, S., Karasoy, D. (2008). İki düzeyli bağımlı değişken modelinin yarı parametrik tahmini. *İstatistikçiler Dergisi: İstatistik ve Aktüerya*, 1(3): 135-143.
- Akkuş, Ö., Demir, S., Tatlıdil, H. (2009). Nadaraya-Watson çekirdek kestiricilerinin yarı parametrik model tahminindeki performansı üzerine bir benzetim çalışması. *İstatistikçiler Dergisi: İstatistik ve Aktüerya*, 2(2): 28-36.
- Akkus, O. (2011). XploRe Package for the Popular Parametric and the Semiparametric Single Index Models. *Gazi University Journal of Science*, 24(4): 753-762.
- Altman N.S. (1990). Kernel Smoothing of Data With Correlated Errors. *Journal of the American Statistical Association*, 85(411): 749–759.
- Altman, N.S. (1992). An Introduction to Kernel and Nearest-Neighbour Nonparametric Regression, *The American Statistician*, 46(3): 175- 85.
- Alpar, R. (2003). *Uygulamalı Çok Değişkenli İstatistiksel Yöntemlere Giriş: I*, Nobel, Ankara.
- Aneiros-Pérez, G., Vieu, P. (2006). Semi-functional partial linear regression. *Statist. Probab. Lett.* 76 (11): 1102–1110.
- Aneiros-Pérez, G., Vieu, P. (2008). Nonparametric time series prediction: A semi-functional partial linear modeling. *J. Multivariate Anal.* 99 (5): 834–857.
- Ata, N., Karasoy, D., Sözer, M. T. (2008). Orantısız hazardlar için parametrik ve yarı parametrik yaşam modelleri. *İstatistikçiler Dergisi: İstatistik ve Aktüerya*, 1(3): 125-134.
- Aydın, D. (2005). Semiparametrik regresyon modellemede splayn düzeltme yaklaşımı ile tahmin ve çıkarsamalar. *Anadolu Üniversitesi, Fen Bilimleri Enstitüsü, İstatistik Anabilim Dalı, Doktora Tezi*.

- Aydın, D., Mammadov, M., Yüzer, A. F., Ağaoğlu, E. (2007). Yarı Parametrik Modellerde Splayn Düzeltme ile Tahmin ve Çıkarsamalar. *Anadolu University Journal of Science and Technology Cilt/Vol.: 8-Sayı/No: 1* : 41-52.
- Aydın, D., Alma, Ö. G. (2015). Diagnostics of influential observations in partially linear models. *Erzincan Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, 8(1): 51-68.
- Aytaç, M. (1991). Uygulamalı Parametrik Olmayan İstatistik Testleri, Uludağ Üniversitesi Basımevi, Bursa.
- Bağdatlı, S. (2010). Semiparametrik Regresyon ve Bir Uygulama. Yayınlanmamış Yüksek Lisans Tezi, İstanbul Ticaret Üniversitesi, Fen Bilimleri Enstitüsü, İstanbul.
- Begun, J., Hall W., Huang W., Wellner J. (1983). Information and Asymptotic Efficiency in Parametric-Nonparametric Models. *The Annals of Statistics*, 11: 432-452.
- Benedetti, J. (1977). On the Nonparametric Estimation of Regression Functions, *Journal of the Royal Statistical Society, Ser.B*, 39(2): 248-253.
- Berry, W. D. (1993). *Understanding Regression Assumptions*, Vol. 92.
- Bowman, A., Adelchi, A. (1997). *Applied Smoothing Techniques for Data analysis: The Kernel Approach with S-Plus Illustrations*, Oxford University Press.
- Box, G.E.P., Jenkins, G.M. (1970). *Time Series Analysis: Forecasting and Control*, Holden-Day, London.
- Brockmann, M., Theo, G., Eva, H. (1993). Locally Adaptive Bandwidth Choice for Kernel Regression Estimators, *Journal of the American Statistical Association*, 88(424): 1302–1309.
- Buckley M. J., Eagleson, G. K., Silwerman G. K. (1988). The Estimation of Residual Variance in Nonparametric Regression, *Biometrika*, 75(2): 189–199.
- Carota, C., Parmigiani, G. (2002). Semiparametric regression for count data. *Biometrika*, 89(2): 265-281.
- Chamberlain, G. (1992). Efficiency bounds for semiparametric regression. *Econometrica: Journal of the Econometric Society*, 567-596.
- Chandran, K., Prajneshu, P. (2004). Modelling Rice Productivity Additive Nonparametric Regression Approach, *National Academy Science Letters*, ISSN 0250-541X, 27: 261-267, National Academy of Sciences, India.
- Chen, H. (1988). Covergance Rates for Parametric Components in a Partially Linear Models, *Annals of Statistics*, 16: 136-146.
- Chen, H., Randall, A. (1997). Semi-nonparametric estimation of binary response models with an application to natural resource valuation. *Journal of Econometrics*. 76: 323–340.

