

**DEĞİŞEN KATSAYILI REGRESYON MODELİ İLE
GEBELERİN ORTALAMA ARTERİYEL KAN BASINCINA
ETKİ EDEN RİSK FAKTÖRLERİNİN BELİRLENMESİ**

**DETERMINATION OF THE RISK FACTORS AFFECTING
THE MEAN ARTERIAL BLOOD PRESSURE OF
PREGNANTS VIA VARYING COEFFICIENT REGRESSION
MODEL**

IŞIL DURMUŞ

DOÇ. DR. SEMRA TÜRKAN

Tez Danışmanı

Hacettepe Üniversitesi
Lisansüstü Eğitim-Öğretim ve Sınav Yönetmeliğinin
İstatistik Anabilim Dalı için Öngördüğü
YÜKSEK LİSANS TEZİ olarak hazırlanmıştır.

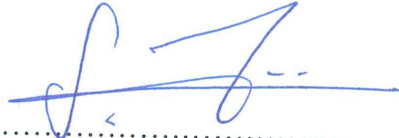
2018

İŞİL DURMUŞ' un hazırladığı “**Değişen Katsayılı Regresyon Modeli ile Gebelerin Ortalama Arteriyel Kan Basıncına Etki Eden Risk Faktörlerinin Belirlenmesi**” adlı bu çalışma aşağıdaki jüri tarafından **İSTATİSTİK ANABİLİM DALI**' nda **YÜKSEK LİSANS TEZİ** olarak kabul edilmiştir.

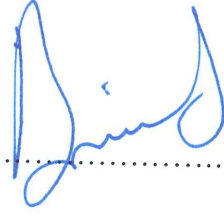
Prof. Dr. Hüseyin TATLIDİL
Başkan



Doç. Dr. Semra TÜRKAN
Danışman



Doç. Dr. Bülent ALTUNKAYNAK
Üye



Doç. Dr. Duygu İÇEN
Üye



Dr. Öğr. Üyesi Esra POLAT
Üye



Bu tez Hacettepe Üniversitesi Fen Bilimleri Enstitüsü tarafından **YÜKSEK LİSANS TEZİ** olarak onaylanmıştır.

Prof. Dr. Menemşe GÜMÜŞDERELİOĞLU
Fen Bilimleri Enstitüsü Müdürü

YAYINLAMA VE FİKRİ MÜLKİYET HAKLARI BEYANI

Enstitü tarafından onaylanan lisansüstü tezimin / raporunun tamamını veya herhangi bir kısmını, basılı (kağıt) ve elektronik formatta arşivleme ve aşağıda verilen koşullarla kullanıma açma iznini Hacettepe Üniversitesine verdiğimi bildiririm. Bu izinle Üniversiteye verilen kullanım hakları dışındaki tüm fikri mülkiyet haklarım bende kalacak, tezimin tamamının ya da bir bölümünün gelecekteki çalışmalarda (makale, kitap, lisans ve patent vb.) kullanım hakları bana ait olacaktır.

Tezin kendi orijinal çalışmam olduğunu, başkalarının haklarını ihlal etmediğimi ve tezimin tek yetkili sahibi olduğumu beyan ve taahhüt ederim. Tezimde yer alan telif hakkı bulunan ve sahiplerinden yazılı izin alınarak kullanılması zorunlu metinlerin yazılı izin alınarak kullandığımı ve istenildiğinde suretlerini Üniversiteye teslim etmeyi taahhüt ederim.

Yükseköğretim Kurulu tarafından yayınlanan “ Lisansüstü Tezlerin Elektronik Ortamda Toplanması, Düzenlenmesi ve Erişime Açılmasına İlişkin Yönerge” kapsamında tezim aşağıda belirtilen koşullar haricinde YÖK Ulusal Tez Merkezi / H. Ü. Kütüphaneleri Açık Erişim Sisteminde erişime açılır.

- o Enstitü / Fakülte yönetim kurulu kararı ile tezimin erişime açılması mezuniyet tarihimden itibaren 2 yıl ertelenmiştir. ⁽¹⁾
- o Enstitü / Fakülte yönetim kurulunun gerekçeli kararı ile tezimin erişime açılması mezuniyet tarihimden itibaren Ay ertelenmiştir. ⁽²⁾
- o Tezimle ilgili gizlilik kararı verilmiştir. ⁽³⁾

28 / 08 / 2018

(İmza)

Öğrencinin Adı SOYADI

ŞİL DURMUŞ

“Lisansüstü Tezlerin Elektronik Ortamda Toplanması, Düzenlenmesi ve Erişime Açılmasına İlişkin Yönerge”

- (1) Madde 6. 1. Lisansüstü teze ilgili patent başvurusu yapılması veya patent alma sürecinin devam etmesi durumunda, tez danışmanının önerisi ve enstitü anabilim dalının uygun görüşü üzerine enstitü veya fakülte yönetim kurulu iki yıl süre ile tezin erişime açılmasının ertelenmesine karar verebilir
- (2) Madde 6. 2. Yeni teknik, materyal ve metotların kullanıldığı, henüz makaleye dönüşmemiş veya patent gibi yöntemlerle korunmamış ve internetten paylaşılması durumunda 3. Şahıslara veya kurumlara haksız kazanç imkanı oluşturabilecek bilgi ve bulguları içeren tezler hakkında tez danışmanının önerisi ve enstitü anabilim dalının uygun görüşü üzerine enstitü ve fakülte yönetim kurulunun gerekçeli kararı ile altı ayı aşmamak üzere tezin erişime açılması engellenebilir.
- (3) Madde 7. 1. Ulusal çıkarları veya güvenliği ilgilendiren, emniyet, istihbarat, savunma ve güvenlik, sağlık vb. konulara ilişkin lisansüstü tezlerle ilgili gizlilik kararı, tezin yapıldığı kurum tarafından verilir*. Kurum ve kuruluşlarla yapılan işbirliği protokolü çerçevesinde hazırlanan lisansüstü tezlere ilişkin gizlilik kararı ise, ilgili kurum ve kuruluşun önerisi ile enstitü veya fakültenin uygun görüşü üzerine üniversite yönetim kurulu tarafından verilir. Gizlilik kararı verilen tezler Yükseköğretim Kuruluna bildirilir.
Madde 7. 2. Gizlilik kararı verilen tezler gizlilik süresince enstitü veya fakülte tarafından gizlilik kuralları çerçevesinde muhafaza edilir, gizlilik kararının kaldırılması halinde Tez Otomasyon Sistemine yüklenir.

* Tez danışmanının önerisi ve enstitü anabilim dalının uygun görüşü üzerine enstitü veya fakülte yönetim kurulu tarafından karar verilir.

Sevgili Ailem ve de Defne'mize...

ETİK

Hacettepe Üniversitesi Fen Bilimleri Enstitüsü, tez yazım kurallarına uygun olarak hazırladığım bu tez çalışmada;

- tez içindeki bütün bilgi ve belgeleri akademik kurallar çerçevesinde elde ettiğimi,
- görsel, işitsel ve yazılı tüm bilgi ve sonuçları bilimsel ahlak kurallarına uygun olarak sunduğumu,
- başkalarının eserlerinden yararlanması durumunda ilgili eserlere bilimsel normlara uygun olarak atıfta bulunduğumu,
- atıfta bulunduğum eserlerin tümünü kaynak olarak gösterdiğimi,
- kullanılan verilerde herhangi bir tahrifat yapmadığımı,
- ve bu tezin herhangi bir bölümünü bu üniversite veya başka bir üniversitede başka bir tez çalışması olarak sunmadığımı

beyan ederim.

28.08.2018



IŞIL DURMUŞ

ÖZET

DEĞİŞEN KATSAYILI REGRESYON MODELİ İLE GEBELERİN ORTALAMA ARTERİYEL KAN BASINCINA ETKİ EDEN RİSK FAKTÖRLERİNİN BELİRLENMESİ

Işıl DURMUŞ

Yüksek Lisans, İstatistik Bölümü

Tez Danışmanı: Doç. Dr. Semra TÜRKAN

Ağustos 2018, 62 Sayfa

Değişen katsayılı modeller (DKM) parametrik olmayan regresyon modelleri arasında çok boyutlu ve rilerin özelliklerini ortaya çıkarmak için kullanılır. Bu modeller, “ indeks değişken” olarak adlandırılan değişken üzerinden katsayıların değişimine izin verdiğinden boyutluluk lanetinden (*curse of dimensionality*) kaçınılmış olur. Çok boyutluluk, parametrik olmayan modeller ile etkili bir tahmin yapılmasını engellemektedir. Değişen katsayı modellerinde, katsayıların indeks değişkeninin her bir tabakasında değişerek düzleştirme yapılır. Buradan indeks değişkeni ve bağımsız değişkenler arasında doğrusal olmayan ilişkilerin olduğu anlaşılmaktadır. DKM, toplamsal modellere alternatif modellerdir. Zaman, sıcaklık, coğrafi konumlar vb. gibi bazı değişkenler üzerinden model katsayılarının değişimine izin verdiği için geniş uygulama alanına sahiptir.

Bu tez çalışmasında, ilk olarak parametrik olmayan regresyon modelleri ve tahmin yöntemleri hakkında bilgi verilmiş, daha sonra değişen katsayılı regresyon modeli tanıtılmış

ve tahmin yöntemleri incelenmiştir. Değişen katsayılı regresyon modeli, 2018 yılında Nisan-Temmuz aylarında muayeneye g elen 450 ge beye ilişkin elde ed ilen v erilere uygulanmıştır. Bu veriler, hasta takip formları aracılığıyla elde edilmiştir. Gebelerin ortalama arteriyel kan basıncı (MAP) değişkeni bağımlı değişken; gestasyonel yaş (gebelik haftası) indeks değişken; daha önceki gebelik öyküsü, aile hipertansiyon öyküsü, vücut kitle indeksi (VKI), yaş, diyabet hastalığı öyküsü, kötü obstetrik öyküsü, hematokrit (HCT) ve kandaki trombosit sayısı (BPC) ise bağımsız değişkenler olarak alınmıştır. MAP ile diğer bağımsız değişkenler arasındaki ilişki, gebelik haftası indeks değişkeni üzerinden değişen katsayılı model ile analiz edilmiştir.

Anahtar k elimeler: Parametrik olmayan regresyon modelleri, Değişen katsayılı model, Cezalandırılmış eğrisel çizgi, Düzleştirilmiş eğrisel çizgi, Yerel polinom tahmin

ABSTRACT

DETERMINATION OF THE RISK FACTORS AFFECTING THE MEAN ARTERIAL BLOOD PRESSURE OF PREGNANTS VIA VARYING COEFFICIENT REGRESSION MODEL

Işıl DURMUŞ

Master, Department of Statistical

Supervisor: Assoc. Prof. Dr. Semra TÜRKAN

August 2018, 62 Pages

Varying Coefficient Models are used to reveal properties of multidimensional data among nonparametric regression models. Since these models allow the coefficients to change over the variable called the "index variable", the curse of dimensionality is avoided. Multidimensionality prevents nonparametric models from making an effective prediction. In varying coefficient model, it can be considered that the coefficients are changed in each layer of the index variable and smoothing is done by this way. From this, it is understood that there are nonlinear relationships between index variable(s) and independent variables. Varying Coefficient Model is an alternative to additive models. As it permits the vary of model coefficients over some of the variables, such as time, temperature, geographical locations, etc. it has a large application area.

In this thesis study, firstly, the nonparametric regression models and estimation methods were given, then the variable coefficient regression model was introduced and the estimation

methods examined. The varying coefficient regression model was applied to the data related to 450 pregnant women coming for the medical examination in 2018 during the April-July. The data were obtained through patient follow-up forms.

Mean arterial blood pressure (MAP) variable was considered as dependent variable and gestational age considered as index variable, the number of previous pregnancies, family history of hypertension, body mass index (BMI), age, diabetes mellitus, bad obstetric history, hematocrit (HCT) and blood platelet count (BPC) were taken as independent variables. The relationship between MAP and other independent variables was analyzed via varying coefficient model as gestational age taking index variable.

Key Words: Nonparametric regression models, Varying Coefficient Model, Penalized splines, Smoothing splines, Local polynomial estimation

TEŐEKKÜR

Tez alıőmamın her aőamasında deęerli katkı ve eleőtirileriyle yol gősteren, sonsuz sabırla beni her zaman alıőmaya teővik eden ve gőven veren danıőmanım ve hocam

Sayın Do. Dr. Semra TÜRKAN' a,

Őnemli yorum ve deęerlendirmeleri ile katkıda bulunan jőri űyelerim Sayın Prof. Dr. Hőseyin TATLIDİL' e, Sayın Do. Dr. Bőlent ALTUNKAYNAK' a, Sayın Do. Dr. Duygu İEN' e ve Sayın Dr. Őęr. Ŭyesi Esra POLAT' a itenlikle teőekkőr ederim.

Bununla birlikte hayatım boyunca her zaman yanımda olan, bu sőrete bana her zamankinden daha ok anlayıő gősteren annem Aynur DURMUŐ' a, babam Sőleyman DURMUŐ' a ve teyzecięim Ayőe DOęAN' a bana olan inan ve destekleri iin sonsuz teőekkőrler.

Ayrıca canım abim Koray DURMUŐ ve ablam Dr. Arzu BOSTANCI DURMUŐ' a da sevgilerimi ve her zaman bana destek oldukları iin teőekkőrlerimi iletmekle birlikte; tıp camiası adına incelenmesi faydalı olacak bu tezde bőyők ölçőde katkısının olmasından dolayı ablama ayrıca takdir ve saygılarımı da iletmek isterim.

İÇİNDEKİLER

Sayfa

ÖZET	i
ABSTRACT	iii
ÇİZELGELER	vii
ŞEKİLLER	viii
SİMGELER VE KISALTMALAR	ix
1. GİRİŞ	1
2. REGRESYON MODELLERİ	5
2.1 Doğrusal Regresyon	5
2.2 Doğrusal Olmayan Regresyon	5
2.2.1 Yerel Modelleme	6
2.2.2 Bant Geniřliđi Seçimi	8
2.3 Tahmin Yöntemleri	9
2.3.1 Yerel (Local) Polinom Düzleřtirici	10
2.3.2 Çekirdek (Kernel) Düzleřtiricisi ve Yerel Doğrusal Düzleřtirici	13
2.3.3 Yerel Olarak Ađırlıklandırılmıř Saçılım Grafiđi Düzleřtiricileri	14
2.3.4 Eğrisel Çizgi Düzleřtiricisi (Spline Smoothing)	17
2.3.5 Cezalandırılmıř Eğrisel Çizgi (Penalized Spline)	22
3. DEĐİŐEN KATSAYILI MODEL	27
3.1 Literatür	29
3.2 Deđiřen Katsayılı Modellerde Tahmin Yöntemleri	30
3.2.1 Yerel (Local) Polinom Düzleřtiricisi	31
3.2.2 Düzleřtirilmif Eğrisel Çizgiler	36
3.2.3 Cezalandırılmıř Eğrisel Çizgi	37
4. UYGULAMA	40
4.1 Veriler	40
4.2 Analiz	44
5. SONUÇ	52
EKLER	54
KAYNAKLAR	58
ÖZGEÇMİŐ	62

ÇİZELGELER

Sayfa

Çizelge 2.1. Çekirdek Fonksiyonları	12
Çizelge 2.2. Noktasal asimptotik yanlılık ve varyans	14
Çizelge 3.1. Literatür taraması özeti.....	30
Çizelge 4.1. Bağımsız Değişkenlere İlişkin Sıklık Tablosu	44
Çizelge 4.2. Cezalandırılmış eğrisel çizgi ile tahmin edilen değişen katsayılı model sonuçları	45
Çizelge 4.3. Düzleştirilmiş eğrisel çizgi ile tahmin edilen değişen katsayılı model sonuçları	46

ŞEKİLLER

Şekil 1.1. Parametrik olmayan yöntemler için bir yol haritası	2
Şekil 2.1. Yerel modelleme yaklaşımı ile tahmin.....	7
Şekil 2.2. Yerel doğrusal kestirim.	8
Şekil 2.3. B-eğrisel çizgi ve örtüşen birkaç B-eğrisel çizginin gösterimi	22
Şekil 4.1. Düzleştirilmiş eğrisel çizgi ile tahmin edilen fonksiyonlarının gebelik haftasına göre grafikleri	49

SİMGELER VE KISALTMALAR

Simgeler

y	Bağımlı değişken vektörü
X	Bağımsız değişkenler matrisi
β	Bilinmeyen parametreler vektörü
σ^2	Varyans
$\hat{\sigma}^2$	σ^2 'nin kestirimi
h	Bant genişliği
$m(x)$	Parametrik olmayan regresyon fonksiyonu
$\hat{m}(x)$	Parametrik olmayan regresyon fonksiyonunun tahmini
$B_j(x; q)$	q dereceli B-eğrisel çizgi
$\hat{\beta}_{p,TA}(u_0)$	Tek adım tahmin edicisi
$\hat{\beta}_{p,1A}(u)$	İki adım tahmin edicisi

Kısaltmalar

AKT	Artık Kareler Toplamı (Residual Sum of Square)
DKM	Değişen Katsayılı Model (Varying Coefficient Model)
CV	Çapraz Geçerlilik (Cross-validation)
GCV	Genelleştirilmiş Çapraz Geçerlilik (Generalized Cross-validation)
\overline{CV}	Düzeltilmiş Çapraz Geçerlilik (Corrected Cross-validation)
\overline{GCV}	Düzeltilmiş Genelleştirilmiş Çapraz Geçerlilik (Corrected Generalized Cross-validation)
K(.)	Çekirdek (Kernel) fonksiyon
LOWESS	Yerel olarak ağırlıklandırılmış saçılım grafiği düzleştiricisi (Locally Weighted Scatter Plot Smoothing)
LOO	Birini-dışarıda-bırak (Leave-One-Out)
AIC	Akaike Bilgi Kriteri (Akaike Information Criterion)
TA	Tek adım tahmin edicisi (One step estimator)

İA	İki adım tahmin edicisi (Two step estimator)
MAP	Ortalama arteriyel kan basıncı (Mean arterial blood pressure)
VKI	Vücut kitle indeksi (Body mass index)
HCT	Hematokrit (Hematocrit)
BPC	Kandaki trombosit sayısı (Blood platet count)

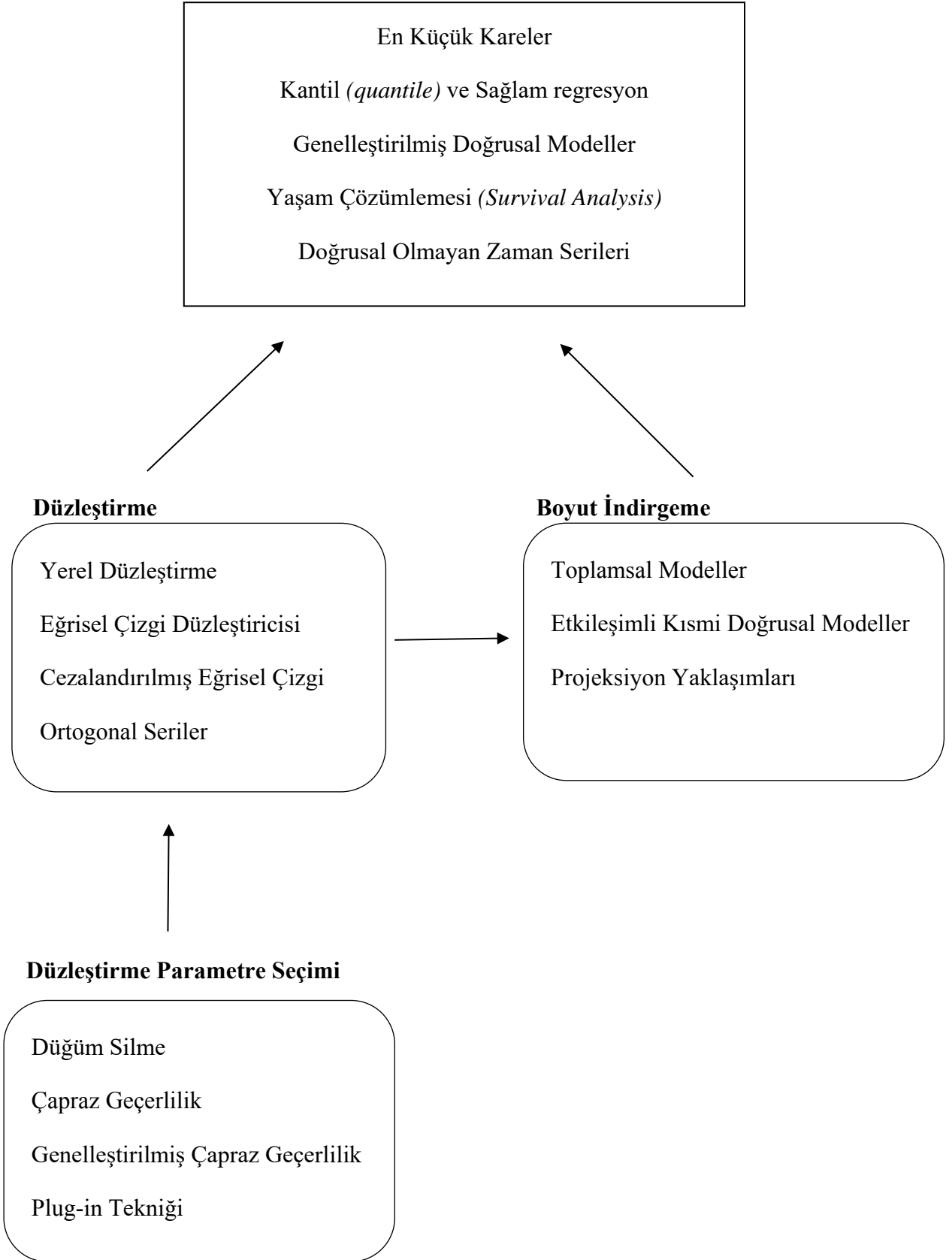
1. GİRİŞ

Regresyon analizi, istatistikte en çok kullanılan yöntemlerden birisidir. Regresyon analizinin amacı bağımsız değişkenlerin katkısını değerlendirerek ve bağımsız değişkenlerin bağımlı değişken üzerindeki etkisini belirleyerek, bağımlı ve bağımsız değişkenler arasındaki ilişkiyi ortaya çıkarmaktır. Bu çerçevede bağımlı ve bağımsız değişkenler arasındaki ilişkinin fonksiyonel şekli biliniyorsa parametrik regresyon modelleri, iki değişken arasındaki ilişkinin fonksiyonel şekli bilinmiyorsa parametrik olmayan regresyon modelleri kullanılır.

Parametrik olmayan yöntemler için bir yol haritası Şekil 1.1'deki gibi verilebilir. Bu yol haritası hangi yöntemin ne amaçla kullanılabileceğini göstermektedir. Şekil 1.1'de yerel düzleştirme yöntemleri (*local smoothing methods*); Nadaraya-Watson çekirdek regresyonu, Gasser-Müller düzleştiricisi (*smoother*), yerel olarak ağırlıklandırılmış saçılım grafiği düzleştiricisi (*Locally Weighted Scatter Plot Smoothing, LOWESS*) vb. yöntemleri içermektedir. Eğrisel çizgi (*spline*) yöntemleri temel olarak; düzleştirilmiş eğrisel çizgi (*smoothing splines*) ve polinom eğrisel çizgilerini (*polynomial splines*) kapsamaktadır. Ortogonal seri yaklaşımları; “Kesilmiş Fourier serileri yöntemi (*truncated Fourier series method*)” ve “Eşik değer dalgacık yöntemi (*thresholded wavelet method*)” gibi yöntemleri içermektedir.

Herhangi bir araştırmacı Şekil 1.1'deki istatistiksel yöntemlerinin hepsini kullanarak yeni bir çıkarsama yapabilir. Örneğin, düzleştirilmiş eğrisel çizgi (*smoothing splines*) ve düzleştirme parametresinin seçim ölçütü olarak da çapraz geçerlilik ölçütünün kullanıldığı toplamsal modelleme yöntemleri (*additive modelling techniques*) kullanılarak kantil (*quantile*) regresyon modeli önerilebilir. Ancak, Şekil 1.1'deki yol haritasının temel amacı yeni regresyon modellerinin ortaya çıkarılması için bir fikir vermesinden ziyade parametrik olmayan yöntemlere ilişkin literatürün genel bir resmini ortaya koymaktır [1].

