

**PARAMETRİK OLMAYAN BULANIK
REGRESYON MODELLERİ ANALİZİ**

Münevvere YILDIZ
Doktora Tezi

İstatistik Anabilim Dalı
Mayıs – 2013

**Bu tez çalışması Anadolu Üniversitesi Bilimsel Araştırma Projeleri
Komisyonu Başkanlığı tarafından desteklenmiştir. Proje No: 1202F031**



JÜRİ VE ENSTİTÜ ONAYI

Münevvere Yıldız'ın "Parametrik Olmayan Bulanık Regresyon Modelleri Analizi" başlıklı İstatistik Anabilim Dalındaki, Doktora Tezi 10.05.2013 tarihinde, aşağıdaki jüri tarafından Anadolu Üniversitesi Lisansüstü Eğitim - Öğretim ve Sınav Yönetmeliğinin ilgili maddeleri uyarınca değerlendirilerek kabul edilmiştir.

	Adı Soyadı	İmza
Üye (Tez Danışmanı)	: Prof. Dr. Memmedağa Memmedli
Üye	: Prof. Dr. Tahir Khaniyev
Üye	: Prof. Dr. Efendi Nasiboğlu
Üye	: Doç. Dr. Sevil Şentürk
Üye	: Doç. Dr. Nihal Erginel

Anadolu Üniversitesi Fen Bilimleri Enstitüsü Yönetim Kurulu'nun..... tarih ve sayılı kararıyla onaylanmıştır.

Enstitü Müdürü



ÖZET

Doktora Tezi

PARAMETRİK OLMAYAN BULANIK REGRESYON MODELLERİ ANALİZİ

Münevvere YILDIZ

**Anadolu Üniversitesi
Fen Bilimleri Enstitüsü
İstatistik Anabilim Dalı**

**Danışman : Prof. Dr. Memmedağa Memmedli
İkinci Danışman : Yard. Doç. Dr. Rabia Ece Omay
2013, 116 sayfa**

Tez çalışmasında, parametrik olmayan bulanık regresyon modelleri incelenmiştir. k -en yakın komşuluk, çekirdek düzeltme ve yerel polinomial düzeltme modelleri bulanık yapıda ifade edilmiştir. Ayrıca bu modeller için, band genişliği seçiminde çapraz geçerlilik ve genelleştirilmiş çapraz geçerlilik kriterleri, şapka matrisi kullanılarak geliştirilmiştir.

Çalışmalarda kullanılan verilerde, yanıt değişkenin bulanık değerli ve açıklayıcı değişkenin kesin değerli olduğu durum dikkate alınmıştır. Analizler R programında locpol paketi kullanılarak yapılmıştır.

Uygulamalarda özellikle, bulanık yerel polinomial regresyon modelinde polinomun derecesinin doğrusal ve kübik alındığı durumlar üzerinde inceleme yapılmıştır. Bu modeller için band genişliği, geliştirilen çapraz geçerlilik ve genelleştirilmiş çapraz geçerlilik kriterleri ile seçilerek, modellerin performansları ortalama karesel hata değerleri kullanılarak karşılaştırılmıştır. Elde edilen sonuçlardan genelleştirme yapabilmek amacıyla farklı türdeki veri setleri üzerinde çalışılmıştır.

Uygulamaların pek çoğunda bulanık yerel kübik modele ait performans değerleri daha yüksek olmasına rağmen, özellikle eğriselliği fazla olan modelleri daha pürüzsüz (dalgalanması az olan) bir şekilde ifade eder. Ayrıca, band genişliği değerinin bulanık yerel doğrusal modele göre daha yüksek seçilmesi ile işlem basamaklarını azalttığı göz önüne alındığında, bulanık yerel kübik modellerin kullanımının daha faydalı olduğu sonucuna varılmıştır.

Anahtar Kelimeler: Parametrik olmayan bulanık regresyon modelleri, k -en yakın komşuluk, Çekirdek düzeltme, Yerel polinomial düzeltme, Çapraz geçerlilik, Genelleştirilmiş çapraz geçerlilik

ABSTRACT

PhD Dissertation

ANALYSIS OF NONPARAMETRIC FUZZY REGRESSION MODELS

Münevvere YILDIZ

Anadolu University
Graduate School of Sciences
Statistical Department

Supervisor : Prof. Dr. Memmedağa Memmedli
Second Supervisor : Assist. Prof. Dr. Rabia Ece Omay
2013, 116 pages

In this thesis, the nonparametric fuzzy regression models are examined. k -nearest neighbour, kernel smoothing and local polynomial smoothing models are expressed in fuzzy structure. Furthermore to these models, cross-validation and generalized cross-validation criteria are developed for bandwidth selection by using the hat matrix.

The data used in the studies, the situation is taken into account as to be the response variable is valued fuzzy and the explanatory variable is valued crisp. The locpol package in R program is used for analysis.

In applications especially the cases, that the degree of the polynomial is linear and cubic in fuzzy local polynomial regression model, are examined. The bandwidth for these models is selected by developed cross validation and generalized cross validation criteria, the performance of the models are compared using averaged squared error values. Studied on different types of data sets in order to be able to generalize the obtained results.

In many applications it is observed that models, especially has more curvature, are expressed more smoothly (less fluctuating) although the fuzzy local cubic model's performance values are higher. Additionally, it is seen that steps of operation are reduced by selecting the higher bandwidth value than fuzzy local linear model, so it is concluded that the usage of fuzzy local cubic models gives better results.

Keywords: Nonparametric fuzzy regression models, k - nearest neighbour, Kernel smoothing, Local polynomial smoothing, Cross validation, Generalized cross validation

TEŞEKKÜR

Doktora çalışmalarım boyunca bilgi ve deneyimlerini benimle paylaşan,
yönlendiriciliğini ve desteğini esirgemeyen, tez çalışmamın ortaya çıkmasında
büyük emek sarfeden çok değerli hocam

Prof.Dr. Memmedağa Memmedli'ye,

R Programında gerçekleştirilen uygulamalarda emeği geçen ve yardımını
esirgemeyen ikinci danışmanım

Yard.Doç.Dr. Rabia Ece Omay'a,

Tüm tez izlemelerim boyunca Ankara'dan gelerek zahmet çeken ve yapmış
olduğu eleştiriler, vermiş olduğu öneriler ile tez çalışmama katkıda bulunan çok
kıymetli hocam

Prof.Dr. Tahir Khaniyev'e,

Ayrıca tez izlemelerimde yer alan, pek çok konuda desteğini gönülden
hissettiğim, her sınav öncesi motivasyon kaynağım olan değerli hocam

Doç.Dr. Sevil Şentürk,

Tez savunma jürimde yer alan ve kıymetli zamanlarını ayırıp tezimi okuyarak
yapmış olduğum çalışmalara daha eleştirel bakabilmemi sağlayan hocalarım,

Prof.Dr. Memmedağa Memmedli, Prof.Dr. Tahir Khaniyev,

Prof.Dr. Efendi Nasiboğlu, Doç.Dr. Sevil Şentürk ve Doç.Dr. Nihal Erginel'e,

Çalışmalarım boyunca pek çok konuda bana destek olan
Anadolu Üniversitesi İstatistik Bölümü hocalarım,

Yard.Doç.Dr. Çiğdem Giriftinoğlu,

Öğr.Gör. Özer Özdemir'e,

R programımı öğrenmem konusunda bilgilerini benimle paylaşan

Akhlitdin Nizamitdinov'a,

Doktora eğitimim boyunca özveri ve desteğini bir an olsun eksik etmeyen
sevgili eşim ***Hüseyin Yıldız***'a,

teşekkür ederim...

Münevvere Yıldız

Mayıs 2013

İÇİNDEKİLER

	<u>Sayfa</u>
ÖZET.....	i
ABSTRACT	ii
TEŞEKKÜR	iii
İÇİNDEKİLER	iv
ÇİZELGELER DİZİNİ	vii
ŞEKİLLER DİZİNİ	x
SİMGELER ve KISALTMALAR DİZİNİ	xiii
1. GİRİŞ	1
2. PARAMETRİK OLMAYAN REGRESYON MODELLERİ	5
2.1. Parametrik Olmayan Regresyon Problemi.....	6
2.1.1. Pürüzlülük ceza yaklaşımı ve splayn düzeltme tekniği.....	7
2.2. Regresyonda Düzeltme Kavramı ve Düzeltme Teknikleri	12
2.2.1. Çekirdek düzeltme	13
2.2.2. k -en yakın komşuluk tahmini	15
2.2.3. Yerel polinomial düzeltme.....	16
2.2.3.1. Etki, Varyans ve Serbestlik Derecesi	19
2.2.3.2. Yerel polinomun derecesi.....	24
2.2.3.3. Ağırlık fonksiyonu	27
2.3. Parametrik Olmayan Regresyonda Düzeltme Parametresinin Seçimi	27
2.3.1. Çapraz geçerlilik kriteri.....	28
2.3.2. Genelleştirilmiş çapraz geçerlilik kriteri.....	29
2.3.3. Geliştirilmiş Akaike Bilgi kriteri	30

2.4. Parametrik Olmayan Regresyon Modellerinde Performans Kriterleri ... 31

3. BULANIK FONKSİYONLAR VE TÜREV 33

3.1. Bulanık Küme ve Bulanık Sayılar 33

3.1.1. Üçgen bulanık sayılar..... 37

3.1.2. Yamuk bulanık sayılar 38

3.1.3. *LR* gösterimli bulanık sayılar 39

3.2. Genişletme prensibi 40

3.3. Bulanık Sıralama..... 41

3.4. Bulanık Fonksiyonlar 44

3.5. Bulanık Fonksiyonun Türevi 47

3.5.1. Bulanık olmayan kümede tanımlanan bulanık fonksiyonun türevi.. 47

3.5.2. *LR* bulanık fonksiyonun türevi..... 48

4. PARAMETRİK OLMAYAN BULANIK REGRESYON MODELLERİ 50

4.1. Parametrik Olmayan Bulanık Regresyon Modellerinin Genel Yapısı... 52

4.1.1. Bulanık *k*-en yakın komşuluk tahmini..... 53

4.1.2. Bulanık çekirdek regresyon tahmini 54

4.1.3. Bulanık yerel polinomiyal regresyon tahmini..... 55

4.2. Parametrik Olmayan Bulanık Regresyon Modellerinde Düzeltme

Parametresi Seçim Kriterlerinin Geliştirilmesi 59

4.2.1. Bulanık çapraz geçerlilik kriterinin geliştirilmesi..... 60

4.2.2. Bulanık genelleştirilmiş çapraz geçerlilik kriterinin geliştirilmesi .. 63

4.3. Parametrik Olmayan Bulanık Regresyon Modellerinde Performans

Kriteri 64

5. SİMÜLASYON ÇALIŞMASI	66
5.1. Veri Seti-1 ve Veri Seti-2 İçin Parametrik Olmayan Bulanık Regresyon Modellerinin Karşılaştırılması	66
5.2. Veri Seti-3 İçin İki Farklı Durum Dikkate Alınarak Parametrik Olmayan Bulanık Yerel Doğrusal ve Yerel Kübik Modellerin Karşılaştırılması	80
5.3. Veri Seti-4 İçin Girdi Değişkeninin İki Farklı Durumunda Parametrik Olmayan Bulanık Yerel Polinomial Regresyon Modeli Tahmini	88
6. GERÇEK VERİ SETLERİ ÜZERİNDE UYGULAMALAR	97
6.1. Lee ve Tanaka (1999) Veri Seti İçin Parametrik Olmayan Bulanık Yerel Polinomial Regresyon Modeli Tahmini	97
6.2. D'Urso ve Gastaldi (2002) Veri Seti İçin Parametrik Olmayan Bulanık Yerel Polinomial Regresyon Modeli Tahmini	102
7. SONUÇ ve ÖNERİLER	108
KAYNAKLAR	111

ÇİZELGELER DİZİNİ

2.1. Çekirdek Fonksiyonu Türleri	14
2.2. p dereceli polinom ve çeşitli çekirdek fonksiyonları için değişkenlikteki artış (Fan ve Gijbels 1996).....	25
5.1. Veri Seti-1 için Gauss çekirdeği kullanıldığında çekirdek düzeltme, $p = 1$ ve $p = 3$ için CV ve GCV kriterleri ile band genişliği seçimi.....	68
5.2. Veri Seti-1 için Epanechnikov çekirdeği kullanıldığında çekirdek düzeltme, $p = 1$ ve $p = 3$ için CV ve GCV kriterleri ile band genişliği seçimi.....	69
5.3. Veri Seti-2 için Gauss çekirdeği kullanıldığında çekirdek düzeltme, $p = 1$ ve $p = 3$ için CV ve GCV kriterleri ile band genişliği seçimi.....	73
5.4. Veri Seti-2 için Epanechnikov çekirdeği kullanıldığında çekirdek düzeltme, $p = 1$ ve $p = 3$ için CV ve GCV kriterleri ile band genişliği seçimi.....	73
5.5. Veri Seti-1 ve Veri Seti-2 için Gauss ve Epanechnikov çekirdekleri kullanılarak çekirdek düzeltme, $p = 1$ ve $p = 3$ için geliştirilen CV ve GCV kriterleri ve Wang ve ark. (2007) çalışmasındaki CV kriteri ile seçilen band genişlikleri.....	74
5.6. Veri Seti-1 için farklı çekirdekler kullanılarak çekirdek düzeltme, $p = 1$ ve $p = 3$ için geliştirilen CV ve GCV kriterleri ve Wang ve ark. (2007) çalışmasındaki CV kriteri ile seçilen band genişliklerine göre model performansları.....	78
5.7. Veri Seti-2 için farklı çekirdekler kullanılarak çekirdek düzeltme, $p = 1$ ve $p = 3$ için geliştirilen CV ve GCV kriterleri ve Wang ve ark. (2007) çalışmasındaki CV kriteri ile seçilen band genişliklerine göre model performansları.....	78
5.8. Veri Seti-3 için Gauss çekirdeği kullanıldığında $p = 1$ ve $p = 3$ için CV ve GCV kriterleri ile band genişliği seçimi.....	81
5.9. Veri Seti-3 için Epanechnikov çekirdeği kullanıldığında $p = 1$ ve $p = 3$ için CV ve GCV kriterleri ile band genişliği seçimi.....	81
5.10. Veri Seti-3 için farklı çekirdekler kullanılarak $p = 1$ ve $p = 3$ için CV ve GCV kriterleri ile seçilen band genişliklerine göre model performansları....	82

5.11. Veri Seti-3 (Durum-2) için Gauss çekirdeği kullanıldığında $p = 1$ ve $p = 3$ için CV ve GCV kriterleri ile band genişliği seçimi.....	85
5.12. Veri Seti-3 (Durum-2) için Epanechnikov çekirdeği kullanıldığında $p = 1$ ve $p = 3$ için CV ve GCV kriterleri ile band genişliği seçimi.....	86
5.13. Veri Seti-3 (Durum-2) için farklı çekirdekler kullanılarak $p = 1$ ve $p = 3$ için CV ve GCV kriterleri ile seçilen band genişliklerine göre model performansları.....	86
5.14. Veri Seti-4 için Gauss çekirdeği kullanıldığında $p = 1$ ve $p = 3$ için CV ve GCV kriterleri ile band genişliği seçimi.....	89
5.15. Veri Seti-4 için Epanechnikov çekirdeği kullanıldığında $p = 1$ ve $p = 3$ için CV ve GCV kriterleri ile band genişliği seçimi.....	89
5.16. Veri Seti-4 için farklı çekirdekler kullanılarak $p = 1$ ve $p = 3$ için CV ve GCV kriterleri ile seçilen band genişliklerine göre model performansları..	90
5.17. Veri Seti-4 (Durum-2) için Gauss çekirdeği kullanıldığında $p = 1$ ve $p = 3$ için CV ve GCV kriterleri ile band genişliği seçimi.....	94
5.18. Veri Seti-4 (Durum-2) için Epanechnikov çekirdeği kullanıldığında $p = 1$ ve $p = 3$ için CV ve GCV kriterleri ile band genişliği seçimi.....	94
5.19. Veri Seti-4 (Durum-2) için farklı çekirdekler kullanılarak $p = 1$ ve $p = 3$ için CV ve GCV kriterleri ile seçile band genişliklerine göre model performansları.....	94
6.1. Lee ve Tanaka (1999) Veri Seti.....	98
6.2. Lee ve Tanaka (1999) veri seti için Gauss çekirdeği kullanıldığında $p = 1$ ve $p = 3$ için CV ve GCV kriterleri ile band genişliği seçimi.....	99
6.3. Lee ve Tanaka (1999) veri seti için Epanechnikov çekirdeği kullanıldığında $p = 1$ ve $p = 3$ için CV ve GCV kriterleri band genişliği seçimi.....	99
6.4. Lee ve Tanaka (1999) veri seti için farklı çekirdekler kullanılarak $p = 1$ ve $p = 3$ için CV ve GCV kriterleri ve Wang ve ark. (2007) kriteri ile seçilen band genişliklerine göre model performansları.....	100
6.5. D'Urso & Gastaldi (2002) Veri Seti.....	103
6.6. D'Urso&Gastaldi (2002) veri seti için Gauss çekirdeği kullanıldığında $p = 1$ ve $p = 3$ için CV ve GCV kriterleri ile band genişliği seçimi.....	104

- 6.7. D'Urso&Gastaldi (2002) veri seti için Epanechniov çekirdeği
kullanıldığında $p = 1$ ve $p = 3$ için CV ve GCV kriterleri ile band
genişliği seçimi.....104
- 6.8. D'Urso&Gastaldi (2002) veri seti için farklı çekirdekler kullanılarak
 $p = 1$ ve $p = 3$ için geliştirilen CV ve GCV kriterleri ve Wang ve ark.
(2007)'nin kullandıkları CV kriteri ile seçilen band genişliklerine
göre model performansları.....105

ŞEKİLLER DİZİNİ

2.1. Motorsiklet veri seti için $x \pm 3.3$ veri noktalarında yerel doğrusal uyumlamanın gösterimi.....	19
2.2. p dereceli polinom ve çeşitli çekirdek fonksiyonları için değişkenlikteki artış grafiği (Fan ve Gijbels 1995).....	26
3.1. Bulanık kümelerde kararlılık prensibi.....	35
3.2. Konveks ve normal bulanık küme.....	36
3.3. Bulanık sayı.....	37
3.4. Üçgen Bulanık Sayı.....	38
3.5. Yamuk Bulanık Sayı.....	39
3.6. LR Bulanık Sayısı.....	40
3.7. \tilde{f} fonksiyonunun α -seviye eğrileri.....	47
3.8. LR bulanık fonksiyonunun türevi.....	49
5.1 Veri Seti-1'de çekirdek düzeltme ve Gauss çekirdeği kullanıldığında CV kriteri ile belirlenen $h = 0.2$ değeri için uyum eğrisi.....	70
5.2. Veri Seti-1'de $p = 1$ ve Gauss çekirdeği kullanıldığında CV kriteri ile belirlenen $h = 0.5$ değeri için uyum eğrisi.....	70
5.3. Veri Seti-1'de $p = 3$ ve Gauss çekirdeği kullanıldığında CV kriteri ile belirlenen $h = 1.3$ değeri için uyum eğrisi.....	71
5.4. Veri Seti-1'de çekirdek düzeltme ve Epanechnikov çekirdeği kullanıldığında CV kriteri ile belirlenen $h = 0.42$ değeri için uyum eğrisi.....	71
5.5. Veri Seti-1'de $p = 1$ ve Epanechnikov çekirdeği kullanıldığında CV kriteri ile belirlenen $h = 1.0$ değeri için uyum eğrisi.....	72
5.6. Veri Seti-1'de $p = 3$ ve Epanechnikov çekirdeği kullanıldığında CV kriteri ile belirlenen $h = 2.5$ değeri için uyum eğrisi.....	72
5.7. Veri Seti-2 için çekirdek düzeltme ve Gauss çekirdeği kullanıldığında CV kriteri ile belirlenen $h = 0.13$ için uyum eğrisi.....	74
5.8. Veri Seti-2 için $p = 1$ ve Gauss çekirdeği kullanıldığında CV kriteri ile belirlenen $h = 0.21$ için uyum eğrisi.....	75
5.9. Veri Seti-2 için $p = 3$ ve Gauss çekirdeği kullanıldığında CV kriteri ile belirlenen $h = 0.6$ için uyum eğrisi.....	75

5.10. Veri Seti-2 için çekirdek düzeltme ve Epanechnikov çekirdeği kullanıldığında CV kriteri ile belirlenen $h = 0.32$ için uyum eğrisi.....	76
5.11. Veri Seti-2 için $p = 1$ ve Epanechnikov çekirdeği kullanıldığında CV kriteri ile belirlenen $h = 0.5$ için uyum eğrisi.....	77
5.12. Veri Seti-2 için $p = 3$ ve Epanechnikov çekirdeği kullanıldığında CV kriteri ile belirlenen $h = 1.8$ için uyum eğrisi.....	77
5.13. Veri Seti-3 için $p = 1$ ve Gauss çekirdeği kullanıldığında CV kriteri ile belirlenen $h = 0.15$ için uyum eğrisi.....	82
5.14. Veri Seti-3 için $p = 3$ ve Gauss çekirdeği kullanıldığında CV kriteri ile belirlenen $h = 0.32$ için uyum eğrisi.....	83
5.15. Veri Seti-3 için $p = 1$ ve Epanechnikov çekirdeği kullanıldığında CV kriteri ile belirlenen $h = 0.38$ için uyum eğrisi.....	84
5.16. Veri Seti-3 için $p = 3$ ve Epanechnikov çekirdeği kullanıldığında CV kriteri ile belirlenen $h = 1.02$ için uyum eğrisi.....	84
5.17. Veri Seti-3 (Durum-2) için Gauss çekirdeği kullanılarak CV kriteri ile seçilen band genişliği kullanılarak bulanık yerel doğrusal ve bulanık yerel kübik modeller için uyum eğrileri.....	87
5.18. Veri Seti-3 (Durum-2) için Epanechnikov çekirdeği kullanılarak CV kriteri ile seçilen band genişliği kullanılarak bulanık yerel doğrusal ve bulanık yerel kübik modeller için uyum eğrileri.....	87
5.19. Veri Seti-4 için $p = 1$ ve Gauss çekirdeği kullanıldığında CV kriteri ile seçilen $h = 0.6$ için uyum eğrisi.....	90
5.20. Veri Seti-4 için $p = 3$ ve Gauss çekirdeği kullanıldığında CV kriteri ile seçilen $h = 1.2$ için uyum eğrisi.....	91
5.21. Veri Seti-4 için $p = 1$ ve Epanechnikov çekirdeği kullanıldığında CV kriteri ile seçilen $h = 1.2$ için uyum eğrisi.....	92
5.22. Veri Seti-4 için $p = 3$ ve Epanechnikov çekirdeği kullanıldığında CV kriteri ile seçilen $h = 2.5$ için uyum eğrisi.....	92
5.23. Veri Seti-4 Durum-2 veri seti için Gauss çekirdeği kullanılarak CV kriteri ile seçilen band genişliği kullanılarak bulanık yerel doğrusal ve bulanık yerel kübik modeller için uyum eğrileri.....	95

5.24. Veri Seti-4 Durum-2 için Epanechnikov çekirdeği kullanılarak <i>CV</i> kriteri ile seçilen band genişliği kullanılarak bulanık yerel doğrusal ve bulanık yerel kübik modeller için uyum eğrileri.....	96
6.1. Lee ve Tanaka (1999) veri seti için $p = 1$ ve Gauss çekirdeği kullanıldığında <i>CV</i> kriteri ile seçilen band genişliği $h = 0.77$ için uyum eğrileri.....	100
6.2. Lee ve Tanaka (1999) veri seti için $p = 3$ ve Gauss çekirdeği kullanıldığında <i>CV</i> kriteri ile seçilen band genişliği $h = 4.5$ için uyum eğrileri.....	101
6.3. Lee ve Tanaka (1999) veri seti için $p = 1$ ve Epanechnikov çekirdeği kullanıldığında <i>CV</i> kriteri ile seçilen band genişliği $h = 1.75$ için uyum eğrileri.....	101
6.4. Lee ve Tanaka (1999) veri seti için $p = 3$ ve Epanechnikov çekirdeği kullanıldığında <i>CV</i> kriteri ile seçilen band genişliği $h = 7.0$ için uyum eğrileri.....	102
6.5. D'Urso&Gastaldi (2002) veri seti için $p = 1$ ve Gauss çekirdeği kullanıldığında <i>CV</i> kriteri ile seçilen $h = 0.66$ için uyum eğrileri.....	105
6.6. D'Urso&Gastaldi (2002) veri seti için $p = 3$ ve Gauss çekirdeği kullanıldığında <i>CV</i> kriteri ile seçilen $h = 2.0$ için uyum eğrileri.....	106
6.7. D'Urso&Gastaldi (2002) veri seti için $p = 1$ ve Epanechnikov çekirdeği kullanıldığında <i>CV</i> kriteri ile seçilen $h = 2.01$ için uyum eğrileri.....	106
6.8. D'Urso&Gastaldi veri seti için $p = 3$ ve Epanechnikov çekirdeği kullanıldığında <i>CV</i> kriteri ile seçilen $h = 3.9$ için uyum eğrileri.....	107

SİMGELER ve KISALTMALAR DİZİNİ

Simgeler:

λ, h	Band genişliği parametresi
$K(.)$	Çekirdek fonksiyonu
$W(.)$	Ağırlık dizini
L, H	Şapka matrisi

Kısaltmalar:

CV	Çapraz geçerlilik kriteri
GCV	Genelleştirilmiş çapraz geçerlilik kriteri
AIC_c	Geliştirilmiş Akaike bilgi kriteri
MSE	Hata kareler ortalaması (Mean squared error)
$MISE$	Entegre hata kareler toplamı (Mean integrated squared error)
ASE	Ortalama karesel hata (Averaged squared error)
$MASE$	Beklenen ortalama karesel hata

1. GİRİŞ

Bir regresyon eğrisi, bir veya birkaç açıklayıcı değişken X ile yanıt değişkeni Y arasındaki genel ilişkiyi tanımlar. Regresyon analizinin temel amacı, bilinmeyen regresyon fonksiyonu için uygun bir tahmin üretmektir. Bu noktada, parametrik ve parametrik olmayan tahminlerden faydalanılır. Parametrik regresyon yaklaşımı, açıklayıcı değişken ve cevap değişkeni arasındaki ilişkinin fonksiyonel şeklinin bilindiği durumlarda kullanılır. Ancak parametrik regresyon tahminleri çeşitli varsayımlara dayalı olarak gerçekleştirilirler ve bu varsayımların sağlanmadığı durumlarda yapılan tahminler iyi bir tahmin olma özelliğine sahip olamamaktadırlar. Parametrik yaklaşımlardaki bu sorunlar nedeniyle varsayımların esnetilmesine olanak sağlayan yaklaşımlara ihtiyaç duyulmuştur. Parametrik olmayan ve yarı parametrik yaklaşımlar bunlara örnektir. Bu çalışmada parametrik olmayan yaklaşımlar dikkate alınacaktır. Parametrik olmayan regresyon modelleri daha çok düzeltme teknikleri üzerinde yoğunlaşmıştır. Çekirdek düzeltme, k -en yakın komşuluk düzeltme, yerel polinomiyal düzeltme ve splayn düzeltme teknikleri en çok kullanılanlardır. Literatürde çekirdek yoğunluk tahmini fikri ilk olarak Fix ve Hodges (1951) tarafından önerilmiştir. Green ve Silverman (2000), Hardle (1994), Fan ve Gijbels (1996), Loader (1999) gibi isimler konunun literatürüne çok önemli katkı sağlamışlardır.

Parametrik olmayan regresyon tahminlemede band genişliği parametresinin optimum değerinin seçimi oldukça önemli bir yer tutmaktadır. Parametrik olmayan regresyon tekniklerinin uygulanması her zaman bir band genişliğinin seçimine ihtiyaç duymaktadır (Rice 1984). Çekirdek regresyonda, çok büyük bir band genişliği alınması fazla düzleştirmeye (oversmoothing), çok küçük bir band genişliği alınması ise az düzleştirmeye (undersmoothing) neden olur. Amaç az düzleştirme ve fazla düzleştirme arasındaki dengeyi (trade-off) sağlayan bir h değeri bulmaktır (Hardle 1994). Band genişliği, örneklem genişliğine bağlıdır. Uyumlamadaki başlıca problem, bu parametrenin seçimidir. Genellikle herhangi bir global hata ölçüsünü en küçükleyecek şekilde varyans ve yan'ın karesi ($bias^2$) arasındaki dengeyi sağlamak amacıyla seçilir (Vieu 1991).

Literatürde, parametrik olmayan regresyon modelleri üzerine geliştirilen pek çok çalışmada, daha çok kesin durumlar dikkate alınmıştır. Kesin durumlara ait bilginin yer aldığı çalışmalarda gerek parametrik gerekse parametrik olmayan regresyon modelleri yaygın olarak çeşitli alanlarda kullanılmaktadır. Ancak bazı veri gruplarını bu modellerle göstermek mümkün olamamakta, bazende bulanık yapıya sahip verilerin kullanılması gerekliliği bu klasik regresyon modellerinin kullanımını olanaksız hale getirmektedir. Bu bakımdan belirsizliğin var olduğu ve klasik regresyon modellerinin yetersiz kaldığı durumlarda yeni arayışlar başlamış ve bulanık regresyon modelleri ileri sürülmüştür. Bulanık tekniklerin amacı, verilerdeki tüm özgün belirsizliği modele dahil etmektir. Bu nedenle bulanık verilere dayalı modeller, verilerdeki özgün belirsizliğin reddedildiği ya da isteğe bağlı olarak ihmal edildiği modellere göre daha fazla bilgi içerirler. Ayrıca, bulanık verilere dayalı modeller çok daha geneldir, çünkü kesin sayı kendisiyle ilişkilendirilmiş bulanıklığın olmadığı özel bir bulanık sayı olarak kabul edilebilir (D'Urso ve Gastaldi 2002).

Literatürde bulanık doğrusal regresyon modelleri üzerine yapılmış kapsamlı çalışmalar bulunmaktadır. Tanaka ve ark. (1982), Tanaka (1987), Tanaka ve Watada (1988), Tanaka ve ark. (1989), Bardossy (1990), Savic ve Pedrycz (1991), Xizhao ve Minghu (1992) ve Chang ve Lee (1994b), Ishibuchi ve Tanaka (1992), Nasrabadi ve Nasrabadi (2004), Diamond (1988) ve Hong ve ark. (2001) tarafından yapılmış olan çalışmalar bunlardan bazılarıdır.

Ancak parametrik olmayan bulanık regresyon modelleri konusunda hem teorik hemde uygulama bazında çok az sayıda çalışmanın bulunması, bu tür modellerin davranışları konusunda merak uyandırmaktadır. Cheng ve Lee (1999), fonksiyonel şeklinin önceden belirlenemediği bulanık regresyonu yani parametrik olmayan bulanık regresyonu incelemişlerdir. Parametrik olmayan yöntemlerden k -en yakın komşuluk ve çekirdek regresyon tahmin edicilerini bulanıklaştırmış ve analiz etmişlerdir. Yine Cheng ve Lee (2001), girdi ve çıktı arasındaki fonksiyonel ilişkinin önceden belirlenemediği bulanık regresyon analizinde radyal tabanlı fonksiyon (RBF) ağlarını kullanmışlardır. Wang ve ark. (2007), bulanık yerel doğrusal regresyon modelini ele almışlar ve deneysel çalışmalar yoluyla bu

yaklaşımı, çekirdek ve k -en yakın komşu düzeltme yaklaşımları ile karşılaştırmışlardır.

Bu çalışmalarda kullanılan çekirdek düzeltme ve k -en yakın komşu düzeltme tekniklerinin, literatürde yerel polinomial modellere göre daha az etkin oldukları ifade edilmiştir. Ancak yerel polinomial modellerin bulanık yapıdaki ifadesi ve tanımlamaları üzerinde eksiklikler bulunmaktadır. Bu amaçla tez çalışmasında, parametrik olmayan bulanık regresyon modellerinden özellikle yerel polinomial modelleri araştırmak, güçlü ve zayıf yönlerini analiz etmek, farklı dereceler kullanarak ve farklı çekirdek fonksiyonları götürülerek modeller için optimum band genişliklerini geliştirilen formüllerle hesaplayarak, geliştirme yapmak amaçlanmıştır.

Çalışma yedi bölümden oluşmaktadır. Bu bölümlerden ilki olan giriş bölümünde tezin amacı ve kapsamına yer verilmiş, konu üzerinde daha önceden yapılmış bazı çalışmalara değinilmiştir, literatür konusunda esas araştırmalara her bölümün kendi içerisinde değinilmiştir.

İkinci bölümde, parametrik olmayan regresyon problemi tanımlanmış ve regresyonda pürüzlülük ceza yaklaşımı ele alınmıştır. Parametrik olmayan regresyon modelleri için düzeltme kavramı üzerinde durularak, çekirdek düzeltme, k -en yakın komşuluk düzeltme, splayn düzeltme teknikleri ile ayrıntılı olarak yerel polinomial düzeltme tekniği incelenmiştir. Ayrıca parametrik olmayan regresyon modellerinde önemli bir konu olan band genişliği seçim kriterlerinden, çapraz geçerlilik kriteri, geliştirilmiş çapraz geçerlilik kriteri ve Akaike Bilgi kriterine yer verilmiştir. Bölümün sonunda parametrik olmayan regresyon modellerinin karşılaştırılmasında kullanılan performans kriterlerine değinilmiştir.

Üçüncü bölüm, bu çalışmada ikinci bölümden dördüncü bölüme geçişte bir köprü niteliği taşımaktadır. İkinci bölümde kesin teorisine değinilen kavram ve tekniklerin dördüncü bölümde bulanık yapıda ifade edilmesinde kullanılacak çeşitli tanımlamalar bu bölümde detaylı olarak incelenmiştir. Bulanık kümeler ve bulanık sayı tanımlamalarının yanı sıra bulanık teoride önemli bir konu olan genişletme prensibi ele alınmıştır. Sonrasında, bulanık sıralama, bulanık fonksiyonlar ile türev alma konularına değinilmiştir.

Çalışmanın özgün bölümlerinin ilki olan dördüncü bölümde, parametrik olmayan regresyon modellerinin bulanık yapıda ifadelerine yer verilmiştir. Çekirdek düzeltme, k -en yakın komşuluk düzeltme ve yerel polinomiyal düzeltme tekniklerinin bulanık yapıda tanımlaması gerçekleştirilmiştir. Sonrasında bu modellerden yerel polinomiyal düzeltme tekniğinde, band genişliği (düzeltme parametresi) seçimi için şapka matrisini kullanarak çapraz geçerlilik ve genelleştirilmiş çapraz geçerlilik kriterlerinin bulanık yapıda tanımları geliştirilmiştir. Parametrik olmayan bulanık regresyon modellerin karşılaştırılması amacıyla kullanılacak olan performans kriterlerinden bahsedilmiştir.

Çalışmanın dördüncü bölümünde teorisi üzerinde durulan parametrik olmayan düzeltme tekniklerinden, çekirdek düzeltme ve yerel polinomiyal düzeltme teknikleri konusunda araştırma yapmak, modellerin farklı veri setleri üzerinde davranışlarını incelemek amacıyla, beşinci bölümde dört farklı türetim fonksiyonu ile farklı veri setleri üretilmiş, simülasyon deneyleri ile elde edilen sonuçlar gerek tablolar gerekse grafiksel gösterimler yardımıyla özetlenmiştir. Analizlerin gerçekleştirilmesinde, R Programı ve bu program içerisinde yer alan *locpol* paketi kullanılmıştır. Bu bölümde özellikle, bulanık yerel polinomiyal modellerde, yerel doğrusal ve yerel kübik modeller arasında karşılaştırmalara yer verilmiştir. Geliştirilen çapraz geçerlilik ve genelleştirilmiş çapraz geçerlilik formülasyonları kullanılarak optimum band genişliği değeri hesaplanmış, sonrasında modellerin performanslarını karşılaştırmak için ortalama karesel hata (*ASE*) değerleri bulunmuştur.

Altıncı bölümde, bir önceki bölümde gerçekleştirilen simülasyon deneylerini desteklemek amacıyla gerçek bulanık veri setleri üzerinde benzer hesaplamalar yapılmıştır. Hesaplamalar yapılırken iki farklı veri seti kullanılmıştır.

Son bölüm olan yedinci bölümde, gerçekleştirilen tüm uygulama çalışmalarından elde edilen sonuçlar yorumlanmıştır ve sonraki dönem çalışmalarında yapılabilecekler konusunda önerilerde bulunulmuştur.

2. PARAMETRİK OLMAYAN REGRESYON MODELLERİ

Bir regresyon eğrisi, bir veya birkaç açıklayıcı değişken X ile yanıt değişkeni Y arasındaki genel ilişkiyi tanımlar. Regresyon analizinin temel amacı, bilinmeyen regresyon fonksiyonu için uygun bir tahmin üretmektir. Bu noktada, parametrik ve parametrik olmayan tahminlerden faydalanılır. Parametrik regresyon yaklaşımı, açıklayıcı değişken ve cevap değişkeni arasındaki ilişkinin fonksiyonel şeklinin bilindiği durumlarda kullanılır. Ancak parametrik regresyon tahminleri çeşitli varsayımlara dayalı olarak gerçekleştirilirler ve bu varsayımların sağlanmadığı durumlarda yapılan tahminler iyi bir tahmin olma özelliğine sahip olamamaktadırlar. Ayrıca, pratik problemlerde yanıt değişkeni ile açıklayıcı değişkenler arasındaki ilişkinin belirli bir parametrik modelle ifade edilmesi çok kısıtlayıcı olmakla birlikte, yanlış sonuçlara neden olabilmektedir. Parametrik yaklaşımlardaki bu sorunlar nedeniyle varsayımların esnetilmesine olanak sağlayan yaklaşımlara ihtiyaç duyulmuştur. Parametrik olmayan ve yarı parametrik yaklaşımlar bunlara örnektir. Bu bölümde parametrik olmayan yaklaşımlar ve düzeltme teknikleri üzerine odaklanılacaktır.

Parametrik olmayan regresyon ve düzeltme teknikleri konusunda pek çok çalışma mevcuttur. Literatürde çekirdek yoğunluk tahmini fikri ilk olarak Fix ve Hodges (1951) çalışmalarında yer almıştır (Wand ve Jones 1995). Sonrasında Rosenblatt (1956) ve Parzen (1962) çekirdek yoğunluk tahmin edicilerini önermişlerdir. Daha sonra çalışmalar geliştirilerek Nadaraya (1964) ve Watson (1964), Nadaraya-Watson çekirdek regresyon tahmin edicisini ortaya koymuşlardır. Priestley ve Chao (1972) ile Gasser ve Müller (1979, 1984) farklı çekirdek tahmin edicileri ile parametrik olmayan regresyon modellerine katkıda bulunmuşlardır. Ayrıca çekirdek fonksiyonunun seçimi ve bunların türevleri konularında araştırmalar da yapılmıştır (Wand ve Jones 1995).

Hardle ve Marron (1985) ise bilinmeyen çok değişkenli regresyon fonksiyonunun tahmini için çekirdek tahmin edicilerinin nasıl kullanıldıklarını ve düzeltme parametresinin seçim tekniklerini araştırmışlardır. Hardle (1994)'nin "*Uygulamalı Parametrik Olmayan Regresyon*" başlıklı kitabı bu alanda kapsamlı bir kaynak oluşturmaktadır. Yine Green ve Silverman (2000) tarafından yazılmış

“Parametrik Olmayan Regresyon ve Genelleştirilmiş Lineer Modeller” kitabı parametrik olmayan modeller hakkında detaylı teorik bilgi içermektedir. Fox (2000) kitabında R paket programında parametrik olmayan regresyon için geliştirmiş olduğu çeşitli fonksiyon ve algoritmalarla araştırmacılara ışık tutmuştur. Fan ve Gijbels (1996) yerel polinomial modeller ve uygulamaları üzerinde yoğunlaştıkları kitaplarında, parametrik olmayan regresyon konusunda derinlemesine incelemelerde bulunmuşlardır. Loader (1999) “Yerel Regresyon ve Olabilirlik” isimli kitabında konunun teorisinin yanı sıra R, S ve S Plus programları için uygulamalara yer vermiştir.

Çalışmanın bu bölümünde parametrik olmayan regreyon problemi ve regresyonda düzeltme kavramı üzerinde durularak parametrik olmayan regresyon yaklaşımlarından bazıları ele alınacaktır. Bu yaklaşımlarda önemli bir yere sahip olan düzeltme parametresi seçim yöntemlerine değinilerek, modellerin performanslarının karşılaştırılmasında kullanılacak kriterler incelenecektir.

