

**DOĞRUSAL OLMAYAN REGRESYON MODELLERİNDE
PARAMETRE TAHMİN YÖNTEMLERİ, ÖNERİLER VE
KARŞILAŞTIRMALARI**

Tarhan SERİN

**DOKTORA TEZİ
İSTATİSTİK**

**GAZİ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

EKİM 2010

ANKARA

Tarhan SERİN tarafından hazırlanan “DOĞRUSAL OLMAYAN REGRESYON MODELLERİNDE PARAMETRE TAHMİN YÖNTEMLERİ, ÖNERİLER VE KARŞILAŞTIRMALARI” adlı bu tezin Doktora tezi olarak uygun olduğunu onaylarım.

Prof. Dr. Hamza GAMGAM

.....

Tez Danışmanı, İstatistik Anabilim Dalı

Bu çalışma, jürimiz tarafından oy birliği ile İstatistik Anabilim Dalında Doktora tezi olarak kabul edilmiştir.

Prof. Dr. Süleyman GÜNAY

.....

İstatistik Anabilim Dalı, Hacettepe Üniversitesi

Prof. Dr. Hamza GAMGAM

.....

İstatistik Anabilim Dalı, Gazi Üniversitesi

Prof. Dr. Fikri ÖZTÜRK

.....

İstatistik Anabilim Dalı, Ankara Üniversitesi

Prof. Dr. Reşat KASAP

.....

İstatistik Anabilim Dalı, Gazi Üniversitesi

Prof. Dr. Hasan BAL

.....

İstatistik Anabilim Dalı, Gazi Üniversitesi

Tarih: 15/10/2010

Bu tez ile G.Ü. Fen Bilimleri Enstitüsü Yönetim Kurulu Doktora derecesini onamıştır.

Prof. Dr. Bilal TOKLU

.....

Fen Bilimleri Enstitüsü Müdürü

TEZ BİLDİRİMİ

Tez içindeki bütün bilgilerin etik davranış ve akademik kurallar çerçevesinde elde edilerek sunulduğunu, ayrıca tez yazım kurallarına uygun olarak hazırlanan bu çalışmada bana ait olmayan her türlü ifade ve bilginin kaynağına eksiksiz atıf yapıldığını bildiririm.

Tarhan SERİN

**DOĞRUSAL OLMAYAN REGRESYON MODELLERİNDE PARAMETRE
TAHMİN YÖNTEMLERİ, ÖNERİLER VE KARŞILAŞTIRMALARI
(Doktora Tezi)**

Tarhan SERİN

**GAZİ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

Ekim 2010

ÖZET

Bu çalışmada parametre tahmin yaklaşımlarında temel olan en küçük kareler yöntemine ve birçok başka parametre tahmin yöntemi için bir yol gösterici olan ağırlıklandırılmış en küçük kareler yöntemine değinilmiştir. Ayrıca en çok olabilirlik tahmini ve bunun bir modifikasyonu olan yarı olabilirlik yöntemi tanıtılmıştır. Sağlam tahmin ediciler ele alınarak bu yöntemlerin geliştirilme nedenlerine değinilmiştir. Sonra parametre tahmini amacıyla Tvrdik ve diğerlerinin geliştirmiş olduğu stokastik yöntem tanıtılmıştır. Diğer yandan yapılan çalışmalar doğrultusunda bazı noktalar farkedilmiş ve bunların üzerine fikir ve görüşler geliştirilerek önerilerde bulunulmuştur. Gauss-Newton temelli algoritmaların bir çatı altında toplanabildiği gösterilmiş ve sonuçlar geliştirilerek bir öneride bulunulmuştur. Ayrıca Gauss-Newton tabanlı algoritmalarından farklı olarak stokastik araştırma stratejisine dayanan bir yöntem önerilmiştir. Birkaç süreç tasarlanıp simülasyon aracılığı ile veriler üretilmiş ve uygulama amacıyla bu çalışmada değinilen yöntemler ve yaklaşımlar kullanılarak parametre tahminleri yapıp standart hataları bulunarak sonuçları karşılaştırılmıştır. Önerilen yöntemler ile bulunan parametre tahminlerinin oldukça az sapmalı ve küçük standart hatalı oldukları görülmüştür.

Bilim Kodu : 205.1.066

Anahtar Kelimeler : Regresyon, Doğrusal Olmayan Regresyon Modeli, Doğrusal Olmayan Parametre Tahmini, Normal Denklemler, Gauss-Newton İterasyonları, Sağlam Tahminler, Tvrdik Yöntemi, Stokastik Araştırma.

Sayfa Adedi : 138

Tez Yöneticisi : Prof. Dr. Hamza GAMGAM

**PARAMETER ESTIMATION METHODS, PROPOSALS AND THEIR
COMPARISONS IN NONLINEAR REGRESSION MODELS**

(PhD. Thesis)

Tarhan SERİN

**GAZİ UNIVERSITY
INSTITUTE OF SCIENCE AND TECHNOLOGY**

October 2010

ABSTRACT

In this study, the least squares method of the parameter estimation approaches and the weighted least squares method, which is a guiding method for many other parameter estimation methods, have been referred to. Moreover, the maximum likelihood estimation and the quasi likelihood method, which is a modified version of the maximum likelihood estimation has been introduced. By addressing robust estimators, the reasons for developing these methods have been mentioned in the study. Then, the stochastic method, which is developed by Tvrdik et al., has been introduced for parameter estimation. On the other hand, certain points have been realized, whereas some theories and opinions developed as well as some suggestions were made on the approaches. It was shown that algorithms based on Gauss-Newton can be grouped together and a suggestion was made via generalizing the results. Besides, a method based on stochastic search algorithm has been offered as distinct from Gauss-Newton based algorithms. A few processes have been designed and data has been gathered via simulation; moreover, the results of the parameter estimation and standard errors have been identified and were compared by using the methods and approaches mentioned in this study for application. It was found that the proposed methods and parameter estimations that were found were slightly biased and with small standard error.

Science Code :205.1.066

Key Words :Regression, Nonlinear Regression, Nonlinear Parameter Estimation, Normal Equations, Gauss-Newton Iterations, Robust Estimations, Tvrdik's Method, Stochastic Search.

Page Number : 138

Adviser : Prof. Dr. Hamza GAMGAM

TEŞEKKÜR

Çalışmalarımnda çok önemli katkıları olan tez danışmanım Prof. Dr. Hamza GAMGAM'a teşekkürü bir borç olarak biliyorum. Tüm tez çalışmaları süresi boyunca değerli katkılarını ve ilgisini esirgemeyen Tez komitesi üyeleri olan Prof. Dr. Fikri ÖZTÜRK'e ve Prof. Dr. Reşat KASAP'a yardımları ve anlayışları için teşekkürlerimi ifade etmek istiyorum. Tez çalışmalarım süresi boyunca moral ve teknik desteklerini esirgemeyen İstatistik bölümü öğretim görevlilerini ayrıca anmak ve onlara teşekkür etmek istiyorum.

Anlayışlarını ve katkılarını hiç bir fırsatta esirgemeyen aileme, anneme, babama ve kardeşime duyduğum minnettarlıkla teşekkür ediyorum.