İstatistiksel Yöntemler



Şekil 1.1. Parametrik olmayan yöntemler için bir yol haritası [1]

Şekil 1.1’de verilen yol haritasından hareketle, bu tez çalışmasının ikinci bölümünde doğrusal olmayan regresyon modelleri ve tahmin yöntemleri tanıtılmıştır. Parametrik olmayan regresyonda kullanılan tahmin yöntemlerinden; yerel düzleştirici, çekirdek düzleştirici, yerel olarak ağırlıklandırılmış saçılım grafiği düzleştiricisi, düzleştirilmiş eğrisel çizgi ve cezalandırılmış eğrisel çizgi yöntemleri ayrıntılı olarak incelenmiştir. Yerel modellemede bant genişliğinin seçimi ve cezalandırılmış eğrisel çizgi yönteminde düzleştirme parametresinin seçimi için önerilen ölçütlerden yaygın olarak kullanılan çapraz geçerlilik ve genelleştirilmiş çapraz geçerlilik ölçütleri verilmiştir.

Değişen katsayılı regresyon modelleri ile ilgili yapılmış çalışmaların bazıları aşağıda verilmiştir.

Zaman serisi analizinde bu modellerin ilk uygulamaları Haggan ve Ozaki (1981) tarafından çalışılmıştır [2].

Chen ve Tsay (1993), Cai ve diğerleri (2000) bu modellerin doğrusal olmayan zaman serisi analizindeki uygulamalarını incelemişlerdir [3, 4].

Cai ve diğerleri (2000) değişen katsayılı otoregresif modellere ilişkin istatistiksel çıkarımları vermişlerdir [5].

Hastie ve Tibshirani (1993) parametrik olmayan yöntemler için bu modelleri kullanmışlardır [6].

Brumback ve Rice (1998), Hoover ve diğerleri (1998), Wu ve diğerleri (1998), Fan ve Zhang (2000), Chiang ve diğerleri (2001) ve Huang ve diğerleri (2002), değişen katsayılı modellerin uzun kesit veri analizindeki uygulamalarını göstermişlerdir [7, 8, 9, 10, 11, 12]. Bu çalışmalarda zaman içinde değişkenlerin hangilerinin bağımlı değişkeni etkilediği değişen katsayılı modeller ile araştırılmıştır.

Cai ve Sun (2003) ve Tian ve diğerleri (2005) zamanla değişen katsayılı Cox modelini incelemişlerdir [13, 14].

Fan ve Huang (2005), değişen katsayılı modellerle ilgili teorik çıkarımlar üzerine çalışmışlardır [15].

Cai ve Tiwari (2000), çevresel çalışmalarda değişen katsayılı modelleri uygulamıştır [16].

Literatür taramasıyla ilgili detaylı bilgi, değişen katsayılı modellerin anlatıldığı üçüncü bölümde ayrıca verilmiştir. Üçüncü bölümde aynı zamanda değişen katsayılı regresyon modeli tanıtılmış, değişen katsayılı regresyon modelinin tahmininde kullanılan yerel

polinom düzleştirici, düzleştirilmiş eğrisel çizgi ve cezalandırılmış eğrisel çizgi yöntemleri detaylı olarak incelenmiştir.

Dördüncü bölümde, değişen katsayılı model gebelere ilişkin verilere uygulanmıştır. Bu bölümde, ortalama arteriyel kan basıncını etkilediği düşünülen risk faktörleri değişen katsayılı model ile incelenmiştir. Gebelerde ortalama arteriyel kan basıncı değerinin incelenmesinin nedeni, bu değer anne ve fetus sağlığı için gebelikte görülen hipertansif sendromların (gebelik hipertansiyonu, preeklampsi, eklampsi vb.) ortaya çıkarılmasında en iyi göstergelerden birisi olmasıdır.

Literatürdeki çalışmalar incelenerek ve uzman görüşü alınarak ortalama arteriyel kan basıncı üzerinde etkili olabilecek değişkenler; yaş, vücut kitle indeksi, gebelik öyküsü, düşük öyküsü, aile hipertansiyon öyküsü, diyabet hastalığı öyküsü, kötü obstetrik öyküsü, hematokrit (HCT) ve kandaki trombosit sayısı (BPC) olarak belirlenmiştir. Modelde kadınların “gestasyonel yaşı (gebelik haftası)” indeks değişken olarak alınmıştır. Model tahmininde düzleştirilmiş eğrisel çizgi ve cezalandırılmış eğrisel çizgi yöntemleri kullanılmıştır. Düzleştirilmiş eğrisel çizgi yöntemi ile elde edilen model daha küçük Akaike Bilgi Kriteri değerine sahip olduğu için bu model ile tahmin edilen sonuçlar di kkte alınmıştır.

Son bölümde ise düzleştirilmiş eğrisel çizgi yöntemi ile elde edilen model sonuçları tartışılmıştır.

2. REGRESYON MODELLERİ

2.1 Doğrusal Regresyon

Doğrusal regresyon modeli iki değişken arasındaki ilişkiyi modellemek için yaygın olarak kullanılan bir modeldir. (x_i, y_i) , $i = 1, \dots, n$, veri çiftleri kullanılarak veriyi en iyi şekilde temsil edecek bir model tahmin edilmeye çalışılır. Modelde açıklanamayan kısım hata olarak değerlendirilir. Genel olarak basit regresyon modeli,

$$y_i = \beta_0 + \beta_1 x_{i1} + \varepsilon_i \quad (2.1)$$

şeklinde ifade edilir. Burada β bilinmeyen parametre ve ε hata terimidir. Hataların genellikle bağımsız, özdeş dağıldığı varsayılır. Regresyon analizinin temel amaçları; iki değişken arasındaki ilişkiyi özetlemek, her x değerinin y bağımlı değişkenine katkısını ölçmek, verilen x değeri için ortalama bağımlı değişken değerini tahmin etmek ve verilen x değerleri dışındaki x değerleri için de tahmin yapabilmektir [1].

Doğrusal regresyon modelinde, bağımlı değişken ortalaması doğrusaldır. Buna göre

$$E(Y|X = x_i) \equiv m(x) = \beta_0 + \beta_1 x_i \quad (2.2)$$

dir. Doğrusal regresyon modelini kestirebilmek için Gauss-Markov varsayımları olarak bilinen bazı varsayımların sağlanması gerekmektedir. Bu varsayımlar; hataların ortalamasının sıfır ve sabit varyansa sahip olması, x_i 'nin sabit olması, x_i ile ε_i arasında ilişki olmaması ve ε_i 'ler arasında ilişki olmamasıdır.

2.2 Doğrusal Olmayan Regresyon

Çoğu zaman parametrik regresyon modeli ile ilgili varsayımlar sağlanmayabilir ve ya verilerin saçılımı doğrusal olmayabilir. Bu gibi durumlarda "Saçılım grafiği düzleştiricisi" (*Scatter plot smoothing*), ilk olarak akla gelebilir. Saçılım grafiği düzleştiricisi saçılım grafiğindeki "temel eğilimi" mümkün mertebede düzleştiren ve yaygın olarak kullanılan en basit düzleştirici yöntemidir. Bu yöntemde saçılım noktaları, bir olasılık modeline bakılmaksızın bir düzlemdeki noktalar topluluğu olarak basitçe ele alınır. Saçılım grafiğindeki temel eğilim genel olarak;

$$m(x) = E(y|x) \quad (2.3)$$

şeklinde ifade edilir. Başka bir deyişle,

$$y_i = m(x_i) + \varepsilon_i, \quad E(\varepsilon_i) = 0 \quad (2.4)$$

dir. (2.4)'teki m model parametrik olmayan regresyon modelinin genel ifadesidir. (2.4)'te fonksiyon m , (x_i, y_i) veri çiftinden tahmin edilmesi gereken belirsiz "düzgün (*smooth*)" ve sürekli bir fonksiyondur. Hata terimlerinin ortalaması 0 ve varyansının σ^2 olduğu varsayımıyla, bilinmeyen m fonksiyonu verilerden tahmin edilmeye çalışılır.

Verilerin saçılımının doğrusal olmadığı durumlarda yaygın olarak kullanılan modellerden birisi de polinom regresyon modelidir. Bu modellerde polinom derecesi seçilerek, modeldeki parametre sayısı artırılır. Bu model sıkça kullanılmasına rağmen uygulamada birçok sorunla karşılaşmaktadır. Bunlardan biri polinom fonksiyonlarının modellemede çok esnek olmamasıdır. Çünkü parametreler tahmin edilirken polinom fonksiyonlarının tüm mertebeden türevlerinin alınması gerekmektedir. Bir diğeri; her bir gözlemin polinom eğrisinin uzak parçaları üzerinde büyük bir etkiye sahip olmasıdır. Üçüncü bir nokta ise polinom derecesinin sürekli olarak kontrol edilemeyecek olmasıdır. Sonuç olarak polinom derecesi arttıkça veriyi temsil edebilecek uygun polinomun seçilmesi zor ve zaman alıcı olacaktır [1].

2.2.1 Yerel Modelleme

Eğri uydurmada polinom kestiriminin sakıncalarını gidermenin birkaç yolu vardır. Bunlardan biri, türev eğrilerinin olası süreksizliklerine izin vermektir. Bu, eğrisel çizgi yaklaşımına (*spline approach*) yönlendirir. "Düğüm" olarak adlandırılan süreksizlik noktalarının yerleri, düzleştirilmiş eğrisel çizgi (*smoothing spline*) veya adımsal silme yöntemi ile elde edilen veriler ile seçilebilir [1].

Diğer bir öneri, regresyon fonksiyonunu bir ortogonal seri olarak ifade etmektir. Daha sonra temel fonksiyonlarının birkaç alt kümesini seçmek ve yaklaşık regresyon fonksiyonunu elde etmek için bunları kullanmaktır. Bu yaklaşıma "ortogonal seri yöntemi" denir. Üçüncü bir yaklaşım ise, parametre sayısını arttırmak yerine doğrusal regresyon modelini yerel olarak tahmin etmektir. Yani seçilen herhangi bir x noktasında, x etrafında doğrusal olarak $m(\cdot)$ fonksiyonunu modellemek; başka bir deyişle doğrusal regresyon yöntemini x etrafındaki verilerin bir kısmına uygulamaktır. Bu yaklaşım yerel (doğrusal) modelleme (*local linear modelling*) yaklaşımı olarak adlandırılır. Bant genişliği olarak adlandırılan yerel komşuluğun genişliği, ya öznel olarak veri analizcileri tarafından ya da nesnel olarak veriler kullanılarak seçilmektedir. Böylece doğrusal olmama sorunu, birden çok parametrelili

parametrik yöntem kullanmak yerine, iki parametrelili birçok doğrusal regresyon modeli kullanılarak çözülmüş olur.

Belirli bir x noktasındaki regresyon eğrisini tahmin etmek için doğrusal regresyon yöntemi x etrafındaki verilerin oluşturduğu bir şeritte uygulanır. Yerel komşuluk genişliği h olsun. Şeritteki veriler,

$$y_i = b_0(x) + b_1(x)x_i + \varepsilon, \quad x_i \in x \pm h \quad (2.5)$$

modeli ile modellenir. x 'ten uzak bir veri noktasının katkısını azaltmak için yerel en küçük kareler toplamı ağırlıklandırılabilir.

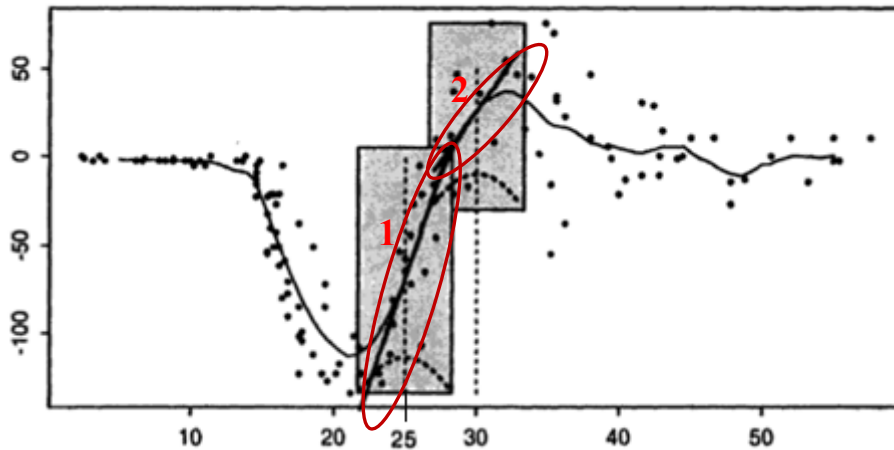
K , tek tepeli negatif olmayan bir fonksiyon olsun. Her bir (x_i, y_i) noktası $K\{(x_i - x)/h\}$ ağırlığı ile ağırlıklandırılır. Buna göre ağırlıklandırılmış en küçük kareler toplamı,

$$\sum_{i=1}^n \{y_i - b_0(x) - b_1(x)x_i\}^2 K\left(\frac{x_i - x}{h}\right) \quad (2.6)$$

şeklinde ifade edilir. x noktasındaki tahmin edilen regresyon eğrisi aşağıdaki gibi ifade edilir:

$$\hat{m}(x) = \hat{b}_0(x) + \hat{b}_1(x)x \quad (2.7)$$

Herhangi bir x noktasında yerel modelleme ile regresyon modeli tahmininin nasıl yapıldığı Şekil 2.1'de gösterilmiştir.

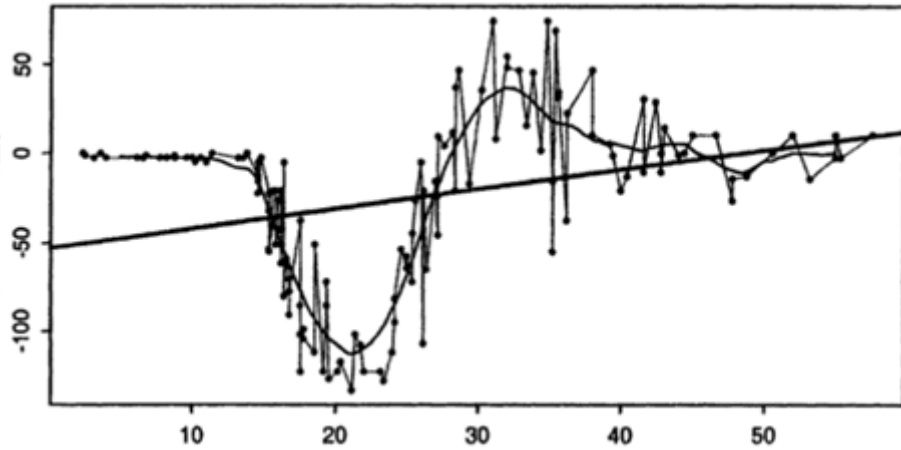


Şekil 2.1. Yerel modelleme yaklaşımı ile tahmin [1].

Şekil 2.1'deki gösterimde $x=25$ noktası için ağırlıklandırılmış doğrusal regresyon modeli tahmini, $x = 25 \pm 3.3$ şeridindeki gözlemler $K(u) = (1 - u^2)_+$ (“+” alt indis, pozitif değerleri ifade eder) çekirdek fonksiyonu ile ağırlıklandırılarak bulunur. Bu tahmin, Şekil 2.1'de 1 numaralı elips içindeki doğrusal çizgi ile gösterilmiştir. $x=30$ noktası komşuluğundaki ağırlıklandırılmış doğrusal regresyon modeli tahmini, $x = 30 \pm 3.3$ şeridindeki gözlemler aynı çekirdek fonksiyonu ile ağırlıklandırılarak elde edilir. Bu tahmin Şekil 2.1'de, 2 numaralı elips içindeki doğrusal çizgi ile gösterilmiştir. Bu işlem birçok x değeri için tekrarlanarak Şekil 2.1'de düz çizgi ile gösterilen yerel doğrusal tahmin eğrisi elde edilir.

2.2.2 Bant Genişliği Seçimi

Yerel modellemede ilk akla gelen soru, yerel modelin geçerli olması için yerel komşuluğun ne kadar genişlikte olması gerekliliğidir. Bu soru (2.6)'daki h bant genişliği parametresinin ne kadar olması gerektiği sorusu ile eşdeğerdir. Bant genişliği parametresi h çok küçük seçilirse model yanlılığı küçük olur. Ancak, yerel komşuluktaki gözlemlerin sayısı da az olacağı için, tahmin edilen yerel parametrenin, $\hat{b}_1(x)$, varyansı da büyük olacaktır. Diğer taraftan, bant genişliği h büyük olursa, model yanlılığı büyüyecektir. Bant genişliği seçimine göre model yanlılığı Şekil 2.2'de gösterilmiştir:



Şekil 2.2. Yerel doğrusal kestirim; $h = 0$ (kesikli çizgiler), $h=3.3$ (ince eğri) ve $h=$ (düz çizgi) için kestirimler. h , 0'dan ∞ 'a, tahmin en karmaşık eğriden düz çizgiye yaklaşır [1].

Şekil 2.2’de, $h=0; 3.3 ; +\infty$ için yerel doğrusal kestirim elde edilmiştir. h küçük olduğunda ($h=0$) tahmin; aslında veri noktalarının interpolasyonudur. h büyük olduğunda ($h=\infty$) ise tahmin, parametrik doğrusal regresyon tahmini ile aynıdır. h parametresine bağlı yerel doğrusal modellerin tüm ailesi, h ’nin değerine göre $\{M_h : h \in (0, \infty)\}$ en karmaşık model M_0 ’dan (interpolasyon modeli) en basit model M_∞ ’a (parametrik doğrusal modele) döner. Görüldüğü gibi bant genişliği parametresi model karmaşıklığını belirleyen bir parametredir. Bu nedenle, bant genişliği seçimi, belirli bir veri seti için $\{M_h\}$ ailesinden $\{M_{h_0}\}$ modelinin seçilmesi olarak değerlendirilebilir. Yani veriye en uygun modelin seçilmesi anlamına gelmektedir. Şekil 2.2’de görüldüğü gibi $h=3.3$ için veriye en uygun model tahmin edilmiştir. Bant genişliği tahmin edilen eğrinin hatasını minimum yapacak şekilde ya da araştırmacı tarafından subjektif olarak da seçilebilir. Bant genişliğinin seçimi için çok sayıda ölçüt geliştirilmiştir. Bu ölçütlerden “çapraz geçerlilik” ve “genelleştirilmiş çapraz geçerlilik” ölçütleri bant genişliği seçiminde yaygın olarak kullanılan ölçütlerdir [1].

2.3 Tahmin Yöntemleri

y ve x arasındaki ilişkiyi veren bilinmeyen $m(x)$ fonksiyonunun düzgün bir şekilde tahmin edilmesi için kullanılan yöntemlere genel olarak parametrik olmayan regresyon yöntemleri denilmektedir. L iteratürde y aygın olarak kullanılan parametrik olmayan regresyon yöntemleri (düzleştiriciler) aşağıda verilmiştir:

- Yerel (*Local*) polinom düzleştirici
- Çekirdek (*Kernel*) düzleştirici
- Yerel olarak ağırlıklandırılmış saçılım grafiği düzleştiricisi (*Locally Weighted Scatter Plot Smoothing, LOWESS*)
- Eğrisel çizgi düzleştiricisi (*Spline smoothing*)
- Cezalandırılmış eğrisel çizgi (*Penalized spline*)

2.3.1 Yerel (Local) Polinom Düzleştirici

$(x_1, y_1), \dots, (x_n, y_n)$ rasgele örneklem değerleri ele alınsın. Parametrik olmayan regresyon modeli

$$y_i = m(x_i) + \varepsilon_i \quad (2.8)$$

biçiminde ifade edilir.

(2.8)'de $E(\varepsilon|x) = 0$ ve $V(\varepsilon|X = x) = \sigma^2(x)$ olup, $m(x) = E(Y|X = x)$ fonksiyonunun düzgün ancak parametrik olmayan bir fonksiyon olduğu varsayılır. Burada amaç, $m(x)$ fonksiyonun tahmini $\hat{m}(x)$ 'yi bulmaktır. Yerel polinom tahminini bulabilmek için ilk olarak x 'in komşuluğundaki z için $m(z)$ 'nin Taylor açılımı bulunur. Buna göre $m(z)$ 'nin Taylor açılımı,

$$m(z) \approx \sum_{j=0}^p \frac{m^{(j)}(x)}{j!} (z-x)^j \equiv \sum_{j=0}^p B_j (z-x)^j \quad (2.9)$$

biçimindedir [1]. x 'in komşuluğundaki herhangi bir veri için

$$m(x_i) \approx \sum_{j=0}^p B_j (x_i - x)^j \equiv \mathbf{x}_i^T \boldsymbol{\beta} \quad (2.10)$$

dir. (2.10)'da $\mathbf{x}_i = (1, (x_i - x), \dots, (x_i - x)^p)^T$ ve $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_p)^T$ 'dir. Sezgisel olarak, x 'e daha yakın gözlem değeri $m(x)$ üzerinde daha etkili olurken; x 'ten uzaktaki gözlem değerinin $m(x)$ üzerinde daha az etkiye sahip olduğu söylenebilir. Bu nedenle $m(x)$ 'in tahmininde yerel olarak ağırlıklandırılmış polinom regresyon modeli kullanılabilir. Yerel olarak ağırlıklandırılmış en küçük kareler toplamı;

$$\sum_{i=0}^n (y_i - \mathbf{x}_i^T \boldsymbol{\beta})^2 K_h(x_i - x) = \sum_{i=0}^n \left\{ y_i - \sum_{j=0}^p \beta_j (x_i - x)^j \right\}^2 K_h(x_i - x) = (\mathbf{y} - \mathbf{X}_x \boldsymbol{\beta})^T \mathbf{W}_x (\mathbf{y} - \mathbf{X}_x \boldsymbol{\beta}) \quad (2.11)$$

biçiminde ifade edilir. Buradaki \mathbf{X}_x ve \mathbf{W}_x matrisleri,

$$\mathbf{X}_x = \begin{bmatrix} 1 & (x_1 - x) & (x_1 - x)^2 & \cdots & (x_1 - x)^p \\ 1 & (x_2 - x) & (x_2 - x)^2 & \cdots & (x_2 - x)^p \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & (x_n - x) & (x_n - x)^2 & \cdots & (x_n - x)^p \end{bmatrix}$$

$$\mathbf{W}_x = \begin{bmatrix} K_h(x_1 - x) & 0 & \cdots & 0 \\ 0 & K_h(x_2 - x) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & K_h(x_n - x) \end{bmatrix}$$

şeklindedir. (2.11)'de $K_h(x_i - x) \equiv h^{-1}K\left(\frac{x_i - x}{h}\right)$, $K(\cdot)$, $\int K(x) = 1$ ve bant genişliği ya da düzleştirme parametresi h pozitif değerler alır.

Genellikle uygulamalarda χ ekirdek f onksiyonu olarak Gaussian χ ekirdek f onksiyonu kullanılmaktadır. Gaussian çekirdek fonksiyonu,

$$K(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right), -\infty < x < \infty \quad (2.12)$$

olup, simetrik beta ailesi

$$K(x) = \frac{1}{\text{Beta}(1/2, \gamma + 1)} (1 - x^2)_+^\gamma, \quad \gamma = 0, 1, 2, \dots, \quad -1 \leq x \leq 1 \quad (2.13)$$

şeklindedir. (2.13)'te $+$, üstelleştirmeden önce alınan pozitif kısmı ve $\text{Beta}(\cdot, \cdot)$, Beta fonksiyonunu belirtmektedir. Beta ailesinde çekirdek fonksiyonları $[-1, 1]$ aralığındadır.

$\gamma = 0, 1, 2$ ve 3 için bu fonksiyon sırasıyla tek biçimli çekirdek, Epanechnikov çekirdek, iki ağırlıklı çekirdek (*Biweight kernel*) ve üç ağırlıklı çekirdek fonksiyonları olarak adlandırılmaktadır. Bu fonksiyonlar ve daha sonraki bölümlerde kullanılacak olan *Tricube* çekirdek fonksiyonu Çizelge 2.1'de verilmiştir [17].