2.1. Parametrik Olmayan Regresyon Problemi

Tek açıklayıcı değişkenin kullanıldığı ve n , $\{(X_i, Y_i)\}_{i=1}^n$ bağımsız gözlemden oluşan, açıklayıcı değişken ile yanıt değişkeni arasındaki parametrik olmayan regresyon ilişkisi (2.1) ile ifade edilir.

$$Y_i = m(X_i) + \varepsilon_i, \quad i = 1, 2, \dots, n \quad (2.1)$$

Denklem (2.1)’de, $m(X)$ parametreleri belirgin olmayan bir fonksiyonel ilişkiyi gösteren bilinmeyen bir fonksiyondur (Green ve Silverman 2000). Yani az sayıda parametre ile basit olarak özetlenemeyen ve belirgin bir şekilde sahip olmayan X değişkenlerinin bir fonksiyonudur. ε ise 0 ortalamalı ve σ^2 sabit varyanslı, bağımsız, özdeş (aynı dağılımlı) olarak dağılan rasgele hata terimleridir. Parametrik olmayan regresyonun amacı parametrelerin tahmini yerine bilinmeyen yanıt fonksiyonu olan $m(X)$ ’i tahmin ederek, açıklayıcı değişken ile yanıt değişkeni arasında genel ilişkiyi açıklayan bir model sağlamaktır. Modelde $m(X)$ için $X = x$ koşulu altında Y değişkeninin beklenen değeri elde edilmek istenir:

$$m(X) = E[Y|X = x] = \frac{\int y f(x, y) dy}{f(x)} \quad (2.2)$$

Amaç, denklem (2.2) ile tanımlanan regresyon fonksiyonunun tahmin edilmesidir. Denklem (2.2)'de $f(x, y)$, (X, Y) 'nin ortak olasılık yoğunluk fonksiyonunu ve $f(x)$, X 'in marjinal olasılık yoğunluk fonksiyonunu göstermektedir (Hardle 1994).

Parametrik olmayan yaklaşım, iki değişken arasındaki genel ilişkinin araştırılmasında çok yönlü bir yöntem sağlaması ve belirli bir parametrik modeli referans almadan tahminler yapabilmesi, hem doğrusal hem de doğrusal olmayan ilişkilerin modellenmesi açısından oldukça yararlıdır. Ancak parametrik olmayan regresyonun bu avantajlarının yanı sıra parametrik regresyona kıyasla artan işlem sayısı ile işlemlerin karmaşıklığı ve bazı durumlarda elde edilen sonuçların açıklanmasında yaşanan zorluklar dezavantajlarından bazılarıdır (Hardle 1994; Fox 2000).

2.1.1. Pürüzlülük ceza yaklaşımı ve splayn düzeltme tekniği

En genel ve basit şekliyle “pürüzlülük ceza yaklaşımı”, polinom regresyondan çok az bir farkla regresyon doğrusu boyunca klasik doğrusal regresyondaki model varsayımlarını esneten bir yöntemdir. Pürüzlülük ceza yaklaşımının esas amacı, hızlı olarak dalgalanan bir eğrinin eğilimini ölçmek ve daha sonra eğri tahmininde oldukça farklı iki amaç olan sabit eğimli uyumlar ve esnek eğimli uyumlar arasında gerekli uzlaşmayı gerçekleştirecek şekilde tahmin problemini ortaya koymaktır (Green ve Silverman 2000). Şöyle ki, pürüzlülük ceza yaklaşımı eğrinin pürüzlülüğünü ölçer ve verilerin eğriye uyumu ile eğrinin pürüzlülüğü arasında bir uzlaşma sağlar.

$[a, b]$ aralığında tanımlanan bir m eğrisi verilsin, m eğrisinin ne kadar “pürüzlü veya dalgalı” olduğunu ölçmenin birkaç yolu vardır. Bunlardan en yaygın olanı, iki kez türevlenebilen bir m fonksiyonunun eğiminin değişim oranı, fonksiyonun ikinci türevi m'' yardımıyla tanımlanır ve uyum fonksiyonunun eğimindeki toplam değişim ölçüsü (2.3) ile hesaplanır.

$$S(m) = \int_a^b m''(x)^2 dx \quad (2.3)$$

λ , düzeltme parametresi (band genişliği) olmak üzere ikinci mertebeden sürekli türeve sahip tüm fonksiyonlar üzerinde minimum olabilen, eğimdeki hızlı değişimler için cezayı ve hata kareler toplamının minimizasyonunu birleştirerek fonksiyonun tahmininin gerçekleştirilebildiği durum (2.4) cezalı en küçük hata-kareler toplamı ile ifade edilir. $\lambda > 0$ olmak üzere, cezalı hata-kareler toplamı,

$$R(m) = \sum_{i=1}^n \{Y_i - m(x_i)\}^2 + \lambda \int_a^b \{m''(x)\}^2 dx \quad (2.4)$$

formülü şeklinde yazılır. Denklem (2.4)'de cezalı en küçük kareler tahmin edicisi \hat{m} , $R(m)$ fonksiyonunun minimize edeni şeklinde tanımlanır. λ düzeltme parametresinin değeri, bir uyumun eğiminin esnekliği üzerinde önemli bir etkiye sahiptir. λ değeri büyükse, (2.4) eşitliğinde ana bileşen pürüzlülük ceza terimi olacak ve minimum veren \hat{m} çok az eğrisellik gösterecektir. Limit durumunda ise λ sonsuza yaklaşacağından ceza terimi sifıra yaklaşacak ve tahmin eğrisi doğrusal regresyon uyumlamasına yaklaşım sergileyecektir. λ değeri çok küçükse, ana katkıyı hata kareler toplamı yapacak ve tahmin eğrisi veri noktalarını yakından izleyecektir. Limit durumunda λ sifıra yaklaştığı için, tahmin eğrisi interpolasyon eğrisine yaklaşacaktır (Green ve Silverman 2000). Çalışmalarda en önemli konulardan bir tanesi, verileri en iyi temsil edecek eğriyi tahmin edebilmek için λ değerinin optimum seçimidir. Düzeltme parametresinin seçim yöntemlerine Bölüm 2.3'te yer verilecektir.

Parametrik olmayan regresyon modelleri için pürüzlülük ceza yaklaşımının kullanıldığı tahmin yöntemlerinden biri splayn düzeltme tekniğidir. Splayn düzeltme ve türlerinin modern gelişimi üzerine yapılan temel çalışmalar Wahba (1975)'ya aittir. Splayn düzeltme tekniğinde tahminleme yapılırken (2.4) eşitliğinde yer alan cezalı en küçük kareler toplamından yararlanır.

$\{x_i, y_i\}_{i=1}^n$ gözlem değerleri olmak üzere, herhangi bir $[a, b]$ aralığında $a = x_0 < x_1 < \dots < x_n < x_{n+1} = b$ koşulunu sağlayan x_i reel sayılarının $(a, x_1), (x_1, x_2), \dots, (x_i, x_{i+1}), \dots, (x_{n-1}, x_n), (x_n, b)$ alt aralıklarına bölündüğünü

ve her bir $x_i \leq x \leq x_{i+1}, i = 0, 1, \dots, n$ alt aralığında k . mertebeden bir $m_i(x)$ polinom fonksiyonu tanımlandığı varsayalım. Burada x_i 'ler “düğüm noktaları” ve k . mertebeden ($k \geq 1$) polinom fonksiyonların birleşiminden oluşan $m(x)$ fonksiyonu “splayn” adını alır. Splayn düzeltme tekniği ile yapılacak tahminleme de $m(x)$ fonksiyonu aşağıdaki özellikleri sağlar.

- i.* $(a, x_1), (x_1, x_2), \dots, (x_i, x_{i+1}), \dots, (x_{n-1}, x_n), (x_n, b)$ alt aralıklarının herbirinde m bir kübik polinomialdır.
- ii.* m fonksiyonunun kendisi, birinci ve ikinci türevleri her bir x_i noktasında, böylece $[a, b]$ aralığının tümü üzerinde süreklidir.
- iii.* m fonksiyonu $[a, b]$ aralığının a ve b uç noktalarında ikinci ve üçüncü türevlerinde sıfır değerini alır.

$m(x)$ fonksiyonu bu özelliklerden *i* ve *ii*'yi taşırsa “kübik splayn”, üç özelliği bir arada taşırsa “doğal kübik splayn” olarak adlandırılır. \hat{m} 'in doğal kübik splayn olması önemli bir konudur. Çünkü sonsuz boyutlu $C^2[a, b]$ pürüzsüz fonksiyonlar kümesinde minimizasyon probleminin üstesinden bu özellik sayesinde gelinir.

Splayn düzeltme ile tahminlemeye geçmeden önce bazı özel gösterim ve tanımlamalar yapılmalıdır. m 'in $x_1 < x_2 < \dots < x_n$ düğüm noktaları ile doğal kübik splayn olduğu varsayımı altında $m_i = m(x_i)$ ve $\gamma_i = m''(x_i)$ $i = 1, \dots, n$ olsun. Doğal kübik splayn tanımından m 'in x_1 ve x_n düğümlerindeki ikinci türevi $\gamma_1 = \gamma_n = 0$ olur. $\mathbf{m} = (m_1, \dots, m_n)'$ ve $\boldsymbol{\gamma} = (\gamma_2, \dots, \gamma_{n-1})'$ vektörleri verilsin. γ_i 'nin $(n - 2)$ elemanı vardır. \mathbf{m} ve $\boldsymbol{\gamma}$ vektörleri m eğrisini tamamen belirler ve herhangi bir x noktasında m 'in türevleri ve değeri için \mathbf{m} ve $\boldsymbol{\gamma}$ 'nin terimlerinde açık formül vermek mümkündür.

Verilen düğüm noktalarında vektörlerin bir doğal kübik splayn göstermesi için gerek ve yeter koşul Q ve R matrisleri ile tanımlayacağımız iki matrise bağlıdır. Q ve R , $x_1 < x_2 < \dots < x_n$ düğüm noktalarında elde edilen band matrisleri olup $h_i = x_{(i+1)} - x_{(i)}$, $i = 1, \dots, n$ 'nin fonksiyonlarıdır. Q , $i = 1, \dots, n; j = 2, \dots, n - 1$ için;

$$q_{j-1,j} = h_{j-1}^{-1}, \quad q_{jj} = -h_{j-1}^{-1} - h_j^{-1}, \quad q_{j+1,j} = h_j^{-1} \text{ ve } |i - j| \geq 2 \text{ için } q_{ij} = 0.$$

Q , $n \times (n - 2)$ tipinde bir matris olup sütunları $\boldsymbol{\gamma}$ 'nin girişlerine benzer olarak $j = 2$ ile başlayarak numaralandırılmıştır, ilk elemanı q_{12} 'dir. R 'nin elemanları için $j = 2, \dots, n - 1$;

$$r_{ii} = \frac{1}{3}(h_{i-1} + h_i), \quad i = 2, \dots, n - 1$$

$$r_{i,i+1} = r_{i+1,i} = \frac{1}{6}h_i, \quad i = 2, \dots, n - 2 \text{ ve } |i - j| \geq 2 \text{ için } r_{ij} = 0$$

olan $(n - 2) \times (n - 2)$ tipinde simetrik bir matristir. R matrisi herbir i için $|r_{ii}| > \sum_{i \neq j} |r_{ij}|$ anlamında kesin köşegen dominanttır. Bu özellikten dolayı R matrisi kesin pozitif tanımlı bir matristir ve tersi vardır (Green ve Silverman 2000). Buradan bir K matrisi aşağıdaki gibi tanımlanabilir.

$$K = QR^{-1}Q^T$$

Cezalı en küçük kareler problemin ifadesinde kullanılmak üzere Teorem 2.1'de pürüzlülük ceza teriminin matrisler yardımıyla ifadesi yer almaktadır.

Teorem 2.1 (İspat için Bknz. Green ve Silverman 2000)

\mathbf{m} ve $\boldsymbol{\gamma}$ vektörlerinin bir doğal kübik splayn belirtmeleri için gerek ve yeter koşul;

$$Q'\mathbf{m} = R\boldsymbol{\gamma} \tag{2.5}$$

(2.5) eşitliğinin sağlanması durumunda pürüzlülük cezası aşağıdaki gibi yazılabilir.

$$\int_a^b m''(x)^2 dx = \boldsymbol{\gamma}'R\boldsymbol{\gamma} = \mathbf{m}'K\mathbf{m} \tag{2.6}$$

$\mathbf{y} = (y_1, \dots, y_n)^T$ bağımlı değişkene ait gözlem değerleri vektörü, x_i düğüm noktalarında $m(x_i)$ değerlerine ait vektör $\mathbf{m} = (m_1, \dots, m_n)^T = (m(x_1), \dots, m(x_n))^T$ olmak üzere (2.4)'te yer alan cezalı en küçük kareler toplamındaki hata kareler toplamı aşağıdaki gibi yazılabilir.

$$\sum \{y_i - m(x_i)\}^2 = (\mathbf{y} - \mathbf{m})^T (\mathbf{y} - \mathbf{m})$$

Pürüzlülük ceza terimi için eşitlik (2.6) ifadesi göz önünde bulundurularak cezalı en küçük kareler toplamı aşağıdaki gibi ifade edilir.

$$\begin{aligned} R(\mathbf{m}) &= \sum_{i=1}^n \{Y_i - m(x_i)\}^2 + \lambda \int_a^b \{m''(x)\}^2 dx \\ &= (\mathbf{y} - \mathbf{m})^T (\mathbf{y} - \mathbf{m}) + \lambda \mathbf{m}'^T K \mathbf{m} \\ &= \mathbf{m}^T (\mathbf{I} + \lambda K) \mathbf{m} - 2\mathbf{y}^T \mathbf{m} + \mathbf{y}^T \mathbf{y} \end{aligned} \quad (2.7)$$

Eşitlik (2.7)'de yer alan $(\mathbf{I} + \lambda K)$ kesin pozitif tanımlı bir matristir ve bu nedenle eşitlik tek bir minimuma sahiptir. (2.7)'ye minimum değer veren \mathbf{m} vektörü, eşitliğin \mathbf{m} 'e göre türev fonksiyonuna sıfır değerini veren vektör olur.

$$\begin{aligned} R'(\mathbf{m}) &= 2\mathbf{m}(\mathbf{I} + \lambda K) - 2\mathbf{y} = 0 \\ 2\mathbf{m}(\mathbf{I} + \lambda K) &= 2\mathbf{y} \\ \mathbf{m} &= (\mathbf{I} + \lambda K)^{-1} \mathbf{y} \end{aligned} \quad (2.8)$$

Denklem (2.8) ifadesinde yer alan $(\mathbf{I} + \lambda K)^{-1}$ matrisi “düzeltme matrisi” ya da “şapka matrisi” olarak adlandırılır ve bu matris sadece verilen $\lambda > 0$ düzeltme parametresi ve $\mathbf{x} = (x_1, \dots, x_n)$ düğüm noktaları ile belirlenir. Böylece \mathbf{y} değerlerini \mathbf{m} değerlerine görüntüleyen $n \times n$ boyutlu şapka matrisi olarak aşağıdaki gibi ifade edilir.

$$S_\lambda = (\mathbf{I} + \lambda K)^{-1} \quad (2.9)$$

Böylece $\mathbf{m} = S_\lambda \mathbf{y}$ olarak yazılır (Green ve Silverman 2000).

2.2. Regresyonda Düzeltme Kavramı ve Düzeltme Teknikleri

Bağımlı ve bağımsız değişken arasındaki ilişkiyi gözlemlemek amacıyla çizilen bir dağılım grafiği ile mevcut veri seti üzerinde yer alan yüzlerce noktanın bulunduğu bir eğriyi yüksek dereceden polinomiyal modellerle bile tahmin etmek oldukça güçtür. Parametrik olmayan regresyon analizinin amacı bilinmeyen yanıt fonksiyonu olan $m(X)$ 'in fonksiyonel yapısına ait analizi gerçekleştirmektir. Gözlenen hataların azaltılması ile yanıt değişkeninin açıklayıcı değişkene göre ortalama bağımlılığının önemli ayrıntılarını vermek yorumu kolaylaştırır. Düzeltme fikrinin temelinde, verileri bir eğriye uydurmak ve daha basit fonksiyonların birleşimi olabilen esnek fonksiyonları kullanmak yatar. Bu eğri yaklaştırma işlemi genel olarak '*düzeltme (smoothing)*' olarak adlandırılır (Hardle 1994). Bir düzelticinin en önemli özelliği, değişkenler arasındaki ilişkinin şeklini kesin bir biçimde belirlememesidir (bir doğru ya da polinom gibi...) ve bu özelliğinden dolayı parametrik olmayan regresyonda sık kullanılan bir araçtır (Hastie ve Tibshirani 1990). Bir düzeltme yöntemi ile tahmini gerçekleştirilen eğri '*düzgün*' (smooth) olarak adlandırılır (Hardle 1994).

Düzeltme teknikleri farklı ölçümler arasındaki fonksiyonel ilişkileri bulmaya yarayan yöntemlerdir. Parametrik regresyonda olduğu gibi verilerin bir ya da daha fazla bağımsız değişken ve bir bağımlı değişken ölçümlerinden oluştuğu varsayılır. Düzeltme teknikleri uydurulmuş eğrilerin şeklini belirlemek için kendileri veri noktaları sağlayan esnek bir yaklaşım sergilerler (Loader 1999). İzleyen bölümde literatürde yaygın olarak kullanılan ve araştırdığımız konu ile ilişkili olan bazı düzeltme tekniklerinden söz edilecektir. Bu düzeltme teknikleri genel olarak (2.10) eşitliğindeki gibi ifade edilmelerinden dolayı doğrusal düzelticiler olarak anılmaktadırlar.

$$\hat{m}(x) = \sum_{i=1}^n W_i(x) \cdot y_i \quad (2.10)$$

Burada $W(x)$ vektörü, x vektörüne bağlı olarak hesaplanan bir ağırlık vektörünü gösterir. Bu ağırlık vektörünün hesaplanmasında örneğin bir çekirdek fonksiyonundan faydalanılabilir. k -en yakın komşu düzeltme, çekirdek (kernel)

düzeltilme ve yerel polinomiyal düzeltilme gibi teknikler (2.10) formülü ile tahminleme gerçekleştirirler (Hardle 1994).

2.2.1. Çekirdek düzeltilme

Denklem (2.10) ile ifade edilen parametrik olmayan regresyon tahmini ilk kez Nadaraya (1964) ve Watson (1964) tarafından önerilmiştir. Eşitlikte x 'e bağlı $W_i(x)$ ağırlıkları yardımıyla, y_i 'lerin bir ağırlıklı toplamı şeklinde x noktasında bilinmeyen regresyon fonksiyonu tahmin edilir. Burada esas fikir, $m(x_0)$ 'ı tahmin ederken x_0 'a yakın gözlemlere daha büyük, uzak gözlemlere ise daha küçük ağırlıklar verilerek tahminlemenin gerçekleştiriliyor olmasıdır. Yerel ortalama ağırlıkları elde etmek için bir ölçek parametresi tarafından kontrol edilen, her iki yönde azalan ve merkezi sıfırda bulunan tek moda sahip bir dağılımdan yararlanır. Bu nedenle “çekirdek (kernel)” olarak adlandırılan olasılık yoğunluk fonksiyonları kullanılır (Yatchew 2003). Çekirdek fonksiyonu, simetrik bir olasılık yoğunluk fonksiyonu ile aşağıdaki benzer özelliklere sahiptir:

- i. $K(u) \geq 0, \forall u$ için,
- ii. $\int_{-\infty}^{+\infty} K(u) du = 1$,
- iii. $K(-u) = K(u)$, yani simetriktir.

Çekirdek düzeltilme tekniği için bir çekirdek fonksiyonunu kullanan ve denklem (2.10)'da yer alan ağırlık dizisi (2.11) ile ifade edilir:

$$W_i(x) = \frac{K\left(\frac{x_i - x}{h}\right)}{\sum_{i=1}^n K\left(\frac{x_i - x}{h}\right)} \quad (2.11)$$

eşitlikte $\sum_{i=1}^n W_i(x) = 1$ 'dir. Denklem (2.11)'de n gözlem sayısını, K seçilen çekirdek fonksiyonunu ve h düzeltilme parametresini, $W_i(x) = W(x, x_i)$, $x - x_i$ uzaklığına bağlı i . gözlem y_i 'ye atanan ağırlığı ifade eder. Genellikle uzaklık küçükse ağırlık yüksek ve uzaklık büyükse ağırlık düşük olur. Ağırlıklar, çekirdek

tarafından belirtilir ve düzeltme parametresi tarafından kontrol edilirler (Hardle 1994). Literatürde yer alan çekirdek fonksiyonu türleri Çizelge 2.1’de görülmektedir.

Çizelge 2.1 Çekirdek Fonksiyonu Türleri

Çekirdek (Kernel)	$K(x)$
Uniform	$\begin{cases} \frac{1}{2}, & x \leq 1 \\ 0, & x > 1 \end{cases}$
Gauss (Normal)	$\frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}x^2)$
Cauchy	$\{\pi(1+x^2)\}^{-1}$
Triangular	$\begin{cases} (1- x), & x \leq 1 \\ 0, & x > 1 \end{cases}$
Epanechnikov	$\begin{cases} \frac{3}{4}(1-x^2), & x \leq 1 \\ 0, & x > 1 \end{cases}$
Quartic (Biweight)	$\begin{cases} \frac{15}{16}(1-x^2)^2, & x \leq 1 \\ 0, & x > 1 \end{cases}$
Triweight	$\begin{cases} \frac{35}{32}(1-x^2)^3, & x \leq 1 \\ 0, & x > 1 \end{cases}$

Denklem (2.11) ile ifade edilen ağırlık fonksiyonu (2.10)’da yerine yazıldığında çekirdek tahmini (2.12) şeklinde elde edilir ve bu tahmin genellikle “Nadaraya-Watson” tahmin edicisi olarak adlandırılır.

$$\hat{m}(x) = \sum_{i=1}^n \left(\frac{K\left(\frac{x_i - x}{h}\right)}{\sum_{i=1}^n K\left(\frac{x_i - x}{h}\right)} \right) \cdot y_i \quad (2.12)$$

Çekirdek fonksiyonlarında düzeltme parametresinin seçimi önemli rol oynar. Büyük düzeltme parametresi değeri için eğri çok yavaş değişir ve düzeltme önemlidir. Bu durumda tahminin varyansı sınırlı fakat tahmin oldukça sapmalıdır. Küçük düzeltme parametresi değeri için eğri oldukça düzensizdir ve sapmalar

sınırlı fakat tahminin varyansı büyüktür. Bu nedenle düzeltme parametresi sapmalar ve tahminin doğruluğu arasında bir uzlaşma sağlar (Slama 2001).

2.2.2. *k*-en yakın komşuluk tahmini

En yakın komşuluk tahminlemenin yapısı çekirdek tahminlemeye göre farklılık göstermektedir. Çekirdek tahminlemesi, çekirdek fonksiyonu ve düzeltme parametresi ile belirlenen x etrafındaki sabit komşuluktan hareketle bağımlı değişkenlerin ağırlıklı ortalaması olarak ifade edilir. En yakın komşuluk tahminleme ise x etrafındaki değişen komşuluktan hareketle elde edilen bağımlı değişkenlerin ağırlıklı ortalamasıdır. Buradaki en önemli fark komşuluk mesafesinin sabit değil, değişken olmasıdır. Diğer bir deyişle belirli bir x noktasında fonksiyon tahminlenmek isteniyorsa X değişkeninin bu noktaya en yakın k tane gözlemine ait Y değerlerinin ortalaması hesaplanır (Hardle 1994).

En yakın komşuluk tahminlemesine ait ağırlık dizisi Loftsgaarden ve Quesenbery (1965) tarafından tanıtılmış ve sınıflandırma amaçları için Cover ve Hart (1967) tarafından kullanılmıştır.

Denklem (2.1)'deki gibi parametrik olmayan bir regresyon modelinde $(x_1, y_1), \dots, (x_n, y_n)$ gözlem değerleri dikkate alınarak $\{x_i\}_{i=1}^n$, x değişkeninin $[a, b]$ aralığında aldığı farklı değerleri gösterebilir. x_i noktasında m fonksiyonun tahmini, x 'e en yakın k tane komşuya ait olan y_i değerlerinin ortalamasıdır. k -en yakın komşu tahmini (2.13) eşitliğindeki gibi ifade edilir.

$$\hat{m}_k(x) = n^{-1} \sum_{i=1}^n W_{ki}(x) \cdot y_i \quad (2.13)$$

Eşitlikte $\{W_{ki}(x)\}_{i=1}^n$ ağırlık dizisi, aşağıdaki gibi bir indeks dizisi yardımıyla (2.14) ile ifade edilir.

$$J_x = \{i: x_i, x' \text{ e en yakın } k \text{ tane gözlemden biri}\}$$

$$W_{ki}(x) = \begin{cases} n/k, & i \in J_x \\ 0, & d.d \end{cases} \quad (2.14)$$

k -en yakın komşu tahmini, değişen komşuluklarda ağırlıklı ortalamadır. Bu komşuluk öklid uzaklığındaki x_i 'nin k -en yakın komşuları arasında bulunan x değişkenleri aracılığıyla tanımlanır. Burada k düzeltme parametresi olup tahmin edilen eğrinin pürüzlülüğünün derecesini düzenler ve çekirdek düzeltmede yer alan düzeltme parametresine benzer rol oynar. Bu yöntemle elde edilen sonuç fonksiyonu süreklidir. Eğer k büyükse, fonksiyon daha yavaş değişir. $k = n$ ise fonksiyon y 'nin ortalamasına eşit ve sabittir. k küçükse fonksiyon daha hızlı değişir (Slama 2001).

2.2.3. Yerel polinomiyal düzeltme

Yerel polinomiyal regresyon fikri 1970'lerden günümüze literatürde yer almaktadır. Stone (1977, 1980, 1982) ve Cleveland (1979) konu üzerine sistematik bir çalışma yapmışlardır. Yerel polinomiyal uyumlama konusunda son çalışmalar ise Fan (1992, 1993), Fan ve Gijbels (1992), Ruppert ve Wand (1994) şeklindedir. Bu çalışmalar yerel polinomiyal uyumlamanın avantajları konusunda detaylı bilgi içermektedirler.

Yerel polinomiyal düzeltme yönteminin genel olarak ifade edilmesinde (X, Y) popülasyonundan bağımsız ve düzgün dağılımlı $(X_1, Y_1), \dots, (X_n, Y_n)$ ikililerini ve tahminleme metodolojisinin daha iyi anlaşılması açısından (2.15) modelinden türetilen verileri dikkate alalım.

$$Y = m(x) + \sigma(x)\varepsilon \quad (2.15)$$

Modelde, $E(\varepsilon) = 0$, $Var(\varepsilon) = 1$ olmak üzere X ile ε bağımsızdır. İlgilenilen regresyon fonksiyonunun tahmini, $m(x_0) = E(Y|X = x_0)$ ve onun türevleri $m'(x_0), m''(x_0), \dots, m^p(x_0)$ 'dir. x_0 noktasında $m(x)$ fonksiyonunun $(p + 1)$. türevinin var olduğu varsayılır. Bilinmeyen regresyon fonksiyonu $m(x)$ 'e yerel olarak bir p dereceli polinomla yaklaşılır. x_0 'ın komşuluğundaki x için Taylor Açılımı (2.16) eşitliğinde görülmektedir.

$$m(x) \approx m(x_0) + m'(x_0)(x - x_0) + \frac{m''(x_0)}{2!}(x - x_0)^2 + \dots + \frac{m^{(p)}(x_0)}{p!}(x - x_0)^p \quad (2.16)$$

Polinom derecesi $p = 1$ için yerel doğrusal, $p = 2$ için yerel kuadratik ve $p = 3$ için yerel kübik model söz konusu olur. Bu polinom yerel olarak ağırlıklandırılmış en küçük kareler regresyonu probleminin minimizasyonu ile uyumlanır.

$$\sum_{i=1}^n \{Y_i - \sum_{j=1}^p \beta_j (x_i - x_0)^j\}^2 K_h(x_i - x_0) \quad (2.17)$$

Burada h , yerel komşuluk büyüklüğünü kontrol eden düzeltme parametresi (band genişliği) ve $K_h(\cdot) = K(\cdot/h)/h$ bir K çekirdek fonksiyonu ile herbir başlangıç noktasına ağırlıkları atar.

Denklem (2.17)'den $\hat{\beta}_j$, $j = 0, 1, \dots, p$ en küçük kareler probleminin çözümünü gösterebilir. Yukarıdaki Taylor Açılımından $\hat{m}_v(x_0) = v! \hat{\beta}_v$, $m^{(v)}(x_0)$, $v = 0, \dots, p$ için bir tahmin edicidir. $m^{(v)}(\cdot)$ fonksiyonunun tamamını tahmin etmek için, ilgili tanım kümesindeki tüm x_0 noktalarında yukarıdaki ağırlıklandırılmış en küçük kareler problemi çözülür (Fan ve Gijbels 1996; Loader 1999). Problemin çözümü matris gösteriminden faydalanılarak ifade edilmek istenildiğinde \mathbf{X} tasarım matrisi, \mathbf{y} bağımlı değişkene ait gözlem değerleri vektörü ve $\hat{\boldsymbol{\beta}}$ tahminleri gösteren vektör olarak aşağıdaki gibi yazılabilir:

$$\mathbf{X} = \begin{pmatrix} 1 & (X_1 - x_0) & \dots & (X_1 - x_0)^p \\ \vdots & \vdots & \ddots & \vdots \\ 1 & (X_n - x_0) & \dots & (X_n - x_0)^p \end{pmatrix}, \mathbf{y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}, \hat{\boldsymbol{\beta}} = \begin{pmatrix} \hat{\beta}_0 \\ \vdots \\ \hat{\beta}_p \end{pmatrix}.$$

Köşegen elemanları $K\left(\frac{X_i - x_0}{h}\right)$ olan $n \times n$ boyutlu \mathbf{W} ağırlık matrisi $\mathbf{W} = \text{diag}\{K_h(X_i - x_0)\}$ aşağıdaki gibi olur:

$$\mathbf{W} = \begin{pmatrix} K\left(\frac{x_1-x_0}{h}\right) & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & 0 & K\left(\frac{x_n-x_0}{h}\right) \end{pmatrix}$$

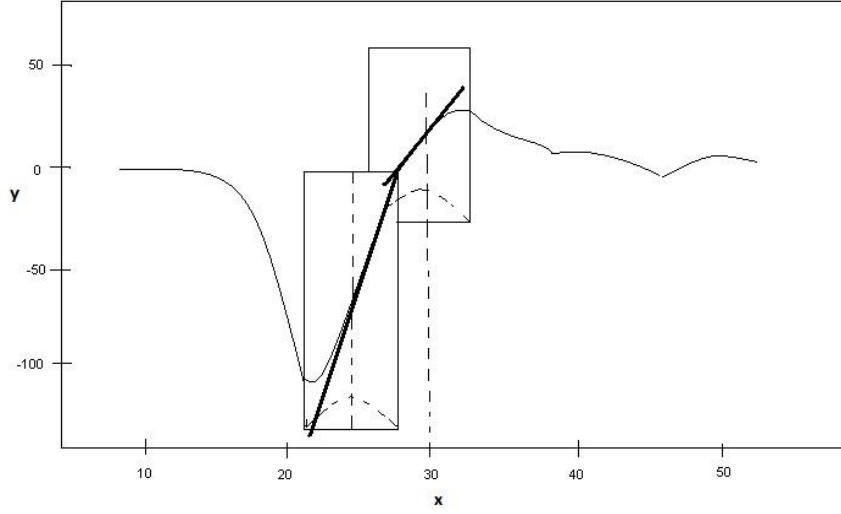
Denklem (2.17)'de yer alan ağırlıklandırılmış en küçük kareler problemi matrisler yardımıyla $\beta = (\beta_0, \dots, \beta_p)^T$ olmak üzere (2.18) ile yazılabilir.

$$\min_{\beta} (\mathbf{y} - \mathbf{X}\beta)^T \mathbf{W} (\mathbf{y} - \mathbf{X}\beta) \quad (2.18)$$

Çözüm vektörü ağırlıklandırılmış en küçük kareler teorisi ile sağlanır ve aşağıdaki gibi verilir.

$$\hat{\beta} = (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W} \mathbf{y} \quad (2.19)$$

Motorsiklet veri seti üzerinde yapılan bir yerel doğrusal uyumlamayı dikkate alalım. Şekil 2.1 verilen bir $x = 25$ noktasındaki regresyon eğrisini belirlemek için çekirdek fonksiyonu $K(t) = (1 - t^2)_+$ (“+” indisi pozitif kısmın alınması anlamına gelir) ve düzeltme parametresi $h = 3.3$ olmak üzere 25 ± 3.3 komşuluğundaki verileri kullanarak uyumlama gerçekleştirilir. Başka bir $x = 30$ noktası için ise benzer şekilde 30 ± 3.3 komşuluğundaki veriler tahmin edilerek regresyon uyumlaması gerçekleştirilir. Tüm eğrinin şekli ise her bir noktada tahminlenen regresyon fonksiyonlarının birleştirilmesi sonucunda elde edilir (Fan ve Gijbels 1996).



Şekil 2.1 Motorsiklet veri seti için $x \pm 3.3$ veri noktalarında yerel doğrusal uyumlanmanın gösterimi

2.2.3.1. Etki, Varyans ve Serbestlik Derecesi

Düzeltilme miktarının tanımlanmasını nasıl gerçekleştirebiliriz? Bu sorunun cevabı olarak band genişliği bir tanımlama getirir. Ancak ideal olan yanıt bu değildir. Çünkü düzeltme gerçekleştirilirken bu tanımlama yerel polinomun derecesi, ağırlık fonksiyonu vb. gibi seçimleri dikkate almaz. Ayrıca band genişliği, diğer düzeltme yöntemleri ya da parametrik modeller ile anlamlı bir karşılaştırma yapılabilmesine olanak vermez. Burada istenilen, yöntemler arasında karşılaştırma yapılabilmesine olanak sağlayan birimsiz bir tanımlamadır. İki tür tanımlama aşağıdaki gibidir.

- Noktasal kriter, tek bir noktada düzeltme miktarını tanımlar. Varyans indirgeme faktörü (variance reducing factor) ve etki fonksiyonunu (influence function) içerir.
- Global kriter, düzeltme miktarının tamamını tanımlar. Uyumlanmış serbestlik derecesidir.

Bu tanımlamalara geçmeden önce yerel polinomial düzeltme tekniği için farklı bir gösterimden faydalanılacaktır. Denklem (2.15)'da yer alan parametrik

olmayan regresyon modelini dikkate alalım. Belirli bir x noktası etrafında yerel olarak, m fonksiyonuna, parametrik fonksiyonların basit bir sınıfının üyesi ile yaklaşılabılır. Örneğin; Taylor teoremi, türevlenebilen bir fonksiyona yerel olarak bir doğru boyunca yaklaşılabileceğini ve iki kez türevlenebilen fonksiyona kuadratik polinom ile yaklaşılabileceğini söyler. Uyumlanan nokta x için, $h(x)$ band genişliği tanımlanır ve tahmini elde etmek için ağırlıklandırılmış gözlemler (2.20) formülü ile elde edilir.

$$w_i(x) = W\left(\frac{x_i - x}{h(x)}\right) \quad (2.20)$$

Burada $W(u)$, x 'in yakınındaki gözlemlere daha büyük ağırlıklar atayan ağırlık fonksiyonudur. $m(u)$ 'ya yerel kuadratik polinom ile yaklaşıldığını dikkate aldığımızda aşağıdaki eşitlik durumu elde edilir.

$$m(u) \approx a_0 + a_1(u - x) + \frac{1}{2}a_2(u - x)^2$$

Burada $|u - x| < h(x)$ 'dir. Polinomlar için kapalı vektör gösterimini kullanarak aşağıdakini yazabiliriz:

$$a_0 + a_1(u - x) + \frac{1}{2}a_2(u - x)^2 = \langle a, A(u - x) \rangle.$$

Burada $a = \begin{pmatrix} a_0 \\ a_1 \\ a_2 \end{pmatrix}$ katsayılar vektörü, $A(v) = \begin{pmatrix} 1 \\ v \\ \frac{v^2}{2} \end{pmatrix}$ uyumlama fonksiyonlarının bir vektörü, $\langle . \rangle$ ise iç çarpım sembolüdür. Katsayılar vektörü a , yerel olarak ağırlıklandırılmış kareler toplamının minimizasyonu ile tahminlenebilir.

$$\sum_{i=1}^n w_i(x)(Y_i - \langle a, A(x_i - x) \rangle)^2 \quad (2.21)$$

Yukarıda gözlemlendiği gibi yerel regresyon tahmini en küçük kareler problemini çözdüğünden, $\hat{m}(x)$ doğrusal bir tahmin olur. Böylece her bir x için

$l(x) = \{l_i(x)\}_{i=1}^n$ şeklinde bir ağırlık diyagramı vektörü mevcuttur. Tahmin aşağıdaki gibi ifade edilebilir.

$$\hat{m}(x) = \sum_{i=1}^n l_i(x) \cdot Y_i \quad (2.22)$$

Ağırlık diyagramı, yerel regresyon tahmininin ortalama ve varyansının kapalı yapıda ifadesini sağlar.

$$\begin{aligned} E(\hat{m}(x)) &= \sum_{i=1}^n l_i(x) m(x_i) = \langle l(x), m \rangle \\ \text{Var}(\hat{m}(x)) &= \sigma^2 \sum_{i=1}^n l_i(x)^2 = \sigma^2 \|l(x)\|^2 \end{aligned} \quad (2.23)$$

Varyans ifadesi Y_i gözlemlerinin bağımsız ve sabit σ^2 varyansına sahip olduğunu varsayar. Varyans indirgeme faktörü $\|l(x)\|^2$, yerel regresyondan dolayı varyanstaki azalışı ölçer. Genellikle bu değer azalırken band genişliği değeri artar.

Verileri uyumlanan değerlere görüntüleyen ve satırları $l(x_i)^T$ olan $n \times n$ boyutundaki şapka matrisi (hat matrix) L olmak üzere;

$$\begin{pmatrix} \hat{m}(x_1) \\ \vdots \\ \hat{m}(x_n) \end{pmatrix} = LY$$

Şapka matrisinin köşegen elemanları $l_i(x_i)$ etki (influence) değerleridir. Bu değerler $infl(x_i)$ ile gösterilirler ve veri noktalarından, uyumlanan eğri $\hat{m}(x_i)$ 'nin duyarlılığını ölçerler.

Yerel uyumlamada serbestlik derecesi, aynı veri setine uygulanan farklı tahminlemeler arasında kıyaslama yapılabilmesi açısından düzeltme miktarının ölçülmesine olanak sağlar. Serbestlik derecesi için mümkün pek çok tanımlama bulunmaktadır, bunlardan en sık kullanılan ikisi aşağıdaki gibidir.

$$\begin{aligned} v_1 &= \sum_{i=1}^n infl(x_i) = tr(L) \\ v_2 &= \sum_{i=1}^n \|l(x_i)\|^2 = tr(L^T L) \end{aligned} \quad (2.24)$$

Parametrik regresyon modeli için şapka matrisi L simetrik ve idempotenttir, serbestlik derecesi tanımları genellikle parametre sayısıyla denk olacak şekilde örtüşmektedir. Yerel regresyon modelleri için iki tanımlama genelde eşit değildir.

Denklem (2.23) ile verilen basit ortalama ve varyans tanımlarını biraz daha geliştirelim. Bu noktada ilk olarak ağırlık diyagramı belirlenir. X , satırları $A(x_i - x)^T$ olan $n \times (p + 1)$ boyutlu tasarım matrisi, W , $w_i(x)$ girişleri ile köşegen matris ve $Y = (Y_1, Y_2, \dots, Y_n)^T$ yanıt vektörüdür. Ağırlıklandırılmış kareler toplamı, matris yapısında aşağıdaki gibi ifade edilir.

$$(Y - Xa)^T W (Y - Xa)$$

Eğer WX full sütun rankına (full column rank) sahip, yani sütunları doğrusal olarak bağımsız ise, en küçük kareler teorisi açık bir ifade verir. Denklem (2.21)'in minimize edicisi;

$$\hat{a} = (X^T W X)^{-1} X^T W Y \quad (2.25)$$

şeklinde hesaplanır. Bu eşitlik, (2.22)'de yer alan ağırlık diyagramını yerel polinomiyal düzeltme için tanımlar,

$$l(x)^T = e_1^T (X^T W X)^{-1} X^T W \quad (2.26)$$

burada $e_1 = (1, 0, \dots, 0)^T$ şeklinde birim vektördür. Aşağıdaki teorem Henderson (1916) tarafından ortaya konulmuştur ve yerel polinomiyal regresyonda ağırlık diyagramlarının tanımlanmasını sağlamaktadır.