İÇİNDEKİLER

	Sayfa
ÖZET.....	iv
ABSTRACT.....	vi
TEŞEKKÜR.....	viii
İÇİNDEKİLER.....	ix
ÇİZELGELERİN LİSTESİ.....	xiii
ŞEKİLLERİN LİSTESİ.....	xvi
SİMGELER VE KISALTMALAR.....	xviii
1. GİRİŞ.....	1
2. DOĞRUSAL OLMAYAN REGRESYON MODELLERİNDE PARAMETRE TAHMİNİ.....	4
2.1. En Küçük Kareler Yöntemi.....	4
2.2. Genelleştirilmiş En Küçük Kareler Yöntemi.....	9
2.3. Parametrelerin Başlangıç Değerlerinin Seçimi.....	12
2.4. Kareler Toplamı Fonksiyonunun Minimizasyonu ve İteratif Algoritmanın Seçimi.....	14
3. DİĞER TAHMİN YÖNTEMLERİ.....	16
3.1. En Çok Olabilirlik Tahmini.....	16
3.2. Yarı (Quasi) Olabilirlik Tahmini.....	19
3.3. L1-Norm Minimizasyonu.....	22
3.4. Sağlam (Robust) Tahmin Ediciler.....	23
3.4.1. Sağlam (Lp) regresyon.....	24
3.4.2. M-tahmin ediciler.....	26

Sayfa

4. PARAMETRE TAHMİNİ İÇİN TVRDİK YÖNTEMİ.....	31
4.1. Nelder-Mead Yöntemi.....	31
4.2. Nelder-Mead Yönteminin Genelleştirilmesi.....	37
4.3. Tvrdik Yöntemi.....	40
4.3.1. Global optimizasyon.....	41
4.3.2. Kontrollü rassal araştırma (Controlled random search, CRS).....	42
5. YAKLAŞIM ÖNERİLERİ.....	47
5.1. Bir Stokastik Yaklaşım Önerisi.....	47
5.2. Yöntemlerin Genelleştirilmesi.....	51
5.3. Hata Varyans-Kovaryans Matrisi ve Optimal Sağlam Fonksiyonu.....	61
6. PARAMETRE TAHMİN YÖNTEMLERİNİN KARŞILAŞTIRILMASI İÇİN SİMÜLASYON ÇALIŞMASI.....	64
6.1. Model 1a Verisi.....	65
6.1.1. Model 1a verisi ve en küçük kareler yöntemi.....	66
6.1.2. Model 1a süreci için simülasyon çalışması.....	68
6.2. Model 1b Verisi.....	69
6.2.1. Model 1b verisi ve en küçük kareler yöntemi.....	70
6.2.2. Model 1b verisi ve mutlak değerler toplamı.....	72
6.2.3. Model 1b verisi ve M-tahmin edici.....	72
6.2.4. Model 1b süreci için simülasyon çalışması.....	77
6.3. Model 1c Verisi.....	78
6.3.1. Model 1c verisi ve genelleştirilmiş en küçük kareler.....	80
6.3.2. Model 1c verisi ve yarı olabilirlik yöntemi.....	87

Sayfa

6.3.3. Model 1c süreci için simülasyon çalışması.....	92
6.4. Model 1d Verisi.....	93
6.4.1. Model 1d verisi ve en küçük kareler yöntemi.....	94
6.4.2. Model 1d verisi ve M-tahmin edici.....	95
6.4.3. Model 1d süreci için simülasyon çalışması.....	96
6.5. Model 2a Verisi.....	97
6.5.1. Model 2a verisi ve en küçük kareler yöntemi.....	98
6.5.2. Model 2a süreci için simülasyon çalışması.....	100
6.6. Model 2b Verisi.....	101
6.6.1. Model 2b verisi ve genelleştirilmiş en küçük kareler.....	102
6.6.2. Model 2b verisi ve yarı olabilirlik yöntemi.....	106
6.6.3. Model 2b süreci için simülasyon çalışması.....	109
6.7. Model 2c Verisi.....	110
6.7.1. Model 2c verisi ve parametre tahmini.....	112
6.7.2. Model 2c süreci için simülasyon çalışması.....	115
6.8. Model 3a Verisi.....	116
6.8.1. Model 3a verisi ve en küçük kareler yöntemi.....	118
6.8.2. Model 3a süreci için simülasyon çalışması.....	118
6.9. Stokastik Araştırma Stratejilerinin Performansına Yönelik Simülasyon Çalışması.....	119
6.9.1. Model 1a verisi ve Tvrdik yöntemi.....	119
6.9.2. Model 1a verisi ve önerilen stokastik yaklaşım.....	120
6.9.3. Model 1a süreci için stokastik araştırmalarının performansına yönelik yapılan simülasyon çalışması.....	122

	Sayfa
6.9.4. Model 1c verisi ve Tvrdik yöntemi.....	122
6.9.5. Model 1c verisi ve önerilen yöntem.....	123
6.9.6. Model 3a süreci için stokastik araştırması simülasyonu çalışması.....	124
7. SONUÇ.....	126
KAYNAKLAR.....	129
EKLER.....	134
EK-1 Normal Denklemler.....	135
ÖZGEÇMİŞ.....	138

ÇİZELGELARİN LİSTESİ

Çizelge	Sayfa
Çizelge 3.1. $z_i = e_i/s$ olmak üzere en çok kullanılan $\rho(z)$ sağlam fonksiyon.....	29
Çizelge 6.1. Model 1 aracılığı ile üretilen Model 1a verisi.....	65
Çizelge 6.2. Model 1a verisi için Gauss-Newton iterasyonu sonuçları.....	67
Çizelge 6.3. Model 1a süreci örneklerine uygulanan yöntemler ve sonuçları.....	68
Çizelge 6.4. Model 1b verisi.....	69
Çizelge 6.5. Model 1b verisi için Gauss-Newton iterasyonu sonuçları.....	70
Çizelge 6.6. Andrew dalga fonksiyonunun kullanıldığı M-tahmin edici ile yapılan iterasyon sonuçları.....	73
Çizelge 6.7. Andrew dalga fonksiyonu kullanılarak M-tahmin edici ile elde edilen ağırlıklar.....	74
Çizelge 6.8. Huber t fonksiyonunun kullanıldığı M-tahmin edici ile yapılan iterasyon sonuçları.....	76
Çizelge 6.9. Huber t fonksiyonu kullanılarak M-tahmin edici ile elde edilen ağırlıklar.....	77
Çizelge 6.10. Model 1b süreci örneklerine uygulanan yöntemler ve sonuçları.....	78
Çizelge 6.11. Model 1 aracılığı ile üretilen Model 1c verisi.....	79
Çizelge 6.12. Model 1c verisi için Gauss-Newton iterasyonu sonuçları.....	80
Çizelge 6.13. Gauss-Newton iterasyonları.....	81
Çizelge 6.14. i. gözlem düzeyine bağlı olarak K matrisinin i. köşegen elemanı.....	85
Çizelge 6.15. Gauss-Newton iterasyonu sonuçları.....	85
Çizelge 6.16. Model 1c verisi için yarı olabilirlik yöntemi iterasyonları.....	89
Çizelge 6.17. Model 1c süreci örneklerine uygulanan yöntemler ve sonuçları.....	93
Çizelge 6.18. İki aykırı değer barındıran 20 gözlemlili Model 1d verisi.....	94
Çizelge 6.19. Model 1d verisi için en küçük kareler iterasyonu sonuçları.....	95

Çizelge	Sayfa
Çizelge 6.20. İteratif tekrar ağırlıklandırılmalı en küçük kareler iterasyonu sonuçları.....	96
Çizelge 6.21. Model 1d süreci örneklerine uygulanan yöntemler ve sonuçları.....	97
Çizelge 6.22. Model 2 yardımıyla üretilen Model 2a verisi.....	98
Çizelge 6.23. Gauss-Newton iterasyonu sonuçları.....	99
Çizelge 6.24. Model 2a süreci örneklerine uygulanan yöntemler ve sonuçları.....	100
Çizelge 6.25. Model 2 yardımıyla üretilen Model 2b verisi.....	101
Çizelge 6.26. Gauss-Newton iterasyonları.....	102
Çizelge 6.27. Gözlem düzeylerine bağlı olarak K matrisinin köşegen elemanları.....	104
Çizelge 6.28. Model 2b verisi için Gauss-Newton iterasyonu sonuçları.....	104
Çizelge 6.29. Model 2b verisi için yarı olabilirlik tahmin yöntemi iterasyonları.....	107
Çizelge 6.30. Değişken varyanslı Model 2b süreci örneklerine uygulanan yöntemler ve sonuçları.....	110
Çizelge 6.31. Model 2 yardımıyla üretilen Model 2c verisi.....	111
Çizelge 6.32. Yöntemlerin genelleştirilmesi yaklaşımının Gauss-Newton iterasyonları.....	114
Çizelge 6.33. Aykırı değerler barındıran değişken varyanslı Model 2c süreci örneklerine uygulanan yöntemler ve sonuçları.....	115
Çizelge 6.34. Model 3 aracılığı ile üretilen Model 3a verisi.....	117
Çizelge 6.35. Model 3a süreci örneklerinin parametre tahminleri.....	119
Çizelge 6.36. Tvrdik yöntemine göre gerçekleşen yakınsama.....	120
Çizelge 6.37. Tvrdik yönteminin son iterasyon tablosu.....	120
Çizelge 6.38. Beşinci bölümde önerilen yaklaşımla gerçekleşen yakınsama.....	121
Çizelge 6.39. Önerilen yaklaşımın son iterasyon tablosu.....	121