Çizelge 2.1. Çekirdek Fonksiyonları

Gaussian çekirdek	$K(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2}, u \in [-\infty; \infty]$
Tek biçimli çekirdek	$K(u) = \frac{1}{2}, u \in [-1, 1]$
Üçgensel çekirdek	$K(u) = (1 - u), u \in [-1, 1]$
Epanechnikov çekirdek	$K(u) = \frac{3}{4}(1 - u^2), u \in [-1, 1]$
İki Ağırlıklı çekirdek	$K(u) = \frac{15}{16}(1 - u^2)^2, u \in [-1, 1]$
Üç Ağırlıklı çekirdek	$K(u) = \frac{35}{32}(1 - u^2)^3, u \in [-1, 1]$
Tricube çekirdek	$K(u) = \frac{70}{81}(1 - u ^3)^3, u \in [-1, 1]$

Yerel polinom tahmin edicisi $\hat{\boldsymbol{\beta}} = (\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_p)^T$, (2.11)'deki yerel ağırlıklandırılmış kareler toplamının $\boldsymbol{\beta}$ 'ya göre türevi alınıp sifira eşitlenerek,

$$\hat{\boldsymbol{\beta}} = \min_{\boldsymbol{\beta}} \sum_{i=0}^n (y_i - \mathbf{x}_i^T \boldsymbol{\beta})^2 K_h(x_i - x) = (\mathbf{X}_x^T \mathbf{W}_x \mathbf{X}_x)^{-1} \mathbf{X}_x^T \mathbf{W}_x \mathbf{y} \quad (2.14)$$

şeklinde elde edilir.

(2.14)'teki ifade (2.10)'da yerine konularak $m(x)$ fonksiyonunun yerel polinom regresyon kestiricisi,

$$\hat{m}(x) = \mathbf{t}^T (\mathbf{X}_x^T \mathbf{W}_x \mathbf{X}_x)^{-1} \mathbf{X}_x^T \mathbf{W}_x \mathbf{y}$$

şeklinde elde edilir. Bu eşitlikte $\mathbf{t}^T = (1, 0, \dots, 0)_{1 \times (p+1)}^T$ 'dir [17].

Tanım olarak, $m(x) \equiv \beta_0$ olduğundan, tahmin edilen regresyon fonksiyonu $\hat{m}(x) = \hat{\beta}_0(x)$ 'dır. $\hat{\beta}_0(x)$, x 'in bir fonksiyonunun $\hat{\beta}_0$ olduğunu vurgulamak için kullanılmıştır. Buna ek olarak, $m(x)$ regresyon fonksiyonunun $\boldsymbol{\beta}$ 'ya göre x noktasında v 'inci dereceden türevinin bir tahmin edicisi aşağıdaki gibidir:

$$\hat{m}^{(v)}(x) = v! \hat{\beta}_v(x) \quad (2.15)$$

2.3.2 Çekirdek (Kernel) Düzleştiricisi ve Yerel Doğrusal Düzleştirici

Çekirdek düzleştiricisi ve yerel doğrusal düzleştirici, yerel polinom düzleştiricisinin özel durumlarıdır. (2.10)'da $p = 0$ olduğunda, yerel polinom düzleştiricisi “çekirdek düzleştiricisi” olarak adlandırılır [18]. Çekirdek düzleştiricisi yerel sabit kestirim olarak düşünülebilir. Bununla birlikte, çekirdek düzleştiricisi, $m(x)$ regresyon fonksiyonunun türevlerini tahmin etmez. Çekirdek düzleştiricisinin tahmin edicisi, Ek-1'de verilen $m(x) = E(Y|X = x)$ koşullu beklenen değeri bulunarak aşağıdaki gibi elde edilir:

$$\widehat{m}_h(x) = \frac{\sum_{i=1}^n K_h(x_i - x)y_i}{\sum_{i=1}^n K_h(x_i - x)} \quad (2.16)$$

(2.16)'daki tahmin edici Nadaraya-Watson çekirdek düzleştiricisinin tahmin edicisidir [19,20].

Yaygın olarak kullanılan diğer çekirdek tahmin edicisi ise Gasser-Müller çekirdek tahmin edicisidir [21,22].

Gasser-Müller çekirdek tahmin edicisi,

$$\widehat{m}_h(x) = \sum_{i=1}^n \int_{s_{i-1}}^{s_i} K_h(u - x) du y_i$$

şeklindedir. Burada $s_i = (x_i + x_{i+1})/2$ ve $X_0 = -\infty, X_{n+1} = +\infty$ olduğu varsayılmaktadır.

(2.10)'da $p = 1$ olduğunda yerel polinom düzleştiricisi “yerel doğrusal düzleştirici” olarak adlandırılır. Yerel doğrusal düzleştirici sadece $m(x)$ fonksiyonunu tahmin etmekle kalmayıp aynı zamanda $m(x)$ fonksiyonunun birinci türevleri olan $m'(x)$ 'i de tahmin eder. $m(x)$ fonksiyonunun yerel doğrusal tahmin edicisi,

$$\widehat{m}(x) = n^{-1} \sum_{i=1}^n \frac{(s_2(x) - s_1(x)(x_i - x))K_h(x_i - x)y_i}{s_2(x)s_0(x) - s_1(x)^2} \quad (2.17)$$

şeklinde elde edilir. (2.17)'de $s_r(x) = n^{-1} \sum_{i=1}^n (x_i - x)^r K_h(x_i - x)$ 'dir.

Nadaraya-Watson tahmin edicisi, Gasser-Müller ve yerel doğrusal tahmin edicilerinin asimptotik yanlılık ve varyans değerleri Çizelge 2.2'de verilmiştir [18].

Çizelge 2.2. Noktasal asimptotik yanlılık ve varyans

Yöntem	Yanlılık	Varyans
Nadaraya-Watson	$\left(m''(x) + \frac{2m'(x)m'(x)}{m(x)} \right) b_n$	V_n
Gasser-Müller	$m''(x)b_n$	$1,5V_n$
Yerel Doğrusal	$m''(x)b_n$	V_n

$$* b_n = \frac{1}{2} \int_{-\infty}^{\infty} u^2 K(u) du h^2 \quad \text{ve} \quad V_n = \frac{\sigma^2(x)}{m(x)nh} \int_{-\infty}^{\infty} K(u) du$$

2.3.3 Yerel Olarak Ağırlıklandırılmış Saçılım Grafiği Düzleştiricileri

Yerel ortalama düzleştiricileri (kernel düzleştiricisi vb.) aykırı değerlerden etkilenirler. Bu etki eğri kestiriminin olması gerekenden çok farklı olarak elde edilmesine neden olur. Bu gibi durumlarda aykırı değerlere karşı daha dayanıklı bir tahmin yöntemi tercih edilmelidir. Bu Alt Bölüm'de parametrik olmayan regresyonda aykırı değerlerin olması durumunda Cleveland [23] tarafından önerilen sağlam kestirim yöntemleri açıklanacaktır.

2.3.3.1 Yerel Olarak Ağırlıklandırılmış Saçılım Grafiği Düzleştiricisi (Locally Weighted Scatter plot Smoothing, LOWESS)

Yerel olarak ağırlıklandırılmış en küçük kareler kestiricisinin sağlam kestiricisi olan yerel olarak ağırlıklandırılmış saçılım grafiği düzleştiricisi (LOWESS), Cleveland [23] tarafından önerilmiştir. Bu düzleştirici ile kestirim yapılırken ilk olarak (2.11)'deki ağırlıklandırılmış kareler toplamı minimum yapılarak kestirim elde edilir. Daha sonra kestirim değerleri kullanılarak artıklar hesaplanır ve büyük artık değerlerine daha küçük ağırlıklar verilerek artıklar ağırlıklandırılır. Tekrar p'inci dereceden yerel polinom kestirimi elde edilir ancak burada başlangıç kestiriminden elde edilen artıklara verilen ağırlıklar ve başlangıç kestiriminde her bir gözleme verilen ağırlıkların çarpımı olan yeni bir ağırlık her gözleme atanır. Yani başlangıç kestiriminde büyük artıklara sahip gözlemlere ikinci kestirimde daha düşük ağırlıklar verilmiş olur. Bu süreç birçok kez tekrarlanarak LOWESS tahmin edicisi elde edilir [1].

Yerel polinom kestirimi, LOWESS yönteminin temel bileşenidir ve bu nedenle bant genişliği, çekirdek fonksiyonu ve polinom derecesinin nasıl seçileceğine dair sorular önem arz etmektedir. Cleveland [23], LOWESS yönteminde Çizelge 2.1'deki Tricube çekirdek (*Tricube kernel*) fonksiyonunu kullanmıştır. Bu özel ağırlık fonksiyonunun seçilmesinin

nedenlerinden biri, sıfıra doğru düzgün bir şekilde azalmasıdır. Yerel komşuluklar ayrıca en yakın komşu bant genişliği ile belirlenmektedir. $0 < d \leq 1$ olsun ve nd , ve r ’in bir bölümünü gösterebilir. nd , en yakın tamsayıya yuvarlansın ve bu sayı “ r ” ile gösterilsin.

Verilen x_1, \dots, x_n gözlemleri için, x_k gözleminin komşuluğu, x_k ’dan r ’inci en yakın komşuya olan mesafe olarak tanımlanan h_k bant genişliği ile belirlenir. Daha doğrusu, $j=1, 2, \dots, n$ için $h_k, |x_k - x_j|$ arasındaki r . en küçük değerdir. Çekirdek fonksiyonu ve bant genişliği belirlendikten sonra, x_k noktasındaki kestirim değeri bulunabilir. Her bir x_i gözlemine

$$K_i(x_k) = K\{h_k^{-1}(x_i - x_k)\} \quad (2.18)$$

ağırlıkları atanır. Bu ağırlıklar, (2.11)’deki ağırlıklandırılmış kareler toplamını minimum yapan $\hat{\beta}_j, j=0, \dots, p$ tahminlerini elde etmek için kullanılır. y_k ’nın başlangıç kestirim değeri $\hat{y}_k = \hat{\beta}_0 = \hat{\beta}_0(x_k)$ olarak elde edilir. Yerel polinom regresyon tahmin edicileri doğrusal düzleştiriciler (*linear smoothers*) olduğundan

$$\hat{y}_k = \sum_{i=1}^n w_i(x_k) y_i \quad (2.19)$$

şeklinde yazılabilir. Burada görüldüğü gibi $w_i(x_k)$ ağırlıkları $\{x_i\}$ noktasına bağlı olduğu kadar x_k ’ya da bağlıdır. Yukarıdaki adım, her x_k gözlemi için yapılır, yani her nokta x_k için bir yerel polinom kestirimi elde edilir. Bu sayede başlangıç kestirim değerleri olan $\hat{y}_1, \dots, \hat{y}_n$ değerleri elde edilir.

Başlangıç kestirimlerinden elde edilen artıklar,

$$r_k = y_k - \hat{y}_k, \quad k = 1, \dots, n. \quad (2.20)$$

şeklinde yazılır. $M, |r_1|, \dots, |r_n|$ dizisinin medyanı olmak üzere sağlamlık ağırlıkları aşağıdaki gibi tanımlanır:

$$\delta_i = B(r_i / (6M)), \quad i = 1, \dots, n, \quad (2.21)$$

Burada $B(t) = (1 - |t|^2)^2 I_{[-1,1]}(t)$, (2.13)’de $\gamma = 2$ olduğundaki iki ağırlıklı çekirdek (*biweight kernel*) fonksiyonudur [1]. Böylelikle ilk adım bitirilir.

$(x_i, y_i), i = 1, \dots, n$ gözlem çifti için $\delta_i K_i(x_k)$ ağırlığı kullanılarak ikinci adımdaki p 'inci dereceden yerel polinom kestirimi \hat{y}_k elde edilir. t iterasyondan sonra, $\hat{y}_k, k = 1, \dots, n$ sağlam yerel olarak ağırlıklandırılmış regresyon tahmin edicisi elde edilir.

LOWESS değerlerinin hesaplanması için kullanılan algoritma aşağıdaki gibi özetlenebilir:

1. Her k için yerel olarak ağırlıklandırılmış regresyon kestirimleri \hat{y}_k elde edilir ve r_k artık değerleri hesaplanır.
2. Eşitlik (2.21)'deki $\delta_i, i = 1, \dots, n$ sağlamlık ağırlıkları hesaplanır.
3. Her k için; (x_i, y_i) için $\delta_i K_i(x_k)$ ağırlıkları kullanılarak yerel ağırlıklandırılmış regresyon yöntemi ile yeni kestirim değeri \hat{y}_k elde edilir.
4. Adım 2 ve 3 toplam t kez tekrarlanır. Elde edilen kestirim değerleri tahmin eğrisini oluşturur.

Cleveland [23], $p = 1$ ve $t = 3$ olarak alınmasını önermiştir. Bant genişliği d 'ye bağlıdır ve 0.2 ile 0.8 arasında bir değer seçilmesi önerilmektedir. $t=0$ olduğunda, LOWESS en yakın komşu bant genişliğini kullanarak yerel polinom kestirime karşılık gelir [1].

2.3.3.2 Loess

Loess, saçılım grafiğinde veri noktalarından geçen düz bir eğri kestirmek için yerel ağırlıklandırılmış regresyonu kullanan, parametrik olmayan bir yöntemdir. Loess eğrileri parametrik olarak modellenmesi zor olan verilerdeki eğilimleri ortaya çıkarır. LOWESS'in çok değişkenli versiyonu olarak düşünülebilir. Loess, Cleveland ve Delvin [24] tarafından önerilmiştir. Cleveland ve Delvin [24] ağırlık fonksiyonu olarak Tricube çekirdek (*Tricube kernel*) fonksiyonunu kullanmışlardır. Ancak herhangi bir ağırlık fonksiyonu da kullanılabilir. Loess, noktalar arasındaki mesafeyi belirlemek için Öklid mesafesini kullanır. $D_k(x)$, x noktasından k 'nci en yakın komşularına (en uzak komşuya) olan Öklid uzaklığı olsun. Dolayısıyla, $h = D_k(x)$, x noktası için bant genişliğidir. Bu durumda Loess ağırlık fonksiyonu,

$$K_h(u) = \frac{1}{h} K\left(\frac{D(u)}{h}\right) \quad (2.22)$$

dır. Burada $u = (x_i - x)$ ve $D(u) = \sqrt{(x_1 - x)^2 + (x_2 - x)^2 + \dots + (x_n - x)^2}$ 'dir [1].

Loess yönteminde bant genişliğinin seçimi için Stone [25] tarafından önerilen bir gözlemin dışarıda bırakılarak kestirimlerin elde edilmesine dayanan birini-dışarıda-bırak (*Leave-One-Out, LOO*) ölçütü yaygın olarak kullanılmaktadır [1]. Bu ölçüt çapraz geçerlilik ölçütünün başka bir çeşitidir. Bu ölçütte,

$$\sum_{i=1}^n (y_i - \hat{m}(x)_{-i})^2 \quad (2.23)$$

kareler toplamını minimum yapan h değeri bant genişliği olarak seçilir. Burada $\hat{m}(x)_{-i}$, i 'nci gözlem çıkarıldıktan sonra elde edilen kestirim değeridir. Bu ölçüt i le bant genişliğini tahmin etmek zaman alıcı olduğu için genellikle $K=5$ ya da $K=10$ olan alt grupların dışarıda bırakılmasına dayanan K -katlı çapraz geçerlilik ölçütü minimize edilir.

2.3.4 Eğrisel Çizgi Düzleştiricisi (Spline Smoothing)

Tüm veri noktalarında türevlenebilen bir polinom fonksiyonu, farklı veri noktalarında farklı düzgünlük dercesine sahip eğrileri kestirmek için esnek bir yöntem değildir. Polinom fonksiyonunun bu dezavantajı yerel olarak düşük dereceli bir polinom yaklaşımı kullanılarak giderilebilir. Diğer bir yol ise, belirli veri noktalarında fonksiyon türevlerinin süreksizliğine izin veren bir yöntem kullanılmasıdır. Bu yöntem parçalı polinom veya eğrilerin kestirilmesine dayanan eğrisel çizgi yöntemidir. Eğrisel çizgi yönteminde fonksiyonunun türevlerinin süreksizliklere sahip olabileceği veri noktaları “düğüm (*knots*)” olarak adlandırılır. Eubank [26] ve Wahba [27] eğrisel çizgi yaklaşımı ve istatistiksel uygulamalarına dair çeşitli çalışmalar yapmışlardır. Green ve Silverman [28], istatistikte çeşitli alanlarda eğrisel çizgi yöntemlerini uygulamışlardır [1].

Bu alt bölümde ilk olarak eğrisel çizgi yöntemi sabit düğüm noktalarına göre verilmiştir. Daha sonra düğüm noktası seçimi için iki yöntem açıklanmıştır:

- İlk yöntem Smith [29] ve Breiman ve diğerleri [30] düğüm silme fikrine dayanmaktadır.
- İkinci yöntem ise Wittaker [31] ve Wahba [32] düzleştirilmiş eğri fikrine dayanmaktadır.

2.3.4.1 Polinom Eğrisel Çizgi

Bilinmeyen regresyon fonksiyonu m , bir eğrisel çizgi (*spline*) fonksiyonu ile tahmin edilsin. Çoğunlukla tahmin için kübik eğrisel çizgi, başka bir deyişle sürekli ilk iki türeve sahip parçalı polinom, kullanılır. Bu nedenle bu alt bölümde tahmin sürecinde sadece kübik eğrisel

çizgiler ele alınmıştır. t_1, \dots, t_J sabit düğüm dizisi $-\infty < t_1 < \dots < t_J < +\infty$ şeklinde olsun. Kübik eğrisel çizgi fonksiyonları, iki kez sürekli olarak türevlenebilen fonksiyonlardır. Öyle ki, $(-\infty, t_1], [t_1, t_2], \dots, [t_{J-1}, t_J], [t_J, +\infty)$ aralıklarının her birinde m kübik polinomdur. Tüm kübik eğrisel çizgi fonksiyonları $(J+4)$ boyutlu bir doğrusal uzay oluşturmaktadır. Bu doğrusal uzay için çok bilinen iki kübik eğrisel çizgi temeli vardır [1]:

- Güç temeli (*Power basis*),
- B-Eğrisel çizgi (*B-spline*).

Güç temeli $(x - t_j)_+^3$, ($j = 1, \dots, J$), $1, x, x^2, x^3$ şeklindedir.

Genellikle, B-Eğrisel çizgi temeli daha durağandır, çünkü temel fonksiyonlar arasındaki çoklu korelasyon daha küçüktür, ancak güç eğrisel temelinin bir avantajı vardır. Güç eğrisel temelinde $(x - t_j)_+^3$ temel işlevinin silinmesi t_j düğümünün silinmesidir.

B_1, \dots, B_{J+4} kübik eğrisel çizgi temeli olsun. Bir kübik eğrisel çizgi fonksiyonu

$$m(x) = \sum_{j=1}^{J+4} \theta_j B_j(x)$$

şeklinde ifade edilebilir. Verilen düğüm noktaları için eğrisel çizgi (*spline*) yöntemi,

$$\min_{\theta} \sum_{i=1}^n \left\{ y_i - \sum_{j=1}^{J+4} \theta_j B_j(x_i) \right\}^2 \quad (2.24)$$

şeklinde ifade edilen en küçük kareler toplamını minimum yapan en i yi eğrisel çizgi tahminini bulmayı amaçlar.

En küçük kareler tahmini $\hat{\theta}_j$, $j = 1, \dots, J+4$, olmak üzere m fonksiyonunun eğrisel çizgi

tahmini $\hat{m}(x) = \sum_{j=1}^{J+4} \hat{\theta}_j B_j(x)$ şekilde yazılabilir.

Eğrisel çizgi (*spline*) yöntemi, düğüm noktaları (J) sayısının seçimi kadar düğüm noktalarının konumlarının seçimine de oldukça duyarlıdır. Düğüm noktaları genellikle eğriliklerin makul ölçüde büyük değişime sahip olduğu noktalara yerleştirilir. Düğüm noktalarını yerleştirmede deneme yanılma yöntemleri çok zaman alıcı olabilir. Bu nedenle düğüm noktalarının yerlerini seçmek için bazı yöntemler önerilmiştir [1]. Bu yöntemlerden birisi düğüm silme yaklaşımıdır. Bu yaklaşım Kooperberg ve diğerleri [33, 34] tarafından önerilmiştir.

Düğüm silme yaklaşımında, t_1, \dots, t_j düğüm noktaları seçim sürecinde silinebilecek ilk düğüm noktaları olsun. Bu düğümler genellikle x değişkeninin sıra istatistiklerine yerleştirilir ve J düğüm noktalarının sayısı $J = \lfloor n/2 \rfloor$ veya $J = \lfloor n/3 \rfloor$ olarak alınır. Başlangıç düğüm noktalarının bir örneği $t_j = x_{(2j)}$, $j = 1, \dots, \lfloor n/2 \rfloor$ 'dir. Bu düğüm noktalarına ilişkin güç temeli,

$$B_j(x) = (x - t_j)_+^3, \quad (j = 1, \dots, J),$$

$$B_{j+1}(x) = 1, \quad B_{j+2}(x) = x, \quad B_{j+3}(x) = x^2, \quad B_{j+4}(x) = x^3$$

şeklinde ifade edilir. (2.24)'den elde edilen en küçük kareler tahmini $\hat{\theta}_j$ ve bu tahminin standart hatasının tahmini $SE(\hat{\theta}_j)$ 'dir. Daha sonra j_0 'ncü düğüm noktası ($1 \leq j_0 \leq J$), en küçük mutlak t değerine ($|\hat{\theta}_j|/SE(\hat{\theta}_j)$, $1 \leq j \leq J$) göre silinir. Bahsedilen silme işlemi her adımda bir düğüm silerek tekrarlanır. Böylelikle j ile indekslenen modeller dizisi elde edilir. En küçük düzeltilmiş Mallows'un C_p değerine sahip j 'inci model seçilir. Düzeltilmiş Mallows'un C_p ölçütü [35],

$$C_j = AKT_j + \lambda(J + 4 - j) \hat{\sigma}^2 \quad (2.25)$$

şeklinde tanımlanır. Burada $\hat{\sigma}$ tam modelin standart sapmasının tahmini, $(J+4-j)$ j 'inci modelin artık kareler toplamının (AKT) serbestlik derecesi ve λ düzleştirme parametresidir. Kooperberg ve Stone [36], log-eğrisel çizgi (*logspline*) yoğunluk tahmininde $\lambda = 2$ değeri yerine $\lambda = 3$ değerinin kullanılmasını önermektedir. Burada kübik eğrisel çizgi için B-eğrisel çizgi ile tahmin incelenmiştir. Daha sonraki 2.3.5 Alt B ölüm'de cezalandırılmış eğrisel çizgiler anlatılırken B-eğrisel çizgileri ayrıntılı olarak ele alınmıştır.

2.3.4.2 Düzleştirilmiş Eğrisel Çizgi (Smoothing Spline)

Düğüm noktalarının otomatik olarak seçilmesine alternatif bir yaklaşım düzleştirilmiş eğrisel çizgi yaklaşımıdır. Bu yöntemi açıklamak için, öncelikle $\sum_{i=1}^n \{y_i - m(x_i)\}^2$ en küçük kareler toplamını minimum yapan m fonksiyonunun tahmin edilmesi problemini ele alalım. Bu problemin en basit çözümü veri noktalarının ara değerlerini bularak m fonksiyonunu tahmin etmektir. Ancak bu tahmin, ve rileri modellemekte yetersizdir ve tahmin edilen

parametrelerdeki deęişkenlięin büyük olmasına neden olur. Bu da modelin aşırı parametrelendirilmesi (*overparametrize*) anlamına gelmektedir. Bu yöntemle elde edilen artık deęerleri $e_i = y_i - \widehat{m}(x_i) = 0$ 'dir. Çünkü ilişkisiz rasgele hataların hepsinin sıfır olması hiçbir zaman beklenmez. Bu, $\sum_{i=1}^n \{y_i - m(x_i)\}^2$ toplamını minimum yapan en küçük kareler yaklaşımının tercih edilmemesinin başka bir nedenidir. Buradaki en küçük kareler yaklaşımında eksik olan, tahmin edilen parametrelerdeki büyük deęişkenliğe bir ceza uygulanmıyor olmasıdır. Böyle bir ceza terimi en uygun şekilde pürüzlülük ile tanımlanabilir. Genel olarak pürüzlülük $\int \{m''(x)\}^2 dx$ ile ölçülmektedir. B una g öre cezalandırılmış en küçük kareler toplamı aşağıdaki gibi ifade edilir [1]:

$$\sum_{i=1}^n \{y_i - m(x_i)\}^2 + \lambda \int \{m''(x)\}^2 dx \quad (2.26)$$

Burada λ düzleştirme parametresi olarak adlandırılır ($\lambda > 0$). (2.26) kareler toplamı minimum yapılarak \widehat{m}_λ elde edilir.