Teorem 2.2 (Henderson Teoremi) (İspatı için bakınız Loader 1999)

p dereceli bir yerel polinomiyal uyum için ağırlık diyagramı aşağıdaki yapıya sahiptir.

$$l_i(x) = W \left(\frac{x_i - x}{h(x)} \right) \langle \alpha, A(x_i - x) \rangle \quad (2.27)$$

Yani en küçük kareler ağırlıkları, p dereceli bir polinom ile çarpılır. Bu gösterim tektir ve $\mathbf{X}^T \mathbf{W} \mathbf{X}$ tekil değilken (non-singular) sağlanır.

Aksine, eğer doğrusal tahmin p dereceli polinomlarla yeniden üretilirse ve ağırlık diyagramı en fazla p işaret değişikliğine sahip ise tahmin p dereceli bir yerel polinomiyal uyum olarak ifade edilebilir.

Henderson teoremi ile (2.23)'de yer alan varyans ifadesi aşağıdaki (2.28) ile yazılabilir (Loader 1999).

$$\text{Var}(\hat{m}(x)) = \sigma^2 e_1^T (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} (\mathbf{X}^T \mathbf{W}^2 \mathbf{X}) (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} e_1 \quad (2.28)$$

Varyans ile yakın ilişkili bir nicelik (2.29)'da verilen etki fonksiyonudur.

$$\text{infl}(x) = e_1^T (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} e_1 W(0) \quad (2.29)$$

Bu tanımlama etki fonksiyonunun daha önceki kullanımı ile genelleştirilir; şapka matrisinin \mathbf{L} köşegen elemanları $l_i(x_i) = \text{infl}(x_i)$ olur.

Denklem (2.22)'de yer alan doğrusal tahminin sapması;

$$E(\hat{m}(x)) - m(x) = \sum_{i=1}^n l_i(x) m(x_i) - m(x) \quad (2.30)$$

olarak ifade edilir. $\hat{m}(x)$, p dereceli bir yerel polinomiyal uyum olsun. $m(x)$ 'in $(p+2)$ kez türevlenebilir olduğu varsayımı ile x etrafında $m(\cdot)$ fonksiyonunu Taylor serisi olarak genişletebiliriz.

$$\begin{aligned} m(x_i) = & m(x) + (x_i - x)m'(x) + \dots + (x_i - x)^p \frac{m^p(x)}{p!} \\ & + (x_i - x)^{p+1} \frac{m^{p+1}(x)}{(p+1)!} + (x_i - x)^{p+2} \frac{m^{p+2}(x)}{(p+2)!} + \dots \end{aligned}$$

Henderson teoreminin bir uygulaması olarak, $1 \leq j \leq p$ için $\sum_{i=1}^n l_i(x)(x_i - x)^j = 0$ 'dır. Buradan;

$$\begin{aligned}
E(\hat{m}(x)) - m(x) &= \\
&= \frac{m^{(p+1)}(x)}{(p+1)!} \sum_{i=1}^n l_i(x)(x_i - x)^{p+1} + \\
&\quad + \frac{m^{(p+2)}(x)}{(p+2)!} \sum_{i=1}^n l_i(x)(x_i - x)^{p+2} + \dots \quad (2.31)
\end{aligned}$$

yazılabilir. Sapma ifadesi, $(p+1)$. türev $\hat{m}^{(p+1)}(x)$ 'i ilk terim olarak içerir. Eşitlikte neden $\hat{m}^{(p+2)}(x)$ terime bakılmıştır sorusunu şöyle açıklayabiliriz; tasarım noktalarının eşit aralıklandırıldığı varsayalım. Bu durumda ağırlık diyagramı uyumlama noktası x etrafında simetrik olur. Eğer p çift ise $p+1$ tek olur ve simetriklikten dolayı $\sum_{i=1}^n l_i(x)(x_i - x)^{p+1} = 0$ olur. Bu nedenle sapma ifadesinde ilk terim ortadan kaybolur. Böyle bir durumda eşitlikte ikinci terim daha baskındır. Düzgün olmayan (nonuniform) veriler ve çift p için, $m^{(p+1)}$ ve $m^{(p+2)}$ terimleri genellikle benzer büyüklükte dirler (Loader 1999).

2.2.3.2. Yerel polinomun derecesi

En küçük kareler eşitliğinde (2.17) kullanılan düzeltme parametresinde olduğu gibi, yerel polinomun derecesi de sapma-varyans dengesi üzerinde etkilidir. Yüksek polinom derecesi ile ortalama $m(x)$ temelinde, düşük dereceli bir polinoma göre daha iyi bir yaklaşım sağlayabilir. Bu nedenle, yüksek dereceli bir polinomla uyumlama genellikle $\hat{m}(x)$ tahminine daha küçük bir sapma ile ulaşılmasını sağlar. Ancak yüksek dereceli polinom ile gerçekleştirilen tahminlemede, tahminleme katsayılarının fazlalığı ve sonuçların değişkenliği söz konusudur. Örneğin; eğer yerel kuadratik tahmin ve yerel doğrusal tahmin aynı band genişliği için hesaplanırsa, yerel kuadratik tahmin çok daha değişken olacaktır. Ancak bu noktada varyanstaki artış band genişliğinin artırılması ile tolere edilebilir.

Bu durumda polinom derecesini düşük seçmek ve band genişliği üzerine yoğunlaşmak, istenilen uyumun elde edilmesi konusunda yeterlidir. En yaygın

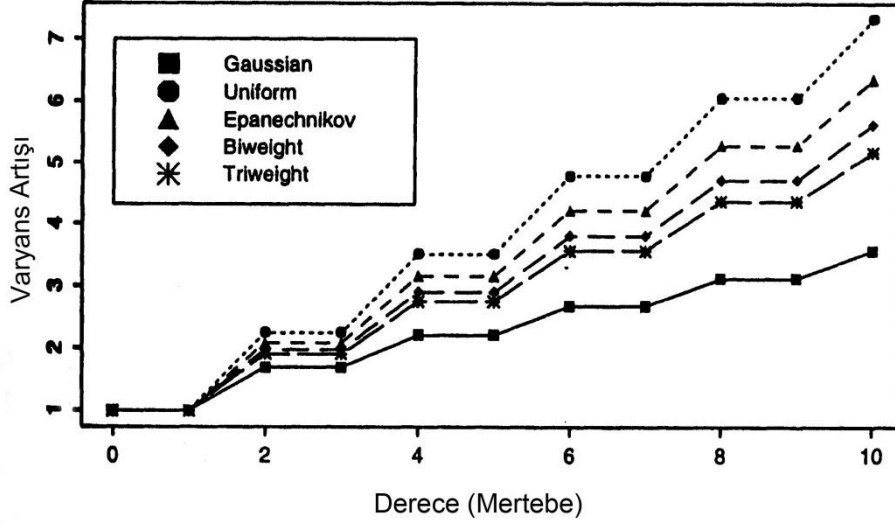
gerçekleştirilen seçimler yerel doğrusal ve yerel kuadratik yapılardır. Yerel kübik ve daha yüksek dereceden uyumlar nadiren daha faydalı sonuçlar üretirler (Loader 1999).

p uyumlanan eğrinin polinom derecesini göstermek üzere Fan ve Gijbels (1996), polinom derecesi ile çekirdek fonksiyonlarında gözlemlenen değişkenliği Nadaraya-Watson tahmin edicisinin ($p = 0$) varyansına göre karşılaştırmalı olarak (V_p/V_0) Çizelge 2.2 ile açıklamışlardır.

Çizelge 2.2 p dereceli polinom ve çeşitli çekirdek fonksiyonları için değişkenlikteki artış (Fan ve Gijbels 1996)

p	Gaussian	Uniform	Epanechnikov	Biweight	Triweight
1	1	1	1	1	1
2	1.6876	2.2500	2.0833	1.9703	1.9059
3	1.6876	2.2500	2.0833	1.9703	1.9059
4	2.2152	3.5156	3.1550	2.8997	2.7499
5	2.2152	3.5156	3.1550	2.8997	2.7499
6	2.6762	4.7852	4.2222	3.8133	3.5689
7	2.6762	4.7852	4.2222	3.8133	3.5689
8	3.1224	6.0562	5.2872	4.7193	4.3753
9	3.1224	6.0562	5.2872	4.7193	4.3753
10	3.5704	7.3281	6.3509	5.6210	5.1744

Şekil 2.2 polinom derecesi ile çeşitli çekirdek fonksiyonları için değişkenlikteki artışı göstermektedir. Örneğin, $p = 2$ ve $p = 3$ dereceli polinomlar için değişim (varyans) aynıdır, ancak polinom derecesinin yüksek seçilmesine bağlı olarak $p = 3$ için aynı varyansta, sapma daha küçük olduğundan kübik polinom tercih edilebilir.



Şekil 2.2 p dereceli polinom ve çeşitli çekirdek fonksiyonları için değişkenlikteki artış grafiği (Fan ve Gijbels 1995)

Çizelge ve Şekil 2.2'den gözlemlendiği gibi tek dereceli yaklaşımdan çift dereceli yaklaşıma geçildiğinde asimtotik varyansın arttığı görülür. Örneğin, Epanechnikov çekirdeğinde, yerel doğrusal uyum yerine yerel kuadratik uyum kullanıldığında varyans 2.0833 faktörü ile artar. Ayrıca Gauss çekirdek fonksiyonunun diğer çekirdek fonksiyonlarına göre değişkenliği daha az olmakla birlikte Uniform çekirdek fonksiyonu en yüksek değişkenliği gösterir (Fan ve Gijbels 1996).

Bir regresyon fonksiyonunun tahmini üzerine odaklanıldığında, yerel doğrusal uyumlar, yerel kübik uyumlar ya da başka bir tek dereceli uyumun gerçekleştirilmesi şeklinde sınırlama getirilebilir. Tek derece ile gerçekleştirilen uyumlar asimtotik olarak daha iyi performans göstermelerinden dolayı yerel sabit ya da yerel kuadratik uyumlar tercih edilmez. Genel olarak iki tek dereceli uyum arasında doğrudan bir karşılaştırma yoktur ancak düşük dereceli bir uyumun sonuçları daha büyük sapmaya ve daha küçük varyansa sahip olurlar, derece arttıkça sapma azalır ancak varyans büyür (Fan ve Gijbels 1996).

2.2.3.3. Ağırlık fonksiyonu

Yerel polinomial düzeltme tekniğinde ağırlık fonksiyonu sapma-varyans dengesi üzerinde daha az etkiye sahip olmakla birlikte uyumlanan regresyon eğrisinin görsel kalitesi açısından önemlidir. En basit ağırlık fonksiyonu dikdörtgen (rectangular) fonksiyondur.

$$W(u) = \begin{cases} |u| < 1, & 1/2 \\ 0, & d.d \end{cases}$$

Ancak bu ağırlık fonksiyonu sürekli olmayan ağırlıklar $w_i(x)$ ve sürekli olmayan uyum eğrisine götürmesinden dolayı nadiren kullanılır. Genellikle sürekli simetrik, sıfır noktası etrafında yoğunlaşan ve $[-1, 1]$ aralığı üzerinde desteklenen $W(u)$ ağırlığı seçilir. En yaygın seçilen ağırlık fonksiyonlarına Çizelge 2.1'de yer verilmiştir (Loader 1999).

2.3. Parametrik Olmayan Regresyonda Düzeltme Parametresinin Seçimi

Parametrik olmayan regresyonda uyumlamanın gerçekleştirilebilmesi için her zaman bir düzeltme parametresi seçimine ihtiyaç duyulmaktadır. İyi bir uyumlamanın gerçekleştirilebilmesi için seçilecek olan düzeltme parametresi değeri önemli bir yer tutmaktadır. Hastie ve Tibshirani (1990) düzeltme yöntemlerinde çekirdek fonksiyonu ve düzeltme parametresi h 'ın seçiminin önemliliği konusuna değinmişler ve çekirdek fonksiyonunun seçiminin düzeltme parametresinin seçimine göre daha önemsiz olduğunu ortaya koymuşlardır. Söz konusu optimal band genişliği seçimi olduğunda, hangi seçim yönteminin daha iyi olduğu konusunda bir uzlaşma sağlanamamıştır.

Düzeltilme parametresi, örneklem genişliğine bağlıdır. Uygulamadaki başlıca problem bu parametrenin seçimidir. Genellikle herhangi bir global hata ölçüsünü en küçükleyecek şekilde varyans ve sapmanın karesi arasındaki ödünleşimi dengelemek için seçilir (Vieu 1991).

Çekirdek regresyonda, çok büyük bir band genişliği alınması fazla düzleştirmeye (oversmoothing), çok küçük bir band genişliği alınması ise az

düzleştirmeye (undersmoothing) neden olur. Amaç, az düzleştirme ve fazla düzleştirme arasındaki dengeyi (trade-off) sağlayan bir h değeri bulmaktır (Hardle 1994).

Literatürde yaygın olarak parametrik olmayan regresyon tahmininde kullanılan düzeltme parametresi seçim yöntemleri çarpaz geçerlilik kriteri (CV), genelleştirilmiş çarpaz geçerlilik kriteri (GCV), geliştirilmiş Akaike Bilgi kriterleri (AIC_c), Mallow'un C_p kriteri, klasik pilotları kullanan risk tahminleri ve yerel risk tahmin kriterleridir. Bu yöntemlerden çarpaz geçerlilik kriteri (CV), genelleştirilmiş çarpaz geçerlilik kriteri (GCV) ve geliştirilmiş Akaike Bilgi kriterleri (AIC_c) izleyen bölümde ayrıntılı olarak incelenecektir. Herhangi bir seçim kriterini minimum yapan h , uygun düzeltme parametresi olarak seçilir.

2.3.1. Çarpaz geçerlilik kriteri

Parametrik olmayan regresyonda optimal band genişliği seçimi için sık kullanılan yöntemlerden bir tanesi çarpaz geçerlilik kriteridir. Bu kriter Wahba ve Wold (1975) tarafından önerilmiştir. Birini dışarıda bırakma çarpaz geçerlilik (Leave one out cross validation) kriteri farklı örneklem büyüklüklerinde oldukça iyi sonuçlar veren ve parametrik olmayan regresyonda en çok kullanılan otomatik düzeltme tekniğidir. Bu teknikte bir gözlem seçilir ve veri setinden çıkartılır. Daha sonra olası modellerden biri budanmış (truncated) veri setine uygulanarak uyumun ölçüsü hesaplanır. Daha sonra başka bir gözlem veri setinden çıkartılarak aynı işlem yapılır. Bu süreç her bir gözlem veri setinden çıkartılıncaya kadar tekrar edilir. Çarpaz geçerlilik skoru ortalama uyum ölçüsüdür ve farklı modelleri karşılaştırmak için kullanılır (Keele 2008).

Çarpaz geçerlilik, $\{x_i, y_i\}_{i=1}^n$ gözlem noktaları olmak üzere, bu gözlemlerden bir veri noktasını çıkartarak kalan $(n - 1)$ veri noktasına dayalı olarak x_i noktasında bilinmeyen fonksiyon için kareli artıkları tahmin eden ve bu tahmini minimum yapan düzeltme parametresini seçen yöntemdir. $y^{(-i)}$, bağımlı değişken vektöründen y_i gözlemini çıkarttıktan sonra kalan gözlemlerin oluşturduğu vektör ve bu vektöre uygun tahmin fonksiyonu $\hat{m}_\lambda^{(-i)}(x)$ olmak üzere tahmin hatasına ait çarpaz geçerlilik kriteri (2.32) ile ifade edilir.

$$CV(\lambda) = n^{-1} \sum_{i=1}^n \{y_i - \hat{m}_\lambda^{(-i)}\}^2 \quad (2.32)$$

Eşitlikten anlaşılacağı üzere optimum düzeltme parametresi λ 'yı elde etmek için n tane \hat{m}^{-i} eğrisini bulmak ve n farklı düzeltme problemini çözmek gerekmektedir. İşlemleri kolaylaştırmak amacıyla eşitlik (2.32)'de yer alan ve y_i gözlem değerleri vektörünü tahmin değerlerine görüntüleyen S_λ şapka matrisi kullanılarak eşitlik aşağıdaki gibi yazılabilecektir.

$$CV(\lambda) = \frac{1}{n} \sum_{i=1}^n \left\{ \frac{y_i - \hat{m}_\lambda(x_i)}{1 - (S_\lambda)_{ii}} \right\}^2 \quad (2.33)$$

Denklem (2.32)'de yer alan $(S_\lambda)_{ii}$, S_λ matrisinin i . köşegen elemanını ifade eder. Çapraz geçerlilik kriterini minimum yapan λ değerinin bulunması ile işlem sonlandırılır (Green ve Silverman 2000).

2.3.2. Genelleştirilmiş çapraz geçerlilik kriteri

Çapraz geçerlilik kriterinin tanımlamasından anlaşılacağı gibi özellikle büyük veri setleri üzerinde yapılan çalışmalarda bu kriterin hesaplanması oldukça güçtür. Bu nedenle Craven ve Wahba (1979) hesaplamasal açıdan daha az zaman isteyen genelleştirilmiş çapraz geçerlilik kriterini önermişlerdir. Genelleştirilmiş çapraz geçerlilik kriteri özünde şapka matrisinin köşegen elemanları yerine, bu matrisin izini kullanarak hesaplamaları gerçekleştirir. Genelleştirilmiş çapraz geçerlilik kriteri denklem (2.34) ile tanımlanır.

$$\begin{aligned} GCV(\lambda) &= \frac{1}{n} \frac{\sum_{i=1}^n \{y_i - \hat{m}_\lambda(x_i)\}^2}{\{1 - n^{-1} \text{tr}(S_\lambda)\}^2} \\ &= \frac{\|(I - S_\lambda)\mathbf{y}\|^2}{\{1 - n^{-1} \text{tr}(S_\lambda)\}^2} \\ &= \frac{\|(I - S_\lambda)\mathbf{y}\|^2}{\{1 - n^{-1} df_{fit}\}^2} \end{aligned} \quad (2.34)$$

Burada, $\mathbf{y} = (y_1, y_2, \dots, y_n)'$ gözlem değerleri vektörünü, $(\hat{m}(x_1), \hat{m}(x_2), \dots, \hat{m}(x_n))' = S_\lambda \mathbf{y}$ tahminlere ait vektörü ve $df_{fit} = tr(S_\lambda)$ modelin serbestlik derecesini ifade eder (Ruppert ve ark. 2003). Çapraz geçerlilik kriterinde olduğu gibi genelleştirilmiş çapraz geçerlilik kriterinde de amaç $GCV(\lambda)$ değerini minimum yapan λ düzeltme parametresini bulmaktır.

2.3.3. Geliştirilmiş Akaike Bilgi kriteri

Hurvich ve ark. (1998) çekirdek regresyon için ilk tanımlanan AIC kriterine göre daha az yanlılık gösteren, düzeltme parametresi seçim ölçütü olarak Geliştirilmiş Akaike Bilgi Kriterini (AIC_c) önermişlerdir. Geliştirilmiş AIC_c kriteri,

$$AIC_c(\lambda) = \log(\hat{\sigma}^2) + \frac{1+tr(S_\lambda)/n}{1-\{tr(S_\lambda)+2\}/n} = \log(\hat{\sigma}^2) + 1 + \frac{2\{tr(S_\lambda)+1\}}{n-tr(S_\lambda)-2} \quad (2.35)$$

formülüyle hesaplanır. Hesaplamalarda şapka matrisinin izinin kullanılıyor olması, düzeltme parametresinin seçimi konusunda uygulanması kolay bir kriterdir. Varyans fonksiyonu tanımından ve basit cebirsel işlemlerin uygulanmasının ardından formül aşağıdaki şekli alır.

$$\begin{aligned} AIC_c(\lambda) &= \log \frac{\sum \{y_i - \hat{m}_\lambda(x_i)\}^2}{n} + 1 + \frac{2\{tr(S_\lambda) + 1\}}{n - tr(S_\lambda) - 2} \\ &= \log \frac{\|(S_\lambda - I)\mathbf{y}\|^2}{n} + 1 + \frac{2\{tr(S_\lambda + 1)\}}{n - tr(S_\lambda) - 2} \end{aligned} \quad (2.36)$$

Diğer düzeltme parametresi seçim kriterlerinde olduğu gibi amaç, AIC_c kriterini minimum yapan λ düzeltme parametresi değerini belirlemektir.

2.4. Parametrik Olmayan Regresyon Modellerinde Performans Kriterleri

Parametrik olmayan regresyon modellerinde tahminleyenin ya da modelin performansı söz konusu olduğunda, tahminleyenin yansız ve varyansının küçük olması istenilen bir durumdur. Ayrıca model performansı veri seti için, tahmin edilen değer ile gözlenen değer birbirine ne kadar yakın olduklarıyla ilişkilidir. Parametrik olmayan modelde böyle bir tahminleyen için yanlılık ve varyans arasındaki dengenin oluşturulması model performansı açısından yeterli olacaktır.

Bu konu aslında araştırmacıyı modellerde band genişliği seçimi ile karşı karşıya getirmektedir. Band genişliği seçim yöntemlerinin temelinde de yan ve varyans arasında kurulan denge ön plandadır. Bu dengeyi kurmak için uygulamada çeşitli kriterler kullanılmaktadır. Bu kriterlerden en sık kullanılanı hata kareler ortalamasıdır (Mean Squared Error) (MSE). $\hat{m}(x)$, $m(x)$ fonksiyonunun tahmini olmak üzere;

$$\begin{aligned} MSE(\hat{m}(x)) &= E\{\hat{m}(x) - m(x)\}^2 \\ &= E[(\hat{m}(x) - M(x) + M(x) - m(x))^2] \\ &= E[(\hat{m}(x) - M(x))^2] + E[(M(x) - m(x))^2] + 2E[(\hat{m}(x) - M(x))(M(x) - m(x))] \\ &= Var(\hat{m}(x)) + (Yan(\hat{m}(x)))^2 + 2E[(\hat{m}(x)M(x) - \hat{m}(x)m(x) - (M(x))^2 + M(x)m(x))] \\ &= Var(\hat{m}(x)) + (Yan(\hat{m}(x)))^2 + 2[(M(x))^2 - M(x)m(x) - (M(x))^2 + M(x)m(x)] \\ &= Var(\hat{m}(x)) + (Yan(\hat{m}(x)))^2 \end{aligned} \quad (2.37)$$

şeklinde ifade edilir. Denklem (2.37)'de $M(x) = E[\hat{m}(x)]$ 'dir ve MSE ölçümünde ilk terim varyansı ölçerken ikinci terim yanlılığın karesini ölçmektedir (Hastie ve Tibshirani 1990). Temel fikir sapma ve varyans arasında dengenin kurularak MSE 'nin minimum değerinin elde edilmesidir. Ancak bu ölçüm tek nokta üzerinde çalışır bu yüzden yalnızca yerel hatayı minimize eder. Bu nedenle değişken band genişliği seçiminde kullanılması daha uygundur. Daha genel bir değerlendirme için entegre hata kareler toplamı (Mean Integrated Squared Error) ($MISE$) kullanılabilir.

$$MISE(\hat{m}(x)) = \int MSE(\hat{m}(x))w(x)dx \quad (2.38)$$

Denklem (2.38)'de yer alan eşitlikte w pozitif bir ağırlık fonksiyonudur. MSE ifadesinin, sapma ve varyansın toplamı şeklinde yazılabilesinden dolayı $MISE$ değerinin hesaplaması da aynı şekilde gerçekleştirilebilir (Fan ve Gijbels 1996).

Uygulama kolaylığı olan diğer bir hata ölçüm kriteri ortalama karesel hatadır (Averaged Squared Error) (ASE). Bu ölçüm denklem (2.39) ile tanımlanır.

$$ASE(\hat{m}(x)) = n^{-1} \sum_{i=1}^n (\hat{m}(x_i) - m(x_i))^2 \quad (2.39)$$

ASE rassal bir değişkendir ve genel bir ölçümü gerçekleştirir, bu özellik ona tüm eğriyi belirleyebilme yeteneği verir. Ayrıca ASE 'nin beklenen değeri de yine hata ölçümünde kullanılabilen bir kriterdir ve (2.40) ile ifade edilir.

$$MASE(\hat{m}(x)) = E(ASE(\hat{m}(x))) \quad (2.40)$$

Eşitliklerde görüldüğü gibi bu kriterler aslında birer uzaklık ölçüsünü gösterirler. Kriterlerin hepsi asimtotik olarak aynı düzeltme seviyesine ulaşmalarından dolayı hatanın ölçümünde herhangi birinin kullanılması uygundur. Hesaplamasal olarak en kolay kriter ASE olmasından dolayı çoğu araştırmada bu kritere yer verilmiştir (Hardle 1991).

3. BULANIK FONKSİYONLAR VE TÜREV

Bulanık mantık kavramı ilk olarak, Lotfi A. Zadeh (1965) tarafından “Fuzzy Sets” adlı makalesinde tanımlanmış ve bu makale bulanık küme teorisinin temellerini oluşturmuştur. Zadeh, niteliklerin ikili üyelik fonksiyonu ile ifade edildiği klasik kümeler yerine, dereceli üyelik fonksiyonu ile ifade edildiği bulanık kümeler tanımlamasını önermiştir. Sonrasında Zadeh (1978) yapmış olduğu çalışmada bulanık teoriye önemli katkılar sağlayan genişletme prensibini ifade etmiştir. En basit şekliyle “bulanıklık” kavramı, bir ifade de yer alan belirsizlik olarak tanımlanabilir. Günlük hayatta kullandığımız ve tanımladığımız pek çok şey özünde bulanıklık içermektedir. Örneğin, “yaşlı insan”, “yüksek ısı”, “küçük sayı” gibi ifadeler kesinlikten ziyade bulanıklığı içerirler (Terano ve ark. 1992). Hayatımızda bu denli çok belirsizlik barındıran durum söz konusu olduğuna göre bulanık mantık teorisinin geliştirilmesi ve kesin teoride yer alan kavram ve tanımlamaların bulanık yapıda ifadesi önemli bir konudur.

Parametrik olmayan bulanık regresyon modellerinin tanımlanması ve incelenmesi için, bulanık mantık teorisinden bazı temel kavram ve bilgilere ihtiyaç duyulacaktır. Bu amaçla, bu bölümde bulanık teoride yer alan bazı gösterimlerin açıklanması bir sonraki bölüm için hazırlık niteliğindedir.

Çalışmanın bu bölümünde literatürde sıkça kullanılan üçgen bulanık sayı, yamuk bulanık sayı ve *LR* bulanık sayılarının tanımları verilecektir. Bulanık mantık teorisinde pek çok tanımlamanın yapılmasında kullanılan genişletme prensibi ele alınarak, ardından bulanık sıralama yöntemleri üzerinde durulacaktır. Son olarak ise bazı bulanık fonksiyonların tanımları ile tanım bölgesi kesin küme olan bulanık fonksiyonlar için bazı türev kavramları verilecektir.

3.1. Bulanık Küme ve Bulanık Sayılar

X klasik bir küme (universe) ve $\mu_A(x):X \rightarrow [0,1]$ üyelik fonksiyonunu göstermek üzere, bir A bulanık kümesi aşağıdaki çiftlerin kümesi olarak tanımlanır:

$$A = \{(x, \mu_A(x)), x \in X\} \quad (3.1)$$

A ve B iki bulanık küme olmak üzere, $\forall x \in X$ için $\mu_A(x) = \mu_B(x)$ ise, A ve B kümeleri eşittir denir. İki bulanık kümenin birleşimi ve kesişimine ait üyelik fonksiyonları şöyle tanımlanır:

$$\forall x \in X, \mu_{A \cup B}(x) = \max(\mu_A(x), \mu_B(x))$$

$$\forall x \in X, \mu_{A \cap B}(x) = \min(\mu_A(x), \mu_B(x))$$

A bulanık kümesinin tamamlayıcısı (tümleyeni) olan \bar{A} kümesi ise,

$$\forall x \in X, \mu_{\bar{A}}(x) = 1 - \mu_A(x)$$

üyelik fonksiyonu ile tanımlanır.

A bulanık kümesinin $\alpha \in [0,1]$ için α -kesimi $A_\alpha = \{x \in X | \mu_A(x) \geq \alpha\}$ olarak tanımlanan bir kesin kümedir. Bulanık kümelerin α -kesimleri ile aşağıdaki özellikler kolayca elde edilebilir:

$$(A \cup B)_\alpha = A_\alpha \cup B_\alpha \quad \text{ve} \quad (A \cap B)_\alpha = A_\alpha \cap B_\alpha$$

Bir bulanık A kümesini onun α -kesimleri yardımıyla ifade edebiliriz. $\alpha \in [0,1]$ olmak üzere bir bulanık A kümesinin α -kesimi, A kümesindeki üyelik derecesi α 'ya eşit ve daha büyük olan tüm elemanlarını içeren bir kesin A_α kümesidir. Bulanık teoride önemli bir yere sahip olan kararlılık prensibi (resolution principle) bir bulanık A kümesinin α -kesimleri bakımından genişletilebileceğini göstermektedir. A bulanık kümesinin üyelik fonksiyonu onun α -kesimlerinin karakteristik fonksiyonları yardımıyla,

$$\mu_A(x) = \sup_\alpha \min(\alpha, \mu_{A_\alpha}(x)), \quad x \in X \quad (3.2)$$

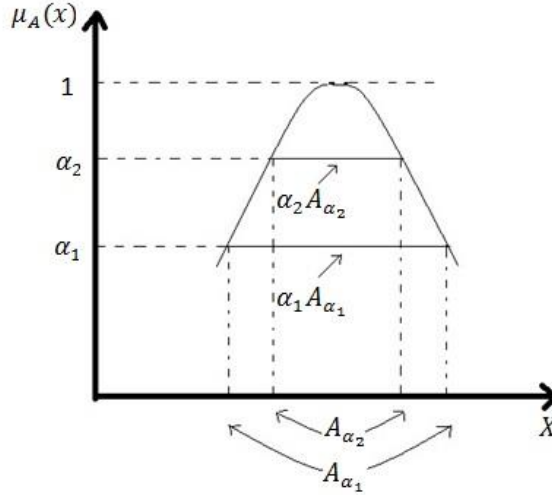
formülü ile ifade edilebilir. Burada,

$$\mu_{A_\alpha}(x) = \begin{cases} 1 & x \in A_\alpha \\ 0 & d.d \end{cases}$$

Üyelik fonksiyonunu $\min(\alpha, \mu_{A_\alpha}(x))$ ile ifade edilen bulanık kümeyi αA_α olarak yazdığımızda A bulanık kümesi için kararlılık prensibi, $\Lambda_\alpha = \{\alpha | \exists x \in X, \mu_{A_\alpha}(x) = \alpha\}$ olmak üzere aşağıdaki şekilde ifade edilebilir.

$$A = \bigcup_{\alpha \in \Lambda_\alpha} \alpha A_\alpha \quad (3.3)$$

Kararlılık prensibinin şekil üzerinde ifadesi 3.1'de verilmiştir.



Şekil 3.1 Bulanık kümelerde kararlılık prensibi

Bir bulanık kümenin maksimum üyelik değeri 1'e eşit ise normallik koşulu sağlanmış olur ve aşağıdaki gösterimle ifade edilir.

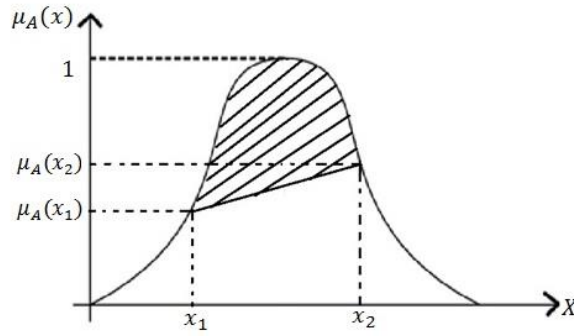
$$\exists x \in X, \mu_A(x) = 1 \quad (3.4)$$

Bir bulanık küme ancak ve ancak her bir α -kesiminin konveks küme olması koşulu ile konveks olarak tanımlanır. $\forall x_1 \in X, \forall x_2 \in X$ ve $\lambda \in [0,1]$ için,

$$\mu_A(\lambda x_1 + (1 - \lambda)x_2) \geq \min(\mu_A(x_1), \mu_A(x_2)) \quad (3.5)$$

ise A bulanık kümesi konvektir denir.

Konveks A bulanık kümesinden alınan $(x_1, \mu_A(x_1))$ ve $(x_2, \mu_A(x_2))$ noktalarını birleştirecek bir doğru çizildiğinde, bu doğru üzerindeki tüm noktaların üyelik dereceleri $\mu_A(x_1)$ ve $\mu_A(x_2)$ 'nin minimumuna eşit veya daha büyük olur (Lin ve Lee 1996). Şekil 3.2'de, verilen bir üyelik fonksiyonu ile ifade edilen bulanık kümenin konveks ve normalliği görülmektedir.



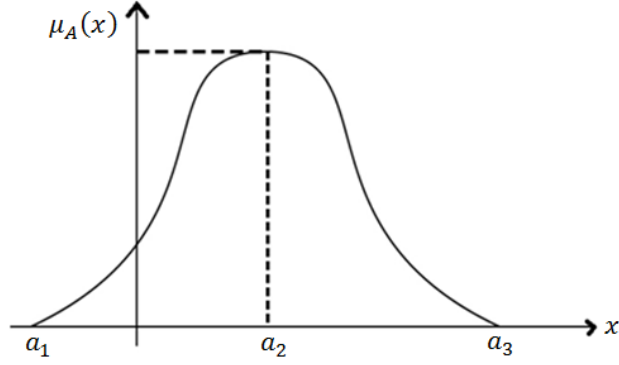
Şekil 3.2 Konveks ve normal bulanık küme

\Re reel sayılar kümesinin normal ve konveks bir bulanık A kümesine bulanık sayı denir. Burada μ_A üyelik fonksiyonunun kısmi-sürekli olması istenir.

Bulanık sayı, \Re gerçel sayılar kümesinde bulanık aralık olarak tanımlanan bir bulanık küme şeklinde ifade edilebilir. Bu aralığa ait sınırlar belirsiz olduğundan, aralıklar aynı zamanda bir bulanık kümedir. Bulanık aralıklar, $[a_1, a_2, a_3]$ şeklinde iki sınır noktası a_1, a_3 ve bir tepe noktası a_2 yardımıyla gösterilirler. Bu açıklamalardan yola çıkılarak görülmektedir ki bulanık sayılara α -kesim işlemi uygulanabilir. Eğer A bulanık sayısı için α -kesim aralığını gösterirsek, tanımlanan A_α aralığı aşağıdaki gibi olur.

$$A_\alpha = [a_1^{(\alpha)}, a_3^{(\alpha)}] \quad (3.6)$$

Tanımı gerçekleştirilen bulanık sayı ifadesi Şekil 3.3 ile gösterilmiştir (Lee 2005).



Şekil 3.3 Bulanık sayı

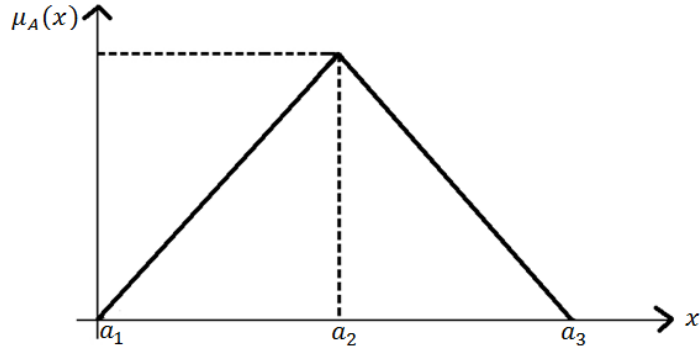
Böylelikle, bulanık sayı, kısmı-sürekli üyelik fonksiyonuna sahip \mathfrak{R} reel sayılar doğrusunda normal ve dışbükey (konveks) bulanık alt kümedir (Lin ve Lee 1996).

3.1.1. Üçgen bulanık sayılar

$A = (a_1, a_2, a_3)$ bir bulanık sayı olmak üzere, a_1 bulanık sayının alt sınırı, a_2 bulanık sayının merkezi ve a_3 bulanık sayının üst sınırını göstermek üzere aşağıdaki üyelik fonksiyonu ile tanımlanması durumunda üçgen bulanık sayı olarak adlandırılır.

$$\mu_A(x) = \begin{cases} 0, & x < a_1 \\ \frac{x-a_1}{a_2-a_1}, & a_1 \leq x \leq a_2 \\ \frac{a_3-x}{a_3-a_2}, & a_2 \leq x \leq a_3 \\ 0, & x > a_3 \end{cases} \quad (3.7)$$

Eğer $a_2 - a_1 = a_3 - a_2$ ise simetrik üçgen bulanık sayı olarak adlandırılır ve üyelik fonksiyonu Şekil 3.4 ile gösterilir (Lee 2005).



Şekil 3.4 Üçgen Bulanık Sayı

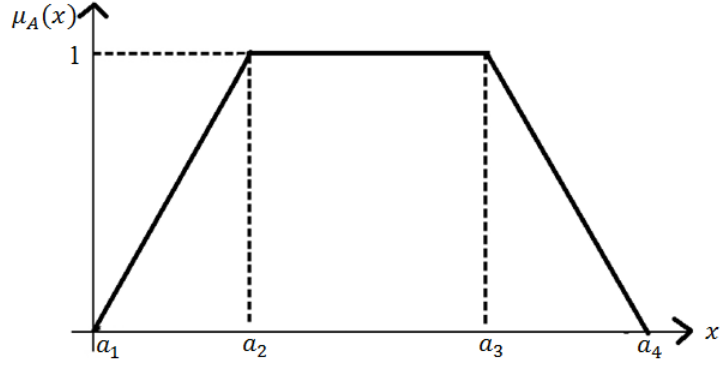
Simetrik üçgen bulanık sayı sol ve sağ uç olarak iki gerçek sayı yardımıyla yazılabilir. (a_1, a_2, a_3) şeklindeki bir simetrik üçgen bulanık sayı (a_1, a_3) şeklinde de yazılabilir ve bu durumda $a_2 = \frac{a_3 + a_1}{2}$ olarak bulunur.

3.1.2. Yamuk bulanık sayılar

$A = (a_1, a_2, a_3, a_4)$ bir yamuk bulanık sayı olmak üzere, yapısında üyelik derecesi 1'e eşit olan pek çok nokta barındırır. Yamuk bulanık sayı aşağıdaki gibi bir üyelik fonksiyonu yardımı ile ifade edilir.

$$\mu_A(x) = \begin{cases} 0, & x < a_1 \\ \frac{x-a_1}{a_2-a_1}, & a_1 \leq x \leq a_2 \\ 1, & a_2 \leq x \leq a_3 \\ \frac{a_4-x}{a_4-a_3}, & a_3 \leq x \leq a_4 \\ 0, & x > a_4 \end{cases} \quad (3.8)$$

Bir yamuk bulanık sayı $a_2 = a_3$ olması durumunda üçgen bulanık sayıya dönüşür. Yamuk bulanık sayıya ait bir üyelik fonksiyonunun gösterimi Şekil 3.5'te görülmektedir (Lee 2005).



Şekil 3.5 Yamuk Bulanık Sayı

3.1.3. LR gösterimli bulanık sayılar

Bulanık sayıların LR gösterimi, onun genelliğini sınırlamaksızın hesaplamasal etkinliğini artırmasından dolayı Dubois ve Prade (1980) tarafından önerilmiştir. Bir M bulanık sayısı için üyelik fonksiyonu $\mu_M(x)$ 'in LR gösterimi eşitlik (3.9) ile tanımlanır.

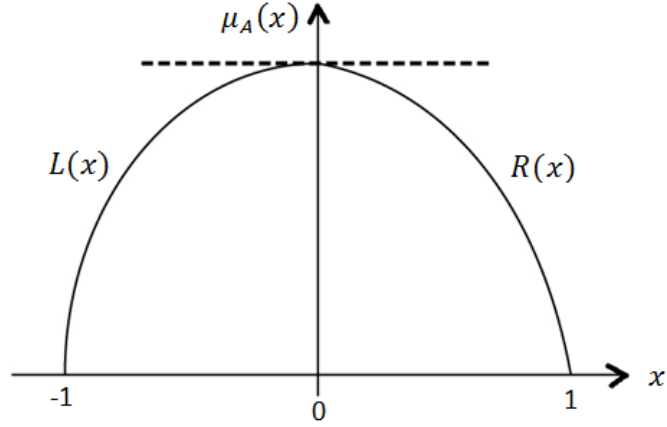
$$\mu_M(x) = \begin{cases} L[(m-x)/\alpha], & x < m, \alpha > 0 \\ 1, & x = m \\ R[(x-m)/\beta], & x > m, \beta > 0 \end{cases} \quad (3.9)$$

Referans fonksiyonları olarak anılan $L(x)$ ve $R(x)$ fonksiyonları, $\mathfrak{R}^+ \rightarrow [0,1]$ olmak üzere aşağıdaki koşulları sağlar.

