



T.C.
NECMETTİN ERBAKAN
ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ



AYKIRI DEĞER VARLIĞINDA
GENELLEŞTİRİLMİŞ EKLEMELİ
MODELLER VE BİR UYGULAMA

Taloutou Yarı DRAMANE

YÜKSEK LİSANS TEZİ

İSTATİSTİK Anabilim Dalı

Kasım-2022
KONYA
Her Hakkı Saklıdır

ÖZET

YÜKSEK LİSANS TEZİ

AYKIRI DEĞER VARLIĞINDA GENELLEŞTİRİLMİŞ EKLEMELİ MODELLER VE BİR UYGULAMA

Taloutou Yarı DRAMANE

Necmettin Erbakan Üniversitesi Fen Bilimleri Enstitüsü
İSTATİSTİK Anabilim Dalı

Danışman: Dr.Öğr. Üyesi Ahmet PEKGÖR

2022,75 Sayfa

Jüri

Dr.Öğr.Üyesi Ahmet PEKGÖR
Prof.Dr. Aşır GENÇ
Prof.Dr. Coşkun KUŞ

Günümüzde teknolojinin hızlı bir şekilde ilerlemesi sebebiyle birçok alanda otomatik olarak veri toplanmaktadır. Bu verileri analiz edebilmek yani verilerden faydalı bilgi çıkartmak önem arz etmektedir. Bu bakımdan literatürde yer alan genelleştirilmiş toplamsal modelleri bu çalışmanın hedefi olmuştur. Ancak veri toplama araçlarında bazen kalibrasyon hataları, veri girişi hatası, işlem sorunu, modeldeki eksik faktörler, rassal hatalar gibi sıkıntılardan dolayı sistem aykırı değerler üretebilir ve bunun sonucunda da yanlış tahminlere ulaşılabilir. Benzeri nedenlerden ötürü bu çalışmada, aykırı değerlerin model fonksiyonuna göre farklı bölgelerde, bozulmuş şiddetleri ve bozulmuş oranlarına göre 5 ayrı yöntemde genelleştirilmiş toplamsal modellerinin bir birlerine karşı performansları incelenmiştir. Aykırı değerler varlığında yöntemler içerisinde PDSR ve TDSR yöntemleri kayda değer bir performans sergilemişlerdir.

Anahtar Kelimeler: Doğrusal düzleştiriciler, Genelleştirilmiş toplamsal modeller, Aykırılık, Monte-Carlo simülasyon.

ABSTRACT

MS THESIS

GENERALIZED ADDITIVE MODELS AND AN APPLICATION IN OUTLIER ENTITY

Taloutou Yarı DRAMANE

**THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCE OF
NECMETTİN ERBAKAN UNIVERSITY
THE DEGREE OF MASTER OF SCIENCE
IN STATISTICS**

Advisor: Asst.Prof.Dr. Ahmet PEKGÖR

2022, 75 Pages

Jury

Asst.Prof.Dr. Ahmet PEKGÖR

Prof.Dr. Aşır GENÇ

Prof.Dr. Coşkun KUŞ

Today, due to the rapid development of technology, data is collected automatically in many areas. It is important to be able to analyze these data, that is, to extract useful information from the data. In this respect, the generalized additive models in the literature have been the target of this study. However, sometimes the system may produce outliers due to problems such as calibration errors, data entry errors, processing problems, missing factors in the model, and random errors in data collection tools, and as result, it is possible to reach biased estimates. For similar reasons, in this study, according to the model function of the outlier in different regions and according to the distortion intensities and distortion rates the performance of five different methods against each other of the generalized additive models was investigated. In the presence of outliers, PDSR and TDSR methods among the methods showed a remarkable performance.

Keywords: Linear smoothings, Generalized additive models, Outliers, Monte-Carlo simulation

ÖNSÖZ

Bu tez çalışmanın hazırlanma sırasında göstermiş olduğu yakın ilgisinden ve her türlü yönlendirici yardımlarından dolayı tez danışmanım Sayın Dr.Öğr. Üyesi Ahmet PEKGÖR' e, tez izleme komitesi üyesi Sayın Prof.Dr. Aşır GENÇ' e, ve Sayın Prof.Dr. Coşkun KUŞ'a, ilgi ve desteklerini her zaman yanımda hissettiğim Dr.Öğr. Üyesi Aydın KARAKOCA'a, tüm değerli hocalarıma ve hem Türk Milletine hem de Türkiye'deki ikinci Aileme çok teşekkür ederim.

Beni bugünlere getirdiği, hiçbir fedakârlıktan kaçınmadığı, tüm dertlerime ortak olduğu, zorlukları benimle paylaştığı, bana her konuda destek olduğu ve can yoldaşı olduğu için canım Anneme ve Babama sonsuz teşekkür ediyorum, hakları nasıl öderim?

İşlerimin yoğun olduğu günlerde bana büyük bir sabır gösteren sevgili eşime ve tüm Benin Milletine en içten teşekkürlerimi sunarım.

Taloutou Yarı DRAMANE
KONYA-2022

İÇİNDEKİLER

ÖZET	iv
ABSTRACT.....	v
ÖNSÖZ	vi
İÇİNDEKİLER	vii
SİMGELER VE KISALTMALAR	ix
1. GİRİŞ	1
2. DOĞRUSAL DÜZLEŞTİRİCİLER	5
i. Yan(Bias) ve varyans dengelemesi(Trade-off).....	6
ii. Düzleştirici Matrisi ve Serbestlik Derecesi	6
iii. Düzleştirici Parametresinin Otomatik Seçimi	7
2.1. Loess Düzleştirici	10
2.1.1. Loess yöntemi ve algoritması	11
2.1.2. R programlama dilinde Loess	12
2.2. Çekirdek Düzleştirici	12
2.2.1. Çekirdek yoğunluk tahmin edicisi (Kernel Density Estimator).....	13
2.2.2. Çekirdek yöntemi.....	14
2.2.3. R programlama dilinde çekirdek düzleştirici.....	15
2.3. Splines.....	15
2.3.1. Spline Regesyonu (Spline Regression).....	15
2.3.2. Düzleştirici Spline (Smoothing Spline)	19
3. GENELLEŞTİRİLMİŞ TOPLAMSAL MODELLERİ.....	24
3.1. Toplamsal Modelleri.....	24
3.2. Genelleştirilmiş Doğrusal Modelleri	25
3.3. GAM Modelleri	27
3.3.1. GAM Algoritması	28
3.3.2. R programlama dilinde GAM	29
3.3.3. Aykırılık.....	30
4. UYGULAMA	32
5. SONUÇLAR VE ÖNERİLER.....	44
5.1. Sonuçlar	44
5.2. Öneriler	44
6. KAYNAKLAR	45
EKLER	49

EK-1 Monte-Carlo simülasyonlarıyla elde edilen 5 ayrı yönteme ilişkin bozulmamış ve bozulmuş verilerdeki Hata Kareler Ortalaması, Ortalama Sıra İndisleri ve Standart Sapma performansları	49
EK-2 Simülasyonlarda kullanılan R programına ait kodlar.....	59
ÖZGEÇMİŞ	64



SİMGELER VE KISALTMALAR

Simgeler

- ε : Epsilon
 g : Düzleştirici fonksiyonu
 \hat{g}_λ : Düzleştirici fonksiyonunun tahmin edicisi
 λ : Düzleştirme parametresi
 S_λ : Düzleştirici matrisi (Smoothing matrix)
 w_j^i : Ağırlık fonksiyonu
 E : Beklenen değer
 Y_i^* : Yeni bir örneklem
 σ : Varyans
 $\hat{\sigma}$: Varyans tahmini
 b_λ : Bias vektörü
 ξ : Düğüm nokta(knots)

Kısaltmalar

- GAM* : Genelleştirilmiş Toplamsal Modelleri (Generalized Additive Models)
GLM : Genelleştirilmiş Doğrusal Modelleri (Generalized Linear Models)
GAMLSS : Genelleştirilmiş Toplamsal Modelleri (Location, Scale ve Shape parameters)
CV : Çapraz doğrulama (Cross Validation)
GCV : Genelleştirilmiş Çapraz doğrulama (Generalized Cross Validation)
ASR : Artık kareler ortalaması
 C_p : Mallows's istatistiği
df : Serbestlik derecesi
AMSE : Average Mean Squared Error
APSE : Average Predictive Squared Error
MSE : Hata Kareler Ortalaması (Mean Squared Error)
TDSR : İnce Tabakalı Düzleştirici Spline
LR : Yerel Regresyon
PDSR : P-spline ile Düzleştirici Spline

KDSR : K bik D zleřtirici Spline
SR : Spline Regresyonu



1. GİRİŞ

İnsanoğlu olayları anlamlandırabilmek için olayları gözlemleyerek gözlemlere dayalı veriler tutarak birtakım işlemlerle bu veriler üzerinden daha kapsamlı bir anlamlandırma oluşturur. Bu işlemlerin başında da veri analizleri yer almaktadır. Veri analizlerinde en yaygın olarak kullanılan araçlardan biri de doğrusal regresyondur. Doğrusal regresyon da amaç, bağımsız değişkenler üzerinden bağımlı değişkeni anlamlandırabilmektir. Ancak bu işlemleri yapabilmek için bazı ön varsayımların sağlanması gerekmektedir. Bu varsayımlar kısaca:

- Bağımlı değişken ile bağımsız değişkenler arasında doğrusal ilişkinin olması
- Bağımlı değişken dağılımının normal dağılıma sahip olması
- Bağımsız değişkenler arasında doğrusal bir ilişkinin olmaması.
- Modelde hata terimlerinin sıfır ortalamalı, sabit varyanslı ve birbirinden bağımsız olması

gerekmektedir. Ancak uygulamalarda bu varsayımlar her zaman gerçekleşmeyebilmektedir. Bu varsayımlardan en az birinin sağlanmaması durumunda ise doğrusal regresyona alternatif olarak diğer yöntemler kullanılmaktadır. Bu yöntemlerden ilki parametrik olmayan regresyon yöntemleridir. Buja ve ark.(1989,p.454), literatürde parametrik olmayan regresyon süreçlerinin, özellikle doğrusal düzleştiricilerin çok geniş olduğunu belirtmişlerdir. Bu alanda Watson(1964), Rosenblatt(1971), Reinsch(1967), Priestly ve Chao(1972), Ston(1977), Craven ve Wahba(1979), Cleveland(1979), Friedman ve Stuetzle(1981), Silverman(1985)'nın kapsamlı çalışmaları bulunmaktadır. Bu yöntem, tahmin fonksiyonunun önceden belirlenmiş bir form almadığı bir regresyon analizi biçimini oluşturmayı mümkün kılmıştır. Bu form verilerden gelen bilgiler üzerine kuruludur. Dolayısıyla Hastie ve Tibshirani(1990,p.1), verilerin en uygun uyum formu göstermesine izin vermek düşüncesini ortaya koymuşlardır. Veriler model yapısını ve modelin tahminini sağlayabilmesi için örneklem hacmi parametrik regresyondan daha büyük olması gerekmektedir. Hardle ve Luckhaus (1984) parametrik olmayan regresyon yöntemlerini sağlam (Robust) yöntemler olarak ifade etmişlerdir. Parametrik olmayan regresyonda pek çok yöntem mevcuttur. Bu yöntemler içerisinde literatürde en yaygın kullanılanı eğri yumuşatmadır (smoothings - düzleştiricilerdir). Takezawa(2005), Düzleştirici, basit parametrik olmayan regresyon için kullanılan bir terimdir. Düzleştiriciler, polinom fonksiyonlarında çok sayıda dönüşüm yöntemlerinin olması sebebiyle, en doğru ilişkiyi

zahmetsiz bir şekilde elde edebilmek için kullanılmaktadır. Düzleştiriciler, bağımlı değişken ile bir veya daha fazla bağımsız değişkenler arasındaki eğrisellik fonksiyonunu özetlemek için kullanılan bir araçtır (Hastie ve Tibshirani, 1990).

Literatürde basit parametrik olmayan regresyonda en yaygın kullanılan düzleştirici yöntemler Loess (locally-weighted running-line) düzleştirici, Çekirdek (Kernel) düzleştirici, spline regresyonu ve spline düzleştiricileridir. Bu yöntemler tahmin edicinin esnekliğini kontrol etmeye izin vermektedir. Ancak bu esneklik, verinin aşırı uyum göstermesi tahmin edicinin varyansının artmasına, aynı biçimde düzleştiricinin düzgünleşmesi de yanın artmasına neden olmaktadır.

Araştırmalarda her zaman asıl hedef, incelenen olaylarda maksimum bilgiye ulaşmaktır. Ancak bağımsız değişkenler aynı anda bağımlı değişken üzerine sahip olduğu etki incelenmek istenildiğinde akla gelen yöntemlerin başında çoklu parametrik olmayan regresyondur. Çoklu parametrik olmayan regresyon modellerinden en sık kullanılanların başında genelleştirilmiş toplamsal modeller (Generalized Additive Model - GAM) gelir.

Genel olarak istatistiksel analizlerde bağımlı değişken kategorik olduğu durumlarda Genelleştirilmiş Doğrusal Modeller (Generalized Linear Models - GLM) tercih edilir. GLM, ilk defa Nelder ve Wedderburn (1972) tarafından önerilmiştir, normal dağılımdan farklı hata dağılımlarına ve model yapısında belli bir dereceye kadar doğrusal olmayanlığa izin verir (Wood,2017). GLM, normal (ordinary) doğrusal regresyonda normallik varsayımı ve değişen varyans sorunları olduklarında çözüm olarak kullanılmaktadır. GLM'de bağımlı değişkenler üstel aileye sahiptir. GLM'ler çoklu doğrusal regresyon yapısına benzer ve sadece bağımlı değişkeni doğrusal hale getirmek amacıyla dönüşüm metotları uygulamaktadır. Fakat uygun polinom terimleri veya parametrik dönüşümler her zaman mümkün olmadığından, GLM 'nin daha geniş bir genelleştirilmesine yol açan Genelleştirilmiş Toplamsal Modelleri ortaya konulmuştur. GAM ilk kez Hastie ve Tibshirani (1986,1990) tarafından literatüre kazandırılmıştır. Hastie ve Tibshirani(1990), Guisan ve ark.(2002), GAM'lar, GLM'ler sınıfının bir uzantısı olduğunu ve açıklayıcı değişkenlerinin doğrusal düzleştirici (smooth) fonksiyonlarının bir toplamı olarak belirtmişlerdir. Parametrik ve parametrik olmayan regresyon modellerinin bir kombinasyonunun yarı parametrik regresyon modelleme olması açısından GAM'lar, literatürde yarı(semi) parametrik modelleri olarak da adlandırılmaktadır. Yani, yarı parametrik regresyon, geleneksel parametrik regresyon analizi ile parametrik olmayan regresyon yöntemlerinin bir bütünüdür. Yarı parametrik

regresyon modelleri verilerdeki karmaşıklığı azaltır ve anlaşılabilir şekilde özetler. Her iki yapıda birbirini tamamlamaktadır. Yarı parametrik regresyon her iki yapıyı kullanarak araştırmacılara en iyi ve tutarlı sonuçları verebilme imkanı sağlamaktadır (Ruppert ve ark., 2009). Yarı parametrik regresyon modeli ilk olarak Engle ve ark. (1986) tarafından elektrik satışları ve hava durumu için kullanılmıştır ve daha sonraları Young ve Bowman (1995), soğan verilerine benzer yarı parametrik modelleri uygulamışlardır.

GAM'lar üzerine literatürde geniş ve farklı alanlarında uygulama çalışmaları da bulunmaktadır. İstatistik alanında başta olmak üzere, Hastie ve Tibshirani (1986), doğrusal formunu değiştiren eklemeli düzleştirici fonksiyonlar ile genelleştirilmiş toplamsal modeller sınıfını tanıtmışlardır. Asfha (2017), çoklu bağlantı ve aykırı değer varlığında üç cezalı regresyon spline'ları, kübik spline, p-spline ve ince tabakalı spline'ların performanslarını incelemiştir. Gressani ve Lambert (2020), Laplace-P-spline metodolojisini tasarlamak için Laplace yaklaşımı ve Bayesian ile P-splines'ı birleştirerek genelleştirilmiş toplamsal modellerin sınıfında sonuç çıkarımı yapmışlardır. Aeberhard ve ark. (2021), Genelleştirilmiş Toplamsal modellerin sınıfındaki konum, şekil ve ölçek (*GAMLSS*) parametreleri için sağlam düzleştirici parametre seçimi önermişlerdir. Konum, ölçek ve şekil için *GAMLSS* gibi çeşitli tahmin edicilere sahip karmaşık modeller için, bazı dağılım parametrelerinin diğerlerinden daha sık güncellendiği tahmin edicilerin dengesiz güncellemeleri, bazı alt modellerin içine uygun şekilde yerleştirilmesini engelleyen bir sorundur. Bu sorunu Zhang ve ark. (2021), *GAMLSS* için döngüsel olmayan bir yükseltme algoritması içinde uyarlamalı adım uzunluğu belirleme kullanan bir yaklaşım önermişlerdir.

GAM'ın uygulama alanı çok geniş olmasından literatürde bu konuda pek çok çalışma yer almaktadır. Ekoloji alanında Guisan ve ark. (2002), tür (species) dağılımlarına çalışmalarında genelleştirilmiş doğrusal ve genelleştirilmiş toplamsal modellerini uygulamışlardır. Tıp biliminde Vandal, N. (2005), meme yoğunluğu ve iki insülin benzeri büyüme faktörü arasındaki ilişkiyi inceleyen bir çalışmada genelleştirilmiş toplamsal modelleri kullanılmıştır. Biyoloji alanında, Gianola ve ark.(2006), genomik desteğiyle genetik değer tahmininde semiparametrik yöntemler üzerine incelenme yapmışlardır. Psikoloji alanında da Savaş (2009), zeka testi yapılan kişilere ait anket verisine Genelleştirilmiş Toplamsal Modelleri uygulamıştır. Adamec ve Drápela (2015), yükseklik eğrisini modellemesinde genelleştirilmiş toplamsal modelleri test etmek için alternatif olarak üç tip düzleştirici fonksiyonları kullanılmıştır. Ateba ve ark. (2020), Mali'nin Dangassa topluluğu üzerinde sıtma iletim dinamiklerini tahmin

etmek için fonksiyonel Genelleştirilmiş Toplamsal modellerini kullanan yeni bir yaklaşım uygulanmıştır. Sorensen ve ark. (2020), altı yaş beyni kohortunda kendi kendine bildirilen uyku ve hipokampal hacim arasındaki ilişkiyi öğrenmek için Genelleştirilmiş Toplamsal modelini kullanmışlardır. Pyrgou ve Santamouris (2020), mortalite uyarının etkisini ölçmek için Genelleştirilmiş Toplamsal modeli tekniği kullanarak soğuk ve aşırı sıcak arasındaki ilişkiyle ilgili ölüm riskini incelemişlerdir. Analitis ve ark. (2020)'nın çalışmasında, hava kirliliğinin sağlığa etkileri incelendiğinde genelleştirilmiş toplamsal modelleri ve makine öğrenimi yöntemleri kullanılmıştır. Ziraat alanında, Dhulipala ve Patil (2020), Hindistan'ın tarım sektöründeki navlun üretiminde çoklu doğrusal regresyon ve genelleştirilmiş toplamsal modelleri kullanılarak *GAM*'ın daha iyi performans verdiğini göstermişlerdir. Erfanian ve ark. (2021), sağlık performansı izlemede, *GAM*'lerin başarılı performans gösterdiğini vurgulamışlardır. Amato ve ark. (2021), elektrik tesislerinin planlanması tahminlerinde yarı parametrik genelleştirilmiş toplamsal modelleri kullanmışlardır. Mühendislik alanında Bakhshi ve Ahmed (2022) uçak kazası tahminlerinde *GLM* ve *GNM* (Genelleştirilmiş Nonlinear Model) modeller'den *GAM* daha iyi performans göstermiştir.

Teknolojinin ilerlemesiyle, verileri analiz etmeyi kolaylaştıran yeni istatistiksel yazılımların uygulanmasına yol açmıştır. Literatürde yapılan çalışmalarda, son derece popüler bir istatistiksel yazılım haline gelen R programlama dili, istatistiksel hesaplamalarda yaygın olarak kullanılmaktadır. Wood (2006)'nin çalışması, R programlama diliyle genelleştirilmiş toplamsal modellerinin çağdaş bir anlatımıdır. Perperoglou ve ark.(2019), R programlama dilinde spline fonksiyon prosedürlerinin incelenmesi yapılmıştır.

Genel olarak modellemede amaç, araştırma kapsamında toplanan çok sayıda değişkenle ilgili derlenen bilgilerin değerlendirilmesi ve özet hale getirilmesidir. Tutarlı bir analiz elde etmenin ilk adımlarından birisi merkezden uzak olan gözlemlerden modelin etkilenmesini azaltmaktır. Aykırı değerler (Outliers) genellikle hata ya da problem olarak dikkate alınsa da bazı araştırmalarda önemli bilgiler taşıyan değişkenler olarak değerlendirilmektedir. Ancak aykırı değerler genellikle yanlış model kurmaya, yanlış parametre tahminlerine ve doğru olmayan sonuçlara yol açan değerlerdir.

Bu tez çalışmasında amaç modellemelerde yaygın kullanılan *GAM*, modelleme de hayati öneme sahip olan aykırı değerlerin etkisinin azaltılması ve aykırı değerlerin varlığı altında daha tutarlı sonuçların elde edilebilmesini sağlamak ve simüle veri setleri üzerinde bunu göstermektedir.

2. DOĞRUSAL DÜZLEŞTİRİCİLER

Sırasıyla x_i ve y_i bağımsız ve bağımlı değişkenlerin i . gözlem değerleri ve (x_i, y_i) , çiftlerden oluşan rastgele bir örneklem olmak üzere parametrik olmayan regresyon modeli

$$y_i = g(x_i) + \varepsilon_i, i = 1, 2, \dots, n \quad (2.1)$$

biçimindedir. Burada tahmin edilen $g(x_i)$ doğrusal regresyonun tersine kısıtları olmayan ve bilinmeyen regresyon fonksiyonu, ε_i ise 0 ortalamalı ve sabit varyanslı bağımsız rastgele hata terimidir. Genel olarak herhangi bir yöntemle $g(x_i)$ tahminine başlamadan önce veriler her zaman bir serpilme dağılım grafiği üzerine görselleştirmesi gerekmektedir. Serpilme dağılım grafikleri örneklemin tüm çift noktalarında temsil edilmektedir. Rastgele örneklemin regresyon fonksiyonunun tahmini, verinin eğrisini en iyi temsil eden fonksiyon olarak tanımlanmaktadır. Bu durumda $g(x_i)$ tahmini parametrik olmayan şekilde temsil edildiği takdirde onu düzleştirici fonksiyonu eşitlik (2.1) (smoothing function) tarzında belirler. Hastie ve Tibshirani (1990,p. 45), düzleştiricileri doğrusal düzleştiriciler (Linear smoothers) ve doğrusal olmayan düzleştiriciler (Nonlinear smoothers) olmak üzere iki sınıf olarak ayırmışlardır. Bu çalışmada sadece doğrusal düzleştiriciler üzerinde durulacaktır. Doğrusal düzleştiriciler içerisinde en çok kullanılan Loess, Çekirdek, Spline Regresyonu ve Düzleştirici Spline düzleştiricileridir. Doğrusal düzleştiriciler, tek açıklayıcı değişken olduğu durumda “serpilme dağılım grafiği düzleştirici- Scatterplot Smoothers” olarak adlandırılmış ve Doğrusal regresyondaki temel varsayımın aksine Y ler X lerin doğrusal bir formu olmak zorunda değildir. Dolayısıyla düzleştiriciler parametrik olmayan regresyon sınıfındadır. Genelde düzleştiricilerin iki ana kullanımı vardır. İlk olarak veriyi betimlemektedir. Yani Y 'ye karşı X 'in dağılım grafiğinin görsel görünümünü geliştirmek ve gözlem değerleri arasındaki eğilimi seçmesine yardımcı olmak için serpilme dağılım grafiği tercih edilmektedir. İkincisi ise Y 'nin ortalamasını X 'ler üzerinden bağımlılığını tahmin etmek ve bu da GAM 'ların tahmini için benzer bir yapıyı oluşturmaktadır. Düzleştiricilerin daha esnek olmasının nedeni düzeltme parametre değerinin seçilebilmesindendir. Genel olarak düzeltme parametresinin alabileceği değerler pozitif olmasının yanında düzeltme parametresi, düzleştirici fonksiyonlara göre farklı değerler almaktadır. Böyle olması sebebiyle $g(x_i)$ için çeşitli esnek tahmin ediciler elde edilebilmektedir. Daha sonra da noktalar arasında basit enterpolasyon yöntemiyle eğrisellik tamamlanmaktadır. Bu konu ile ilgili R programında uygulamalar Pekgör (2022) çalışmasında mevcuttur.

i. Yan(Bias) ve varyans dengelemesi(Trade-off)

Düzleştirme sınıfının seçimi ve düzleştirme parametresinin seçimi, düzleştiricilerin temel özellikleriydi. Bu özellikler yan-varyans dengelemesi içermektedir. Birinin azalması çoğu zaman diğerinin artmasına neden olmaktadır. Verinin aşırı uyumu (Overfitting) kavramı, düzleştiriciler için büyük bir tehlike arz etmektedir ve bu durumda her ne kadar yan azalsa da varyansın artmasına sebep olmaktadır. Yan ve varyans arasındaki dengeleme λ (düzleştirme parametresi) tarafından yönetilmektedir (Hastie ve Tibshirani, 1990, p.40). Düzleştirme parametresi, $\lambda \rightarrow 0$ iken yanın azalmasına karşılık varyansın artmasına neden olurken, $\lambda \rightarrow \infty$ iken her ne kadar yanı artırsa da varyansı azaltmaktadır. Dolayısıyla küçük bir düzleştirme parametresinin kullanılması daha fazla varyans ve daha az yana sahip olan tahmin edici elde edilmesini sağlamaktadır. Öte yandan λ değeri arttıkça daha küçük varyanslı ve yanlı bir tahmin edici elde edilmektedir. Optimal bir λ seçimi ile alakalı literatürde Çapraz doğrulama (Cross-Validation - CV), Genelleştirilmiş Çapraz doğrulama (Generalized Cross Validation- GCV), artık kareler ortalaması (Average Squared Residual- ASR) ve Mallows's C_p istatistiği gibi birçok kriterler bulunmaktadır. Bu yöntemler hata kareler ortalamasını en küçüklenme esasına dayalı yöntemler sınıfındadır.

ii. Düzleştirici Matrisi ve Serbestlik Derecesi

Doğrusal regresyonda serbestlik derecesi (Degrees of Freedom- df) basitçe Şapka matrisinin (Hat matrix) izi olarak tanımlanmaktadır. Bu tanım doğrusal düzleştiriciler için de geçerlidir. Doğrusal düzleştiricinin türüne göre düzleştirici matrisi tasarlanmaktadır. Örneğin Yerel ağırlıklı koşu hattı düzleştiriciler (Locally Weighted running- line smoothers) için S_λ (düzleştirici matris)' in elemanları

$$S_{ij} = \frac{w_j^i}{\sum_{k=1}^m N(x_i)^{w_k^i}} + \frac{(x_i - \bar{w}_j^i)(x_j - \bar{x}_w^i)w_j^i}{\sum_{k=1}^m N(x_i)^{w_k^i} (x_j - \bar{x}_w^i)^2}, \quad i, j = 1, 2, \dots, n \quad m \leq n$$

biçimindedir. Burada $S_{ij} = S_\lambda$ olarak tanımlanmaktadır. Dolayısıyla serbestlik derecesi S_λ ' nin özdeğerlerinin toplamıdır.

$$df = tr(S_\lambda) \quad (2.2)$$

Ancak doğrusal düzeltiriciler için eşitlik (2.2) dışında iki yaygın df tanımı daha vardır. Bunlar, $n - tr(2S_\lambda - S_\lambda S_\lambda^T)$ ve $tr(S_\lambda S_\lambda^T)$ dır. Bu tanımlar doğrusal regresyon modeline benzetilerek uyarlanabilir ve farklı amaçlar için de kullanılabilir. Dolayısıyla serbestlik derecesi Y 'nin dağılımına bağlı olmasına rağmen bu mantık da doğrusal olmayan düzeltiricilere genişletilebilmektedir. Her ne kadar düzeltirme parametresi kullanımı serbestlik derecesi kullanımına göre daha iyi sonuç verse de, uygulamalarda düzeltirme parametresi seçiminin zorluğundan dolayı serbestlik derecesi kullanımı daha çok tercih edilmektedir.

iii. Düzeltirici Parametresinin Otomatik Seçimi

Düztirme parametresinin optimal değerini seçmek için literatürde çeşitli kriterler vardır. Bu kriterler eşitlik (2.1) deki \hat{g}_λ yi tahmin etmek için kullanılmaktadır.