(2.26) ifadesi iki kısımdan oluşmaktadır. İlk kısım, uyum eksiklięini cezalandırmaktadır. Bu da bir anlamda modelleme yanlılıęıdır. İkinci kısım, aşırı parametrelendirme ile ilgili pürüzlülüęü cezalandırmaktadır. $\lambda = 0$ için ve ri no ktaların ara deęerleri bulunarak m fonksiyonu tahmin edilir. $\lambda = +\infty$ için ise m fonksiyonunun tahmini $m(x) = \beta_0 + \beta_1 x$ şeklinde doğrusal regresyon modelidir. λ sıfırdan sonsuza doğru deęiştikçe, m fonksiyonunun tahmini en karmaşık modelden (interpolasyon) en basit modele (doğrusal model) deęişmektedir. Böylece, düzleştirilmiş eğrisel çizgi yaklaşımının model karmaşıklıęı, düzleştirme parametresi λ ile etkili bir şekilde kontrol edilir. \widehat{m}_λ , düzleştirilmiş eğrisel çizgi tahmin edicisi olarak adlandırılır [1].

(2.26) ile belirtilen cezalandırılmış en küçük kareler toplamını minimum yapan m fonksiyonunun tahmin edilmesinde $[X_{(1)}, X_{(n)}]$ aralıęındaki kübik eğrisel çizgilerden yararlanılır. $[X_{(1)}, X_{(n)}]$ aralıęında çözüm tektir. Ayrıca, \widehat{m}_λ tahmin edicisi y bağımlı deęişkeninin doğrusal bir kombinasyonudur. Buna göre,

$$\widehat{m}_\lambda(x) = n^{-1} \sum_{i=1}^n W_i(x, \lambda; x_1, \dots, x_n) y_i \quad (2.27)$$

şeklinde ifade edilebilir. Burada W_i ağırlığı, y_i bağımlı değişkeni değerlerine bağlı değildir. Çekirdek regresyon ve düzleştirilmiş eğrisel çizgi regresyonu arasındaki bağlantı teorik olarak Silverman [37, 38] tarafından geliştirilmiştir.

Silverman [37], düzleştirilmiş eğrisel çizgi yaklaşımının temelde bant genişliği değişkeni ile yerel bir çekirdek ortalaması olduğuna işaret etmiştir. Sınır değerinden uzak x_i için, n büyük olduğunda ve λ küçük olduğunda,

$$W_i(x, \lambda; x_1, \dots, x_n) \approx m(x_i)^{-1} h(x_i)^{-1} K_s \left\{ (x_i - x) / h(x_i) \right\}$$

dir. Burada $h(x_i) = \left[\lambda / \{n m(x_i)\} \right]^{1/4}$ ve $K_s(u) = 0.5 \exp(-|u|/\sqrt{2}) \sin(|u|/\sqrt{2} + \pi/4)$ 'dir.

Bu yaklaşım Messer [39] tarafından da kanıtlandığı gibi, düzleştirilmiş eğrisel çizgi tahmin edicisinin $\widehat{m}_\lambda(x)$ ortalamasını ve varyansını hesaplamak için de geçerlidir. (2.27) yaklaşımı, düzleştirilmiş eğrisel çizgi yaklaşımında bir x noktası etrafındaki yerel komşuluğa ağırlıkların nasıl atandığını anlamak için sezgisel bir bakış sağlar [1].

Düzleştirme parametresi λ , veriler kullanılarak seçilebilir. λ 'yı seçmek için diğer bir yaklaşım, çapraz geçerlilik ölçütünün minimum yapılmasıdır. Çapraz geçerlilik ölçütü,

$$CV(\lambda) = n^{-1} \sum_{i=1}^n \left\{ y_i - \widehat{m}_{\lambda, -i}(x_i) \right\}^2 \quad (2.28)$$

şeklinde ifade edilir. Burada $\widehat{m}_{\lambda, -i}$, i 'nci gözlemi çıkarıldıktan sonra elde edilen tahmin değeridir.

Çapraz geçerlilik ölçütünün sayısal olarak hesaplanması zaman alıcıdır. Bu nedenle Wahba [40] ve Craven ve Wahba [41] tarafından çapraz geçerlilik ölçütünün geliştirilmiş versiyonu "Genelleştirilmiş Çapraz Geçerlilik (GCV)" ölçütü önerilmiştir.

$$\left(\widehat{m}_\lambda(x_1), \dots, \widehat{m}_\lambda(x_n) \right)^T = H(\lambda) y$$

şeklinde de ifade edilebilir. Burada $H(\lambda)$ sadece x değişkenine bağlı olan $(n \times n)$ boyutlu şapka matrisidir ve $y = (y_1, \dots, y_n)^T$ 'dir. Buna göre GCV ölçütü,

$$GCV(\lambda) = \left[n^{-1} \text{tr} \{ I - H(\lambda) \} \right]^{-2} AKO(\lambda) \quad (2.29)$$

şeklinde yazılır. Burada $AKO(\lambda) = n^{-1} \sum_{i=1}^n \left\{ Y_i - \widehat{m}_\lambda(X_i) \right\}^2$ dir [1].

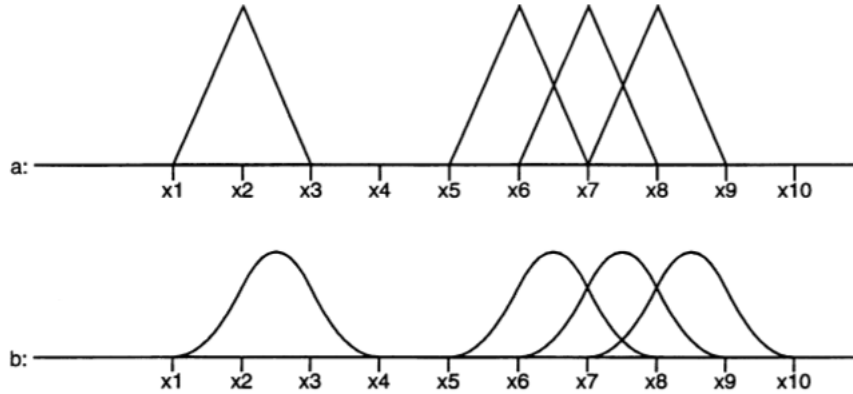
λ 'nın seçimi için literatürde başka ölçütler de önerilmiştir. Bu ölçütler, Wahba ve Wang [42]'in çalışmasında detaylı bir şekilde açıklanmıştır.

2.3.5 Cezalandırılmış Eğrisel Çizgi (Penalized Spline)

Parametrik olmayan regresyonda bilinmeyen tek değişkenli m fonksiyonu cezalandırılmış B-eğrisel çizgi ile tahmin edilebilir. B-eğrisel çizgi, düğüm noktaları olarak adlandırılan x 'in belirli değerlerinde birleştirilmiş parçalı polinomlardan oluşur. Düğüm noktaları seçildikten sonra, herhangi bir polinom derecesi için B-eğrisel çizgi yinelemeli olarak kolaylıkla elde edilir. Düğüm noktalarının seçimi birçok araştırmanın konusu olmuştur. Çünkü düğüm sayısının fazla olması verilerin gerçek fonksiyonel yapısından farklı, aşırı kestirilmesine (*overfitting*), az olması ise gerçek fonksiyonel yapısından daha yetersiz kestirilmesine (*underfitting*) neden olmaktadır.

Düğüm sayısının ve konumunun seçimi için birçok araştırmacı seçim ölçütleri önermişlerdir. Ancak, O'Sullivan [43, 44] diğer araştırmacılara göre daha çok sayıda düğüm noktasının kullanılmasını önermiştir.

B-eğrisel çizgi Şekil 2.3 ile gösterilmiştir.



Şekil 2.3. Ayrı tutulan bir B-eğrisel çizgi ve örtüşen birkaç B-eğrisel çizginin gösterimi (a) polinom derecesi 1; (b) Polinom derecesi 2 [45].

Şekil 2.3 (a) 'nın sol kısmında polinom derecesi 1 olan B-eğrisel çizgi gösterilmektedir. B-eğrisel çizgi iki doğrusal parçadan oluşmaktadır; ilk parça x_1 'den x_2 'ye, ikinci parça x_2 'den x_3 'e. Burada düğüm noktaları x_1 , x_2 ve x_3 'tür. x_1 'in solunda ve x_3 'ün sağında B-eğrisel çizgi

değeri sıfırdır. Şekil 2.3 (a) 'nın sağ kısmında polinom derecesi 1 olan ve her biri üç düğüm noktasından geçen üç tane daha B-eğrisel çizgi gösterilmektedir. Daha fazla düğüm noktası seçilerek istenilen aralıkta B-eğrisel çizgileri oluşturulabilir [45].

Şekil 2.3 (b) 'nin sol kısmında polinom derecesi 2 olan bir B-eğrisel çizgi gösterilmektedir. İki düğüm noktasında birleştirilen üç karesel parçadan oluşmaktadır. Birleştirme noktalarında, sadece polinom parçalarının ordinatları eşleşmez, aynı zamanda ilk türevleri de eşittir (ancak ikinci türevleri eşit değildir). B-eğrisel çizgiyi oluşturmak için dört komşu düğüm noktası, x_1, \dots, x_4 , seçilmiştir. Şekil 2.3 (b) 'nin sağ tarafında polinom derecesi 2 olan üç B -eğrisel çizgi daha gösterilmiştir. B-eğrisel çizgilerin birbiriyle örtüştüğü dikkate alınmalıdır. Birinci dereceden B-eğrisel çizgiler iki komşu düğüm noktasında çakışmakta, ikinci dereceden B-eğrisel çizgiler dört komşu düğüm noktasında çakışmaktadır. En soldaki ve en sağdaki eğrisel çizgilerin çakışması daha az olacaktır. Bir x değerinde, 2 tane birinci derece (veya 3 tane ikinci derece) B-eğrisel çizgi sıfırdan farklıdır [45].

Bu örnekler, polinom derecesi q olan B -eğrisel çizgi için genişletilebilir. Buna göre q dereceli bir B-eğrisel çizgi,

- Her birinin derecesi q olan $q+1$ polinom parçasından oluşur
- Polinom parçaları, q tane iç düğüm noktasında birleşir
- Birleşim noktalarında $q-1$ derecesine kadar olan türevler süreklidir
- B-eğrisel çizgi, $q + 2$ düğüm noktası komşuluğunda pozitif, diğer veri noktalarında sıfırdır
- Sınırlar hariç, düğüm noktaları komşuluğundaki $2q$ polinom parçaları çakışır
- Herhangi bir x değeri için, $q+1$ B-eğrisel çizgi değeri sıfırdan farklıdır.

x_{\min} ile x_{\max} aralığındaki veriler $n'+1$ düğüm noktası seçilerek n' eşit aralığa bölünsün. Her aralık polinom derecesi q olan $q+1$ B-eğrisel çizgi ile tahmin edilsin. B-eğrisel çizgilerin tahmin edilmesi için $n' + 2q + 1$ tane düğüm noktası seçilmelidir. Regresyon modelinde B-eğrisel çizgi sayısı $n = n' + q$ 'dir. Bu, Şekil 2.3' teki gibi grafikler oluşturarak kolayca görülebilir [45].

B-eğrisel çizgiler parametrik olmayan tek değişkenli regresyonda temel fonksiyon olarak kullanılabilir. Üçüncü dereceden B-eğrisel çizgilerin doğrusal bileşimi düzgün bir eğri (*smooth curve*) oluşturur. B-eğrisel çizgi kullanılarak regresyon modelinin tahmini, polinom regresyon modelinin tahmininden daha zor değildir. De Boor [46], düşük dereceli B-eğrisel çizgilerden herhangi bir dereceli B -eğrisel çizgilerin hesaplaması için bir algoritma

önermiştir. Sıfır dereceli B-eğrisel çizgi, iki düğüm noktası arasındaki bir aralıkta sabit olduğu için, herhangi bir derecedeki B-eğrisel çizgileri tahmin etmek kolaydır.

Burada düğümlerin eşit mesafede yerleştirildiği durum için algoritma verilmiştir. Ancak De Boor [46]' un algoritması düğüm noktalarının farklı şekilde yerleştirildiği durumlar için de kolaylıkla uygulanabilir.

$B_j(x; q)$, eşit mesafede yerleştirilen düğüm noktaları için q dereceli j 'inci B-eğrisel çizginin herhangi bir x noktasındaki değeridir. \hat{y} , (x_i, y_i) veri çiftinin kestirilen doğrusal bir kombinasyonudur. Buna göre,

$$\hat{y}(x) = \sum_{j=1}^n \hat{\gamma}_j B_j(x; q)$$

şeklinde ifade edilir. Bazı çalışmalarda $B_j(x; q)$ yerine $B_j(x)$ gösterimi kullanılmaktadır. Burada kolaylık sağlama açısından $B_j(x)$ gösterimi seçilmiştir. De Boor [46], B-eğrisel çizgilerin türevleri için aşağıdaki formülü vermiştir:

$$h \sum_j \gamma_j B_j'(x; q) = \sum_j \gamma_j B_j(x; q-1) - \sum_j \gamma_{j+1} B_{j+1}(x; q-1) = -\sum_j \Delta \gamma_{j+1} B_j(x; q-1) \quad (2.30)$$

Burada h , düğüm noktaları arasındaki uzaklıktır ve $\Delta \gamma_j = \gamma_j - \gamma_{j-1}$ 'dir. İkinci türev için

$$h^2 \sum_j \gamma_j B_j''(x; q) = \sum_j \Delta^2 \gamma_j B_j(x; q-2) \quad (2.31)$$

eşitliği kullanılır. Burada $\Delta^2 \gamma_j = \Delta \Delta \gamma_j = \gamma_j - 2\gamma_{j-1} + \gamma_{j-2}$ 'dir [45].

p tane B-eğrisel çizginin, $B_j(\cdot)$ oluşturduğu n tane (x_i, y_i) veri noktasına ilişkin regresyon denklemi ele alındığında bu regresyona ilişkin en küçük kareler toplamı,

$$S = \sum_{i=1}^n \left\{ y_i - \sum_{j=1}^p \gamma_j B_j(x_i) \right\}^2 \quad (2.32)$$

şeklinde ifade edilir. Düğüm sayısı fazla seçildiğinde, kestirilen eğri verinin gerçek fonksiyonel yapısından daha fazla değişkenlik gösterecektir. Kestirim esnekliğini azaltmak için O'Sullivan [43, 44] kestirilen eğrinin ikinci türevi üzerine bir ceza uygulamıştır ve cezalandırılmış en küçük kareler toplamını,

$$S = \sum_{i=1}^n \left\{ y_i - \sum_{j=1}^p \gamma_j B_j(x_i) \right\}^2 + \lambda \int_{x_{\min}}^{x_{\max}} \left\{ \sum_{j=1}^p \gamma_j B_j''(x) \right\}^2 dx \quad (2.33)$$

şeklinde ifade etmiştir. Kestirilen fonksiyonun ikinci türevinin karesinin integrali, Reinsch [47] 'in düzleştirilmiş eğrisel çizgi çalışmaları sonrasında “pürüzsüzlük cezası” olarak ifade edilmiştir.

İkinci türev ile ilgili özel bir durum yoktur. Daha düşük veya daha yüksek mertebeden türevler kullanılabilir. Düzleştirilmiş eğrisel çizgiler dikkate alındığında, ilk türev basit denklemlere ve parçalı doğrusal bir kestirime yol açarken; yüksek mertebeden türevler büyük bant genişlikli oldukça karmaşık denklemlere ve çok düz bir kestirime neden olur. Ceza terimi olarak, komşu B-eğrisel çizgilerin katsayılarının sonlu farklarının (yüksek mertebeli) da alınabileceği önerilmiştir. Buna göre cezalandırılmış en küçük kareler toplamı,

$$S = \sum_{i=1}^n \left\{ y_i - \sum_{j=1}^p \gamma_j B_j(x_i) \right\}^2 + \lambda \sum_{j=k+1}^p (\Delta^k \gamma_j)^2 \quad (2.34)$$

şeklinde ifade edilir. Burada $\{B_j(x)\}_{j=1}^p$ q dereceli B-eğrisel çizgi temeli, $p = q + d + 1$ ve d

düğüm sayısıdır. Δ^k k'inci fark operatörü ve $\sum_{j=k+1}^p (\Delta^k \gamma_j)^2$ değeri $\int_0^1 [m^{(k)}(x)]^2 dx$ 'in

yaklaşık değeridir [45]. (2.34)'ün ilk terimi uyum iyiliğini ölçmekte, ikinci terim ise m fonksiyonunun pürüzlülüğünü cezalandırmaktadır. Düzleştirme parametresi λ iki am arasında dengiyi kontrol etmektedir. Genellikle $q = 3$ ve $k = 2$ olarak alınır. Böylece elde edilen tahmin $m(\cdot)$, ikinci dereceden sürekli türeve sahip olur.

(2.34)'teki cezalandırılmış en küçük kareler toplamı matris formunda

$$(\mathbf{y} - \mathbf{B}\boldsymbol{\gamma})^T (\mathbf{y} - \mathbf{B}\boldsymbol{\gamma}) + p\lambda \boldsymbol{\gamma}^T \mathbf{D}\boldsymbol{\gamma} \quad (2.35)$$

şeklinde ifade edilir. (2.35)'teki cezalandırılmış en küçük kareler toplamını minimum yapan $\hat{\boldsymbol{\gamma}}$,

$$\hat{\boldsymbol{\gamma}} = (\mathbf{B}^T \mathbf{B} + \lambda \mathbf{D}_k^T \mathbf{D}_k)^{-1} \mathbf{B}^T \mathbf{y}$$

dir. Burada \mathbf{D}_k fark operatörü Δ^k 'nin matris gösterimidir ve \mathbf{B} 'nin elemanları $b_{ij} = B_j(x_i)$ 'dir. $\lambda = 0$ olduğunda B-eğrisel çizgi fonksiyonunun standart normal dağılımlı doğrusal regresyon denklemleri elde edilir. $k = 0$ ise ridge regresyonun özel bir modeli elde edilir.

2.3.5.1 Düzleştirme Parametresinin Seçimi

Düzleştirme parametresinin seçimi için yaygın olarak kullanılan ölçütler çapraz geçerlilik ölçütü ve genelleştirilmiş çapraz geçerlilik ölçütüdür. B-eğrisel çizgi ile kestirilen modelde düzleştirme parametresinin seçimi için çapraz geçerlilik ölçütü [45],

$$CV(\lambda) = \sum_{i=1}^n \left(\frac{y_i - \hat{y}_i}{1 - h_{ii}} \right)$$

şeklinde tanımlanır. Burada h_{ii} , $\mathbf{H} = \mathbf{B}(\mathbf{B}^T \mathbf{B} + \lambda \mathbf{D}_k^T \mathbf{D}_k)^{-1} \mathbf{B}^T$ şapka matrisinin köşegen elemanlarıdır.

Genelleştirilmiş çapraz geçerlilik ölçütü ise,

$$GCV(\lambda) = \sum_{i=1}^n \frac{(y_i - \hat{y}_i)^2}{\left(n - \sum_{i=1}^n h_{ii} \right)} \quad (2.36)$$

şeklinde tanımlanır. Her iki ölçüt arasındaki fark genellikle küçüktür. En iyi λ , çapraz geçerlilik veya genelleştirilmiş çapraz geçerliliği minimum yapan değerdir.

Düzeltilmiş çapraz geçerlilik (\overline{CV}) ve düzeltilmiş genelleştirilmiş çapraz geçerlilik (\overline{GCV}) ölçütleri ise,

$$\overline{CV}(\lambda) = \sqrt{CV(\lambda) / n}$$

$$\overline{GCV}(\lambda) = \sqrt{n GCV(\lambda)} \quad (2.37)$$

şeklinde ifade edilir.

3. DEĞİŞEN KATSAYILI MODEL

y bağımlı değişken vektörü ve $\mathbf{X} = (1, x_2, x_3, \dots, x_p)$ bağımsız değişkenler matrisi olmak üzere doğrusal regresyon modeli

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (3.1)$$

şeklinde yazılır. (3.1)'de $\mathbf{X} = (1, x_2, x_3, \dots, x_p)$ $n \times p$ boyutlu tasarım matrisi, $\boldsymbol{\beta} = (\beta_0, \dots, \beta_p)$ $p \times 1$ boyutlu bilinmeyen parametre vektörü ve $\boldsymbol{\varepsilon}$ $n \times 1$ boyutlu hata terimleri vektörüdür. Ayrıca $E(\boldsymbol{\varepsilon}|\mathbf{X}) = 0$ olduğu varsayılır.

(3.1)'de $\boldsymbol{\beta}$ katsayılar vektörünün sabit olduğu varsayılır. Bazı durumlarda bu varsayım doğrusal regresyon modellerinin uygulanmasını sınırlandırabilir. Örneğin; uzun ke sit verilerin analizinde, finansal piyasa araştırmalarında, ekonomi ve bazı ekolojik araştırmalarda $\boldsymbol{\beta}$, bazı bağımsız değişkenlere göre değişebilir. Bu değişkenler genellikle zaman, sıcaklık veya coğrafi konumdur.

Doğrusal regresyon modellerinin esnekliğini artırmak ve doğrusal regresyon modellerinin uygulama alanlarını genişletmek amacıyla, $\boldsymbol{\beta}$ katsayılarının bazı bağımsız değişkenlerinin düzgün fonksiyonları olarak ifade edildiği değişken katsayılı modelleri kullanılır. Değişken katsayılı model [18]

$$y_i = \beta_0 + \beta_1(u)x_{i1} + \beta_2(u)x_{i2} + \beta_3(u)x_{i3} + \dots + \beta_p(u)x_{ip} + \varepsilon_i \quad (3.2)$$

şeklinde ifade edilir. (3.2)'deki model matris-vektör gösterimi ile

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta}(u) + \boldsymbol{\varepsilon} \quad (3.3)$$

şeklinde yazılır. Burada $E(\boldsymbol{\varepsilon}|\mathbf{X}, u) = 0$ 'dır.

(3.3)'te $\beta_0, \beta_1, \beta_2, \dots, \beta_p$ katsayıları u bağımsız değişkeninin düzgün ve belirsiz fonksiyonlarıdır.

$\boldsymbol{\beta}$ katsayılarının bazı bağımsız değişkenlerinin çok değişkenli düzgün fonksiyonları olarak ifade edildiği değişken katsayılı modeli ise

$$y_i = \beta_0 + \beta_1(\mathbf{u}_1)x_{i1} + \beta_2(\mathbf{u}_2)x_{i2} + \beta_3(\mathbf{u}_3)x_{i3} + \dots + \beta_p(\mathbf{u}_p)x_{ip} + \varepsilon_i \quad (3.4)$$

şeklinde yazılır. Burada $\mathbf{u}_i = (u_{i1}, u_{i2}, \dots, u_{ip})^T$ 'dır.

Yukarıda bahsedilen değişken katsayılı modellerde u ve X_1, X_2, \dots, X_p değişkenleri arasında etkileşim vardır. u 'nun farklı değerlerinde, y ile x değişkenleri arasındaki ilişki bu değerlere bağlı olarak değişir. Bu nedenle katsayı fonksiyonları regresyon denklemindeki gibi yorumlanabilir [18].