- i. $L(x) = L(-x), R(x) = R(-x)$
- ii. $L(0) = 1, R(0) = 1$
- iii. L ve R $[0, \infty]$ aralığında artmayandır ($x \geq 0$ için $L(x)$ ve $R(x)$ kesin azalandır)

Denklem (3.9) ifadesinde m , M 'nin orta değeri, α ve β ise uygun olarak sol ve sağ saçılımlar (yayımlar) olarak adlandırılırlar. Saçılımlar sıfıra yaklaştığında sayı bulanıklığını kaybeder, saçılımlar arttıkça sayı daha fazla bulanıklaşır (Lin ve

Lee 1996). M , LR bulanık sayısına ait üyelik fonksiyonu referans fonksiyonlarının yapısına bağlı olarak Şekil 3.6'daki gibi ifade edilebilir.



Şekil 3.6 LR Bulanık Sayısı

Sembolik olarak LR bulanık sayısını $M = (m, \alpha, \beta)_{LR}$ şeklinde yazacağız ve gerekli olmadığı durumlarda LR indisini göstermeyeceğiz. LR bulanık sayıları üzerinde bir dizi genişletilmiş aritmetik işlemler geliştirilmiştir (Dubois ve Prade 1980). Örneğin, $(m, \alpha, \beta)_{LR}$ ve $(n, \gamma, \delta)_{LR}$ sayılarının genişletilmiş toplamı $(m + n, \alpha + \gamma, \beta + \delta)_{LR}$, $(m, \alpha, \beta)_{LR}$ ve $(n, \gamma, \delta)_{RL}$ sayılarının farkı $(m - n, \alpha + \delta, \beta + \gamma)_{LR}$ formülleri ile hesaplanır, $(m, \alpha, \beta)_{LR}$ sayısının toplamsal tersi ise $-(m, \alpha, \beta)_{LR} = (-m, \beta, \alpha)_{RL}$ olarak bulunur.

3.2. Genişletme prensibi

Zadeh (1978) tarafından ortaya konulan genişletme prensibi (extension principle) bulanık küme teorisinde en önemli araçlardan biridir. Bu prensip, kesin matematiksel kavramların bulanık küme çerçevesinde genelleştirilmesini sağlar ve noktasal görüntülemeyi bulanık kümeler için görüntülemeye genişletir.

Bu ilke, U kesin kümesindeki n 'li (x_1, x_2, \dots, x_n) 'leri, V kesin kümesindeki bir noktaya görüntüleyen her f fonksiyonunun, U 'daki n bulanık alt kümenin V 'deki bir bulanık alt kümeye görüntülemesine genelleştiren bir araç sağlar.

Böylece, bulanık olmayan elemanlar arasındaki her matematiksel ilişki, bulanık elemanlar ile ilgili olarak genişletilebilir.

$f: U \rightarrow V$ verilen bir fonksiyon ve $A = \mu_1/x_1 + \mu_2/x_2 + \dots + \mu_n/x_n$, U 'da bir bulanık küme olmak üzere, genişletme prensibi aşağıdaki gibi ifade edilir.

$$\begin{aligned} f(A) &= f(\mu_1/x_1 + \mu_2/x_2 + \dots + \mu_n/x_n) \\ &= \mu_1/f(x_1) + \mu_2/f(x_2) + \dots + \mu_n/f(x_n) \end{aligned} \quad (3.10)$$

Bu durumda V kümesinde tanımlanan $f(A)$ bulanık kümesinin üyelik fonksiyonu şöyle ifade edilir,

$$\mu_{f(A)}(x) = \max_{\substack{x_i \in U \\ f(x_i)=y}} [\mu_A(x_i)] \quad (3.11)$$

U kesin kümesi evrenlerin bir kartezyen çarpımı $U = U_1 \times U_2 \times \dots \times U_n$ olmakla, kesin f fonksiyonu U 'daki n -liyi, V 'deki bir noktaya görüntüleyen n değişkenli bir fonksiyon olsun, yani $y = f(x_1, x_2, \dots, x_n)$. Uygun olarak A_1, A_2, \dots, A_n U_1, U_2, \dots, U_n 'de, n sayıda bulanık kümeler olsun. Genişletme prensibi $f(x_1, x_2, \dots, x_n)$ fonksiyonunu, U 'daki n bulanık A_1, A_2, \dots, A_n alt kümelere şöyle genişletir. $B = f(A)$, $f(\cdot)$ 'in A_1, A_2, \dots, A_n 'de bulanık görüntüsü olmak üzere, bu bulanık küme şöyle tanımlanır:

$$\begin{aligned} B &= \{(y, \mu_B(y)) \mid y = f(x_1, x_2, \dots, x_n), (x_1, x_2, \dots, x_n) \in U\} \\ \mu_B(y) &= \sup_{\substack{(x_1, x_2, \dots, x_n) \in U \\ y=f(x_1, x_2, \dots, x_n)}} \min[\mu_{A_1}(x_1), \mu_{A_2}(x_2), \dots, \mu_{A_n}(x_n)], \end{aligned} \quad (3.12)$$

eğer $y = f(x_1, x_2, \dots, x_n)$ şeklinde hiçbir $(x_1, x_2, \dots, x_n) \in U$ elemanı yoksa $\mu_B(y) = 0$ 'dır (Dubois ve Prade 1980; Lin ve Lee 1996).

3.3. Bulanık Sıralama

Bulanık sayıların karşılaştırılması ya da sıralanması (ranking) uygulama çalışmaları için oldukça önemli bir konudur. Matematiksel modellerin

kurulmasında, bulanık küme teorisi kullanıldığında bulanık değişkenlerin ya da bulanık parametrelerin işlenmesinin sıralama / karşılaştırma problemlerini içermesi kaçınılmazdır. Ancak bulanık sayılar doğrudan sıralanamazlar. Bu nedenle direkt olarak karşılaştırılabilir değillerdir. Literatürde bulanık sayıların sıralanması konusunda çeşitli yöntemler önerilmiştir. Literatürde yer alan bulanık sıralama yöntemleri üzerinde çalışmalar Lai ve Hwang (1992), MaCahon ve Lee (1990), Chang ve Lee (1993) tarafından ileri sürülmüştür. Sistematik ve kapsamlı bir sınıflandırma Chang ve Lee (1993) tarafından sağlanmıştır (Cheng 1998):

- i. α -kesimi kullanan yöntemler: Bu yöntemlerde bulanık sayılar α -kesimlerinin basit karşılaştırılması ile sıralanırlar. Adamo (1980) tarafından ileri sürülen α -kesim yönteminde A ve B gibi iki bulanık sayı karşılaştırılmak istenildiğinde $\alpha \in [0,1]$ için α -kesimleri $A^\alpha = [a_1, a_2]$ ve $B^\alpha = [b_1, b_2]$ olmak üzere; eğer $a_2 \leq b_2$ ise $A \leq B$ olur. Bu tanımlama α değerine bağlıdır ve genellikle $\alpha > 0.5$ olması istenir (Cheng 1999).
- ii. İntegralleme yöntemleri: Bu yaklaşımla bulanık sıralama yöntemi temel olarak bir bulanık sayıyı kendi ortalama değeri ile ölçer.
- iii. Çoklu indisleri kullanan yöntemler: Bu yöntemler bulanık sayıların sıralanmasında, çoklu sıralama ya da karşılaştırma fonksiyonlarını referans olarak kullanırlar.
- iv. Sezgisel yaklaşımlar: problemin bulanık özelliklerini korumak amacıyla, sezgisel yaklaşımlar önerilmiştir. Bu yaklaşımlar daha çok araştırmacının sezgisel olarak bir sıralama yapabileceğini ileri sürmüşlerdir.

Herbir bulanık sıralama yöntemi avantajlara sahip olmalarının yanında çeşitli dezavantajlara da sahiptirler. α -kesimi kullanan yöntemler karşılaştırmayı kolaylaştırma eğilimindedirler ve bu nedenle çoğu zaman sapmalı sonuçlar üretirler. İntegralleme yöntemleri oldukça sabittirler ve sıralama sürecinin bulanık doğasını göz ardı ederler (Chang ve Lee 1994a). Bu bölümde Chang ve Lee (1994a) tarafından önerilen varlığa dayalı bulanık sıralama yöntemi incelenecektir.

Bulanık sayılarda karşılaştırma genellikle, evrensel kümede bir elemanın aidiyetlik derecesine ve elemanın konumuna dayalıdır. Karşılaştırmadaki varlık çatışması, bulanık sıralama yöntemlerinde karışıklığa ve istenmeyen sonuçlara neden olur. Chang ve Lee (1994a), bulanık kümelerin, sonrasında tam anlamıyla hatalı sonuçlara neden olabileceği için evrensel kümedeki aynı elemana dayalı olarak değil, aynı varlık seviyesine (existence level) dayalı olarak karşılaştırılabileceklerini göstermişlerdir.

Verilen bir varlık seviyesi v için, A bulanık sayısının üyelik fonksiyonu $\mu_A(x)$ 'in ters görüntüsü aşağıdaki gibi olur.

$$\mu_A^{-1}(v) = \{x: \mu_A(x) = v\} \quad (3.13)$$

Eğer A bulanık sayısının üyelik fonksiyonu konveks ve sürekli ise $\{\mu_A^{-1}(v)\}$, A 'da iki grup eleman içerir, $0 < v < 1$ için $\{\mu_A^{-1}(v = 1)\}$ modu ile bu iki grup eleman, $v \in [0,1]$ için eşitlik (3.14) ve (3.15) ile tanımlanır.

$$x'(v) = \min\{x: \mu_A(x) = v\} \quad (3.14)$$

$$x''(v) = \max\{x: \mu_A(x) = v\} \quad (3.15)$$

Denklem (3.13) ile verilen karşılaştırmanın eksikliği yalnızca bir v noktasını dikkate almasıdır, bu nedenle yeterli bilgi kullanılmaz. Mümkün tüm bilgiyi kullanmak için, tam varlık seviyesi v dikkate alınmalıdır. Tam varlık indeksi aşağıdaki gibi tanımlanır.

$$I = \int_0^1 g(\{\mu_A^{-1}(v)\}) dv \quad (3.16)$$

g , üyelik fonksiyonunun ters bir fonksiyonu olmak üzere (3.17) ile tanımlanabilir:

$$g(\{\mu_A^{-1}(v)\}) = \varpi(v)[\chi_1(v)x'(v) + \chi_2(v)x''(v)] \quad (3.17)$$

burada, $\varpi(v)$, v için bir subjektif ağırlıklandırma fonksiyonu, $\chi_1(v)$ ve $\chi_2(v)$, $\mathfrak{x}'(v)$ ve $\mathfrak{x}''(v)$ için v derecesinde subjektif ağırlıklandırma faktörleridir.

Yukarıdaki tanımlar LR bulanık sayıları için basitleştirilebilir. A bir LR bulanık sayısı olmak üzere, LR sayısının sol kısmı A_L ve sağ kısmı A_R olsun. Bu durumda,

$$\mathfrak{x}'(v) = \mu_{A_L}^{-1}(v) \text{ ve } \mathfrak{x}''(v) = \mu_{A_R}^{-1}(v) \quad (3.18)$$

olur. Yani, $\mathfrak{x}'(v)$ ve $\mathfrak{x}''(v)$, sağ ve sol referans fonksiyonlarının dolayısıyla A 'nın üyelik fonksiyonunun ters görüntüsü olurlar. A , LR bulanık sayısı için tam varlık ölçüsü (OM) eşitlik (3.19) ile tanımlanır.

$$OM(A) = \int_0^1 \varpi(v) [\chi_1(v) \mu_{A_L}^{-1}(v) + \chi_2(v) \mu_{A_R}^{-1}(v)] dv \quad (3.19)$$

Basitlik açısından (3.19) eşitliğinde farklı olmayan ağırlıklandırmaları dikkate aldığımızda, tüm $v \in [0,1]$ için $\varpi(v) = 1$, $\chi_1(v) = \chi_2(v) = 1/2$ olmak üzere, $A = (a^m, e^L, e^R)_{LR}$ şeklinde tanımlı bir üçgen bulanık sayı olduğunda (3.19) eşitliği aşağıdaki gibi daha basit şekilde ifade edilebilir.

$$\begin{aligned} OM(A) &= \frac{1}{2} \int_0^1 [(a^m - e^L(1-v)) + (a^m + e^R(1-v))] dv \\ &= (4a^m - e^L + e^R)/4 \end{aligned} \quad (3.20)$$

3.4. Bulanık Fonksiyonlar

Bulanık fonksiyon tanımı, belirsizliğin olduğu yere göre çeşitli şekillerde tanımlanabilir. Esas olarak üç tür bulanık fonksiyon tanımlamasından söz edilebilir.

- Bulanık özelliklere sahip sıradan fonksiyonlar ya da yeterli bulanık kısıtlar,

- Kendilerinde ekstra bir belirsizlik üretmeksizin argümanlarının belirsizlik taşıdığı fonksiyonlar: bulanık olmayan bir elemanın görüntüsü bulanık olmayan bir elemandır,
- Bulanık olmayan argümanların kötü bilinen (ill-known) fonksiyonları.

Üç tür bulanık fonksiyon tanımlamalarının nasıl gerçekleştirildiği aşağıda ayrıntılı olarak ele alınacaktır (Dubois ve Prade 1980).

Bulanık Olarak Sınırlanan Fonksiyonlar

X ve Y iki evrensel küme olmak üzere X 'den Y 'ye geleneksel bir f fonksiyonu: $x \in X \rightarrow f(x) \in Y$ şeklinde tanımlanır. A ve B sırasıyla X ve Y üzerinde iki bulanık küme olsun. Eğer,

$$\forall x \in X, \mu_B(f(x)) \geq \mu_A(x) \quad (3.21)$$

ise, f 'ye bulanık tanım kümesi (fuzzy domain) A ve bulanık değer kümesi (fuzzy range) B olan bir fonksiyondur denir.

Bulanık tanım kümesi B ve bulanık değer kümesi C olmak üzere Y 'den Z 'ye bir g fonksiyonunu da dikkate alalım. Bu durumda $g \circ f$ bileşkesi, bulanık tanım kümesi A ve bulanık değer kümesi C olan bir fonksiyondur. Çünkü $\mu_B(f(x)) \geq \mu_A(x)$, $\mu_C(g(y)) \geq \mu_B(y)$ ve $y = f(x)$, olmak üzere, $\mu_C(g(f(x))) \geq \mu_A(x)$ olur.

Bulanık Olmayan Bir Fonksiyonun Bulanık Genişletimi

f , X 'den Y 'ye bulanık olmayan bir fonksiyon olsun. X üzerinde bir \tilde{x} bulanık kümesinin görüntüsü genişletme ilkesi ile tanımlanır. $f(\tilde{x})$ bulanık kümesi, $f^{-1}(y)$, y 'nin ters görüntü kümesi olmak üzere aşağıdaki gibi tanımlanır (Dubois ve Prade 1980):

$$\mu_{f(\tilde{x})}(y) = \begin{cases} \sup_{x \in f^{-1}(y)} \mu_{\tilde{x}}(x), & f^{-1}(y) \neq \emptyset \\ 0, & f^{-1}(y) = \emptyset \end{cases} \quad (3.22)$$

Bir örnek üzerinde bu tanımın bir uygulamasını gösterelim. $f(x) = 3x + 1$ şeklinde bir kesin fonksiyon olsun. $\tilde{x} = \{(0,0.9), (1,0.8), (2,0.7), (3,0.6), (4,0.5)\}$ bulanık kümesi için (3.22) formülü ile tanımlanan $f(\tilde{x})$ bulanık değeri şöyle olacaktır (Lee 2005).

$$f(\tilde{x}) = \{(1,0.9), (4,0.8), (7,0.7), (10,0.6), (13,0.5)\}$$

Bulanık Olmayan Değişkenin Bulanık Fonksiyonu

Y üzerinde $x \in X$ 'in görüntüsünün bir bulanık $\tilde{f}(x)$ kümesi olması durumunu dikkate alalım. $\tilde{\mathcal{P}}(Y)$, Y 'de bulanık kümelerin bir kümesi olarak, X 'den Y 'ye bir bulanıklaştırma fonksiyonu $\tilde{f}: x \rightarrow \tilde{f}(x) \in \tilde{\mathcal{P}}(Y)$, olarak tanımlanıyor. \tilde{f} fonksiyonu aşağıdaki gibi bir bulanık R ilişkisinin kesimi olarak da tanımlanabilir (Dubois ve Prade 1980).

$$\forall (x, y) \in X \times Y, \mu_{\tilde{f}(x)}(y) = \mu_R(x, y) \quad (3.23)$$

$X = \{2,3,4\}$ ve $Y = \{2,3,4,6,8,9,12\}$ olarak tanımlanan iki kesin küme olsun. Bir bulanıklaştırma fonksiyonu \tilde{f} , X 'de yer alan elemanları güç kümesi $\tilde{\mathcal{P}}(Y)$ 'ye aşağıdaki şekilde görüntüler: $\tilde{\mathcal{P}}(Y) = \{\tilde{y}_1, \tilde{y}_2, \tilde{y}_3\}$ olmak üzere;

$$\tilde{f}(2) = \tilde{y}_1, \tilde{f}(3) = \tilde{y}_2, \tilde{f}(4) = \tilde{y}_3$$

$$\tilde{y}_1 = \{(2,0.5), (4,1), (6,0.5)\},$$

$$\tilde{y}_2 = \{(3,0.5), (6,1), (9,0.5)\},$$

$$\tilde{y}_3 = \{(4,0.5), (8,1), (12,0.5)\}.$$

Görüntüleme ayrıntılarını inceleyecek olursak \tilde{f} fonksiyonu $2 \in X$ elemanını $2 \in \tilde{y}_1$ 'e 0.5 derecesiyle, $4 \in \tilde{y}_1$ 'e 0.1 derecesiyle ve $6 \in \tilde{y}_1$ 'e 0.5 derecesiyle görüntüler. Burada esas olan \tilde{f} bulanıklaştırma fonksiyonun ifadesinden

faydalanarak, bulanık olmayan bir deęişkeni bulanık fonksiyonla tanımlamaktır (Lee 2005).

3.5. Bulanık Fonksiyonun Türevi

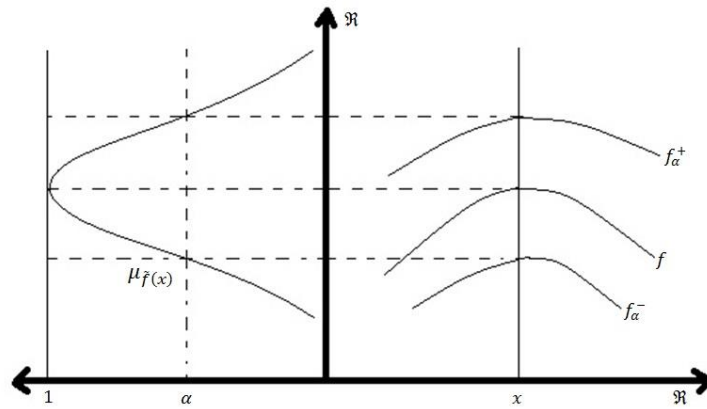
Bu bölümde ilk olarak bulanık olmayan kümede tanımlanan bulanık fonksiyonun türev tanımı gerçekleştirilecek. Sonrasında çalışmanın 4. Bölüm'ünde kullanılan LR bulanık sayıları için türev tanımlaması yapılacaktır.

3.5.1. Bulanık olmayan kümede tanımlanan bulanık fonksiyonun türevi

\tilde{f} , $[a, b] \in \mathfrak{R}$ 'de tanımlanmış $\forall x \in [a, b]$ için $\tilde{f}(x)$ bulanık sayı deęerini alan bulanıklaştırma fonksiyonu olsun. $\forall \alpha \in (0,1]$ sayısı için, x ve α parametreleri ile,

$$\mu_{\tilde{f}(x)}(y) = \alpha$$

eşitlięi $\alpha \neq 1$ için $y = f_{\alpha}^{+}(x)$ ve $y = f_{\alpha}^{-}(x)$ şeklinde yalnız iki ve $\alpha = 1$ için ise yalnız bir $y = f(x)$ şeklinde sürekli çözümlere sahiptir. Bu fonksiyonlara, \tilde{f} fonksiyonunun α -seviye eęrileri denir. \tilde{f} fonksiyonunun α -seviye eęrileri, Şekil 3.7'de gösterilmiştir.



Şekil 3.7 \tilde{f} fonksiyonunun α -seviye eęrileri

$f_{\alpha}^{+}(x)$ ve $f_{\alpha}^{-}(x)$ fonksiyonları için aşağıdaki monotonluk özelliği geçerlidir.

$$\forall \alpha_1, \alpha_2 \quad \alpha_1 \leq \alpha_2 \text{ için } f_{\alpha_1}^{+}(x) \geq f_{\alpha_2}^{+}(x) \geq f(x) \geq f_{\alpha_2}^{-}(x) \geq f_{\alpha_1}^{-}(x)$$

\tilde{f} , \mathfrak{R} 'den \mathfrak{R} 'ye bulanıklaştırma fonksiyonu olsun. Her $x \in D \subseteq \mathfrak{R}$ 'nin görüntüsünün bir bulanık sayı olduğu ve \tilde{f} 'in herbir α -seviye eğrisi f_{α} 'nın tüm $x_0 \in D$ noktalarında türevlenebildiği varsayalım. \tilde{f} 'in x_0 noktasındaki türevi $(d\tilde{f}/dx)(x_0)$ şeklinde gösterilir ve aşağıdaki üyelik fonksiyonu ile tanımlanır:

$$\mu_{(d\tilde{f}/dx)(x_0)}(P) = \sup_{f_{\alpha}: (df_{\alpha}/dx)(x_0)=P} \mu(f_{\alpha}) \quad (3.24)$$

eşitlikte tanıma göre $\mu(f_{\alpha}) = \alpha$ olur. ($\nexists \alpha \quad (df_{\alpha}/dx)(x_0) = P$ ise, $\mu_{(d\tilde{f}/dx)(x_0)}(P) = 0$)

Böylece $(d\tilde{f}/dx)(x_0)$ 'a göre P 'nin üyelik değeri, x_0 'daki eğimi P olan tüm α seviye eğrilerinden daha büyük değerlidir.

$(d\tilde{f}/dx)(x_0)$, x_0 noktasındaki seviye eğrileri demetinin paralelliği için bir tahmindir. Daha az bulanık olan $(d\tilde{f}/dx)(x_0)$ için seviye eğrileri daha fazla paralel olurlar (Dubois ve Prade 1980).

3.5.2. LR bulanık fonksiyonun türevi

$\forall x \in D$ için \tilde{f} fonksiyonunun $\tilde{f}(x)$ bulanık değerlerinin $\tilde{f}(x) = (f(x), s(x), t(x))_{LR}$ şeklinde LR bulanık sayılar olduğunu varsayalım.

$\forall \alpha \in (0,1]$, $\alpha \neq 1$ için tanımlanan $f_{\alpha}^{+}(x)$ ve $f_{\alpha}^{-}(x)$ α -seviye eğrileri, bu durumda aşağıdaki gibi bulunabilirler:

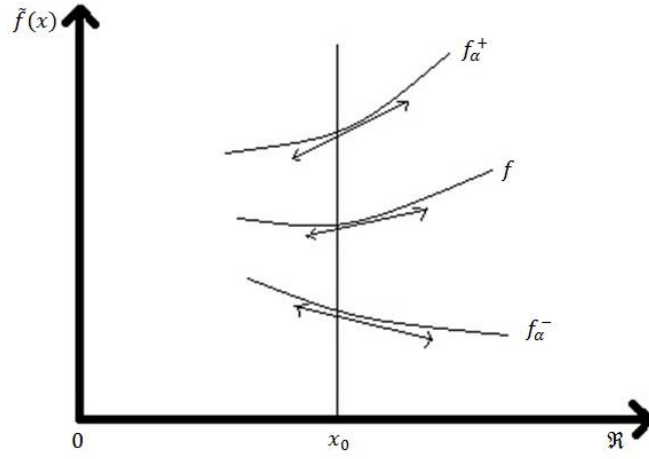
$$f_{\alpha}^{+}(x) = f(x) + R^{-1}(\alpha)t(x) \text{ ve } f_{\alpha}^{-}(x) = f(x) + L^{-1}(\alpha)s(x)$$

$\alpha = 1$ için, $f_1(x) = f(x)$ olur. $\tilde{f}(x)$ 'in kesin konveks olduğunu ve $f(x), s(x), t(x)$ fonksiyonlarının $\forall x \in D$ için türevi olduğunu varsayalım. Bu

koşullar sağlandığında, $s(x)$ ve $t(x)$ fonksiyonlarının $x_0 \in D$ noktası civarında artan fonksiyonlar olması durumunda $(d\tilde{f}/dx)(x_0)$ bulanık türevi şöyle tanımlanabilir (Dubois ve Prade 1980):

$$\left(\frac{d\tilde{f}}{dx}\right)(x_0) = \left(\frac{df}{dx}(x_0), \frac{ds}{dx}(x_0), \frac{dt}{dx}(x_0)\right)_{LR} \quad (3.25)$$

Şekil 3.8, LR bulanık fonksiyonunun türev tanımının grafiksel olarak gösterimini ifade etmektedir.



Şekil 3.8 LR bulanık fonksiyonunun türevi

4. PARAMETRİK OLMAYAN BULANIK REGRESYON MODELLERİ

Pek çok problemde iki durumu karşılaştırmak, aralarındaki ilişki araştırılmak istenildiğinde, girdi (bağımsız) ve çıktı (bağımlı) değişkenleri arasındaki durumu belirlemek oldukça zor olabilmektedir. Bu nedenle araştırmacılar daha esnek uygulamaya sahip parametrik olmayan regresyon modellerini kullanmaktadırlar. Diğer pek çok teknikte olduğu gibi konunun kesin teorisi üzerinde çalışmaların yer aldığı literatürde, bulanık yapı ve kavramlar üzerinde çok az çalışma bulunmaktadır. Araştırmalarda kullanılan veri setlerinin pek çoğunun bulanıklığı içerdiği göz önüne alındığında, tüm yöntemlerin bulanık versiyonlarının geliştirilmesinin önemliliği bir kez daha karşımıza çıkmaktadır.

Bulanık tekniklerin amacı, verilerdeki tüm özgün belirsizliği modele dahil etmektir. Bu nedenle bulanık verilere dayalı modeller, verilerdeki özgün belirsizliğin reddedildiği ya da isteğe bağlı olarak ihmal edildiği modellere göre daha fazla bilgi içerirler. Ayrıca, bulanık verilere dayalı modeller çok daha geneldir, çünkü kesin sayı kendisiyle ilişkilendirilmiş bulanıklığın olmadığı özel bir bulanık sayı olarak kabul edilebilir (D'Urso ve Gastaldi 2002).

Literatürde bulanık regresyon üzerine pek çok çalışma yer almaktadır. Konu üzerine ilk çalışma Tanaka ve ark. (1982) tarafından, girdilerin kesin sayı çıktının bulanık sayı olduğu durumu dikkate alarak, bulanık doğrusal regresyonu ileri sürmüşlerdir. Bulanık regresyon daha sonraları Tanaka (1987), Tanaka ve Watada (1988), Tanaka ve ark. (1989), Bardossy (1990), Savic ve Pedrycz (1991), Xizhao ve Minghu (1992) ve Chang ve Lee (1994b) tarafından ele alınmıştır.

Ishibuchi ve Tanaka (1992), bulanık regresyon için sinir ağlarını kullanarak basit fakat güçlü bir yöntem önermişlerdir. Bulanık regresyon analizini aralık regresyon analizine indirgeyip, çıktı aralığının alt ve üst sınırlarını iki bağımsız sinir ağının eğitilmesi ile tanımlamışlardır.

Bulanık regresyona matematiksel temelli yaklaşımlar Tanaka ve ark. (1982) ile Nasrabadi ve Nasrabadi (2004)'nin çalışmalarında yer verilmiştir.

Diamond (1988) ve Hong ve ark. (2001) yaptıkları çalışmalarda bulanık regresyon modelleri için en küçük kareler yaklaşımlarını kullanmışlardır.

Kao ve Chyu (2002), Sanchez ve Gomez (2003), Hong ve ark. (2004) bulanık regresyon çalışmaları konusunda uygulama çalışmaları yapmışlardır.

Ishibuchi ve ark. (2001), simetrik üçgensel bulanık katsayıya sahip doğrusal bulanık modellere dayalı bulanık regresyon yöntemlerinin çeşitli tiplerini açıklamışlardır.

Parametrik olmayan regresyon modellerinin karmaşık ve zor yapısı bu modellerin bulanık yapıda ifadesi konusunda zorlukları da beraberinde getirmiştir. Literatürde parametrik olmayan bulanık regresyon modelleri konusunda oldukça az çalışma yer almaktadır. Bu nedenle bu modeller üzerinde araştırma yapılması önemli bir konudur.

Cheng ve Lee (1999), fonksiyonel şeklinin önceden belirlenemediği bulanık regresyonu yani parametrik olmayan bulanık regresyonu incelemiştir. Parametrik olmayan yöntemlerden k -en yakın komşuluk ve çekirdek regresyon tahmin edicilerini bulanıklaştırmış ve analiz etmişlerdir.

Cheng ve Lee (2001), girdi ve çıktı arasındaki fonksiyonel ilişkinin önceden belirlenemediği bulanık regresyon analizinde radyal tabanlı fonksiyon (RBF) ağlarını kullanmışlardır.

Wang ve ark. (2007), bulanık yerel doğrusal regresyon modelini ele almışlar ve deneysel çalışmalar yoluyla bu yaklaşımı, çekirdek ve k -en yakın komşu düzeltme yaklaşımları ile karşılaştırmışlardır.

Kesin teoride üzerinde detaylı araştırmaların yapıldığı parametrik olmayan regresyon çalışmalarının (Green ve Silverman (2000); Hardle (1994); Loader (1999); Fan ve Gijbels (1996); Wood(2006); Hastie ve Tibshirani (1990)) bulanık teorideki uygulamaları oldukça sınırlı kalmıştır.

Parametrik olmayan yerel regresyon modelinin bulanık versiyonun geliştirilmesinin temel amacı, doğasında bulanıklığı barındıran verilerin analizinde modelin etkinliğini gözlemlemektir. Bulanıklığı içeren veri setleri bu tür bulanık modeller ile ifade edildiklerinde daha iyi ve daha etkin sonuçlar vermektedirler.

Çalışmanın bu bölümünde; ilk olarak parametrik olmayan bulanık regresyon modellerinin genel yapısı ele alınmış, daha sonra literatürde sıkça kullanılan yöntemlerden; k -en yakın komşuluk tahmini, çekirdek düzeltme ve yerel

polinomiyal düzeltme yöntemlerinin bulanık versiyonları tanıtılmıştır. Çalışmanın özgün bölümlerinden birisini oluşturan bu bölümde, band genişliği seçim yöntemleri olan çapraz geçerlilik ve genelleştirilmiş çapraz geçerlilik kriterlerinin bulanık yapıda ifadesi şapka matrisi kullanılarak gerçekleştirilmiştir. Son olarak modellerin karşılaştırılması konusunda faydalanılabilecek kriterlerden bahsedilmiştir.

4.1. Parametrik Olmayan Bulanık Regresyon Modellerinin Genel Yapısı

Bu bölümde, 2. Bölümde kesin teorisi ayrıntılı olarak incelenen parametrik olmayan regresyon modellerinin bulanık yapıda ifadesi gerçekleştirilecektir. Parametrik olmayan bulanık regresyon modelinin ifadesinde girdi değişkeni X 'in kesin sayı ve çıktı değişkeni Y 'nin bulanık sayı olduğu durum göz önüne alınarak, tek açıklayıcı değişkeni olan bir parametrik olmayan bulanık regresyon modeli aşağıdaki şekilde ifade edilir.

$$y = \mu(x) \{+\} \varepsilon \quad (4.1)$$

Denklem (4.1)'in bazen aşağıdaki gibi ifade edilmesi de uygun olur.

$$\varepsilon = y \{-\} \mu(x) \quad (4.2)$$

Burada $\mu(x) = (l(x), c(x), u(x))_{LR}$ kesin \mathcal{D} kümesinde tanımlanmış ve değerleri LR bulanık sayılar olan bulanık fonksiyon, ε bulanık hata terimidir. $\{+\}$ yada $\{-\}$ sembolleri, gözlenen ve tahminlenen bulanık hata arasındaki farkın ölçümünü ve kullanılan bulanık sıralama yöntemine bağlı olarak tanımlanan operatörleri göstermektedirler (Cheng ve Lee 1999; Wang ve ark. 2007).

Bölüm 3'de anlatıldığı gibi, alt limiti, merkezi ve üst limiti uygun olarak l_A, c_A ve u_A gerçek sayıları olan bir LR bulanık $A = (l_A, c_A, u_A)_{LR}$ sayısının üyelik fonksiyonu aşağıdaki gibi ifade edilir:

$$\mu_A(t) = \begin{cases} L\left(\frac{c_A-t}{c_A-l_A}\right), & l_A \leq t \leq c_A \\ R\left(\frac{t-c_A}{u_A-c_A}\right), & c_A \leq t \leq u_A \\ 0, & d.d. \end{cases} \quad (4.3)$$

Burada $L(\cdot)$ ve $R(\cdot)$ fonksiyonları $[0,1]$ aralığında sürekli, kesin azalan fonksiyonlardır ve $L(0) = R(0) = 1$, $L(1) = R(1) = 0$ 'dır. Bu durumda (4.1) eşitliği aşağıdaki şekle gelir.

$$y = \mu(x) \{+\} \varepsilon = (l(x), c(x), u(x))_{LR} \{+\} \varepsilon \quad (4.4)$$

Böylece (4.4) biçiminde tek değişkenli parametrik olmayan bir bulanık regresyon modelini elde etmiş oluruz. Genel yapısı ifade edilen modelin, parametrik olmayan teknikler bazında özelleştirilmiş hali izleyen bölümlerde tanımlanmıştır.

4.1.1. Bulanık k -en yakın komşuluk tahmini

Düzeltilmenin temel düşüncesi, eğer bilinmeyen fonksiyon yeteri kadar düzgün ise, x gözlemine yakın olan gözlemlerin, bilinmeyen fonksiyonunun x noktasındaki değeri hakkında bilgiyi içermeleridir. Böylece bilinmeyen fonksiyonun bir tahmininin elde edilmesinde x 'in etrafındaki verilerin yerel ortalamasını kullanmak mümkün olabilir. Eşitlik (4.1) ile verilen parametrik olmayan bulanık regresyon modelinde yer alan $\mu(x)$ için bu tahmin edici, düzeltici (smoothing) olarak adlandırılır.

k -en yakın komşuluk tahmini, değişen bir komşulukta ağırlıklandırılmış ortalamadır. Komşuluk, öklit uzaklığında x 'in k -en yakın komşusu olarak tanımlanır. Eşitlik (4.1) ile verilen parametrik olmayan bulanık regresyon modeli için k -en yakın komşuluk düzeltme tahmin edicisi aşağıdaki gibi tanımlanacaktır:

$$\begin{aligned} \hat{\mu}(x_0) &= \left(\hat{l}(x_0), \hat{c}(x_0), \hat{u}(x_0) \right)_{LR} \\ &= \left(\sum_{i=1}^n w_i(x_0) l_{y_i}, \sum_{i=1}^n w_i(x_0) c_{y_i}, \sum_{i=1}^n w_i(x_0) u_{y_i} \right)_{LR} \end{aligned} \quad (4.5)$$

Denklem (4.5)'de, $\hat{\mu}(x_0)$, $\mu(x_0)$ 'ın tahmini ve $w_i(x_0)$ ($i = 1, 2, \dots, n$) ağırlık dizisidir. Bu dizi aşağıdaki indeksler kümesi kullanılarak (4.6) formülü ile tanımlanır:

$$J(x_0) = \{i: x_i, x_0 \text{ 'a } k - \text{ en yakın gözlemlerden biri}\}$$

$$w_i(x_0) = w_i(x_0; k) = \begin{cases} 1/k & i \in J(x_0) \\ 0 & d.d. \end{cases} \quad (4.6)$$

(4.5) eşitliğinde, ağırlıkların yerine (4.6)'yı yazarak

$$\hat{\mu}(x_0) = (\hat{l}(x_0), \hat{c}(x_0), \hat{u}(x_0))_{LR} = \left(\frac{1}{k} \sum_{i \in J(x_0)} l_{y_i}, \frac{1}{k} \sum_{i \in J(x_0)} c_{y_i}, \frac{1}{k} \sum_{i \in J(x_0)} u_{y_i} \right)_{LR} \quad (4.7)$$

tahminini elde ederiz.

Düzeltilme tekniğini daha iyi anlamak için bir örnek üzerinde inceleme yapalım. $(X_i, Y_i)_{i=1, \dots, 5}$ ikilileri $(1, (5,2))$, $(7, (12,4))$, $(2, (3,1))$, $(3, (1,1))$, $(6, (4,2))$ olsun, burada Y_i 'ler üyelik fonksiyonu üçgen biçiminde olan bulanık sayılardır. $x = X_4 = 3$ için $k = 3$ en yakın komşu tahminini gerçekleştirmek istediğimizde $J(x_0 = 3) = \{1, 3, 4\}$ olur ve

$$w_1(x_0 = 3) = \frac{1}{3}, \quad w_2(3) = 0, \quad w_3(3) = \frac{1}{3}, \quad w_4(3) = \frac{1}{3}, \quad w_5(3) = 0$$

ağırlıkları hesaplanır. Bu ağırlıklar kullanılarak tahmin aşağıdaki gibi elde edilir (Cheng 1998):

$$\hat{Y}_4 = \frac{1}{3} \cdot (5,2) + 0 \cdot (12,4) + \frac{1}{3} \cdot (3,1) + \frac{1}{3} \cdot (1,1) + 0 \cdot (4,2) = \left(3, \frac{4}{3}\right)$$

4.1.2. Bulanık çekirdek regresyon tahmini

Parametrik olmayan yöntemlerde ağırlık dizilerinin gösteriminde, belirli bir x noktasına bağlı olarak, diğer noktaların konumlarının ağırlık yapısını ve

büyükliğini ayarlayan ölçek parametresini içeren bir yoğunluk fonksiyonu kullanılır. Bu yoğunluk fonksiyonu çekirdek fonksiyonu olarak adlandırılır. Çekirdek tahmininin yapısı k -en yakın komşu tahmininin yapısından farklıdır. k -en yakın komşu tahmini değişen komşulukta ağırlıklandırılmış ortalama olarak tanımlanır ve komşulukta yer alan noktalara eşit ağırlıklar verir. Çekirdek tahmini ise band genişliği h ve çekirdek fonksiyonu K ile x noktası etrafındaki sabit komşulukta yanıt değişkeninin ağırlıklandırılmış ortalamasıdır.

Çekirdek tahmininde, eşitlik (4.5)'te yer alan bilinmeyen bulanık fonksiyonun yerel tahmininde kullanılan ağırlık dizisi aşağıdaki gibi ifade edilir:

$$w_i(x_0) = w_i(x_0; h) = \frac{K_h(|x_i - x_0|)}{\sum_{j=1}^n K_h(|x_j - x_0|)}, \quad i = 1, 2, \dots, n \quad (4.8)$$

Çekirdek tahmini (4.8) ağırlık dizisi kullanılarak (4.9) ile tanımlanır:

$$\begin{aligned} \hat{\mu}(x_0) &= \left(\hat{l}(x_0), \hat{c}(x_0), \hat{u}(x_0) \right)_{LR} \\ &= \left(\frac{\sum_{i=1}^n K_h(|x_i - x_0|) l_{y_i}}{\sum_{j=1}^n K_h(|x_j - x_0|)}, \frac{\sum_{i=1}^n K_h(|x_i - x_0|) c_{y_i}}{\sum_{j=1}^n K_h(|x_j - x_0|)}, \frac{\sum_{i=1}^n K_h(|x_i - x_0|) u_{y_i}}{\sum_{j=1}^n K_h(|x_j - x_0|)} \right)_{LR} \end{aligned} \quad (4.9)$$

k -en yakın komşu tahmini, çekirdek tahminine göre özellikle sınır noktalarında güçlü bir sınır etkisine (boundary effect) sahiptir. Bu nedenle de daha dalgalı tahminler üretir. k -en yakın komşu tahmini sabit sayıda gözlem kullanır, bu nedenle sınır yakınındaki ortalama prosedür daha fazla nokta içerir. Ancak bu durum tahmin yöntemlerinden birinin diğerinden daha iyi olduğunu göstermek için yeterli değildir (Cheng ve Lee 1999; Wang ve ark. 2007).