Düztirme parametresini seçmek için her x_i 'de yerel hata kareler ortalaması en küçükleme yerine genel hata kareler ortalaması (Average Mean Squared Error, $AMSE$) en küçüklendir. Dolayısıyla $AMSE(\lambda)$ formülü aşağıdaki gibidir.

$$AMSE(\lambda) = \frac{1}{n} \sum_{i=1}^n E\{g(x_i) - \hat{g}_\lambda(x_i)\}^2 \quad (2.3)$$

Ancak literatürde $AMSE$ ' den biraz daha farklı bir kriter de kullanılmaktadır ve farklılık sadece σ^2 'nin sabit bir fonksiyonu iledir. Bu kriter tahmin hata karelerin ortalaması (Average Predictive Squared Error, $APSE$) olarak adlandırılmaktadır. En iyi tahmin edici elde etmek için gözlem değerlerinin belli orandaki kısmı (training data) alınarak elde edilen $\hat{g}_\lambda(x_i)$ ile verinin geri kalan kısmıyla yeni bir örneklem (test data) kullanılarak tahmin hata kareleri en küçüklendir.

$$APSE(\lambda) = \frac{1}{n} \sum_{i=1}^n E\{Y_i^* - \hat{g}_\lambda(x_i)\}^2 \quad (2.4)$$

Burada Y_i^* , x_i 'deki olan bir gözlemdir ve $Y_i^* = g(x_i) + \epsilon_i^*$ olarak tanımlanmaktadır. ϵ_i^* , ϵ_i 'lerden bağımsızdır ve genel biçimde

$$APSE = AMSE + \sigma^2 \quad (2.5)$$

olarak ifade edilmektedir. Alternatif olarak ve daha uygun bir strateji X 'in gerçek dağılımının ortalaması üzerinden tahmin hata kareler ortalamasının beklenen değerini (Expected Squared prediction errors averaged) en küçüklemeektir. Bunun için de bazı ekstra bilgi veya varsayımlar gerekmektedir. Dolayısıyla PSE tanımlandığı gibi kullanılmaktadır. Dolayısıyla APSE tahmini oluşturması mantığına göre otomatik düzeltirme parametresi seçimi yöntemleri elde edilebilmektedir (Hastie ve Tibshirani, 1990, p.42).

Hem AMSE hem de APSE için gerekli olan düzeltirme parametresi seçiminde yan ve varyans ilişkisini kurgulayarak optimal değerin elde edilmesinde kullanılan yöntemlerden biri olan Çapraz doğrulama (Cross-Validation - CV), en çok kullanılan kriterlerden biridir. (x_i, y_i) noktaları birer birer dışlanarak ve geride kalan $n - 1$ noktalarına dayalı olarak x_i noktasındaki \hat{y}_i tahmin ederek kullanılmaktadır. Çapraz doğrulama eğitim ve test veri setlerinin kullanımına benzer olarak hata karelerin ortalaması olarak oluşturmaktadır.

$$CV(\lambda) = \frac{1}{n} \sum_{i=1}^n \{y_i - \hat{g}_{\lambda}^{-i}(x_i)\}^2 \quad (2.6)$$

$\hat{g}_{\lambda}^{-i}(x_i)$ ayrılmış i . veri noktası bırakılarak ve x_i noktasındaki hesaplanmış tahminidir. CV, uygun bir aralıkta λ değerlerini hesaplayarak bunlar içerisinde en küçük değeri seçmektedir. Dolayısıyla bu yöntem yaklaşık olarak aşağıdaki gibi bulunmaktadır.

$$E\{cv(\lambda)\} \approx APSE(\lambda) \quad (2.7)$$

Düzleştirici matrisinin köşegen elemanları basitçe elde edilebildiğinde CV'nin hesaplanması kolaydır. Örnek olarak Yerel ağırlıklı koşu hattı düzleştiriciler için S (düzleştirici matris)'in elemanları şu şekilde hesaplanmıştır.

$$S_{ij} = \frac{w_j^i}{\sum_{j \in N(x_i)} w_j^i} + \frac{(x_i - \bar{w}_j^i)(x_j - \bar{x}_w^i)w_j^i}{\sum_{j \in N(x_i)} w_j^i (x_j - \bar{x}_w^i)^2} \quad (2.8)$$

Burada w_j^i , i . hedef noktasının komşuluğundaki j . noktasının aldığı ağırlığı temsil ederken \bar{x}_w^i ise $N(x_i)$ bölgesindeki x_j 'nin ağırlıklı ortalamasıdır.

Fakat şu ana kadar spline düzleştirici için S_{ij} köşegen elemanlarının hesaplanması bilinmedikleri sebebiyle CV'in iki varyasyon ortaya koymuştur. Dolayısıyla

Genelleştirilmiş Çapraz doğrulama (Generalized Cross Validation- *GCV*) *CV*'nin bir uzantısı olarak tanımlanmaktadır. Bu problemi çözmek için *GCV* kriteri, S_{ij} yerine daha kolay olsun ki onun ortalamasının değeri $tr(S)/n$ olarak kullanılabilir. Hastie ve Tibshirani (1990,p. 49). *GCV* aşağıdaki gibi hesaplanmaktadır.

$$GCV = \frac{1}{n} \sum_{i=1}^n \left\{ \frac{y_i - \hat{g}_\lambda(x_i)}{1 - tr(S_\lambda)/n} \right\}^2 \quad (2.9)$$

Hastie ve Tibshirani (1990,p. 49), *GCV*, *CV*'ye asimptotik olarak yakınsadığını ispatlamışlardır.

Hastie ve Tibshirani(1990,p.44), artık kareler ortalaması (Average Squared Residual- *ASR*) $APSE(\lambda)$ 'nin saf (naive) tahmini olmasına rağmen iyi bir tahmin olmadığını ifade etmişlerdir.

$$ASR(\lambda) = \frac{1}{n} \sum_{i=1}^n \{y_i - \hat{g}_\lambda(x_i)\}^2 \quad (2.10)$$

Doğrudan $APSE$ tahmini oluşturması için $ASR(\lambda)$ kriteri'nin kalibre edilmesi gerekmektedir. Örneğin, $\hat{g}_\lambda = S_\lambda y$ varsayarak ve bias vektörü $b_\lambda = g - E(S_\lambda y) = g - S_\lambda g$ olarak alınmaktadır. Buna göre $E\{ASR(\lambda)\}$ 'nin $APSE(\lambda)$ 'dan biraz daha farklı olduğu görülmektedir. Bu aşağıdaki gibi gösterilebilir.

$$E\{ASR(\lambda)\} = \left\{ 1 - \frac{tr(2S_\lambda - S_\lambda S_\lambda^T)}{n} \right\} \sigma^2 + \frac{b_\lambda^T b_\lambda}{n} \quad (2.11)$$

Burada küçük bir yan ile tahmin etmek istenildiğinde σ^2 nadiren bilinmektedir. Dolayısıyla çözüm, Mallows's C_p istatistiği formundadır ve aşağıdaki gibidir (Hastie ve Tibshirani, 1990, p.48).

$$C_p(\lambda) = ASR(\lambda) + 2 tr(S_\lambda) \frac{\hat{\sigma}^2}{n} \quad (2.12)$$

Burada;

$$\hat{\sigma}^2 = ASR(\lambda^*) / \{n - tr(2S_{\lambda^*} - S_{\lambda^*} S_{\lambda^*}^T)\} \quad (2.13)$$

λ^* optimal düzleştirme parametresi ve $ASR(\lambda^*)$ ise artık kareler toplamıdır.

Son olarak Hastie ve Tibshirani (1990, p.52), düzleştirme parametresi seçmek için literatürde var olan CV, GCV, ASR ve C_p kriterlerine rağmen pratikte bu kriterlerin performansları bazen sorgulanabilmektedir. Düzleştirme parametresini seçmede kullanılan kriterlerin otomatik seçimi, sonuçların daha az güvenilir olmasına neden olmaktadır. Bundan dolayı özellikle toplamsal modeller için kullanımları uygun değildir. Dolayısıyla bu durumda Hastie ve Tibshirani (1990, p.52) ve Silverman (1985), daha çok grafiksel yöntemlerin kullanımına güvenmişlerdir ve bu yüzden literatürde optimal kriter değeri seçmede serbestlik derecesi kullanımı tercih edilmektedir. Bu konuda Härdle, Hall ve Marron (1988)'nın çalışmalarında ayrıntılı bilgiler yer almaktadır.

2.1. Loess Düzleştirici

Loess düzleştirici, istatistik literatüründe çeşitli isimler altında bulunmaktadır. Bunlardan bazıları Yerel Regresyon (Local Regression) (Cleveland ve Loader, 1996, p.1 - Loader, 1999, P.1 - Ryan, 1996, p.331, Chambers ve Hastie, 1992, p.314); Yerel polinom regresyon (Local Polynomial regression) (Wand ve Jones, 1990, p.4 - Opsomer ve ark., 2000, p.6); Yerel ağırlıklı regresyon (Locally weighted regression) (Cleveland, 1979, p.5); Yerel ağırlıklı polinom regresyon (Locally weighted polynomial Regression) (Cleveland ve Devlin, 1988); Yerel ağırlıklı koşu hattı düzleştiriciler (Locally Weighted running-line smoother-loess) (Hastie ve Tibshirani, 1990, p.16)'dır. Takezawa (2005, p.186), Loess teriminin kullanımında karışıklığı önlemek için daha iyi bir tanım kullanımı En yakın komşu yerel doğrusal regresyon (Nearest-Neighbor local-linear regression) dur. Ayrıca Lowess (LOcal Weighted Scatterplot Smoother (Yerel Ağırlıklı Serpilme Düzleştiriciler) ise Loess'in sağlam(robust) bir versiyonu olarak tanımlanmaktadır (Takezawa, 2005, p.186). Loess, parametrik olmayan regresyon bir tekniktir. Ama "Yerel regresyon" olarak kısaltmasını ifade etmişlerdir (Cleveland, 1993, p.94 - Fox, 2000, p.19 - Ryan, 1996, p.332). Bu çalışmada Cleveland (1979) tarafından önerilen yerel ağırlıklı regresyon ve özellikle R, S gibi programlama dillerinde loess olarak yer almaktadır.

Yerel polinom regresyon, ilk defa Cleveland (1979) tarafından önerilmiş ve yerel ağırlıklı regresyon olarak adlandırılmıştır (Ryan, 1997, p.331). Ama daha sonra Cleveland ve Devlin (1988) tarafından geliştirilmiştir. Loess kelimesi almanca löss'ten geliyor ve İngilizcede "loose" anlamına gelmektedir (Chamber ve ark., 1990, p.314- Cleveland, 1993, p.94).

2.1.1. Loess yöntemi ve algoritması

Loess yönteminde, $g(x_0)$ 'ı tahmin etmek için birkaç adımda hesaplanan k en yakın komşu (k nearest-neighbours) kullanılmaktadır.

i. x_0 'in en yakın k komşusu tanımlanır ve $N(x_0)$ ile gösterilir.

ii. x_0 'dan en uzak yakın komşunun uzaklığı, $\Delta(x_0) = \max |x_0 - x_i|$ hesaplanır

iii. w_i ağırlıkları, $tri - küp$ ağırlık fonksiyonu

$$W = \left(\frac{|x_0 - x_i|}{\Delta(x_0)} \right)^3 \quad (2.14)$$

kullanılarak $N(x_0)$ 'deki her bir noktaya atanır, burada,

$$W(u) = \begin{cases} (1 - u^3)^3, & 0 \leq u < 1 \\ 0, & \text{diğer yerlerde} \end{cases}$$

şeklindedir.

iv. $g(x_0)$, *iii.* deki hesaplanan ağırlıklar kullanılarak $N(x_0)$ bölgesinde x_0 'hedef noktasındaki tahmin fonksiyonu değeridir ve y 'nin x 'e göre uyumunu ağırlıklı en küçük kareler yöntemi ile elde edilmektedir (Hastie ve Tibshirani, 1990, p.30).

Genellikle ifade edilen en yakın komşu sayısı verinin yüzdesi veya veri noktalarının aralığı (*span*) düzleştirme parametresidir. Prensipde en yakın komşu yöntemi simetrik yakın komşu yöntemine göre daha çok tercih edilir. Çünkü en yakın komşu noktalarından hedef noktaya olan ortalama uzaklığı daha azdır (bağımsız değişkenler eşit aralıklı olmadıkça) ve tahmin edicinin yanlılığını azaltmaktadır. Uç nokta olan en yakın komşu, verilerin ortasındaki komşulukla aynı sayıda veri noktası içerir, böylece oraya en küçük kareler yöntemi uygulandığında, hedef noktadan uzaktaki noktalara çok fazla ağırlık verilmektedir. Loess için en yakın komşu yöntemi tri-küp fonksiyonuyla uzak noktaların ağırlığını azaltma işini yapmasından ötürü daha uygundur. Loess yöntemi algoritmasında ağırlıklı en küçük kareler tahmin edicisi elde etmek için kullanılan polinomun derecesini belirtmez. Bir veya iki dereceli polinomlar ile en kısa sürede tahmin edicinin sonuçları elde etmeye izin verdikleri için genellikle daha tercih edilmektedir. Genel olarak yerel ağırlıklı düzleştiriciler popülerdir. Çekirdek düzleştiricilerin pürüzsüzlüğü özelliklerine sahipken, bağımsız değişkenlerin yerel yoğunluğuna en yakın komşu düzleştiricilerin

yetenekleriyle bant genişliğini uyarlamaktadır (Hastie ve Tibshirani, 1990,p.29). Ryan(1997,p.331), Loess, çekirdek regresyonunda olduğu gibi bir ağırlık fonksiyonunun kullanılması gerçeğinden kaynaklanmaktadır. Dolayısıyla Loess düzleştirici, kernel regresyonu ve koşu hattı pürüzsüz (running line smoother) bir kombinasyonu olarak düşünebilmektedir (Ryan, 1997, p.331).

2.1.2. R programlama dilinde Loess

Loess yöntemi uyumunu, R programlama dilinde farklı fonksiyonlarla elde edilebilmektedir. Bunlar `{stat}` paketinde `loess.smooth()`, `loess()`, `scatter.smooth()`, ve `{gam}` paketinde yer alan `gam()`'in `lo()` alt fonksiyonlarıdır. Bu fonksiyonlar birbirine benzer olmaları yanında bazı seçenekleri ve varsayılan ayarlarında değişiklikler olabilmektedir. Dolayısıyla kullanıcılar tarafından dikkatli olunması gerekir. Nathalie(2005,p.19), `gam()` ile elde ettiği sonuçların `scatter.smooth()` ve `loess()` sonuçlarından biraz farklı olmasına karşın genel olarak benzer eğri şekline sahip olduklarını göstermiştir. Bu fonksiyonlarda “*span*” (bölgeye ait gözlemlerin yüzdesini) değerinin seçimi yanında polinom derecesi(*degree*) de belirtmek gerekir. Bu parametrelerin varsayılan değerleri birbirinden farklıdır. `loess.smooth()` ve `scatter.smooth()` fonksiyonlarda *span*=2/3 ve *degree*=1; `loess()` fonksiyonda *span*=0.75 ve *degree*=2; `lo()` fonksiyonda ise *span*=0.5 ve *degree*=1 olarak kullanılmaktadır. Son olarak optimal “*span*” değeri belirtmek için her 4 fonksiyonda da doğrudan otomatik düzleştirici parametresi kriteri kullanmak mümkün değildir. Dolayısıyla önceden otomatik düzleştirici parametresi kriteri ayarlanması gerekmektedir. Regresyon analizlerde `loess()` fonksiyonu, maksimum sadece 4 bağımsız değişken kullanabilmektedir.

2.2. Çekirdek Düzleştirici

Literatüde Çekirdek regresyonu (Kernel Regression), çekirdek düzleştiriciler (Kernel Smoothers) veya çekirdek yerel regresyon olarak adlandırılmıştır ve X rastgele değişken verildiğinde Y rastgele değişkeninin koşullu beklenen değerini tahmin etmede kullanılan tamamen (purely) parametrik olmayan regresyon yöntemidir. Bu yöntem, bağımsız değişkenler arasında ilişki olmaması gibi teorik varsayımlar gerektirmemektedir. Dolayısıyla Çekirdek Regresyon yöntemi, veri modellemede kullanıcıya büyük bir özgürlük sağlamaktadır.

2.2.1. Çekirdek yoğunluk tahmin edicisi (Kernel Density Estimator)

İstatistiksel çalışmalarda olasılık yoğunluk fonksiyonu tahmin etmek için çok kavramlar ve teknikler geliştirilmiştir. Bunlardan bazıları histogram, çekirdek yoğunluk tahmini edicisi sayılabilir. Çekirdek yoğunluk tahmin edicisi genelleştirilmiş histogram ailesine sahiptir. Ancak histogramın bazı eksikliklerinden dolayı birçok açıdan çekirdek yoğunluğu tahmini histograma göre daha çok tercih edilmektedir (Härdle ve ark., 2004). Genel olarak çekirdek yoğunluk tahmin edicisi kullanarak rastgele değişkenlerin olasılık yoğunluk fonksiyonu genel biçimini aşağıdaki gibidir.

$$\hat{f}_h(x) = \frac{1}{n} \sum_{i=1}^n K_h(x - X_i) \quad (2.15)$$

$$K_h(\bullet) = \frac{1}{h} K(\bullet | h) \quad (2.16)$$

Burada $K(\bullet)$ seçili çekirdek fonksiyonudur ve h ise bant genişliğini göstermektedir. Çekirdek fonksiyonu aynı anda ağırlık fonksiyonu (weighted function) anlamına gelmektedir ve denklem (2.15)'te $f(x)$ fonksiyonu ise çekirdek yoğunluk tahmin edicisidir (Kernel density Estimator) (Hardle ve ark., 2004)

Tablo 3.1 Literatürde yer alan bazı ağırlık fonksiyonları

Kernel	$K(u)$
Uniform	$\frac{1}{2} I(u \leq 1)$
Epanechnikov	$\frac{3}{4} (1 - u^2) I(u \leq 1)$
Triweight	$\frac{35}{32} (1 - u^2)^3 I(u \leq 1)$
Gaussian	$\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} u^2\right)$
Cosine	$\frac{\pi}{4} \cos\left(\frac{\pi}{4} u\right) I(u \leq 1)$
Quartic (Biweight)	$\frac{15}{16} (1 - u^2)^2 I(u \leq 1)$

2.2.2. Çekirdek yöntemi

Bir X rassal değişkeni bilindiğinde Y rassal değişkeninin koşullu beklenen değeri $E(Y|X) = g(X)$ 'e eşittir. Bilinmeyen bir fonksiyon olan $g(X)$ 'i tahmin etmede çekirdek regresyonu kullanılabilir.

$$g(x) = E(Y|X = x) = \int y f(y|x) dy = \int y \frac{f(x, y)}{f_X(x)} dy \quad (2.17)$$

Burada $f_X(x)$, X 'in marjinal olasılık yoğunluk fonksiyonu ve $f(x, y)$ ise X ve Y rasgele değişkenlerinin ortak olasılık yoğunluk fonksiyonudur (joint probability density function). Çekirdek yoğunluk tahmin edicisi kullanarak hem $f_X(x)$ hem de $f(x, y)$ tahmin edilebilmektedir ve aşağıdaki biçimdedir. Burada d boyut durumudur Hardle ve ark. 2004, p. 94,118)

$$\hat{f}(x) = \frac{1}{n} \sum_{i=1}^n \frac{1}{h^d} K\left(\frac{x - x_i}{h}\right) \quad (2.18)$$

$$\hat{f}_{h,g}(x, y) = \frac{1}{n} \sum_{i=1}^n K_h\left(\frac{x - x_i}{h}\right) K_g\left(\frac{y - y_i}{g}\right) \quad (2.19)$$

Eşitlik (2.17) deki işlem karmaşıklığını Nadaraya (1964) ve Watson (1964) daha basit halde düzenleme yaparak $g(X)$ tahmincisini eşitlik (2.20) haline dönüştürmüşlerdir. Bu düzenlemede koşullu beklenen değer fonksiyon tahmin etme problemine çözüm olarak çekirdek tahmin edicisini kullanmışlardır (Hardle, 1994, p.32).

$$\hat{g}_h(x) = \frac{\sum_{i=1}^n K_h(x - x_i) y_i}{\sum_{i=1}^n K_h(x - x_i)} \quad (2.20)$$

$g(X)$ tahmincisi, Loess yönteminde olduğu gibi belirli bir bölge için tahmin edici bulup ve daha sonra tüm etki alan için interpolasyon yöntemiyle tamamlanmaktadır. Bu sebeple yerel polinom çekirdek regresyonu olarak adlandırılmıştır. Çekirdek regresyon alanındaki literatüre göre çekirdek fonksiyonu seçimi düzleştirici parametresi seçimine göre daha önemsiz olduğu gösterilmektedir (Hastie ve Tibshirani, 1990, p.19).

2.2.3. R programlama dilinde çekirdek düzleştirici

R’da standart paketlerden $\{stats\}$ paketinde *ksmooth()* fonksiyonu, $\{ibr\}$ paketinde *npregress()* fonksiyonu, $\{fdm2id\}$ paketinde *KERREG()* fonksiyonu ve $\{KernSmooth\}$ paketinde *locpoly()* fonksiyonu kullanarak çekirdek regresyon uyumu elde edilebilmektedir. Her bir fonksiyonla veriler arasında optimal uyum olması için “bandwith” (düzleştirici parametresi) değerlerinin ve “çekirdek fonksiyon” larının ayarlanması gereklidir. Ancak bu yöntem sadece ve sadece tek açıklayıcı değişken olması durumunda kullanılmaktadır.

2.3. Splines

Uygulamalarda Spline, noktalar kümesi tarafından belirlenen düzgün bir eğri oluşturmak için kullanılan esnek bir yapıdır. Spline, dağıtılmış düğüm(knot) adı verilen sabit noktalar üzerinden elde edilen eğridir. Burada düğüm noktaları sırasıyla $\xi_1 < \xi_2 < \dots < \xi_k$ olmak üzere her bir düğüm noktaları arasında parçalı polinom fonksiyonu kullanılmaktadır. Genelde spline’ler üç sınıfa ayrılmaktadır. Bunlar sırasıyla Astronomi veri analizlerde en çok kullanılan spline interpolasyonu (interpolating spline), İstatistikçilerin dikkatini çeken düzleştirme yöntemlerinden spline regresyonu (spline regression) ve spline düzleştirici (Smoothing spline) dir. Burada sadece GAM modellerinde kullanılan spline regresyon ve spline düzleştiriciler ele alınacaktır.

2.3.1. Spline Regesyonu (Spline Regression)

Spline regresyonların kullanımları kolaydır ve diğer tahmin süreçlerine zorlanmadan dahil edilebildiklerinden dolayı çok tercih edilen bir yöntemdir. Regresyon spline’larda bilinmeyen g ’nin genel formu şu şekildedir.

$$g(x) = \sum_{i=1}^q b_i(x)\beta_i \quad (2.21)$$

Burada $b_i(x)$, i . taban fonksiyonu ve β_i ’ ler bilinmeyen parametrelerdir.

Regresyon spline’ların uyumu, parçalı polinomlar olarak temsil edilmektedir. Bu parçaları tanımlayan bölgeler bir dizi düğümle (knots) ayrılmaktadır. Dolayısıyla düzleştirici, verilen herhangi bir düğüm noktaların kümesi için uygun taban vektörler kümesinde çoklu regresyon ile hesaplanmaktadır. Bu vektörler taban fonksiyonlar (Basis

fuctions) olarak adlandırılmaktadır (Hastie ve Tibshirani, 1990, p.25). Kırpılmış güç serisi tabanları (Truncated power series) ve B-spline tabanları, spline regresyonlarında ki en çok kullanılan tabanlardır. Bunun yanında B-spline tabanları uygulamalarda daha iyi sonuç vermesi sebebiyle tercih edilmektedir. Örneğin, $\{\xi_1 < \dots < \xi_k\}$ kümesine iç düğümler kümesi ve daha basit olması için ξ_0 ve ξ_{k+1} iki sınır düğüm noktalar da eklemektedir. Parçalı kübik spline düzleştirici için kesik güç serisi taban fonksiyonları kullanarak $g(x)$ fonksiyonu parametrik olarak şu şekilde hesaplanmaktadır.

$$g(x) = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \sum_{j=1}^k \theta_j (x - \xi_j)_+^3 \quad (2.22)$$

Örneğin $a = x - \xi_j$; burada a_+ , a 'nın pozitif kısmını göstermek üzere $g(x)$ fonksiyonu aşağıdaki özelliklere sahiptir:

- i. g , $[\xi_j, \xi_{j+1})$ ' in herhangi bir aralıkta kübik polinomdur.
- ii. g 'nin sürekli türevi olması gerekir
- iii. g , ξ_1, \dots, ξ_k noktalarda atlayan bir adım fonksiyonu olan üçüncü türevine sahiptir.