- Clark, R. M. (1977). Nonparametric Estimation of a Smooth Regression Function, *Journal of the Royal Statistical Society, Ser. B*, 39: 107–113.
- Craven, P., Wahba, G. (1979). Smoothing Noisy Data With Spline Functions: Estimating the Correct Degree of Smoothing By the Method of the Generalized Cross-Validation, *Numer. Math.*, 31: 377-403.
- Cula, S. (1998). Çok Değişkenli Olasılık Yoğunluk Fonksiyonunun Çekirdek Fonksiyonlarıyla Kestirimi, Hacettepe Üniversitesi, Fen Bilimleri Enstitüsü, İstatistik Anabilim Dalı, Doktora Tezi, Ankara.
- Cuzick, J. (1992). Efficient estimates in semiparametric additive regression models with unknown error distribution. *The Annals of Statistics*, 20(2): 1129-1136.
- Çağlayan, E. (2002). Yarı parametrik regresyon modelleri ile yaşam boyu sürekli gelir hipotezinin Türkiye uygulaması. Marmara Üniversitesi, Sosyal Bilimler Enstitüsü, Ekonometri Anabilim Dalı, Doktora Tezi.
- Çağlayan, E. (2012). Nonparametrik regresyon modelleri. İstanbul, Derin Yayınları.
- Çağlayan, E., Dayioğlu, T. (2011). Comparing the parametric and semiparametric logit models: Household poverty in Turkey. *International Journal of Economics and Finance*, 3(5), 197-207.
- Demir, S. (2005). Regresyon Fonksiyonlarının Uyarlanabilir Nadaraya-Watson Çekirdek Kestirimleri, Hacettepe Üniversitesi, Fen Bilimleri Enstitüsü, İstatistik Anabilim Dalı, Doktora Tezi, Ankara.
- Dhekale, B. S., Sahu, P. K., Vishwajith, K. P., Mishra, P., Narsimhaiah, L. (2017). Application of parametric and nonparametric regression models for area, production and productivity trends of tea (*Camellia sinensis*) in India. *Indian Journal of Ecology*, 44(2): 192-200.
- Duran, E.A., Akdeniz, F., Hu, H. (2011). Efficiency of a Liu-type Estimator in Semiparametric Regression Models. *J. Computational Applied Mathematics* 235(5): 1418-1428.
- Engle, R. F., Granger, C. W., Rice, J., Weiss, A. (1986). Semiparametric estimates of the relation between weather and electricity sales. *Journal of the American statistical Association*, 81(394): 310-320.
- Eubank, L. R. (2011). *Spline Regression, Smoothing and Regression: Approaches, Computation and Application*, Edited By Micheal G. Schimek.
- Evren, A., Tuna, E. (2011). Yarı parametrik kısmı doğrusal panel veri modelleriyle uluslar arası göç. *Ekonometri ve İstatistik Dergisi*, (13): 95.
- Fan, J., Yao, Q. (2003). *Nonlinear time series: nonparametric and parametric methods*. Springer: New York.
- Fit, A. (2012). Parametrik Olmayan Regresyonda Düzeltme Parametresi Seçimi, Yüksek Lisans Tezi, Muğla Sıtkı Koçman Üniversitesi, Muğla, 93s.