4.1.3. Bulanık yerel polinomial regresyon tahmini

Tek değişkenli parametrik olmayan bulanık regresyon modelini ele alalım:

$$y = \mu(x) \{+\} \varepsilon \quad (4.10)$$

Burada $\mu(x) = (c(x), \alpha(x), \beta(x))$ kesin \mathcal{D} kümesinde tanımlanmış LR bulanık değerler alan bulanık fonksiyon, ε bulanık hata terimi, $\{+\}$ - kullanılan bulanık sıralama yöntemine bağlı olarak tanımlanan operatör'dür (Cheng ve Lee 1999; Wang ve ark. 2007).

(x_i, y_i) 'lerin ($i = 1, 2, \dots, n$), kesin girdilere ve LR bulanık çıktılara sahip (4.10) modeli için gözlem verileri oldukları varsayımı altında, her bir y çıktısı, c_y merkez, α_y ve β_y sol ve sağ yayılım (saçılım) bileşenleri ile (c_y, α_y, β_y) şeklinde bir LR bulanık sayıdır. Bu sayı $l_y = c_y - \alpha_y$ ve $u_y = c_y + \beta_y$ alınarak, sol uç, merkez ve sağ uç noktaları yardımıyla $(l_y, c_y, u_y)_{LR}$ şeklinde yazılabilir.

$l(x), c(x), u(x)$ fonksiyonları sırasıyla LR - fonksiyonun sol ucu, merkezi ve sağ ucu olarak, \mathcal{D} tanım kümesinde $(p + 1)$. mertebeye kadar sürekli türeve sahip olduklarında, bu fonksiyonlar verilen bir $x_0 \in \mathcal{D}$ noktasının komşuluğunda yaklaşık olarak p . mertebeden Taylor polinomu şeklinde ele alınarak aşağıdaki gibi yazılırlar.

$$\begin{aligned} l(x) &\approx l(x_0) + l'(x_0)(x - x_0) + \dots + \frac{l^{(p)}(x_0)}{p!}(x - x_0)^p \\ c(x) &\approx c(x_0) + c'(x_0)(x - x_0) + \dots + \frac{c^{(p)}(x_0)}{p!}(x - x_0)^p \\ u(x) &\approx u(x_0) + u'(x_0)(x - x_0) + \dots + \frac{u^{(p)}(x_0)}{p!}(x - x_0)^p \end{aligned} \quad (4.11)$$

Böylece $x_0 \in \mathcal{D}$ komşuluğunda $(l(x), c(x), u(x))$ fonksiyonunun tahmini için $3 \times (p + 1)$ boyutundaki $l(x_0), l'(x_0), \dots, l^{(p)}(x_0); c(x_0), c'(x_0), \dots, c^{(p)}(x_0); u(x_0), u'(x_0), \dots, u^{(p)}(x_0)$ parametrelerinin tahmini yeterli olur. Bu amaçla ağırlıklandırılmış en küçük kareler yöntemi kullanılabilir.

Burada bulanık kümeler arasındaki uzaklığın ölçümünde Diamond uzaklığı dikkate alınmıştır (Diamond 1988). $A = (l_A, c_A, u_A)_{LR}$ ve $B = (l_B, c_B, u_B)_{LR}$ gibi iki LR bulanık sayısı olmak üzere, bu sayılar arasındaki uzaklık için Diamond uzaklığı kullanılarak (4.12) ile tanımlanır.

$$d^2(A, B) = (l_A - l_B)^2 + (c_A - c_B)^2 + (u_A - u_B)^2 \quad (4.12)$$

$d^2(A, B) = 0$ olması durumunda, A ve B 'nin üyelik fonksiyonları eşittir. Diamond uzaklığı yardımıyla aşağıdaki yerel ağırlıklandırılmış hata kareler toplamının,

$$\begin{aligned}
& \sum_{i=1}^n d^2(\mu(x_i), y_i) K_h(|x_i - x_0|) = \\
& \sum_{i=1}^n d^2 \left((l(x_i), c(x_i), u(x_i)), (l_{y_i}, c_{y_i}, u_{y_i}) \right) K_h(|x_i - x_0|) = \\
& \sum_{i=1}^n \left(l_{y_i} - l(x_0) - l'(x_0)(x - x_0) - \dots - \frac{l^{(p)}(x_0)}{p!} (x - x_0)^p \right)^2 K_h(|x_i - x_0|) + \\
& \sum_{i=1}^n \left(c_{y_i} - c(x_0) - c'(x_0)(x - x_0) - \dots - \frac{c^{(p)}(x_0)}{p!} (x - x_0)^p \right)^2 K_h(|x_i - x_0|) + \\
& \sum_{i=1}^n \left(u_{y_i} - u(x_0) - u'(x_0)(x - x_0) - \dots - \frac{u^{(p)}(x_0)}{p!} (x - x_0)^p \right)^2 K_h(|x_i - x_0|)
\end{aligned} \tag{4.13}$$

minimizasyonu problemi şeklinde ifade edilir. Burada verilen bir çekirdek fonksiyonu $K(\cdot)$ ve düzeltme parametresi h için, $K_h(|x_i - x_0|) = \frac{1}{h} K\left(\frac{|x_i - x_0|}{h}\right)$, $i = 1, 2, \dots, n$, ağırlık fonksiyonunun rolü verilen bir x_0 noktasına yakın olan gözlemlere daha fazla, uzakta olan noktalara daha az ağırlık atamaktır. Denklem (4.11)'de yer alan $l(x_0), l'(x_0), \dots, l^{(p)}(x_0)$; $c(x_0), c'(x_0), \dots, c^{(p)}(x_0)$; $u(x_0), u'(x_0), \dots, u^{(p)}(x_0)$ polinom katsayıları, ağırlıklandırılmış en küçük kareler problemi (4.13)'ün çözümü ile tahmin edilirler. Denklem (4.13)'teki amaç fonksiyonu aslında üç ayrı bölümün toplamıdır ve her bir bölüm birbirinden bağımsız olan farklı gruptaki parametreleri içerir. Bu nedenle eşitlikteki minimizasyon problemi, birbirinden bağımsız bu üç bölümün ağırlıklandırılmış en küçük kareler problemi olarak aşağıdaki şekilde ifade edilebilir:

$$\sum_{i=1}^n \left(l_{y_i} - l(x_0) - l'(x_0)(x - x_0) - \dots - \frac{l^{(p)}(x_0)}{p!} (x - x_0)^p \right)^2 K_h(|x_i - x_0|) \rightarrow \min$$

$$\begin{aligned}
& \sum_{i=1}^n \left(c_{y_i} - c(x_0) - c'(x_0)(x - x_0) - \dots \right. \\
& \quad \left. - \frac{c^{(p)}(x_0)}{p!} (x - x_0)^p \right)^2 K_h(|x_i - x_0|) \rightarrow \mathbf{min} \\
& \sum_{i=1}^n \left(u_{y_i} - u(x_0) - u'(x_0)(x - x_0) - \dots \right. \\
& \quad \left. - \frac{u^{(p)}(x_0)}{p!} (x - x_0)^p \right)^2 K_h(|x_i - x_0|) \rightarrow \mathbf{min}
\end{aligned} \tag{4.14}$$

Problemi farklı bir yapıda ifade etmek amacıyla aşağıdaki matris ve vektörleri tanımlayalım:

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

$$l_y = \begin{pmatrix} l_{y_1} \\ \vdots \\ l_{y_n} \end{pmatrix}, c_y = \begin{pmatrix} c_{y_1} \\ \vdots \\ c_{y_n} \end{pmatrix}, u_y = \begin{pmatrix} u_{y_1} \\ \vdots \\ u_{y_n} \end{pmatrix}.$$

$K_h(z) = \frac{1}{h} K\left(\frac{z}{h}\right)$ olmak üzere, $n \times n$ boyutundaki

$\mathbf{W}(x_0; h) = \text{Diag}(K_h(|x_1 - x_0|), \dots, K_h(|x_n - x_0|))$ köşegen ağırlık matrisini kullanarak, denklem (4.14)'deki minimizasyon problemlerini matris notasyonu yardımıyla şöyle yazabiliriz:

$$l_y^T \mathbf{X}^T(x_0) \mathbf{W}(x_0; h) \mathbf{X}(x_0) l_y \rightarrow \mathbf{min}$$

$$c_y^T \mathbf{X}^T(x_0) \mathbf{W}(x_0; h) \mathbf{X}(x_0) c_y \rightarrow \mathbf{min}$$

$$u_y^T \mathbf{X}^T(x_0) \mathbf{W}(x_0; h) \mathbf{X}(x_0) u_y \rightarrow \mathbf{min}$$

Bu problemlerin çözümü olarak aşağıdaki formülleri elde ederiz:

$$\begin{aligned}
(\hat{l}(x_0), \hat{l}'(x_0), \dots, \hat{l}^{(p)}(x_0))^T &= \left(\mathbf{X}^T(x_0) \mathbf{W}(x_0; h) \mathbf{X}(x_0) \right)^{-1} \mathbf{X}^T(x_0) \mathbf{W}(x_0; h) l_y \\
(\hat{c}(x_0), \hat{c}'(x_0), \dots, \hat{c}^{(p)}(x_0))^T &= \left(\mathbf{X}^T(x_0) \mathbf{W}(x_0; h) \mathbf{X}(x_0) \right)^{-1} \mathbf{X}^T(x_0) \mathbf{W}(x_0; h) c_y \\
(\hat{u}(x_0), \hat{u}'(x_0), \dots, \hat{u}^{(p)}(x_0))^T &= \left(\mathbf{X}^T(x_0) \mathbf{W}(x_0; h) \mathbf{X}(x_0) \right)^{-1} \mathbf{X}^T(x_0) \mathbf{W}(x_0; h) u_y \quad (4.15)
\end{aligned}$$

(4.15) formülleri $Z(x_0; h) = \left(\mathbf{X}^T(x_0) \mathbf{W}(x_0; h) \mathbf{X}(x_0) \right)^{-1} \mathbf{X}^T(x_0) \mathbf{W}(x_0; h)$ alınarak daha basit gösterimle yazılabilir:

$$\begin{aligned}
(\hat{l}(x_0), \hat{l}'(x_0), \dots, \hat{l}^{(p)}(x_0))^T &= Z(x_0; h) l_y \\
(\hat{c}(x_0), \hat{c}'(x_0), \dots, \hat{c}^{(p)}(x_0))^T &= Z(x_0; h) c_y \\
(\hat{u}(x_0), \hat{u}'(x_0), \dots, \hat{u}^{(p)}(x_0))^T &= Z(x_0; h) u_y \quad (4.16)
\end{aligned}$$

Bölüm 2.'de yer alan çekirdek fonksiyonları benzer şekilde bulanık yapıda da kullanılmaktadır. Gauss çekirdek fonksiyonu aşağıdaki gibi ifade edilir:

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

Epanechnikov çekirdek fonksiyonu ise,

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

olarak ifade edilir.

4.2. Parametrik Olmayan Bulanık Regresyon Modellerinde Düzeltme Parametresi Seçim Kriterlerinin Geliştirilmesi

Çalışmanın 2. Bölüm'ünde değinildiği gibi parametrik olmayan regresyon modellerinde düzeltme parametresinin seçiminde literatürde sıkça kullanılan yöntemler; çapraz geçerlilik kriteri, genelleştirilmiş çapraz geçerlilik kriteri ve

geliştirilmiş Akaike bilgi kriteridir. Parametrik olmayan bulanık regresyon modellerinde ise bu kriterlerden çapraz geçerlilik kriteri üzerine Wang ve ark. (2007), Cheng ve Lee (1999) tanımlamalar yapmışlardır. Çalışmanın özgün kısımlarından birini oluşturan bu bölümde, kesin teoride yer alan CV ve GCV kriterlerinin tanımlarından yola çıkılarak, girdilerin kesin değerli ve çıktının bulanık LR sayısı olarak ifade edildiği durum göz önüne alınarak kriterlerin bulanık yapıda ifadesi gerçekleştirilmiştir. Ayrıca bu yeni formüllerin geliştirilmesinde yerel polinomiyal düzeltme tekniği dikkate alınmıştır.

4.2.1. Bulanık çapraz geçerlilik kriterinin geliştirilmesi

Yerel polinomiyal regresyon tahmin edicisi en küçük kareler probleminin çözümünden elde edildiği için $\hat{\mu}(x)$ doğrusal bir tahmin edicidir. Şöyle ki; her bir x değeri için $\alpha(x) = (\alpha_1(x), \alpha_2(x), \dots, \alpha_n(x))^T$ şeklinde bir ağırlık vektörü mevcuttur ve $\hat{\mu}(x) = \sum_{i=1}^n \alpha_i(x)y_i$ olarak yazılır (Loader, 1999). Burada $\alpha^T(x)$ vektörü (4.17)'deki gibi hesaplanır:

$$\begin{aligned}\alpha^T(x) &= e_1^T (X^T(x)W(x, h)X(x))^{-1} X^T(x)W(x, h) \\ &= e_1^T Z(x, h), \quad e_1 = (1, 0, \dots, 0)^T\end{aligned}\quad (4.17)$$

Verileri uyumlanan değerlere haritalayan şapka (hat) matrisi H , satırları $\alpha^T(x_i)$ vektörleri olan $n \times n$ boyutunda bir matris olmak üzere aşağıdaki gibi ifade edilir.

$$H = \begin{pmatrix} \alpha_1(x_1) & \alpha_2(x_1) & \dots & \alpha_n(x_1) \\ \vdots & \vdots & & \vdots \\ \alpha_1(x_n) & \alpha_2(x_n) & \dots & \alpha_n(x_n) \end{pmatrix}\quad (4.18)$$

Şapka matrisi yardımıyla bilinmeyen fonksiyon μ 'nün tahmini,

$$\hat{\mu} = (\hat{\mu}(x_1), \dots, \hat{\mu}(x_n))^T = Hy\quad (4.19)$$

şeklinde bulunur. H şapka matrisinin elemanları kesin sayılardır ve x_i 'nin $i = 1, 2, \dots, n$ değerleri ile band genişliğine bağlıdır. H matrisinin i . köşegen elemanları $h_{ii} = \alpha_i(x_i)$ etki (influence) noktalarıdır. Düzeltmenin miktarını ölçmede kullanılan düzeltme parametresi serbestlik derecesi ile ilişkili olmasından dolayı, düzeltme, serbestlik derecesinin uygun değeri ayarlanarak gerçekleştirilebilir. Yerel polinomiyal regresyon için iki farklı serbestlik derecesi aşağıdaki gibi tanımlanır (Loader, 1999).

$$v_1 = \sum_{i=1}^n h_{ii} = \text{tr}H \quad (4.20)$$

ve

$$v_2 = \sum_{i=1}^n \|\alpha(x_i)\|^2 = \text{tr}(H^T H) \quad (4.21)$$

Yerel polinomiyal regresyonda sabit band genişliği h 'ın seçiminde kullanılacak olan çapraz geçerlilik kriterinin bulanık versiyonunu şapka matrisi yardımıyla tanımlayalım. Kriterin kesin durumdaki ifadesi (4.22) eşitliğindeki gibidir.

$$CV(h) = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{\mu}_{(i)}(x_i, h))^2 = \frac{1}{n} \sum_{i=1}^n \frac{(y_i - \hat{\mu}(x_i, h))^2}{(1 - h_{ii})^2} \quad (4.22)$$

Denklem (4.22)'de $\hat{\mu}_{(i)}(x_i, h)$, düzeltme parametresi h seçildiğinde, veri kümesinden i . gözlem değeri (x_i, y_i) çıkarılmakla x_i noktasında tahmin edilen kesin regresyon fonksiyonudur. Diamond uzaklığı kullanılarak bu formülün bulanık yapıdaki uygun ifadesi tanımlanabilir. $\hat{\mu}_{(i)}(x_i, h)$, düzeltme sürecinde i . gözlem değeri (x_i, y_i) çıkartılarak, h düzeltme parametresi ile x_i girdisi için parametrik olmayan bulanık regresyon modelinde tahminlenen bulanık fonksiyon olsun:

$$\hat{\mu}_{(i)}(x_i, h) = \left(\hat{l}_{(i)}(x_i, h), \hat{c}_{(i)}(x_i, h), \hat{u}_{(i)}(x_i, h) \right) \quad (4.23)$$

Çapraz geçerlilik fonksiyonunun tanımına göre,

$$\begin{aligned}
CV(h) &= \frac{1}{n} \sum_{i=1}^n d^2(y_i, \hat{\mu}_{(i)}(x_i, h)) \\
&= \frac{1}{n} \sum_{i=1}^n d^2 \left[\left(\hat{l}_{(i)}(x_i, h), \hat{c}_{(i)}(x_i, h), \hat{u}_{(i)}(x_i, h) \right), (l_{y_i}, c_{y_i}, u_{y_i}) \right] \\
&= \frac{1}{n} \sum_{i=1}^n \left[\left(l_{y_i} - \hat{l}_{(i)}(x_i, h) \right)^2 + \left(c_{y_i} - \hat{c}_{(i)}(x_i, h) \right)^2 + \left(u_{y_i} - \hat{u}_{(i)}(x_i, h) \right)^2 \right] \quad (4.24)
\end{aligned}$$

Denklem (4.24), H şapka matrisinin kesin değerler almasından dolayı, l, c, u değişkenlerine uygun ifadelerin üçünün bir toplamı şeklinde yazılabilir ve (4.22) aşağıdaki gibi ifade edilir:

$$\begin{aligned}
CV(h) &= \frac{1}{n} \sum_{i=1}^n \left[\left(l_{y_i} - \hat{l}_{(i)}(x_i, h) \right)^2 + \left(c_{y_i} - \hat{c}_{(i)}(x_i, h) \right)^2 + \left(u_{y_i} - \hat{u}_{(i)}(x_i, h) \right)^2 \right] \\
&= \frac{1}{n} \sum_{i=1}^n \frac{\left(l_{y_i} - \hat{l}_{(i)}(x_i, h) \right)^2}{(1-h_{ii})^2} + \frac{1}{n} \sum_{i=1}^n \frac{\left(c_{y_i} - \hat{c}_{(i)}(x_i, h) \right)^2}{(1-h_{ii})^2} + \frac{1}{n} \sum_{i=1}^n \frac{\left(u_{y_i} - \hat{u}_{(i)}(x_i, h) \right)^2}{(1-h_{ii})^2} \\
&= \frac{1}{n} \sum_{i=1}^n \frac{\left(l_{y_i} - \hat{l}_{(i)}(x_i, h) \right)^2 + \left(c_{y_i} - \hat{c}_{(i)}(x_i, h) \right)^2 + \left(u_{y_i} - \hat{u}_{(i)}(x_i, h) \right)^2}{(1-h_{ii})^2} \quad (4.25)
\end{aligned}$$

Sonuçta (4.24) eşitliği aşağıdaki gibi yazılabilir.

$$CV(h) = CV(l, h) + CV(c, h) + CV(u, h) \quad (4.26)$$

Burada,

$$\begin{aligned}
CV(l, h) &= \frac{1}{n} \sum_{i=1}^n \frac{\left(l_{y_i} - \hat{l}_{(i)}(x_i, h) \right)^2}{(1-h_{ii})^2}, \quad CV(c, h) = \frac{1}{n} \sum_{i=1}^n \frac{\left(c_{y_i} - \hat{c}_{(i)}(x_i, h) \right)^2}{(1-h_{ii})^2}, \\
CV(u, h) &= \frac{1}{n} \sum_{i=1}^n \frac{\left(u_{y_i} - \hat{u}_{(i)}(x_i, h) \right)^2}{(1-h_{ii})^2}
\end{aligned}$$

Sabit band genişliği h 'ın optimum değeri yukarıdaki eşitlikler yardımıyla belirlenebilir. Optimum $h = h_0$ düzeltme parametresinin seçimi için $CV(h_0) = \min_h CV(h)$ minimizasyon probleminin çözülmesi gerekir.

4.2.2. Bulanık genelleştirilmiş çapraz geçerlilik kriterinin geliştirilmesi

Şapka matrisi yardımıyla çapraz geçerlilik fonksiyonunun bir düzeltilmiş şekli olarak genelleştirilmiş çapraz geçerlilik fonksiyonu geliştirilmiştir. Denklem (4.22) formülünde şapka matrisinin i . köşegen elemanı yerine köşegenlerin ortalaması yazılarak aşağıdaki genelleştirilmiş çapraz geçerlilik formülü elde edilir:

$$GCV(h) = \frac{1}{n} \sum_{i=1}^n \frac{(y_i - \hat{\mu}(x_i, h))^2}{\left(1 - \frac{1}{n} tr(H)\right)^2} \quad (4.27)$$

Denklem (4.27) formülünün bulanık versiyonunu elde etmek için Diamond uzaklığı kullanalım:

$$\begin{aligned} GCV(h) &= \frac{1}{n} \sum_{i=1}^n \frac{d^2(y_i - \hat{\mu}(x_i, h))}{\left(1 - \frac{1}{n} trH\right)^2} \\ &= \frac{1}{n} \sum_{i=1}^n \frac{\left(l_{y_i} - \hat{l}(x_i, h)\right)^2 + \left(c_{y_i} - \hat{c}(x_i, h)\right)^2 + \left(u_{y_i} - \hat{u}(x_i, h)\right)^2}{\left(1 - \frac{1}{n} trH\right)^2} \end{aligned} \quad (4.28)$$

Denklem (4.28) ifadesi H şapka matrisinin elemanlarının kesin sayılar olmasından dolayı, l, c, u değişkenlerine uygun ifadelerin üçünün bir toplamı şeklinde yazılabilir:

$$GCV(h) = GCV(l, h) + GCV(c, h) + GCV(u, h) \quad (4.29)$$

Burada,

$$\begin{aligned} GCV(l, h) &= \frac{1}{n} \sum_{i=1}^n \frac{\left(l_{y_i} - \hat{l}(x_i, h)\right)^2}{\left(1 - \frac{1}{n} trH\right)^2}, & GCV(c, h) &= \frac{1}{n} \sum_{i=1}^n \frac{\left(c_{y_i} - \hat{c}(x_i, h)\right)^2}{\left(1 - \frac{1}{n} trH\right)^2}, \\ GCV(u, h) &= \frac{1}{n} \sum_{i=1}^n \frac{\left(u_{y_i} - \hat{u}(x_i, h)\right)^2}{\left(1 - \frac{1}{n} trH\right)^2} \end{aligned}$$

Düzeltilme parametresinin optimum $h = h_0$ değerinin seçimi için $GCV(h_*) = \min_h GCV(h)$ minimizasyon problemini çözmek gerekir.

4.3. Parametrik Olmayan Bulanık Regresyon Modellerinde Performans Kriteri

Çalışmanın 2. Bölüm'ünde verilen kesin teorideki modellerin performanslarının karşılaştırılmasında kullanılan kriterlerin tamamının bulanık yapıda ifadesi henüz gerçekleşmemiştir. Bulanık yapıda ifade edilen parametrik olmayan regresyon modellerinin karşılaştırılmasında kullanılacak kriterlerden biri Cheng ve Lee (1999) tarafından tanımlanmıştır. Bu tanımlama gerçekleştirilirken uzaklık ya da iki bulanık sayı arasındaki fark dikkate alınmıştır. Literatürde yer alan çeşitli bulanık sıralama yaklaşımlarından Chang ve Lee (1994a) tarafından önerilen yaklaşımı, tanımlamada kullanmışlardır. Buna göre yan (bias) ifadesi aşağıdaki gibi yapılmıştır.

$$B = \frac{1}{N} \sum_{i=1}^N [y_i - \hat{y}_i]^2 \quad (4.30)$$

Eşitlikte y_i bağımlı değişkene ait gözlem değerlerini ve \hat{y}_i ise bu gözlemlere ait tahmin değerlerini göstermektedir. Kriterin minimum olması istenilen durumu ifade eder.

Diğer bir kriter, iki bulanık sayı arasındaki uzaklık ölçüsü olarak Diamond uzaklığı kullanılarak tanımlanmıştır. Kesin teorideki yapıya benzer olarak bulanık teorideki performans kriterleri de gözlem değerleri ile uyum değerleri arasındaki farka yani hataya dayalıdır. Bölüm 2'de kesin teori için tanımlanan ortalama karesel hata (ASE) kriterinin bulanık yapıda tanımlaması aşağıdaki gibi olur.

$$\begin{aligned} ASE(h) &= \frac{1}{n} \sum_{i=1}^n d^2(\mu(x_i), \hat{\mu}(x_i)) \\ &= \frac{1}{n} \sum_{i=1}^n [(l(x_i) - \hat{l}(x_i))^2 + (c(x_i) - \hat{c}(x_i))^2 + (u(x_i) - \hat{u}(x_i))^2] \end{aligned} \quad (4.31)$$

Bu performans kriterinin minimum deęeri modeller için istenilen durumu gösterir. *ASE* kriterinin daha küçük olduęu modeli, daha iyi bir model olarak tanımlamak mümkündür.

5. SİMÜLASYON ÇALIŞMASI

Çalışmanın 4. Bölüm'ünde geliştirilen *CV* ve *GCV* band genişliği seçim kriterlerinin bulanık yapıda uygulanması amacıyla bu bölümde çeşitli simülasyon deneylerine yer verilmiştir. Bu amaçla farklı veri türetim fonksiyonlarından faydalanılmıştır. Gerçekleştirilen simülasyon ve gerçek veriler üzerindeki uygulamaların temel amacı; parametrik olmayan bulanık regresyon modelleri konusunda hem teorik hemde uygulama bazında çok az sayıda çalışmanın bulunmasından dolayı bu tür modellerin davranışlarını araştırmaktır.

Çalışmanın temel amacı doğrultusunda, parametrik olmayan regresyon modellerinin bulanık versiyonları geliştirilmiş, band genişliği seçim kriterleri ve modellerin karşılaştırılmasında kullanılacak kriterlerin bulanık versiyonlarının geliştirilerek farklı türdeki veri setleri üzerinde uygulamalar yapılmıştır. 4. Bölüm'de teorisi üretilen kriterler, bu bölümde simülasyon çalışmalarında ve 6. Bölüm'de gerçek veri setleri üzerinde kullanılarak sonuçlar genelleştirilmiş ve mevcut parametrik olmayan bulanık regresyon modelleri teorisine katkı sağlanmıştır.

Bu bölümde dört farklı fonksiyon ile türetilen bulanık veriler kullanılarak simülasyon deneyleri gerçekleştirilmiştir. Herbir uygulama çalışmasında, geliştirilmiş olan *CV* ve *GCV* formülasyonları için R programında *lcpol* paketi ile program kodları yazılmış ve uygulamalar gerçekleştirilmiştir.

5.1. Veri Seti-1 ve Veri Seti-2 İçin Parametrik Olmayan Bulanık Regresyon Modellerinin Karşılaştırılması

Bu çalışmada, çekirdek düzeltme, bulanık yerel doğrusal ve bulanık yerel kübik modeller için uygun düzeltme parametresinin seçiminde çapraz geçerlilik (*CV*) ve genelleştirilmiş çapraz geçerlilik (*GCV*) kriterleri kullanarak, bu modellerin performanslarını karşılaştırmak amaçlanmıştır. Cheng ve Lee (1999) çalışmasındaki fonksiyonlar kullanılarak, boyutları 100 olan veri setleri türetilmiştir ve simülasyon denemelerinde 200 tekrarlıma gerçekleştirilmiştir. Bulanık yanıt çıktıları, merkez ve sağ-sol saçılımları Veri Seti-1 ve

Veri Seti-2'deki gibi türetilen, simetrik üçgen bulanık sayılardır. Simülasyon çalışmasında gürültü, merkez için $[-0.5, 0.5]$, simetrik saçılımlar için ise $[-0.25, 0.25]$ aralığında düzgün dağılımdan elde edilmektedir.

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

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

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

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

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

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

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

Veri Seti-2 $[0,10]$ aralığında veri üretiminde kullanılan ikinci fonksiyon aşağıdaki gibi tanımlanmıştır.

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

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

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

$$\sigma_i = \frac{1}{3}g_2(x_i) + \text{rand}[-0.25, 0.25] \quad i = 1,2, \dots, 100$$

olarak hesaplanırlar. Bulanık çıktılar aşağıdaki gibi ifade edilir.

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

Simülasyon çalışmasında her iki fonksiyon yardımıyla türetilen veri setleri için, teorisi geliştirilen *CV* ve *GCV* kriterleri yardımıyla band genişliği (düzeltme parametresi) seçimi çekirdek düzeltme, bulanık yerel doğrusal (*deg* = 1) ve bulanık yerel kübik (*deg* = 3) modeller için yapılmış daha sonrasında modellerin performansları karşılaştırılmıştır. Veri Seti-1 için *CV* ve *GCV* kriterleri yardımıyla band genişliğinin seçiminde belirli aralıklarda yaklaşım gerçekleştirilmiş ve Çizelge 5.1'de Gauss çekirdeği kullanılarak ve Çizelge 5.2'de Epanechnikov çekirdeği kullanılarak her bir *h* değerlerine karşılık gelen *CV* ve *GCV* skorları hesaplanmıştır.

Çizelge 5.1 Veri Seti-1 için Gauss çekirdeği kullanıldığında çekirdek düzeltme, $p = 1$ ve $p = 3$ için *CV* ve *GCV* kriterleri ile band genişliği seçimi

Veri Seti-1								
Çekirdek Düzeltme Gauss			Deg = 1 Gauss			Deg = 3 Gauss		
<i>h</i>	<i>CV</i>	<i>GCV</i>	<i>h</i>	<i>CV</i>	<i>GCV</i>	<i>h</i>	<i>CV</i>	<i>GCV</i>
0.10	0.4034	0.3986	0.2	0.3458	0.3388	0.5	0.3374	0.3298
0.15	0.3757	0.3694	0.3	0.3312	0.3296	0.7	0.3242	0.3207
0.16	0.3755	0.3676	0.4**	0.3229	0.3236	0.9	0.3203	0.3157
0.17	0.3738	0.3673	0.5*	0.3217	0.3257	1.1	0.3133	0.3127
0.18	0.3718	0.3662	0.6	0.3268	0.3302	1.3*	0.3091	0.3120
0.19**	0.3712	0.3628	0.7	0.3399	0.3407	1.5**	0.3087	0.3092
0.20*	0.3706	0.3660	0.8	0.3578	0.3616	2.0	0.3061	0.3072
0.21	0.3752	0.3698	0.9	0.3810	0.3866	3.0	0.3045	0.3062
0.25	0.3818	0.3749	1.0	0.4163	0.4208	5.0	0.3037	0.3062

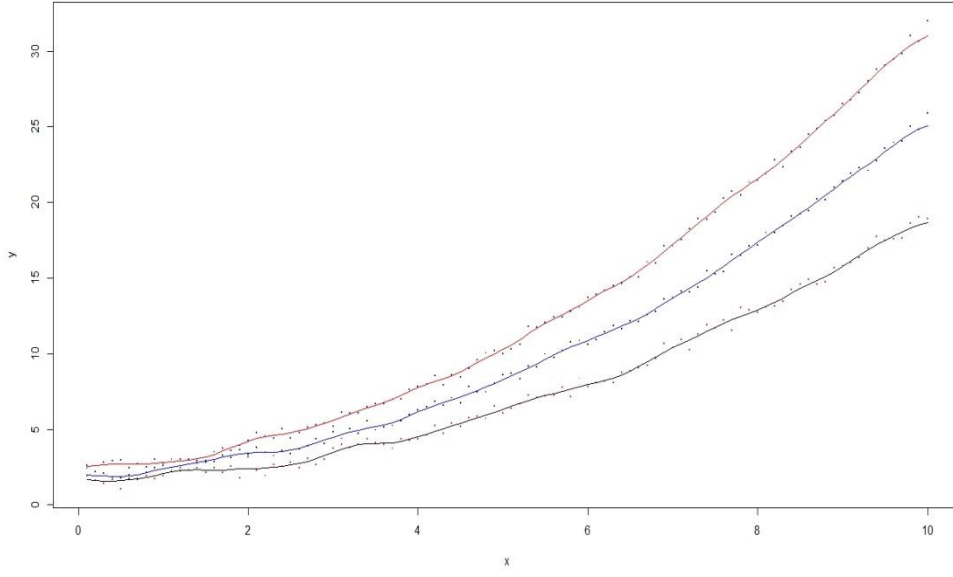
** *GCV* Kriteri ile belirlenen optimum *h* band genişliği değeri

* *CV* Kriteri ile belirlenen optimum *h* band genişliği değeri

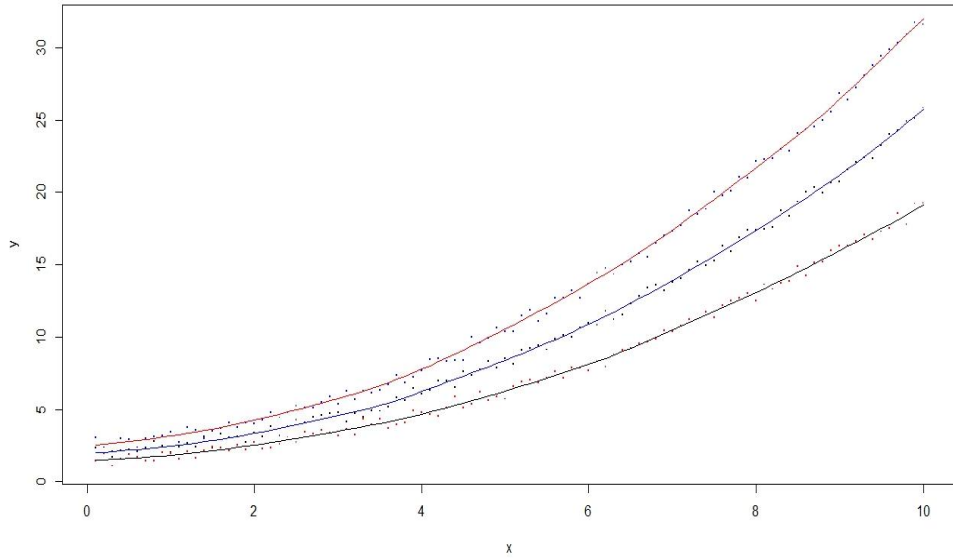
Çizelge 5.2 Veri Seti-1 için Epanechnikov çekirdeği kullanıldığında çekirdek düzeltme, $p = 1$ ve $p = 3$ için CV ve GCV kriterleri ile band genişliği seçimi

Veri Seti-1								
Çekirdek Düzeltme Epanechnikov			Deg = 1 Epanechnikov			Deg = 3 Epanechnikov		
<i>h</i>	<i>CV</i>	<i>GCV</i>	<i>h</i>	<i>CV</i>	<i>GCV</i>	<i>h</i>	<i>CV</i>	<i>GCV</i>
0.30	0.3796	0.3794	0.5	0.3369	0.3425	0.8	0.4074	0.3588
0.35	0.3758	0.3684	0.6	0.3323	0.3315	1.0	0.3659	0.3449
0.38	0.3732	0.3669	0.8	0.3223	0.3208	1.2	0.3514	0.3348
0.39	0.3728	0.3663	0.9**	0.3209	0.3203	1.5	0.3368	0.3317
0.40	0.3717	0.3659	1.0*	0.3195	0.3209	1.7	0.3294	0.3241
0.41**	0.3706	0.3627	1.1	0.3225	0.3224	1.9	0.3288	0.3235
0.42*	0.3701	0.3649	1.2	0.3227	0.3237	2.0	0.3222	0.3214
0.43	0.3720	0.3650	1.3	0.3287	0.3260	2.5**(**)	0.3195	0.3150
0.45	0.3727	0.3662	1.5	0.3391	0.3358	3.0	0.3148	0.3147
0.50	0.3787	0.3687	1.7	0.3403	0.3397	5.0	0.3075	0.3090

Elde edilen sonuçlar çerçevesinde *CV* kriteri ile seçilen uygun band genişlikleri için uyum eğrileri çizdirilmiştir. Şekil 5.1 bulanık çekirdek düzeltme için, Şekil 5.2 bulanık yerel doğrusal ve Şekil 5.3 bulanık yerel kübik modeller için çizdirilmiştir. Bundan sonraki tüm grafiklerde mavi ile gösterilen eğri merkez eğrisini, kırmızı ile gösterilen eğri sağ saçılım (merkez+saçılım) eğrisini ve siyah ile gösterilen eğri sol saçılım (merkez-saçılım) eğrisini göstermektedir. Ayrıca grafik tanımlamalarında kullanılan $p = 1$ bulanık yerel doğrusal modeli, $p = 3$ bulanık yerel kübik modeli ifade etmektedir.

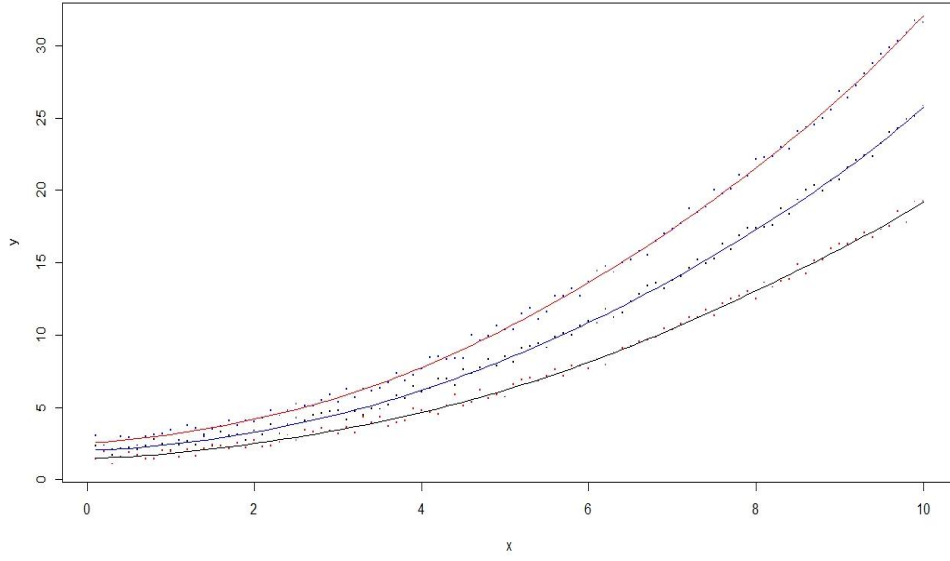


Şekil 5.1 Veri Seti-1’de çekirdek düzeltme ve Gauss çekirdeği kullanıldığında CV kriteri ile belirlenen $h = 0.2$ değeri için uyum eğrisi¹



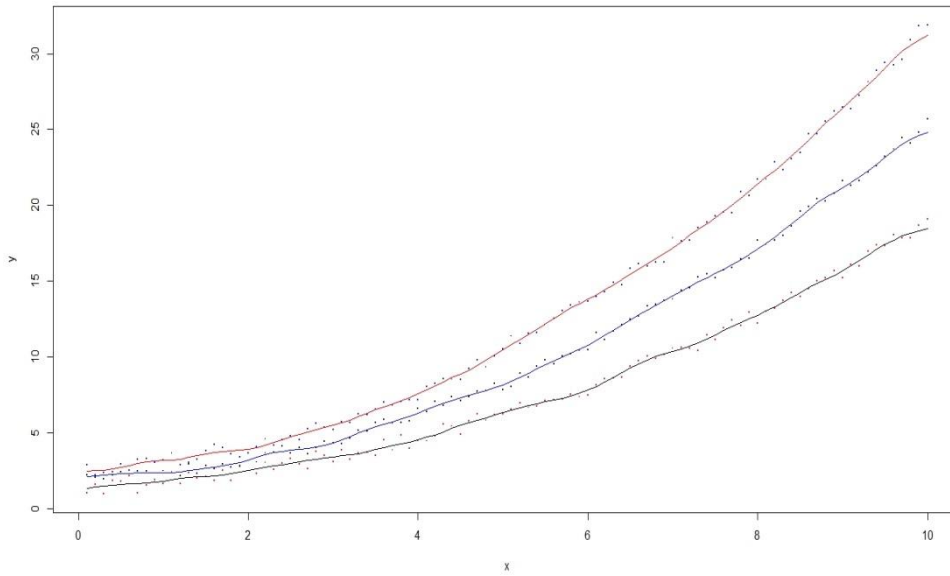
Şekil 5.2 Veri Seti-1’de $p = 1$ ve Gauss çekirdeği kullanıldığında CV kriteri ile belirlenen $h = 0.5$ değeri için uyum eğrisi

¹ Şekillerdeki **mavi eğri** “merkez” için yapılan tahmini, **kırmızı eğri** “merkez+saçılımın” tahminini, **siyah eğri** ise “merkez-saçılım” tahminini göstermektedir.