Burada herhangi kullanabilecek bir fonksiyonu taban fonksiyonu olarak nitelendirmek için bu üç koşulun sağlanması gerekmektedir.

Eşitlik (2.22)'deki denklemin cebirsel avantajları olmasına rağmen regresyon spline hesaplamak için önerilen bir form değildir (Hastie ve Tibshirani, 1990, p.25). Eubank (1999, p.299), kırpılmış güç serisi tabanının kötü koşullu olma eğiliminden dolayı daha iyi koşullandırılmış diğer temel fonksiyonların kullanımının araştırılmasına ihtiyaç duyulmuştur. Bu özelliğe sahip olan tabanlardan biri de B-spline tabanlarıdır. B-spline tabanlar grafik modellemede yaygın olarak kullanılmaktadır. Regresyon spline'larıyla alakalı daha ayrıntılı bilgi Eubank (1999) çalışmasında mevcuttur.

Regresyon spline yöntemi uygulandığında herhangi bir taban fonksiyonu kullanılabilir. Ancak B-spline taban fonksiyonu seçildiğinde düğüm noktalarının sayısı ve konumlarının belirlenmesine ihtiyaç duyulacaktır. Geleneksel ve alışlagelmiş polinom terimleri, trigonometrik ve diğer parametrik dönüşümler kullanabilmektedir (Hastie ve Tibshirani, 1990, p.247).

Regresyon spline da düğüm noktaların sayısı ve konum seçimi sebebiyle bir dezavantaj oluşturmaktadır. Düğüm noktalarının sayısı az kullanıldığı zaman pürüzsüz,

yerel olmayan bazı uyumsuz davranışlar gösterebilmektedir (Hastie ve Tibshirani, 1990, p.26).

2.3.1.1. B-splines

B-spline taban fonksiyonu, kırpılmış güç serilerine göre sayısal olarak üstün alternatif bir tabandır ve matematiksel tanımı da oldukça basittir. B-spline tabanı oluşturulmak istenildiğinde öncelikle $m - 1$ dereceli spline fonksiyonu için mevcut olan $\Delta = \{\xi_1 < \xi_2 < \dots < \xi_k\}$ düğüm noktalarına

$$\xi_{-(m-1)} = \dots = \xi_{-1} = \xi_0 \quad (X \text{ bölgesindeki minimum})$$

$$\xi_{k+1} = \dots = \xi_{k+m} \quad (X \text{ bölgesindeki maksimum})$$

sağlayacak biçimde $2m$ düğüm noktaları $\xi_{-(m-1)}, \dots, \xi_{-1}, \xi_0, \xi_{k+1}, \dots, \xi_{k+m}$ eklenir.

Daha sonra da B-spline olan m dereceli ξ_1, \dots, ξ_k düğüm noktaları yeniden

$$N_{i,m}(x) = \frac{x - \xi_i}{\xi_{i+m-1} - \xi_i} N_{i,m-1}(x) + \frac{\xi_{i+m} - x}{\xi_{i+m} - \xi_{i+1}} N_{i+1,m-1}(x) \quad (2.23)$$

$i = -(m-1), \dots, k$ için ve

$$N_{i,1}(x) = \begin{cases} 1, & x \in [\xi_i, \xi_{i+1}), \\ 0, & \text{diğer yerlerde} \end{cases}$$

bu şekilde tanımlanabilmektedir (Eubank,1999,p.300). Üstelik tahmin edici elde etmek için verilen $N'_y N_y c = N'_y y$ normal denklemler c katsayı vektörüne göre çözülürse

$$N_y = \{N_{j,m}(x_i)\} i = 1, n, j = -(m-1), k \quad (2.24)$$

matrix oluşturabilmektedir.

B-spline tabanları oluşturmasıyla ilgi Eubank(1999), Boor (1978) çalışmalarında ayrıntılı bilgiler mevcuttur. B-spline tabanları ile spline regresyon fonksiyonu

$$g(x) = \sum_{i=1}^q N_{i,m}(x) \beta_i \quad (2.25)$$

biçimindedir.

2.3.1.2. Dügüm noktaları'nın seçimi

Tek değişkenli spline problemlerinde düğüm noktalarını seçmek için kullanılan en basit yöntem genelleştirilmiş kardinal spline stratejisidir (Hastie ve Tibshirani, 1990, p.247). Bu yöntem de her bağımsız değişken için p_j düğüm nokta sayısı seçilerek daha sonra eşit olarak bireysel aralıklarda yerleştirmektedir. Yani kübik B-spline tabanları kullanıldığında p_j düğümleri her j . bağımsız değişken için $k_j = p_j + 4$ taban fonksiyonlarına karşılık gelmektedir. Öyle ise her bağımsız değişken için kaç tane düğüm noktaları seçilmesi gerekir sorusuna önceden serbestlik derecesi belirleme veya genelleştirilmiş çapraz doğrulama GCV , C_p istatistik kriterleri veya Atılğan (1988) yaptığı gibi AIC kriteri kullanılmasıyla otomatik olarak yönlendirmektedir.

Stone ve Koo (1985) çalışmalarında biraz farklı bir yöntem incelemişlerdir. Bu yöntemde her bağımsız değişken (X_j)'nin beşinci en büyük ve en küçük değerine düğüm atanmaktadır. Ardından bunların arasında eşit aralıklarla yerleştirilmiş üç düğüm daha eklenmektedir. Düğüm sayısını ve konumunu optimize etmek için uyarlanabilecek yol üzerine bir algoritma ortaya çıkarmaktadır. Algoritmanın temel mantığı, düğüm yerlerini belirlemek için önceden kullanılan kübik taban fonksiyonlarının yerine parçalı doğrusal taban fonksiyonlarını kullanmayı tercih etmişlerdir. Bu yöntemde düğüm pozisyonları belirlendikten sonra her düğüm noktasında köşe yuvarlama yoluyla parçalı doğrusal fonksiyonları, parçalı kübik fonksiyonlarına dönüştürülmektedir (Friedman ve Silverman ,1989). Hastie ve Tibshirani (1990, p.250), çoklu durumlar için de süreçlerin değişmediğini göstermişlerdir.

Parametrik taban fonksiyonları kullanmanın avantajlarından dolayı B-spline tabanları seçimi akla uygun görülmesine karşılık düğüm noktaları belirleme stratejilerinde bazı olumsuz yönleri de bulunmaktadır. Bunun yanında B-spline ile ilgili bir başka sorun ise, serbestlik derecesi arttığında hem tüm düğüm nokta kümesini hem de otomatik olarak uyum da tutarsız bir şekilde değişmesidir. Bu durum düzleştirici spline'da ters olarak bulunmaktadır. Buradaki olumsuz durumu düzeltmek için düzleştirme parametresini değiştirmek yeterlidir. Dolayısıyla regresyon spline larının olumsuz yönleri bulunmaktadır (Hastie ve Tibshirani,1990,p.253)

2.3.1.3. R programlama dilinde spline regresyonu

R’da Spline regresyon fonksiyonu tahmininde hem B-splines(**bs**) hem de doğal tabanları (natural splines; **ns**) tercih edilebilmektedir. Ancak `{splines}` paketindeki **bs()** fonksiyonu en çok kullanılan olması sebebi, tüm spline’lar için taban fonksiyonların genel matrisini oluşturmasıdır. Dolayısıyla **lm()** ve **gam()** modellerinde onu alt fonksiyon olarak kullanabilmektedir. Ancak **bs(...,df, knots,degree,)** argümanları ayarlamakla beraber optimal düzleştirici fonksiyonu elde edilir. Fonksiyonların varsayımlarda **bs()** fonksiyonu iki sınır düğümü ve verinin tam medyan noktasından bir iç düğüm ile kübik B-spline tabanları oluşturmaktadır.

2.3.2. Düzleştirici Spline (Smoothing Spline)

Düğüm noktalarının yerleştirilmesi problemi regresyon spline’nın temel dezavantajlarından biridir. Bu problemi ortadan kaldırmak için verinin tüm gözlemleri düğüm noktaları olarak sayılabilmektedir ve hata kareler ortalamasına düzleştirici cezası eklenir. Buna da düzleştirici spline veya Cezalı spline (Penalized spline) adı verilmektedir. Eşitlik (2.1)’de \hat{g} tahmininde hata kareler ortalamasının (Mean Squares Error, *MSE*) en küçük olması istenilmektedir.

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{g}(x_i))^2 \quad (2.26)$$

Bu durumda da tüm noktalar birleştirildiğinde aşırı uyum (overfitting) problemiyle karşılaşılacaktır. Dolayısıyla bunun yerine \hat{g} tahmininde modifiye edilmiş en küçük kareler kriteri kullanılır ve

$$\frac{1}{n} \sum (Y_i - \hat{g}(x_i))^2 + \lambda \int [\hat{g}''(x)]^2 dx \quad (2.27)$$

biçiminde ifade edilir. Burada düzleştirici parametresi ($\lambda > 0$) ve tahmin edicinin esnekliği sağlanmaktadır. Loess yöntemindeki span rölünü eşitlik (2.27) deki λ parametresi üstlenmektedir. Yine eşitlik (2.27) deki $\int [\hat{g}''(x)]^2 dx$ ifadesi aşırı uyum cezasıdır (Roughness penalty) ve $\int_0^1 [\hat{g}''(x)]^2 dx$ biçiminde de olabilir (Wood, 2017, p.126). Üstelik g aşırı uyumluyken cezası çok büyüktür ancak aşırı uyumdan kurtuldukça cezası da küçülecektir. Dolayısıyla kriterin ikinci kısmı aşırı uyuma karşı denge uyumu

sağlamaktadır. Bu da literatürde spline düzleştirici uyumu (smoothing spline fitting) olarak adlandırılmaktadır.

Faraway (2006, p.239), aşırı uyum cezasın da optimal çözüme \hat{g} 'yi kübik spline olarak elde etmiştir. Yani $X_{(i)}$ 'ler sıralı değerler olmak üzere \hat{g} tahmin fonksiyonu her $(x_{(i)}, x_{(i+1)})$ aralığında parçalı kübik polinomdur ve özellikle \hat{g} , \hat{g}' ve \hat{g}'' türevleri süreklidir. Düzleştirici spline'larda taban değişikliği de mümkündür.

Regresyon spline'lar ve düzleştirici spline'lar arasında ufak bir farklılık bulunmaktadır. Regresyon spline'da B-spline'nın kullandığı düğüm noktaları örneklem boyutundan daha az olmasına rağmen seçilen düğüm sayısı da aşırı uyumu kontrol etmektedir. Buna karşılık düzleştirici spline'lar da ise tüm gözlemlenen X değerleri otomatik biçimde düğüm noktaları olarak alınmakta ve aşırı uyum kontrolünde de λ 'yi kullanmaktadır (Faraway, 2006, p.239).

2.3.2.1. Düzleştirici spline tabanları

Literatürde düzleştirici spline'lar için bir ceza terimi içeren tabanlar daha çok tercih edilmektedir. Bunlar içerisinde en yaygın olanları; kübik(cubic) spline, p-spline ve ince tabakalı (thin plate) spline tabanlardır.

2.3.2.1.1. Kübik spline

Kübik spline tabanları doğal olarak (2.26) eşitliğinde optimal düzleştirici belirtilmesinde yardımcı olmaktadır. Yani kübik spline regresyon: model uyumunu, pürüzsüzlüğünü ve taban fonksiyonunun önceden seçilmemesinin yanında (2.26) eşitliğini en küçükleyen fonksiyonun aranılmasıyla ortaya çıkmaktadır. Genel olarak kübik spline tabanlarıyla optimal $g(x)$ fonksiyonu aşağıdaki biçimde hesaplanmaktadır.

$$g(x) = \frac{x_{j+1} - x}{h_j} \beta_j + \frac{x - x_j}{h_j} \beta_{j+1} + \frac{\frac{(x_{j+1} - x)^3}{h_j} - h_j(x_{j+1} - x)}{6} \omega_j + \frac{\frac{(x - x_j)^3}{h_j} - h_j(x - x_j)}{6} \omega_{j+1}, \quad x_j \leq x \leq x_{j+1} \quad (2.28)$$

Burada $\frac{x_{j+1} - x}{h_j}$ kübik spline taban fonksiyonudur ve

$h_j = x_{j+1} - x_j$; $\theta_j = g(x_j)$; $\omega_j = g''(x_j)$ olarak ispatlanmaktadır (Asfha, 2017).

2.3.2.1.2. P-spline

B-spline tabanı, özellikle büyük ölçekli spline enterpolasyonu için geliştirilen bir tabandır. Ancak çoğu istatistiksel çalışmalarda küçük mertebede cezalı regresyon spline'lar iyi bir sonuç vermemektedir. Bu problemi çözmek için daha gelişmiş ve sağlam tabanlara ihtiyaç duyulmaktadır. Dolayısıyla uygulamalarda (Eilers ve Marx, 1996) B-spline tabanları modifiye edilerek p-spline tabanları geliştirilmiştir. P-spline'lar cezalı B-spline tabanları anlamına gelmektedir. P-spline'lar, B-spline'lara göre küçük dereceli düzleştiriciler olarak tanımlanması yanında eşit aralıklı düğümler ve taban fonksiyonlarını kontrol etmek için parametrelere doğrudan farklı ceza uygulanmaktadır. Eşitlik (2.25)'de β_i ' ler kareler farkını cezalandırmak için cezayı şu şekilde

$$p = \sum_{i=1}^{k-1} (\beta_{i+1} - \beta_i)^2 = \beta_1^2 - 2\beta_1\beta_2 + 2\beta_2^2 - 2\beta_2\beta_3 + \dots + \beta_k^2, \quad (2.29)$$

ve matris biçiminde

$$p = \boldsymbol{\beta}^T \begin{bmatrix} 1 & -1 & 0 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ -1 & 2 & -1 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & -1 & 2 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{bmatrix} \boldsymbol{\beta} \quad (2.30)$$

olarak elde edilebilmektedir Wood(2017, p.150). Bu tür cezaların oluşturulması R programlama dilinde çok kolaydır. P-spline tabanları oluşturulması ve kullanımının kolay olmasının yanında daha fazla esneklik sağlamaktadır. Dolayısıyla P-spline'lar; esas olarak B-spline tabanlarla beraber herhangi bir ceza birleştirilebilecek bir tabandır. P-spline'lar esnek bir taban olmalarına rağmen düzensiz düğüm aralığı olduğu durumlarda basitliğin bir miktar azalması gibi olumsuz tarafları bulunmaktadır.

2.3.2.1.3. İnce tabakalı spline

İnce tabakalı spline'lar özellikle çoklu düzleştirici fonksiyonların tahmin problemlerini çözmek için geliştirilmiştir (Duchon,1977) ve kullanımında her taban için düğüm konumlarını seçmek gerekirken; tabanlar da sadece her bağımsız değişken temsil

etmek için kullanılabilir. Genel olarak (2.1) eşitliğindeki $g(x_i)$ optimal düzleştirici fonksiyonu tahmini de

$$\|y - \hat{g}\|^2 + \lambda f_{md}(g) \quad (2.31)$$

biçimindedir. Burada $g = (g(x_1), g(x_2), \dots, g(x_n))^T$, λ düzleştirici parametresi ve $f_{md}(g)$ ise g 'nin pürüzsüzlüğünün cezalı bir fonksiyondur. Örneğin, iki bağımsız değişkenin ikinci türevleri kullanmakla beraber $m = 2$ ve $d = 2$ olacak

$$f_{22} = \iint \left(\frac{\partial^2 g}{\partial x_1^2} \right)^2 + \left(\frac{\partial^2 g}{\partial x_1 \partial x_2} \right)^2 + \left(\frac{\partial^2 g}{\partial x_2^2} \right)^2 dx_1 dx_2 \quad (2.32)$$

biçiminde hesaplanmaktadır Wood(2017, p.150). İnce tabakalı spline tabanları, pürüzsüzlüğün tam olarak ne demek olduğunu tanımlamak, optimal düzleştirme fonksiyonuna ne kadar ağırlık vermesinin yanında hedefine göre en iyi düzleştirici fonksiyonu bulmak ve cezaları modeldeki tüm düzgün terimleri temsil etmek için de kullanılmaktadır. Ancak hesaplamalarda birçok bilinmeyen parametreler sahip bir düzleştirici fonksiyonuyla karşılaştırabilmektedir. Bu konu ile alakalı ayrıntılı bilgiler Wood (2017, p.151) çalışmasında yer almaktadır.

2.3.2.2. Düzleştirici spline hesaplanması

Genel olarak eşitlik (2.26)'daki denklem çözümünde $n - 2$ iç düğümlü olan doğal kübik spline (Natural cubic spline) taban kullanılmaktadır. Ancak hesaplama daha kolay olması için B-spline tabanı kullanımı tercih edilmektedir ve aşağıdaki şekilde oluşturulur.

$$g(x) = \sum_{j=1}^{n+2} \gamma_j B_j(x) \quad (2.32)$$

Burada γ_j katsayıları ve B_j ise kübik B-spline taban fonksiyonlarını ifade etmektedir.

2.3.2.3. R programlama dilinde düzleştirici spline

Düzleştirici spline yöntemiyle uyumu R programlama dilinde `{stats}`, `{npreg}`, `{gam}` ve `{mgcv}` paketlerinde sırasıyla bulunan **`smooth.spline()`**, **`ss()`** ve **`s()`** fonksiyonları da kullanabilmektedir. Bu fonksiyonlar, farklı özelliklere sahip olmalarına rağmen kullanımlarından elde edilen hata kareler ortalamaları birbirine yakındır. Her bir fonksiyon için de düzleştirme parametresi ayarlayabilmektedir.



3. GENELLEŞTİRİLMİŞ TOPLAMSAL MODELLERİ

Genelleştirilmiş toplamsal modelleri, toplamsal modelleri ve genelleştirilmiş doğrusal modellerinin birleşimi olarak tanımlanmaktadır. Sırasıyla bu modeller alt kısımda tanıtılacaktır.

3.1. Toplamsal Modelleri

Uygulamalarda doğrusal modellerin birçok dönüşüm yöntemlerinin olmasının yanında özellikle her bağımsız değişken için ayrı dönüşüm fonksiyonu bulunması ve uygun modeli elde etmek genellikle zahmetli hale gelmektedir. Bu problemleri çözmek için toplamsal modelleri, parametrik varsayımlar olmadan ve aynı anda en iyi dönüşümler elde etmek için kullanılabilir. Toplamsal modelleri, 1984 yıllarında ilk defa Stone tarafından geliştirilmiştir. Genel olarak toplamsal modelleri eşitlik (3.1) deki biçimde tanımlanmaktadır.

$$Y = \alpha + \sum_{j=1}^p g_j(x_j) + \varepsilon \quad (3.1)$$

Burada $E(\varepsilon) = 0$ ve $var(\varepsilon) = \sigma^2$ bağımsız hata terimleridir ve x_j ' ler $[0,1]$ aralığında değer aldığı varsayılmaktadır. g_j ' ler ise her bağımsız değişken için keyfi bir doğrusal düzleştirici fonksiyonudur (Stone, 1984, p.1)

Fakat modelde güçlü etkileşim olduğu durumlarda toplamsal modelleri iyi uyum göstermemektedir (Faraway, 2006, p.254). Dolayısıyla değişken sayısının yeteri kadar olması durumunda $g_{ij}(x_i, x_j)$ şeklinde modele eklemeler düşünülebilmektedir. Bu sayede kategorik değişkenler, model içerisine kolayca uyum sayılabilecektir. Bu uyum için

$$Y = \alpha + \sum_{j=1}^p g_j(x_j) + \gamma Z + \varepsilon \quad (3.2)$$

kullanılabilir (Faraway, 2006, p.225). Burada Z toplamsal olarak modellenmeyecek değişkenlerin tasarım matrisi, elemanları nitel ve nicel olarak karmaşık olabilmektedir. γ ' lar ise regresyon katsayılarıdır.

R programlama dilinde üç farklı toplamsal modelleri uydurulabilmektedir. Birinci olarak Hastie ve Tibshirani (1990) taraflarından geliştirilen $\{gam\}$ paketi. İkinci olarak

Simon Wood (2000) tarafından geliştirilen $\{mgcv\}$ paketidir. Ancak $\{gam\}$ paketinde düzleştiricilerin seçeneği daha fazla iken $\{mgcv\}$ paketin de ise bu işlemler otomatik olarak hesaplatılmaktadır. Üçüncü ise Gu (2013) tarafından geliştirilen spline tabanlı bir yaklaşım $\{gss\}$ paketidir.

Genelde toplamsal modellerin uyum algoritması seçilen paketler ile alakalıdır. Faraway (2006) $\{gam\}$ paketinde *backfitting* algoritmasını kullanmıştır.

3.2. Genelleştirilmiş Doğrusal Modelleri

Genelleştirilmiş Doğrusal modelleri (*GLM*) geleneksel doğrusal modellerin bir genellemesidir. Spesifik olarak bağımsız değişkenlerin etkileri parametrelerde doğrusal olması varsayılmaktadır. Ancak sadece bağımlı değişkeninin dağılımı ile bağımsız değişkenlerin arasında bağlantı olabilmektedir. Genel olarak *GLM*'nin iki temel özelliği vardır. Birincisi bağımlı değişkenin dağılımının üstel aileye sahip olması, ikincisi ise bağımlı değişkeninin ortalamasıyla doğrusal kombinasyonu olan bağımsız değişkenlerin arasında ilişkiyi açıklayan bağlantı fonksiyonu kullanılmasıdır.

Üstel ailesi genel olarak şu şekildedir.

$$g(y|\theta, \Phi) = \exp \left[-\frac{y\theta - b(\theta)}{a(\Phi)} + c(y, \Phi) \right] \quad (3.3)$$

Burada θ kanonik katsayısıdır ve konum parametresini temsil eder. Φ Parametresi ise yayım parametresi ve ölçeği temsil eder. Ailenin farklı üyelerini belirtmek için a , b ve c fonksiyonlarını tanımak gerekir. Dolayısıyla örneklerde en çok ortak kullanılan üstel aileye sahip 3 dağılım:

i. Normal veya Gaussian

$$\begin{aligned} g(y|\theta, \Phi) &= \frac{1}{\sqrt{2\pi\sigma}} \exp \left[-\frac{(y - \mu)^2}{2\sigma^2} \right] \\ &= \exp \left[\frac{y\mu - \mu^2/2}{\sigma^2} - \frac{1}{2} \left(\frac{y^2}{\sigma^2} + \log(2\pi\sigma^2) \right) \right] \end{aligned}$$

Burada $\theta = \mu$, $\Phi = \sigma^2$, $a(\Phi) = \Phi$, $b(\theta) = \frac{\theta^2}{2}$ ve $c(y, \Phi) = -\left(\frac{y^2}{\Phi} + \log(2\pi\Phi)\right)/2$

ii. Poisson

$$\begin{aligned} g(y|\theta, \Phi) &= e^{-\mu} \mu^y / y! \\ &= \exp(y \log(\mu) - \mu - \log(y!)) \end{aligned}$$

$\theta = \log(\mu)$, $\Phi \equiv 1$, $a(\Phi) = 1$, $b(\theta) = \exp(\theta)$ ve $c(y, \Phi) = -\log(y!)$ olarak yazabilmektedir.

iii. Binom

$$\begin{aligned} g(y|\theta, \Phi) &= \binom{n}{y} \mu^y (1 - \mu)^{n-y} \\ &= \exp(y \log \mu + (n - y) \log(1 - \mu)) + \log \binom{n}{y} \\ &= \exp \left(y \log \frac{\mu}{1 - \mu} + n \log(1 - \mu) \right) + \log \binom{n}{y} \end{aligned}$$

$$\theta = \log \frac{\mu}{1 - \mu}, \quad b(\theta) = -n \log(1 - \mu) = n \log(1 + \exp \theta)$$

ve

$$c(y, \Phi) = \log \binom{n}{y}$$

olarak görülmektedir.

Gamma ve Ters Gaussian dağılımları üstel aile olarak daha az kullanılan dağılımlardır. Not olarak Φ parametresi dağılımın türüne göre farklı isimler almaktadır. İki parametresi olan Normal ve Gamma dağılımları için bu parametre serbest iken, tek parametreye sahip Poisson ve Binom dağılımları için bu parametre 1 olarak alınmaktadır. Ancak üstel aileye sahip olmayan dağılımlar da *GLM* – tarzı (*GLM* – Style) yaklaşımla üstel ailesine dönüştürüle bilinir.

Link Fonksiyonu (Bağlantı fonksiyonu)

Aşağıdaki model doğrusal modeli olarak varsaymaktadır.

$$\eta = X^T \beta ; \eta = g(\mu) \tag{3.4}$$

Link Fonksiyonu g , bağımlı değişken ortalaması $EY = \mu$ ile doğrusal bağımsız değişkenler arasında nasıl bir bağlantı olduğunu açıklamaktadır. Kanonik bağlantı fonksiyonu dağılım türüne göre farklı şekilde hesaplanmaktadır.

Tablo 3.1 GLM için kanonik bağlantılar

Aile (<i>Family</i>)	Bağlantı (<i>Link</i>)	Varyans Fonksiyonu
Normal	$\eta = \mu$	1
Poisson	$\eta = \log \mu$	μ
Binomial	$\eta = \log \left(\frac{\mu}{1 - \mu} \right)$	$\mu(1 - \mu)$
Gamma	$\eta = \mu^{-1}$	μ^2
Ters Gaussian	$\eta = \mu^{-2}$	μ^3

Faraway (2006, p.128), kanonik bağlantı fonksiyonu kullanıldığında $X^T Y$, β için yeterli olduğunu göstermiştir. Genelde kanonik bağlantı fonksiyonları matematiksel ve hesaplama açısından uygun olması da doğal seçimdir. Fakat kanonik bağlantı fonksiyonu kullanmak zorunlu değildir. Dolayısıyla başka bağlantılar da kullanabilmektedir.