Çizelge	Sayfa
Çizelge 6.40. Tvrdik yöntemi ve önerilen yöntemle elde edilen tahminlerin karşılaştırılması.....	122
Çizelge 6.41. Tvrdik yöntemine göre gerçekleşen yakınsama.....	123
Çizelge 6.42. Tvrdik yönteminin son iterasyon tablosu.....	123
Çizelge 6.43. Önerilen yöntemle gerçekleşen yakınsama.....	124
Çizelge 6.44. Önerilen yöntemin son iterasyon tablosu.....	124
Çizelge 6.45. Tvrdik yöntemi ve önerilen yöntemle elde edilen tahminlerin karşılaştırılması.....	125

ŞEKİLLERİN LİSTESİ

Şekil	Sayfa
Şekil 4.1: Nelder-Mead yönteminde LGH üçgeni, C orta noktası ve R yansıma noktası.....	34
Şekil 4.2: LGH üçgeni, R noktası ve genişletmeyle elde edilen E noktası.....	35
Şekil 4.3: Nelder-Mead yönteminde P_1 ve P_2 karşılaştırılarak bulunan P büzülme noktası.....	36
Şekil 4.4: L noktasına doğru çekilen üçgen.....	37
Şekil 6.1. Model 1a verisinin diyagramı.....	66
Şekil 6.2. Model 1b verisi için diyagram.....	70
Şekil 6.3. Model 1b verisi ile en küçük kareler eğrisi.....	71
Şekil 6.4. Model 1b verisi için L1-norm regresyon yöntemi ile bulunan eğri.....	72
Şekil 6.5. Model 1b verisi için Andrew dalga fonksiyonu kullanılarak M-tahmin edici aracılığı ile bulunan eğri.....	74
Şekil 6.6. Model 1b verisi için Huber t fonksiyonu kullanılarak M-tahmin edici aracılığı ile bulunan eğri.....	76
Şekil 6.7. Model 1c verisinin diyagramı.....	79
Şekil 6.8. Ağırlıklandırma yapılmadan elde edilen artıklar için serpm diyagramı.....	82
Şekil 6.9. Artıkların mutlak değerleri saçılma grafiği.....	82
Şekil 6.10. Mutlak artıklar, A_t merkezi hareketli ortalamaları ve m_t serisi.....	84
Şekil 6.11. Mutlak ağırlıklandırılmış artıklar, A_t merkezi hareketli ortalamaları ve m_t serisi.....	86
Şekil 6.12. Artıkların mutlak değeri, A_t merkezi hareketli ortalamaları ve m_t serisi.....	89
Şekil 6.13. Ağırlıklandırılmış artıkların mutlak değerleri grafiği.....	90
Şekil 6.14. Model 1d verisinin diyagramı.....	94

Şekil	Sayfa
Şekil 6.15. Model 2a verisinin diyagramı.....	98
Şekil 6.16. Model 2b verisinin diyagramı.....	102
Şekil 6.17. Ağırlıklandırma yapılmadan elde edilen model artıkları serpmeye diyagramı.....	103
Şekil 6.18. Mutlak artıklar, A_t merkezi hareketli ortalamaları ve m_t serisi.....	103
Şekil 6.19. Mutlak ağırlıklandırılmış artıklar, A_t ortalamaları ve m_t serisi.....	105
Şekil 6.20. Artıkların mutlak değeri, A_t merkezi hareketli ortalamaları ve m_t serisi.....	108
Şekil 6.21. Ağırlıklandırılmış artıkların mutlak değeri grafiği.....	108
Şekil 6.22. Model 2c verisinin diyagramı.....	111
Şekil 6.23. Model 3a verisinin diyagramı.....	117

SİMGELER VE KISALTMALAR

Bu çalışmada kullanılmış bazı simgeler ve kısaltmalar, açıklamaları ile birlikte aşağıda sunulmuştur.

Simgeler	Açıklama
$E(\cdot)$	Beklenen değer operatörü
$V(\cdot)$	Varyans-kovaryans operatörü
θ_*	Parametre gerçek değeri
$\hat{\theta}$	Parametre tahmini
θ	Parametre uzayındaki herhangi bir değer
β, δ	Parametre fark vektörü
ε	Gerçek parametre değeri ile teorik modelin hataları
e	Parametre tahmini ile tahmin modelinin artıkları
r	Herhangi bir θ değeri ile modelin artıkları
σ^2	Hata varyansı
s^2	Hata varyansı tahmini
V	Hata varyans-kovaryans matrisi
K	V matrisinin kare kök matrisi
ι	Dönüştürülmüş hata vektörü
n	Model için kullanılan gözlem sayısı, örnek hacmi
p	Model içindeki parametre sayısı
Z	Kısmi türevler matrisi ile tahmin uzayı
λ	Levenberg-Marquardt algoritması λ katsayısı
R	Dönüştürülmüş kısmi türevler matrisi
P	Projeksiyon matris operatörü
L	Log olabilirlik fonksiyonu
l	Log yarı-olabilirlik fonksiyonu
I	Bilgi matrisi

Simgeler**Açıklama** ρ

Sağlam (robust) fonksiyonu

 ψ

Etki fonksiyonu

 w

Ağırlık fonksiyonu

 k

Alt indiste iterasyon numarası

 X_k Bir X matrisinin k numaralı iterasyonu

1. GİRİŞ

Bu çalışma doğrusal olmayan regresyon modellerinin parametre tahminleri üzerinedir. Doğrusal olmayan regresyon modellerinde model parametreleri doğrusal değildir ancak bağımsız değişkenler doğrusal ya da doğrusal olmayan yapıda olabilir. Gündelik yaşamda çoğu fonksiyonel bağıntı, katsayıları açısından doğrusal olmayan bir yapıdadır. Doğrusal olmamanın iki kaynağı vardır. Doğrusal olmama durumu, modelin geometrisinden veya model parametrelerinden kaynaklanmaktadır [Seber ve Wild, 1989].

Konuyla ilgili olarak Günay (1972) tarafından yapılan çalışmada bazı parametre tahmin edicileri dahil en çok olabilirlik ile en küçük kareler tahmin edicileri incelenerek bu tahmin ediciler yeterlilikleri, yanlılıkları, etkinlikleri ve tutarlılıkları açısından karşılaştırılmıştır. Aydın (1992) tarafından yapılan çalışmada genel doğrusal olmayan modellerin normal dağılıma sahip hata terimlerine ilişkin varyans-kovaryans matrisi için ileri sürülen önsel dağılımın olabilirlik fonksiyonu üzerindeki etkileri incelenmiştir. Candan (1995) tarafından yapılan çalışmada sağlam yöntemleri ve sağlam regresyon yaklaşımları incelenmiştir. Çağlar (1995) tarafından yapılan çalışmada doğrusal olmayan regresyon modellerinde Gauss-Newton yaklaşımı yardımıyla olabilirlik fonksiyonu maksimize edilip model parametreleri tahmin edilmiştir. Aksoy (1996) tarafından yapılan çalışmada ise doğrusal olmayan regresyon modeli kurulduktan sonra model artıkları arasında beliren otokorelasyonlar da modele dahil edilip doğrusal olmayan parametre tahminleri doğrusallaştırma yöntemiyle yapılmıştır.

Doğrusal olmayan modeller konusunda yapılan çalışmalarda, doğrusal olmayan modellerin verilere uydurulmasının zor ve uğraştırıcı olduğu görülmüştür. Bu modellerde tahmin için kullanılan algoritmalar yakınsama konusunda birçok durumda yetersiz kalmaktadır [Box, 1971; Gallant, 1987]. Bu modellerin tahmininde kullanılan yaklaşımlardan bir tanesi modelin doğrusallaştırılmasıdır. Bu tür modeller aslen doğrusal olarak anılır [Bates ve Watts, 1988]. Bu durumda eğer tüm varsayımlar sağlanıyorsa olağan en küçük kareler yöntemi ile bu modellerin

parametreleri tahmin edilebilir [Miller, 1981; McKee ve McClendon, 1994]; fakat varsayımlar sağlanmıyorsa bulunan tahminler yanıltıcı bulunur [Chatterjee ve Price, 1991; Hamilton, 1992]. Ayrıca, bu yönde yapılan doğrusallaştırma girişimleri ile elde edilen yapıların uygunluğunda bazı belirsizlikler bulunabilmektedir. Bununla birlikte parametre tahminlerinin bazen istenmeyen özellikleri olabilmektedir [Gallant, 1987].