Uzun kesit verilerin olduğu çalışmalarda n tane gözlem olduğunda, i 'nci gözlem için veri kümesi $\{y_{i1}(t), x_{i1}(t), x_{i2}(t), \dots, x_{ip}(t)\}$, $t = t_{ij}$, $j = 1, 2, \dots, n_j$ biçiminde olup zamanla değişen katsayılı bir model

$$y_i(t) = \beta_1(t)x_{i1} + \beta_2(t)x_{i2} + \beta_3(t)x_{i3} + \dots + \beta_p(t)x_{ip} + \varepsilon_i(t) \quad (3.5)$$

şeklinde tanımlanır. Bu çalışmalarda zaman içinde değişkenlerin hangilerinin bağımlı değişkeni etkilediği araştırılmıştır.

(3.2)'deki model, literatürde en çok bilinen ve çalışılmış olan değişken katsayılı modeldir. Bu modelde $\beta_0, \beta_1, \beta_2, \dots, \beta_p$ katsayıları ortak bir u bağımsız değişkeninin fonksiyonu olarak ifade edilir. Örneğin, zamanla değişken katsayılı modelde u bağımsız değişkeni zaman; sıcaklık ile değişken katsayılı modelde u bağımsız değişkeni sıcaklıktır. Daha detaylı bilgi için [6, 10, 8] çalışmaları incelenebilir.

(3.2)'deki modelde $\beta_0, \beta_1, \beta_2, \dots, \beta_p$ katsayılarının her biri u_1, u_2, \dots, u_p değişkenlerine bağlı olduğunda (3.4)'teki model elde edilir. Örneğin, sırasıyla X_1 ve X_2 değişkenleri için t sıcaklık ve r radyasyon olsun. (3.4)'teki $\beta_0, \beta_1, \beta_2, \dots, \beta_p$ katsayıları, $\mathbf{u}_i = (u_{i1}, u_{i2}, \dots, u_{ip})^T$, $i=1, 2, \dots, p$, değişkenler vektörünün bir fonksiyonudur. Değişken katsayılı modellerde mekân ve ya coğrafi konum da değişken katsayı olarak alınır.

Bu tez çalışmasında, u 'nun tek değişken ve $\beta_0, \beta_1, \beta_2, \dots, \beta_p$ katsayıları için sabit olduğu (3.2)'deki model ele alınmıştır. Bu modelin bir avantajı, doğrusal modele benzer şekilde katsayıların yorumlanabilmesidir.

3.1 Literatür

Değişen katsayılı regresyon modelleri; genelleştirilmiş doğrusal modeller, zaman serileri analizi ve uzun kesit veri analizi gibi çok boyutlu parametrik olmayan regresyon modelleri olarak ifade edilebilir.

Fan ve Zhang (1999), solunum ve dolaşım problemi yaşayan hasta sayısı değişkenini, kükürt dioksit (mg / m^3), a zot di oksit (mg / m^3), t oz (mg / m^3) değişkenlerinden hangilerinin etkilediğini indeks değişken zaman olmak üzere değişen katsayılı model ile incelemiştir [48].

Huang ve diğerleri (2004), HIV virüslü hastaların sigara içme durumu (içiyor/içmiyor), HIV virüsü geçirdiği yaş ve HIV enfeksiyonu öncesi CD4 hücre yüzdesi değişkenlerinin indeks değişken zaman o lmak ü zere enfeksiyon sonrasındaki C D4 yüz desini etkileyip etkilemediğini değişen katsayılı model ile analiz etmişlerdir [49].

Huang (2004) çalışmasında; şehirlerdeki kişi başına suç oranı ile konut başına ortalama oda sayısı, 10000\$ başına tam değerli emlak vergisi oranı, 10 milyonda nitrik oksit konsantrasyonu, şehirlere ilişkin öğrenci başına düşen öğretmen oranı, 1940 yılından önce inşa edilmiş sahipli evlerin oranı değişkenleri ile 1970 yılı Boston bölgesindeki nüfus sayımına göre sahipli evlerin medyan değeri arasındaki ilişkileri görmek için değişen katsayılı modeli kullanmıştır [50].

Wang (2007), bir ekosistem için kullanılabilir fotosentetik aktif radyasyon değişkeni ile birlikte indeks değişkeni sıcaklık olarak ele alarak CO_2 'nin net ekosistem değişiminine etkisini değişen katsayılı model ile incelemiştir [18].

Sterrantino (2013), yakıt-hava karışımının bir ölçüsü olan “eşitlik oranı (E)”, “motorun sıkıştırma oranı (C)” değişkenlerinin, motor egzozunda nitrik oksit ve nitrojen dioksit konsantrasyonu olan “ NO_x ” değişkenini etkileyip etkilemediğini değişen katsayılı modeli yeniden parametrelendirerek ele almıştır [51].

Park ve diğerleri (2015), koroner kalp hastalığı öyküsünü; koroner kalp hastalığı aile öyküsü, sistolik (max) kan basıncı, vb. değişkenlerinin etkileyip etkilemediğini yaş indeks değişkeniyle birlikte genelleştirilmiş değişen katsayılı modeller ile incelemiştir [52].

Assaf ve Campostrini (2015), zaman indeks değişken olmak üzere sigara içme durumunu, yaş, cinsiyet, medeni hal, eğitim düzeyi, gelir düzeyi, çalışma durumu, sosyo-ekonomik ve sosyo-demografik düzeyi gösteren bölge değişkeni, alkol tüketimi, fiziksel aktivite, depresyon durumu gibi değişkenlerin ne ölçüde etkilediğini genelleştirilmiş değişen

katsayılı model ile araştırmışlardır [53]. Çizelge 3.1’de değişen katsayılı modeller ile ilgili son yıllarda yapılan çalışmalar özetlenmiştir:

Çizelge 3.1. Literatür taraması özeti

	Göstergeler	İndeks Değişken	Bağımlı Değişken	Ülke
Fan ve Zhang (1999)	<ul style="list-style-type: none"> • Kükürt dioksit (mg / m³) • Azot dioksit (mg / m³) • Toz (mg / m³) 	Zaman	Solunum ve dolaşım problemi yaşayan hasta sayısı	Çin
Huang, Wu ve Zhou (2004)	<ul style="list-style-type: none"> • Sigara içme durumu (içiyor/içmiyor) • HIV enfeksiyon geçirdiği yaş • HIV enfeksiyon öncesi CD4 hücre yüzdesi 	Zaman	Enfeksiyon sonrası belirli bir zamandaki CD4 yüzdesi	USA
Huang (2004)	<ul style="list-style-type: none"> • Şehirlerdeki kişi başına suç oranı • Konut başına ortalama oda sayısı • 10000 \$ başına tam değerli emlak vergisi oranı • 10 milyonda nitrik oksit konsantrasyonu • Şehirlere ilişkin öğrenci başına düşen öğretmen oranı • 1940 yılından önce inşa edilmiş sahipli evlerin oranı 	Nüfusun düşük olduğu bölgeler	1970 yılı Boston bölgesindeki nüfus sayımına göre sahipli evlerin medyan değeri	USA
Wang (2007)	<ul style="list-style-type: none"> • Bir ekosistem için kullanılabilir fotosentetik aktif radyasyon 	Sıcaklık	CO ₂ 'nin net ekosistem değişimi	USA
Cai, Fan ve Li (2000)	<ul style="list-style-type: none"> • Cinsiyet • log (yanık alanı+1) • Oksijen 	Yaş	Yanıklarından kurtulanlar / kurtulamayanlar	USA
Sterrantino (2013)	<ul style="list-style-type: none"> • Motorun sıkıştırma oranı 	Eşitlik oranı	Motor egzozunda nitrik oksit ve nitrojen dioksit konsantrasyonu	İtalya
Park, Mammen, Lee ve Lee (2015)	<ul style="list-style-type: none"> • Koroner kalp hastalığı aile öyküsü • Psiko-sosyal stresin bir ölçütü olan Tip-A davranışı • Sistolik (max) kan basıncı • Düşük yoğunluklu lipoprotein kolesterol 	Yaş	Koroner kalp hastalığı öyküsü	Güney Afrika
Assaf and Campostrini (2015)	<ul style="list-style-type: none"> • Yaş • Cinsiyet • Medeni hal • Eğitim düzeyi • Gelir düzeyi (veya ekonomik sıkıntılar) • Çalışma durumu • Sosyo-ekonomik ve sosyo-demografik • Değişkenleri temsil eden bölge • Alkol tüketimi • Fiziksel aktivite • Depresyon durumunun sağlık riski ve davranış değişkenlerini temsil etmesi 	Zaman	Sigara içme durumu	İtalya

3.2 Değişen Katsayılı Modellerde Tahmin Yöntemleri

Değişen katsayılı modellerde parametreler

- Yerel (*Local*) polinom düzleştiricisi,
- Düzleştirilmiş eğrisel çizgi (*Smoothing spline*),
- Cezalandırılmış eğrisel çizgiler (*Penalized spline*),

yöntemleri ile tahmin edilebilir.

Bu yöntemlerden bazılarının zayıf yönleri vardır. Örneğin; düzleştirilmiş eğrisel çizgi (*smoothing spline*) düzleştiricisinde birden çok düzleştirme parametresi seçmek zordur. Çekirdek ve yerel doğrusal düzleştiriciler için tüm katsayı fonksiyonlarının, $\beta_j(u)$ $j=1, \dots, p$, aynı düzgünlük (*smoothness*) derecesine sahip olduğu varsayılır. Bu varsayım katsayı fonksiyonlarının u değişkeninin aynı aralığında eşit olarak tahmin edilmesini sağlar. Değişken katsayılı modelde çekirdek veya yerel doğrusal düzleştiriciler “tek adım yöntemi” olarak d a i simlendirilir. Uygulamada düzgünlük derecesinin aynı olması varsayımının sağlanması zordur ve farklı katsayı fonksiyonları genellikle farklı düzgünlük derecelerine sahiptir. Bu nedenle Fan ve Zhang [48], farklı katsayı fonksiyonlarının farklı düzgünlük derecelerine sahip olduğu varsayımına dayanan iki adım yöntemini önermişlerdir [54]. Teorik olarak, değişken katsayı fonksiyonlarının cezalandırılmış eğrisel çizgi tahminleri, birden çok düzleştirme parametresinin kullanılmasından dolayı katsayı fonksiyonları farklı düzgünlük derecelerine sahiptir. Lu ve diğerleri [55] önerdikleri yaklaşımda sabit ve çok sayıda düğüm noktası seçerek, değişken katsayılı modelde cezalandırılmış eğrisel çizgi (*penalized spline*) tahminlerinin tutarlı ve asimptotik olarak normal olduğunu göstermişlerdir. Bu çalışmada ayrıca çok sayıda düzleştirme parametresinin seçimi için sistematik bir optimizasyon algoritması da geliştirilmiştir.

3.2.1 Yerel (Local) Polinom Düzleştiricisi

(3.2) modelindeki $\beta_j(u)$, $j=1, 2, \dots, p$, katsayı fonksiyonlarını tahmin etmek için kullanılacak yöntemlerden birisi yerel doğrusal regresyon düzleştiricisidir. Verilen u_0 noktası için yerel ve doğrusal olarak $\beta_j(u)$, $j=1, 2, \dots, p$, katsayı fonksiyonları tahmin edilir.

u_0 komşuluğundaki u için Taylor açılımı kullanılarak $\beta_j(u)$ fonksiyonu (2.9)'daki eşitliğe benzer şekilde yaklaşık olarak

$$\beta_j(u) \approx \beta_j(u_0) + \beta_j^{(1)}(u_0)(u - u_0) \equiv a_j + b_j(u - u_0) \quad (3.6)$$

şeklinde yazılır. (3.6)'da $b_j = \beta_j^{(1)}(u_0)$ 'dır. $u = u_0$ noktasında yerel olarak ağırlıklandırılmış en küçük kareler toplamı (2.11)'deki eşitlikte olduğu gibi

$$\sum_{i=1}^n \left[y_i - \sum_{j=1}^p x_{ij} \{ a_j + b_j(u_i - u_0) \} \right]^2 K_h(u_i - u_0) \quad (3.7)$$

şeklinde ifade edilir [18]. Yerel en küçük kareler tahmin edicisini elde etmek için (3.7)'deki en küçük kareler toplamını minimum yapan $\mathbf{a} = (a_1, a_2, a_3, \dots, a_p)^T$ ve $\mathbf{b} = (b_1, b_2, b_3, \dots, b_p)^T$ vektörleri bulunur. Buna göre her u_0 için

$$\begin{aligned} (\hat{\mathbf{a}}, \hat{\mathbf{b}})^T &\equiv (\hat{a}_1, \hat{a}_2, \hat{a}_3, \dots, \hat{a}_p, \hat{b}_1, \hat{b}_2, \hat{b}_3, \dots, \hat{b}_p)^T \\ &\equiv (\hat{\beta}_1(u_0), \hat{\beta}_2(u_0), \dots, \hat{\beta}_p(u_0), \hat{\beta}_1^{(1)}(u_0), \hat{\beta}_2^{(1)}(u_0), \dots, \hat{\beta}_p^{(1)}(u_0))^T \end{aligned}$$

olarak ifade edilir. Burada $\hat{\mathbf{a}} \equiv (\hat{a}_1, \hat{a}_2, \hat{a}_3, \dots, \hat{a}_p)^T \equiv (\hat{\beta}_1(u_0), \hat{\beta}_2(u_0), \dots, \hat{\beta}_p(u_0))^T$ 'dir. Bu yöntem tek adım tahmin yöntemidir.

İki adımlı tahmin yönteminde; genel varsayımlar değişmeksizin, katsayı fonksiyonunun, $\beta_p(u)$, diğer katsayı fonksiyonlarından daha düzgün olduğu ve $\beta_p(u)$ 'nin dördüncü dereceden bir türeve sahip olduğu varsayılır. Bu durumda $\beta_p(u)$ 'nin kübik fonksiyonu kullanılabilir. $\beta_p(u)$ 'nin kübik fonksiyonu,

$$\begin{aligned} \beta_p(u) &\approx \beta_p(u_0) + \beta_p^{(1)}(u_0)(u - u_0) + \beta_p^{(2)}(u_0)(u - u_0)^2 + \beta_p^{(3)}(u_0)(u - u_0)^3 \\ &\equiv a_p + b_p(u - u_0) + c_p(u - u_0)^2 + d_p(u - u_0)^3 \end{aligned} \quad (3.8)$$

şeklinde yazılır. u_0 komşuluğunda yerel en küçük kareler toplamı

$$\sum_{i=1}^n \left[y_i - \sum_{j=1}^{p-1} \{ a_j + b_j(u_i - u_0) \} x_{ij} - \{ a_p + b_p(u - u_0) + c_p(u - u_0)^2 + d_p(u - u_0)^3 \} x_{ip} \right]^2 K_{h_1}(u_i - u_0)$$

şeklinde yazılır [36]. Burada $\hat{a}_j, \hat{b}_j, j = 1, 2, 3, \dots, p-1$ ve $\hat{a}_p, \hat{b}_p, \hat{c}_p, \hat{d}_p$ yerel en küçük kareler fonksiyonunu minimum yapan tahmin değerleridir. $\hat{\beta}_{p,TA}(u_0) \equiv \hat{a}_p, \hat{\beta}_p(u_0)$ 'ın tek adım tahmin edicisi olarak adlandırılır.

İki adım tahmin yönteminin ilk adımında [18],

$$\sum_{i=1}^n \left[y_i - \sum_{j=1}^p \{ a_j + b_j(u_i - u_0) \} x_{ji} \right]^2 K_{h_0}(u_i - u_0) \quad (3.9)$$

şeklinde ifade edilen yerel en küçük kareler toplamı, verilen h_0 başlangıç bant genişliği için minimum yapılarak $\beta_j(u)$, $j=1,2,\dots,p$, katsayı fonksiyonunun başlangıç tahminleri bulunur. Bu tahminler $\hat{\beta}_1(u_0), \hat{\beta}_2(u_0), \dots, \hat{\beta}_p(u_0)$ olarak belirtilir. İkinci adımda, (3.9)'daki en küçük kareler toplamında ilk adımda bulunan $\hat{\beta}_1(\cdot), \hat{\beta}_2(\cdot), \dots, \hat{\beta}_{p-1}(\cdot)$ değerleri yerine konulursa en küçük kareler toplamı

$$\sum_{i=1}^n \left[y_i - \sum_{j=1}^{p-1} \hat{\beta}_j(u_0) x_{ij} - \{a_p + b_p(u - u_0) + c_p(u - u_0)^2 + d_p(u - u_0)^3\} x_{ip} \right]^2 K_{h_2}(u_i - u_0)$$

şeklinde elde edilir. Bu en küçük kareler toplamı a_p, b_p, c_p, d_p 'ye göre minimize edilir.

Burada h_2 bant genişliği kullanılmıştır. İkinci adım, iki adım tahmin edicisi $\hat{\beta}_p(\cdot)$ 'nin yerel en küçük kareler kübik tahmini olarak düşünülebilir. $\hat{\beta}_{p,IA}(u_0) \equiv \hat{a}_p, \hat{\beta}_p(u_0)$ 'ın iki adım tahmin edicisidir. Yerel polinom düzleştiricisinde iki adım tahmin yöntemi matris gösterimleri ile aşağıdaki gibi özetlenebilir:

$$\mathbf{Y} = [y_1 \ y_2 \ \dots \ y_n]^T,$$

$$\mathbf{X}_0 = \begin{pmatrix} x_{11} & (u_1 - u_0)x_{11} & x_{12} & (u_1 - u_0)x_{12} & \dots & x_{1p} & (u_1 - u_0)x_{1p} \\ x_{21} & (u_2 - u_0)x_{21} & x_{22} & (u_2 - u_0)x_{22} & \dots & x_{2p} & (u_2 - u_0)x_{2p} \\ \cdot & \cdot & \cdot & \cdot & \dots & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \dots & \cdot & \cdot \\ x_{n1} & (u_n - u_0)x_{n1} & x_{n2} & (u_n - u_0)x_{n2} & \dots & x_{np} & (u_n - u_0)x_{np} \end{pmatrix},$$

$$\mathbf{W}_0 = \text{diag} \{ K_{h_0}(u_1 - u_0), K_{h_0}(u_2 - u_0), \dots, K_{h_0}(u_n - u_0) \}$$

olmak üzere

$$\sum_{i=1}^n \left[y_i - \sum_{j=0}^{p-1} \{a_j + b_j(u_i - u_0)\} x_{ji} \right]^2 K_{h_0}(u_i - u_0)$$

şeklindeki yerel en küçük kareler toplamı minimum yapıldığında

$\hat{\beta}_{j,0}(u_0) = \mathbf{e}_{2j-1,2(p-1)}^T (\mathbf{X}_0^T \mathbf{W}_0 \mathbf{X}_0)^{-1} \mathbf{X}_0^T \mathbf{W}_0 \mathbf{y}$ olarak elde edilir. Burada $u = u_0$, $j=0,1,2,\dots,p-1$ ve

$\mathbf{e}_{k,m}$, m uzunluğuna sahip ve k 'nci değeri 1 olan birim vektördür. Yerel en küçük kareler

toplamı,

$$\sum_{i=1}^n \left[y_i - \sum_{j=0}^{p-2} \{a_j + b_j(u_i - u_0)\} x_{ji} - \{a_{p-1} + b_{p-1}(u - u_0) + c_{p-1}(u - u_0)^2 + d_{p-1}(u - u_0)^3\} x_{i,p-1} \right]^2 K_{h_1}(u_i - u_0)$$

şeklinde ifade edilirse

$$\mathbf{X}_2 = \begin{pmatrix} x_{1,p} & (u_1 - u_0)x_{1p} & (u_1 - u_0)^2 x_{1p} & (u_1 - u_0)^3 x_{1p} \\ x_{2,p} & (u_2 - u_0)x_{2p} & (u_2 - u_0)^2 x_{2p} & (u_2 - u_0)^3 x_{2p} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ x_{n,p} & (u_n - u_0)x_{np} & (u_n - u_0)^2 x_{np} & (u_n - u_0)^3 x_{np} \end{pmatrix}$$

ve

$$\mathbf{X}_3 = \begin{pmatrix} x_{11} & (u_1 - u_0)x_{11} & x_{12} & (u_1 - u_0)x_{12} & \cdots & x_{1,p-1} & (u_1 - u_0)x_{1,p-1} \\ x_{21} & (u_2 - u_0)x_{21} & x_{22} & (u_2 - u_0)x_{22} & \cdots & x_{2,p-1} & (u_2 - u_0)x_{2,p-1} \\ \cdot & \cdot & \cdot & \cdot & \cdots & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdots & \cdot & \cdot \\ x_{n1} & (u_n - u_0)x_{n1} & x_{n2} & (u_n - u_0)x_{n2} & \cdots & x_{n,p-1} & (u_n - u_0)x_{n,p-1} \end{pmatrix}$$

$$\mathbf{X}_1 = (\mathbf{X}_2, \mathbf{X}_3) \text{ ve } \mathbf{W}_1 = \text{diag} \{K_{h_1}(u_1 - u_0), K_{h_1}(u_2 - u_0), \dots, K_{h_1}(u_n - u_0)\}$$

dir. Yukarıda belirtilen yerel en küçük kareler toplamı minimum yapılırsa,

$$\hat{\beta}_{p-1,1}(u_0) = \mathbf{e}_{2p-2,2p}^T (\mathbf{X}_1^T \mathbf{W}_1 \mathbf{X}_1)^{-1} \mathbf{X}_1^T \mathbf{W}_1 \mathbf{y}$$

olarak elde edilir. $u = u_0$ için iki adım tahmin edicisi,

$$\hat{\beta}_{p-1,2}(u_0) = (1, 0, 0, 0) (\mathbf{X}_2^T \mathbf{W}_2 \mathbf{X}_2)^{-1} \mathbf{X}_2^T \mathbf{W}_2 (\mathbf{y} - \mathbf{V})$$

olarak elde edilir. Burada

$$\mathbf{W}_2 = \text{diag} \{K_{h_2}(u_1 - u_0), K_{h_2}(u_2 - u_0), K_{h_2}(u_3 - u_0), \dots, K_{h_2}(u_n - u_0)\},$$

$$\mathbf{V} = (v_1, v_2, \dots, v_n)^T \text{ ve } v_i = \sum_{j=1}^{p-2} \hat{\beta}_{j,0}(u_i) x_{ij} \text{ 'dir.}$$

İki adım tahmin edicisi $\hat{\beta}_{p-1,2}(\cdot)$, Y_1, Y_2, \dots, Y_n için ağırlıklandırılmış bir toplam olduğu için verilen h_0 ve h_2 bant genişliğinde doğrusal bir tahmin edicidir.

3.2.1.1 Yerel Polinom Yaklaşımında Bant Genişliğinin Seçimi

Yerel polinom yaklaşımlarında, h bant genişliğinin seçimi önemlidir. Literatürde bant genişliğinin seçimi için birçok ölçüt önerilmiştir. Bu ölçütlerden en yaygın kullanılanları çapraz geçerlilik ölçütü ve genelleştirilmiş çapraz geçerlilik ölçütüdür.

Değişken katsayılı modellerde yerel polinom düzleştiricisi kullanıldığında h bant genişliğinin seçimi önemlidir. Veri ilk olarak d_j ile gösterilen veri noktalarını içeren q tane gruba bölünür. d_j ,

$$d_j = \{q_k + j, k = 1, 2, 3, \dots\}, \quad j = 0, 1, 2, \dots, q-1$$

olarak tanımlanır. j. grup silindikten sonra geriye kalan veriler kullanılarak j'nci kestirim değeri elde edilir. $\hat{y}_{-d_j}(u_i, x_i)$, j. grup silindikten sonra elde edilen kestirim değerini göstermektedir [18].

Fonksiyonlar farklı düzgünlük derecelerine sahipken tek bir bant genişliği ile bütün fonksiyonlar aynı anda tahmin edilemez. Bu durumda h'nin seçimi analizlerde oldukça önem taşımaktadır. Değişken katsayılı modellerde çapraz geçerlilik ölçütü:

$$CV(h) = \sum_{j=0}^{q-1} \sum_{i \in d_j} \left[y_i - \hat{y}_{-d_j}(u_i, x_i) \right]^2 \quad (3.10)$$

şeklinde ifade edilir. CV(h) değerini minimum yapan h değeri seçilir. Genelleştirilmiş değişken katsayılı modellerde uyum iyiliği kriteri, sapma veya Pearson artık kareler toplamından hesaplanmaktadır. Ardından, iki katlı çapraz geçerlilik ölçütü gelmektedir:

Uygulamalarda q genellikle 20 olarak seçilir. Çapraz geçerlilik ölçütü, genellikle verilerin bölümlere ayrılmasına çok hassas değildir [4].