GLM’de, en çok olabilirlik tahmin yöntemi (Maximum Likelihood) kullanılarak β parametreleri tahmin edilebilmektedir. Burada $a(\Phi) = \Phi/w$ iken tek değişkenli en çok olabilirlik tahmini şu şekildedir.

$$\log L(\theta_i, \Phi; y_i) = w_i \left[\frac{y_i \theta_i - b(\theta_i)}{\Phi} \right] + c(y_i, \Phi) \quad (3.5)$$

Bağımsız değişkenler için en çok olabilirlik tahmini $\sum_i \log L(\theta_i, \Phi; y_i)$ olarak bulunmaktadır.

3.3. GAM Modelleri

Hastie ve Tibshirani (1990) tarafından geliştirilen Genelleştirilmiş Toplamsal Modelleri (Generalized Additive Models- *GAM*), çoklu regresyon fonksiyonunu tahmin etmek amacıyla doğrusal düzleştirici yöntemlerini kullanır ve geleneksel GLM’nin

parametrik olmayan bir çeşiddir (Nathalie, 2005, p.32). Genel olarak modelin yapısı şu şekildedir (Wood, 2006)

$$Y_i = f(\mu_i) = X_i^* \theta + g_1(x_{1i}) + g_2(x_{2i}) + g_3(x_{3i}, x_{4i}) + \dots \quad (3.6)$$

Burada $\mu_i = \mathbb{E}(Y_i)$ ve Y_i , üstel aile dağılımlarına sahip bağımlı değişkendir. X_i^* Bağımsız değişkenlerin parametrik kısmının tasarım matrisinin i . satırını ve θ da onun vektör parametresini ifade etmektedir. g_i Bağımsız değişkenlerin düzleştirici fonksiyonları ve $f(\mu_i)$ ise bağlantı fonksiyonudur. *GAM*'lar özel olarak esnek bir özelliğe de sahiptirler. Ayrıntılı olarak parametrik ilişkiler yerine sadece düzleştirici fonksiyonları da kullanılabilmektedir. Bu durum eşitlik (3.7) de ifade edilmiştir.

$$f(\mu_i) = \alpha + \sum_{j=1}^p g_j(x_j) + \varepsilon \quad (3.7)$$

Bu esneklik literatürde iki türlü kullanılmaktadır. Birincisi doğrusal düzleştirici fonksiyonunu temsil etmesi ve ikincisi ise optimal düzleştirme derecesini seçmesidir (Wood, 2006, p.119).

3.3.1. GAM Algoritması

Literatürde GAM modelleri tahmininde yerel regresyon, çekirdek regresyon, regresyon spline ve düzleştirici spline yaklaşımlarının kullanımları önerilmiştir. *GAM* modellerin amacı g_j fonksiyonunu aynı anda parametrik olmayan şekilde tahmin etmektir. Bu fonksiyonlar bireysel doğrusal olmayan form veya güçlü etkileşim olduğu durumlarda da birleşim formu biçiminde de olabilmektedir. Modelde her g_j için farklı bir düzleştirici kullanılabilmektedir. Genel olarak *GAM* modelleri “*backfitting*” algoritmasıyla tahmin edilmektedir. Bu algoritmayı ilk olarak Friedman ve Stuetzle (1981) önermişlerdir. Hastie ve Tibshirani (1990, p.90), bu algoritmayı *GAM* modellerinde ilk olarak uygulamışlardır. Bu algoritma Hastie ve Tibshirani (1990, p.91) ve Faraway (2006, p.255) de verilmiştir. Bu algoritma içerisinde yer alan $S_j(y|x_j)$, y 'e karşı x_j düzleştirici fonksiyonu olmak üzere araştırmacılar için S_j seçimi serbest bırakılmıştır. Dolayısıyla S_j Splayn ve Loess gibi parametrik olmayan düzleştiriciler veya doğrusal ve polinom gibi de parametrik form şeklinde olabilmektedir.

3.3.2. R programlama dilinde GAM

Genel olarak literatürde *GAM* modelleri uydurmasında en çok kullanılan yazılımların başında R ve S-PLUS yazılımları gelmektedir. R programlama dilinde *GAM* modelleri için $\{gam\}$ ve $\{mgcv\}$ paketleri mevcuttur. Her bir paket içerisinde, özellikleri farklı olan **gam** fonksiyonları bulunmaktadır. İlk olarak $\{gam\}$ paketinde bulunan **gam** fonksiyonu “backfitting” algoritmasını kullanarak fonksiyona ait “formula” parametresinde loess (*lo*), smoothing spline (*s*), B-spline(*bs*) ve natural spline(*ns*) gibi farklı düzleştirici terimlerini kullanmaktadır ve kullanılan bu fonksiyonlar Tablo 3.2’de yer almaktadır. Düzleştirici spline ile GAM modelleri uydurmasında literatürde taban fonksiyonları seçimi yerine serbestlik derecesi kullanımı tercih edilmektedir. Sembolik *lo* fonksiyonu için varsayılan değerler “span” parametre değeri için 0.5 ve polinom derecesi “degree” parametresinde 1 olarak çalışmaktadır. Sembolik *s* fonksiyonu için serbestlik derecesi 1’den büyük olması ve “spar” parametre değerinin (0,1] aralığında olması gerekmektedir.

Tablo 3.2. R programın da kullanılan $\{gam\}$ paketi ve fonksiyonları

Yöntemler	R Paketleri ve fonksiyonları	Yardımcı Fonksiyonlar
Loess	$\{gam\}$ gam	lo (...,span=0.5, degree=1)
Smoothing Spline	$\{gam\}$ gam	s (x, df=4, spar=1)
Regression Spline B-Spine	$\{gam\}$ gam	bs (x, df = NULL, knots = NULL, degree = 3, intercept = FALSE, Boundary.knots = range(x))
Regression Spline Natural Spline	$\{gam\}$ gam	ns (x, df = NULL, knots = NULL, intercept = FALSE, Boundary.knots = range(x))

İkinci olarak $\{mgcv\}$ paketinde “en çok olabilirlik(likelihood)” algoritmasını kullanan **gam** fonksiyonudur. Bu fonksiyonun “formula” parametresi, yalnızca düzleştirici spline(*s*) için geliştirilmiş ve taban fonksiyonlarını doğrudan seçebilmesinin yanında sadece genelleştirilmiş çapraz doğrulama(*GCV*) aracılığıyla otomatik düzleştirme parametresi seçimi yapabilmektedir. **s()** düzleştirici fonksiyonu içerisinde varsayımlara göre ayarlanmış seçenekler bulunmaktadır. İlk olarak düzleştirici terimleri temsil eden taban fonksiyonu ince tabakalı tabanlardır. Bunun arkasında cezalı kübik regresyon spline **s(...,bs="cr")** veya cezalı spline **s(...,bs="ps")** gibi tabanlar tercih

edilebilmektedir. Üstelik düzleştirici terimleri temsil etmede taban fonksiyonlarının boyutu(b) seçilebilmektedir. Varsayılan olarak $b=10$ olarak kullanılmaktadır. Optimal taban boyutu, maksimum serbestlik derecesi elde etme anlamına gelmektedir. Son olarak aşırı uyuma karşı “ γ ” parametresi 1.4 olarak da seçilebilmektedir. Wood (2017, p.220), modelde farklı tabanlar kullanımının sorun olmadığını ifade etmiştir. Ayrıca çoklu tahmin edicilerin düzleştiricileri ve özellikle farklı birimlerde ölçülen ortak değişkenlerin veya değişkenler arasında büyük etkileşim olduğu durumlarda ise sırasıyla $te()$ ve $ti()$ fonksiyonları kullanılabilir. Üstelik $t2()$ fonksiyonu, $te()$ fonksiyonu için bir alternatif olarak da seçilebilmektedir.

Tablo 3.3. R programında kullanılan {*mgcv*} paketi ve fonksiyonları

Yöntemler	R paketleri ve fonksiyonları	Yardımcı fonksiyonlar
Düzleştirici spline (Smoothing Spline)	{ <i>mgcv</i> } gam	$s(..., k = 1, fx = F, bs = "tp", m = NA, by = NA, xt = NULL, id = NULL, sp = NULL, pc = NULL)$
tensor product smooths	{ <i>mgcv</i> } gam	$te(..., k=NA, bs="cr", m=NA, d=NA, by=NA, fx=F, np=T, xt=NULL, id=NULL, sp=NULL, pc=NULL)$
tensor product interaction	{ <i>mgcv</i> } gam	$ti(..., k = NA, bs = "cr", m = NA, d = NA, by = NA, fx= F, np = T, xt=NULL, id=NULL, sp = NULL, mc=NULL, pc=NULL)$
tensor product smooth	{ <i>mgcv</i> } gam	$t2(..., k=NA, bs="cr", m=NA, d=NA, by=NA, xt=NULL, id=NULL, sp=NULL, full=F, ord=NULL, pc=NULL)$

3.3.3. Aykırılık

Aykırı değer(outlier), veri setinde bulunan diğer gözlem değerlerinden önemli ölçüde farklılık gösteren bir veri noktası olarak düşünülebilir. Aykırı değerler, istatistik literatüründe uyumsuzluklar, sapmalar veya anormallikler olarak da adlandırılır. Aykırı değerlerin varlığı örneklem ya da ölçme hataları en başta olmak üzere birçok nedenden kaynaklanabilir. Bir aykırı değer, istatistiksel analizlerde ve modellemelerde ciddi sorunlara neden olabilir, bu nedenle aykırı değerlerin tespiti önem arz etmektedir. Çoğu uygulamada veriler, sistemdeki aktiviteyi veya varlıklar hakkında toplanan gözlemleri yansıtabilen bir veya daha fazla üretici mekanizma tarafından oluşturulur. Oluşturma

süreci olağandışı bir şekilde davrandığında, aykırı değerlerin yaratılmasıyla sonuçlanır. Bu nedenle, bir aykırı değer genellikle, veri oluşturma sürecini etkileyen sistemlerin ve varlıkların anormal özellikleri hakkında faydalı bilgiler de içerebilmektedir. Bu tür olağandışı özelliklerin tanınması, uygulamaya özel yararlı bilgiler de sağlayabilir. Bunlara örnek olarak kanserli hücrelerin tespiti, bilgisayar sistemlerinde virüslerin tespiti, güvenlik olarak erken uyarı sistemleri söylenebilir.



4. UYGULAMA

Bu kısımda bölüm 2 de anlatılan yöntemlerin, Monte-Carlo Simülasyonu ile 2000 lik denemelerle farklı bozulma durumlarındaki performansları incelenecektir. Simülasyonlarda R programlama dilinden faydalanılmıştır ve bu kodlar Ek-2’te yer almaktadır. Bu uygulamada Tablo 4.1 de yer alan 3 ayrı modelde ve sırasıyla,

Örneklem hacimleri: 40, 80, 100 ve 500

Bozulma oranları: %5, %10, %20 ve %40

Bozulma şiddetleri: 5σ ve 20σ

Bozulma bölgeleri: pozitif, negatif ve merkez

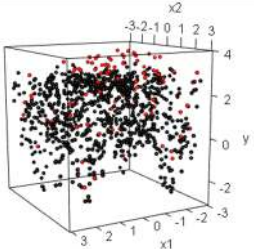
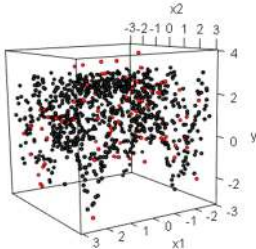
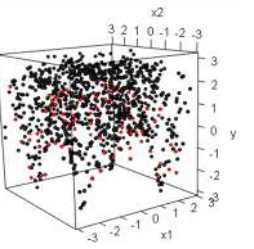
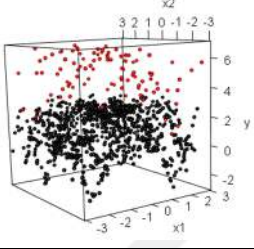
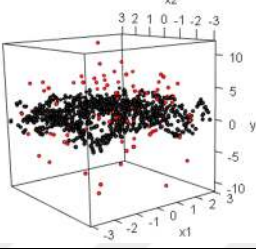
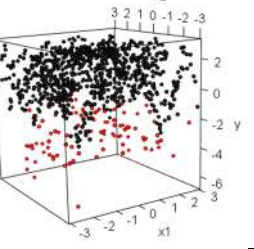
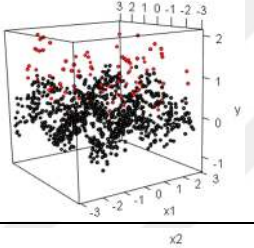
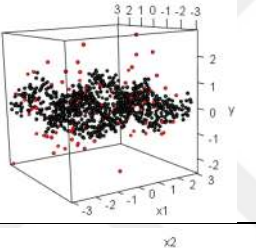
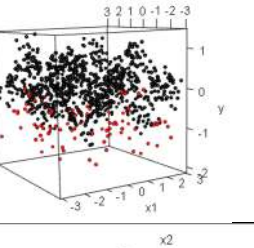
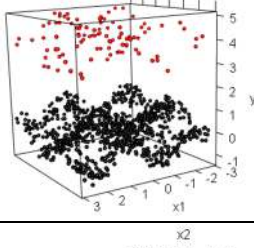
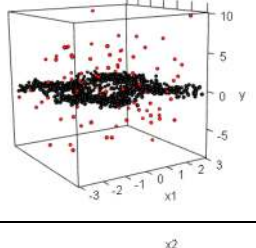
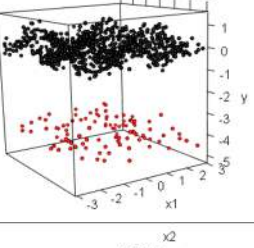
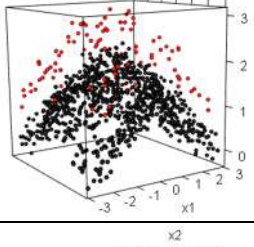
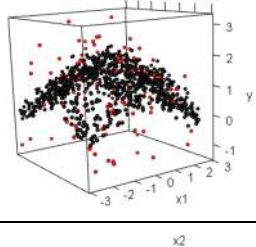
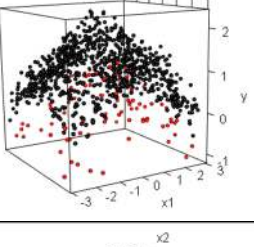
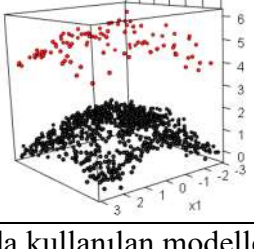
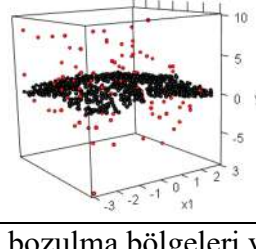
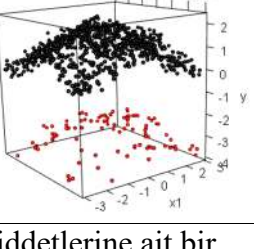
alınarak toplamda 12’si normal ve 288’i bozdurulmuş veri seti olmak üzere 300 farklı durum göz önünde bulundurulmuştur. Simülasyonlarda kullanılan veri setleri içerisinde modellere ait bozulma şiddetleri ve bölgelere göre 18 farklı veri setinin görünümü Şekil 4.1 de verilmiştir.

Model fonksiyonlarında normal hata terimi $\epsilon \sim N(0, \sigma)$ iken bozdurulmuş hata terimlerin de bozulma şiddeti k olmak üzere bozulma bölgelerinde (pozitif, merkez ve negatif) sırasıyla hata terimlerinin dağılımları $N(k\sigma, \sigma)$, $N(0, k\sigma)$ ve $N(-k\sigma, \sigma)$ olarak ele alınmıştır.

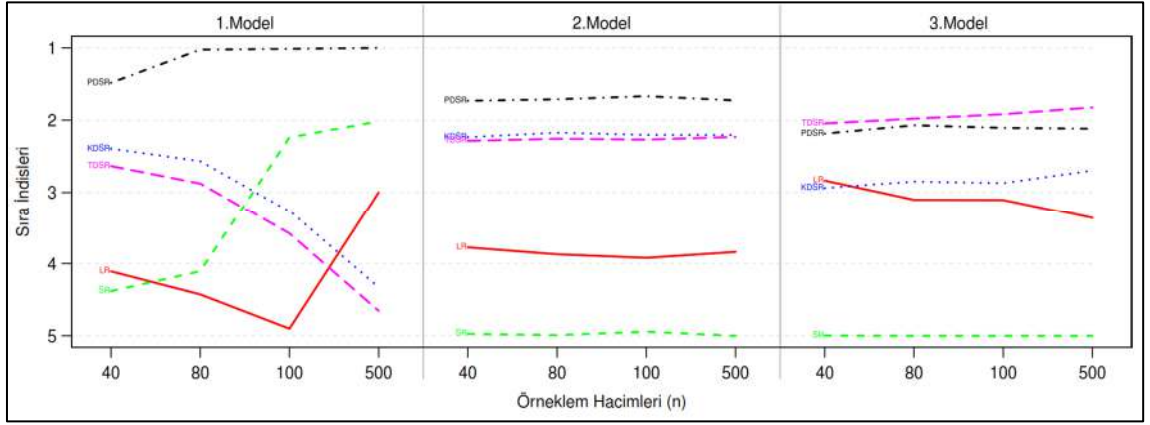
Yöntemlerde band genişlikleri k-fold çapraz sorgulama ile elde edilmiş olup performans değerlendirmelerinde model de aşırı uyumu önlemek amacıyla verilerin %80’i eğitim, %20’si test için kullanılmıştır. Burada asıl amaç 5 yöntemin farklı aykırı değerlerin varlığı durumlarında ortaya koyacakları performansları incelemek olduğundan, performans değerlendirmelerinde her bir simülasyon sonrası bu yöntemlerin hata kareler ortalama değerlerine göre iyilik sıralamaları dikkate alınmıştır. Elde edilen sonuçlar da bu sıralamalara göre değerlendirildi ve simülasyonlardan elde edilen hata kareler ortalamaları ve standart sapmaları Ek-1 bölümünde verilmiştir.

Tablo 4.1. Simülasyonlar da kullanılan model fonksiyonları

Model Numarası	Model Fonksiyonları
1	$y = \sin(x_1^2) + \cos(x_1^2) + \sin(x_2^2) + \cos(x_2^2) + \epsilon$
2	$y = \sin(x_1) * \cos(x_1) + \sin(x_2) * \cos(x_2) + \epsilon$
3	$y = (\sin(x_1)/x_1)^2 + (\sin(x_2)/x_2)^2 + \epsilon$

Model Numarası	Bozulma Şiddeti(k)	Pozitif Bölge	Merkez Bölge	Negatif Bölge
1	5			
	20			
2	5			
	20			
3	5			
	20			

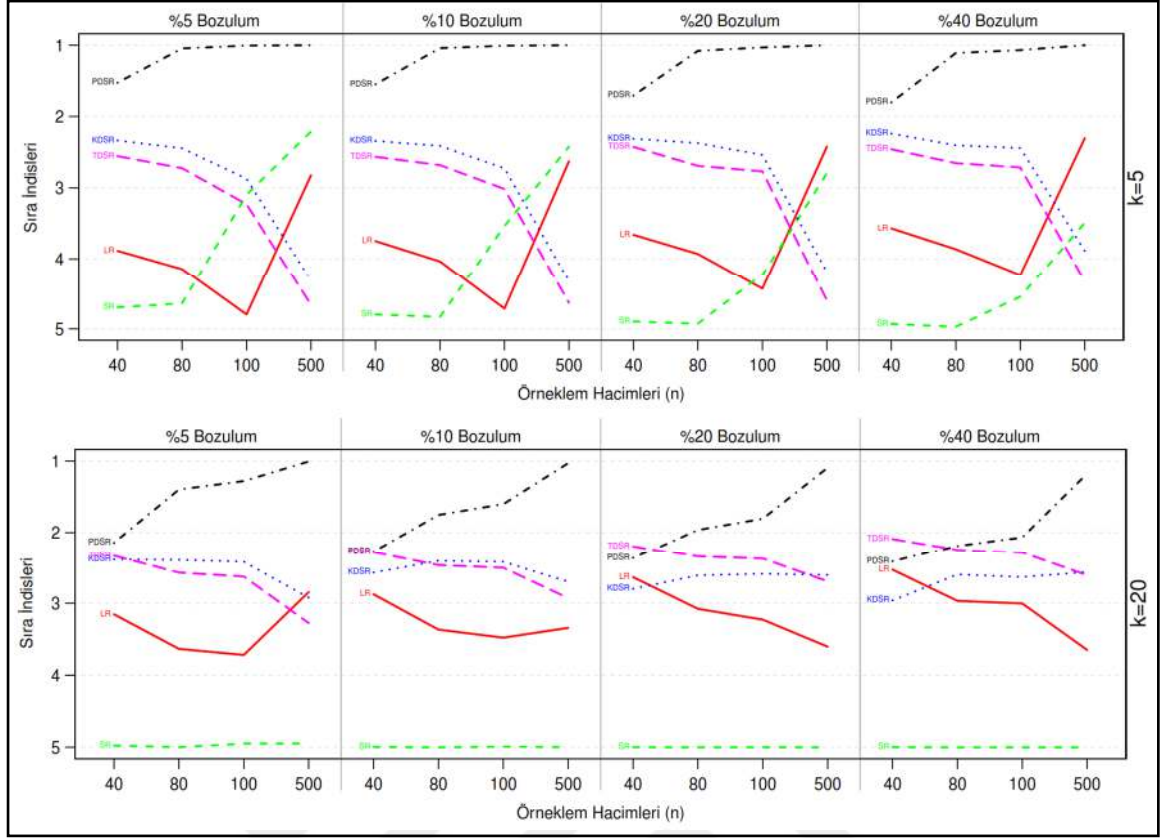
Şekil 4.1. Senaryolarda kullanılan modellerin bozulma bölgeleri ve şiddetlerine ait bir veri setinden görüntüler



Şekil 4.2. Bozdurulmamış yapay veriler de yöntemlerin modellere göre performansları

Şekil 4.2. de verilen grafikler de:

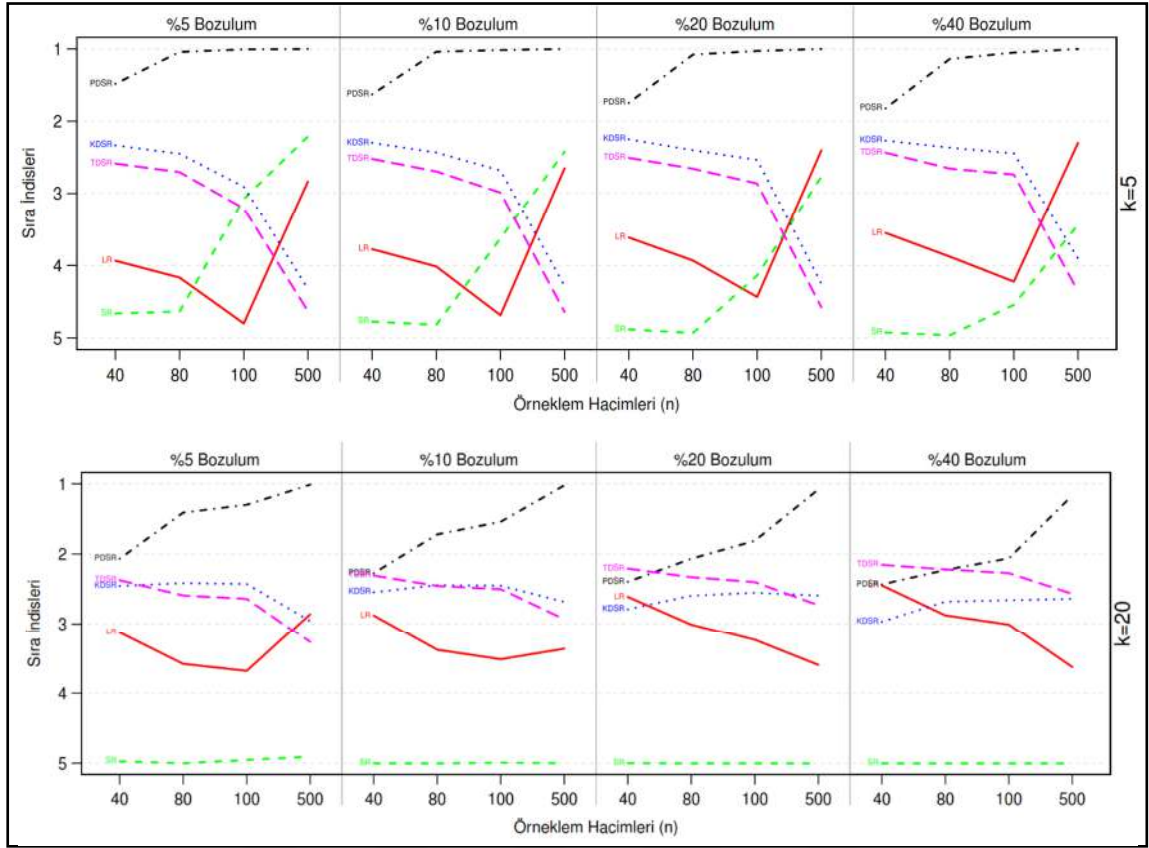
- 1. Modelde n değeri artıkça TDSR ve KDSR yöntemlerinin performanslarının eğilimi düşmekteyken PDSR yöntemi tüm örneklem hacimlerinde en iyi performansa sahiptir. Bunun yanında SR yöntemi, örneklem hacmi artmakla birlikte iyi performansa doğru giderken, LR yöntemi örneklem hacmi 100'den büyük olduğu durumda performansı iyileşmeye başlamaktadır.
- 2. Modelde tüm örneklem hacimlerinde performans sıralamaları en iyiden düşük seviyeye sırasıyla PDSR, KDSR, TDSR, LR ve SR olarak görülmektedir. Ancak KDSR ve TDSR yöntemleri aralarındaki farklılık düşük seviyededir.
- 3. Modelde örneklem hacimleri 80 ve üzeri olduğu durumlarda performans sıralamaları en iyiden düşük seviyeye sırasıyla TDSR, PDSR, KDSR, LR ve SR olarak görülmektedir. Düşük örneklem hacminde LR ile KDSR'nin sıralamaları yer değiştirmektedir.