Doğrusal olmayan modelin parametrelerine göre doğrusallaştırma işlemi cebirsel olarak gerçekleştirilemiyorsa bu modeller doğrusal olmayan modeller olarak anılır [Gallant, 1987; Bates ve Watts, 1988]. Bu modellerin tahmininde iki yöntem söz konusudur. Birincisi, model uydurma için güçlü ve etkili bir iteratif algoritmanın kullanılmasıdır. Bu algoritmalar karmaşık modeller için tasarlanmıştır. İkincisi, Gauss-Newton ve Newton-Raphson algoritmalarının uyarlamalarına dayanır [Seber ve Wild, 1989; Bronson ve Naadimuthu, 1997]. Uyarlanmamış algoritmalar yakınsama sorunları dolayısıyla güvenli değildir. Bundan dolayı algoritmalar farklı modeller için farklı şekillerde uyarlanarak kullanılmaktadır [Seber ve Wild, 1989].

Doğrusal olmayan model tahminlerinde Gauss-Newton ve Newton-Raphson algoritmaları yaygın olarak kullanılmasına rağmen bu algoritmalarda başlangıç noktasının seçimi oldukça önemlidir. Uygun olmayan başlangıç noktaları uygun olmayan çözümler elde edilmesine neden olabileceği gibi çözüm zamanı bakımından da ekonomik olmayabilir. Bu nedenle başlangıç noktasının seçiminden çok fazla etkilenmeyen bazı sezgisel yöntemler önerilmiştir. Çok değişkenli bir fonksiyonun yerel minimumunu bulduran bir algoritma Nelder ve Mead (1964) tarafından geliştirilmiştir. Bu algoritmada Nelder ve Mead tepe noktalarıyla ortaya çıkan basit geometrik şekle simpleks adını vermişlerdir. Bundan dolayı bir fonksiyonun minimizasyonunu araştıran bu metoda “Nelder-Mead Simpleks Algoritması” yada kısaca “simpleks algoritması” da denir. Ancak bu adlandırma doğrusal programlama problemlerinin çözümünde kullanılan Dantzig ve diğerlerinin metoduyla karıştırılabildiğinden bu yönteme “Nelder-Mead Politop Algoritması” da denmektedir [Nash ve Walker-Smith, 1987]. Daha sonraları Nelder-Mead algoritmasına dayalı farklı yöntemler önerilmiştir. Bu yöntemlerden biri, Tvrdik ve arkadaşları (2006) tarafından önerilen stokastik bir algoritmadır. Bu algoritma

uyarlamış populasyon-tabanlı yaklaşıma dayalıdır. Bu yöntem, fonksiyonun belirli bir bölgesinde birden fazla noktaya (populasyona) dayalı olarak arama yapan stokastik bir yaklaşımdır. Dolayısıyla bu yöntemin yerel minimum ve yerel maksimum noktalara takılma olasılığı oldukça düşüktür. Bu yöntemlere ilişkin ayrıntılar çalışmanın dördüncü bölümünde verilmiştir.

Bu çalışmada doğrusal olmayan regresyon modellerinde parametre tahmini için iki farklı yöntem önerilmektedir. Bu yöntemlerden biri stokastik arama stratejisi dahilinde olan stokastik yöntemdir. Diğer Gauss-Newton temeline dayalı iteratif bir yöntemdir.

Tvrdik algoritmasında Nelder-Mead yöntemindeki tepe noktalarının (simpleks) her bir iterasyonda farklı seçimleri söz konusudur. Bu yöntemde tepe noktalarının her bir iterasyonda değişmesi elde edilen sonuçlarda iyileşmenin oldukça yavaş olmasını beraberinde getirmektedir. Bu çalışmada önerilen stokastik yöntemde ise her bir iterasyonda mevcut tepe noktaları korunmaktadır. Bu sayede yöntem uygun çözüme daha hızlı yakınsamaktadır. Çalışmada önerilen iteratif yöntemde ise aykırı değerler ve değişen varyans durumlarında parametre tahmini yapılabilir. Bu yöntem diğer bazı iteratif yöntemleri içinde barındıran genel bir algoritmaya dayalıdır. Önerilen yöntemlere ilişkin ayrıntılar çalışmanın beşinci bölümünde yer verilmiştir.

Bu çalışmanın ikinci bölümünde doğrusal olmayan regresyon modelleri genel olarak tanıtılmıştır. Ayrıca bu bölümde en küçük kareler ve ağırlıklandırılmış en küçük kareler yöntemine değinilmiştir. Üçüncü bölümde doğrusal olmayan regresyon modellerinde parametre tahminleri için kullanılan iteratif yöntemlerden bazıları tanıtılmıştır. Dördüncü bölümde Tvrdik ve diğerlerinin (2006) doğrusal olmayan regresyon modeli parametrelerinin tahminine yönelik yöntemi ele alınmıştır. Beşinci bölümde önerilen tahmin yöntemleri açıklanmıştır. Altıncı bölümde değinilen tahmin yöntemleri ile önerilen yöntemler simülasyon sonuçlarına göre karşılaştırılmıştır. Son bölümde elde edilen bulgular ve sonuçlar tartışılmıştır.

2. DOĞRUSAL OLMAYAN REGRESYON MODELLERİNDE PARAMETRE TAHMİNİ

Parametreler açısından doğrusal olmayan regresyon modeli

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

şeklinde ifade edilir. Bu modelde ε_i rassal hata terimi olup, bu rassal değişkenlerin bağımsız ve $\varepsilon_i \sim (0, \sigma^2)$ oldukları varsayılır. Modelde f fonksiyonu beklenen fonksiyonu, x_i vektörü bağımsız değişkenleri ve θ vektörü p tane bilinmeyen doğrusal olmayan parametreleri ifade eder. Parametre vektörünün gerçek değeri θ_* şeklinde ifade edilmiştir. Doğrusal olmayan regresyon modellerinde $E(y_i)$ parametrelerin doğrusal olmayan bir fonksiyonu olmak üzere parametrelerine göre türevlerinden en az bir tanesi en az bir parametreye bağımlıdır [Bates ve Watts, 1988; Bierens ve Gallant, 1997], bu özellik parametrelere göre doğrusal olmamanın bir göstergesidir. Varsayımları bu şekilde verilen modelin parametrelerinin tahminleri en küçük kareler yöntemiyle bulunabilir.

2.1. En Küçük Kareler Yöntemi

Herhangi bir θ için $r_i = y_i - f(x_i, \theta)$ artıkları üzerinden doğrusal olmayan model için artık kareler toplamı fonksiyonu

$$S(\theta) = \sum_{i=1}^n [y_i - f(x_i, \theta)]^2 \quad (2.2)$$

şeklinde ifade edilir. Buna bağlı olarak en küçük kareler tahmini $\hat{\theta}$ ile gösterilir ve $S(\theta)$ kareler toplamı fonksiyonunu minimize eder. $S(\theta)$ fonksiyonunda $f(x_i, \theta)$ fonksiyonları p bileşenli θ vektörüne göre türevlenebilirse minimum noktada

$\frac{\partial S(\theta)}{\partial \theta_j} \Big|_{\hat{\theta}} = 0, j = 1, 2, \dots, p$ sonuçları geçerlidir. Bu denklemlerin sayısal yöntemlerle çözümüyle $\hat{\theta}$ parametre tahmin vektörünün bileşenleri olan $\hat{\theta}_j$ parametre tahminleri elde edilmiş olunur. Bununla birlikte $\hat{\theta}$ parametre tahminine bağlı olarak $e_i = y_i - f(x_i, \hat{\theta})$ model artıklarını göstermektedir. Parametre vektörünün gerçek değeri θ_* şeklinde ve herhangi bir değer θ olarak ifade edilmiştir. Doğrusal olmayan regresyon modellerinin parametre tahminleri aşamalı olarak yapıldığından herhangi bir θ vektörünün bir değişken gibi ele alınarak incelenmesi gerekmektedir. Parametre vektörünün iyileştirilmesi arka arkaya yapılan iteratif tahminlerle gerçekleştirilmektedir. Bu tahminler ilgili iterasyonu bağladığından iyileştirilen θ değerleri birer ara tahmin olmakla birlikte sadece iterasyon numarası şeklinde indislenmiştir. Bu değerlerin sonuç $\hat{\theta}$ parametre tahmininden ve gerçek θ_* parametre değerinden farklı olduğu ifade edilir.