Zhang ve Lee [56], değişken katsayılı modellerin yerel polinom kestirimine dayanan değişken bant genişliği seçim ölçütü önermişlerdir. Ayrıca, bilinmeyen katsayı fonksiyonları mekânsal homojen ise, sabit bir bant genişliğinin yeterli olduğu iddiasını tartışmışlar buna karşın karmaşık katsayı fonksiyonlarının olduğu değişken katsayılı modellerini tahmin etmek için değişken bant genişliğinin gerekliliği olduğunu vurgulamışlardır.

3.2.2 Düzleştirilmiş Eğrisel Çizgiler

Değişen katsayılı modelde bilinmeyen katsayı fonksiyonlarını tahmin etmek için kullanılan bir diğer yöntem Hastie ve Tibshirani [6] tarafından önerilen cezalandırılmış en küçük kareler yöntemidir.

x_{ij} ve u_{ij} , x_j ve u_j 'nin i 'nci gözlem değerleri olmak üzere değişen katsayılı modelin

$$y_i = x_{i1}\beta_1(u_{i1}) + x_{i2}\beta_2(u_{i2}) + \dots + x_{ip}\beta_p(u_{ip}) + \varepsilon_i, \quad i=1, \dots, n$$

şeklinde ifade edildiği bilinmektedir. Bu modelde $\beta_j(\cdot)$ 'leri tahmin etmek için

$$\sum_{i=1}^n \left\{ y_i - \sum_{j=1}^p x_{ij}\beta_j(u_{ij}) \right\}^2 + \sum_{j=1}^p \lambda_j \int \beta_j''(u_j)^2 du_j \quad (3.11)$$

(2.26)'daki eşitliğe benzer şekilde tanımlanan cezalandırılmış en küçük kareler toplamı minimum yapılır. (3.11)'deki eşitlikte ilk kısım uyum iyiliğini ölçerken, ikinci kısım ceza terimi olarak bilinmektedir. $\beta_j(\cdot)$ katsayı fonksiyonlarının tahminlerini bulmak için $\beta_j(\cdot)$ 'ler temel fonksiyonlarına bağlı olarak

$$\beta_j(u) = \sum_{k=1}^{n_j} \gamma_{jk} B_{jk}(u) \quad (3.12)$$

şeklinde ifade edilir. Burada n_j , u 'nun tekrar etmeyen değerlerinin sayısıdır (yani düğüm sayısıdır). $B_{jk}(u)$, γ_{jk} parametresiyle j . değişken için temel fonksiyonunu ifade etmektedir ve polinom temelli doğal kübik eğrisel çizgi veya B-eğrisel çizgi fonksiyonlarından birisi olabilir [6].

Kolaylık sağlanması için, denklem doğal kübik spline temelinde parametreleştirilir. $\beta_j(\cdot)$ fonksiyonları doğal kübik eğrileri kullanılarak tahmin edilmek istenirse,

$$\beta_j(u_{ij}) = \sum_{i=1}^{n_j} \gamma_{ij} N_j^l(u_{ij})$$

$$\beta_j = \mathbf{N}_j \boldsymbol{\gamma}_j$$

şeklinde tanımlanır.

Doğal kübik eğrilerin kullanılmasının temel nedeni, (3.11)'deki ceza terimi olarak bilinen integralin toplamsal bir eşitliğe dönüştürebilmesidir. Bu sayede (3.11)'deki ifadenin türevi alınabilir ve tahminler elde edilebilir.

Cezalandırılmış en küçük kareler toplamı doğal kübik eğriler matrisi ile

$$R = \left\| \mathbf{y} - \sum_1^p \mathbf{V}_j \mathbf{N}_j \boldsymbol{\gamma}_j \right\|^2 + \sum_1^p \lambda_j \boldsymbol{\gamma}_j^T \boldsymbol{\Omega}_j \boldsymbol{\gamma}_j \quad (3.13)$$

şeklinde ifade edilir. Burada $\boldsymbol{\Omega}_j$, $\int \mathbf{N}_j^{l''}(u) \mathbf{N}_j^{k''}(u) du$ integralinin j'nci elemanıdır. \mathbf{V}_j , x_j değerlerinden oluşan köşegen matristir. u_j 'nin tekrar etmeyen değerlerinin sayısı n_j olup j'inci değişken için doğal kübik eğrisel çizgi temel fonksiyonları $\mathbf{N}_j^1(u_j), \dots, \mathbf{N}_j^{n_j}(u_j)$ 'dir. \mathbf{N}_j ise ik'inci elemanı $\mathbf{N}_j^k(u_{ij})$ olan temel matristir (basis matrix).

(3.13)'ün $\boldsymbol{\gamma}_j$ 'ye göre türevi alınırsa $\beta_j(\cdot)$ fonksiyonlarının tahmini doğal kübik eğriler kullanılarak aşağıdaki gibi elde edilir:

$$\frac{\partial R}{\partial \boldsymbol{\gamma}_j} = (\mathbf{N}_j^T \mathbf{V}_j^2 \mathbf{N}_j + \lambda_j \boldsymbol{\Omega}_j) \boldsymbol{\gamma}_j = \mathbf{N}_j^T \mathbf{V}_j (\mathbf{y} - \sum_{k \neq j} \mathbf{V}_k \mathbf{N}_k \boldsymbol{\gamma}_k), \quad j = 1, \dots, p$$

$$\hat{\boldsymbol{\beta}}_j = \mathbf{N}_j \hat{\boldsymbol{\gamma}}_j = \mathbf{N}_j (\mathbf{N}_j^T \mathbf{V}_j^2 \mathbf{N}_j + \lambda_j \boldsymbol{\Omega}_j)^{-1} \mathbf{N}_j^T \mathbf{V}_j \left(\mathbf{y} - \sum_{k \neq j} \mathbf{V}_k \mathbf{N}_k \boldsymbol{\gamma}_k \right) = \mathbf{S}_j(\lambda_j) \mathbf{V}_j^- \left(\mathbf{y} - \sum_{k \neq j} \mathbf{V}_k \mathbf{N}_k \boldsymbol{\gamma}_k \right)$$

Burada λ_j , $j=1,2,\dots,p$ düzleştirme parametreleridir (*smoothing parameters*) ve

$\mathbf{S}_j(\lambda_j) = \mathbf{N}_j (\mathbf{N}_j^T \mathbf{V}_j^2 \mathbf{N}_j + \lambda_j \boldsymbol{\Omega}_j)^{-1} \mathbf{N}_j^T \mathbf{V}_j^2$ düzleştirme matrisidir.

3.2.3 Cezalandırılmış Eğrisel Çizgi

Değişken katsayılı modelde, Huang ve diğerleri [12] model ke stiminde cezalandırılmış eğrisel çizgileri kullanmışlardır.

Cezalandırılmış eğrisel çizgi (Penalized Spline, P-spline) kestiriminde (bir tü r cezalandırılmış regresyon eğrisel çizgi kestirimi), B-eğrisel çizgi katsayılarına bir fark cezası konularak oluşturulan farklı bir ceza türü kullanılır. P-eğrisel çizgileri Eilers ve Marx'ın [45] tanımladığı gibi kullanmanın başlıca yararı P-eğrisel çizgilerin hiçbir sınır etkisi i çermemesidir. B-eğrisel çizgilerin temel fonksiyonlar olarak kullanımı, sayısal özelliklerinden dolayı literatürde tercih edilmektedir [18, 6].

$\{y_i, \mathbf{X}_i^T, u_i\}_{i=1}^n$ rasgele örnekleminde tüm $u \in [a, b]$ için $\beta_r(u)$ 'nin, birinci mertebeden türevlenebilir olduğu varsayalım. $\beta_r(u)$, B-eğrisel çizgiler kullanıldığında

$$\beta_r(u_i) = c_{r1}B_{r1}(u_i) + \dots + c_{rn_r}B_{rn_r}(u_i) = \mathbf{B}_r^T(u_i)\mathbf{c}_r$$

şeklinde ifade edilir. $\mathbf{c}_r = (c_{r1}, \dots, c_{rn_r})$ ve $\mathbf{B}_r(u_i) = (B_{r1}(u_i), \dots, B_{rn_r}(u_i))^T$ B-eğrisel çizgi

temelidir. $\beta_r(u)$ 'nin cezalandırılmış eğrisel çizgi tahmin edicisi $\hat{\beta}_r(u)$,

$$\frac{1}{n} \sum_{i=1}^n (y_i - x_{i1}B_{r1}^T(u_i)\mathbf{c}_1 - \dots - x_{ip}B_{rp}^T(u_i)\mathbf{c}_p)^2 + \sum_{r=1}^p \lambda_r \mathbf{c}_r^T \mathbf{Q}_r \mathbf{c}_r \quad (3.14)$$

şeklindeki cezalandırılmış artık kareler toplamı minimum yapılarak elde edilir. Burada λ_r

negatif olmayan düzeltme parametresi ve \mathbf{Q}_r 'ler, (2.35) eşitliğindeki \mathbf{D} matrisine benzer

şekilde tanımlanır. $\mathbf{z}_i = (x_{i1}B_{r1}^T(t_i), \dots, x_{ip}B_{rp}^T(t_i))^T$ ve $\boldsymbol{\gamma} = (c_1^T, \dots, c_n^T)^T$ olmak üzere (3.14),

$$\frac{1}{n} \sum_{i=1}^n (y_i - \mathbf{z}_i^T \boldsymbol{\gamma})^2 + \sum_{r=1}^p \lambda_r \boldsymbol{\gamma}^T \mathbf{S}_r \boldsymbol{\gamma}$$

$$(\mathbf{y} - \mathbf{Z}\boldsymbol{\gamma})^T (\mathbf{y} - \mathbf{Z}\boldsymbol{\gamma}) + \sum_{r=1}^p \lambda_r \boldsymbol{\gamma}^T \mathbf{S}_r \boldsymbol{\gamma} \quad (3.15)$$

şeklinde ifade edilir. (3.15)'te $\mathbf{Z} = \sum_{i=1}^n \mathbf{z}_i \mathbf{z}_i^T$, $\mathbf{S}_r = \text{diag}(0, \dots, 0, \mathbf{Q}_r, 0, \dots, 0)$ ve $\boldsymbol{\gamma}^T \mathbf{S}_r \boldsymbol{\gamma} = \mathbf{c}_r^T \mathbf{Q}_r \mathbf{c}_r$

'dir. λ_r düzeltme parametresi ile ilişkili olan \mathbf{S}_r 'ler pozitif tanımlı cezaları belirten

katsayı matrisleridir. (3.15)'teki cezalandırılmış kareler toplamını minimum yapan $\boldsymbol{\gamma}$

$$\hat{\boldsymbol{\gamma}} = (\mathbf{Z}^T \mathbf{Z} + \mathbf{S})^{-1} \mathbf{Z}^T \mathbf{y} \quad (3.16)$$

olarak tahmin edilir. Burada $\mathbf{S} = (\mathbf{S}_1, \dots, \mathbf{S}_p)$ blok köşegen matristir. Buna göre β 'nin B-

eğrisel çizgi temeli kullanılarak cezalandırılmış eğrisel çizgi tahmini

$$\hat{\boldsymbol{\beta}} = \mathbf{B}\hat{\boldsymbol{\gamma}}$$

dır [54]. P-eğrisel çizgi kestiriminde ceza, farklı derecelere sahip olabilir. Böylece birinci,

ikinci veya üçüncü farklar alınabilir. Eilers ve Marx [45], en azından kuadratik veya kübik

B-eğrisel çizgi temelli ikinci dereceden bir fark cezasının kullanılmasını önermektedir. P-

eğrisel çizgi tahmin yöntemi, ceza derecesinin seçilmesine izin verdiği için daha fazla

esneklik sağlar.

3.2.3.1 Düzleştirme Parametrelerinin Seçimi

Sabit bir düğüm sayısı verildiğinde, genellikle tek düğüm değerleri yerleştirilir. Ruppert [54] düğüm sayısının seçimi ile ilgili ayrıntılı bir çalışma yapmıştır. Fonksiyonları düzgünleştirmek için, monoton ve ya t ek tepeli, 10-15 arasında düğüm noktasının seçilmesini önermiştir. Ayrıca katsayı fonksiyonlarının birden çok yerel minimum ve maksimum değere sahip olması durumunda 20'den fazla düğüm noktasının seçilmesi gerektiğini belirtmiştir. Cezalandırılmış eğrisel çizgilerin avantajı düğüm noktalarının yerleştirilmesi probleminin, regresyon eğrisel çizgilerini kullanmak yerine cezalandırılmış regresyon eğrisel çizgilerini kullanarak kısmen hafifletmesidir [54]. Fakat bu durumda model esnekliği düzleştirme parametresi ile kontrol edilir. Bununla birlikte düzleştirme parametrelerinin seçimi önemli bir konudur.

λ düzleştirme parametresinin seçimi için geliştirilmiş çapraz geçerlilik ölçütü

$$GCV = \frac{n \|y - \mathbf{A}y\|^2}{[\text{tr}(\mathbf{I} - \mathbf{A})]^2} \quad (3.17)$$

şeklindedir. Burada $\mathbf{A} = \mathbf{Z} \left(\mathbf{Z}\mathbf{Z}^T + \sum_{r=1}^p \lambda_r \mathbf{S}_r \right)^{-1} \mathbf{Z}^T$ 'dir. GCV ölçütüne dayanan kullanışlı ve etkili çoklu düzleştirme parametre tahminleri Gu ve Wahba [57] tarafından önerilmiştir. Wood [58], geliştirilmiş toplamsal modelde düzleştirme derecesinin seçimi için bir ölçüt önermiştir. Wood [59] sabit ceza terimine izin veren istikrarlı ve etkili bir, çoklu düzleştirme parametre tahmini önermiştir. Çoklu düzleştirme parametresi seçimine ilişkin bu çalışmaların ışığında, Lu ve diğerleri [55] DKM için cezalandırılmış eğrisel tahmininin düzleştirme parametrelerinin seçiminde etkili bir yöntem geliştirmişlerdir.

4. UYGULAMA

Klinik deneyimlere göre, risk faktörlerinin gebelik sırasında kadınların kan basıncı üzerindeki etkilerini bilmek önemlidir. Bu amaçla yürütülen çalışmada, 2018 yılı Mart-Temmuz aylarında muayeneye gelen gebelere ilişkin veriler, Ek 6 'da verilen hasta takip formları doldurularak elde edilmiştir. Gebelere ait klinik verilerde, risk faktörlerinin gestasyonel yaş (hamilelik haftası) indeks değişken alınarak, ortalama arteriyel kan basıncı (MAP) üzerine etkilerini araştırmak için değişken katsayılı model kullanılmıştır. Risk faktörleri; yaş, diyabet hikâyesi, obezite, eğitim düzeyi ve genetik vb. değişkenleri içerir [60]. Çalışma kapsamında benzer sosyo-ekonomik özelliklere sahip olduğu bilinen 450 gebe incelenmiştir.

4.1 Veriler

Gebelerin kan basıncı ve risk faktörleri arasındaki ilişkiyi belirlemek için, ortalama arteriyel kan basıncı (MAP) bağımlı değişken; düşük öyküsü, aile hipertansiyon öyküsü, vücut kitle indeksi, yaş, diyabet hastalığı öyküsü, kötü obstetrik öyküsü, hematokrit (HCT) ve kandaki trombosit sayısı (BPC) bağımsız değişkenler olarak alınmıştır. Çalışmada kullanılan değişkenlere ilişkin açıklayıcı bilgiler aşağıda verilmiştir:

Ortalama arteriyel kan basıncı (*The Mean Arterial blood Pressure, MAP*): Damarın içindeki kanın damar duvarına yaptığı yüksek basınca hipertansiyon denir. Kan basıncı; kalbin kanı pompalarken oluşturduğu sistolik basınç ve kalbin kanı pompalama sırasında ara verdiği dönemdeki diastolik basınç olarak iki farklı değerden oluşur.

Ortalama arteriyel kan basıncı ise sistolik (büyük tansiyon) ve diastolik (küçük tansiyon) kan basıncından,

$$MAP = \frac{(2 * \text{diastolik}) + \text{sistolik}}{3}$$

formülü ile hesaplanmıştır [61].

Normal kan basıncı değerleri sırasıyla sistolik için en çok 130 mmHg, diastolik için ise en çok 80 mmHg olmalıdır.

Preeklampsi; gebeliğin 20'nci haftasından sonra ortaya çıkan, proteinürinin (24 saatlik idrarda 300 mg'ın üzerinde protein bulunması) eşlik ettiği hipertansiyon ile karakterize edilen bir sendromdur.

Eklampsi; gebelik ya da loğusalık sırasında preeklampsi olarak tanımlanan hastalarda nörolojik hasar olmadan gelişen koma durumudur. Eklampsi nöbetleri sonrasında koma ve ölüm, geçici kortikal körlük, retinal yırtılma vb. olabilir.

Preeklampsi veya eklampsi olan kadınlarda intraserebral kanama en sık ölüm nedenini oluşturmaktadır.

Gebelikte görülen hipertansiyon, daha önce hiç yüksek tansiyon yakınması olmayıp, gebeliğin 20'inci haftasından sonra ortaya çıkan, preeklampsinin proteinüri benzeri sistemik bulgularının olmadığı hipertansiyon gelişimine verilen isimdir. Bu sürece sistolik kan basıncı 140 mmHg, diastolik kan basıncı 90 mmHg'nin üzerine çıkar. Gebelik hipertansiyonu her ne kadar tehlikesiz olsa da vakaların yaklaşık yarısında gebelik hipertansiyonu preeklampsiye (bu hasta grubunda yüksek risk teşkil eder %15-45), eklampsiye ve hatta anne ve fetüste ölüme neden olabilecek tehlikeli bir duruma dönüşür.

Özetle, anne ve fetus için büyük risk taşıyan gebelikte görülen hipertansiyonun risk faktörlerini belirlemek ve hastalığın komplikasyonlarının kötüye gidişini engellemek gerekmektedir.

Gebeliğin 10'uncu ya da 20'nci haftasındaki ortalama arteriyel kan basıncının, preeklampsi teşhisinde diastolik ya da sistolik kan basıncına göre daha iyi bir ölçüt olduğu yapılan çalışmalarda kanıtlanmıştır. Bu nedenle bu çalışmada preeklampsi vb. gibi sendromların ortaya çıkarılmasında gestasyonel yaşa (gebelik haftasına) göre değiştiği bilinen ortalama arteriyel kan basıncı, **bağımlı değişken (y)** olarak alınmıştır [62].

Gestasyonel yaş: Gebelik yaşı (haftası) anlamına gelmektedir. Bu çalışmada kolaylık sağlaması için gebelik haftası ifadesi kullanılacaktır. Her gebelik haftasında, bazı risk faktörleri, diğer faktörlerden daha önemlidir. Örneğin 20 haftalık gebede; hematokritin kan basıncı üzerinde büyük bir etkisi vardır, ancak bu risk faktörleri ile ilişkili gebelik haftalarında önemsiz olabilir. Bu nedenle gebelik haftası değişkeni üzerinde diğer değişkenlerin etkisi incelenmek istenmiştir. Çalışmada **indeks değişken (u)** olarak alınmıştır.

Literatür taraması ve uzman görüşüne göre, ortalama arteriyel kan basıncını etkileyen değişkenler (**bağımsız değişkenler, x_i**) **yaş, vücut kitle indeksi, gebelik öyküsü, düşük**

öyküsü, aile hipertansiyon öyküsü, diyabet hastalığı öyküsü, kötü obstetrik öykü, hematokrit, kandaki trombosit sayısı olarak alınmıştır. Bu değişkenlere ilişkin detaylı bilgi aşağıda verilmiştir.

Yaş: Yapılan çalışmalarda yaş ilerledikçe gebelik hipertansiyonuna sahip olma riskinin arttığı görüldüğü için yaş değişkeni, 35 yaşından küçük ve 35 yaşından büyük olmak üzere iki kategoride incelenmiştir.

Vücut kitle indeksi (VKI): Vücut kitle indeksi,

$$VKI = \frac{\text{kilo (kg)}}{(\text{boy(m)})^2}$$

formülünden elde edilir. Türkiye değerlerine göre uzman görüşü alınarak VKI değerleri “<25”, “25-30” ve “>30” olmak üzere sınıflandırılmıştır. Bu sınıflandırmada ilk sınıf gebenin normal kiloda olduğunu, ikinci sınıf gebenin kilolu olduğunu ve üçüncü sınıf gebenin obez olduğunu belirtmektedir. Gebelerde hipertansif bozuklukların ileri yaş annelerde olduğu gibi obezlerde de daha sık görüldüğü bilinmektedir. Örneğin; VKI 20’nin altında olanlarda preeklampsi oranı %4,3 iken 35 üzerinde olanlarda %13,3’tür.

Gebelik öyküsü: Preeklampsinin ilk gebelikte görülme oranı ortalama %85’tir. Bununla birlikte aşırı doğurganlık (özellikle 4. gebelik sonrası) gebelikte hipertansiyon insidansında artışa neden olmaktadır. Bu durumda ilk gebelik ve çoklu gebeliklerin negatif etkisi beklenmektedir. Tekil gebeliklerde %6 iken, çoğul gebeliklerde %13’tür.

Düşük öyküsü: Aşırı doğurganlıkta olduğu gibi, düşük sayısının artması gebelerde ortalama arteriyel kan basıncı üzerinde negatif etkiye sahip olacaktır.

Aile hipertansiyon öyküsü: Hipertansiyonu, özellikle aile öyküsü olan hastalarda erken tanımlamak, ileride oluşacak komplikasyonları önlemek açısından son derece önemlidir.

Diyabet hastalığı öyküsü: Hipertansif bir bozukluk olan preeklampsi için, risk faktörleri arasında gebelik öncesi diyabet öyküsünün bulunmasının da bir etken olduğu çalışmalarda belirtilmiştir.

Kötü obstetrik öykü: Önceki üç veya daha fazla gebelikte değişik veya aynı sebeplerden dolayı gebeliklerin sağlıklı bir şekilde sona ermeyip erken dönemde düşük, bozulmuş gebelik, anne karnında bebek ölümü veya anomalili doğum ile sonuçlanmasıdır. Kötü obstetrik öyküsü olan gebeler, gebelikleri esnasında daha sık, sıkı ve ciddi bir şekilde takip edilmeliler ve doğumları da planlı bir şekilde olmalıdır. Kötü obstetrik öykü olarak

sınıflanabilecek durumlar aşağıda sıralanmaktadır (Art arda en az 2 veya 3 tanesine bağlı olarak gebelik kaybı oluşması).

- Erken dönem düşükler
- Habitual abortus (tekrarlayan düşükler)
- Erken gebelik döneminde bozulmuş gebelik kesesi
- Anembriyonik (boş) gebelik
- Ektopik (dış) gebelik
- Mol gebeliği (Üzüm gebeliği; 72, 96 vb. kromozomlu bozuk hücreler oluşması ve çoğalması)
- Serviks (rahim ağzı) yetersizliğine bağlı ikinci trimester (13-26 gebelik haftası) gebelik kayıpları
- Kromozom anomalili doğumlar
- Erken doğuma bağlı gebelik kayıpları
- Konjenital marformasyonlu doğumlar (Doğumsal bozukluklar; down sendromu, tek bacak eksikliği vb.)
- Gelişme geriliğine bağlı gebelik kayıpları
- Sebebi bilinmeyen anne karnındaki bebek ölümleri
- Gebelikte geçirilmiş kaza, travma veya ameliyatlara bağlı bebek kayıpları
- Doğum travmasına (vakum, forseps uygulamaları) bağlı bebek kayıpları
- Doğum sırasında fetal asfiksi (doğum anı ve sonrasında herhangi bir durumdan ötürü boğulma; kordon sıkışması, plesentadan ayrılma vb.) gibi durumlara bağlı bebek kayıpları
- Neonatal ölüm, sebebi bilinmeyen yenidoğan ölümleri

Hematokrit (*Hematocrit, HCT*): Kanda bulunan alyuvarların hacminin toplam kan hacmine denir. Bu oran yüzdelik olarak ifade edilir. Uzman görüşü alınarak HCT değerleri “23.6-31.3”, “31.4-39.1” ve “39.2-46.9” olmak üzere sınıflandırılmıştır. Hematokritin kan basıncı üzerinde gebeliğin 20’nci haftasından itibaren büyük bir etkisinin olduğu bilinmektedir.