Şekil 4.3. Pozitif bölgede model 1'e göre bozulma şiddetleri sırasıyla 5 ve 20 olan yapay veriler de yöntemlerin bozulum oranlarına göre performansları

Şekil 4.3. de verilen grafikler de:

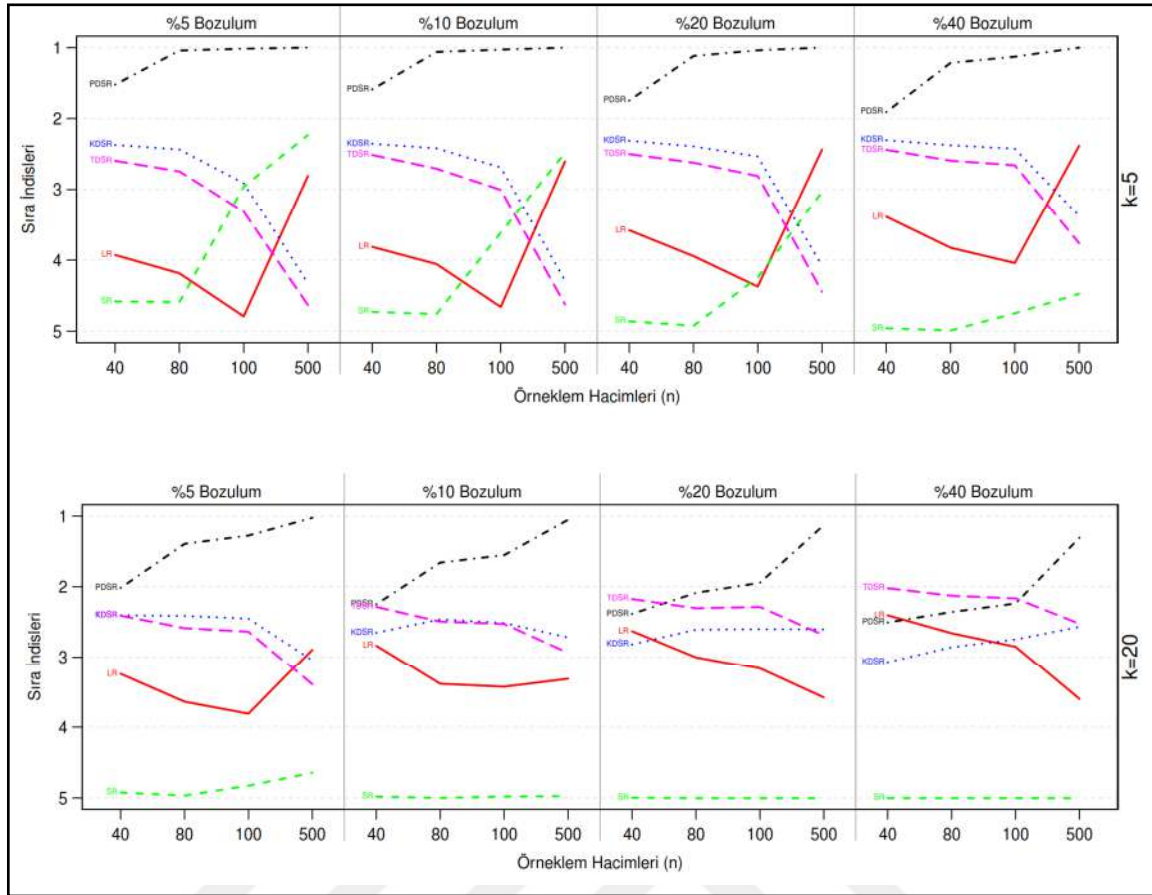
- Bozulum şiddeti 5 iken SR haricindeki yöntemlerin birbirlerine göre olan performansları değişmezken SR yöntemi bozulum oranı arttıkça performansı azalmaktadır. PDSR yöntemi bozulum şiddeti 5 iken tüm bozulum oranlarında en iyi performansı göstermiştir.
- Bozulum şiddeti 20 iken SR yöntemi tüm bozulum oranlarında en düşük performansı sergilemiştir. Aynı bozulum şiddetinde %10 luk bozulum oranına kadarki tüm örneklem hacimlerinde PDSR yöntemi en iyi performansı göstermekteyken, bozulum oranı %20'nin üzerinde ve örneklem hacmi 80'nin üzerindeyken yine PDSR yöntemi en iyi performansı göstermektedir. TDSR yöntemi de örneklem hacmi 40 ve bozulum oranı %20'nin üzerindeyken en iyi performansı vermiştir.



Şekil 4.4. Negatif bölgede model 1'e göre bozulma şiddetleri sırasıyla 5 ve 20 olan yapay verilerde yöntemlerin bozulma oranlarına göre performansları

Şekil 4.4. de verilen grafikler de:

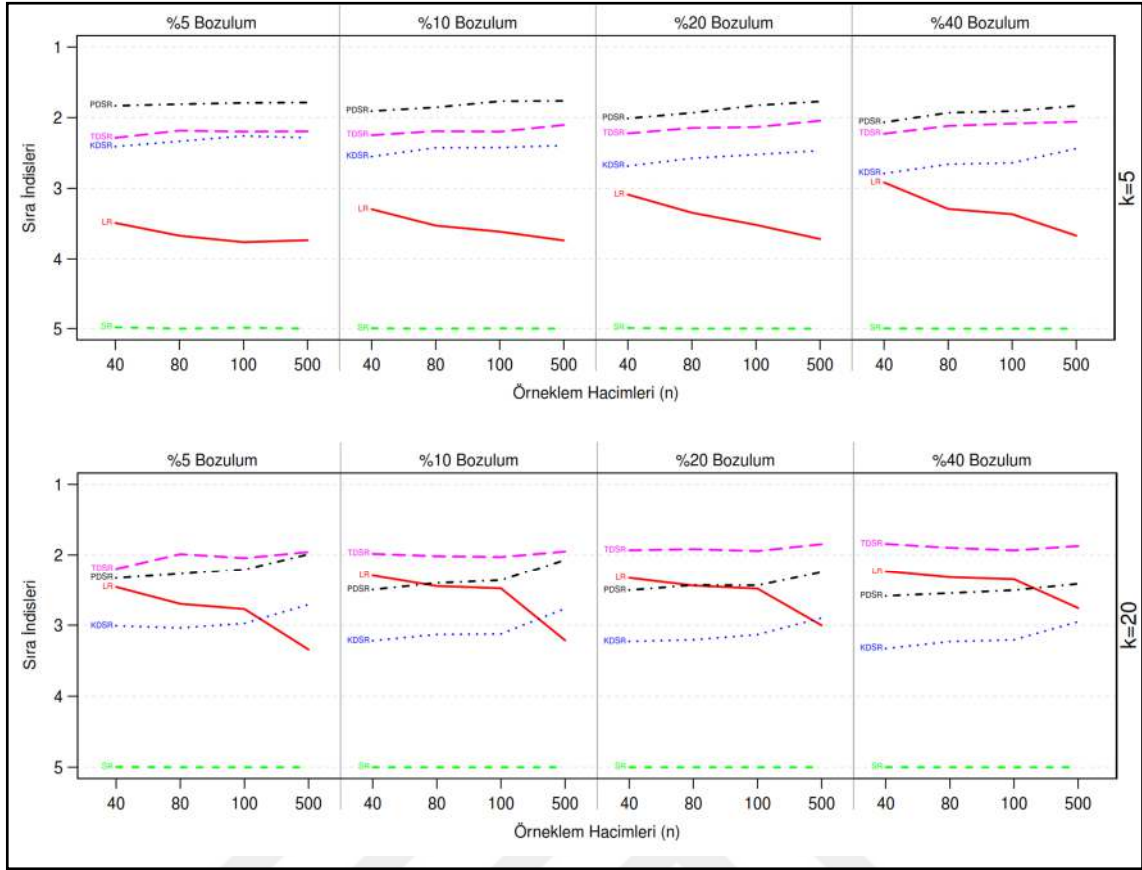
- Bozulma şiddeti 5 iken SR yöntemi haricinde diğer yöntemlerin birbirlerine göre olan performansları değişmezken SR yöntemi bozulma oranı arttıkça performansı düşmektedir. PDSR yöntemi bozulma şiddeti 5 iken tüm bozulma oranlarında en iyi performansı sergilemektedir.
- Bozulma şiddeti 20 iken SR yöntemi tüm bozulma oranlarında en düşük performansı göstermiştir. Aynı bozulma şiddetinde %10 luk bozulma oranına kadarki bütün örneklem hacimlerinde PDSR yöntemi en iyi performansı göstermekteyken, bozulma oranı %20'nin üzerinde ve örneklem hacmi 80'nin üzerindeyken yine PDSR yöntemi en iyi performansı göstermektedir. TDSR yöntemi de örneklem hacmi 40 ve bozulma oranı %20'nin üzerindeyken en iyi performansı vermiştir.



Şekil 4.5. Merkez bölgede model 1'e göre bozulma şiddetleri sırasıyla 5 ve 20 olan yapay verilerde yöntemlerin bozulum oranlarına göre performansları

Şekil 4.5. de verilen grafikler de:

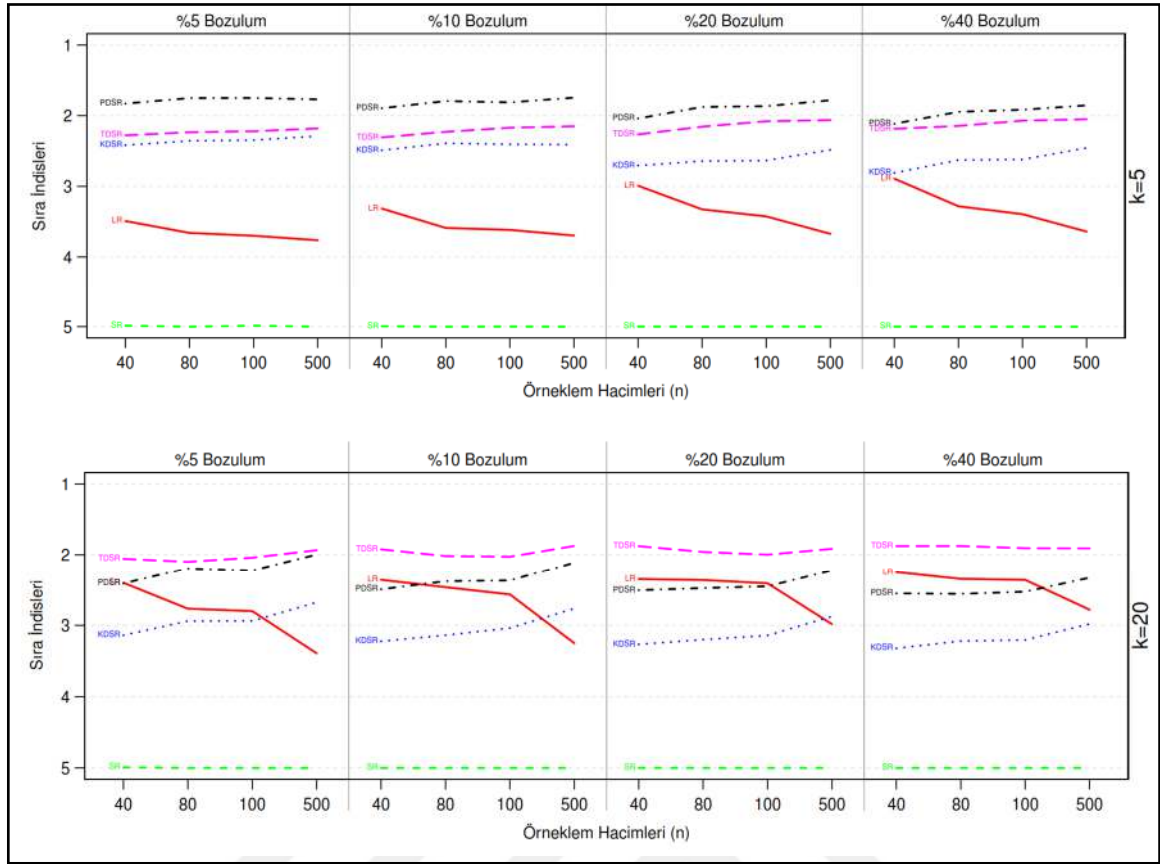
- Bozulum şiddeti 5 iken SR haricindeki yöntemlerin birbirlerine göre olan performansları değişmezken SR yöntemi bozulum oranı arttıkça performansı azalmaktadır. PDSR yöntemi bozulum şiddeti 5 iken tüm bozulum oranlarında en iyi performansı göstermiştir.
- Bozulum şiddeti 20 iken SR yöntemi tüm bozulum oranlarında en düşük performansı sergilemiştir. Aynı bozulum şiddetinde %10 luk bozulum oranına kadarki tüm örneklem hacimlerinde PDSR yöntemi en iyi performansı göstermekteyken, bozulum oranı %20'nin üzerinde ve örneklem hacmi 100'nün üzerindeyken yine PDSR yöntemi en iyi performansı göstermektedir. TDSR yöntemi de örneklem hacmi 40 ve bozulum oranı %20'nin üzerindeyken en iyi performansı vermiştir.



Şekil 4.6. Pozitif bölgede model 2'ye göre bozulma şiddetleri sırasıyla 5 ve 20 olan yapay veriler de yöntemlerin bozulum oranlarına göre performansları

Şekil 4.6. de verilen grafikler de:

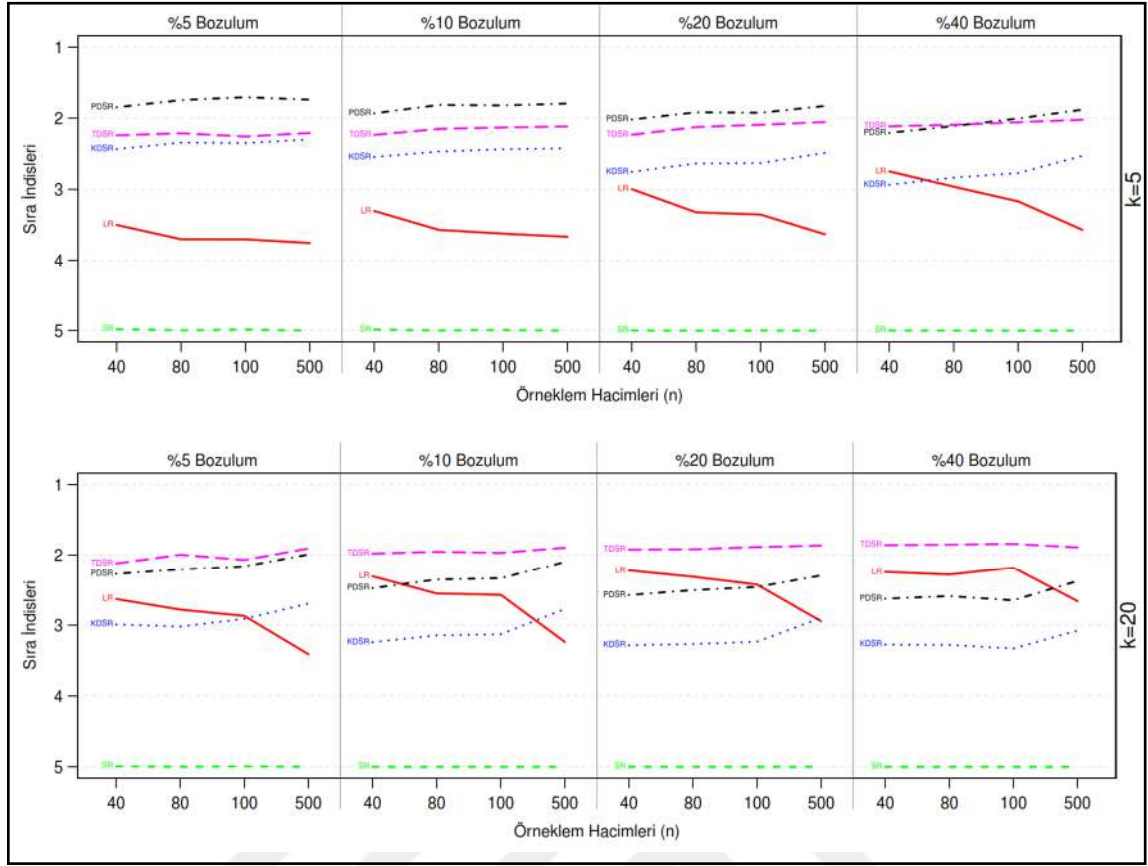
- Bozulma şiddeti 5 iken yöntemlerin performans dereceleri iyiden kötüye sırasıyla PDSR, TDSR, KDSR, LR ve SR olarak görülmektedir. Burada Şekil 4.2 de yer alan bozulmamış verilerdeki performanslar göz önünde tutulduğunda KDSR ve TDSR yöntemleri arasında sıra değişikliği olmuştur.
- Bozulma şiddeti 20 iken ve tüm bozulum oranlarında TDSR yöntemi en iyi performansa sahiptir. Örneklem hacmi 500'ün altındayken, LR yönteminin performansı bozulum oranının artmasıyla beraber PDSR'nin önüne geçmektedir.



Şekil 4.7. Negatif bölgede model 2'ye göre bozulma şiddetleri sırasıyla 5 ve 20 olan yapay veriler de yöntemlerin bozulum oranlarına göre performansları

Şekil 4.7. de verilen grafikler de:

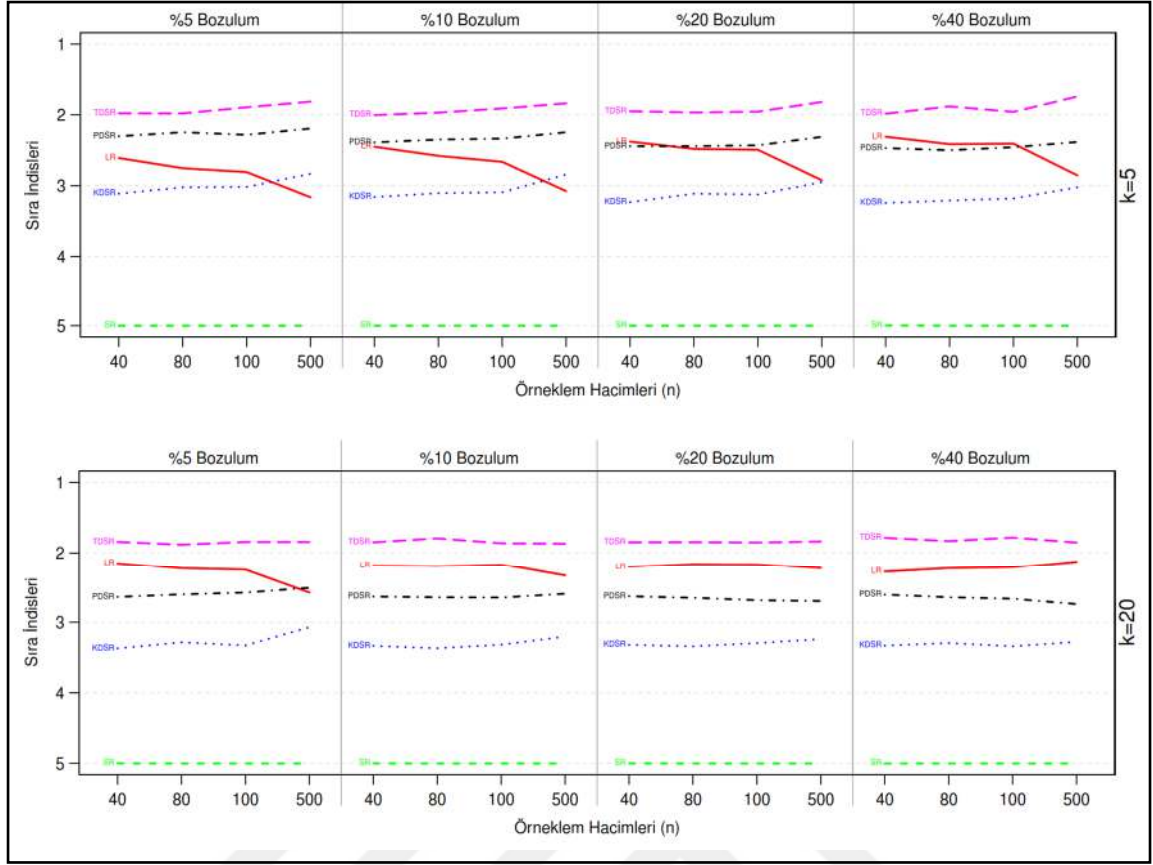
- Bozulma şiddeti 5 iken ve tüm bozulum oranlar için en iyi performanstan en düşük performans sırasıyla PDSR, TDSR, KDSR, LR ve SR yöntemlerin performansları görülmektedir. Ancak bozulum oranı %40'da PDSR ve TDSR yöntemlerin arasında farklılık düşük seviyededir.
- Bozulma şiddetinin 5 ten 20'ye çıkmasıyla en iyi performansı veren yöntem TDSR olarak değişmiştir. LR ve PDSR haricindeki diğer yöntemlerin birbirlerine karşı olan performansları bozulum oranlarının artımında benzerlik göstermektedir. LR yönteminin performansı bozulum oranının artmasıyla beraber PDSR yöntemiyle daha rekabetçidir.



Şekil 4.8. Merkez bölgede model 2'ye göre bozulma şiddetleri sırasıyla 5 ve 20 olan yapay verilerde yöntemlerin bozulum oranlarına göre performansları

Şekil 4.8. de verilen grafikler de:

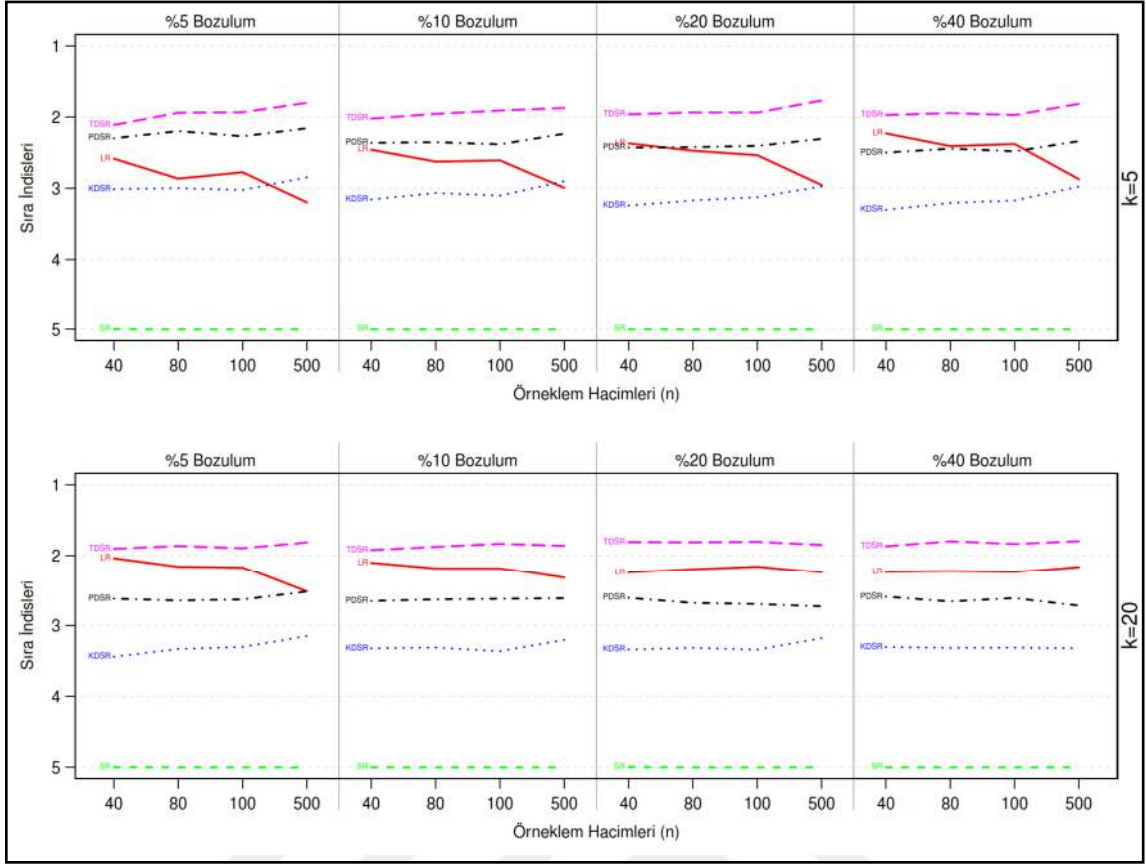
- Bozulma şiddeti 5 iken ve bozulum oran %5 ten %20'e kadar PDSR yöntemi en iyi performansa sahiptir. Bu durumda yöntemlerin performansları sırasıyla PDSR, TDSR, KDSR, LR ve SR olarak görülmektedir. Ancak yüksek bozulma oranında ve küçük örneklem hacimlerinde TDSR yöntemi PDSR yönteminden daha iyi performans göstermiştir.
- Bozulma şiddeti 20 iken TDSR yöntemi en iyi performansa sahiptir. Bozulum oranları %10'nun üzerinde arttıkça LR yöntemi küçük örneklemelerde PDSR yönteminden daha iyi performans göstermektedir.