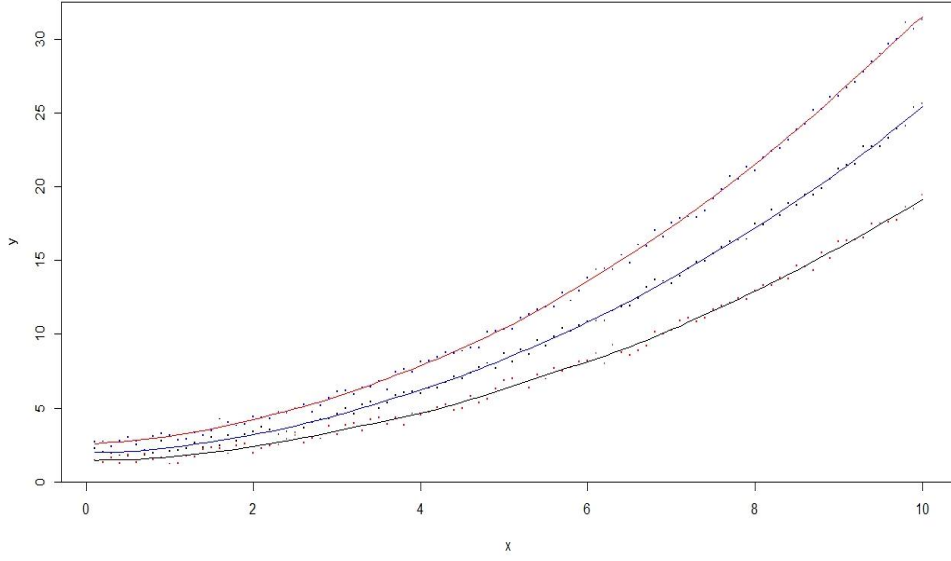


Şekil 5.3 Veri Seti-1’de $p = 3$ ve Gauss çekirdeği kullanıldığında CV kriteri ile belirlenen $h = 1.3$ değeri için uyum eğrisi

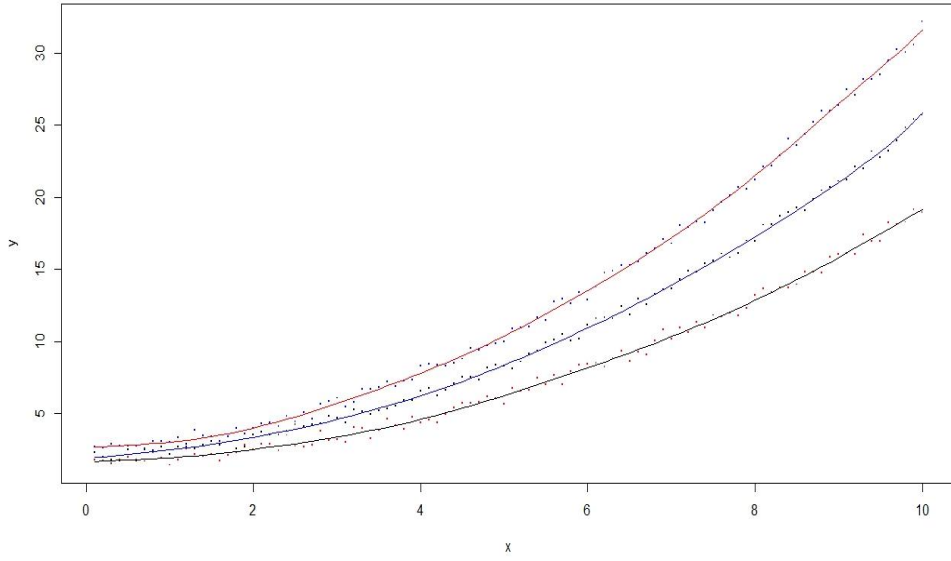
Gerçekleştirilen tahminlere ait eğrilerden Şekil 5.4 bulanık çekirdek düzeltme, Şekil 5.5 bulanık yerel doğrusal ve Şekil 5.6 bulanık yerel kübik modellerde Epanechnikov çekirdeğinin kullanımı ile elde edilmiştir.



Şekil 5.4 Veri Seti-1’de çekirdek düzeltme ve Epanechnikov çekirdeği kullanıldığında CV kriteri ile belirlenen $h = 0.42$ değeri için uyum eğrisi



Şekil 5.5 Veri Seti-1’de $p = 1$ ve Epanechnikov çekirdeği kullanıldığında CV kriteri ile belirlenen $h = 1.0$ değeri için uyum eğrisi



Şekil 5.6 Veri Seti-1’de $p = 3$ ve Epanechnikov çekirdeği kullanıldığında CV kriteri ile belirlenen $h = 2.5$ değeri için uyum eğrisi

Veri Seti-2 için CV ve GCV kriterleri yardımıyla band genişliğinin seçiminde belirli aralıklarda yaklaşım gerçekleştirilmiş ve Çizelge 5.3 ve Çizelge

5.4'de görülen her bir h değerlerine karşılık gelen CV ve GCV skorları belirlenmiştir.

Çizelge 5.3 Veri Seti-2 için Gauss çekirdeği kullanıldığında çekirdek düzeltme, $p = 1$ ve $p = 3$ için CV ve GCV kriterleri ile band genişliği seçimi

Veri Seti-2								
Çekirdek Düzeltme Gauss			Deg = 1 Gauss			Deg = 3 Gauss		
h	CV	GCV	h	CV	GCV	h	CV	GCV
0.10	0.4192	0.4081	0.1	0.4108	0.4033	0.2	0.4642	0.3891
0.11	0.4173	0.4041	0.15	0.3740	0.3711	0.5**	0.3409	0.3324
0.12	0.4021	0.3942	0.18	0.3681	0.3606	0.6*	0.3362	0.3332
0.13*(**)	0.4020	0.3910	0.2**	0.3637	0.3590	0.7	0.3379	0.3341
0.14	0.4067	0.3945	0.21*	0.3631	0.3618	0.8	0.3500	0.3477
0.15	0.4070	0.3947	0.22	0.3680	0.3651	1.0	0.4045	0.4027
0.16	0.4079	0.3950	0.25	0.3737	0.3713	1.2	0.5345	0.5344
0.17	0.4148	0.3958	0.29	0.3912	0.3901	1.5	0.8635	0.8796
0.20	0.4304	0.4134	0.3	0.4012	0.3978	1.7	0.9802	0.9967

Çizelge 5.4 Veri Seti-2 için Epanechnikov çekirdeği kullanıldığında çekirdek düzeltme, $p = 1$ ve $p = 3$ için CV ve GCV kriterleri ile band genişliği seçimi

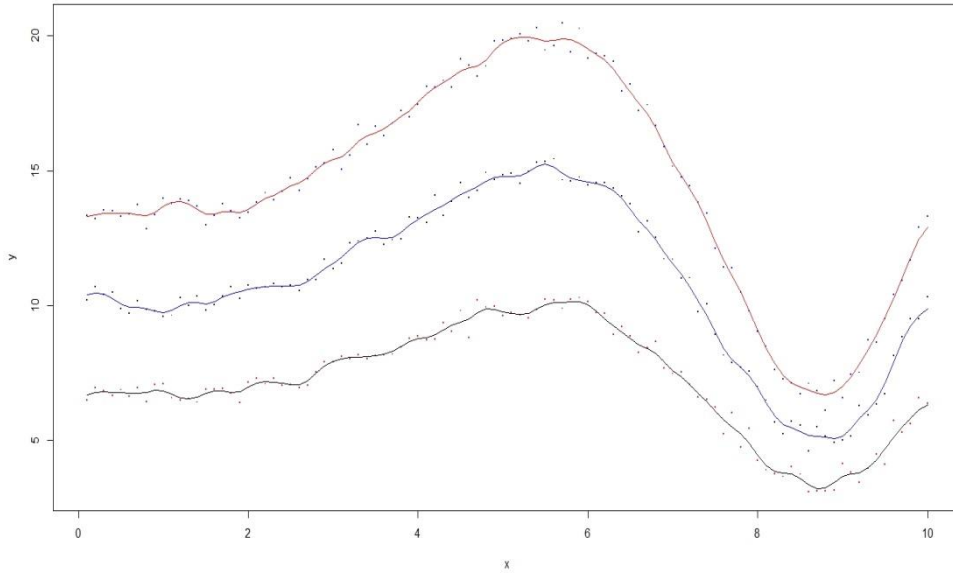
Veri Seti-2								
Çekirdek Düzeltme Epanechnikov			Deg = 1 Epanechnikov			Deg = 3 Epanechnikov		
h	CV	GCV	h	CV	GCV	h	CV	GCV
0.20	0.4563	0.4491	0.4**	0.3712	0.3600	0.9	0.3843	0.3489
0.25	0.4101	0.3987	0.5*	0.3632	0.3627	1.0	0.3686	0.3445
0.27	0.4099	0.3982	0.51	0.3634	0.3639	1.1	0.3581	0.3434
0.28	0.4098	0.3963	0.52	0.3667	0.3639	1.5	0.3437	0.3322
0.30	0.4051	0.3945	0.53	0.3668	0.3653	1.6	0.3383	0.3312
0.32*(**)	0.4026	0.3922	0.6	0.3782	0.3764	1.7**	0.3381	0.3306
0.34	0.4097	0.3934	0.7	0.4050	0.4057	1.8*	0.3370	0.3322
0.35	0.4102	0.3977	0.8	0.4507	0.4501	2.0	0.3381	0.3375
0.40	0.4146	0.4037	0.9	0.4923	0.4891	2.5	0.3738	0.3692

Çizelge 5.5'te Veri Seti-1 ve Veri Seti-2 için Gauss ve Epanechnikov çekirdekleri kullanılarak, geliştirilen CV ve GCV kriterleri ve Wang ve ark. (2007) çalışmasında kullanılan CV kriteri ile seçilen band genişlikleri değerleri bulanık çekirdek düzeltme, bulanık yerel doğrusal (LLS) (local linear smoothing) ve bulanık yerel kübik (LCS) (local cubic smoothing) modeller için verilmiştir.

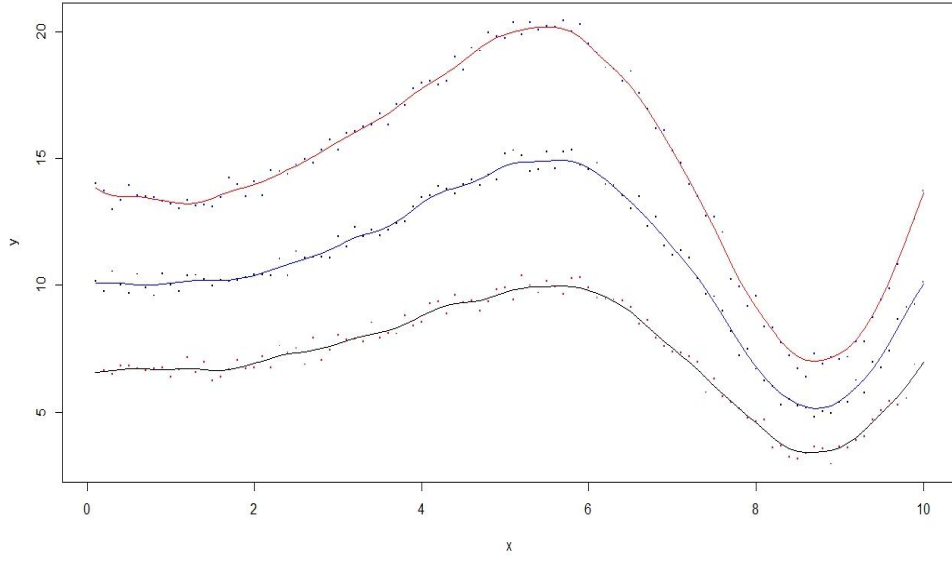
Çizelge 5.5 Veri Seti-1 ve Veri Seti-2 için Gauss ve Epanechnikov çekirdekleri kullanılarak çekirdek düzeltme, $p = 1$ ve $p = 3$ için geliştirilen CV ve GCV kriterleri ve Wang ve ark.(2007) çalışmasındaki CV kriteri ile seçilen band genişlikleri

	Çekirdek Fonksiyonu	Çekirdek Düzeltme		LLS		LCS		Wang ve ark. (2007) (LLS)
		$CV(h)$	$GCV(h)$	$CV(h)$	$GCV(h)$	$CV(h)$	$GCV(h)$	$CV(h)$
Veri Seti-1	Gauss	0.2	0.19	0.5	0.4	1.3	1.5	0.51
Veri Seti-1	Epanechnikov	0.42	0.41	1.0	0.9	2.5	2.5	1.20
Veri Seti-2	Gauss	0.13	0.13	0.21	0.2	0.6	0.5	0.21
Veri Seti-2	Epanechnikov	0.32	0.32	0.5	0.4	1.8	1.7	0.52

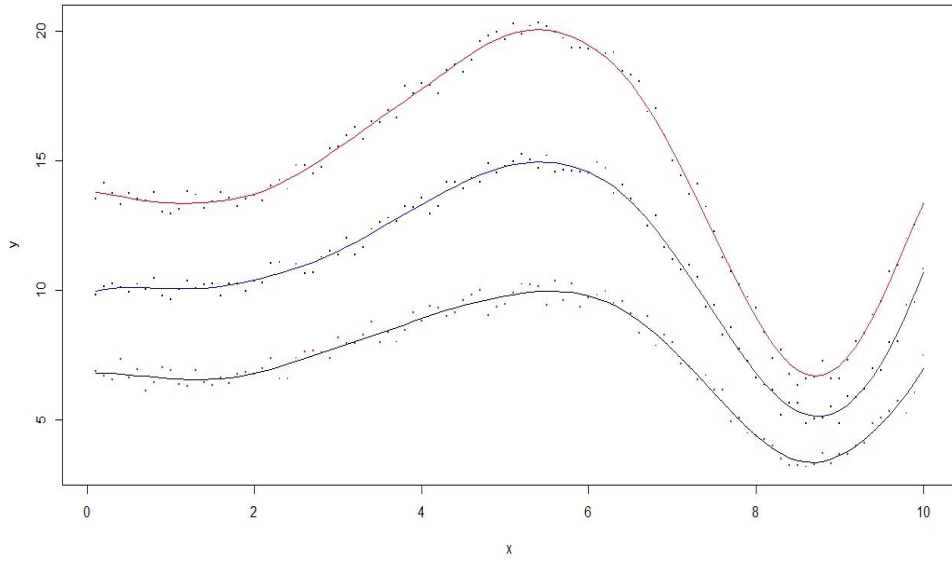
Veri Seti-2 için elde edilen sonuçlar çerçevesinde CV kriteri ile seçilen uygun band genişlikleri için Gauss çekirdeği kullanılarak çizdirilen uyum eğrileri Şekil 5.7, Şekil 5.8 ve Şekil 5.9'da görülmektedir.



Şekil 5.7 Veri Seti-2 için çekirdek düzeltme ve Gauss çekirdeği kullanıldığında CV kriteri ile belirlenen $h = 0.13$ için uyum eğrisi

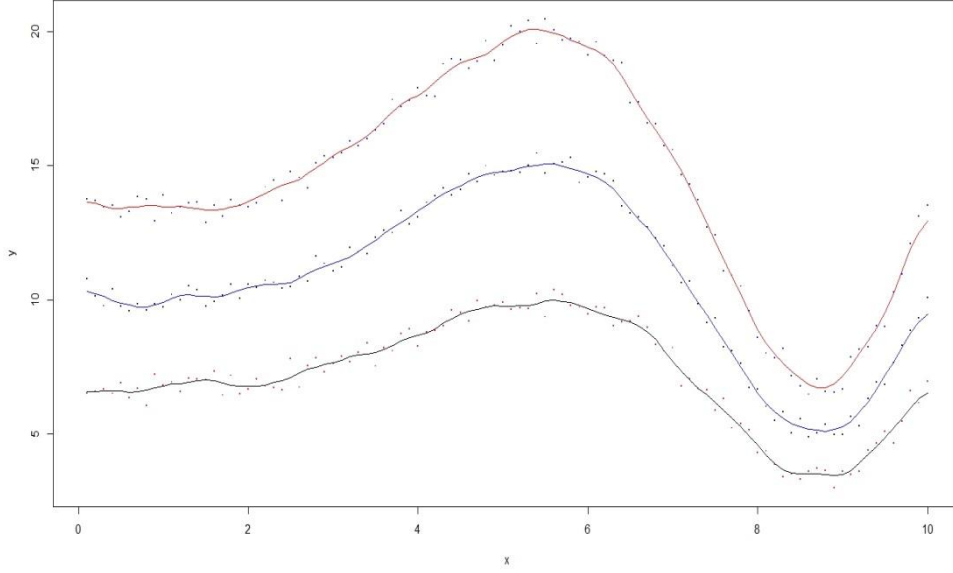


Şekil 5.8 Veri Seti-2 için $p = 1$ ve Gauss çekirdeği kullanıldığında CV kriteri ile belirlenen $h = 0.21$ için uyum eğrisi

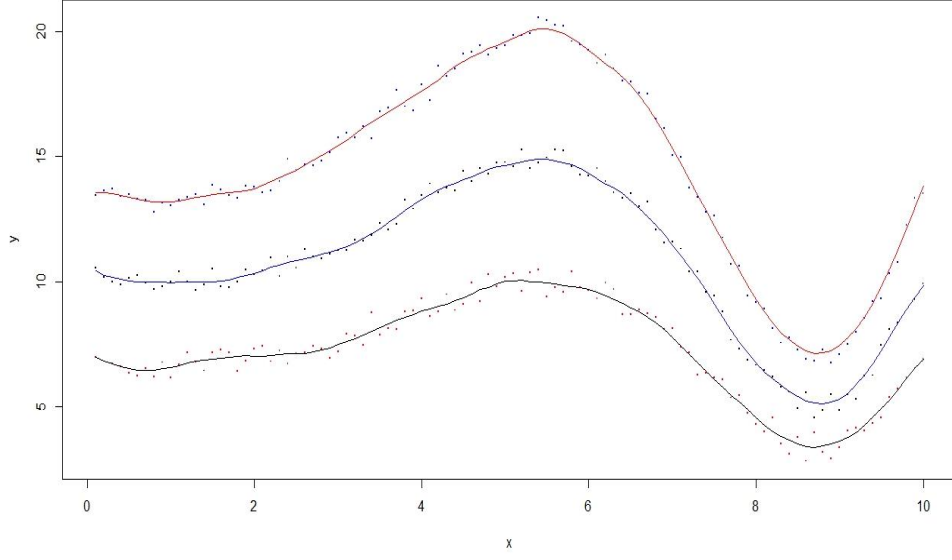


Şekil 5.9 Veri Seti-2 için $p = 3$ ve Gauss çekirdeği kullanıldığında CV kriteri ile belirlenen $h = 0.6$ için uyum eğrisi

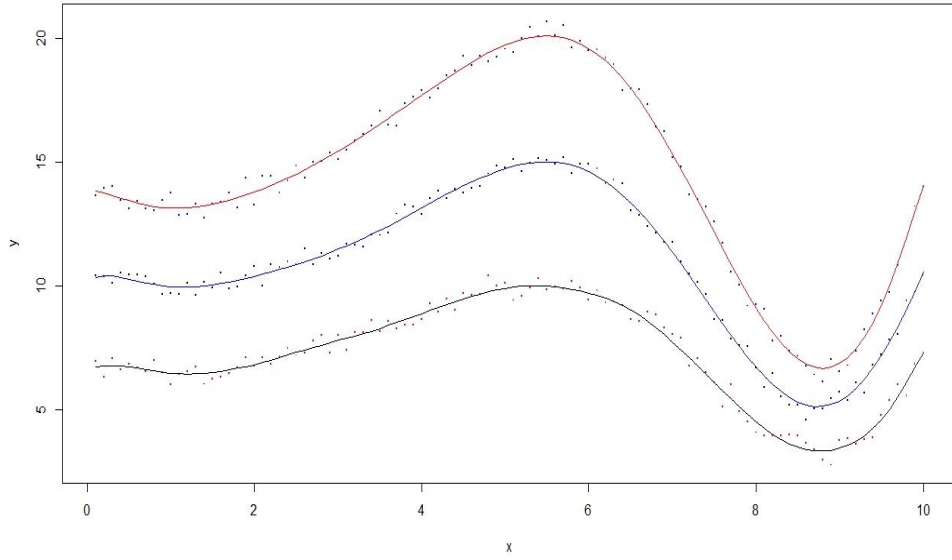
Veri Seti-2 için elde edilen sonuçlar çerçevesinde CV kriteri ile seçilen uygun band genişlikleri için Epanechnikov çekirdeği kullanılarak çizdirilen uyum eğrileri Şekil 5.10, Şekil 5.11 ve Şekil 5.12’de görülmektedir.



Şekil 5.10 Veri Seti-2 için çekirdek düzeltme ve Epanechnikov çekirdeği kullanıldığında CV kriteri ile belirlenen $h = 0.32$ için uyum eğrisi



Şekil 5.11 Veri Seti-2 için $p = 1$ ve Epanechnikov çekirdeği kullanıldığında CV kriteri ile belirlenen $h = 0.5$ için uyum eğrisi



Şekil 5.12 Veri Seti-2 için $p = 3$ ve Epanechnikov çekirdeği kullanıldığında CV kriteri ile belirlenen $h = 1.8$ için uyum eğrisi

Modellerin performanslarını karşılaştırmak amacıyla ortalama karesel hata değeri *ASE* kullanılmış, Veri Seti-1 ve Veri Seti-2 için *CV* ve *GCV* kriterleri ile seçilen band genişlikleri ile oluşturulmuş modellerin performanslarını gösteren özet bilgilere Çizelge 5.6 ve Çizelge 5.7’de yer verilmiştir.

Çizelge 5.6 Veri Seti-1 için farklı çekirdekler kullanılarak çekirdek düzeltme, $p = 1$ ve $p = 3$ için geliştirilen *CV* ve *GCV* kriterleri ve Wang ve ark. (2007) çalışmasındaki *CV* kriteri ile seçilen band genişliklerine göre model performansları

Yöntem	Seçim Kriteri	Gauss Çekirdeği	<i>ASE</i>	Epanechnikov Çekirdeği	<i>ASE</i>
Çekirdek Düzeltme	<i>CV</i>	0.20	0.2324987	0.42	0.2453053
	<i>GCV</i>	0.19	0.2247401	0.41	0.2428460
<i>LLS</i>	<i>CV</i>	0.50	0.2646376	1.00	0.2653438
	<i>GCV</i>	0.40	0.2553885	0.90	0.2626948
<i>LCS</i>	<i>CV</i>	1.30	0.2699719	2.50	0.2650361
	<i>GCV</i>	1.50	0.2710324	2.50	0.2653659
Wang ve ark. (2007) (<i>LLS</i>)	<i>CV</i>	0.51	0.2662507	1.20	0.2760315

Çizelge 5.7 Veri Seti-2 için farklı çekirdekler kullanılarak çekirdek düzeltme, $p = 1$ ve $p = 3$ için geliştirilen *CV* ve *GCV* kriterleri ve Wang ve ark. (2007) çalışmasındaki *CV* kriteri ile seçilen band genişliklerine göre model performansları

Yöntem	Seçim Kriteri	Gauss Çekirdeği	<i>ASE</i>	Epanechnikov Çekirdeği	<i>ASE</i>
Çekirdek Düzeltme	<i>CV</i>	0.13	0.1871375	0.32	0.2285608
	<i>GCV</i>	0.13	0.1879244	0.32	0.2285440
<i>LLS</i>	<i>CV</i>	0.21	0.2308442	0.50	0.2503436
	<i>GCV</i>	0.20	0.2264042	0.40	0.2293865
<i>LCS</i>	<i>CV</i>	0.60	0.2536422	1.80	0.2660677
	<i>GCV</i>	0.50	0.2457798	1.70	0.2623458
Wang ve ark. (2007) (<i>LLS</i>)	<i>CV</i>	0.21	0.2308016	0.52	0.2579763

Yapılan uygulama çalışması sonucunda, grafiklerinde de görüldüğü gibi Veri Seti-1 daha doğrusal bir yapıya sahip fonksiyonla ifade edilmesinden dolayı, bulanık yerel kübik modelde band genişliği seçim kriterlerinden *CV* ve *GCV* skorları sürekli azalmakta ve band genişliği olarak çok büyük bir değer seçimi söz konusu olabilmektedir. Bunun nedeni zaten model yapısının doğrusala yakın olması ve band genişliğinin ne kadar büyük seçilirse seçilsin sonuç üzerinde önemli bir etki yaratmayacak olmasıdır. Çekirdek düzeltme kullanıldığında

seçilen band genişliği değeri çok daha küçüktür. Bölüm 2 ve Bölüm 4'te değinildiği üzere band genişliği ne kadar büyükse pürüzlülük o kadar az olacaktır ve böyle bir doğrusal model büyük band genişliği ile de rahatlıkla ifade edilebilecektir. Bu bakımdan *CV* ve *GCV* skorlarına ait değerlerin olduğu çizelgeden skor değerlerindeki düşüşün azaldığı noktalar dikkate alınmış ve band genişliği olarak o skora karşılık gelen *h* değeri belirlenmiştir. Çizelge 5.6 ve Çizelge 5.7'de görülen *ASE* değerleri çekirdek düzeltme için her ne kadar band genişliğine bağlı olarak küçük hesaplanmış olsa da modeller için grafiksel gösterimler dikkate alındığında çok eğrisel uyum eğrileri yarattığı görülmektedir. Bulanık yerel doğrusal ve bulanık yerel kübik modeller için grafiksel ifadeler çok daha pürüzsüz (dalgalanması az olan) bir yapı sergilemektedir. Çekirdek düzeltme, ciddi sınır etkisine maruz kalmasından dolayı çok daha dalgalı tahmin eğrileri üretmektedir. Wang ve ark. (2007) çalışmalarında bulanık doğrusal modelin, çekirdek düzeltmeye göre daha pürüzsüz bir tahminleme gerçekleştirdiğini ortaya koymuşlardır. Ayrıca, yerel düzeltme tekniğinde sınır etkisi anlamlı bir şekilde azalmıştır.

Veri Seti-2 daha eğrisel bir yapıya sahiptir. Modellerin performans tablolarına göre bulanık yerel doğrusal model için her ne kadar *ASE* değeri daha küçük olsa da, aynı çekirdek fonksiyonları için oluşturulan grafikler incelendiğinde bulanık yerel kübik modelde dalgalanmanın daha az olduğu ve veri setini açıklamada daha başarılı olduğu söylenebilir. Veri Seti-1'e benzer şekilde yine bulanık çekirdek düzeltmede band genişliği küçük seçilmiştir, ancak çok daha dalgalı eğri oluşturmuştur.

Aynı veri setleri Wang ve ark. (2007) çalışmasında kullanılmış bulanık çekirdek düzeltme ve bulanık yerel doğrusal model dikkate alınarak *CV*'nin genel tanımlaması yapılmış ve band genişliği hesaplanmıştır. Bu çalışmada geliştirilen *CV* kriteri yardımıyla bulanık yerel doğrusal model için hesaplanan band genişlikleri ile Wang ve ark. (2007) çalışmalarında elde edilen band genişliklerinin birbirine oldukça yakın oldukları gözlemlenmiştir.

5.2. Veri Seti-3 İçin İki Farklı Durum Dikkate Alınarak Parametrik Olmayan Bulanık Yerel Doğrusal ve Yerel Kübik Modellerin Karşılaştırılması

Fan ve Gijbels (1995) çalışmalarında kullanmış oldukları (2) nolu fonksiyondan faydalanarak, bu fonksiyonun bulanık yapıda ifadesi gerçekleştirilmiş ve x girdi değişkeninin $[0,10]$ aralığında $x_i = 0.1i$, ($i = 1,2, \dots, 100$) değerlerini ve $[-2, 2]$ aralığında $x_i = -2 + 0.04i$, $i = 0, 1, 2, \dots, 100$ değerlerini aldığı iki durum dikkate alınarak analizler yapılmıştır.

Durum-1 $g_3(x)$, $[0,10]$ aralığında aşağıdaki gibi tanımlanmış bir fonksiyon olsun:

$$g_3(x) = \sin(2x) + 2\exp(-16x^2)$$

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

$$y_i = g_3(x_i) + \text{rand}[-0.5, 0.5]$$

$$\sigma_i = \left| \frac{1}{3} g_3(x_i) + \text{rand}[-0.25, 0.25] \right|$$

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

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

Veri Seti-3 için Gauss ve Epanechnikov çekirdekleri kullanılarak CV ve GCV kriterleri ile hesaplanan band genişlikleri Çizelge 5.8 ve Çizelge 5.9'da görülmektedir. Çizelge 5.10'da ise bulanık yerel doğrusal ve bulanık yerel kübik

modeller için *CV* ve *GCV* kriterleri ile seçilmiş olan band genişliklerine göre modellerin performans değerleri yer almaktadır.

Çizelge 5.8 Veri Seti-3 için Gauss çekirdeği kullanıldığında $p = 1$ ve $p = 3$ için *CV* ve *GCV* kriterleri ile band genişliği seçimi

Veri Seti-3					
<i>Deg = 1 Gauss</i>			<i>Deg = 3 Gauss</i>		
<i>h</i>	<i>CV</i>	<i>GCV</i>	<i>h</i>	<i>CV</i>	<i>GCV</i>
0.08	0.4402891	0.4181709	0.2	0.4570003	0.3949112
0.1	0.4169853	0.3995223	0.28	0.3922297	0.3684027
0.12	0.3975523	0.3873080	0.3**	0.3894418	0.3655387
0.13	0.3886253	0.3842345	0.32*	0.3835013	0.3658378
0.14	0.3885327	0.3822197	0.34	0.3850496	0.3680264
0.15*(**)	0.3800287	0.3757649	0.35	0.3871625	0.3701007
0.16	0.3818987	0.3780425	0.37	0.3875013	0.3744188
0.18	0.3860575	0.3797156	0.38	0.3867715	0.3746143
0.2	0.3878293	0.3801732	0.4	0.3910980	0.3805197
0.3	0.4461288	0.4395628	0.5	0.4299485	0.4222734

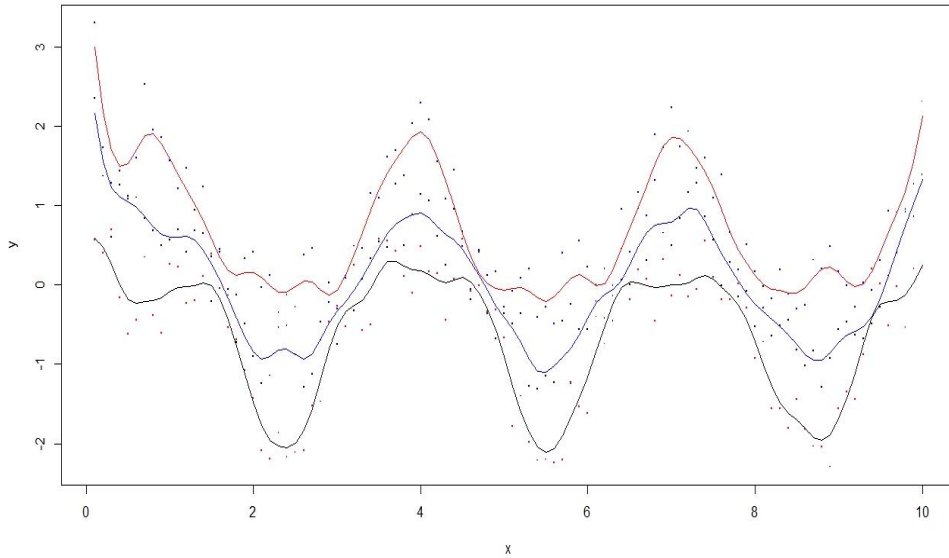
Çizelge 5.9 Veri Seti-3 için Epanechnikov çekirdeği kullanıldığında $p = 1$ ve $p = 3$ için *CV* ve *GCV* kriterleri ile band genişliği seçimi

Veri Seti-3					
<i>Deg = 1 Epanechnikov</i>			<i>Deg = 3 Epanechnikov</i>		
<i>h</i>	<i>CV</i>	<i>GCV</i>	<i>h</i>	<i>CV</i>	<i>GCV</i>
0.3	0.4070957	0.3825600	0.4	1.455825	0.4529197
0.32	0.3889494	0.3803540	0.5	0.8021971	0.4087424
0.34	0.3875548	0.3789032	0.6	0.5297677	0.3913634
0.35	0.3872863	0.3759302	0.8	0.4165838	0.3715183
0.36**	0.3868317	0.3712751	0.9**	0.4039611	0.3706466
0.38*	0.3821925	0.3749383	1.0	0.3956912	0.3750558
0.4	0.3853781	0.3755784	1.02*	0.3929532	0.3771889
0.41	0.3869868	0.3770249	1.05	0.3958869	0.3793286
0.42	0.3887270	0.3782136	1.08	0.3969558	0.3834527
0.5	0.3971166	0.391669	1.1	0.3972190	0.3858563

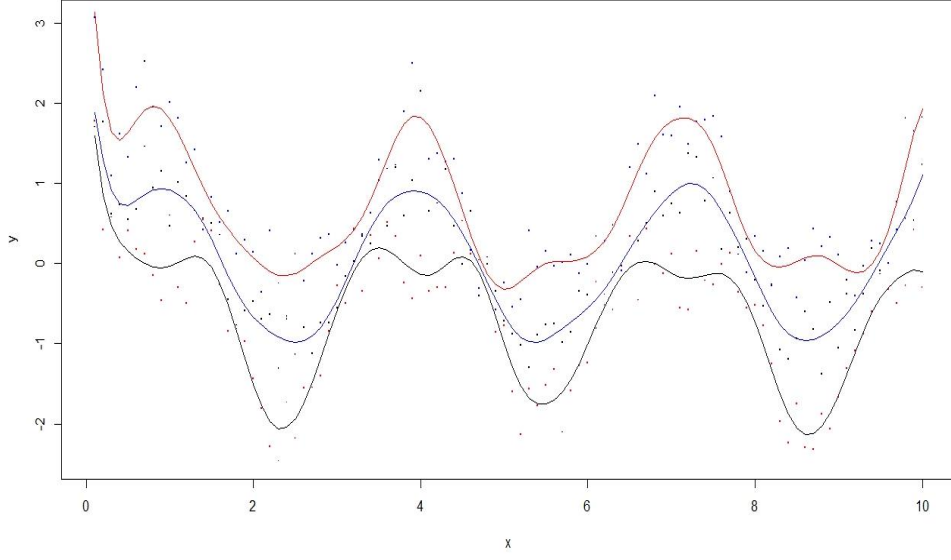
Çizelge 5.10 Veri Seti-3 için farklı çekirdekler kullanılarak $p = 1$ ve $p = 3$ için CV ve GCV kriterleri ile seçilen band genişliklerine göre model performansları

Yöntem	Seçim Kriteri	Gauss Çekirdeği	ASE	Epanechnikov Çekirdeği	ASE
<i>LLS</i>	<i>CV</i>	0.15	0.196666	0.38	0.2325479
	<i>GCV</i>	0.15	0.1970035	0.36	0.2289633
<i>LCS</i>	<i>CV</i>	0.32	0.2274435	1.02	0.2611587
	<i>GCV</i>	0.30	0.2219668	0.90	0.2454555

Çizelge 5.10 incelendiğinde hem Gauss çekirdeği hem de Epanechnikov çekirdekleri ile seçilen band genişlikleri için bulanık yerel doğrusal model ile yapılan seçimde band genişliği oldukça küçük olmakla birlikte modelin performansının daha iyi olduğu gözlenmektedir. Ancak bunun yanı sıra band genişliğinin bulanık yerel kübik modelde, bulanık yerel doğrusal modele göre yaklaşık iki kat daha büyük olduğu dikkate alındığında model performansındaki az miktardaki değişiklik gözardı edilebilir durumdadır. Şekil 5.13, Veri Seti-3 için $p = 1$ ve Gauss çekirdeği kullanıldığında CV kriteri ile belirlenen $h = 0.15$ için uyum eğrilerini ve Şekil 5.14, $p = 3$ ve Gauss çekirdeği kullanıldığında CV kriteri ile belirlenen $h = 0.32$ için uyum eğrilerini vermektedir.

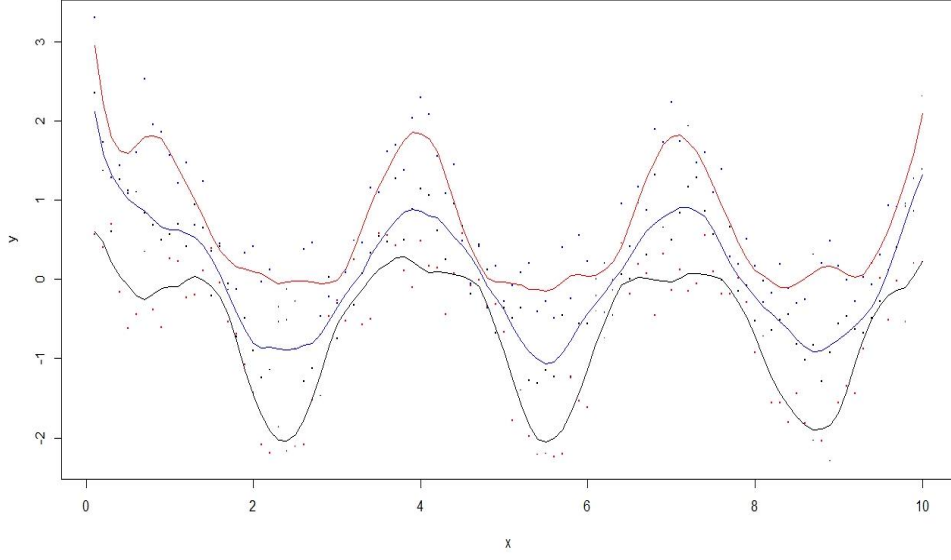


Şekil 5.13 Veri Seti-3 için $p = 1$ ve Gauss çekirdeği kullanıldığında CV kriteri ile belirlenen $h = 0.15$ için uyum eğrisi

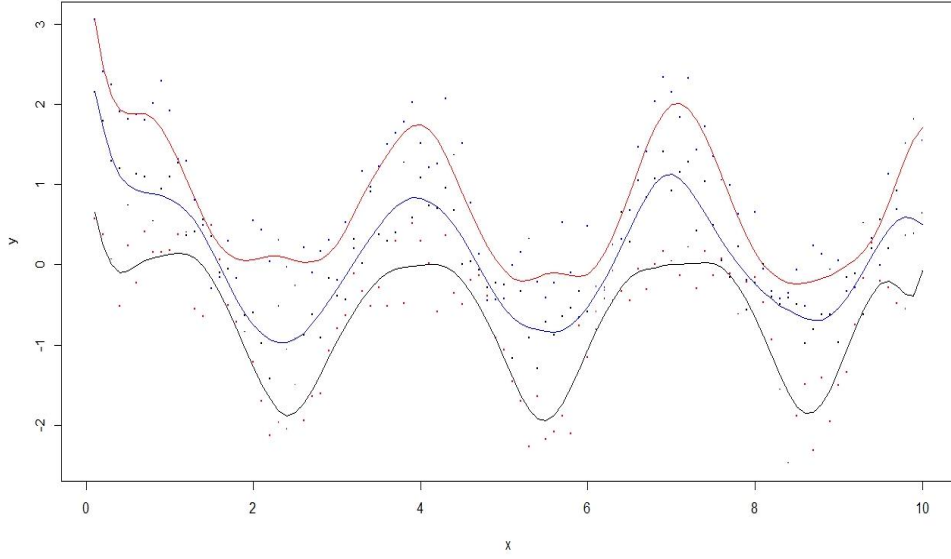


Şekil 5.14 Veri Seti-3 için $p = 3$ ve Gauss çekirdeği kullanıldığında CV kriteri ile belirlenen $h = 0.32$ için uyum eğrisi

Şekil 5.15, Veri Seti-3 için $p = 1$ ve Epanechnikov çekirdeği kullanıldığında CV kriteri ile belirlenen $h = 0.38$ için uyum eğrilerini ve Şekil 5.16, $p = 3$ ve Epanechnikov çekirdeği kullanıldığında CV kriteri ile belirlenen $h = 1.02$ için uyum eğrilerini göstermektedirler.