Kandaki trombosit sayısı (*Blood Platelet Count, BPC*): Trombositler kan pıhtılaşmasına yardımcı olan küçük kan hücresi parçalarıdır. Uzman görüşü alınarak BPC değerleri “<150”, “150-300” ve “>300” olmak üzere sınıflandırılmıştır. Örneğin; ağır preeklampsi vakalarında

genellikle düşük BPC değerleri gözlenmektedir. BPC aşırı yükselirse tansiyon da artmaktadır.

Bağımsız değişkenlere ilişkin sıklık tablosu Çizelge 4.1'de verilmiştir:

Çizelge 4.1. Bağımsız Değişkenlere İlişkin Sıklık Tablosu

Bağımsız Değişken	Kısaltma	Kategori	f	%
Yaş	Yaş	<35	403	89,6
		≥ 35	47	10,4
Vücut Kitle İndeksi	VKI	<25	20	4,40
		25-30	186	41,3
		>30	244	54,2
Gebelik haftası	GH	-	-	-
Gebelik öyküsü	GO	0: Yok	244	54,2
		1: Var	206	45,8
Düşük öyküsü	DO	0: Yok	336	74,7
		1: Var	114	25,3
Aile hipertansiyon öyküsü	AHP	0: Yok	252	56,0
		1: Var	198	44,0
Diyabet hastalığı öyküsü	DY	0: Yok	328	72,9
		1: Var	122	27,1
Kötü obstetrik öyküsü	KO	0: Yok	432	96,0
		1: Var	18	4,00
Hematokrit	HCT	23,6-31,3	30	6,70
		31,4-39,1	344	76,4
		39,2-46,9	76	16,9
Kandaki trombosit sayısı	BPC	<150	29	6,40
		150-300	351	78,0
		>300	70	15,6

4.2 Analiz

Bu çalışmada gebelerin ortalama kan basıncını etkileyen risk faktörlerini belirlemek için gebelik haftasının değişen değişken olarak alındığı

$$MAP_i = \beta_1(GH)Yaş + \beta_2(GH)VKI + \beta_3(GH)GO + \beta_4(GH)DO + \beta_5(GH)AHP + \beta_6(GH)DY + \beta_7(GH)KO + \beta_8(GH)HCT + \beta_9(GH)BPC$$

şeklindeki değişen katsayılı model tahmin edilmiştir.

Değişen katsayılı model tahmininde düzleştirilmiş eğrisel çizgi ve cezalandırılmış eğrisel çizgi yöntemleri kullanılmıştır. Düzleştirilmiş eğrisel çizgi ve cezalandırılmış eğrisel çizgi yöntemleri ile tahmin yapılırken düğüm noktalarının seçilmesine gerek yoktur. Bu

yöntemlerde düğüm noktası sayısı olarak indeks değişkenin tekrar etmeyen değerlerinin olabilecek en büyük değeri düğüm noktası sayısı olarak alınır. Bu nedenle bu çalışmada düğüm noktası sayısı gebelik haftasına göre belirlenmiştir. Düzleştirme parametresinin seçiminde genelleştirilmiş çapraz geçerlilik ölçütü kullanılmıştır. Düzleştirilmiş eğrisel çizgi ve cezalandırılmış eğrisel çizgi ile tahmin edilen modellerden en küçük Akaike Bilgi Kriter değerine sahip değişken katsayılı model seçilmiştir. Analiz sonuçları R programında “mgcv” paketi kullanılarak elde edilmiştir.

Cezalandırılmış eğrisel çizgi ile tahmin edilen değişken katsayılı model sonuçları Çizelge 4.2’de verilmiştir.

Çizelge 4.2. Cezalandırılmış eğrisel çizgi ile tahmin edilen değişken katsayılı model sonuçları

Kısaltma	Kategori	p- değeri
Yaş (GH)	<35	0,0219**
	>35	0,5345
VKI (GH)	<25	0,2445
	25-30	0,5452
	>30	0,2089
GO (GH)	0: Yok	0,2686
	1: Var	0,2680
DO (GH)	0: Yok	0,8057
	1: Var	0,1822
AHP (GH)	0: Yok	0,3228
	1: Var	0,1317
DY (GH)	0: Yok	0,5211
	1: Var	0,1806
KO (GH)	0: Yok	0,0321**
	1: Var	0,0286**
HCT (GH)	23,6-31,3	0,6172
	31,4-39,1	0,2985
	39,2-46,9	0,0876*
BPC(GH)	<150	0,6096
	150-300	0,0444**
	>300	0,4898
AIC=3557,417		

*0,1 anlamlılık düzeyinde; **0,05 anlamlılık düzeyinde

Çizelge 4.2 incelendiğinde cezalandırılmış eğrisel çizgi tahmin sonuçlarına göre, istatistiksel olarak gebelerin ortalama arteriyel kan basıncını (MAP) etkileyen değişken katsayıya sahip

değişkenler; yaş, kötü obstetrik öyküsü (KO), hematokrit (HCT) ve kandaki trombosit sayısı (BPC) olarak bulunmuştur.

Düzleştirilmiş eğrisel çizgi ile tahmin edilen sonuçlar Çizelge 4.3'te verilmiştir.

Çizelge 4.3. Düzleştirilmiş eğrisel çizgi ile tahmin edilen değişken katsayılı model sonuçları

Kısaltma	Kategori	p- değeri
Yaş (GH)	<35	0,2604
	>35	0,2313
VKI (GH)	<25	0,1041
	25-30	0,2358
	>30	0,8731
GO (GH)	0: Yok	0,6623
	1: Var	0,3932
DO (GH)	0: Yok	0,2147
	1: Var	0,1470
AHP (GH)	0: Yok	0,0795*
	1: Var	0,7263
DY (GH)	0: Yok	0,8560
	1: Var	0,0964*
KO (GH)	0: Yok	0,0020**
	1: Var	0,0274**
HCT (GH)	23,6-31,3	0,0952*
	31,4-39,1	0,3763
	39,2-46,9	0,8321
BPC(GH)	<150	0,0426**
	150-300	0,0250**
	>300	0,0027**
AIC=3552,623		

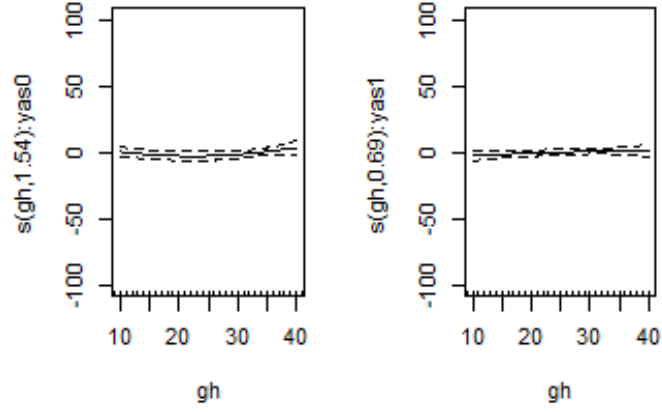
*0,1 anlamlılık düzeyinde; **0,05 anlamlılık düzeyinde

Çizelge 4.3 incelendiğinde düzleştirilmiş eğrisel çizgi tahmin sonuçlarına göre, istatistiksel olarak gebelerin ortalama arteriyel kan basıncını (MAP) etkileyen değişken katsayıya sahip değişkenler; aile hipertansiyon öyküsü (AHP), diyabet hastalığı öyküsü (DY), kötü obstetrik öyküsü (KO), hematokrit (HCT) ve kandaki trombosit sayısı (BPC) olarak bulunmuştur.

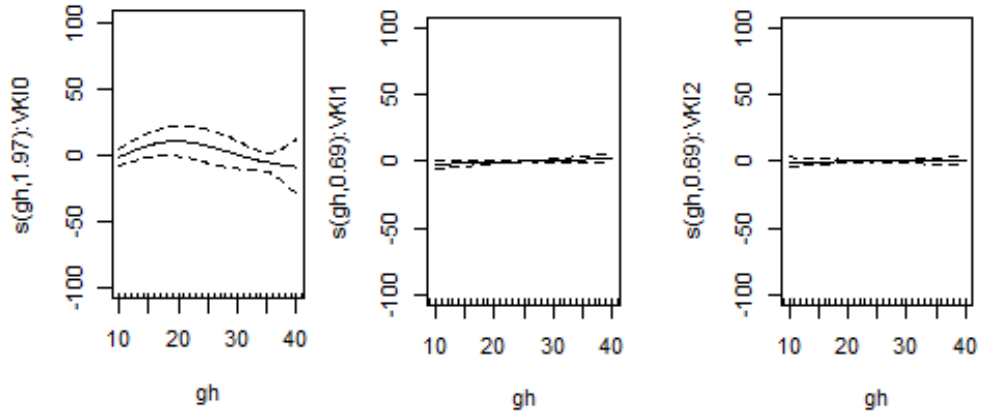
Düzleştirilmiş eğrisel çizgi ile tahmin edilen değişken katsayılı model, cezalandırılmış eğrisel çizgi ile tahmin edilen modele göre daha küçük AIC değerine sahiptir. Buna denkle düzleştirilmiş eğrisel çizgi ile tahmin edilen değişken katsayılı model tercih edilmiştir.

Düzleştirilmiş eğrisel çizgi ile tahmin edilen katsayı fonksiyonlarının gebelik haftalarına göre grafikleri Şekil 4.1’de verilmiştir.

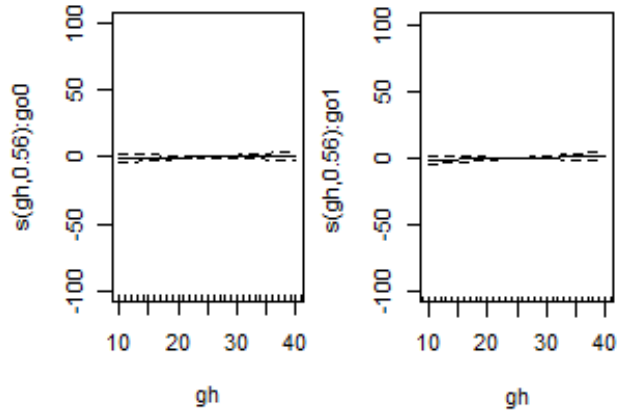
(a)



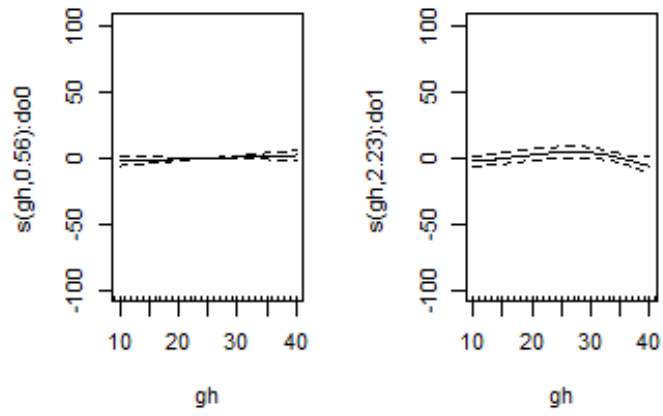
(b)



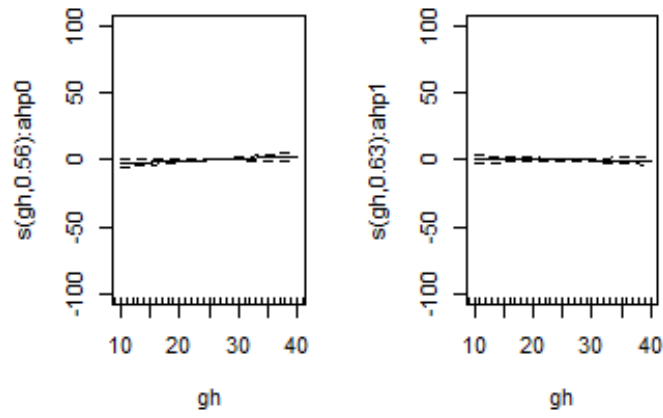
(c)



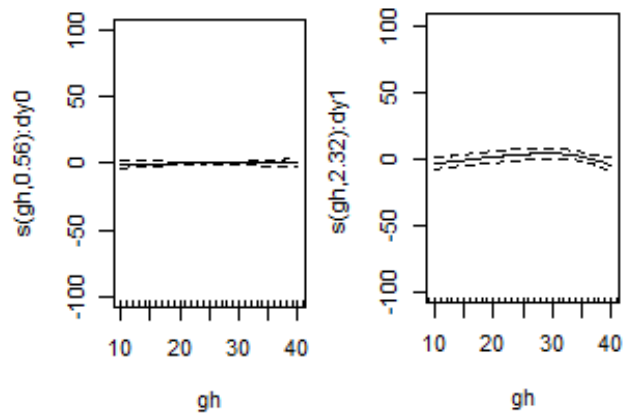
(d)



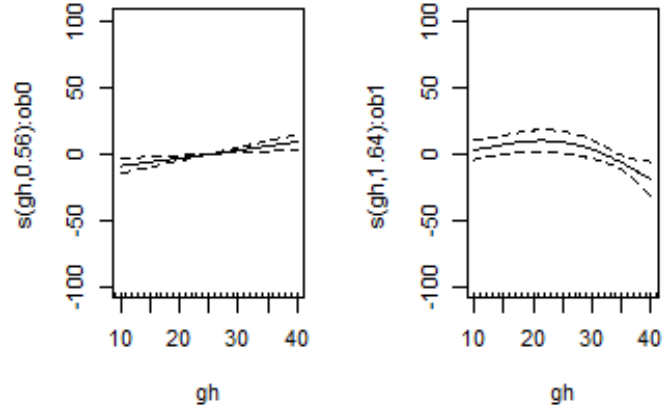
(e)



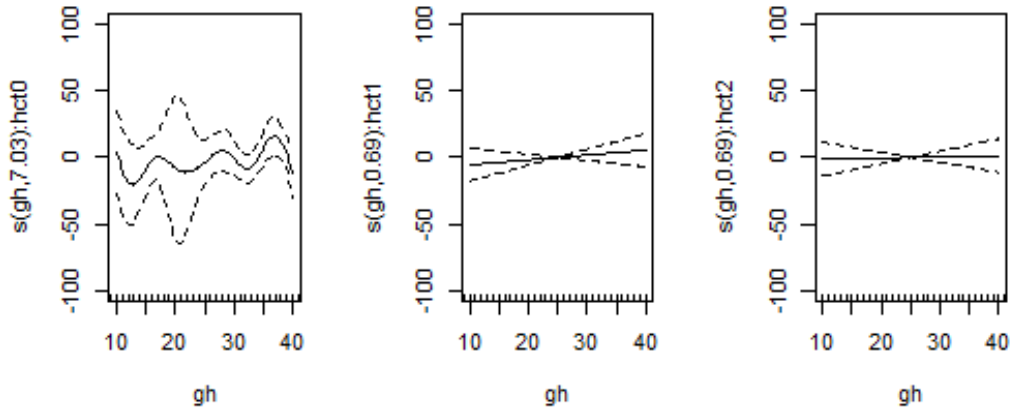
(f)



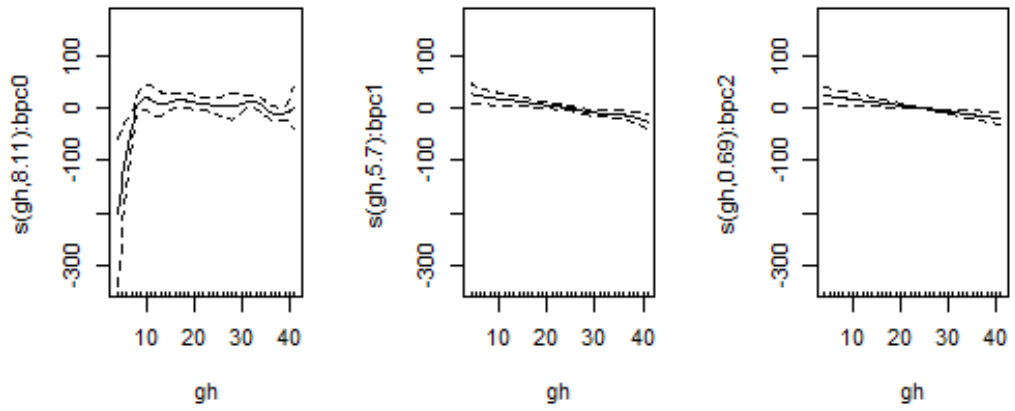
(g)



(h)



(i)



Şekil 4.1. Düzleştirilmiş eğrisel çizgi ile tahmin edilen fonksiyonlarının gebelik haftasına göre grafikleri (a, b, c, d, e, f, g, h, i)

Düzleştirilmiş eğrisel çizgi tahmin sonuçlarına göre istatistiksel olarak:

Şekil 4.1 (a)'da görüldüğü gibi, değişen katsayılı modelde **yaş** değişkeninin ortalama arteriyel kan basıncı üzerine etkisi incelendiğinde, 35 yaşından küçük ve 35 yaşından büyük gebelerde ortalama arteriyel kan basıncı (MAP) değişiminin gebelik haftasına göre çok farklı olmadığı, aynı seviyede kaldığı görülmektedir. Çizelge 4.3'te de yaş değişkeninin %95 güven düzeyinde istatistiksel olarak MAP üzerine anlamlı bir etkisi olmadığı görülmektedir.

Şekil 4.1 (b)'de görüldüğü gibi, değişen katsayılı modelde **vücut kitle indeksi (VKI)** değişkeninin ortalama arteriyel kan basıncı üzerine etkisi incelendiğinde, VKI<25 olan yani zayıf olan gebelerin ortalama arteriyel kan basıncı (MAP) ilk haftalarda az miktar artsa da genelde aynı seviyede kaldığı görülmektedir. VKI değeri 25-30 arasında olan normal gebelerde ve VKI>30 olan obez gebelerde ise ortalama arteriyel kan basıncı değerlerinin gebelik haftasına göre çok farklı olmadığı, aynı seviyede kaldığı görülmektedir. Çizelge 4.3'te de VKI değişkeninin %95 güven düzeyinde istatistiksel olarak MAP üzerine anlamlı bir etkisi olmadığı görülmektedir.

Şekil 4.1 (c)'de görüldüğü gibi, değişen katsayılı modelde **gebelik öyküsü** değişkeninin ortalama arteriyel kan basıncı (MAP) üzerine etkisi incelendiğinde, ilk gebeliği olan ve daha önce gebelik geçirmiş olanlar için MAP değişiminin haftalara göre çok farklı olmadığı, aynı seviyede kaldığı görülmektedir. Çizelge 4.3'te de gebelik öyküsü değişkeninin %95 güven düzeyinde istatistiksel olarak MAP üzerine anlamlı bir etkisi olmadığı görülmektedir.

Şekil 4.1 (d)'de görüldüğü gibi, değişen katsayılı modelde **düşük öyküsü** değişkeninin ortalama arteriyel kan basıncı (MAP) üzerine etkisi incelendiğinde, MAP değerinin düşük öyküsü olan gebelerde 25'nci haftaya kadar artma eğiliminde olduğu görülse de genel olarak çok ciddi bir artış gözlenmemektedir. Daha önce düşük yapmamış ve düşük yapmış gebelerin MAP değerlerinin haftalara göre çok değişmediği, aynı seviyede kaldığı görülmektedir. Çizelge 4.3'te de düşük öyküsü değişkeninin %95 güven düzeyinde istatistiksel olarak MAP üzerine anlamlı bir etkisi olmadığı görülmektedir.

Şekil 4.1 (e)'de görüldüğü gibi, değişen katsayılı modelde **aile hipertansiyon öyküsü** değişkeninin ortalama arteriyel kan basıncı (MAP) üzerine etkisi incelendiğinde, MAP değerlerinin değişiminin aile hipertansiyon öyküsü olan ve olmayan gebelerde haftalara göre çok farklı olmadığı, aynı seviyelerde kaldığı görülmektedir. Çizelge 4.3'te de aile hipertansiyon öyküsü değişkeninin %95 güven düzeyinde istatistiksel olarak anlamlı olmadığı görülmektedir.

Şekil 4.1 (f)'de görüldüğü gibi, değişen katsayılı modelde **diyabet hastalığı öyküsü** değişkeninin ortalama arteriyel kan basıncı (MAP) üzerine etkisi incelendiğinde, MAP değerlerinin diyabet hastalığı olan ve olmayan gebeler için çok fazla değişmediği görülmektedir. Ancak Şekil 4.1 (f)'de diyabet hastalığı olan gebelerde MAP değerinin 25'nci haftaya kadar az da olsa arttığı, sonrasında azaldığı görülmektedir. Bu nedenle, Çizelge 4.3 incelendiğinde gebelerde diyabet hastalığının olmasının MAP değerleri üzerinde %90 güven düzeyinde istatistiksel olarak anlamlı olduğu söylenir.

Şekil 4.1 (g)'de görüldüğü gibi, değişen katsayılı modelde **kötü obstetrik öykü** değişkeninin ortalama arteriyel kan basıncı (MAP) üzerine etkisi incelendiğinde, MAP değerlerinin haftalara göre değiştiği, 30'ncü haftadan itibaren ise düşmeye başladığı görülmektedir. Çizelge 4.3'te de kötü obstetrik öykü değişkeninin MAP değerleri üzerinde %95 güven düzeyinde istatistiksel olarak anlamlı olmadığı görülmektedir.

Şekil 4.1 (h)'de görüldüğü gibi, değişen katsayılı modelde **hematokrit (HCT)** değişkeninin ortalama arteriyel kan basıncı (MAP) üzerine etkisi incelendiğinde, HCT değeri "23,6-31,3" aralığında olan gebelerin MAP değerinin haftalara göre değiştiği, 20'nci haftada en düşük seviyede olduğu görülmektedir. Çizelge 4.3'te de HCT değişkeninin, "23,6-31,3" düzeyinin MAP değerleri üzerinde %95 güven ile istatistiksel olarak anlamlı olduğu söylenir.

Şekil 4.1 (ı)'da görüldüğü gibi, değişen katsayılı modelde **kandaki trombosit sayısı (BPC)** değişkeninin ortalama arteriyel kan basıncı (MAP) üzerine etkisi incelendiğinde, gebelerin MAP değerinin BPC'nin tüm düzeylerinde 20'nci haftadan sonra hızla düştüğü görülmektedir. Çizelge 4.3'te de BPC değişkeninin MAP değerleri üzerinde %95 güven düzeyinde istatistiksel olarak anlamlı olduğu söylenir.

5. SONUÇ

Bu tez çalışmasında doğrusal regresyon modellerinin esnekliğini artırmak ve doğrusal regresyon modellerinin uygulama alanlarını genişletmek amacıyla, bilinmeyen katsayılarının bazı bağımsız değişkenlerinin düzgün fonksiyonları olarak ifade edildiği değişen katsayılı modelin tanıtılması amaçlanmıştır. Değişen katsayılı model parametrik olmayan regresyon modellerinden birisidir. Bu amaçla ilk olarak tek değişkenin olduğu basit parametrik olmayan regresyon modeli tanıtılmıştır. Daha sonra bu modelin tahmininde yaygın olarak kullanılan çekirdek kestiricisi, yerel polinom regresyon kestiricisi, düzleştirilmiş eğrisel çizgi kestiricisi ve cezalandırılmış eğrisel çizgi kestiricileri ile ilgili detaylı bilgilere yer verilmiştir. Değişen katsayılı model tahmininde kullanılan yerel polinom regresyon kestiricisi, düzleştirilmiş eğrisel çizgi kestiricisi ve cezalandırılmış eğrisel çizgi kestiricileri tanıtılmıştır.