Şekil 4.9. Pozitif bölgede model 3'e göre bozulma şiddetleri sırasıyla 5 ve 20 olan yapay veriler de yöntemlerin bozulum oranlarına göre performansları

Şekil 4.9. de verilen grafikler de:

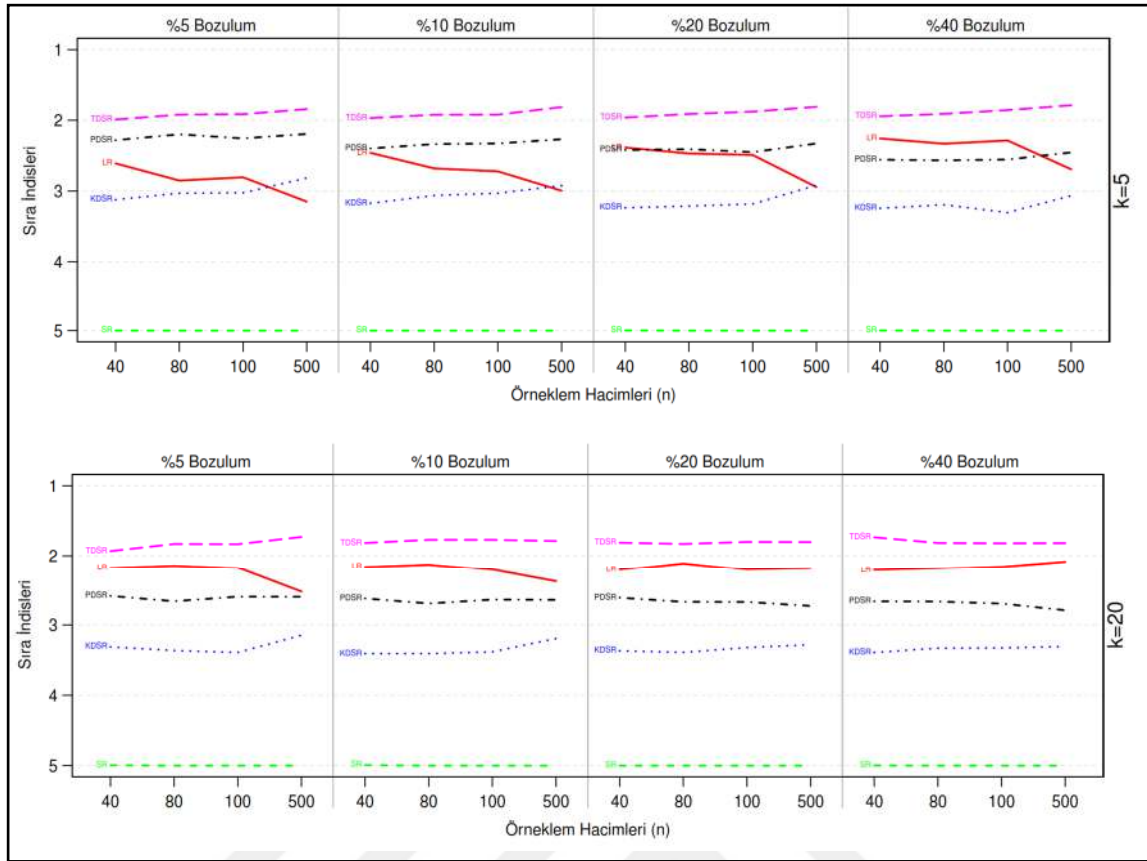
- Bozulma şiddeti 5 iken, düşük bozulma oranlarında en iyi seviyeden en düşüğe kadar yöntemlerin performansları sırasıyla TDSR, PDSR, LR, KDSR ve SR olarak görülmektedir. Bozulma oranları %20 den daha fazla olduğu durumlarda LR yöntemi PDSR yönteminden daha iyi performans sergilemektedir.
- Bozulma şiddeti 20 iken tüm bozulma oranları için en iyiden düşüğe kadar yöntemlerin performansları sırasıyla TDSR, LR, PDSR, KDSR ve SR olarak görülmektedir.



Şekil 4.10. Negatif bölgede model 3'e göre bozulma şiddetleri sırasıyla 5 ve 20 olan yapay veriler de yöntemlerin bozulum oranlarına göre performansları

Şekil 4.10. de verilen grafikler de:

- Bozulma şiddeti 5 iken, düşük bozulma oranlarında en iyi seviyeden en düşüğe kadar yöntemlerin performansları sırasıyla TDSR, PDSR, LR, KDSR ve SR olarak görülmektedir. Bozulma oranları %20'nin üzerinde olduğu durumlarda LR yöntemi PDSR yöntemiyle daha rekabetçidir.
- Bozulma şiddeti 20 iken tüm bozulma oranları için en iyiden düşüğe kadar yöntemlerin birbirlerine karşı performansları sırasıyla TDSR, LR, PDSR, KDSR ve SR olarak görülmektedir.



Şekil 4.11. Merkez bölgede model 3'e göre bozulma şiddetleri sırasıyla 5 ve 20 olan yapay verilerde yöntemlerin bozulma oranlarına göre performansları

Şekil 4.11. de verilen grafikler de:

- Bozulma şiddeti 5 iken, düşük bozulma oranlarında en iyiden en düşüğe kadar yöntemlerin birbirlerine karşı performansları sırasıyla TDSR, PDSR, LR, KDSR ve SR olarak görülmektedir. Bozulma oranları %20'nin üzerinde olduğu durumlarda LR yöntemi PDSR yöntemiyle daha rekabetçidir.
- Bozulma şiddeti 20 iken tüm bozulma oranları için en iyiden düşüğe kadar yöntemlerin performansları sırasıyla TDSR, LR, PDSR, KDSR ve SR olarak görülmektedir.

5. SONUÇLAR VE ÖNERİLER

Bu çalışmada aykırılığa karşı geliştirilmiş toplamsal model yöntemlerinin performansları incelenmiştir. Simülasyondan elde edilen sonuçlar ve bu sonuçlara bağlı öneriler alt kısımda yer almaktadır.

5.1. Sonuçlar

GAM’da kullanılan 5 ayrı yöntemin farklı aykırılık senaryolarında Monte-Carlo simülasyon yöntemiyle birbirlerine karşı performanslarının incelendiği bu çalışmada, model 1 de düşük bozulum şiddetinde PDSR, yüksek bozulum şiddetinde ise bozulum oranlarına göre PDSR ve TDSR en iyi performanslara sahip olduğu görülmüştür. Model 2 de düşük bozulum şiddetinde PDSR, yüksek bozulum şiddetinde ise TDSR yöntemi en iyi performansa sahip olduğu ve Model 3 de ise tüm durumlar için TDSR yöntemi en iyi performansa sahip olduğu tespit edilmiştir.

Genel olarak tüm modellerde çoğunlukla yüksek bozulum şiddetinde TDSR en iyi performansa sahipken SR yöntemi model 1de düşük bozulum şiddeti ve düşük bozulum oranı durumu haricinde diğer tüm durumlarda en düşük performansa sahip olmuştur. Bunun yanında bozulma bölgelerinden pozitif ve negatif bölgelerin, yöntemlerin birlerine karşı olan performans değerlerine önemli derecede etkisi olmamıştır.

Çıkan sonuçlara göre yöntemlerin birbirlerine karşı performanslarının, modelin yapısından ve aykırılık oranıyla şiddetinden önemli derecede etkilendiği görülmüştür.

5.2. Öneriler

Simülasyon sonuçlarıyla elde edilen yöntemlerin birbirlerine karşı performanslarında, modelin yapısı ve aykırılık oranıyla şiddetinin öneminden dolayı, çok daha farklı model yapılarının da incelenmesi daha uygun olacaktır. Bunun yanında gelecek çalışmalarda çok daha büyük örneklem hacimleriyle PDSR ve TDSR yöntemlerinin performanslarına bakılması ve daha yoğun çalışmalara önem verilmesi gerekmektedir.

İleriki çalışmalarda araştırmacılar için belli bir platformda mevcut simülasyon çalışmalarının arttırılması ve depolanmasıyla, araştırmacı mevcut veriyi bu veri tabanına aktararak en uygun olan modele göre bozulma raporlarındaki performanslara göre en uygun yöntemi tercih edebilir.

6. KAYNAKLAR

- Hastie, T. J., & Tibshirani, R. J. (1990). Generalized Additive Models, vol. 43 CRC Press.
- Buja, A., Hastie, T., & Tibshirani, R. (1989). Linear smoothers and additive models. *The Annals of Statistics*, 453-510.
- Wood, S. N. (2006). Generalized additive models: an introduction with R. Chapman and Hall/CRC.
- Wood, S. N. (2017). Generalized additive models: an introduction with R. CRC press.
- Pyrgou, A., & Santamouris, M. (2020). Probability Risk of Heat-and Cold-Related Mortality to Temperature, Gender, and Age Using GAM Regression Analysis. *Climate*, 8(3), 40.
- Analitis, A., Barratt, B., Green, D., Beddows, A., Samoli, E., Schwartz, J., & Katsouyanni, K. (2020). Prediction of PM_{2.5} concentrations at the locations of monitoring sites measuring PM₁₀ and NO_x, using generalized additive models and machine learning methods: A case study in London. *Atmospheric Environment*, 240, 117757.
- Dhulipala, S., & Patil, G. R. (2020). Freight production of agricultural commodities in India using multiple linear regression and generalized additive modelling. *Transport Policy*, 97, 245-258.
- Ateba, F. F., Febrero-Bande, M., Sagara, I., Sogoba, N., Touré, M., Sanogo, D., ... & Doumbia, S. (2020). Predicting malaria transmission dynamics in Dangassa, Mali: a novel approach using functional generalized additive models. *International journal of environmental research and public health*, 17(17), 6339.
- Adamec, Z., & Drápela, K. (2015). Generalized additive models as an alternative approach to the modelling of the tree height-diameter relationship. *Journal of Forest Science*, 61(6), 235-243.
- Aeberhard, W. H., Cantoni, E., Marra, G., & Radice, R. (2021). Robust fitting for generalized additive models for location, scale and shape. *Statistics and Computing*, 31(1), 1-16.
- Amato, U., Antoniadis, A., De Feis, I., Goude, Y., & Lagache, A. (2021). Forecasting high resolution electricity demand data with additive models including smooth and jagged components. *International Journal of Forecasting*, 37(1), 171-185.
- Khoda Bakhshi, A., & Ahmed, M. M. (2022). Real-time crash prediction for a long low-traffic volume corridor using corrected-impurity importance and semi-parametric

- generalized additive model. *Journal of transportation safety & security*, 14(7), 1165-1200.
- Erfanian, M., Sadeghpour Gildeh, B., & Reza Azarpazhooh, M. (2021). A new approach for monitoring healthcare performance using generalized additive profiles. *Journal of Statistical Computation and Simulation*, 91(1), 167-179.
- Savaş, N., & Cengiz, M. Yaşam Memnuniyetini Etkileyen Faktörlerin Belirlenmesinde Genelleştirilmiş Toplamsal Modellerin Kullanımı. *Erzincan University Journal of Science and Technology*, 2(1), 109-124.
- Asfha, H. D. (2017). Performance of spline-based GAM in the presence of outliers and multicollinearity= Splayn-tabanlı GAM'ın çoklu bağlantı ve aykırı değer varlığında performansları (Master's thesis, Anadolu Üniversitesi).
- Zhang, B., Hepp, T., Greven, S., & Bergherr, E. (2021). Adaptive step-length selection in gradient boosting for generalized additive models for location, scale and shape. *arXiv preprint arXiv:2102.09248*.
- Sørensen, Ø., Brandmaier, A. M., Macià, D., Ebmeier, K., Ghisletta, P., Kievit, R. A., Athanasia M. Mowinckel, Kristine B. Walhovd, Rene Westerhausen, Fjell, A. (2021). Meta-analysis of generalized additive models in neuroimaging studies. *NeuroImage*, 224, 117416.
- Ruppert, D., Wand, M. P., & Carroll, R. J. (2009). Semiparametric regression during 2003–2007. *Electronic journal of statistics*, 3, 1193.
- Vandal, N. (2005). La régression non paramétrique multidimensionnelle. Théorie et application une étude portant sur la densité mammaire. *Mémoire d'Etude de la Maîtrise en Statistique de l'Université de Laval*, 108p.
- Guisan, A., Edwards Jr, T. C., & Hastie, T. (2002). Generalized linear and generalized additive models in studies of species distributions: setting the scene. *Ecological modelling*, 157(2-3), 89-100.
- Gressani, O., & Lambert, P. (2021). Laplace approximations for fast Bayesian inference in generalized additive models based on P-splines. *Computational Statistics & Data Analysis*, 154, 107088.
- Gianola, D., Fernando, R. L., & Stella, A. (2006). Genomic-assisted prediction of genetic value with semiparametric procedures. *Genetics*, 173(3), 1761-1776.
- Perperoglou, A., Sauerbrei, W., Abrahamowicz, M., & Schmid, M. (2019). A review of spline function procedures in R. *BMC medical research methodology*, 19(1), 1-16.

- Pekgör, A (2022). Doğrusal regresyonda alternatif uygulamalar: Parametrik olmayan regresyon. P.15
- Engle, R. F., Granger, C. W., Rice, J., & Weiss, A. (1986). Semiparametric estimates of the relation between weather and electricity sales. *Journal of the American statistical Association*, 81(394), 310-320.
- Young, S. G., & Bowman, A. W. (1995). Non-parametric analysis of covariance. *Biometrics*, 920-931.
- Nelder, J. A., & Wedderburn, R. W. (1972). Generalized linear models. *Journal of the Royal Statistical Society: Series A (General)*, 135(3), 370-384.
- Pyrgou, A., & Santamouris, M. (2020). Probability Risk of Heat-and Cold-Related Mortality to Temperature, Gender, and Age Using GAM Regression Analysis. *Climate*, 8(3), 40.
- Hardle, W., & Luckhaus, S. (1984). Uniform consistency of a class of regression function estimators. *The Annals of Statistics*, 12(2), 612-623.
- Chambers, J., Hastie, T., & Pregibon, D. (1992). Statistical models in S. In *Compstat* (pp. 317-321). Physica-Verlag HD.
- Cleveland, W. S., & Loader, C. (1996). Smoothing by local regression: Principles and methods. In *Statistical theory and computational aspects of smoothing* (pp. 10-49). Physica-Verlag HD.
- Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American statistical association*, 74(368), 829-836.
- Cleveland, W. S., & Devlin, S. J. (1988). Locally weighted regression: an approach to regression analysis by local fitting. *Journal of the American statistical association*, 83(403), 596-610.
- Fox, J. (2000). *Nonparametric simple regression: Smoothing scatterplots* (No. 130). Sage.
- Ryan, T. P. (2008). *Modern regression methods* (Vol. 655). John Wiley & Sons.
- Loader, C. (1999). Local Regression Methods. *Local Regression and Likelihood*, 15-44.
- Wand, M. P., & Jones, M. C. (1994). *Kernel smoothing*. CRC press.
- Opsomer, J., Wang, Y., & Yang, Y. (2000). Nonparametric regression with correlated errors. *Statistical Science*, 134-153.
- Stone, C. J. (1985). Additive regression and other nonparametric models. *The annals of Statistics*, 13(2), 689-705.

- Silverman, B. W. (1985). Some aspects of the spline smoothing approach to non-parametric regression curve fitting. *Journal of the Royal Statistical Society: Series B (Methodological)*, 47(1), 1-21.
- Takezawa, K. (2005). *Introduction to nonparametric regression*. John Wiley & Sons.
- Friedman, J. H., & Stuetzle, W. (1981). Projection pursuit regression. *Journal of the American statistical Association*, 76(376), 817-823.
- Härdle, W., Hall, P., & Marron, J. S. (1988). How far are automatically chosen regression smoothing parameters from their optimum?. *Journal of the American Statistical Association*, 83(401), 86-95.
- Härdle, W. (1994). *Applied nonparametric regression* (No. 19). Cambridge university press.
- Härdle, W., Müller, M., Sperlich, S., & Werwatz, A. (2004). *Nonparametric and semiparametric models* (Vol. 1). Berlin: Springer.(p. 62)
- Faraway, J. J. (2006). *Extending the linear model with R*, Texts in Statistical Science Series.
- Duchon, J. (1977). Splines minimizing rotation-invariant semi-norms in Sobolev spaces. In *Constructive theory of functions of several variables* (pp. 85-100). Springer, Berlin, Heidelberg.
- Eubank, R. L. (1999). *Nonparametric regression and spline smoothing*. CRC press.
- De Boor, C., & De Boor, C. (1978). *A practical guide to splines* (Vol. 27, p. 325). New York: springer-verlag.
- Atilgan, T. (1988). Basis selection for density estimation and regression. submitted for publication.
- Stone, C. J., & Koo, C. Y. (1985, August). Additive splines in statistics. In *Proceedings of the Statistical Computing Section ASA*. Washington, DC (pp. 45-48).
- Friedman, J. H., & Silverman, B. W. (1989). Flexible parsimonious smoothing and additive modeling. *Technometrics*, 31(1), 3-21.
- Eilers, P. H., & Marx, B. D. (1996). Flexible smoothing with B-splines and penalties. *Statistical science*, 11(2), 89-121.
- Gu, C., & Gu, C. (2013). *Smoothing spline ANOVA models* (Vol. 297). New York: Springer.
- Wood, S. N. (2000). Modelling and smoothing parameter estimation with multiple quadratic penalties. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 62(2), 413-428.

EKLER

EK-1 Monte-Carlo simülasyonlarıyla elde edilen 5 ayrı yöntemle ilişkin bozulmamış ve bozulmuş verilerdeki Hata Kareler Ortalaması, Ortalama Sıra İndisleri ve Standart Sapma performansları

Ek 1.1. Bozulmamış Veriler ile Farklı Model ve Örneklem Hacimlerine Göre Yöntemlerin Simülasyonlardaki Performansları

Model Numarası	n	Hata Kareler Ortalaması					Ortalama Sıra İndisler					Standart Sapmalar				
		LR	SR	KDSR	PDSR	TDSR	LR	SR	KDSR	PDSR	TDSR	LR	SR	KDSR	PDSR	TDSR
1	40	1.534	81.566	0.594	0.308	0.674	4.108	4.382	2.393	1.483	2.634	0.145	0.246	0.089	0.061	0.095
1	80	1.122	6.643	0.447	0.088	0.495	4.427	4.106	2.566	1.024	2.877	0.064	0.061	0.040	0.018	0.042
1	100	0.986	0.267	0.427	0.072	0.472	4.902	2.238	3.265	1.014	3.581	0.048	0.017	0.032	0.013	0.033
1	500	0.172	0.091	0.425	0.049	0.444	3.001	2.016	4.329	1.000	4.654	0.004	0.002	0.006	0.002	0.007
2	40	0.166	7157.035	0.067	0.062	0.069	3.776	4.974	2.234	1.731	2.285	0.048	0.445	0.031	0.030	0.031
2	80	0.117	8.090	0.052	0.050	0.052	3.872	4.990	2.172	1.710	2.256	0.021	0.059	0.014	0.014	0.014
2	100	0.097	0.240	0.050	0.048	0.050	3.921	4.943	2.202	1.667	2.267	0.015	0.016	0.011	0.011	0.011
2	500	0.047	0.088	0.042	0.041	0.042	3.839	5.000	2.204	1.726	2.231	0.002	0.002	0.002	0.002	0.002
3	40	0.069	38.161	0.066	0.056	0.054	2.836	4.996	2.938	2.185	2.045	0.031	0.183	0.031	0.028	0.028
3	80	0.057	13.759	0.051	0.048	0.047	3.103	5.000	2.849	2.070	1.978	0.014	0.065	0.014	0.013	0.013
3	100	0.052	0.241	0.048	0.046	0.045	3.106	5.000	2.871	2.106	1.917	0.011	0.016	0.011	0.010	0.010
3	500	0.043	0.087	0.042	0.041	0.041	3.364	5.000	2.696	2.118	1.822	0.002	0.002	0.002	0.002	0.002

Ek 1.2. 1. Model de Pozitif Bölgedeki Bozulma Şiddetleri ve Bozulma Oranlarına Göre Yöntemlerin Simülasyonlardaki Performansları

Bozulma Şiddeti(k)	Bozulma Oranı(%)	n	Hata Kareler Ortalaması					Ortalama Sıra İndisler					Standart Sapmalar				
			LR	SR	KDSR	PDSR	TDSR	LR	SR	KDSR	PDSR	TDSR	LR	SR	KDSR	PDSR	TDSR
5	5	40	1.609	107.663	0.708	0.445	0.784	3.886	4.696	2.336	1.526	2.556	0.150	0.301	0.097	0.075	0.103
		80	1.193	10.380	0.517	0.160	0.567	4.145	4.640	2.445	1.046	2.724	0.066	0.087	0.043	0.024	0.045
		100	1.087	3.948	0.499	0.134	0.539	4.796	3.096	2.873	1.005	3.230	0.051	0.027	0.034	0.018	0.036
		500	0.255	0.199	0.470	0.100	0.491	2.823	2.210	4.308	1.000	4.659	0.005	0.003	0.007	0.003	0.007
	10	40	1.689	208.268	0.798	0.541	0.878	3.749	4.796	2.342	1.548	2.565	0.151	0.392	0.103	0.083	0.108
		80	1.253	102.208	0.572	0.213	0.624	4.035	4.831	2.412	1.041	2.681	0.067	0.132	0.046	0.028	0.048
		100	1.164	0.875	0.558	0.187	0.599	4.714	3.536	2.726	1.007	3.017	0.052	0.030	0.036	0.021	0.038
		500	0.322	0.295	0.527	0.144	0.545	2.628	2.423	4.316	1.000	4.633	0.006	0.004	0.007	0.004	0.007
	20	40	1.833	42808.106	0.925	0.699	0.992	3.660	4.896	2.312	1.706	2.426	0.159	1.143	0.112	0.096	0.116
		80	1.369	33.430	0.658	0.315	0.719	3.930	4.926	2.373	1.079	2.692	0.071	0.134	0.049	0.034	0.051
		100	1.240	1.255	0.638	0.278	0.675	4.431	4.230	2.538	1.032	2.769	0.054	0.038	0.039	0.026	0.040
		500	0.425	0.451	0.591	0.217	0.612	2.422	2.797	4.185	1.000	4.596	0.006	0.005	0.008	0.005	0.008
	40	40	1.893	141.821	1.029	0.837	1.092	3.570	4.930	2.239	1.802	2.459	0.163	0.476	0.119	0.108	0.124
		80	1.483	41.618	0.777	0.432	0.826	3.864	4.969	2.403	1.110	2.654	0.074	0.167	0.054	0.040	0.055
		100	1.346	1.931	0.729	0.383	0.775	4.229	4.546	2.442	1.069	2.714	0.056	0.045	0.042	0.030	0.043
		500	0.539	0.634	0.676	0.303	0.695	2.303	3.492	3.887	1.000	4.318	0.007	0.006	0.008	0.005	0.008
20	5	40	2.473	1810.740	1.881	1.720	1.854	3.155	4.975	2.390	2.143	2.337	0.184	1.015	0.158	0.152	0.158
		80	2.151	113.918	1.481	1.158	1.532	3.635	4.998	2.396	1.398	2.573	0.088	0.264	0.072	0.063	0.074
		100	2.000	5.830	1.390	1.060	1.436	3.719	4.949	2.423	1.280	2.629	0.068	0.075	0.056	0.048	0.057
		500	1.219	1.816	1.233	0.859	1.254	2.846	4.948	2.923	1.007	3.276	0.011	0.009	0.011	0.009	0.011
	10	40	3.400	3260.462	3.011	2.817	2.807	2.876	4.995	2.572	2.270	2.287	0.214	1.394	0.201	0.194	0.194
		80	2.958	103.429	2.311	2.059	2.343	3.367	5.000	2.410	1.754	2.469	0.104	0.338	0.091	0.086	0.092
		100	2.803	8.543	2.197	1.905	2.224	3.479	4.991	2.423	1.602	2.505	0.081	0.098	0.072	0.067	0.072
		500	2.054	3.330	1.948	1.574	1.964	3.343	4.997	2.700	1.031	2.929	0.014	0.013	0.014	0.012	0.014
	20	40	4.872	21381.641	4.914	4.435	4.235	2.638	4.997	2.800	2.369	2.196	0.259	1.799	0.260	0.248	0.243
		80	4.286	245.669	3.796	3.499	3.715	3.078	5.000	2.612	1.964	2.346	0.126	0.465	0.119	0.114	0.117
		100	4.075	15.717	3.568	3.269	3.519	3.229	4.998	2.590	1.806	2.377	0.099	0.133	0.092	0.089	0.092
		500	3.340	5.808	3.106	2.740	3.114	3.604	5.000	2.607	1.094	2.695	0.018	0.017	0.018	0.016	0.018
	40	40	6.680	7425.602	7.231	6.403	5.915	2.533	4.997	2.962	2.416	2.092	0.307	2.018	0.318	0.302	0.291
		80	5.753	898.514	5.332	5.063	5.127	2.968	5.000	2.599	2.189	2.244	0.147	0.613	0.142	0.138	0.139
		100	5.644	24.890	5.266	4.965	5.112	3.004	5.000	2.636	2.068	2.292	0.117	0.165	0.113	0.110	0.112
		500	4.765	8.663	4.447	4.095	4.455	3.650	5.000	2.564	1.176	2.610	0.022	0.021	0.021	0.020	0.021

Ek 1.3. 1. Model de Negatif Bölgedeki Bozulma Şiddetleri ve Bozulma Oranlarına Göre Yöntemlerin Simülasyonlardaki Performansları

Bozulma Şiddeti(k)	Bozulma Oranı(%)	n	Hata Kareler Ortalaması					Ortalama Sıra İndisler					Standart Sapmalar				
			LR	SR	KDSR	PDSR	TDSR	LR	SR	KDSR	PDSR	TDSR	LR	SR	KDSR	PDSR	TDSR
5	5	40	1.619	35.788	0.703	0.416	0.791	3.936	4.662	2.334	1.482	2.586	0.149	0.262	0.096	0.073	0.103
		80	1.213	42.381	0.516	0.158	0.569	4.168	4.636	2.450	1.041	2.705	0.066	0.097	0.043	0.024	0.046
		100	1.093	0.569	0.509	0.132	0.547	4.800	3.078	2.907	1.006	3.209	0.051	0.025	0.035	0.018	0.036
		500	0.257	0.198	0.476	0.099	0.495	2.835	2.210	4.324	1.000	4.631	0.005	0.003	0.007	0.003	0.007
	10	40	1.697	257.033	0.792	0.540	0.859	3.776	4.775	2.299	1.629	2.521	0.153	0.393	0.103	0.084	0.108
		80	1.260	22.570	0.588	0.222	0.643	4.015	4.818	2.433	1.039	2.695	0.068	0.110	0.046	0.028	0.048
		100	1.137	0.892	0.545	0.185	0.585	4.686	3.627	2.682	1.015	2.990	0.052	0.031	0.036	0.021	0.037
		500	0.325	0.294	0.521	0.144	0.541	2.650	2.421	4.286	1.000	4.643	0.006	0.004	0.007	0.004	0.007
	20	40	1.711	201.993	0.881	0.695	0.970	3.616	4.881	2.250	1.747	2.506	0.154	0.435	0.110	0.096	0.115
		80	1.353	133.114	0.674	0.308	0.722	3.933	4.931	2.401	1.078	2.657	0.071	0.157	0.049	0.034	0.051
		100	1.265	1.625	0.651	0.282	0.695	4.435	4.139	2.535	1.028	2.863	0.055	0.039	0.039	0.026	0.041
		500	0.427	0.452	0.597	0.217	0.615	2.403	2.769	4.251	1.000	4.577	0.006	0.005	0.008	0.005	0.008
	40	40	1.828	671.523	1.003	0.816	1.075	3.552	4.924	2.268	1.821	2.435	0.159	0.585	0.117	0.106	0.122
		80	1.488	52.314	0.772	0.428	0.824	3.879	4.962	2.365	1.139	2.655	0.074	0.162	0.053	0.040	0.055
		100	1.342	1.836	0.742	0.384	0.790	4.223	4.544	2.445	1.050	2.738	0.057	0.045	0.042	0.031	0.044
		500	0.539	0.634	0.681	0.302	0.700	2.299	3.438	3.906	1.000	4.357	0.007	0.006	0.008	0.005	0.008
20	5	40	2.506	5180.696	1.915	1.718	1.881	3.133	4.973	2.455	2.066	2.373	0.184	1.095	0.160	0.151	0.159
		80	2.089	197.274	1.446	1.155	1.502	3.584	4.999	2.415	1.410	2.592	0.087	0.275	0.071	0.063	0.073
		100	1.965	4.964	1.375	1.049	1.410	3.683	4.953	2.426	1.299	2.639	0.068	0.073	0.056	0.048	0.057
		500	1.199	1.787	1.214	0.845	1.235	2.859	4.903	2.957	1.012	3.269	0.011	0.009	0.011	0.009	0.011
	10	40	3.390	557.125	2.987	2.800	2.851	2.878	4.999	2.544	2.273	2.306	0.216	0.973	0.202	0.195	0.197
		80	2.922	340.469	2.293	2.022	2.305	3.382	5.000	2.443	1.720	2.455	0.103	0.382	0.091	0.085	0.091
		100	2.829	9.526	2.211	1.895	2.237	3.517	4.991	2.451	1.540	2.501	0.082	0.100	0.072	0.067	0.073
		500	2.051	3.328	1.948	1.574	1.966	3.367	4.998	2.683	1.022	2.930	0.014	0.013	0.014	0.012	0.014
	20	40	4.845	1961.143	4.922	4.478	4.278	2.611	4.997	2.789	2.396	2.207	0.262	1.596	0.263	0.251	0.246
		80	4.247	511.904	3.843	3.592	3.735	3.009	5.000	2.594	2.067	2.330	0.126	0.511	0.120	0.116	0.118
		100	4.189	33.794	3.650	3.340	3.615	3.238	4.999	2.552	1.811	2.400	0.100	0.141	0.094	0.090	0.093
		500	3.337	5.800	3.105	2.733	3.119	3.598	5.000	2.592	1.080	2.730	0.018	0.017	0.018	0.016	0.018
	40	40	6.329	4070.080	7.123	6.316	5.800	2.447	5.000	2.966	2.433	2.154	0.303	2.011	0.321	0.302	0.291
		80	5.721	546.037	5.435	5.159	5.177	2.875	5.000	2.679	2.227	2.219	0.147	0.598	0.144	0.140	0.140
		100	5.532	30.592	5.168	4.853	4.991	3.008	5.000	2.659	2.061	2.272	0.116	0.167	0.112	0.109	0.110
		500	4.759	8.639	4.448	4.077	4.444	3.629	5.000	2.638	1.165	2.568	0.022	0.021	0.021	0.020	0.021