- Fox, J. (2000). Multiple and Generalized Nonparametric Regression. Thousand Oaks: A Sage University Paper.
- Fung, W.K., Zhu, Z.Y., Wei, B.C., He, X. (2002). Influence Diagnostics and Outlier Tests for Semiparametric Mixed Models, *J. Roy. Statist. Soc. Ser. B*, 64: 565-579.
- Green, P.J., Jennison, C., Seheult, A. (1985). Analysis of Field Experiments by Least Square Smoothing, *J. Roy. Statist. Soc. B* 47: 299-315.
- Green, P.J., Silverman, B.W. (1994). Nonparametric Regression and Generalized Linear Models, Chapman&Hall, NewYork.
- Hardle, W., Muller, M., Sperlich, S. and Werwatz, A. 2004. Nonparametric and semiparametric models. Springer, 299, New York.
- Harezlak, J., David, R., Wand, P. (2008), Semiparametric Regression with R, ISBN 978-1-4939-8853-2, New York.
- Harezlak, J., Ruppert, D., Wand, M. P. 2018. Semiparametric regression with R. Springer New York.
- Hastie, T.J. Tibshirani, R.J. (1999). Generalized Additive Models, Chapman&Hall/CRC, NewYork.
- Hong, S.Y. (1998). Automatic Bandwidth Selection and Data-Driven Estimators in a Semiparametric Regression Model, A Dissertation Degree of Doctor of Philosophy, Northwestern University.
- Horowitz, J. L. (1993). Semiparametric Estimation of a Work-Trip Mode Choice Model. *Journal of Econometrics*, 58: 49-70.
- Hu, H. (2005). Ridge estimation of semiparametric regression model. *Journal of Computational and Applied Mathematics*, 176:215–222.
- Hurvich, C. M., Simonoff, J. S., Tasi, C. L. (1988). Smoothing Parameter Selection in Nonparametric Regression Using An Improved Akaike Information Criterion, *J.R. Statist. Soc. B.*, 60: 271-293.
- Hurvich, C. M., Tsai, C. L. (1989). Regression and Time Series Model Selection in Small Samples. *Biometrika*, 76(2): 297-307.
- Ichimura, H. (1993). Semiparametric least squares (SLS) and weighted SLS estimation of single-index models. *Journal of Econometrics*, 58: 71-120.
- Kaki, B. (2004). Yarı Parametrik Regresyon Yönteminin Hayvancılıkta Kullanımı. Yüzüncü Yıl Üniversitesi, Fen Bilimleri Enstitüsü, Zootekni Anabilim Dalı, Yüksek Lisans Tezi.
- Keele, L. J. (2008). Semiparametric regression for the social sciences. John Wiley & Sons.
- Kemp, G. C. R. (2000). Semiparametric Estimation of a Logit Model. University of Essex.