Uygulamada 2018 yılı Nisan-Temmuz aylarında muayene ol an gebelere ilişkin veriler değişen katsayılı model ile analiz edilmiştir. Bu verilere değişen katsayılı modelin uygulanmasının nedeni her gebelik haftasında, bazı risk faktörlerinin diğer risk faktörlerine göre daha önemli olmasıdır. Örneğin 20 haftalık gebede; bazı faktörlerin ortalama kan basıncı üzerinde büyük bir etkisi varken bu risk faktörleri ilerleyen gebelik haftalarında önemsiz olabilir. Bu nedenle gebelik haftası değişkenine göre diğer değişkenlerin gebelerin ortalama kan basıncına etkileri değişen katsayılı model ile incelenmiştir. Gebelerin ortalama kan basıncını etkilediği düşünülen risk faktörleri, yapılan literatür taraması ve uzman görüşü sonucunda gebelik öyküsü, düşük öyküsü, aile hipertansiyon öyküsü, vücut kitle indeksi, yaş, diyabet hastalığı öyküsü, kötü obstetrik öyküsü, hematokrit (HCT) ve kandaki trombosit sayısı (BPC) olarak belirlenmiştir. İndeks değişken ise gebelik haftasıdır.

Değişen katsayılı model tahmininde düzleştirilmiş eğrisel çizgi ve cezalandırılmış eğrisel çizgi yöntemleri kullanılmıştır. Düzleştirme parametresinin seçiminde her iki yöntemde de genelleştirilmiş çapraz geçerlilik ölçütü kullanılmıştır. Düzleştirilmiş eğrisel çizgi ile tahmin edilen değişen katsayılı modele ilişkin AIC değeri daha küçük olduğu için verilerin analizinde bu model tercih edilmiştir. Bu modele ilişkin sonuçlar incelendiğinde gebelik haftasına göre istatistiksel olarak gebelerin ortalama arteriyel kan basıncını (MAP) etkileyen değişen katsayıya sahip değişkenler; aile hipertansiyon öyküsü (AHP), diyabet hastalığı

öyküsü (DY), kötü obstetrik öyküsü (KO), hematokrit (HCT) ve kandaki trombosit sayısı (BPC) olarak bulunmuştur. Bulunan sonuçlar genel olarak değerlendirildiğinde düzleştirilmiş eğrisel çizgi tahmin sonuçlarına göre istatistiksel olarak, gebelerin ortalama arteriyel kan basıncını (MAP) etkileyen değişen katsayıya sahip değişkenlerin çoğunun gebeliğin 20'nci haftasından sonra düşmeye başladığı ya da aynı seviyede kaldığı görülmektedir. Çalışmada bahsedilen gebelikte görülen hipertansiyon, preeklampsi ve eklampsi vb. hipertansif hastalıkların belirlenmesinde değerlendirilebilen ortalama arteriyel kan basıncı (MAP); gestasyonel yaşa (gebelik haftası) göre her bir risk faktörü için değişen etkiye sahiptir. Bu nedenle seçilen model sonuçlarında da, 20'nci haftanın bu hastalıklarda beklenen değişim noktası olduğu görülmektedir.

Daha sonraki çalışmalarda, preeklampsi veya eklampsi olan gebelere ilişkin veriler elde edilebilirse bu gebelerde ortalama arteriyel kan basıncının gebelik haftasına göre hangi değişkenlerden etkilendiği araştırılıp, normal gebelerdeki sonuçlar ile karşılaştırılması amaçlanmaktadır.

EKLER

EK 1. Nadaraya-Watson tahmin edicisinin elde edilmesi

Koşullu beklenen değer ile tanımlanan regresyon fonksiyonu

$$\begin{aligned} m(x) &= E(Y|X = x) \\ &= \int yf(y|x)dy \\ &= \frac{\int yf(x, y)dy}{f(x)} \end{aligned} \quad (1)$$

Çarpımsal çekirdek tahmin edicisi;

$$\hat{f}_{h_1, h_2}(x, y) = \frac{1}{n} \sum_{i=1}^n \frac{1}{h_1} K\left(\frac{x_i - x}{h_1}\right) \frac{1}{h_2} K\left(\frac{y_i - y}{h_2}\right) \quad (2)$$

Eşitlik (1) için pay kısmının tahmini,

$$\begin{aligned} \int y \hat{f}(x, y) dy &= \int y \frac{1}{n} \sum_{i=1}^n \frac{1}{h_1} K\left(\frac{x_i - x}{h_1}\right) \frac{1}{h_2} K\left(\frac{y_i - y}{h_2}\right) dy \\ &= \frac{1}{n} \sum_{i=1}^n \frac{1}{h_1} K\left(\frac{x_i - x}{h_1}\right) \int \frac{y}{h_2} K\left(\frac{y_i - y}{h_2}\right) dy \end{aligned} \quad (3)$$

(1.3)'teki integralde $\left(\frac{y_i - y}{h_2}\right) = t$ dönüşümü ile çekirdek fonksiyonun sağladığı

$\int tK(t)dt = 0$ ve $\int K(t)dt = 1$ özellikler yoluyla eşitlik (3),

$$\int y \hat{f}(x, y) dy = \frac{1}{n} \sum_{i=1}^n \frac{y_i}{h_1} K\left(\frac{x_i - x}{h_1}\right) \quad (4)$$

olarak elde edilir.

Sonuç olarak Nadaraya-Watson çekirdek tahmin edicisi,

$$\begin{aligned} \widehat{m}_h(x) &= \frac{\hat{f}_{h_1, h_2}(x, y)}{\hat{f}_h(x)} \\ &= \frac{\frac{1}{n} \sum_{i=1}^n y_i \frac{1}{h_1} K\left(\frac{x_i - x}{h_1}\right)}{\hat{f}_h(x)} \end{aligned} \quad (5)$$

şeklinde elde edilir. $h_1 = h$ sabit bant genişliği kullanılarak Nadaraya-Watson tahmin edicisi,

$$\widehat{m}_h(x) = \frac{\frac{1}{nh} \sum_{i=1}^n y_i K\left(\frac{x_i - x}{h}\right)}{\frac{1}{nh} \sum_{i=1}^n K\left(\frac{x_i - x}{h}\right)} = \frac{1}{n} \sum_{i=1}^n W_i(x) y_i \quad (6)$$

Burada $W_i(x)$ ağırlık fonksiyonudur.

EK 2. B-Eğrisel çizgi tahmininde ceza teriminin elde edilmesi

O'Sullivan (1986, 1988) üçüncü dereceden B-eğrisel çizgi kullanmış ve görülen cezayı elde etmiştir:

$$h^2P = \lambda \int_{x_{\min}}^{x_{\max}} \left\{ \sum_j \gamma_j B_j''(x; 3) \right\}^2 dx \quad (1)$$

B-eğrisel çizgilerinin türev özelliklerinden yola çıkarak,

$$h^2P = \lambda \int_{x_{\min}}^{x_{\max}} \left\{ \sum_j \Delta^2 \gamma_j B_j(x; 1) \right\}^2 dx \quad (2)$$

Bu denklem aşağıdaki şekilde yazılabilir:

$$h^2P = \lambda \int_{x_{\min}}^{x_{\max}} \sum_j \sum_k \Delta^2 \gamma_j \Delta^2 \gamma_k \cdot B_j(x; 1) B_k(x; 1) dx \quad (3)$$

$B_j(x; 1)$ ve $B_k(x; 1)$ 'nin çapraz çarpımlarının çoğu kaybolur, çünkü 1. dereceden B-eğrisel çizgileri yalnızca $j; k-1, k$ veya $k+1$ olduğunda çakışır. Öyle ki;

$$h^2P = \lambda \int_{x_{\min}}^{x_{\max}} \left[\left\{ \sum_j \Delta^2 \gamma_j B_j(x; 1) \right\}^2 + 2 \sum_j \Delta^2 \gamma_j \Delta^2 \gamma_{j-1} \cdot B_j(x; 1) B_{j-1}(x; 1) \right] dx \quad (4)$$

ya da

$$h^2P = \lambda \sum_j (\Delta^2 \gamma_j)^2 \int_{x_{\min}}^{x_{\max}} B_j^2(x; 1) dx + 2\lambda \sum_j \Delta^2 \gamma_j \Delta^2 \gamma_{j-1} \int_{x_{\min}}^{x_{\max}} B_j(x; 1) B_{j-1}(x; 1) dx \quad (5)$$

$$h^2P = \lambda \left\{ c_1 \sum_j (\Delta^2 \gamma_j)^2 + c_2 \sum_j \Delta^2 \gamma_j \Delta^2 \gamma_{j-1} \right\} \quad (6)$$

burada c_1 ve c_2 verilen (eşit uzaklıktaki) düğümler için sabitlerdir:

$$\begin{aligned} c_1 &= \int_{x_{\min}}^{x_{\max}} B_j^2(x; 1) dx; \\ c_2 &= \int_{x_{\min}}^{x_{\max}} B_j(x; 1) B_{j-1}(x; 1) dx. \end{aligned} \quad (7)$$

EK 3. Hasta Takip Formu

Hasta Takip Formu

Ad-Soyad:

Yaş:

Dosya No:

Tarih:

Boy/ Kilo:

Gebeliğinizin kaçınıcı haftasındasınız:

Daha önceki gebelik sayınız: (Nullparite de kapsanmıştır.)

Düşük sayınız:

Gebeliğiniz hangisiyse işaretleyiniz:

- Normal
 Çoğul gebelik
 Mol gebelik

Hipertansiyonunuz olup olmadığını belirtiniz:

Var Yok

Gebelik hipertansiyonun (GHT) olup olmadığını belirtiniz:

Var Yok

Ailenizde hipertansiyon olup olmadığını belirtiniz:

Var Yok

Kronik böbrek hastalığınız olup olmadığını belirtiniz:

Var Yok

Dişabet hastalığınız olup olmadığını belirtiniz:

Var Yok

Kötü obstetrik öykü olup olmadığını belirtiniz:

Var Yok

MAP: Ortalama arteriyel kan basıncı (Mean Arterial Blood Pressure)	
HCT: Hematokrit (Hematocrit)	
MPV: Ortalama trombosit hacimleri (Mean Platelet Volumes)	
BPC: Kan trombosit sayısı (Blood Platelet Count)	

KAYNAKLAR

- [1] Fan, J., Gijbels, I., *Local Polynomial Modelling and Its Applications*, Chapman & Hall, London, **1996**.
- [2] Haggan, V ., Ozaki, T ., Modeling Nonlinear Vibrations Using an Amplitude-Dependent Autoregressive Times Series Model, *Biometrika*, 68, 189-196, **1981**.
- [3] Chen, R., Tsay, R.S., Nonlinear Additive ARX Models, *Journal of the American Statistical Association*, 88(423), 955–967, **1993**.
- [4] Cai, Z., Fan, J., Li, R., Efficient Estimation and Inferences for Varying-Coefficient Models, *Journal of the American Statistical Association*, 95(451), 888-902, **2000**.
- [5] Cai, Z., Fan, J., Li, R., Functional-Coefficient Regression Models for Nonlinear Time Series, *Journal of the American Statistical Association*, 95(451), 941-956, **2000**.
- [6] Hastie, T.J., Tibshirani, R.J., Varying-coefficient models (with discussion), *Journal of the Royal Statistical Society, B*, 55 , 757-796, **1993**.
- [7] Brumback, B.A., Rice, J.A., Smoothing Spline Models for the Analysis of Nested and Crossed Samples of Curves, **1998**.
- [8] Hoover, D .R., Rice, J .A., Wu, C .O., Yang, L .-P., Nonparametric Smoothing Estimates of Time-varying Coefficient Models with Longitudinal Data, *Biometrika*, 85, 809-822, **1998**.
- [9] Wu, C.O., Chiang, C.T., Hoover, D.R., Asymptotic Confidence Regions for Kernel Smoothing of a Varying-Coefficient Model with Longitudinal Data, *Journal of the American Statistical Association*, 93, 1388-1402, **1998**.
- [10] Fan, J., Zhang, W., Simultaneous Confidence Bands and Hypothesis Testing in Varying-coefficient Models, 27, 715-731, **2000**.
- [11] Chiang, C.T., Rice, J.A., Wu, C.O., Smoothing Spline Estimation for varying coefficient models with repeatedly measured dependent variables, *Journal of the American Statistical Association*, 96, 605-619, **2001**.
- [12] Huang J.Z., Wu, C.O., Zhou, L., Varying-coefficient Models and Basis Function Approximations for the Analysis of Repeated Measurements, *Biometrika*, 89, 111-128, **2002**.
- [13] Cai, Z., Sun, Y., Local Linear Estimation for Time-Dependent Coefficients in Cox's Regression Models, *Scandinavian Journal of Statistics*, **2003**.
- [14] Tian, L., Zucker, D., Wei, L.J., On the Cox Model with Time-varying Regression Coefficients, *Journal of the American Statistical Association*, 100, 172-183, **2005**.
- [15] Fan, J., Huang, T., Profile Likelihood Inferences on Semiparametric Varying-coefficient Partially Linear Models, *Bernoulli*, 11, 1031-1057, **2005**.
- [16] Cai, Z., Tiwari, R.C., Application of a Local Linear Autoregressive Model to BOD Time Series, *Environmetrics: The Official Journal of the International Environmetrics Society*, 11(3), 341-350, **2000**.
- [17] Türkan, S., *Yarı Parametrik Regresyon Modelinde Etkili Gözlem Analizi*, Doktora Tezi, Hacettepe Üniversitesi Fen Bilimleri Enstitüsü, Ankara, **2012**.

- [18] Wang, Y., *Varying-Coefficient Models: New Models, Inference Procedures and Applications*, Doktora Tezi, The Pennsylvania University, **2007**.
- [19] Nadaraya, E. A., On Estimating Regression, *Theory of Probability and Its Applications*, 9, 141-142, **1964**.
- [20] Watson, G. S., Smooth Regression Analysis, *Sankhya A*, 26, 359-372, **1964**.
- [21] Gasser, T., Müller, H. G., *Kernel Estimation of Regression Functions, Smoothing Techniques for Curve Estimation*, Lecture Notes in Mathematics 757, 23-68, **1979**.
- [22] Gasser, T., Müller, H. G., Estimating Regression Functions and Their Derivatives by the Kernel Method, *Scandinavian Journal of Statistics*, 11(3), 171-185, **1984**.
- [23] Cleveland, W. S., Robust locally weighted regression and Smoothing scatterplots, *Journal of the American Statistical Association*, 74, 829-836, **1979**.
- [24] Cleveland, W. S., Devlin, S. J., Locally weighted regression: An approach to regression analysis by local fitting, *Journal of the American Statistical Association*, 83(403), 596-610, **1988**.
- [25] Stone, C.J., Consistent Nonparametric Regression, *Annual Statistics*, 5, 595-645, **1977**.
- [26] Eubank, R. L., *Spline Smoothing and Nonparametric Regression*, Marcel Dekker, New York, **1988**.
- [27] Wahba, G., *Spline Models for Observational Data*, Society for Industrial and Applied Mathematics, Philadelphia, **1990**.
- [28] Green, P. J., Silverman, B. W., *Nonparametric Regression and Generalized Linear Models: a Roughness Penalty Approach*, Chapman & Hall, London, **1994**.
- [29] Smith, P. L., *Curve fitting and modeling with splines using statistical variable selection methods*, NASA Report 166034, NASA, Langley Research Center, Hampton, VA, **1982**.
- [30] Brieman, L., Friedman, J.H., Olshen, R.A., Stone, C.J., *CART: Classification and Regression Trees (first edition, 1984)*, Wadsworth, Belmont, **1993**.
- [31] Wittaker, E.T., On a new method of graduation, *Proceedings of the Edinburgh Mathematical Society*, 41, 63-75, **1923**.
- [32] Wahba, G., Smoothing noisy data with spline functions, *Numberische Mathematik*, 24, 383-393, **1975**.
- [33] Kooperberg, C., Stone, C.J., Truong, Y.K., Hazard Regression, *Journal of the American Statistical Association*, 90, 78-94, **1995a**.
- [34] Kooperberg, C., Stone, C.J., Truong, Y.K., Log-spline estimation of a possibly mixed spectral distribution, *Journal Time Series Analysis*, 16, 359-388, **1995b**.
- [35] Mallows, C.L., Some comments on C_p , *Technometrics*, 15, 661-675, **1973**.
- [36] Kooperberg, C., Stone, C.J., A Study of Log-spline Density Estimation, *Computational Statistics and Data Analysis*, 12, 327-347, **1991**.
- [37] Silverman, B.W., Spline smoothing: the equivalent variable kernel method, *Annual Statistics*, 12, 898-916, **1984**.

- [38] Silverman, B.W, Some aspects of the Spline Smoothing approach to nonparametric regression curve fitting (with discussion), *Journal of the Royal Statistical Society*, B, 47, 1-52, **1985**.
- [39] Messer, K., A comparison of a spline estimate to equivalent kernel estimate, *Annual Statistics*, 19, 817-829, **1991**.
- [40] Wahba, G., A survey of some Smoothing problems and the method of generalized cross-validation for solving them, *Applications of Statistics (P.R. Krisnaiah, ed.)*, 507-523, North Holland, Amsterdam, **1977**.
- [41] Craven, P., Wahba, G., Smoothing noisy data with Spline functions: estimating the correct degree of S smoothing by the method of generalized cross-validation, *Numberische Mathematik*, 31, 377-403, **1979**.
- [42] Wahba, G., Wang, Y., When is the optimal regularization parameter insensitive to the choice of the loss function?, *Communications in Statistics*, 5, 1685-1700, **1990**.
- [43] O'Sullivan, F., A statistical perspective on ill-posed inverse problems (with discussion), *Statistical Science*, 1, 505-527, **1986**.
- [44] O'Sullivan, F., Fast computation of fully automated logdensity and log-hazard estimators, *Society for Industrial and Applied Mathematics Journal on Scientific and Statistical Computing*, 9, 363-379, **1988**.
- [45] Eilers, P. H. C., Marx, B. D., Flexible Smoothing with B-Splines and Penalties, *Statistical Science*, 11(2), 89-121, **1996**.
- [46] De Boor, C., *A Practical Guide to Splines*, Springer, Berlin, **1978**.
- [47] Reinsch, C., Smoothing by Spline functions, *Numberische Mathematik*, 10, 177-183, **1967**.
- [48] Fan, J., Zhang, W., Statistical Estimation in Varying Coefficient Models, *The Annals of Statistics*, 27(5), 1491-1518, **1999**.
- [49] Huang, J. Z., Wu, C. O., Zhou, L., Polynomial Spline Estimation and Inference for Varying Coefficient Models with Longitudinal Data, *Statistica Sinica*, 14, 763-788, **2004**.
- [50] Huang, T., *New Developments in Varying-Coefficient Partially Linear Models*, Doktor Tezi, University of North Carolina at Chapel Hill, **2004**.
- [51] Sterrantino, A. F., *Varying Coefficient Models as Mixed Models: Reparametrization Methods and Bayesian Estimation*, Alma Mater Studiorum Universita di Bologna, **2013**.
- [52] Park, B.U., Mammen, E., Lee, Y.K., Lee, E.R., Varying Coefficient Regression Models: A Review and New Developments, *International Statistical Review*, 83(1), 36-64, **2015**.
- [53] Assaf, S., Campostrini, S., Application of the Varying Coefficient Model to the Behaviour Risk Factor Surveillance Data in Italy: A Study of Changing Smoking Prevalence Among Sub-populations, *BMC Public Health*, **2015**.
- [54] Ruppert, D., Selecting The Number of Knots for Penalized Splines, *Journal of Computational and Graphical Statistics*, 11(4), 735-757, **2002**.

- [55] Lu, Y., Zhang, R., Zhu, L., Penalized Spline Estimation for Varying-Coefficient Models, *Communications in Statistics – Theory and Methods*, 37(14), 2249-2261, **2008**.
- [56] Zhang, W., Lee, S.Y., Variable Bandwidth Selection in Varying-Coefficient Models, *Journal of Multivariate Analysis*, 74, 116-134, **2000**.
- [57] Gu, C., Wahba, G., Minimizing GCV/GML Scores with Multiple Smoothing Parameters via the Netwon Method, *Society for Industrial and Applied Mathematics Journal on Scientific and Statistical Computing*, 12, 383–398, **1991**.
- [58] Wood, S. N., Modelling and Smoothing Parameter Estimation with Multiple Quadratic Penalties, *Journal of the Royal Statistical Society, B*, 62, 413–428, **2000**.
- [59] Wood, S. N., Stable and Efficient Multiple Smoothing Parameter Estimation for Generalized Additive Models, *Journal of the American Statistical Association*, 99, 673–686, **2004**.
- [60] Cheng, W., Yang, L., Zhao, H., Wang, P., Yan, J., Varying Coefficient Models for Analyzing the Effects of Risk Factors on Pregnant Women’s Blood Pressure, *13th International Conference on Machine Learning and Applications*, December, Detroit, MI USA, **2014**.
- [61] Anonim, Mean arterial pressure (MAP) Calculator, <http://www.globalrph.com/map.htm> (Agustos, **2018**).
- [62] Cnossen, J.S., Vollebregt, K.C., Vrieze, N., Riet, G., Mol, B.W.J, Franx, A., Khan, K.S., Post, J.A.M., Accuracy of Mean Arterial Pressure and Blood Pressure Measurements in Predicting Pre-eclampsia: Systematic Review and Meta-Analysis, *British Medical Journal*, Mayis, **2008**.

ÖZGEÇMİŞ

Kimlik Bilgileri

Adı Soyadı : Işıl DURMUŞ

Doğum Yeri : Adıyaman

Medeni Hali : Bekar

E-posta : isil.durmus@tuik.gov.tr

Adresi : Türkiye İstatistik Kurumu, Necatibey Cad. No:114 Çankaya / ANKARA

Eğitim

Lise : Fethiye Kemal Mumcu Anadolu Lisesi

Lisans : Hacettepe Üniversitesi, İstatistik Bölümü

Yabancı Dil ve Düzeyi

İngilizce : 85 (YDS)

İş Deneyimi

Ekim 2015- : Türkiye İstatistik Kurumu Uzman Yardımcısı

Deneyim Alanları

-

Tezden Üretilmiş Projeler ve Bütçesi

-

Tezden Üretilmiş Yayınlar

-

Tezden Üretilmiş Tebliğ ve/veya Poster Sunumu ile Katıldığı Toplantılar

-



HACETTEPE ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
YÜKSEK LİSANS/DOKTORA TEZ ÇALIŞMASI ORJİNALLİK RAPORU

HACETTEPE ÜNİVERSİTESİ
FEN BİLİMLER ENSTİTÜSÜ
İSTATİSTİK ANABİLİM DALI BAŞKANLIĞI'NA

Tarih: 23/09/2018

Tez Başlığı: DEĞİŞEN KATSAYILI REGRESYON MODELİ İLE GEBELERİN ORTALAMA ARTERİYEL KAN BASINCINA ETKİ EDEN RİSK FAKTÖRLERİNİN BELİRLENMESİ

Yukarıda başlığı/konusu gösterilen tez çalışmamın a) Kapak sayfası, b) Giriş, c) Ana bölümler d) Sonuç kısımlarından oluşan toplam 62 sayfalık kısmına ilişkin, 23/09/2018 tarihinde tez ~~şahsım~~/danışmanım tarafından Turnitin adlı intihal tespit programından aşağıda belirtilen filtrelemeler uygulanarak alınmış olan orijinallik raporuna göre, tezimin benzerlik oranı % 7 'dir.

Uygulanan filtrelemeler:

- 1- Kaynakça hariç
- 2- Alıntılar hariç/dahil
- 3- 5 kelimedenden daha az örtüşme içeren metin kısımları hariç

Hacettepe Üniversitesi Fen Bilimleri Enstitüsü Tez Çalışması Orjinallik Raporu Alınması ve Kullanılması Uygulama Esasları'nı inceledim ve bu Uygulama Esasları'nda belirtilen azami benzerlik oranlarına göre tez çalışmamın herhangi bir intihal içermediğini; aksinin tespit edileceği muhtemel durumda doğabilecek her türlü hukuki sorumluluğu kabul ettiğimi ve yukarıda vermiş olduğum bilgilerin doğru olduğunu beyan ederim.

Gereğini saygılarımla arz ederim.

23/09/2018

Adı Soyadı: İŞİL DURMUŞ
Öğrenci No: N14129385
Anabilim Dalı: İSTATİSTİK
Programı: TEZLİ YÜKSEK LİSANS
Statüsü: Y.Lisans Doktora Bütünleşik Dr.

DANIŞMAN ONAYI

UYGUNDUR.

DOÇ. DR. SEMRA TÜRKAN