n gözlemlili, p parametrelili doğrusal olmayan regresyon modeli kısaca $y = f(\theta_*) + \varepsilon$ şeklinde gösterilerek $f(\theta)$ vektör fonksiyonunun bileşenlerinin θ parametrelerine göre kısmi türevleri Z matrisinin satırlarını oluşturmak üzere aşağıdaki matris notasyonu verilmiştir.

$$f(x_i, \theta) = f_i(\theta)$$

$$f(\theta) = (f_1(\theta), f_2(\theta), \dots, f_n(\theta))^t \quad (2.3)$$

$$Z = \frac{\partial f(\theta)}{\partial \theta^t} = \left[\left(\frac{\partial f_i(\theta)}{\partial \theta_j} \right)_{ij} \right]$$

θ parametre vektörünün gerçek değeri θ_* ve parametre vektörünün tahmini $\hat{\theta}$ olmak üzere Z matrisi aşağıdaki gibi gösterilmiştir.

$$Z_* = Z(\theta_*) = \frac{\partial f(\theta)}{\partial \theta^t} \Big|_{\theta_*} = \left[\left(\frac{\partial f_i(\theta)}{\partial \theta_j} \right)_{ij} \right] \Big|_{\theta_*} \quad (2.4)$$

$$\hat{Z} = Z(\hat{\theta}) = \frac{\partial f(\theta)}{\partial \theta^t} \Big|_{\hat{\theta}} = \left[\left(\frac{\partial f_i(\theta)}{\partial \theta_j} \right)_{ij} \right] \Big|_{\hat{\theta}}$$

$S(\theta)$ artık kareler toplamı fonksiyonu

$$S(\theta) = (y - f(\theta))^t (y - f(\theta)) = \|y - f(\theta)\|^2 \quad (2.5)$$

şeklinde yazılarak θ_k değeri civarında $f(\theta)$ vektörünün Taylor serisi açılımı

$Z_k = Z(\theta_k) = \frac{\partial f(\theta)}{\partial \theta^t} \Big|_{\theta_k}$ ile $f(\theta) \cong f(\theta_k) + Z_k(\theta - \theta_k)$ kullanılarak artık kareler toplamı $S(\theta) \cong \|y - f(\theta_k) - Z_k(\theta - \theta_k)\|^2$ şekline getirilir ve θ vektörüne göre minimize edilmesiyle çözüm θ_{k+1} şeklinde gösterilerek aşağıdaki sonuç elde edilir.

$$\theta_{k+1} = \theta_k + (Z_k^t Z_k)^{-1} Z_k^t [y - f(\theta_k)] \quad (2.6)$$

$$\lim_{k \rightarrow \infty} \theta_k = \hat{\theta}$$

Bu iterasyonlar Gauss-Newton yöntemi olarak bilinir [Bates ve Watts, 1988]. Bu iteratif işlemler örneğin $\delta = 10^{-6}$ için

$$\left| \frac{\theta_{j,k+1} - \theta_{j,k}}{\theta_{j,k}} \right| < \delta, \quad j = 1, 2, \dots, p \quad (2.7)$$

şartı sağlanana kadar devam ettirilir [Denison ve dig., 2003].

$S(\theta)$ artık kareler toplamı fonksiyonu minimum noktada

$$\frac{\partial S(\theta)}{\partial \theta} \Big|_{\hat{\theta}} = -2 \left(\frac{\partial f(\theta)}{\partial \theta^t} \right)^t \Big|_{\hat{\theta}} (y - f(\hat{\theta})) = 0 \quad (2.8)$$

olduğundan artık vektörü $y - f(\hat{\theta}) = e$ ile doğrusal olmayan regresyon modeli için normal denklemler $\hat{Z}^t (y - f(\hat{\theta})) = 0 \rightarrow \hat{Z}^t e = 0$ şeklinde yazılır. Bu ifade doğrusallaştırma dolayısı ile doğrusal regresyon modelindeki normal denklemler ile karşılaştırma kurar [Gallant, 1987]. Bu şekilde doğrusal olmayan regresyon modelleri için normal denklemler oluşturulmuş olunur.

Gerçek θ_* değeri civarında $f(\theta)$ vektörünün $Z_* = Z(\theta_*)$ ile Taylor serisi açılımı $f(\theta) \cong f(\theta_*) + Z_*(\theta - \theta_*)$ kullanılarak $S(\theta) \cong \|y - f(\theta_*) - Z_*(\theta - \theta_*)\|^2$ artık kareler toplamı θ vektörüne göre minimize edilmesiyle

$$\hat{\theta} = \theta_* + (Z_*^t Z_*)^{-1} Z_*^t \varepsilon \quad (2.9)$$

elde edilir. Bu ifade kullanılarak

$$E(\hat{\theta}) = \theta_* \quad , \quad V(\hat{\theta}) = \sigma^2 (Z_*^t Z_*)^{-1} \quad (2.10)$$

sonuçları bulunur. Büyük hacimli n için $\hat{Z} = Z(\hat{\theta})$, $S(\hat{\theta}) = \|y - f(\hat{\theta})\|^2$ ve $s^2 = S(\hat{\theta})/(n - p)$ olmak üzere $\hat{\theta}$ tahmininin varyans-kovaryans matrisi tahmini $\hat{V}(\hat{\theta}) = s^2 (\hat{Z}^t \hat{Z})^{-1}$ olarak bulunur. Burada $\hat{Z}$ matrisi (yaklaşık olarak) en son iterasyonla elde edilen $\hat{\theta}$ en küçük kareler tahmini vektörüne bağlı olarak elde edilen kısmi türevler $\lim_{k \rightarrow \infty} Z_k = \hat{Z} = Z(\hat{\theta})$ matrisidir. Eğer n örnek çapı yeterince

büyükse asimptotik sonuçlar doğrultusunda en küçük varyanslı parametre tahminleri bulunmuş olur. Örnek çapı arttıkça ve sonsuza yaklaştıkça yansız ve en küçük varyans özelliklerine yaklaşırlar [Bates ve Watts, 1988]. Bu şekilde istatistiksel testler ve güven aralıkları yaklaşık bir uygunlukta bulunabilir [Gallant, 1987; Bierens ve Gallant, 1997].

Gauss-Newton yöntemi bazı durumlarda çok yavaş ilerleyebilir ve bunun için birçok iterasyonun yapılması gerekebilir. Bazen $S(\theta_k)$ değerindeki artışa bağlı olarak yanlış bir yönde ilerlenebilir. Gauss-Newton algoritmasının başarısının yükseltilebilmesi için birçok yaklaşım önerilmiştir. Bunlardan biri kesirli artışların kullanılmasıdır [Bates ve Watts, 1988]. Buna göre k . iterasyon $\hat{\beta}_k = (Z_k^t Z_k)^{-1} Z_k^t [y - f(\theta_k)]$ standart artış vektörü ile $\theta_{k+1} = \theta_k + \hat{\beta}_k$ olmak üzere $S(\theta_{k+1}) < S(\theta_k)$ ise bir sonraki iterasyona geçilir; ancak $S(\theta_{k+1}) > S(\theta_k)$ ise $\hat{\beta}_k$ yerine $\frac{1}{2} \hat{\beta}_k$ artış vektörü olarak kullanılır. Belirli sayıda denemeden sonra $S(\theta_{k+1})$ değerinde bir azalma sağlanmıyorsa işlemler durdurulur. Temel Gauss-Newton algoritmasında yapılan diğer bir uyarılama Levenberg-Marquardt algoritmasıdır. Levenberg-Marquardt algoritmasına göre $k+1$ iterasyonu

$$\theta_{k+1} = \theta_k + (Z_k^t Z_k + \lambda_k I_p)^{-1} Z_k^t [y - f(\theta_k)] \quad (2.11)$$