- Kim, C., Park, B.U., Kim, W., (2002). Influential Diagnostics in Semiparametric Regression Models, *Statistics & Probability Letters*, 60: 49-58.
- Klein, R. W., Spady, R. H. (1993). An efficient semiparametric estimator for binary response models. *Econometrica: Journal of the Econometric Society*, 387-421.
- Koop, G., Poirier, J. D. (2004). Bayesian Variants of Some Classical Semiparametric regression Techniques, *Journal of Econometrics*, 123: 259-282.
- Kou, S. C., Efron, B. (2002). Smoothers and the Cp, Generalized Maximum Likelihood, and Extended Exponential Criteria: A Geometric Approach, *JASA*, 97: 766-782.
- Lee, K. C. (1990). Avoiding misspecifications and improving efficiency in hedonic and consumption models: applications of semiparametric methods. Doctoral dissertation, London School of Economics and Political Science, United Kingdom.
- Liang, H., Hardle, W., Werwatz, A. (1997). Asymptotic properties of nonparametric regression estimation in partly linear models. Technical report no 55. Humboldt University of Berlin.
- Liang, H., Hardle, W. (1999). Large sample theory of the estimation of the error distribution for a semiparametric model. *Communications in Statistics-Theory and Methods*, 28(9): 2025-2036.
- Linton, O. B. (2009). Semiparametric and nonparametric ARCH modeling. *Handbook of financial time series*, 157-167.
- Liu, J., Zhang, R., Zhao, W. (2013). A robust and efficient estimation method for single index models. *J. Multivariate Anal.* 122 (122): 226–238.
- Liu, G., Du, X., Wang, M., M., Lin, J., Gao, Q. (2018). Semiparametric jump-preserving estimation for single-index models. *Journal of Nonparametric Statistics*, 30(3): 556-580.
- Mahmoud, H. F. F., Kim, I., Kim, H. (2016). Semiparametric single index multi change points model with an application of environmental health study on mortality and temperature. *Environmetrics*, 27(8): 494-506.
- Mahmoud, H. F. (2019). Parametric versus Semi and Nonparametric Regression Models. arXiv preprint arXiv: 1906.10221.
- Mahmoud, H. F. F., Kim, I., (2019). Semiparametric spatial mixed effects single index models. *Computational Statistics & Data Analysis*, 136: 108-112.
- Mammadov, M., Yüzer, A. F., Aydın, D. (2005). Splayn Düzeltme Regresyonu ve Düzeltme Parametresinin Seçimi, 4. İtatistik Kongresi bildiri ve poster özetleri kitabı, Belek- Antalya, 148-149.
- Manski, C. F. (1986). Semiparametric Analysis Of Binary Response From Response-Based Samples. *Journal of Econometrics*. 31(1): 31-40.

- Nash, M. S., Bradford, D. F. (2001). Parametric And Nonparametric Logistic Regressions For Predictions Of Presence/ Ansence of An Amphibian. US Enviromental Protection Agency Office of Research and Developent, EPA/600/R-01/081, 1-35.
- Newey, W. K. (1989). The Asymptotic Variance of Semiparametric Estimators. Princeton University, Econometric Research Program Memo. No. 346.
- Omay, E.R., Aydın, D., Mammadov, M. (2006). Splayn Düzeltme Yöntemi ile Semiparametrik Adittive Modellerin Kestirimi, 5. İstatistik Günleri Sempozyumu Bildiriler Kitabı, Konyaaltı-Antalya.
- Omay, R.E. (2007). Regresyonda Pürüzlülük Ceza Yaklaşımı, Anadolu Üniversitesi, Fen Bilimleri Enstitüsü, İstatistik Anabilim Dalı, Doktora Tezi, Eskişehir.
- Omay, R. E., Aydın, D., Mammadov, M. (2007). Semiparametrik toplamsal regresyon modeli ile tahmin: Eskişehir'deki evlerin kira fiyatları ve özellikleri arasındaki ilişkilerin analizi. Anadolu Üniversitesi Bilim ve Teknoloji Dergisi, 8(1), 153-159.
- Omay, R. E., Memmedli, M. (2008). Hava Kirliliğinin Sağlık Üzerindeki Etkisine Parametrik Olmayan Regresyon Yaklaşımı. Hava Kirliliği ve Kontrolü Ulusal Sempozyumu.
- Omay, R. E., Canpolat, E. (2013). Parametrik Olmayan Sabit Etkiler Panel Veri Modelleri: Türkiye için Hava Kirliliği ve Gelir İlişkisi. Anadolu University of Sciences and Technology-A: Applied Sciences and Engineering, 14(1).
- Peterson, A. V. (1977). Expressing the Kaplan-Meier Estimator as a Function of Emprical Subsurvival Functions, Journal of American Statistical Association, 72: 854-858.
- Rajarathinan, A., Parmar, R. S. (2011). Application pf parametric and nonparametric regression models for area, production and productivity trends of castor corn. Asian Journal of Applied Sciences, 4(1): 42-52.
- Raz, J. (1990). Testing for no Effect When estimating a Smooth Function By Nonparametric Regression: A Randomization Approach, J. Amer. Statist. Assoc., 85: 132-138.
- Ruppert, D., Wand, M. P., Carroll, R. J. (2003). Semiparametric Regression, Cambridge University Press, s.57
- Robinson, M. P. (1988). Root-N-Consistent Semiparametric Regression, Econometrica, 56(4): 931-954.
- Powell, J.L., Stock, J.H., Stoker, T.M. (1989). Semiparametric estimation of index coefficients. Econometrica 1403–1430.
- Severini, T. A., Staniswalis, J. G. (1994). Quasilikelihood estimation in semiparametric models. J. Amer. Statist. Assoc., 89: 501-511.
- Schennach, S. M. (2004). Nonparametric regression in the presence of measurement error. Econometric Theory, 20: 1046-1093.