Şekil 5.15 Veri Seti-3 için $p = 1$ ve Epanechnikov çekirdeği kullanıldığında CV kriteri ile belirlenen $h = 0.38$ için uyum eğrisi



Şekil 5.16 Veri Seti-3 için $p = 3$ ve Epanechnikov çekirdeği kullanıldığında CV kriteri ile belirlenen $h = 1.02$ için uyum eğrisi

Durum-2 $g_3(x)$, fonksiyonu çıktı değişkeninin üretilmesinde kullanılan bir fonksiyon olmak üzere aşağıdaki gibi tanımlansın:

$$g_3(x) = \sin(2x) + 2\exp(-16x^2)$$

$[-2, 2]$ aralığında $x_i = -2 + 0.04i$, $i = 0, 1, 2, \dots, 100$ noktaları için;

$$y_i = g_3(x_i) + \text{rand}[0, 0.4^2]$$

$$\sigma_i = \left| \frac{1}{3} g_3(x_i) + \text{rand}[-0.25, 0.25] \right|$$

olarak hesaplanırlar. Burada $\text{rand}[0, 0.4^2]$, herbir i için ortalaması 0 ve standart sapması 0.4 olan normal dağılımdan bağımsız olarak türetilen bir rassal sayıyı göstermektedir. Saçılım değerleri için $\text{rand}[-0.25, 0.25]$ ise $[-0.25, 0.25]$ aralığında düzgün dağılımdan rassal olarak türetilen sayıyı göstermektedir. Gözlenen bulanık çıktıların simetrik üçgen bulanık sayılar olduğu varsayılarak, bu çıktılara ait gösterim aşağıdaki gibi tanımlanır.

$$Y_i = (l_{y_i}, c_{y_i}, u_{y_i})_T = (y_i - \sigma_i, y_i, y_i + \sigma_i)_T \quad i = 0, 1, \dots, 100$$

Çizelge 5.11 ve Çizelge 5.12, Veri Seti-3 (Durum-2) için Gauss ve Epanechnikov çekirdeği kullanıldığında $p = 1$ ve $p = 3$ için CV ve GCV kriterleri ile band genişliği seçim değerlerini göstermektedir.

Çizelge 5.11 Veri Seti-3 (Durum-2) için Gauss çekirdeği kullanıldığında $p = 1$ ve $p = 3$ için CV ve GCV kriterleri ile band genişliği seçimi

Veri Seti-3 (Durum-2)					
Deg = 1 Gauss			Deg = 3 Gauss		
h	CV	GCV	h	CV	GCV
0.02	0.8235694	0.7987018	0.09	0.7655538	0.6725551
0.04	0.7451286	0.7047266	0.10	0.7312689	0.6582501
0.06	0.6646457	0.6393982	0.15	0.6468126	0.6199142
0.08	0.6351090	0.6237268	0.16	0.6403379	0.6172967
0.09*	0.6172738	0.6163922	0.17**	0.6399719	0.6153879
0.10**	0.6299501	0.6134867	0.18*	0.6338975	0.6186464
0.15	0.6411563	0.6383085	0.20	0.6362659	0.6228629
0.20	0.7019469	0.7039584	0.21	0.6394354	0.6284232
0.30	0.8708560	0.8773310	0.25	0.6575068	0.6430415
0.40	1.063139	1.0903710	0.30	0.6871946	0.6818113

Çizelge 5.12 Veri Seti-3 (Durum-2) için Epanechnikov çekirdeği kullanıldığında $p = 1$ ve $p = 3$ için CV ve GCV kriterleri ile band genişliği seçimi

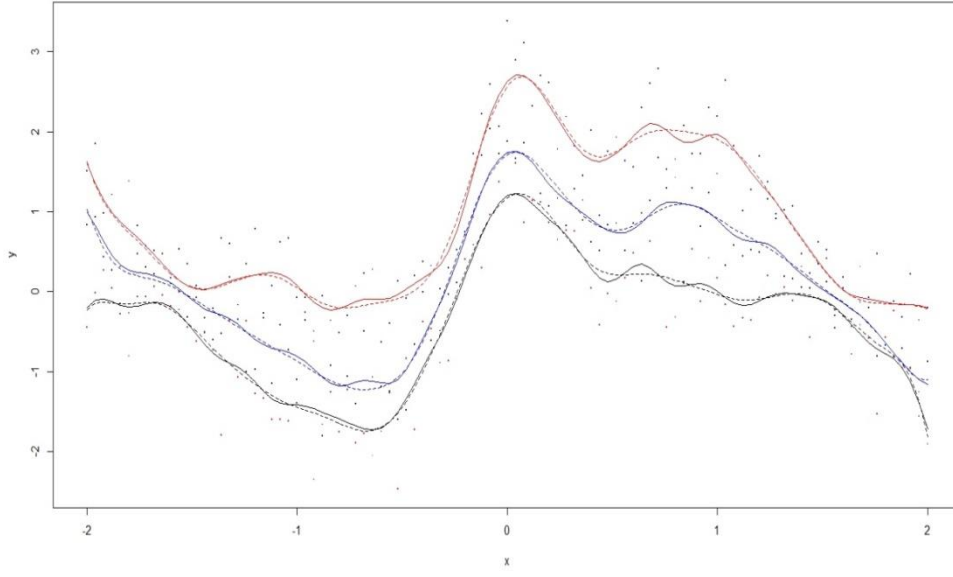
Veri Seti-3 (Durum-2)					
$Deg = 1$ Epanechnikov			$Deg = 3$ Gauss Epanechnikov		
h	CV	GCV	h	CV	GCV
0.10	0.7287000	0.6911264	0.30	0.7319782	0.6490874
0.20	0.6283867	0.6290965	0.40	0.6631265	0.6227521
0.22	0.6167920	0.6135808	0.48	0.6465255	0.6225617
0.23 ^{*(**)}	0.6165962	0.6130227	0.50 ^{**}	0.6394594	0.6225377
0.24	0.6218431	0.6140448	0.51	0.6393297	0.6239972
0.26	0.6251461	0.6190423	0.52 [*]	0.6367363	0.6256435
0.28	0.6293816	0.6294064	0.53	0.6475402	0.6275211
0.30	0.6347423	0.6333916	0.55	0.6498816	0.6292312
0.40	0.6966662	0.6854339	0.58	0.6526632	0.6360447
0.50	0.7645842	0.7645622	0.60	0.6540719	0.641242

Önceki uygulamalara benzer şekilde Veri Seti-3 (Durum-2) için Çizelge 5.11 ve Çizelge 5.12’de CV ve GCV skorlarına karşılık gelen h değerleri kullanılarak Çizelge 5.13’te modellerin performansları özetlemektedir.

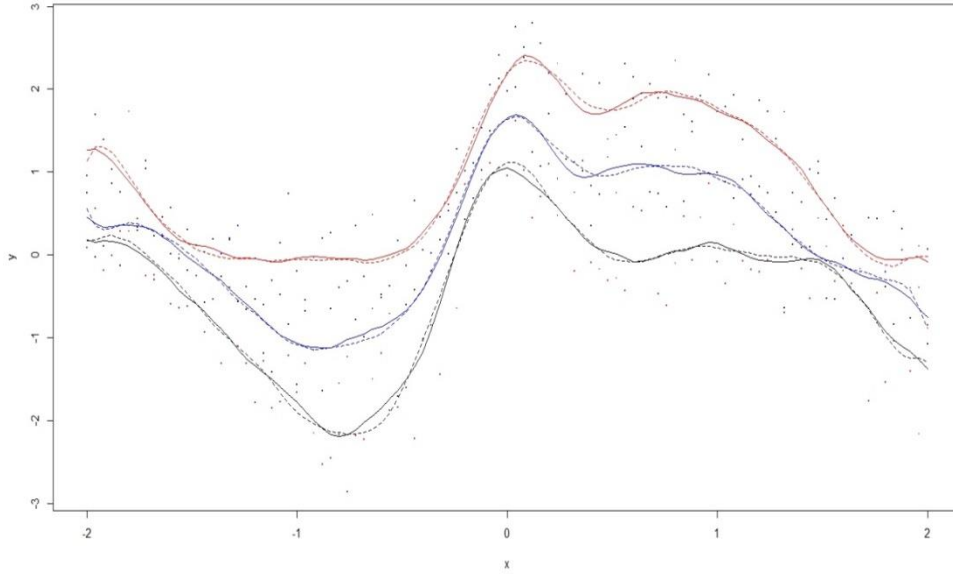
Çizelge 5.13 Veri Seti-3 (Durum-2) için farklı çekirdekler kullanılarak $p = 1$ ve $p = 3$ için CV ve GCV kriterleri ile seçilen band genişliklerine göre model performansları

Yöntem	Seçim Kriteri	Gauss Çekirdeği	ASE	Epanechnikov Çekirdeği	ASE
LLS	CV	0.09	0.4062414	0.23	0.4514972
	GCV	0.10	0.4225954	0.23	0.4515546
LCS	CV	0.18	0.4391495	0.52	0.4683489
	GCV	0.17	0.4290457	0.50	0.4626342

Şekil 5.17 ve Şekil 5.18 Veri Seti-3’te (Durum-2) CV kriteri ile seçilmiş olan band genişlikleri bakımından bulanık yerel doğrusal ve bulanık yerel kübik modelleri ile gerçekleştirilen uyumlama eğrilerini Gauss ve Epanechnikov çekirdekleri için göstermektedir.



Şekil 5.17 Veri Seti-3 (Durum-2) için Gauss çekirdeği kullanılarak CV kriteri ile seçilen band genişliği kullanılarak bulanık yerel doğrusal ve bulanık yerel kübik modeller için uyum eğrileri²



Şekil 5.18 Veri Seti-3 (Durum-2) için Epanechnikov çekirdeği kullanılarak CV kriteri ile seçilen band genişliği kullanılarak bulanık yerel doğrusal ve bulanık yerel kübik modeller için uyum eğrileri

² – düzgün eğri bulanık yerel doğrusal modeli
 -- kesikli eğri bulanık yerel kübik modeli göstermektedir.

Şekillerden de görüldüğü üzere bulanık yerel kübik model ile gerçekleştirilen uyumlamada eğri daha pürüzsüz bir yapı sergilemektedir.

5.3. Veri Seti-4 İçin Girdi Değişkeninin İki Farklı Durumunda Parametrik Olmayan Bulanık Yerel Polinomiyal Regresyon Modeli Tahmini

Fan ve Gijbels (1995) çalışmalarında kullanmış oldukları (3) nolu fonksiyondan faydalanılarak bu fonksiyonun bulanık yapıda ifadesi gerçekleştirilmiş ve x girdi değişkeninin $[0,10]$ aralığında $x_i = 0.1i$, ($i = 1,2, \dots, 100$) değerlerini ve $[-2, 2]$ aralığında $x_i = -2 + 0.04i$, $i = 0, 1, 2, \dots, 100$ değerlerini aldığı iki durum dikkate alınarak analizler yapılmıştır.

Durum-1 $g_4(x)$, $[0,10]$ aralığında aşağıdaki gibi tanımlanmış bir fonksiyon olsun:

$$g_4(x) = 0.3 \exp(-4(x+1)^2) + 0.7 \exp(-16(x-1)^2)$$

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

$$y_i = g_4(x_i) + \text{rand}[-0.5, 0.5]$$
$$\sigma_i = \left| \frac{1}{3} g_4(x_i) + \text{rand}[-0.25, 0.25] \right|$$

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

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

Veri Seti-4 için Gauss ve Epanechnikov çekirdekleri kullanılarak CV ve GCV kriterleri ile seçilen band genişliklerine ait sonuçlar Çizelge 5.14 ve Çizelge 5.15'te görülmektedir.

Çizelge 5.14 Veri Seti-4 için Gauss çekirdeği kullanıldığında $p = 1$ ve $p = 3$ için CV ve GCV kriterleri ile band genişliği seçimi

Veri Seti-4					
<i>Deg = 1 Gauss</i>			<i>Deg = 3 Gauss</i>		
<i>h</i>	<i>CV</i>	<i>GCV</i>	<i>h</i>	<i>CV</i>	<i>GCV</i>
0.1	0.3871957	0.3686705	0.5	0.3251748	0.3226408
0.2	0.3276533	0.3254221	0.7	0.3245902	0.3188162
0.3	0.3187369	0.3178944	0.9**	0.3192417	0.3125777
0.4	0.3163374	0.3131209	1.1	0.3180045	0.3153847
0.5**	0.3140655	0.3125672	1.2*	0.3171098	0.3146441
0.6*	0.3120245	0.3144125	1.3	0.3174572	0.3136009
0.7	0.3120754	0.3134060	1.4	0.3174935	0.3162453
0.8	0.3138966	0.3151195	1.5	0.3173233	0.3151692
0.9	0.3140037	0.3137522	1.6	0.3175483	0.3158080
1.0	0.3142516	0.3150397	1.7	0.3187412	0.3146858

Çizelge 5.15 Veri Seti-4 için Epanechnikov çekirdeği kullanıldığında $p = 1$ ve $p = 3$ için CV ve GCV kriterleri ile band genişliği seçimi

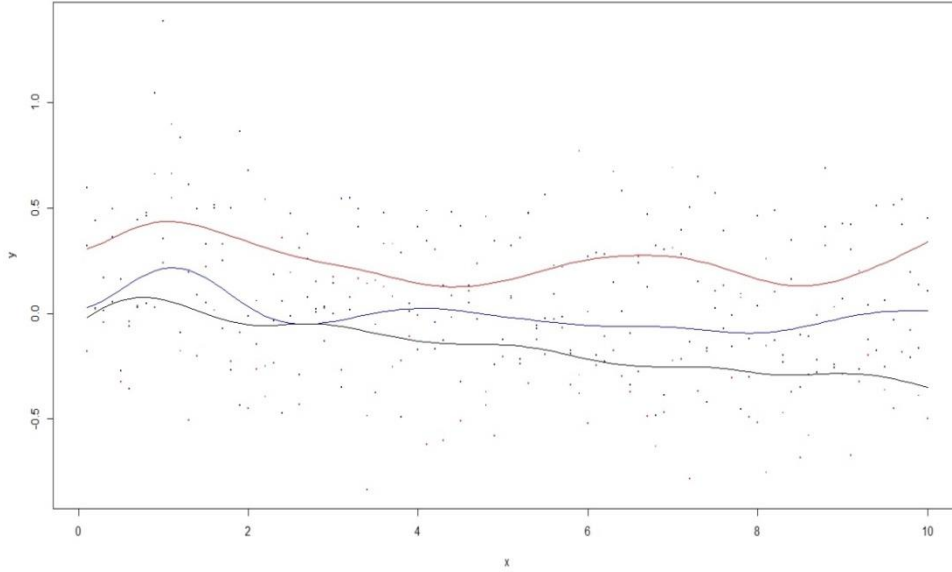
Veri Seti-4					
<i>Deg = 1 Epanechnikov</i>			<i>Deg = 3 Gauss Epanechnikov</i>		
<i>h</i>	<i>CV</i>	<i>GCV</i>	<i>h</i>	<i>CV</i>	<i>GCV</i>
0.5	0.3277299	0.3220275	1.0	0.3509093	0.3301955
0.6	0.3207170	0.3203570	1.2	0.3397622	0.3244282
0.8	0.3172449	0.3182984	1.4	0.3375901	0.3208538
0.9	0.3179808	0.3179707	1.5	0.3300838	0.3222092
1.0	0.3180053	0.3175439	1.7**	0.3249194	0.3197175
1.1**	0.3181440	0.3162035	2.0	0.3272613	0.3194528
1.2*	0.3171618	0.3181414	2.2	0.3284525	0.3209728
1.3	0.3183002	0.3194679	2.4	0.3269149	0.3208325
1.4	0.3182751	0.3168114	2.5*	0.3234212	0.3201606
1.5	0.3178162	0.3175013	3.0	0.3252913	0.3196340

Çizelge 5.16, Veri Seti-4 için farklı çekirdekler kullanılarak $p = 1$ ve $p = 3$ için CV ve GCV kriterleri ile seçilen band genişliklerine göre model performanlarına ait değerleri göstermektedir.

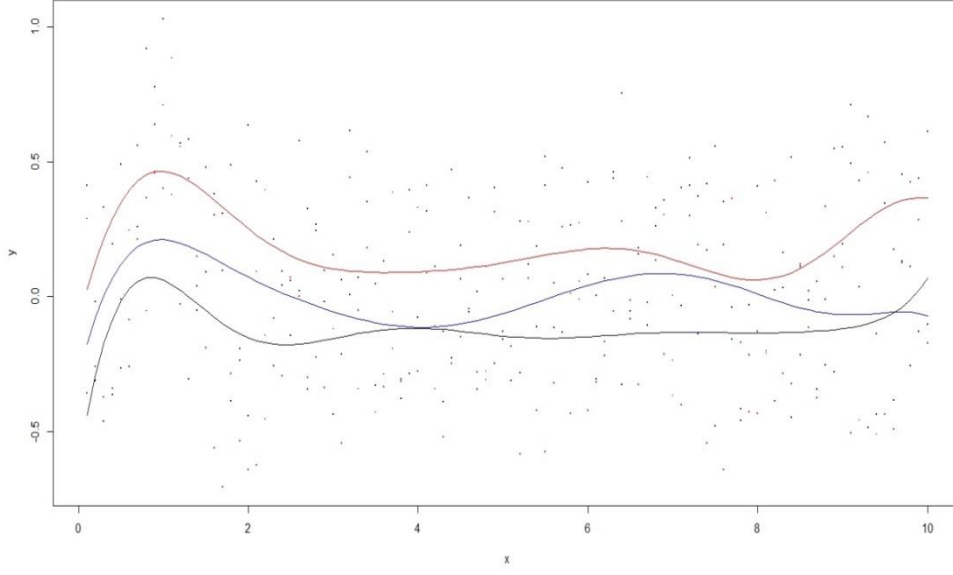
Çizelge 5.16 Veri Seti-4 için farklı çekirdekler kullanılarak $p = 1$ ve $p = 3$ için CV ve GCV kriterleri ile seçilen band genişliklerine göre model performansları

Yöntem	Seçim Kriteri	Gauss Çekirdeği	ASE	Epanechnikov Çekirdeği	ASE
LLS	<i>CV</i>	0.6	0.2628472	1.2	0.2698860
	<i>GCV</i>	0.5	0.2577808	1.1	0.2665418
LCS	<i>CV</i>	1.2	0.2674377	2.5	0.2671454
	<i>GCV</i>	0.9	0.2597413	1.7	0.2522198

Şekil 5.19 Veri Seti-4 için $p = 1$ ve Gauss çekirdeği kullanıldığında CV kriteri ile seçilen $h = 0.6$ için uyum eğrilerini ve Şekil 5.20, $p = 3$ ve Gauss çekirdeği kullanıldığında CV kriteri ile seçilen $h = 1.2$ için uyum eğrilerini göstermektedir.

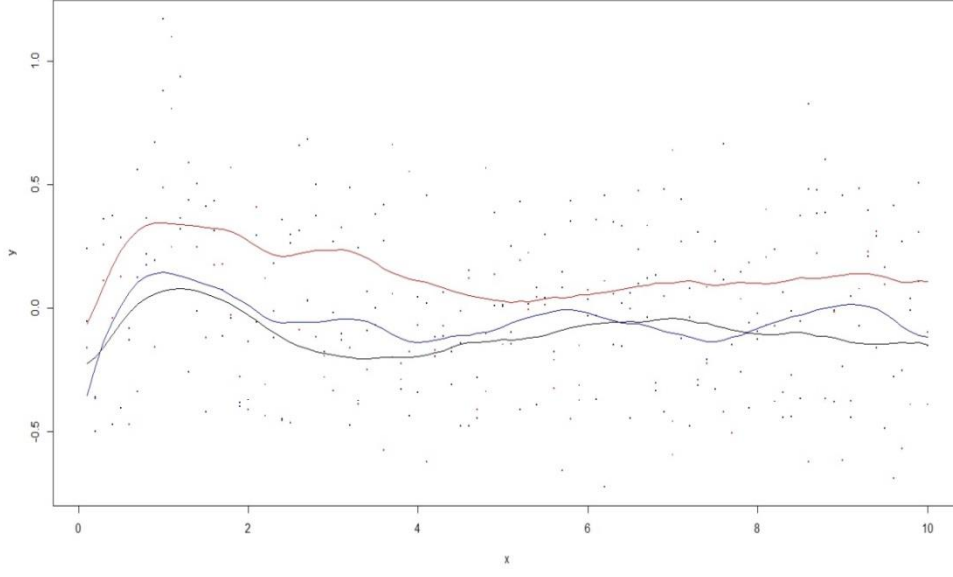


Şekil 5.19 Veri Seti-4 için $p = 1$ ve Gauss çekirdeği kullanıldığında CV kriteri ile seçilen $h = 0.6$ için uyum eğrisi

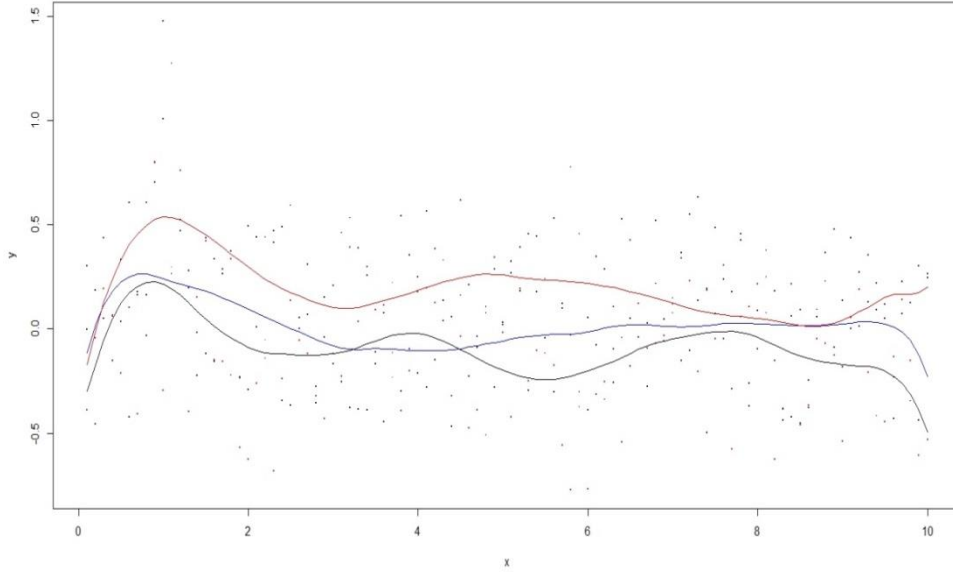


Şekil 5.20 Veri Seti-4 için $p = 3$ ve Gauss çekirdeği kullanıldığında CV kriteri ile seçilen $h = 1.2$ için uyum eğrisi

Şekil 5.21, Veri Seti-4 için $p = 1$ ve Epanechnikov çekirdeği kullanıldığında CV kriteri ile seçilen $h = 1.2$ için uyum eğrilerini ve Şekil 5.22, $p = 3$ ve Epanechnikov çekirdeği kullanıldığında CV kriteri ile seçilen $h = 2.5$ için uyum eğrilerini göstermektedir.



Şekil 5.21 Veri Seti-4 için $p = 1$ ve Epanechnikov çekirdeği kullanıldığında CV kriteri ile seçilen $h = 1.2$ için uyum eğrisi



Şekil 5.22 Veri Seti-4 için $p = 3$ ve Epanechnikov çekirdeği kullanıldığında CV kriteri ile seçilen $h = 2.5$ için uyum eğrisi

Eğriselliği çok az olan sıfır etrafında çok az bir eğrisellik gösteren böyle bir veri setinde CV ve GCV kriterleri ile band genişliğinin seçiminde çizelgeler incelendiğinde skor değerleri birbirine çok yakındır ve band genişliği değerini seçmekte güçlük çekilmektedir. Çizelgelerde skorların değerlerinin minimum olduğu noktalara karşılık gelen h değerleri band genişliği olarak seçilmiştir. Ayrıca grafiksel gösterimler incelendiğinde bazı noktalarda eğrilerin birbirini kestiği görülmektedir. Bunun nedeni yanıt değişkeninin değerlerinin küçük olması ve rassalığın etkisinden kaynaklanmaktadır.

Durum-2 $g_4(x)$, fonksiyonu çıktı değişkeninin üretilmesinde kullanılan bir fonksiyon olmak üzere aşağıdaki gibi tanımlansın:

$$g_4(x) = 0.3 \exp(-4(x + 1)^2) + 0.7 \exp(-16(x - 1)^2)$$

$[-2, 2]$ aralığında $x_i = -2 + 0.04i$, $i = 0, 1, 2, \dots, 100$ noktaları için;

$$y_i = g_4(x_i) + \text{rand}[0, 0.1^2]$$

$$\sigma_i = \left| \frac{1}{3} g_4(x_i) + \text{rand}[-0.25, 0.25] \right|$$

olarak hesaplanırlar. Burada $\text{rand}[0, 0.1^2]$, herbir i için ortalaması 0 ve standart sapması 0.1 olan normal dağılımdan bağımsız olarak türetilen bir rassal sayıyı göstermektedir. Saçılım değerleri için $\text{rand}[-0.25, 0.25]$ ise $[-0.25, 0.25]$ aralığında düzgün dağılımdan rassal olarak türetilen sayıyı göstermektedir. Gözlenen bulanık çıktıların simetrik üçgen bulanık sayılar olduğu varsayılarak, bu çıktılara ait gösterim aşağıdaki gibi tanımlanır.

$$Y_i = (l_{y_i}, c_{y_i}, u_{y_i})_T = (y_i - \sigma_i, y_i, y_i + \sigma_i)_T \quad i = 0, 1, \dots, 100$$

Veri Seti-4 Durum-2 için band genişliğinin seçimi Çizelge 5.17 ve Çizelge 5.18'te görülmektedir.

Çizelge 5.17 Veri Seti-4 (Durum-2) için Gauss çekirdeği kullanıldığında $p = 1$ ve $p = 3$ için CV ve GCV kriterleri ile band genişliği seçimi

Veri Seti-4 Durum-2					
Deg = 1 Gauss			Deg = 3 Gauss		
h	CV	GCV	h	CV	GCV
0.02	0.07179044	0.07105994	0.08	0.07055741	0.06216172
0.03	0.0697142	0.06694967	0.10	0.06307588	0.05886437
0.05	0.06097823	0.05981956	0.12	0.05890721	0.05660605
0.08	0.05566826	0.05502312	0.15**	0.05698954	0.05488938
0.09**	0.05548016	0.05432859	0.18	0.05644749	0.05511901
0.10*	0.05461840	0.05477093	0.20*	0.05609307	0.05620740
0.12	0.05571423	0.05595382	0.22	0.05678865	0.05660873
0.15	0.05730919	0.05780734	0.24	0.05737172	0.05734757
0.18	0.06157438	0.06094869	0.27	0.06033252	0.06005743
0.20	0.06322493	0.06374143	0.30	0.06213535	0.0623531

Çizelge 5.18 Veri Seti-4 (Durum-2) için Epanechnikov çekirdeği kullanıldığında $p = 1$ ve $p = 3$ için CV ve GCV kriterleri ile band genişliği seçimi

Veri Seti-4 Durum-2					
Deg = 1 Epanechnikov			Deg = 3 Epanechnikov		
h	CV	GCV	h	CV	GCV
0.10	0.06357496	0.06076148	0.30	0.06317951	0.05673085
0.12	0.06212045	0.05909174	0.32	0.06271122	0.05690012
0.14	0.05776314	0.05707719	0.35	0.05909533	0.05559540
0.16	0.05708263	0.05553227	0.38	0.05803051	0.05522974
0.17	0.05574823	0.05565975	0.40	0.05775958	0.05499882
0.18**	0.0552960	0.05415368	0.42**	0.05720430	0.05434230
0.19*	0.05493639	0.05438071	0.45	0.05707262	0.05544035
0.20	0.05537448	0.05473639	0.46	0.05673719	0.05515394
0.21	0.05515374	0.05466103	0.48*	0.05671228	0.05544200
0.22	0.05505668	0.05440032	0.50	0.05715588	0.05503750

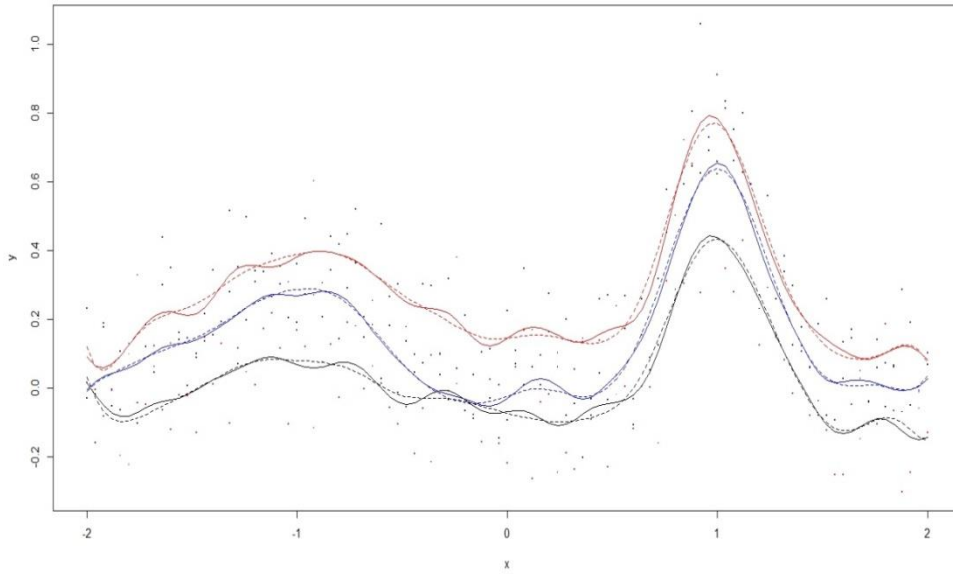
Çizelge 5.19 Veri Seti-4 (Durum-2) için farklı çekirdekler kullanılarak $p = 1$ ve $p = 3$ için CV ve GCV kriterleri ile seçile band genişliklerine göre model performansları

Yöntem	Seçim Kriteri	Gauss Çekirdeği	ASE	Epanechnikov Çekirdeği	ASE
LLS	CV	0.10	0.03747611	0.19	0.03757792
	GCV	0.09	0.03572922	0.18	0.03709936
LCS	CV	0.20	0.04061259	0.48	0.04053092
	GCV	0.15	0.03668817	0.42	0.03890653

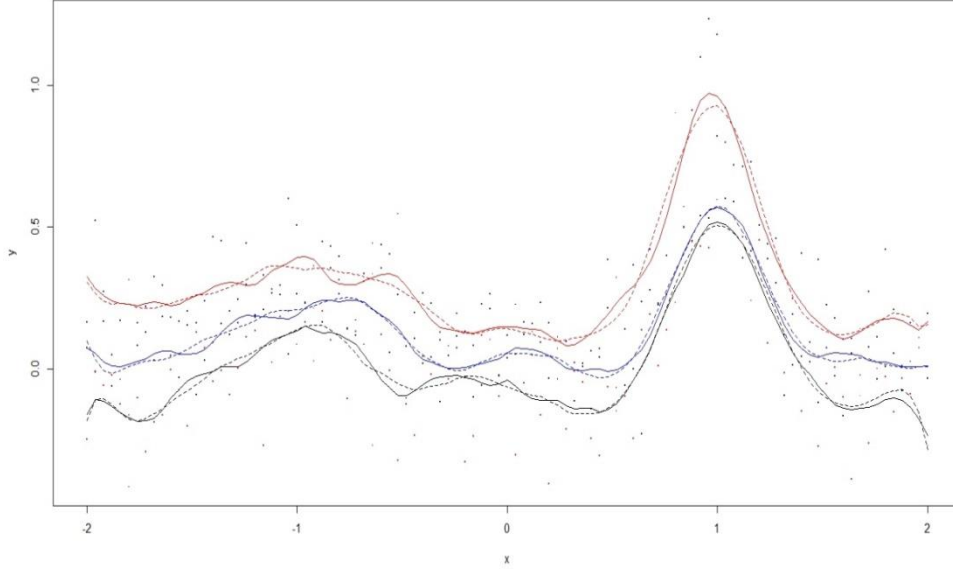
Veri Seti-4 Durum-2 için oluşturulmuş olan Çizelge 5.19 model performansları tablosu dikkate alındığında *ASE* değerlerinin oldukça küçük olduğu gözlenmekte ve her iki model türü kullanılarak veri setinin iyi bir şekilde açıklanabileceği yorumu yapılabilmektedir.

Veri Seti-4 Durum-2 için Gauss çekirdeği kullanılarak *CV* kriteri ile band genişliklerinin hesaplanması sonucunda bulanık yerel doğrusal ve bulanık yerel kübik modeller için Şekil 5.23’de ve Epanechnikov çekirdeği kullanılarak yapılan uyumlamalarda Şekil 5.24’de görülmektedir.

Şekil 5.23 ve Şekil 5.24 incelendiğinde eğriselliği fazla olan böyle bir veri setinde bulanık yerel kübik model kurmak daha pürüzsüz bir eğri ile verileri tanımlamaya öncülük ederken, bulanık yerel doğrusal model çok daha dalgalı bir yapı sergileyerek veri setini tanımlamaktadır.



Şekil 5.23 Veri Seti-4 Durum-2 veri seti için Gauss çekirdeği kullanılarak *CV* kriteri ile seçilen band genişliği kullanılarak bulanık yerel doğrusal ve bulanık yerel kübik modeller için uyum eğrileri



Şekil 5.24 Veri Seti-4 Durum-2 için Epanechnikov çekirdeği kullanılarak CV kriteri ile seçilen band genişliği kullanılarak bulanık yerel doğrusal ve bulanık yerel kübik modeller için uyum eğrileri

6. GERÇEK VERİ SETLERİ ÜZERİNDE UYGULAMALAR

Çalışmanın bu bölümünde, bir önceki bölümde yapılan simülasyon deneylerine ilave olarak iki gerçek veri seti üzerinde uygulama gerçekleştirilmiştir. Lee ve Tanaka (1999) çalışmalarında yer alan girdi değişkeninin kesin değerli ve çıktı değişkeninin bulanık değerli olduğu veri seti dikkate alınmıştır. D'Urso ve Gastaldi (2002) çalışmasındaki veri seti kullanılarak, analizler gerçek veri setleri üzerinde de gerçekleştirilmiştir.

6.1. Lee ve Tanaka (1999) Veri Seti İçin Parametrik Olmayan Bulanık Yerel Polinomial Regresyon Modeli Tahmini

Bu uygulama çalışmasında, Lee ve Tanaka'nın (1999) çalışmalarında kullanmış oldukları gerçek veri seti dikkate alınmıştır. Girdi değişkeni x bir bileyici dişlinin besleme hızıdır ve çıktı değişkeni çalışma yüzeyinin pürüzlülüğünü gösteren simetrik üçgen bulanık sayıdır. Uygulamada, simülasyon deneylerine benzer şekilde CV ve GCV kriteri kullanılarak optimum band genişlikleri hesaplanmış, bulanık yerel doğrusal ve bulanık yerel kübik modellerin performanslarını karşılaştırmak amacıyla ASE değerleri hesaplanmıştır. Çizelge 6.1, veri setine ait değerleri göstermektedir.

Çizelge 6.1 Lee ve Tanaka (1999) Veri Seti

No	x	y1 (sol)	y2 (merkez)	y3(sağ)	e (saçılım)
1	0,75	0,27	0,290	0,310	0,020
2	1,00	0,19	0,240	0,290	0,050
3	1,25	0,20	0,240	0,280	0,040
4	1,50	0,245	0,280	0,315	0,035
5	1,75	0,23	0,280	0,330	0,050
6	2,00	0,20	0,235	0,270	0,035
7	2,25	0,17	0,230	0,290	0,060
8	2,50	0,20	0,330	0,460	0,130
9	2,75	0,20	0,275	0,350	0,075
10	3,00	0,22	0,300	0,380	0,080
11	3,25	0,26	0,335	0,410	0,075
12	3,50	0,22	0,275	0,330	0,055
13	3,75	0,3	0,4	0,5	0,1
14	4,00	0,35	0,455	0,56	0,105
15	4,25	0,34	0,42	0,5	0,08
16	4,50	0,37	0,485	0,6	0,115
17	4,75	0,4	0,5	0,6	0,1
18	5,00	0,41	0,65	0,89	0,24
19	5,25	0,48	0,64	0,8	0,16

Çizelge 6.2 veri seti için Gauss çekirdeği kullanıldığından bulanık yerel doğrusal ve bulanık yerel kübik modeller için band genişliği seçiminde kullanılan seçim kriterlerine ait skor değerlerini ve Çizelge 6.3, Epanechnikov çekirdeği kullanıldığından bulanık yerel doğrusal ve bulanık yerel kübik modeller için band genişliği seçiminde kullanılan seçim kriterlerine ait skor değerlerini göstermektedir.

Çizelge 6.2 Lee ve Tanaka (1999) veri seti için Gauss çekirdeği kullanıldığında $p = 1$ ve $p = 3$ için CV ve GCV kriterleri ile band genişliği seçimi

Lee ve Tanaka (1999) Veri Seti					
<i>Deg = 1 Gauss</i>			<i>Deg = 3 Gauss</i>		
<i>h</i>	<i>CV</i>	<i>GCV</i>	<i>h</i>	<i>CV</i>	<i>GCV</i>
0.2	0.01817601	0.01105804	1.0	0.003817402	0.009303734
0.5	0.01010767	0.009022893	1.5	0.001863774	0.008542141
0.7**	0.008946176	0.008671683	2.0	0.001373535	0.008289765
0.77*	0.008889275	0.008704399	2.5	0.001160253	0.008176546
0.9	0.009025928	0.008915462	3.0	0.001047486	0.008115857
1.0	0.009268489	0.00916835	3.5**	0.0009805849	0.008079474
1.1	0.009587743	0.009471037	4.0	0.0009376263	0.008055921
1.3	0.009762764	0.009640122	4.5*	0.0009083925	0.008039792
1.5	0.009956405	0.098021680	5.0	0.000887594	0.008028261

Çizelge 6.3 Lee ve Tanaka (1999) veri seti için Epanechnikov çekirdeği kullanıldığında $p = 1$ ve $p = 3$ için CV ve GCV kriterleri ile band genişliği seçimi

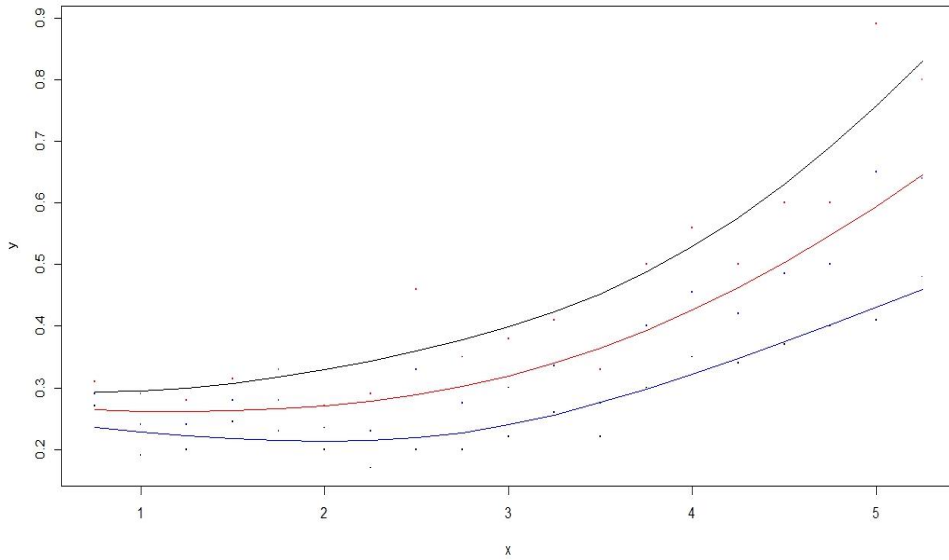
Lee ve Tanaka (1999) Veri Seti					
<i>Deg = 1 Epanechnikov</i>			<i>Deg = 3 Epanechnikov</i>		
<i>h</i>	<i>CV</i>	<i>GCV</i>	<i>h</i>	<i>CV</i>	<i>GCV</i>
0.5	-	0.01171968	1.0	0.1555907	0.01311383
0.8	0.0139368	0.009758859	2.0	0.0319741	0.01047545
1.0	0.0118578	0.009148114	3.0	0.01110802	0.008732646
1.2	0.01054138	0.008983597	4.0	0.009855845	0.008370318
1.5	0.009135601	0.008597761	5.0	0.008760976	0.008120107
1.6	0.00876908	0.008498644	6.0**	0.008598969	0.008062371
1.7**	0.008643823	0.008479249	7.0*	0.008534242	0.008036422
1.75*	0.008612166	0.008480432	8.0	0.008498617	0.008021437
1.8	0.008689818	0.008576969	10.0	0.008461264	0.008005191
2.0	0.008971018	0.008899979	12.0	0.008426947	0.008003856

Çizelge 6.4'te geliştirilen CV ve GCV kriterleri ve Wang ve ark. (2007) çalışmasında kullanılan CV kriteri ile seçilen band genişlikleri için modellerin performans değerleri yer almaktadır.