şeklinde bulunur. Her bir aşamada artık kareler toplamını düşürecek şekilde bir $\lambda_k > 0$ değeri bulunulmaya çalışılır. Bunun için genel olarak iki farklı yaklaşım bulunmaktadır. Örneğin her bir iterasyonda $\lambda_k = 10^{-8}$ den başlanarak $S(\theta_{k+1}) < S(\theta_k)$ koşulu sağlanana kadar λ_k tekrar tekrar 10 ile çarpılıp artırılarak deneme yanılmayla hesaplanır. Bu şekilde minimum λ_k bulunur. Yada her bir iterasyonda $S(\theta_{k+1}) < S(\theta_k)$ koşulu sağlandığı sürece λ_k büyük bir değerden onar onar küçültülür. Buradaki genel strateji her bir iterasyonda artık kareler toplamını azaltacak şekilde λ_k değerini mümkün olan en küçük değerinde tutmaktır [Bates ve

Watts, 1988; Seber ve Wild, 1989; Denison ve dig., 2003]. Bu yaklaşım Marquardt uzlaşması olarak bilinir, çünkü artış vektörü Gauss-Newton ile en hızlı azalış yönü arasında bir yerde bulunur [Nash, 1979; Bates ve Watts, 1988].

2.2. Genelleştirilmiş En Küçük Kareler Yöntemi

Bu yöntem ağırlıklandırılmış en küçük kareler yöntemi olarak da bilinir. Genelleştirilmiş yada ağırlıklandırılmış en küçük kareler yöntemi çatısı altında birçok yaklaşım toplanmaktadır [Denison ve dig., 2003]. Önerilen regresyon modelinin hataları homojenlik (bağımsız ve sabit varyanslı) şartlarını sağlamadığında en küçük kareler yöntemiyle elde edilen parametre tahminleri gerçek değerleriyle ilgisiz bulunur [Bates ve Watts, 1988]. Hataların varyanslarının değişkenlikleri ve hatalar arasında belirgin korelasyonlar mevcut ise hatalar uygun bir şekilde ağırlıklandırılarak yeni elde edilen hataların homojen olmaları sağlanır. Bu yaklaşım matris manipulasyonu temeline dayanmakla birlikte hatalar üzerinden dönüşüm gerçekleştirilmiş olunur. Bu şekilde homojen hatalara sahip hale getirilen regresyon modelinden parametre tahminleri olağan en küçük kareler yöntemi ile bulunur.

Tanım olarak $y_i = f(x_i, \theta_*) + \varepsilon_i, i = 1, 2, \dots, n$ modelinin hatalarının $E(\varepsilon) = 0$ ve $V(\varepsilon) = E(\varepsilon\varepsilon^t) = \sigma^2 V$ özellikleri bulunuyorsa, bu durumda V matrisi bilinmemekle birlikte hataların temsili olarak önerilen regresyon modelinin artıkları incelenerek V matrisinin tahmini yapılabilmektedir. Bu açıdan V matrisi tahmini regresyon modeline bağımlı olduğu gibi tahmin için kullanılan yöntemlere de bağımlıdır. Genelleştirilmiş en küçük kareler yönteminin bütünlüğünü bozmamak için V matrisinin bilindiği varsayılarak $V = K^2$ olacak şekilde simetrik $K^t = K$ matrisi bulunur. Sabit bir matris olarak alınan K matrisi aracılığı ile hatalar homojenlik şartlarını sağlar hale getirilir. Bu şekilde ι (yota) hataları $E(\iota) = 0$ ve $V(\iota) = \sigma^2 I$ ile homojenlik şartlarını sağlar.

$$K^{-1}\varepsilon = \iota \rightarrow \begin{cases} E(\iota) = K^{-1}E(\varepsilon) = 0 \\ V(\iota) = K^{-1}V(\varepsilon)K^{-1} = K^{-1}\sigma^2VK^{-1} = \sigma^2I \end{cases} \quad (2.12)$$

K^{-1} dönüşümü ile $y = f(\theta_*) + \varepsilon$ modeli şu şekilde getirilir:

$$\begin{aligned} y = f(\theta_*) + \varepsilon &\rightarrow K^{-1}y = K^{-1}f(\theta_*) + K^{-1}\varepsilon \Rightarrow z = g(\theta_*) + \iota \\ K^{-1}y = z, \quad K^{-1}f(\theta_*) = g(\theta_*) &, \quad K^{-1}\varepsilon = \iota \end{aligned} \quad (2.13)$$

Bu şekildeki bir dönüştürmeden sonra elde edilen modele olağan en küçük kareler yöntemi uygulanarak genelleştirilmiş en küçük kareler yöntemi tamamlanır. Daha sonra homojenlik şartlarını sağlayan hatalara sahip olan dönüştürülmüş modelin parametre tahminleri yapılır [Seber ve Wild, 1989].

$$R = \frac{\partial g(\theta)}{\partial \theta^t} = K^{-1} \frac{\partial f(\theta)}{\partial \theta^t} = K^{-1}Z \quad (2.14)$$

aracılığı ile $R_* = R(\theta_*)$, $\hat{R} = R(\hat{\theta})$ ve $R_k = R(\theta_k)$ kullanılarak ağırlıklandırılmış en küçük kareler yöntemi sonuçları benzer şekilde oluşturulur.

Ağırlıklandırılmış $z = g(\theta_*) + \iota$ modeli doğrusallaştırılarak olağan en küçük kareler yöntemi ile Gauss-Newton iterasyonu bulunur. $R_k = R(\theta_k)$ ile θ_k civarında $g(\theta) \cong g(\theta_k) + R_k(\theta - \theta_k)$ açılımı yardımıyla kareler toplamı

$$S(\theta) = \|z - g(\theta)\|^2 \cong \|z - g(\theta_k) - R_k(\theta - \theta_k)\|^2 \quad (2.15)$$

fonksiyonunu minimize eden çözüm

$$\theta_{k+1} = \theta_k + (R_k^t R_k)^{-1} R_k^t (z - g(\theta_k)) \quad (2.16)$$

şeklinde bulunur. Buradan $R_k = K^{-1}Z_k$ ile birlikte Gauss-Newton algoritması

$$\theta_{k+1} = \theta_k + \left(Z_k^t V^{-1} Z_k \right)^{-1} Z_k^t V^{-1} [y - f(\theta_k)] \quad (2.17)$$

$$\lim_{k \rightarrow \infty} \theta_k = \hat{\theta}_G$$

yardımıyla $\hat{\theta}_G$ genelleştirilmiş en küçük kareler tahminine ulaşılır.

Veri kaybının bulunduğu tıbbi alanlar için geliştirilmekle birlikte ağırlıklandırma yaklaşımlarının genelleştirildiği teorik bir çalışma olan Lipsitz ve dig. (2009) tarafından yazılan makaleye bakılabilir.

Bazı kaynaklarda farklı amaçlara yönelik her çeşit ağırlıklandırmayı içine alan “ağırlıklandırılmış en küçük kareler yöntemi” başlığı daha genel bir adlandırmayı ifade ederken, “genelleştirilmiş en küçük kareler yöntemi” başlığı ağırlıklandırma amacıyla sadece hata varyans-kovaryans matrisinin (tahmininin) kullanımına işaret etmektedir [Jureckova, 1977; Bunke ve Bunke, 1989].