- Schick, A. (1996). Efficient estimation in a semiparametric additive regression model with autoregressive errors. *Stochastic processes and their applications*, 61(2): 339-361.
- Schimek, G., Michael. (2000). Estimation and Inference in Partially Linear Models with Smoothing Splines, *Journal of Statistical Planning and Inference*, 91: 525-540.
- Sheather, S. J., Jones, M. C. (1991). A Reliable Data-Based Bandwidth Selection Method for Kernel Density Estimation. *Journal of the Royal Statistical Society: Series B (Methodological)*, 53(3): 683-690.
- Shen, C. W., Chen, Y. H. (2018). Model selection for semiparametric marginal mean regression accounting for within-cluster subsampling variability and informative cluster size. *Biometrics*, 74(3): 934-943.
- Shi, X. (2009). Applications of Nonparametric and Semiparametric Methods in Economics and Finance, Dissertation Degree of Doctor of Philosophy, Economics in the Graduate School of Binghamton University, New York.
- Slama, R. (2001). Using Nonparametric and Semiparametric Regression in Epidemiology, England et al., *Am. J. Epidemiol*, 154.
- Speckman, P. (1988). Kernel Smoothing in Partially Linear Model, *J. Royal Statist., Soc. B.*, 50: 413-436.
- Stefanski, L. A. (1985). The effects of measurement error on parameter estimation. *Biometrika* 72: 583-592.
- Sinan, A. (2010). Panel verili yarı parametrik regresyon modelleri. Selçuk Üniversitesi, Fen Bilimleri Enstitüsü, Matematik Anabilim Dalı, Doktora Tezi.
- Tabakan, G. (2009). Yarı Parametrik Regresyonda Tahmin Metodları. Çukurova Üniversitesi, Fen Bilimleri Enstitüsü, İstatistik Anabilim Dalı, Doktora Tezi.
- Tezcan, N. (2011). Parametrik Olmayan Regresyon Analizi. Atatürk Üniversitesi İktisadi ve İdari Bilimler Dergisi, 25.
- Thomas, W. (1991). Influence Diagnostics for The Cross-validated Smoothing Parameter in Spline Smoothing. *J. Amer. Statist. Assoc.* 86: 693-698.
- Toprak, S. (2015). Ölçüm hatalı değişkenli yarı parametrik regresyon modelleri. Dicle Üniversitesi, Fen Bilimleri Enstitüsü, Matematik Anabilim Dalı, Doktora Tezi.
- Turanlı, M. Bağdatlı, S., (2011). Semiparametrik regresyon. Öneri.C.9.S.35. Ocak 2011.207-213.
- Turanlı, M., Bağdatlı Kalkan, S. (2012). Site içerisindeki daire fiyatlarını etkileyen unsurların semiparametrik regresyon analizi ile belirlenmesi. İstanbul Ticaret Üniversitesi Sosyal Bilimler Dergisi, 11(21): 383-402.
- Türkan, S. (2012). Yarı Parametrik Regresyon Modelinde Etkili Gözlem Analizi. Hacettepe Üniversitesi, Fen Bilimleri Enstitüsü, İstatistik Anabilim Dalı, Doktora Tezi.