Ek 1.4. 1. Model de Merkez Bölgedeki Bozulma Şiddetleri ve Bozulma Oranlarına Göre Yöntemlerin Simülasyonlardaki Performansları

Bozulma Şiddeti(k)	Bozulma Oranı(%)	n	Hata Kareler Ortalaması					Ortalama Sıra İndisler					Standart Sapmalar				
			LR	SR	KDSR	PDSR	TDSR	LR	SR	KDSR	PDSR	TDSR	LR	SR	KDSR	PDSR	TDSR
5	5	40	1.638	50.873	0.712	0.436	0.794	3.929	4.581	2.372	1.522	2.596	0.150	0.264	0.097	0.073	0.103
		80	1.214	21.146	0.515	0.161	0.574	4.186	4.591	2.436	1.041	2.746	0.066	0.094	0.043	0.023	0.046
		100	1.076	0.553	0.498	0.134	0.546	4.790	2.964	2.913	1.016	3.317	0.050	0.024	0.034	0.017	0.036
		500	0.257	0.196	0.474	0.098	0.492	2.811	2.231	4.326	1.000	4.632	0.005	0.003	0.007	0.003	0.007
	10	40	1.662	137.226	0.779	0.534	0.845	3.815	4.727	2.356	1.589	2.513	0.151	0.381	0.102	0.081	0.106
		80	1.278	153.154	0.590	0.221	0.639	4.056	4.757	2.419	1.061	2.707	0.068	0.118	0.046	0.028	0.048
		100	1.141	5.828	0.543	0.191	0.588	4.658	3.616	2.690	1.029	3.007	0.052	0.033	0.036	0.021	0.037
		500	0.332	0.314	0.533	0.153	0.550	2.605	2.483	4.289	1.001	4.622	0.006	0.004	0.007	0.004	0.007
	20	40	1.802	350.203	0.973	0.762	1.046	3.582	4.859	2.312	1.746	2.501	0.157	0.499	0.113	0.098	0.118
		80	1.419	45.161	0.707	0.360	0.758	3.945	4.921	2.393	1.118	2.623	0.072	0.141	0.051	0.036	0.053
		100	1.301	1.723	0.688	0.327	0.731	4.372	4.244	2.535	1.038	2.811	0.055	0.041	0.040	0.027	0.042
		500	0.476	0.529	0.632	0.253	0.650	2.436	3.038	4.081	1.000	4.445	0.007	0.005	0.008	0.005	0.008
	40	40	2.058	42052.012	1.283	1.127	1.347	3.390	4.956	2.305	1.908	2.441	0.169	1.097	0.133	0.124	0.136
		80	1.676	40.031	0.940	0.607	0.996	3.827	4.987	2.377	1.215	2.594	0.078	0.188	0.058	0.047	0.060
		100	1.532	2.741	0.926	0.582	0.969	4.041	4.747	2.425	1.128	2.659	0.060	0.054	0.047	0.037	0.048
		500	0.728	0.956	0.832	0.453	0.851	2.385	4.473	3.377	1.001	3.764	0.008	0.007	0.009	0.007	0.009
20	5	40	2.485	598.622	1.909	1.733	1.848	3.248	4.922	2.404	2.017	2.409	0.178	0.749	0.151	0.140	0.150
		80	2.087	149.318	1.461	1.168	1.496	3.642	4.965	2.413	1.393	2.587	0.085	0.244	0.068	0.058	0.069
		100	1.989	15.148	1.395	1.080	1.435	3.811	4.826	2.451	1.275	2.637	0.066	0.077	0.053	0.044	0.054
		500	1.260	1.903	1.280	0.900	1.300	2.893	4.644	3.043	1.024	3.396	0.011	0.009	0.011	0.009	0.011
	10	40	3.347	1601.526	3.260	2.950	2.881	2.837	4.979	2.651	2.248	2.285	0.209	1.059	0.200	0.189	0.189
		80	3.186	985.119	2.603	2.308	2.625	3.389	4.998	2.461	1.658	2.494	0.105	0.430	0.092	0.085	0.093
		100	2.872	10.067	2.359	2.042	2.354	3.430	4.978	2.513	1.552	2.527	0.080	0.103	0.071	0.065	0.071
		500	2.193	3.612	2.086	1.708	2.102	3.318	4.973	2.719	1.053	2.937	0.015	0.013	0.014	0.013	0.014
	20	40	5.193	4274.560	5.511	4.953	4.676	2.630	4.995	2.817	2.384	2.174	0.260	1.599	0.265	0.250	0.243
		80	4.916	3073.017	4.562	4.276	4.383	3.002	5.000	2.608	2.086	2.304	0.131	0.633	0.125	0.121	0.123
		100	4.910	20.045	4.482	4.184	4.373	3.169	5.000	2.601	1.944	2.286	0.106	0.148	0.100	0.097	0.099
		500	4.031	7.181	3.760	3.394	3.767	3.582	5.000	2.604	1.128	2.686	0.020	0.019	0.019	0.018	0.019
	40	40	9.091	76174.978	10.885	9.547	8.455	2.404	5.000	3.071	2.505	2.020	0.348	3.657	0.382	0.356	0.334
		80	8.462	1035.892	8.531	8.081	7.874	2.659	5.000	2.857	2.357	2.127	0.173	0.788	0.174	0.169	0.167
		100	8.534	44.779	8.446	8.040	7.969	2.851	5.000	2.747	2.237	2.165	0.141	0.209	0.140	0.137	0.136
		500	7.569	14.382	7.155	6.786	7.145	3.603	5.000	2.568	1.305	2.524	0.027	0.027	0.027	0.026	0.027

Ek 1.5. 2. Model de Pozitif Bölgedeki Bozulma Şiddetleri ve Bozulma Oranlarına Göre Yöntemlerin Simülasyonlardaki Performansları

Bozulma Şiddeti(k)	Bozulma Oranı(%)	n	Hata Kareler Ortalaması					Ortalama Sıra İndisler					Standart Sapmalar				
			LR	SR	KDSR	PDSR	TDSR	LR	SR	KDSR	PDSR	TDSR	LR	SR	KDSR	PDSR	TDSR
5	5	40	0.250	96.271	0.151	0.137	0.145	3.491	4.982	2.409	1.832	2.286	0.059	0.300	0.045	0.043	0.044
		80	0.182	26.704	0.114	0.108	0.112	3.671	5.000	2.335	1.810	2.184	0.026	0.093	0.020	0.020	0.020
		100	0.166	0.510	0.110	0.106	0.109	3.763	4.988	2.260	1.791	2.198	0.020	0.024	0.016	0.016	0.016
		500	0.101	0.193	0.091	0.091	0.091	3.735	5.000	2.285	1.786	2.194	0.003	0.003	0.003	0.003	0.003
	10	40	0.301	631.642	0.215	0.191	0.201	3.297	4.996	2.551	1.907	2.249	0.064	0.440	0.054	0.051	0.052
		80	0.237	1282.561	0.169	0.161	0.165	3.528	5.000	2.426	1.855	2.191	0.030	0.151	0.025	0.024	0.025
		100	0.222	0.963	0.164	0.156	0.160	3.614	4.996	2.423	1.768	2.199	0.023	0.030	0.020	0.019	0.019
		500	0.147	0.285	0.135	0.134	0.134	3.738	5.000	2.395	1.763	2.104	0.004	0.004	0.004	0.004	0.004
	20	40	0.399	121.966	0.333	0.292	0.297	3.089	4.991	2.685	2.011	2.224	0.075	0.405	0.068	0.064	0.065
		80	0.321	2497.009	0.256	0.242	0.244	3.346	5.000	2.575	1.933	2.146	0.034	0.194	0.031	0.030	0.030
		100	0.314	1.355	0.248	0.236	0.240	3.518	4.998	2.523	1.827	2.134	0.027	0.037	0.024	0.024	0.024
		500	0.225	0.441	0.209	0.207	0.207	3.718	5.000	2.467	1.771	2.044	0.005	0.005	0.005	0.005	0.005
	40	40	0.507	677.790	0.462	0.406	0.407	2.918	4.997	2.791	2.066	2.228	0.085	0.534	0.082	0.077	0.077
		80	0.434	42.416	0.371	0.346	0.349	3.292	5.000	2.660	1.931	2.117	0.040	0.164	0.037	0.036	0.036
		100	0.410	15.597	0.346	0.328	0.331	3.367	5.000	2.640	1.908	2.085	0.031	0.048	0.029	0.028	0.028
		500	0.316	0.619	0.295	0.292	0.293	3.670	5.000	2.438	1.834	2.058	0.006	0.006	0.005	0.005	0.005
20	5	40	1.165	423.072	1.353	1.187	1.114	2.462	4.996	3.010	2.334	2.198	0.118	0.776	0.130	0.118	0.113
		80	1.009	1140.000	1.046	0.978	0.946	2.700	5.000	3.039	2.274	1.987	0.058	0.324	0.059	0.056	0.055
		100	0.980	4.874	0.985	0.933	0.916	2.773	5.000	2.976	2.207	2.044	0.046	0.073	0.046	0.044	0.044
		500	0.889	1.804	0.857	0.849	0.848	3.344	5.000	2.710	1.989	1.957	0.009	0.009	0.009	0.009	0.009
	10	40	1.895	5770.077	2.388	2.035	1.784	2.299	5.000	3.219	2.499	1.983	0.157	1.415	0.179	0.163	0.150
		80	1.746	167.654	1.889	1.758	1.668	2.449	5.000	3.131	2.404	2.016	0.079	0.352	0.082	0.079	0.076
		100	1.797	9.094	1.907	1.805	1.730	2.481	5.000	3.124	2.366	2.029	0.065	0.100	0.067	0.064	0.063
		500	1.597	3.274	1.560	1.543	1.539	3.214	5.000	2.766	2.070	1.950	0.013	0.013	0.012	0.012	0.012
	20	40	3.427	8326.791	4.279	3.645	3.140	2.332	5.000	3.229	2.507	1.932	0.217	1.866	0.245	0.225	0.208
		80	3.053	161.010	3.377	3.137	2.912	2.442	5.000	3.206	2.435	1.917	0.106	0.437	0.112	0.108	0.104
		100	2.977	16.554	3.187	3.015	2.848	2.485	5.000	3.134	2.438	1.943	0.084	0.131	0.087	0.085	0.082
		500	2.749	5.713	2.722	2.693	2.675	3.003	5.000	2.899	2.252	1.846	0.017	0.017	0.016	0.016	0.016
	40	40	5.034	2868.528	6.565	5.546	4.573	2.241	4.999	3.330	2.587	1.843	0.268	1.969	0.304	0.280	0.257
		80	4.438	953.426	4.982	4.679	4.271	2.323	5.000	3.231	2.548	1.898	0.130	0.627	0.137	0.133	0.128
		100	4.392	22.919	4.782	4.522	4.249	2.353	5.000	3.207	2.507	1.933	0.104	0.160	0.108	0.105	0.102
		500	4.064	8.558	4.073	4.034	3.993	2.759	5.000	2.952	2.418	1.871	0.020	0.021	0.020	0.020	0.020

Ek 1.6. 2. Model de Negatif Bölgedeki Bozulma Şiddetleri ve Bozulma Oranlarına Göre Yöntemlerin Simülasyonlardaki Performansları

Bozulma Şiddeti(k)	Bozulma Oranı(%)	n	Hata Kareler Ortalaması					Ortalama Sıra İndisler					Standart Sapmalar				
			LR	SR	KDSR	PDSR	TDSR	LR	SR	KDSR	PDSR	TDSR	LR	SR	KDSR	PDSR	TDSR
5	5	40	0.249	41.750	0.147	0.132	0.141	3.491	4.986	2.417	1.828	2.278	0.059	0.263	0.045	0.042	0.044
		80	0.184	9.577	0.115	0.109	0.112	3.659	5.000	2.356	1.751	2.234	0.026	0.084	0.020	0.020	0.020
		100	0.162	0.527	0.110	0.105	0.108	3.699	4.986	2.346	1.749	2.220	0.020	0.024	0.016	0.016	0.016
		500	0.101	0.193	0.091	0.091	0.091	3.763	5.000	2.288	1.769	2.180	0.003	0.003	0.003	0.003	0.003
	10	40	0.298	579.727	0.212	0.191	0.201	3.314	4.994	2.490	1.894	2.308	0.064	0.405	0.054	0.051	0.053
		80	0.239	21.221	0.168	0.159	0.164	3.589	5.000	2.390	1.793	2.228	0.030	0.105	0.025	0.024	0.024
		100	0.220	0.787	0.160	0.153	0.156	3.616	4.998	2.405	1.811	2.170	0.023	0.029	0.019	0.019	0.019
		500	0.146	0.284	0.134	0.133	0.134	3.696	5.000	2.409	1.745	2.150	0.004	0.004	0.004	0.004	0.004
	20	40	0.381	222.199	0.327	0.288	0.295	2.991	4.998	2.706	2.040	2.265	0.073	0.464	0.068	0.064	0.065
		80	0.326	62.549	0.263	0.246	0.249	3.326	5.000	2.643	1.876	2.155	0.035	0.136	0.031	0.030	0.030
		100	0.310	1.307	0.251	0.238	0.240	3.425	4.997	2.634	1.865	2.079	0.027	0.037	0.025	0.024	0.024
		500	0.227	0.444	0.211	0.209	0.210	3.672	5.000	2.484	1.782	2.062	0.005	0.005	0.005	0.005	0.005
	40	40	0.516	2337.219	0.471	0.414	0.416	2.891	4.999	2.810	2.114	2.186	0.085	0.661	0.082	0.077	0.077
		80	0.427	18.722	0.360	0.338	0.340	3.282	5.000	2.628	1.945	2.145	0.040	0.141	0.037	0.036	0.036
		100	0.406	1.864	0.347	0.329	0.330	3.394	5.000	2.620	1.916	2.070	0.031	0.044	0.029	0.028	0.028
		500	0.315	0.617	0.294	0.291	0.292	3.640	5.000	2.456	1.854	2.050	0.006	0.006	0.005	0.005	0.005
20	5	40	1.011	929.895	1.251	1.049	0.948	2.403	4.991	3.139	2.411	2.056	0.109	0.874	0.125	0.111	0.104
		80	0.982	75.959	0.997	0.935	0.909	2.767	5.000	2.941	2.194	2.098	0.057	0.256	0.058	0.055	0.054
		100	0.981	5.583	0.979	0.933	0.915	2.801	5.000	2.938	2.222	2.039	0.046	0.073	0.046	0.044	0.044
		500	0.871	1.762	0.837	0.828	0.826	3.393	5.000	2.679	1.995	1.933	0.009	0.009	0.009	0.009	0.009
	10	40	1.998	4998.059	2.450	2.093	1.834	2.361	5.000	3.224	2.494	1.921	0.161	1.426	0.182	0.166	0.152
		80	1.766	1756.644	1.926	1.791	1.698	2.465	5.000	3.139	2.380	2.016	0.079	0.440	0.083	0.079	0.077
		100	1.764	8.529	1.856	1.756	1.689	2.566	5.000	3.039	2.369	2.026	0.064	0.099	0.066	0.064	0.062
		500	1.588	3.253	1.548	1.532	1.525	3.251	5.000	2.765	2.111	1.873	0.013	0.013	0.012	0.012	0.012
	20	40	3.402	924.466	4.292	3.615	3.080	2.350	4.999	3.268	2.507	1.876	0.218	1.380	0.247	0.227	0.208
		80	3.029	279.044	3.374	3.135	2.922	2.364	5.000	3.201	2.477	1.958	0.106	0.466	0.112	0.108	0.104
		100	3.014	15.996	3.241	3.069	2.925	2.410	5.000	3.142	2.452	1.996	0.085	0.132	0.088	0.086	0.084
		500	2.749	5.725	2.728	2.699	2.679	2.984	5.000	2.875	2.227	1.914	0.017	0.017	0.016	0.016	0.016
	40	40	4.951	280987.922	6.387	5.429	4.578	2.252	5.000	3.322	2.551	1.875	0.267	3.799	0.302	0.280	0.258
		80	4.459	34799.197	5.037	4.669	4.277	2.347	5.000	3.221	2.558	1.874	0.130	0.830	0.139	0.134	0.128
		100	4.354	22.162	4.782	4.515	4.223	2.362	5.000	3.205	2.528	1.905	0.103	0.158	0.108	0.105	0.102
		500	4.057	8.515	4.064	4.019	3.982	2.781	5.000	2.981	2.331	1.907	0.020	0.021	0.020	0.020	0.020

Ek 1.7. 2. Model de Merkez Bölgedeki Bozulma Şiddetleri ve Bozulma Oranlarına Göre Yöntemlerin Simülasyonlardaki Performansları

Bozulma Şiddeti(k)	Bozulma Oranı(%)	n	Hata Kareler Ortalaması					Ortalama Sıra İndisler					Standart Sapmalar				
			LR	SR	KDSR	PDSR	TDSR	LR	SR	KDSR	PDSR	TDSR	LR	SR	KDSR	PDSR	TDSR
5	5	40	0.236	59.281	0.145	0.129	0.136	3.501	4.979	2.434	1.846	2.240	0.057	0.261	0.043	0.040	0.042
		80	0.182	16.353	0.115	0.110	0.113	3.703	4.997	2.344	1.745	2.211	0.025	0.094	0.020	0.019	0.020
		100	0.165	0.541	0.112	0.107	0.110	3.704	4.986	2.349	1.704	2.257	0.019	0.024	0.016	0.015	0.016
		500	0.102	0.196	0.093	0.092	0.093	3.756	5.000	2.298	1.739	2.207	0.003	0.003	0.003	0.003	0.003
	10	40	0.292	138.681	0.222	0.197	0.205	3.304	4.986	2.545	1.931	2.234	0.062	0.335	0.053	0.050	0.051
		80	0.248	1425.309	0.180	0.171	0.174	3.570	4.998	2.469	1.813	2.150	0.030	0.174	0.025	0.024	0.024
		100	0.225	1.043	0.167	0.160	0.162	3.624	4.992	2.436	1.819	2.129	0.023	0.030	0.019	0.019	0.019
		500	0.155	0.301	0.143	0.141	0.142	3.668	5.000	2.423	1.794	2.115	0.004	0.004	0.004	0.004	0.004
	20	40	0.425	181.378	0.386	0.335	0.341	2.997	4.999	2.754	2.017	2.233	0.075	0.465	0.071	0.066	0.066
		80	0.359	60.868	0.298	0.280	0.283	3.324	5.000	2.636	1.917	2.123	0.036	0.152	0.032	0.031	0.031
		100	0.337	1.373	0.287	0.273	0.274	3.355	4.999	2.631	1.923	2.092	0.028	0.039	0.025	0.025	0.025
		500	0.261	0.513	0.244	0.241	0.242	3.633	5.000	2.487	1.827	2.053	0.005	0.005	0.005	0.005	0.005
	40	40	0.647	309.796	0.666	0.571	0.548	2.747	4.998	2.936	2.206	2.113	0.093	0.592	0.094	0.087	0.085
		80	0.575	66.904	0.545	0.511	0.507	2.963	5.000	2.835	2.109	2.093	0.046	0.204	0.044	0.043	0.043
		100	0.574	2.874	0.519	0.492	0.492	3.169	5.000	2.771	2.004	2.056	0.037	0.053	0.035	0.034	0.034
		500	0.470	0.933	0.443	0.439	0.439	3.571	5.000	2.529	1.879	2.021	0.007	0.007	0.007	0.007	0.007
20	5	40	1.080	1173.418	1.317	1.136	1.020	2.629	4.992	2.989	2.271	2.119	0.103	0.809	0.114	0.103	0.096
		80	1.017	127.668	1.085	1.011	0.970	2.779	4.999	3.020	2.207	1.995	0.053	0.245	0.054	0.051	0.049
		100	1.077	5.816	1.098	1.048	1.025	2.868	4.993	2.910	2.160	2.069	0.045	0.071	0.045	0.043	0.042
		500	0.895	1.818	0.863	0.854	0.852	3.411	5.000	2.694	1.990	1.905	0.009	0.009	0.009	0.009	0.009
	10	40	2.114	526058.730	2.692	2.283	1.987	2.306	5.000	3.240	2.475	1.979	0.150	3.242	0.174	0.156	0.142
		80	1.957	1951.985	2.146	2.009	1.895	2.551	5.000	3.144	2.352	1.953	0.077	0.414	0.081	0.077	0.074
		100	1.899	10.598	2.039	1.923	1.832	2.569	5.000	3.128	2.334	1.969	0.061	0.102	0.064	0.061	0.060
		500	1.751	3.611	1.718	1.699	1.691	3.236	5.000	2.771	2.098	1.895	0.013	0.013	0.013	0.013	0.013
	20	40	3.901	4757.126	5.034	4.300	3.638	2.223	4.999	3.283	2.573	1.922	0.218	1.725	0.253	0.230	0.207
		80	3.587	1183.523	4.070	3.778	3.496	2.313	5.000	3.267	2.503	1.917	0.108	0.571	0.116	0.111	0.106
		100	3.495	20.551	3.842	3.606	3.386	2.424	5.000	3.233	2.458	1.885	0.087	0.145	0.092	0.088	0.085
		500	3.373	7.060	3.356	3.321	3.295	2.944	5.000	2.897	2.295	1.864	0.018	0.018	0.018	0.018	0.018
	40	40	7.852	48802.546	10.265	8.780	7.409	2.244	5.000	3.274	2.624	1.858	0.321	2.975	0.370	0.341	0.309
		80	7.016	10923.086	7.998	7.466	6.783	2.281	5.000	3.279	2.588	1.852	0.158	0.959	0.169	0.163	0.155
		100	7.089	38.329	7.958	7.508	6.926	2.183	5.000	3.329	2.647	1.841	0.128	0.205	0.136	0.132	0.126
		500	6.764	14.329	6.821	6.749	6.680	2.659	5.000	3.076	2.374	1.891	0.026	0.026	0.026	0.026	0.026

Ek 1.8. 3. Model de Pozitif Bölgedeki Bozulma Şiddetleri ve Bozulma Oranlarına Göre Yöntemlerin Simülasyonlardaki Performansları