Herhangi bir θ ile artık kareler toplamı fonksiyonu

$$S(\theta) = \|z - g(\theta)\|^2 = [z - g(\theta)]^t [z - g(\theta)] = [y - f(\theta)]^t V^{-1} [y - f(\theta)] \quad (2.18)$$

şeklinde ağırlıklandırılarak minimize edilmesiyle $\hat{Z}^t V^{-1} [y - f(\hat{\theta})] = 0$ normal denklemlerine varılır [Seber ve Wild, 1989]. $g(\theta)$ veya $f(\theta)$ beklenen fonksiyonu θ_* civarında $g(\theta) \cong g(\theta_*) + R_*(\theta - \theta_*)$ veya $f(\theta) \cong f(\theta_*) + Z_*(\theta - \theta_*)$ Taylor açılımı yerine konup $S(\theta)$ kareler toplamı fonksiyonunu minimize eden sonuç

$$\begin{aligned}\hat{\theta}_G - \theta_* &= (R_*^t R_*)^{-1} R_*^t \epsilon \\ &= (Z_*^t V^{-1} Z_*)^{-1} Z_*^t V^{-1} \epsilon\end{aligned}\quad (2.19)$$

bulunarak $\hat{\theta}_G$ genelleştirilmiş en küçük kareler tahmininin beklenen değeri ve varyansı sırası ile

$$E(\hat{\theta}_G) = \theta_* \quad , \quad V(\hat{\theta}_G) = \sigma^2 (R_*^t R_*)^{-1} = \sigma^2 (Z_*^t V^{-1} Z_*)^{-1} \quad (2.20)$$

şeklinde elde edilir. Eğer gözlem sayısı yeteri kadar büyükse $R_* = R(\theta_*)$ yerine $\hat{R} = R(\hat{\theta}_G)$ kullanılarak parametre tahminlerinin varyans-kovaryans matrisi tahmini

$$\hat{V}(\hat{\theta}_G) = s^2 (\hat{R}^t \hat{R})^{-1} = s^2 (\hat{Z}^t V^{-1} \hat{Z})^{-1}$$

$$\begin{aligned}s^2 &= S(\hat{\theta}_G) / (n - p) \\ &= \frac{1}{n - p} [z - g(\hat{\theta}_G)]^t [z - g(\hat{\theta}_G)] \\ &= \frac{1}{n - p} [y - f(\hat{\theta}_G)]^t V^{-1} [y - f(\hat{\theta}_G)]\end{aligned}\quad (2.21)$$

şeklinde bulunur (Burada ağırlık matrisi $V^{-1} = W$ şeklinde ifade edilir).

2.3. Parametrelerin Başlangıç Değerlerinin Seçimi

Eğrisellik dolayısıyla birden fazla yerel minimumun mevcut olabilmesine bağlı olarak parametrelerin θ_0 başlangıç değerlerinin seçimi iterasyonla bulunacak olan parametrelerin sonuç tahminlerini etkiler. Başlangıç değerlerine bağlı olarak iterasyonla farklı bir noktaya yakınsama gerçekleşebilir veya hiçbir yakınsama gerçekleşmeyebilir [Bates ve Watts, 1988]. Ayrıca hata varyansı büyük ise birden fazla yerel minimum nokta mevcut olabilir [Nash ve Walker-Smith, 1987]. Bununla

birlikte araştırmanın en zor sonuçlarından biri elde edilen tahminler fiziksel olarak mümkün olan sonuçlardan biri olabileceği gibi birden fazla minimum nokta bu sonuçları destekleyebilir [Agarwall ve O'Regan, 2003]. Bu nedenlerden dolayı istenen sonuca ulaşmak için başlangıç değerleri rasgele seçilmemelidir. Doğru parametre değerlerine yakın olan başarılı başlangıç değerleri yakınsama zorluklarını azaltır [Nash, 1979]. Gauss-Newton modifikasyonu olan Levenberg-Marquardt algoritması başlangıç değerlerinin seçimine daha az duyarlı olmakla birlikte θ_0 başlangıcının dikkatli seçimi iterasyonların sayısını düşürebildiğinden bir avantaj sağlar.

Doğrusal olmayan regresyon modellerinde parametrelerin çoğu zaman fiziksel anlamları bulunur. Bu durum θ_0 başlangıç değerlerinin seçiminde oldukça yardımcı olabilmektedir [Gallant, 1987]. Modelin davranışlarını anlamak açısından parametrelerin farklı farklı değerleri için beklenen fonksiyonun grafiğinin çizimi oldukça yardımcı olabilmektedir; parametre değerlerindeki değişimlere bağlı olarak model davranışlarındaki etkiler gözlenebilir [Hamilton, 1992]. Limitler veya özel bazı noktalar parametrelerin değerleri konusunda önemli bilgiler verebilir. Bunun yanında, verilerin genel eğilimleri göz önünde bulundurularak parametrelerin başlangıç değerleri fiziksel ve geometrik olarak bulunabilir [Bates ve Watts, 1988]. Bu yolun dışında, yardımcı denklemler çıkarılarak parametrelerin başlangıç değerleri bulunabilir [Gallant, 1987].

Bazı durumlarda beklenen fonksiyonun yapısı bazı dönüşümlere imkan verebilir. Bu sayede parametreler için θ_0 başlangıç değerleri elde edilebilir. Bu dönüşümlerden biri ile model parametrelerine göre doğrusallaştırılabiliyorsa doğrusal en küçük kareler ile dönüştürülmüş parametrelerin başlangıç değerleri bir uygunlukta bulunabilir. Bu başlangıç değerleri doğrusal olmayan θ parametrelerinin başlangıç değerlerinin elde edilesinde kullanılabilir [Bates ve Watts, 1988]. Ancak doğrusallaştırma işlemi hata yapısının değişmesine neden olur. Bununla birlikte parametrelerin başlangıç değerleri model parametrelerine göre doğrusallaştırılarak olağan veya ağırlıklandırılmış en küçük kareler yöntemi ile veya sağlam yaklaşımları

[Huber, 1972, 1977] ile bulunabilir. Eğer doğrusallaştırma mümkün değilse bir veya birkaç parametre etkisizleştirilerek model geri kalan parametrelerine göre doğrusallaştırılabilir ve olağan veya ağırlıklandırılmış en küçük kareler yöntemi ile bu parametrelerin başlangıç değerleri tahmin edilebilir [Huber, 1977; Gallant, 1987; Bates ve Watts, 1988; Bunke ve Bunke, 1989].

2.4. Kareler Toplamı Fonksiyonunun Minimizasyonu ve İteratif Algoritmanın Seçimi

Birbirinden bağımsız sabit varyanslı beklenen değerleri sıfır olan hatalar varsayımı altında en küçük kareler yaklaşımıyla bulunan parametre tahminleri artık kareleri toplamı fonksiyonunu minimize eder. Regresyon modelinin tek parametresi bulunduğunda kareler toplamının optimal noktasında ilk türev sıfır, minimum nokta olduğunun göstergesi için ikinci türev pozitif olmalıdır [Bronson ve Naadimuthu, 1997; Bates ve Watts, 1988]. Birden fazla parametre arandığında incelemeler çokdeğişkenli fonksiyonların ekstremumlarının incelenmesi problemine dönüşür. Çok boyutlu uzayda çok daha karmaşık yapılanmalar olabileceğinden ekstremum incelemeleri çok daha zorlaşır. Bu tür problemlerin çözümü için birçok algoritma geliştirilmiş olmakla birlikte optimum çözümler garanti edilememektedir [Nash, 1979; Nash ve Walker-Smith, 1987; Bronson ve Naadimuthu, 1997; Rardin, 1998]. Parametrelerin tahminleri için kareler toplamının minimum noktasında parametreler açısından birinci mertebeden kısmi türevler sıfır, ikinci mertebeden kısmi türevlerden oluşan Hessian matrisi pozitif definit (bütün özdeğerleri pozitif) olmalıdır.

İstatistiksel araştırmalarda minimizasyon problemi için temel olarak en hızlı azalış yöntemi ve Gauss-Newton yönteminin uyarlamaları veya Newton-Raphson metodunun yaklaşıklıkları kullanılır. İstatistiksel araştırmalar için geliştirilen algoritmaların bir çoğu bu yöntemlerden türetilmiştir. Parametrelerin güvenilir tahminlerine ulaşmak için artık kareler toplamı fonksiyonu dahilinde çok değişkenli fonksiyonların ekstremumlarını araştıran algoritmalar kullanılabilir. İstatistiksel amaçlar için uyarlanmış Levenberg-Marquardt yöntemi, Quasi-Newton yöntemi [Rardin, 1998], Simpleks yaklaşımı, Hooke-Jeeves yolu araştırması, Rosenbrock

yolu araştırması, Fletcher-Powell yöntemi, modife edilmiş yol araştırması algoritmaları kullanılabilir [Nash ve Walker-Smith, 1987; Bates ve Watts, 1988; Bronson ve Naadimuthu, 1997; Rardin ,1998].