- Van Heerde, H. J., Leeflang, P. S., Wittink, D. R. (2001). Semiparametric analysis to estimate the deal effect curve. *Journal of Marketing Research*, 38(2): 197-215.
- Yatchew, A. (2003). *Semiparametric Regression for the Applied Econometrician*, Cambridge University Press.
- Wahba, G. (1990). *Spline Models of Observational Data*, University of Wisconsin at Madison, Pennsylvania.
- Wand, M. P., Gutierrez, R. G. (1997). Exact risk approaches to smoothing parameter selection. *J. Nonparametric Statist.*, 8: 337-354.
- Wang, Y. (2011). *Smoothing splines: methods and applications*. FL: CRC Press, Boca Raton.
- Wood, S. N. (2000). Modeling and Smoothing Parameter Estimation with Multiple Quadratic Penalties. *J. R. Statist. Soc. B* 62, Part 2: 413-428.
- Qingguo, T. Kong, L., (2017). Quantile regression in functional linear semiparametric model. *Statistics*, 51(6): 1342-1358.
- Zeger, S. L. Diggle, P. J., (1994). Semi-parametric models for longitudinal data with application to CD4 cell numbers in HIV seroconverters. *Biometrics*, 50: 689-699.
- Zhongyi, Z. Baocheng, W., (2001). Diagnostic and Influence Analysis for Semiparametric Nonlinear Regression Models. *Acta. Math. Appl. Sinica* vol. 24 (4): 568-581.
- Zhu, L., Cui, H. (2003). A semi-parametric regression model with errors in variables. *Scandinavian Journal of Statistics*, 30(2): 429-442.

## EKLER

### Ek 7.1. SAS paket programına ait kodlar

```
Program Editor - (Untitled)
data esra;
input
isletme Fiyat Alan havalandirma_alan havalandirma_sayisi karantina_bolgesi
karantina_alan revir_alan gubre_cukur sagim_unitesi Dgum_unitesi merk_uzaklik;
fiyat1=input(fiyat, best16.);
cards;
1 2679145 2160 96 10 1 36 142 291 1 1 155
2 2774055 2604 100.8 28 1 25 100 200 1 1 145
3 1155453 780 86.35 12 1 80 0 217 0 0 145
...
run;
data esra; set esra;
lfiyat=log(fiyat1);
run;

ods graphics on/width=800
Title "Detect the level of spline by algorithm";
proc gampl data=esra plots=all seed=9354 ;
class Dgum_unitesi sagim_unitesi karantina_bolgesi;
  model lFiyat = param(Fiyat) param(sagim_unitesi) param(karantina_bolgesi)
  spline(Alan) spline(havalandirma_alan) spline(havalandirma_sayisi)
  spline(karantina_alan) spline(revir_alan) spline(gubre_cukur) spline(merk_uzaklik)
/dist= normal
normalize;
*weight weight;
performance details;
  output out=maliyet xbeta pred ;
  id isletme;
run;
```

Ek 7.2. İstatistiksel terimler ve Türkçe karşılıkları

<b>Fit Statistics</b>	<b>Uyum İstatistikleri</b>
Penalized Log Likelihood	Cezalı Maksimum Olabilirlik
Roughness Penalty	Pürüzlük Ceza Yaklaşımı
Effective Degrees of Freedom	Etkili Serbestlik Derecesi
Effective Degrees of Freedom for Error	Hatanın Etkili Serbestlik Derecesi
Akaike Information Criteria	Akaike Bilgi Kriteri
Adjusted Akaike Information Criteria	Düzeltilmiş Akaike Bilgi Kriteri
Bayesian Information Criteria	Bayesian Bilgi Kriteri
Generated Cross- Validation	Genelleştirilmiş Çapraz Geçerlilik