Çizelge 6.4 Lee ve Tanaka (1999) veri seti için farklı çekirdekler kullanılarak $p = 1$ ve $p = 3$ için CV ve GCV kriterleri ve Wang ve ark. (2007) kriteri ile seçilen band genişliklerine göre model performansları

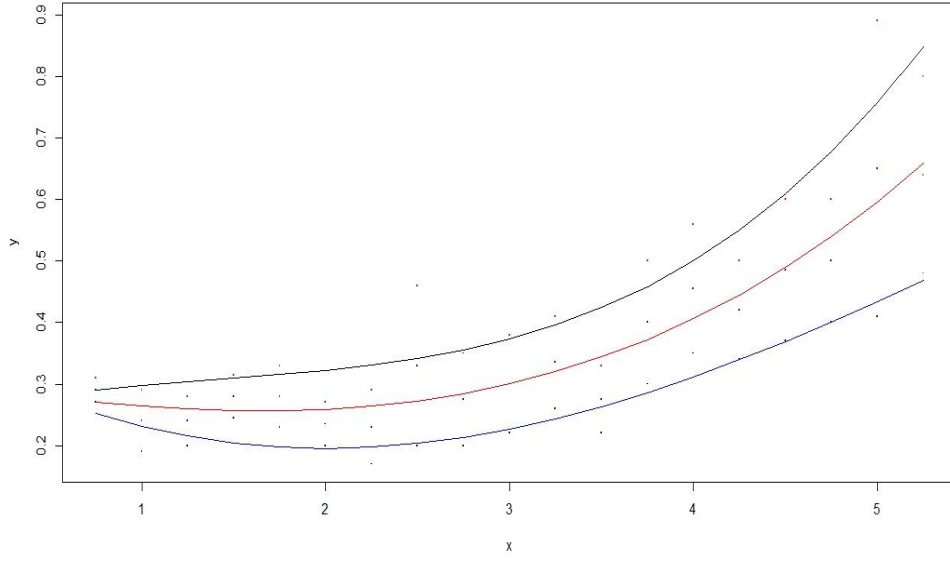
Yöntem	Seçim Kriteri	Gauss Çekirdeği	ASE	Epanechnikov Çekirdeği	ASE
LLS	<i>CV</i>	0.77	0.005891200	1.75	0.007683992
	<i>GCV</i>	0.70	0.004779285	1.70	0.007253257
LCS	<i>CV</i>	4.50	6.297887e-10	7.00	3.789276e-11
	<i>GCV</i>	3.50	5.285123e-09	6.00	4.386862e-10
Wang ve ark. (2007)	<i>CV</i>	0.78	0.006049552	-	-

Şekil 6.1’de bulanık yerel doğrusal modelde Gauss çekirdeği kullanılarak CV kriteri ile seçilen band genişliğine göre çizilen uyum eğrileri görülmektedir.



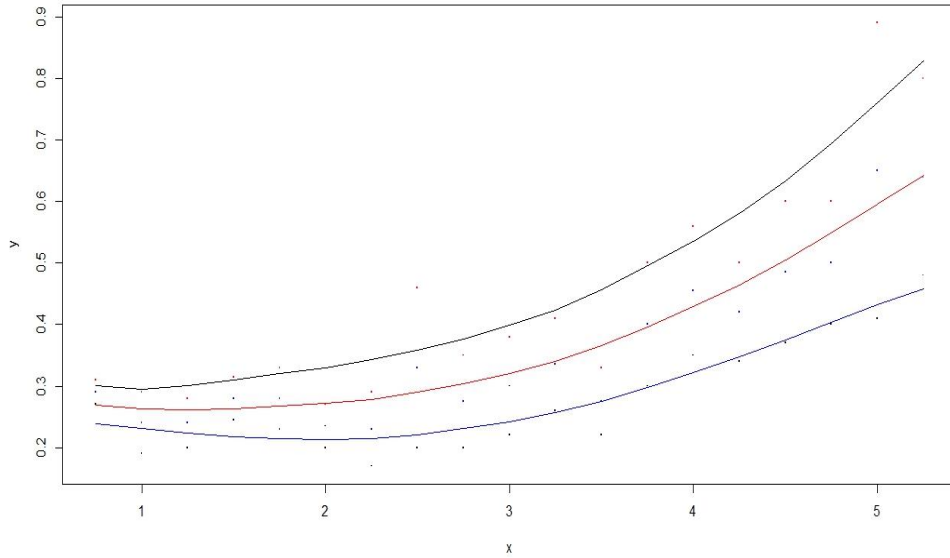
Şekil 6.1 Lee ve Tanaka (1999) veri seti için $p = 1$ ve Gauss çekirdeği kullanıldığında CV kriteri ile seçilen band genişliği $h = 0.77$ için uyum eğrileri

Şekil 6.2’de bulanık yerel kübik modelde Gauss çekirdeği kullanılarak CV kriteri ile seçilen band genişliğine göre çizilen uyum eğrileri görülmektedir.

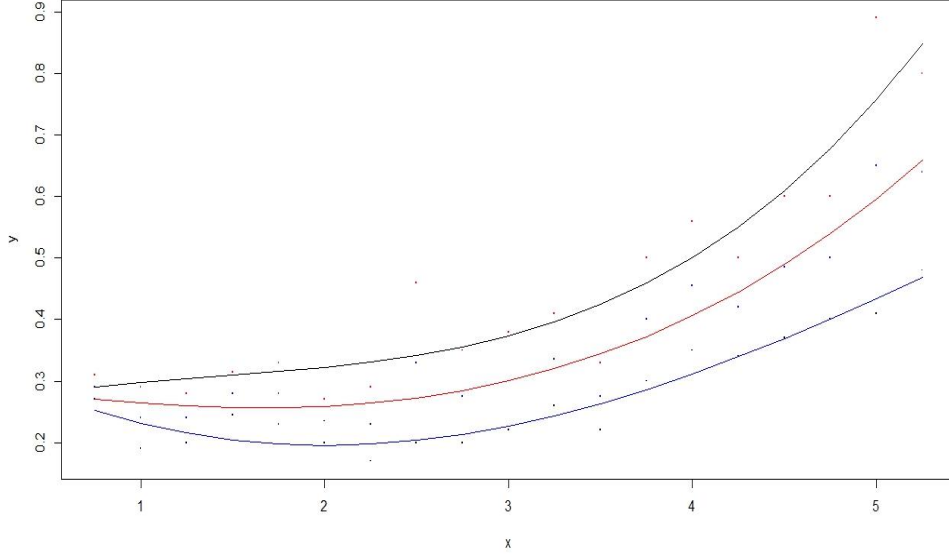


Şekil 6.2 Lee ve Tanaka (1999) veri seti için $p = 3$ ve Gauss çekirdeği kullanıldığında CV kriteri ile seçilen band genişliği $h = 4.5$ için uyum eğrileri

Şekil 6.3 ve Şekil 6.4 benzer olarak bulanık yerel doğrusal ve bulanık yerel kübik modeller için Epanechnikov çekirdeği kullanılarak CV kriteri ile seçilen band genişliğine göre çizilen uyum eğrilerini göstermektedir.



Şekil 6.3 Lee ve Tanaka (1999) veri seti için $p = 1$ ve Epanechnikov çekirdeği kullanıldığında CV kriteri ile seçilen band genişliği $h = 1.75$ için uyum eğrileri



Şekil 6.4 Lee ve Tanaka (1999) veri seti için $p = 3$ ve Epanechnikov çekirdeği kullanıldığında CV kriteri ile seçilen band genişliği $h = 7.0$ için uyum eğrileri

Gerçek veri seti üzerinde yapılan uygulamada, veri setinin doğrusal bir yapı sergilemesi bulanık yerel kübik modelde band genişliği seçiminde problem yaratmıştır. Band genişliğinin seçiminde CV ve GCV skorlarındaki değişimi gösteren çizelgelerde, skorlardaki düşüşün azaldığı noktalar dikkate alınarak onlara karşılık gelen h değerleri uygun band genişliği olarak seçilmiştir. Çizelge 6.4 incelendiğinde bulanık yerel kübik modelde band genişliği, bulanık yerel doğrusal modele göre çok daha büyük seçilmesinin yanında model performans kriteri ASE değerinin çok küçüldüğü de dikkate alınmalıdır.

6.2. D'Urso ve Gastaldi (2002) Veri Seti İçin Parametrik Olmayan Bulanık Yerel Polinomiyal Regresyon Modeli Tahmini

D'Urso ve Gastaldi (2002) çalışmalarında yer alan veri seti, bir http sunucusunda internet üzerinden sanal alışveriş merkezi çalıştırmak amacıyla geliştirilmiş yazılım sisteminin, geliştirilmesi esnasında üretilmiştir. Gözlenen çıktılar, geliştirilmekte olan sistemin performansının belirlenmesi için 30 aylık ortalama global skorlardır ve herbir aydaki skor simetrik olmayan üçgen bulanık

sayılar şeklinde ifade edilmiştir. Skor ve ay indeksi arasındaki regresyon ilişkisinin kurulmasında kullanılan veri seti Çizelge 6.5'te görülmektedir.

Çizelge 6.5 D'Urso & Gastaldi (2002) Veri Seti

x	y1 (sol)	y2(merkez)	y3 (sağ)	left	right
1	4	6	14	2	8
2	2	5	11	3	6
3	3	8	16	5	8
4	1	10	19	9	9
5	4	13	21	9	8
6	14	19	26	5	7
7	12	20	25	8	5
8	14	23	28	9	5
9	20	25	32	5	7
10	21	27	35	6	8
11	19	26	32	7	6
12	21	25	31	4	6
13	2	5	8	3	3
14	3	8	11	5	3
15	2	9	17	7	8
16	6	10	13	4	3
17	6	12	20	6	8
18	10	13	22	3	9
19	14	20	27	6	7
20	21	28	37	7	9
21	21	29	38	8	9
22	26	30	37	4	7
23	33	35	38	2	3
24	41	44	49	3	5
25	39	45	51	6	6
26	38	43	47	5	4
27	44	47	56	3	9
28	44	48	53	4	5
29	44	50	54	6	4
30	44	49	53	5	4

D'Urso ve Gastaldi (2002) veri seti kullanılarak *CV* ve *GCV* kriterleri yardımıyla Gauss ve Epanechnikov çekirdekleri için band genişlikleri seçimlerini gösteren bilgiler Çizelge 6.6 ve Çizelge 6.7'de verilmiştir.

Çizelge 6.6 D'Urso&Gastaldi (2002) veri seti için Gauss çekirdeği kullanıldığında $p = 1$ ve $p = 3$ için CV ve GCV kriterleri ile band genişliği seçimi

D'Urso&Gastaldi (2002) Veri Seti					
Deg = 1 Gauss			Deg = 3 Gauss		
h	CV	GCV	h	CV	GCV
0.4	44.74988	47.46785	0.7	78.59073	59.89703
0.66*	44.72523	47.35591	0.9	76.08064	57.92012
0.73**	44.74071	47.32062	1.2	63.44823	54.65933
0.8	44.80903	47.36256	1.5	52.34280	53.55255
1.0	45.46510	48.15014	1.6**	50.56276	53.51068
1.1	46.08339	48.91188	1.7	49.43173	53.57089
1.2	46.88957	49.87881	1.8	48.75178	53.70733
1.3	47.87302	51.02546	1.9	48.39444	53.90399
1.4	49.02478	52.33577	2.0*	48.28139	54.15439
1.5	50.33794	53.80105	2.5	48.19633	54.06890

Çizelge 6.7 D'Urso&Gastaldi (2002) veri seti için Epanechnikov çekirdeği kullanıldığında $p = 1$ ve $p = 3$ için CV ve GCV kriterleri ile band genişliği seçimi

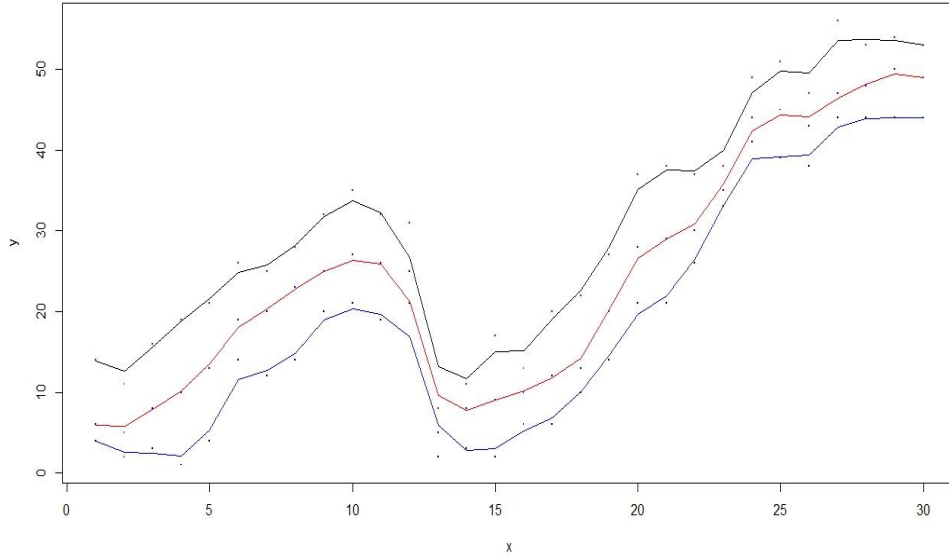
D'Urso&Gastaldi (2002) Veri Seti					
Deg = 1 Epanechnikov			Deg = 3 Epanechnikov		
h	CV	GCV	h	CV	GCV
2.01*	44.73833	47.36374	3.1**	602.0765	55.47809
2.03	44.75043	47.23899	3.3	51.06489	56.40593
2.05**	45.03244	47.19680	3.5	157.5381	57.10375
2.5	47.70257	50.17009	3.7	47.31681	57.57425
3.0	49.75383	52.50683	3.8	164.9059	57.75843
3.5	52.91063	56.91159	3.9*	55.17545	57.91867
3.8	54.63016	58.94914	4.0	76.90517	58.05979
3.9	55.72130	59.10269	4.2	76.41985	55.83657
4.0	56.20194	59.86955	4.5	76.90517	56.48394
4.2	58.36428	61.25864	5.0	77.54478	57.30578

D'Urso ve Gastaldi (2002) veri seti kullanılarak CV ve GCV kriterleri yardımıyla Gauss ve Epanechnikov çekirdekleri için band genişlikleri seçimleri için bulanık yerel doğrusal, bulanık yerel kübik ve Wang ve ark. (2007) çalışmasındaki modellere ait performans değerleri Çizelge 6.8'de görülmektedir.

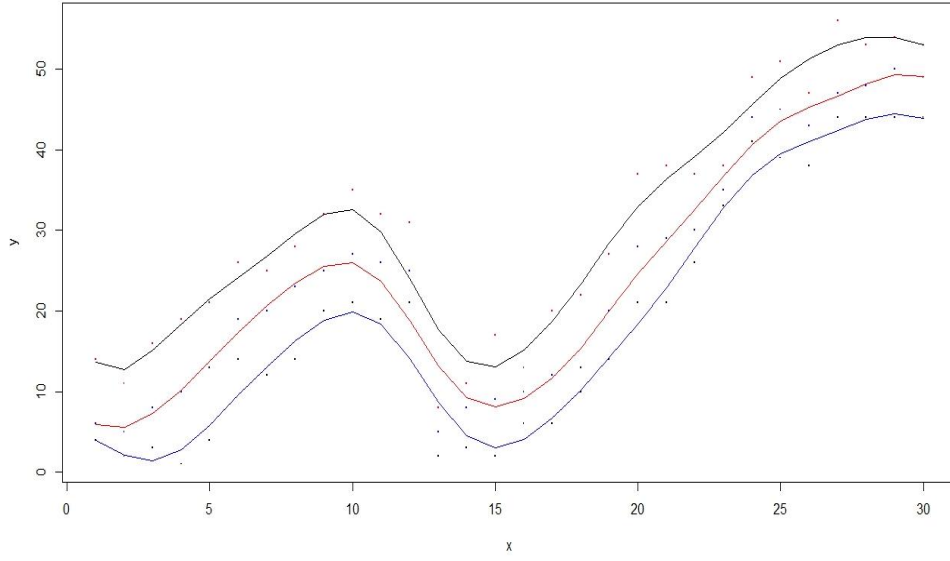
Çizelge 6.8 D'Urso&Gastaldi (2002) veri seti için farklı çekirdekler kullanılarak $p = 1$ ve $p = 3$ için geliştirilen CV ve GCV kriterleri ve Wang ve ark. (2007)'nin kullandıkları CV kriteri ile seçilen band genişliklerine göre model performansları

Yöntem	Seçim Kriteri	Gauss Çekirdeği	ASE	Epanechnikov Çekirdeği	ASE
LLS	<i>CV</i>	0.66	0.007831292	2.01	0.02059755
	<i>GCV</i>	0.73	0.00795834	2.05	0.01253825
LCS	<i>CV</i>	2.0	0.10541000	3.9	0.11336700
	<i>GCV</i>	1.6	0.005260489	3.1	0.06919009
Wang ve ark. (2007)	<i>CV</i>	0.63	0.007337764	-	-

Çizelge 6.6 ve Çizelge 6.7'de seçilen band genişlikleri için Gauss çekirdeği kullanıldığında uyum eğrileri Şekil 6.5 ve Şekil 6.6'da görülmektedir.

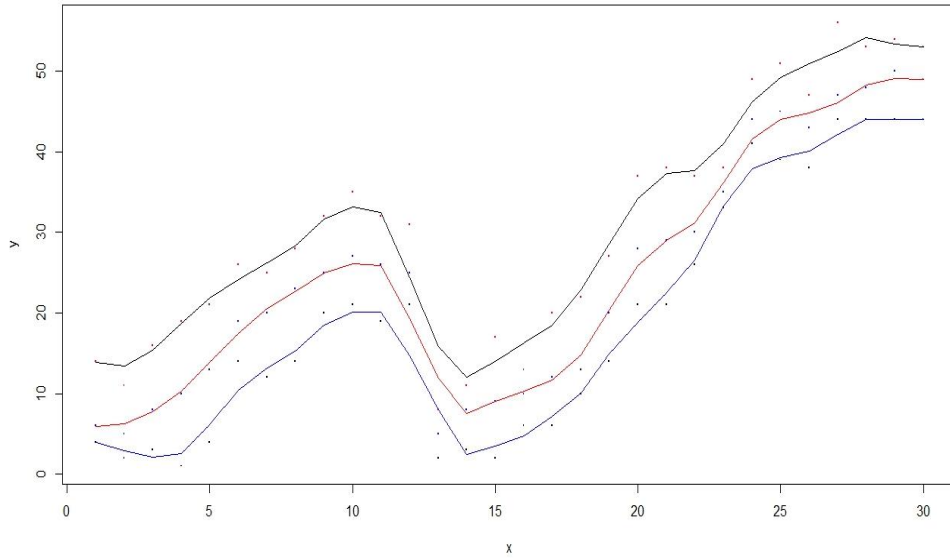


Şekil 6.5 D'Urso&Gastaldi (2002) veri seti için $p = 1$ ve Gauss çekirdeği kullanıldığında CV kriteri ile seçilen $h = 0.66$ için uyum eğrileri

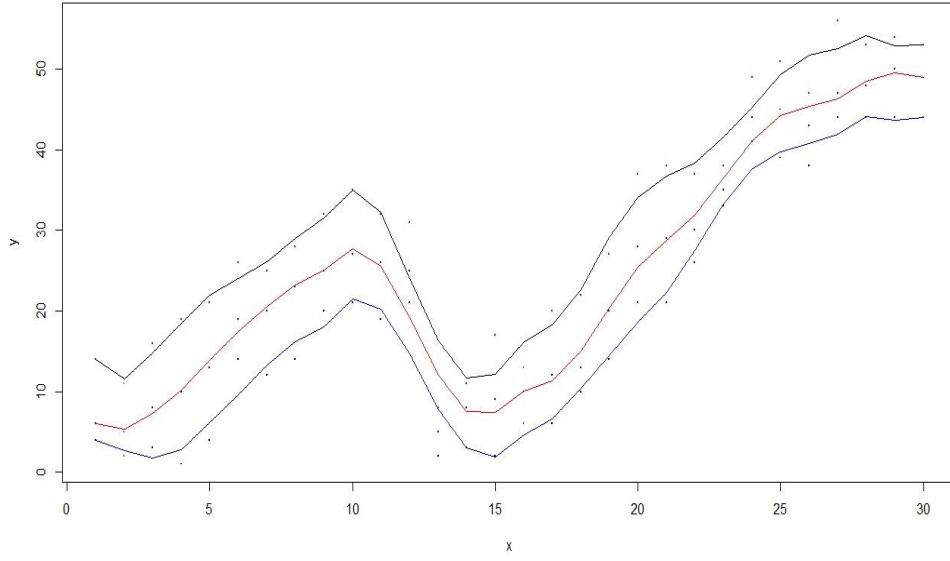


Şekil 6.6 D'Urso&Gastaldi (2002) veri seti için $p = 3$ ve Gauss çekirdeği kullanıldığında CV kriteri ile seçilen $h = 2.0$ için uyum eğrileri

Çizelge 6.6 ve Çizelge 6.7'de seçilen band genişlikleri için Epanechnikov çekirdeği kullanıldığında uyum eğrileri Şekil 6.7 ve Şekil 6.8'de görülmektedir.



Şekil 6.7 D'Urso&Gastaldi (2002) veri seti için $p = 1$ ve Epanechnikov çekirdeği kullanıldığında CV kriteri ile seçilen $h = 2.01$ için uyum eğrileri



Şekil 6.8 D'Urso&Gastaldi veri seti için $p = 3$ ve Epanechnikov çekirdeği kullanıldığında CV kriteri ile seçilen $h = 3.9$ için uyum eğrileri

D'Urso ve Gastaldi (2002) veri setine ait şekillerde görüldüğü üzere, veri seti oldukça dalgalı bir yapıya sahip olması ve veri sayısının az olmasından dolayı genel olarak CV ve GCV skorlarının değerleri yüksek çıkmıştır. Ayrıca Epanechnikov çekirdeği ile bulanık yerel kübik model için CV kriteri ile band genişliği değerinin seçiminde skor değerinde oldukça dalgalı iniş ve çıkışlar gözlemlenmiştir.

7. SONUÇ ve ÖNERİLER

Parametrik olmayan regresyon modelleri bağımlı değişken ile bağımsız değişken arasındaki fonksiyonel ilişkinin bilinmediği durumlarda oldukça kullanışlıdır. Elbette bu durum kesin veri setleri için önemli olduğu kadar bulanık veri setleri için de önemlidir. Doğasında bulanıklığı barındıran verilerin analizinde modellerin etkinliğini gözlemlemek ve kesin teoride yer alan her türlü yöntemin bulanık yapıda da ifade edilmesi gerekmektedir. Çünkü doğada pek çok veri özünde bulanıklığı içermekte ve bulanık veri setleri, kesin veri setlerine göre daha genel olarak ifade edilebilmektedirler.

Parametrik olmayan regresyon modellerinden, çekirdek düzeltme ve yerel polinomiyal regresyon modellerinin bulanık halleri incelenmiştir. Parametrik olmayan regresyon modellerinde band genişliği seçimi oldukça önemli bir konudur. Yerel polinomiyal modellerde, yerel polinomun derecesi ile ilişkili olarak band genişliğinin (h) seçimi gerçekleştirilir. Bu modellerde polinomun derecesi arttıkça yerel band genişliği de artar. Literatürde band genişliği seçimi için çeşitli pratik yöntemler bulunmaktadır (Fan ve Gijbel 1996; Wood 2006).

Çalışmanın 2. Bölüm'ünde band genişliği seçim yöntemlerinin kesin teoride ifadesi ve 4. Bölüm'de bulanık teorideki ifadesine yer verilmiştir. Literatürde daha önce yapılmış çalışmalardan Cheng ve Lee (1999), k -en yakın komşuluk ve çekirdek düzeltme tekniklerini karşılaştırmıştır. Çekirdek düzeltme tekniğinin, k -en yakın komşuluk düzeltme tekniğine göre daha iyi sonuçlar verdiği sonucuna ulaşmışlardır. Wang ve ark. (2007) çalışmalarında bu iki yöntemin yanı sıra bulanık yerel doğrusal düzeltme tekniğini dikkate almışlardır. Bulanık yerel doğrusal düzeltme tekniğinin diğer iki tekniğe göre daha iyi sonuçlar verdiği sonucuna ulaşılmıştır. Çalışmada band genişliği seçim kriterlerinden çapraz geçerlilik kriterinin yalnızca tanımlanmasına yer verilmiş ayrıntılı olarak ifade edilmemiştir. Yerel polinomiyal düzeltme tekniğinin bulanık yapıda ifadesini ve band genişliği seçim kriterlerinden çapraz geçerlilik ve genelleştirilmiş çapraz geçerlilik kriterlerinin ifadesi tez çalışması kapsamında ayrıntılı olarak incelenmiştir.

Tez çalışmasının amacı doğrultusunda parametrik olmayan bulanık regresyon modellerinden, çekirdek düzeltme ve yerel polinomiyal regresyon modelleri için çeşitli veri setleri kullanılarak inceleme yapılmıştır. Genel teknik olarak ağırlıklı en küçük kareler metodu, bulanık uzaklık olarak ise Diamond uzaklığı kullanılmıştır. Yapılan simülasyon çalışmaları ile gerçek veri setleri üzerinde çekirdek düzeltme, bulanık yerel doğrusal ve bulanık yerel kübik regresyon modelleri için 4. Bölüm'de geliştirilen çapraz geçerlilik (*CV*) ve genelleştirilmiş çapraz geçerlilik (*GCV*) kriterleri yardımıyla uygun band genişlikleri seçilmiştir. Yerel polinomiyal regresyon modellerinde, çekirdek fonksiyonunun seçiminin band genişliği seçimine göre daha az önem arz etmesinden dolayı bu çalışmada iki çekirdek fonksiyonu Gauss ve Epanechnikov çekirdek fonksiyonları modellerin uyumlanmasında kullanılmıştır. Tüm bu kriterler göz önüne alınarak çekirdek düzeltme, bulanık yerel doğrusal ve bulanık yerel kübik regresyon modelleri *ASE* değerleri hesaplanarak karşılaştırılmıştır.

Veri seti olarak girdi değişkenin kesin değerlerle ifade edildiği ve çıktı değişkeninin üçgen bulanık sayı olduğu durum dikkate alınmıştır.

Kesin teoride yapılan karşılaştırmalar dikkate alınarak *k*-en yakın komşuluk düzeltme tekniğine uygulamalarda yer verilmemiştir. Wang ve ark. (2007) yapmış oldukları çalışmalarda olduğu gibi 5.1'deki uygulamalarda bulanık yerel polinomiyal modellerin, çekirdek düzeltme tekniğine göre daha pürüzsüz eğri tahmini gerçekleştirdiği gözlemlendiğinden sonraki uygulamalarda bu tekniğe yer verilmemiştir.

Uygulama çalışmalarında, literatürde de değinildiği gibi bulanık yerel polinomiyal modelde polinomun derecesi artarken band genişliğinin de arttığı görülmüştür. Bu durum işlemsel olarak, yerel uyumlama noktalarının azalmasından dolayı işlem basamaklarını azaltmakta ve zamandan kazanç sağlanmaktadır.

Özellikle daha eğrisel olan veri setlerinde bulanık yerel doğrusal modellerin kullanılmasının çok uygun olmadığı gözlenmiştir. Çünkü bu tür durumlarda band genişliği oldukça küçük seçilmekte ve işlem karmaşasına yol açmaktadır, ayrıca uyum eğrisi oldukça dalgalı bir yapıya sahip olmaktadır.

Doğrusal bir şekle sahip olan veri setlerinde bulanık yerel doğrusal ya da bulanık yerel kübik model kullanılması önemli bir etkiye neden olmamakla birlikte, bu tür veri setlerinde band genişliğinin bulanık yerel kübik model için seçiminde skor değerlerinin band genişliği arttıkça azaldığı gözlemlenmiştir.

Genel olarak uygulamaların çoğunda *GCV* kriteri ile seçilen band genişliğinin, *CV* kriteri ile seçilen band genişliğine göre her iki model için çoğu zaman daha küçük olduğu görülmüştür.

Modellerde kullanılan çekirdek fonksiyonlarından Epanechnikov çekirdeği kullanılarak seçilen band genişliği değeri, Gauss çekirdeği kullanılarak yapılan seçime göre her zaman daha büyüktür.

Uygulamaların pek çoğunda bulanık yerel kübik modele ait performans değerleri bir miktar daha yüksek olmasına rağmen, özellikle eğriselliği fazla olan modelleri daha pürüzsüz bir şekilde ifade etme yeteneği ve band genişliği değerinin bulanık yerel doğrusal modele göre daha yüksek seçilmesi ile işlem basamaklarını azalttığı göz önüne alındığında, performans değerindeki çok az bir yüksekliğin gözardı edilebileceği sonucuna varılabilir.

Bu çalışmada, farklı veri setleri üzerinde uygulamalar gerçekleştirilerek bulanık yerel polinomial modellerin yapıları üzerinde genelleştirme yapmak ve davranışları üzerinde bir sonuca varmak amaçlanmıştır. İleri aşamalarda girdi ve çıktı değişkenlerinin farklı durumları üzerine incelemeler yapılabilir. Ayrıca çok değişkenli veri setleri için bu tür modeller üzerine araştırma yapılması literatüre önemli katkılar sağlayacaktır. Band genişliği seçim kriterlerinden çapraz geçerlilik ve genelleştirilmiş çapraz geçerlilik kriteri bulanık yapıda ifade edilmiştir. Farklı seçim kriterlerinin bulanık ifadeleri gerçekleştirilerek uygulamalarda kullanılabilir. Modellerin, performans karşılaştırma kriterlerinin bulanık yapıdaki ifadeleri geliştirilerek, modeller farklı kriterler bazında kıyaslanabilecektir.

KAYNAKLAR

- Adamo J. M. (1980), "Fuzzy Decision Trees", *Fuzzy Sets and Systems* 4(3), 207-220.
- Bardossy A. (1990), "Note On Fuzzy Regression", *Fuzzy Sets and Systems*, 37, 65-75.
- Chang P.T., Lee, E.S. (1994a), "Ranking of Fuzzy Sets Based on the Concept of Existence", *Computers Math. Applic.*, 27, 1-21.
- Chang P.T., Lee, E.S. (1994b), "Fuzzy Linear Regression with Spreads Unrestricted in Sign", *Computers and Mathematics with Applications*, Vol. 28, pp. 61-70.
- Cheng C.B. (1998), *Nonparametric Fuzzy Regression and It's Application*, PhD Dissertation, Kansas University.
- Cheng C.H. (1999), "Evaluating Weapon Systems Using Ranking Fuzzy Numbers", *Fuzzy Sets And Systems* 107, 25-35.
- Cheng C.B., Lee E.S. (1999), "Nonparametric Fuzzy Regression – k-NN and Kernel Smoothing Techniques", *Computers and Mathematical with Applications* 38, 239-251.
- Cheng C.B., Lee E.S. (2001), "Fuzzy Regression with Radial Basis Function Networks", *Fuzzy Sets and Systems* 119, 291-301.
- Cleveland W.S. (1979), "Robust Locally Weighted Regression and Smoothing Scatterplots", *J. Amer. Statist. Assoc.*, 74, 828-836.
- Cover T. M., Hart P. E. (1967), "Nearest Neighbour Pattern Classification", *IEEE Transaction on Information Theory*, 13, 12-17.
- Craven P., Wahba G. (1979), "Smoothing Noisy Data with Spline Functions", *Numerische Mathematik*, 31:377M403.
- Diamond P. (1988), "Fuzzy Least Squares", *Information Sciences* 46, 141-157.
- Dubois D., Prade H. (1980), *Fuzzy Sets and Systems Theory and Applications*, Academic Press Inc. New York.
- D'Urso P., Gastaldi T. (2002), "An "orderwise" Polynomial Regression Prosedure for Fuzzy Data", *Fuzzy Sets and Systems*, 130, 1-19.

- Fan J. (1992), “Design-Adaptive Nonparametric Regression”, J. Amer. Statist. Assoc., 87, 998-1004.
- Fan J. (1993), “Local Linear Regression Smoothers and Their Minimax Efficiency”, Ann. Statist., 21, 196-216.
- Fan J. and Gijbels I. (1992), “Variable Bandwidth and Local Linear Regression Smoothers”, Ann. Statist., 20, 2008-2036.
- Fan J. and Gijbels I. (1995), “Adaptive Order Polynomial Fitting: Bandwidth Robustification and Bias Reduction”, J. Comp. Graph. Statist., 4, 213-227.
- Fan J. and Gijbels I. (1996), *Local Polynomial Modeling and Its Applications*, Chapman & Hall/CRC.
- Fix E. ve Hodges J.L. (1951), *Discriminatory Analysis, Nonparametric Estimation: Consistency Properties*, Report No. 4, Project No.21-49-004. USAF School of Aviation Medicine, Randolph Field, Texas.
- Fox, J. (2000), *Nonparametric Simple Regression: Smoothing Scatterplots*. Sage Publications, California, 83 s.
- Green P.J., Silverman B. W. (2000), *Nonparametric Regression and Generalized Linear Models*, Chapman & Hall/CRC.
- Hardle W. (1991), *Smoothing Techniques With Implementation in S*, Springer-Verlag, New York.
- Hardle W. (1994), *Applied Nonparametric Regression*, Cambridge University Press, New York.
- Hardle W., Marron J.S. (1985), “Optimal Bandwidth Selection In Nonparametric Regression Function Estimation”, The Annals of Statistics, Vol.13, No.4, 1465-1481.
- Hastie T., Tibshirani R. (1990), *Generalized Additive Models*, Chapman&Hall, London.
- Henderson R. (1916), “Note on Graduation by Adjusted Average”, Transactions of the Actuarial Society of America, 17, 43-48.
- Hong Y. Y., Chao Z. T., Yang M. S. (2004), “A Fuzzy multiple Linear Regression Based Loss Formula in Electric Distribution Systems”, Fuzzy Sets and Systems 142, 293-306.

- Hong H., Song J.-K., Do H. (2001), “Fuzzy Least Squares Linear Regression Analysis Using Shape Preserving Operations”, *Information Sciences* 138, 185-193.
- Hurvich C.M., Simonoff J.S., Tsai C.L. (1998), “Smoothing Parameter Selection in Nonparametric Regression Using An Improved Akaike Information Criterion”, *J.R. Statist. Soc. B.*, 60, 271-293.
- Ishibuchi H., Nii M. (2001), “Fuzzy Regression Using Asymmetric Fuzzy Coefficients and Fuzzified Neural Networks”, *Fuzzy Sets and Systems*, 119, 273-290.
- Ishibuchi H., Tanaka H. (1992), “Fuzzy Regression Analysis Using Neural Networks”, *Fuzzy Sets and Systems* 50, 257-265.
- Kao C., Chyu C.-L., (2002), “A Fuzzy Linear Regression Model with Better Explanatory Power”, *Fuzzy Sets and Systems* 126, 401-409.
- Keele, L. (2008), *Semiparametric Regression for the Social Sciences*, John Wiley&Sons Ltd., Ohio, 213 s.
- Lai Y.J., Hwang C.L. (1992), *Fuzzy Mathematical Programming: Methods and Applications*, Springer-Verlag, Berlin.
- Lee K. H. (2005), *First Course on Fuzzy Theory and Applications*, Springer-Verlag, Berlin Heidelberg New York.
- Lee H., Tanaka H., (1999), “Fuzzy Approximations with Non-Symmetric Fuzzy Parameters in Fuzzy Regression Analysis”, *Journal of the Operational Research Society of Japan* 42, 98-112.
- Lin C.-T., Lee G.C.S. (1996), *Neural Fuzzy Systems A Neuro-Fuzzy Synergism to Intelligent Systems*, Prentice-Hall, Inc.
- Loader C. (1999), *Local Regression and Likelihood*. Springer.
- Loftsgaarden D.O., Quesenbery, G. P. (1965), “A Nonparametric Estimate of a Multivariate Density Function”, *Annals of Mathematical Statistics*, 36, 1049-1051.
- MaCahon C.S., Lee E.S. (1990), “Comparing Fuzzy Numbers: The Proportion of the Optimum Method”, *International Journal of Approximation Reasoning*, 4, 159-181.

- Nadaraya E. A. (1964), "On Estimating Regression", Theory Prob. Appl., 10, 186-190.
- Nasrabadi M., Nasrabadi E. (2004), "A Mathematical-Programming Approach to Fuzzy Linear Regression Analysis", Applied Mathem. and Computation 155, 673-688.
- Parzen E. (1962), "On estimation of a Probability Density Function and Mode", Annals of Mathematical Statistics 33: 1065M1070.
- Priestley M.B., Chao M.T. (1972), "Non-Parametric Function Fitting," Journal of the Royal Statistical Society. Series B (Methodological), Vol.34, No:3:385-392.
- Rice J. (1984), "Bandwidth Choice for Nonparametric Regression", The Annals of Statistics, Vol.12, No.4, 1215-1230.
- Rosenblatt, M. (1956), "Remarks on Some Nonparametric Estimates of a Density Function", Annals of Mathematical Statistics 27: 832M835.
- Ruppert D., Wand M. P. (1994), "Multivariate Weighted Least Squares Regression", Ann. Statist., 22, 1346-1370.
- Ruppert D., Wand M. P., Carroll R.J. (2003), *Semiparametric Regression*, Cambridge University Press.
- Sanchez J. D., Gomez A. T. (2003), "Applications of Fuzzy Regression in Actuarial Analysis", Journal of Risk Insurance 70, 665-699.
- Savic D. A., Pedrycz W. (1991), "Evaluation of Fuzzy Linear Regression Model", Fuzzy Sets and Systems, 39, 51-63.
- Slama R. (2001), "Using Nonparametric and Semiparametric Regression in Epidemiology", England et al., Am. J. Epidemiol, 154.
- Stone C. J. (1977), "Consistent Nonparametric Regression", Ann. Statist., 5, 595-645.
- Stone C. J. (1980), "Optimal Rates of Convergence for Nonparametric Estimators", Ann. Statist., 8, 1348-1360.
- Stone C. J. (1982), "Optimal Global Rates of Convergence for Nonparametric Regression", Ann. Statist., 10, 1040-1053.

- Tanaka H. (1987), "Fuzzy Data Analysis by Possibilistic Linear Models", *Fuzzy Sets and Systems*, 24, 363-375.
- Tanaka H., Uejima S., Asai K. (1982), "Linear Regression Analysis with Fuzzy Model", *IEEE Transactions on Systems, Man, and Cybernetics* 12, 903-907.
- Tanaka H., Hayashi I., Watada J. (1989), "Possibilistic Linear Regression Analysis For Fuzzy Data", *European Journal of Operational Research*, 40, 389-396.
- Tanaka H., Watada J. (1988), "Possibilistic Linear Systems and Their Application to the Linear Regression Model", *Fuzzy Sets and Systems*, 27, 275-289.
- Terano T., Asai K., Sugeno M. (1992), *Fuzzy Systems Theory and Its Applications*, Academic Press Inc.
- Vieu P. (1991), "Nonparametric Regression: Optimal Local Bandwidth Choice", *Journal of the Royal Statistical Society Series B (Methodological)*, Vol.53, No.2, 453-466.
- Wahba G. (1975), "Smoothing Noisy Data By Spline Function", *Numer. Math.*, 24, 383-393.
- Wahba G., Wold S. (1975), "A Completely Automatic French Curve: Fitting Spline Function by Cross-Validation", *Communication in Statistics*, 4, 1-17.
- Wand M.P., Jones M.C. (1995), *Kernel Smoothing*, Chapman and Hall, London.
- Wang N., Zhang W.-X., Mei C.-L. (2007), "Fuzzy Nonparametric Regression Based on Local Linear Smoothing Technique", *An International Journal Information Sciences*, 177, 3882-3900.
- Watson G. S. (1964), "Smooth Regression Analysis", *Sankhya, Series A*, 26, 359-372.
- Wood S. N. (2006), *Generalized Additive Models (An Introduction with R)*, Chapman&Hall/CRC.
- Xizhao W., Minghu H. (1992), "Fuzzy Linear Regression Analysis", *Fuzzy Sets and Systems*, 51, 179-188.
- Yatchew A. (2003), *Semiparametric Regression for the Applied Econometrician*, Cambridge University Press.
- Zadeh L.A. (1965), "Fuzzy Sets", *Information and Control*, 8, 338-353.

Zadeh L.A. (1978), "Fuzzy Sets As a Basis for a Theory of Possibility", Fuzzy Sets Syst. 1(1), 3-28.