Bozulma Şiddeti(k)	Bozulma Oranı(%)	n	Hata Kareler Ortalaması					Ortalama Sıra İndisler					Standart Sapmalar				
			LR	SR	KDSR	PDSR	TDSR	LR	SR	KDSR	PDSR	TDSR	LR	SR	KDSR	PDSR	TDSR
5	5	40	0.130	50.251	0.143	0.121	0.111	2.609	5.000	3.113	2.301	1.977	0.042	0.272	0.044	0.040	0.038
		80	0.111	8.522	0.113	0.106	0.102	2.753	5.000	3.023	2.245	1.979	0.020	0.085	0.020	0.019	0.019
		100	0.109	1.462	0.109	0.103	0.100	2.808	5.000	3.018	2.281	1.893	0.016	0.025	0.016	0.016	0.015
		500	0.093	0.193	0.092	0.091	0.090	3.164	5.000	2.832	2.192	1.812	0.003	0.003	0.003	0.003	0.003
	10	40	0.182	130.767	0.213	0.182	0.165	2.449	5.000	3.160	2.388	2.003	0.050	0.371	0.054	0.050	0.047
		80	0.159	45.637	0.168	0.156	0.149	2.578	5.000	3.106	2.348	1.968	0.024	0.118	0.025	0.024	0.023
		100	0.155	0.805	0.160	0.151	0.144	2.663	5.000	3.093	2.335	1.909	0.019	0.029	0.020	0.019	0.018
		500	0.138	0.285	0.136	0.135	0.134	3.077	5.000	2.841	2.245	1.837	0.004	0.004	0.004	0.004	0.004
	20	40	0.269	106.007	0.328	0.279	0.245	2.377	5.000	3.232	2.442	1.949	0.061	0.419	0.068	0.062	0.059
		80	0.239	19.440	0.258	0.241	0.227	2.480	5.000	3.113	2.441	1.966	0.030	0.132	0.031	0.030	0.029
		100	0.228	1.110	0.243	0.229	0.218	2.492	5.000	3.124	2.429	1.955	0.023	0.036	0.024	0.023	0.023
		500	0.210	0.438	0.209	0.207	0.205	2.923	5.000	2.950	2.310	1.817	0.005	0.005	0.005	0.005	0.005
	40	40	0.371	661.940	0.465	0.396	0.345	2.308	4.996	3.244	2.469	1.983	0.073	0.602	0.082	0.076	0.071
		80	0.327	590.710	0.363	0.335	0.311	2.413	5.000	3.209	2.498	1.880	0.035	0.211	0.037	0.035	0.034
		100	0.324	1.715	0.351	0.331	0.312	2.406	5.000	3.180	2.456	1.958	0.028	0.043	0.029	0.028	0.028
		500	0.294	0.614	0.294	0.291	0.287	2.855	5.000	3.023	2.382	1.740	0.005	0.006	0.005	0.005	0.005
20	5	40	0.970	751.307	1.274	1.087	0.920	2.149	4.998	3.371	2.640	1.842	0.103	0.862	0.127	0.113	0.098
		80	0.882	245.251	1.011	0.945	0.865	2.230	5.000	3.284	2.604	1.882	0.053	0.272	0.058	0.055	0.052
		100	0.876	16.238	0.974	0.915	0.855	2.250	5.000	3.329	2.578	1.843	0.043	0.077	0.046	0.044	0.042
		500	0.832	1.766	0.842	0.834	0.823	2.577	5.000	3.070	2.510	1.843	0.009	0.009	0.009	0.009	0.009
	10	40	1.864	19201.587	2.453	2.096	1.744	2.184	4.998	3.335	2.635	1.848	0.154	1.610	0.182	0.166	0.146
		80	1.615	1821.213	1.899	1.761	1.577	2.194	5.000	3.369	2.645	1.792	0.075	0.437	0.082	0.079	0.073
		100	1.643	13.955	1.844	1.747	1.617	2.171	5.000	3.318	2.649	1.862	0.061	0.100	0.065	0.063	0.060
		500	1.526	3.257	1.557	1.540	1.514	2.333	5.000	3.204	2.595	1.868	0.012	0.013	0.012	0.012	0.012
	20	40	3.180	3318.917	4.195	3.567	2.933	2.204	4.998	3.320	2.631	1.847	0.210	1.616	0.242	0.222	0.200
		80	2.927	276.071	3.365	3.137	2.841	2.161	5.000	3.341	2.653	1.845	0.104	0.472	0.112	0.108	0.103
		100	2.879	25.876	3.268	3.080	2.826	2.166	5.000	3.296	2.687	1.851	0.083	0.134	0.089	0.086	0.082
		500	2.669	5.720	2.732	2.704	2.656	2.226	5.000	3.239	2.699	1.836	0.016	0.017	0.016	0.016	0.016
	40	40	4.992	13431.678	6.430	5.492	4.494	2.278	5.000	3.331	2.607	1.784	0.268	2.139	0.301	0.279	0.255
		80	4.337	761.394	4.996	4.659	4.177	2.229	5.000	3.295	2.645	1.831	0.129	0.620	0.137	0.133	0.126
		100	4.269	23.733	4.809	4.545	4.139	2.213	5.000	3.340	2.666	1.781	0.102	0.159	0.109	0.106	0.101
		500	3.967	8.518	4.071	4.030	3.957	2.126	5.000	3.279	2.743	1.852	0.020	0.021	0.020	0.020	0.020

Ek 1.9. 3. Model de Negatif Bölgedeki Bozulma Şiddetleri ve Bozulma Oranlarına Göre Yöntemlerin Simülasyonlardaki Performansları

Bozulma Şiddeti(k)	Bozulma Oranı(%)	n	Hata Kareler Ortalaması					Ortalama Sıra İndisler					Standart Sapmalar				
			LR	SR	KDSR	PDSR	TDSR	LR	SR	KDSR	PDSR	TDSR	LR	SR	KDSR	PDSR	TDSR
5	5	40	0.129	27.699	0.143	0.122	0.115	2.584	4.997	3.012	2.297	2.110	0.042	0.242	0.044	0.040	0.039
		80	0.110	15.257	0.110	0.102	0.098	2.865	5.000	2.998	2.197	1.940	0.020	0.087	0.020	0.019	0.019
		100	0.110	0.587	0.110	0.104	0.101	2.775	5.000	3.024	2.269	1.932	0.016	0.024	0.016	0.016	0.015
		500	0.094	0.193	0.092	0.091	0.090	3.199	5.000	2.846	2.156	1.799	0.003	0.003	0.003	0.003	0.003
	10	40	0.186	13922.172	0.213	0.182	0.167	2.458	5.000	3.158	2.362	2.022	0.050	0.758	0.054	0.050	0.047
		80	0.160	24.725	0.166	0.155	0.148	2.626	5.000	3.066	2.353	1.955	0.024	0.107	0.025	0.024	0.023
		100	0.157	0.755	0.162	0.153	0.147	2.608	5.000	3.102	2.382	1.908	0.019	0.029	0.020	0.019	0.019
		500	0.137	0.285	0.136	0.134	0.133	2.995	5.000	2.900	2.233	1.872	0.004	0.004	0.004	0.004	0.004
	20	40	0.273	242.666	0.337	0.285	0.252	2.369	5.000	3.242	2.428	1.961	0.062	0.437	0.069	0.063	0.059
		80	0.242	36.470	0.261	0.243	0.229	2.472	5.000	3.170	2.422	1.936	0.030	0.136	0.031	0.030	0.029
		100	0.235	2.248	0.248	0.235	0.223	2.535	5.000	3.125	2.404	1.936	0.024	0.038	0.024	0.024	0.023
		500	0.211	0.439	0.209	0.207	0.205	2.954	5.000	2.974	2.305	1.767	0.005	0.005	0.005	0.005	0.005
	40	40	0.369	519.722	0.466	0.391	0.342	2.229	5.000	3.302	2.498	1.971	0.073	0.519	0.081	0.075	0.070
		80	0.329	125.209	0.362	0.335	0.314	2.408	5.000	3.205	2.443	1.944	0.035	0.179	0.037	0.035	0.034
		100	0.320	12.268	0.345	0.326	0.308	2.378	5.000	3.171	2.480	1.971	0.028	0.049	0.029	0.028	0.027
		500	0.294	0.614	0.293	0.290	0.287	2.874	5.000	2.975	2.337	1.814	0.005	0.006	0.005	0.005	0.005
20	5	40	0.961	3940.355	1.310	1.106	0.931	2.037	4.998	3.442	2.621	1.902	0.103	0.996	0.128	0.113	0.099
		80	0.872	149.818	1.006	0.937	0.850	2.157	5.000	3.333	2.647	1.863	0.052	0.273	0.058	0.055	0.051
		100	0.899	5.741	1.009	0.954	0.887	2.169	5.000	3.304	2.632	1.895	0.043	0.074	0.047	0.045	0.042
		500	0.821	1.741	0.833	0.824	0.812	2.517	5.000	3.151	2.520	1.812	0.009	0.009	0.009	0.009	0.009
	10	40	1.807	2314.690	2.406	2.071	1.739	2.101	5.000	3.324	2.654	1.921	0.151	1.191	0.180	0.165	0.147
		80	1.655	597.499	1.897	1.767	1.605	2.181	5.000	3.313	2.631	1.875	0.075	0.409	0.082	0.079	0.074
		100	1.608	8.424	1.836	1.719	1.578	2.181	5.000	3.364	2.623	1.832	0.060	0.098	0.065	0.063	0.059
		500	1.518	3.239	1.550	1.534	1.508	2.321	5.000	3.204	2.615	1.860	0.012	0.013	0.012	0.012	0.012
	20	40	3.267	1397.153	4.331	3.652	2.992	2.246	4.996	3.341	2.610	1.807	0.214	1.489	0.249	0.228	0.205
		80	3.003	923.197	3.442	3.207	2.886	2.192	5.000	3.316	2.681	1.811	0.106	0.516	0.114	0.110	0.104
		100	2.861	61.707	3.222	3.047	2.786	2.157	5.000	3.342	2.697	1.804	0.083	0.140	0.088	0.086	0.082
		500	2.653	5.685	2.714	2.689	2.641	2.242	5.000	3.180	2.730	1.848	0.016	0.017	0.016	0.016	0.016
	40	40	4.875	25888.167	6.429	5.480	4.502	2.236	5.000	3.305	2.592	1.867	0.268	2.481	0.305	0.282	0.258
		80	4.352	583.123	5.055	4.700	4.193	2.225	5.000	3.317	2.661	1.797	0.129	0.589	0.138	0.134	0.127
		100	4.236	23.695	4.790	4.505	4.136	2.238	5.000	3.315	2.613	1.834	0.102	0.161	0.108	0.105	0.101
		500	3.967	8.511	4.067	4.025	3.952	2.162	5.000	3.323	2.720	1.795	0.020	0.021	0.020	0.020	0.020

Ek 1.10. 3. Model de Merkez Bölgedeki Bozulma Şiddetleri ve Bozulma Oranlarına Göre Yöntemlerin Simülasyonlardaki Performansları

Bozulma Şiddeti(k)	Bozulma Oranı(%)	n	Hata Kareler Ortalaması					Ortalama Sıra İndisler					Standart Sapmalar				
			LR	SR	KDSR	PDSR	TDSR	LR	SR	KDSR	PDSR	TDSR	LR	SR	KDSR	PDSR	TDSR
5	5	40	0.132	103.519	0.150	0.127	0.118	2.609	4.999	3.122	2.281	1.989	0.040	0.271	0.043	0.039	0.037
		80	0.112	1043.535	0.113	0.105	0.101	2.852	5.000	3.031	2.198	1.919	0.020	0.134	0.020	0.019	0.018
		100	0.108	0.507	0.109	0.103	0.099	2.806	5.000	3.025	2.256	1.913	0.015	0.023	0.015	0.015	0.015
		500	0.094	0.193	0.092	0.091	0.091	3.150	5.000	2.816	2.192	1.842	0.003	0.003	0.003	0.003	0.003
	10	40	0.183	546.848	0.216	0.185	0.168	2.460	5.000	3.174	2.398	1.968	0.048	0.429	0.052	0.048	0.045
		80	0.169	18.884	0.176	0.164	0.157	2.680	5.000	3.063	2.336	1.921	0.024	0.107	0.024	0.023	0.023
		100	0.166	1.058	0.171	0.162	0.156	2.722	5.000	3.031	2.327	1.920	0.019	0.030	0.020	0.019	0.018
		500	0.145	0.301	0.144	0.142	0.141	2.996	5.000	2.924	2.267	1.813	0.004	0.004	0.004	0.004	0.004
	20	40	0.290	9424.220	0.355	0.300	0.265	2.385	4.997	3.237	2.420	1.961	0.060	0.697	0.068	0.061	0.057
		80	0.273	37.214	0.301	0.280	0.262	2.469	5.000	3.213	2.408	1.910	0.031	0.137	0.032	0.031	0.030
		100	0.267	1.459	0.286	0.270	0.256	2.490	5.000	3.183	2.450	1.877	0.024	0.039	0.025	0.025	0.024
		500	0.245	0.511	0.244	0.241	0.239	2.944	5.000	2.919	2.328	1.809	0.005	0.005	0.005	0.005	0.005
	40	40	0.538	5386.024	0.682	0.588	0.508	2.257	4.998	3.245	2.557	1.943	0.085	0.991	0.096	0.089	0.082
		80	0.488	96.841	0.546	0.512	0.472	2.331	5.000	3.194	2.566	1.909	0.042	0.203	0.044	0.043	0.041
		100	0.475	2.366	0.523	0.493	0.460	2.285	5.000	3.305	2.555	1.855	0.033	0.052	0.035	0.034	0.033
		500	0.442	0.929	0.444	0.439	0.434	2.693	5.000	3.066	2.454	1.787	0.007	0.007	0.007	0.007	0.007
20	5	40	0.994	10687.895	1.364	1.157	0.968	2.175	4.994	3.313	2.588	1.930	0.092	0.918	0.116	0.103	0.089
		80	0.944	327.867	1.093	1.023	0.940	2.143	5.000	3.365	2.664	1.828	0.049	0.269	0.054	0.051	0.047
		100	0.889	4.988	0.991	0.938	0.874	2.180	4.999	3.392	2.598	1.831	0.038	0.069	0.042	0.040	0.037
		500	0.841	1.789	0.853	0.845	0.831	2.524	5.000	3.148	2.600	1.728	0.009	0.009	0.009	0.009	0.009
	10	40	1.821	1960.959	2.505	2.092	1.723	2.160	4.993	3.409	2.624	1.814	0.135	1.159	0.168	0.149	0.127
		80	1.763	151.765	2.040	1.904	1.727	2.127	5.000	3.409	2.695	1.769	0.071	0.351	0.079	0.075	0.069
		100	1.690	15.118	1.910	1.794	1.660	2.208	5.000	3.383	2.640	1.769	0.057	0.102	0.062	0.060	0.056
		500	1.703	3.635	1.735	1.718	1.690	2.376	5.000	3.194	2.644	1.786	0.013	0.013	0.013	0.013	0.013
	20	40	4.095	18753.531	5.317	4.516	3.780	2.208	5.000	3.369	2.613	1.810	0.218	2.228	0.258	0.233	0.205
		80	3.603	6482.204	4.229	3.927	3.525	2.109	5.000	3.391	2.672	1.828	0.108	0.628	0.119	0.114	0.107
		100	3.385	19.920	3.813	3.600	3.308	2.205	5.000	3.322	2.674	1.799	0.085	0.143	0.092	0.088	0.084
		500	3.303	7.071	3.388	3.353	3.290	2.186	5.000	3.282	2.731	1.801	0.018	0.018	0.018	0.018	0.018
	40	40	7.809	8959.419	10.513	8.904	7.291	2.214	4.997	3.392	2.664	1.733	0.320	2.592	0.376	0.343	0.305
		80	7.030	1745.680	8.159	7.608	6.842	2.187	5.000	3.331	2.668	1.814	0.158	0.747	0.171	0.165	0.155
		100	7.006	295.120	7.889	7.475	6.848	2.157	5.000	3.327	2.697	1.819	0.127	0.219	0.135	0.132	0.125
		500	6.503	13.969	6.689	6.626	6.490	2.084	5.000	3.308	2.792	1.816	0.025	0.026	0.026	0.026	0.025

EK-2 Simülasyonlarda kullanılan R programına ait kodlar

```
rm(list=ls())
set.seed(NULL)
library(splines)
library(mgcv)

klasor="E:/GAM/"
setwd(klasor)
sig=0.2#sabit
kfold=5

n=500 #40,80,100,500
bozsiz=5 #sig*bozsiz için k=5,20
bozoran=0 #5,10,20,40
simsay=10000

depo=depo2=depo3=NULL
sim=1;silsay=0

while(sim <=simsay+silsay)
{ x1=runif(n,-3,3)
  x2=runif(n,-3,3)
  E=rnorm(n,0,sig)
  boz=sample(1:n,trunc(n/100*bozoran))
  E[boz]=E[boz]+sig*bozsiz;senaryo=1 #pozitif bölge
  #E[boz]=E[boz]-sig*bozsiz;senaryo=2 #negatif bölge
  #E[boz]=E[boz]*bozsiz;senaryo=3 #merkez bölge
  y=sin(x1^2)+cos(x1^2)+sin(x2^2)+cos(x2^2)+E;model=1
  #y=sin(x1)*cos(x1)+sin(x2)*cos(x2)+E;model=2
  #y=(sin(x1)/x1)^2+(sin(x2)/x2)^2+E;model=3

  veri=data.frame(y=y,x1=x1,x2=x2)

  yn1=yn2=yn3=yn4=yn5=NULL
  ys1=ys2=ys3=ys4=ys5=NULL
  e1=e2=e3=e4=e5=NULL
  hko1=hko2=hko3=hko4=hko5=NULL
  dev=T
  rst=sample(1:n)
  nson=trunc(n*.2)
  severi=veri[rst[1:trunc(n*0.8)],]
  stveri=veri[rst[trunc(n*0.8+1):n],]

  cat("\14",sim,sim-silsay,"Y1...")
  #yöntem-1 (yerel regresyon)
  kcv1=NULL
  for(dp in seq(0.001,0.999,0.001))
  { t=0;say1=0
    for(kf in 1:kfold)
    { test=((kf-1)*(n/kfold)+1):(kf*(n/kfold))
```

```

egt=(1:n)[-test]
everi=veri[rst[egt],]
tveri=veri[rst[test],]
nt=NROW(tveri)
yn1=try(loess(y~x1+x2,data=everi,span=dp) ,silent = T)
if (!is.character(yn1))
{ys1=try(predict(yn1,data.frame(x1=tveri$x1,x2=tveri$x2)),silent = T)
if (!is.character(ys1))
{ bak1=try(sum((tveri$y-ys1)^2)/nt,silent = T)
if (!is.character(bak1) && !is.na(bak1) ) {say1=say1+1;t=t+bak1}
}
}
}
if (say1>=(kfold-3))
{prf1=t/say1
kcv1=rbind(kcv1,c(dp,prf1))
}
}
if (!is.null(kcv1))
{ if (NROW(kcv1)>10)
{opt_dp1=kcv1[which.min(kcv1[,2]),1]
yn1=try(loess(y~x1+x2,data=severi,span=opt_dp1) ,silent = T)
ys1=try(predict(yn1,data.frame(x1=stveri$x1,x2=stveri$x2)),silent = T)
el=try(tveri$y-ys1,silent = T)
hkol=try(sum((el)^2)/nson,silent = T)
}else dev==F
} else dev==F
if (is.null(hkol)) dev=F

cat("\14",sim,sim-silsay,"Y2...")
#yöntem-2 (spline regresyon)
kcv2=NULL
for(dp in 2:30)
{ t=0;say2=0
for(kf in 1:kfold)
{ test=((kf-1)*(n/kfold)+1):(kf*(n/kfold))
egt=(1:n)[-test]
everi=veri[rst[egt],]
tveri=veri[rst[test],]
nt=NROW(tveri)
yn2=try(gam(y~bs(x1,df=dp)+bs(x2,df=dp),family = gaussian(),data = everi) ,silent = T)
if (!is.character(yn2))
{ ys2=NULL
for(al in rst[test])
{
ys2=c(ys2,predict.gam(yn2,newdata=data.frame(x1=veri[al,2],x2=veri[al,3] )))
}
bak2=try(sum((tveri$y-ys2)^2)/nt,silent = T)
if (!is.character(bak2) && !is.na(bak2) ) {say2=say2+1;t=t+bak2}
}
}
}
if (say2>(kfold-3))
{ prf2=t/say2

```

```

    kcv2=rbind(kcv2,c(dp,prf2))
  }
}
if (!is.null(kcv2))
{ if (NROW(kcv2)>3)
  { opt_dp2=kcv2[which.min(kcv2[,2]),1]
    yn2=try(gam(y~bs(x1,df=opt_dp2)+bs(x2,df=opt_dp2),family = gaussian(),data = severi) ,silent = T)
    for(al in rst[trunc(n*0.8+1):n])
    { ys2=c(ys2,predict.gam(yn2,newdata=data.frame(x1=veri[al,2],x2=veri[al,3])))
    }
    e2=try(stveri$y-ys2,silent = T)
    hko2=try(sum((e2)^2)/nson,silent = T)
  }else dev==F
} else dev==F
if (is.null(hko2)) dev=F

cat("\14",sim,sim-silsay,"Y3...")
#yöntem-3 (spline düzleştirici-cubic)
kcv3=NULL
for(dp in seq(0.001,0.999,0.001))
{ t=0;say3=0
  for(kf in 1:kfold)
  { test=((kf-1)*(n/kfold)+1):(kf*(n/kfold))
    egt=(1:n)[-test]
    everi=veri[rst[egt],]
    tveri=veri[rst[test],]
    nt=NROW(tveri)
    yn3=try(gam(y~s(x1,sp=dp,bs="cr")+s(x2,sp=dp,bs="cr"),data=everi) ,silent = T)
    if (!is.character(yn3))
    { ys3=NULL
      ys3=try(predict.gam(yn3,newdata=data.frame(x1=tveri$x1,x2=tveri$x2)),silent = T)
      if (!is.character(ys3))
      { bak3=try(sum((tveri$y-ys3)^2)/nt,silent = T)
        if (!is.character(bak3) && !is.na(bak3) ) {say3=say3+1;t=t+bak3}
      }
    }
  }
}
if (say3>(kfold-3))
{ prf3=t/say3
  kcv3=rbind(kcv3,c(dp,prf3))
}
}
if (!is.null(kcv3))
{ if (NROW(kcv3)>3)
  { opt_dp3=kcv3[which.min(kcv3[,2]),1]
    yn3=try(gam(y~s(x1,sp=opt_dp3,bs="cr")+s(x2,sp=opt_dp3,bs="cr"),data=severi) ,silent = T)
    ys3=try(predict.gam(yn3,newdata=data.frame(x1=stveri$x1,x2=stveri$x2)),silent = T)
    e3=try(stveri$y-ys3,silent = T)
    hko3=try(sum((e3)^2)/nson,silent = T)
  }else dev==F
} else dev==F
if (is.null(hko3)) dev=F

```

```

cat("\14",sim,sim-silsay,"Y4...")
#yöntem-4 (spline düzleştirici-pspline)
kcv4=NULL
for(dp in seq(0.001,0.999,0.001))
{ t=0;say4=0
  for(kf in 1:kfold)
  { test=((kf-1)*(n/kfold)+1):(kf*(n/kfold))
    egt=(1:n)[-test]
    everi=veri[rst[egt],]
    tveri=veri[rst[test],]
    nt=NROW(tveri)
    yn4=try(gam(y~s(x1,sp=dp,bs="ps")+s(x2,sp=dp,bs="ps"),data=everi) ,silent = T)
    if (!is.character(yn4))
    { ys4=NULL
      ys4=try(predict.gam(yn4,newdata=data.frame(x1=tveri$x1,x2=tveri$x2)),silent = T)
      if (!is.character(ys4))
      { bak4=try(sum((tveri$y-ys4)^2)/nt,silent = T)
        if (!is.character(bak4) && !is.na(bak4) ) {say4=say4+1;t=t+bak4}
      }
    }
  }
  if (say4>(kfold-3))
  { prf4=t/say4
    kcv4=rbind(kcv4,c(dp,prf4))
  }
}
if (!is.null(kcv4))
{ if (NROW(kcv4)>3)
  { opt_dp4=kcv4[which.min(kcv4[,2]),1]
    yn4=try(gam(y~s(x1,sp=opt_dp4,bs="ps")+s(x2,sp=opt_dp4,bs="ps"),data=severi) ,silent = T)
    ys4=try(predict.gam(yn4,newdata=data.frame(x1=stveri$x1,x2=stveri$x2)),silent = T)
    e4=try(stveri$y-ys4,silent = T)
    hko4=try(sum((e4)^2)/nson,silent = T)
  }else dev==F
} else dev==F
if (is.null(hko4)) dev=F

cat("\14",sim,sim-silsay,"Y5...")
#yöntem-5 (spline düzleştirici-thin plate spline)
kcv5=NULL
for(dp in seq(0.001,0.999,0.001))
{ t=0;say5=0
  for(kf in 1:kfold)
  { test=((kf-1)*(n/kfold)+1):(kf*(n/kfold))
    egt=(1:n)[-test]
    everi=veri[rst[egt],]
    tveri=veri[rst[test],]
    nt=NROW(tveri)
    yn5=try(gam(y~s(x1,sp=dp,bs="tp")+s(x2,sp=dp,bs="tp"),data=everi) ,silent = T)
    if (!is.character(yn5))
    { ys5=NULL
      ys5=try(predict.gam(yn5,newdata=data.frame(x1=tveri$x1,x2=tveri$x2)),silent = T)
      if (!is.character(ys5))

```

```

    { bak5=try(sum((tveri$y-ys5)^2)/nt,silent = T)
      if (!is.character(bak5) && !is.na(bak5) ) {say5=say5+1;t=t+bak5}
    }
  }
}
if (say5>(kfold-3))
{ prf5=t/say5
  kcv5=rbind(kcv5,c(dp,prf5))
}
}

if (!is.null(kcv5))
{ if (NROW(kcv5)>3)
  { opt_dp5=kcv5[which.min(kcv5[,2]),1]
    yn5=try(gam(y~s(x1,sp=opt_dp5,bs="tp")+s(x2,sp=opt_dp5,bs="tp"),data=severi) ,silent = T)
    ys5=try(predict.gam(yn5,newdata=data.frame(x1=stveri$x1,x2=stveri$x2)),silent = T)
    e5=try(stveri$y-ys5,silent = T)
    hko5=try(sum((e5)^2)/nson,silent = T)
  }else dev==F
} else dev==F
if (is.null(hko5)) dev=F
if (dev==T)
{ sim=sim+1
  if (is.na(hko1)) silsay=silsay+1
  depo=rbind(depo,c(hko1,hko2,hko3,hko4,hko5))
  depo2=rbind(depo2,rank(c(hko1,hko2,hko3,hko4,hko5)))
  depo3=rbind(depo3,(c(sd(e1),sd(e2),sd(e3),sd(e4),sd(e5))/nson ))
  write.csv2(depo ,
file=paste0("ham_depo_model",model,"_snry",senaryo,"_kfold",kfold,"_n",n,"_bozsid",bozsid,"_bozoran",bozoran,".csv"),row.names = F)
  write.csv2(depo2,
file=paste0("ham_depo2_model",model,"_snry",senaryo,"_kfold",kfold,"_n",n,"_bozsid",bozsid,"_bozoran",bozoran,".csv"),row.names = F)
  write.csv2(depo3,
file=paste0("ham_depo3_model",model,"_snry",senaryo,"_kfold",kfold,"_n",n,"_bozsid",bozsid,"_bozoran",bozoran,".csv"),row.names = F)
  }
}

sil1=which(is.na(depo[,1]))
sil2=which(is.na(depo[,2]))
sil3=which(is.na(depo[,3]))
sil4=which(is.na(depo[,4]))
sil5=which(is.na(depo[,5]))
silhep=union(union(union(sil1,sil2),sil3),sil4),sil5)

h1=1-NROW(sil1)/sim;h1=trunc(h1*10000)/10000
h2=1-NROW(sil2)/sim;h2=trunc(h2*10000)/10000
h3=1-NROW(sil3)/sim;h3=trunc(h3*10000)/10000
h4=1-NROW(sil4)/sim;h4=trunc(h4*10000)/10000
h5=1-NROW(sil5)/sim;h5=trunc(h5*10000)/10000
ho=c(h1,h2,h3,h4,h5)

write.csv2(rbind(c(colMeans(depo[-silhep,]),colMeans(depo2[-silhep,]),colMeans(depo3[-silhep,]),ho)),
file=paste0("model",model,"_snry",senaryo,"_kfold",kfold,"_n",
n,"_bozsid",bozsid,"_bozoran",bozoran,".csv"),row.names = F)

```