İterasyonlar ilerledikçe model verilerle uyumlu hale gelir. Adım aralıkları küçüldükçe yakınsama için daha çok iterasyon gereklidir. Adım aralıkları büyüdükçe daha az iterasyon anlamına gelebilir, ancak parametrelerin optimal tahminleri bulunamayabilir veya bir yakınsama gerçekleşmeyebilir. Yakınsama gerçekleşmediğinde adım aralıkları küçültülmelidir. Uzayın eğriselliği göz önünde bulundurularak uyarlamalı/adaptif adım aralıklı iterasyonlar da geliştirilmiştir [Seber ve Wild, 1989]. Uygun olmayan başlangıç değerleri, modelde fazla sayıda parametrenin mevcut olması veya örnek çapının yeterince büyük olmaması sonucu yakınsama gerçekleşmeyebilir [Bates ve Watts, 1988].

3. DİĞER TAHMİN YÖNTEMLERİ

Hataların bağımsızlığı veya sabit varyans koşulları sağlanamadığında parametre tahminleri olabilirlik yaklaşımlarıyla bulunabilir. Eğer veri içinde aykırı değerler bulunuyorsa parametre tahminlerinin bozulmadan yapılmasını olanaklı hale getiren sağlam yaklaşımları da mevcuttur.

3.1. En Çok Olabilirlik Tahmini

Parametre tahmini olabilirlik fonksiyonunun maksimizasyonu şeklinde de bulunabilir. Bunun için hataların ortak olasılık yoğunluk fonksiyonunun biliniyor olması gereklidir. Eğer hatalar bağımsız ise olabilirlik fonksiyonu hataların olasılık yoğunluk fonksiyonlarının çarpımı şeklinde ifade edilir. Olabilirlik fonksiyonunun maksimizasyonu yukarıda ifade edilen iteratif algoritmalar aracılığı ile bulunabilir. Eğer hataların varyansları sabit ve normal dağılımlı ise ve büyük hacimli örnekler mevcutsa en küçük kareler tahmini ile en çok olabilirlik tahmini aynı değerlerde bulunur [Gallant, 1987; Seber ve Wild, 1989]. Büyük hacimli örneklerde en çok olabilirlik tahmin edicilerinin yanlılıkları azalmakta etkinlik ve tutarlılık özellikleri kesinleşmektedir [Günay, 1972].

ε_i hataları bağımsız ve $N(0, \sigma^2)$ dağılımından geliyorsa olabilirlik fonksiyonu

$$p(y | \theta, \sigma^2) = (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2} \sum_{i=1}^n \frac{[y_i - f(x_i, \theta)]^2}{\sigma^2}\right) \quad (3.1)$$

şeklinde ifade edilir. Olabilirlik fonksiyonunun logaritması alınarak

$$\begin{aligned} L(\theta, \sigma^2) &= -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln(\sigma^2) - \frac{1}{2} \sum_{i=1}^n \frac{[y_i - f(x_i, \theta)]^2}{\sigma^2} \\ &= -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} S(\theta) \end{aligned} \quad (3.2)$$

fonksiyonu elde edilir. Kolaylık için $\sigma^2 = v$ denilerek $L(\theta, v)$ fonksiyonu θ ve v parametrelerine göre maksimize edilerek parametrelerin en çok olabilirlik tahminleri elde edilir.

$$\frac{\partial L}{\partial v} = -\frac{n}{2v} + \frac{1}{2v^2} S(\theta) = 0 \rightarrow v = \sigma^2 = \frac{1}{n} S(\theta) \quad (3.3)$$

İkinci mertebeden türev negatif olduğundan elde edilen sonuç maksimumdur. $L(\theta, v)$ fonksiyonunun θ parametre vektörüne göre türevinin sıfıra eşitlenmesi ile $S(\theta)$ artık kareler toplamı fonksiyonu minimize edilir.

$$\frac{\partial L}{\partial \theta} = -\frac{1}{2\sigma^2} \frac{\partial S}{\partial \theta} = 0 \rightarrow \frac{\partial S}{\partial \theta} = 0 \quad (3.4)$$

Bu durumda $\hat{\theta}$ tahmini ve $\hat{\sigma}^2 = S(\hat{\theta})/n$ varyans tahmini en çok olabilirlik tahminleri olarak bulunur. Bu tahminlerle olabilirlik fonksiyonunun maksimum değeri

$$p(y | \hat{\theta}, \hat{\sigma}^2) = (2\pi\hat{\sigma}^2)^{-n/2} e^{-n/2} \quad (3.5)$$

şeklinde hesaplanır. θ parametre vektörü ve $\sigma^2 = v$ ile artırılmış parametre vektörü $\theta_+ = (\theta^t, v)^t$ yardımıyla bilgi matrisi ve beklenen bilgi matrisi

$$I = -\frac{\partial^2 L}{\partial \theta_+ \partial \theta_+^t} \quad (3.6)$$

$$E(I) = -E\left(\frac{\partial^2 L}{\partial \theta_+ \partial \theta_+^t}\right) = \begin{bmatrix} \frac{1}{2\sigma^2} E\left(\frac{\partial^2 S}{\partial \theta \partial \theta^t}\right) & -E\left(\frac{\partial^2 L}{\partial \theta \partial v}\right) \\ -E\left(\frac{\partial^2 L}{\partial v \partial \theta^t}\right) & -E\left(\frac{\partial^2 L}{\partial v^2}\right) \end{bmatrix} = \begin{bmatrix} \frac{1}{\sigma^2} Z(\theta_*)^t Z(\theta_*) & 0 \\ 0^t & \frac{n}{2\sigma^4} \end{bmatrix} \quad (3.7)$$

şeklinde verilir [Seber ve Wild, 1989]. Buna bağlı olarak parametre tahminlerinin varyans-kovaryans matrisi

$$V(\hat{\theta}_+) = V\left(\begin{matrix} \hat{\theta} \\ \hat{v} \end{matrix}\right) = [E(I)]^{-1} = \left[-E\left(\frac{\partial^2 L}{\partial \theta_+ \partial \theta_+^t}\right)\right]^{-1} \quad (3.8)$$

olarak bulunur. Burada olduğu gibi beklenen bilgi matrisinin tersi $\hat{\theta}_+$ en çok olabilirlik tahmininin asimptotik varyans-kovaryans matrisidir; asimptotik varyans-kovaryans matrisi aracılığı ile büyük hacimli n için varyans-kovaryans matrisinin tahmininin çıkarımında kullanılabilir. Bu durumda $\hat{\theta}_+ = (\hat{\theta}^t, \hat{\sigma}^2)^t$ en çok olabilirlik tahminleri ile

$$\frac{\partial S(\theta)}{\partial \theta} \Big|_{\hat{\theta}} = 0 \quad (3.9)$$

olacağından varyans-kovaryans matrisinin tahmini

$$\hat{V}(\hat{\theta}_+) = \left(-\frac{\partial^2 L}{\partial \theta_+ \partial \theta_+^t}\right)_{\hat{\theta}_+}^{-1} = \begin{bmatrix} \frac{1}{2\hat{\sigma}^2} \frac{\partial^2 S}{\partial \theta \partial \theta^t} \Big|_{\hat{\theta}} & 0 \\ 0^t & \frac{n}{2\hat{\sigma}^4} \end{bmatrix}^{-1} = \begin{bmatrix} 2\hat{\sigma}^2 \left(\frac{\partial^2 S}{\partial \theta \partial \theta^t}\right)_{\hat{\theta}}^{-1} & 0 \\ 0^t & \frac{2\hat{\sigma}^4}{n} \end{bmatrix} \quad (3.10)$$

şeklinde verilir.

3.2. Yarı (Quasi) Olabilirlik Tahmini

Bu parametre tahmini en çok olabilirlik tahmini ile bulunan tahmine benzemekle birlikte ilk iki momenti haricinde bağımlı değişkenin dağılımının bilinmesine gerek duyulmaz [Wedderburn, 1974; McCullagh, 1983; Seber ve Wild, 1989]. Yarı/quasi tanımının kullanılmasının nedeni en çok olabilirlik yaklaşımına benzemekle birlikte benzemenin tam olmamasından dolayıdır.

Konuya giriş açısından şu problem üzerinde durulabilir. Hatalar normal dağılıyorsa ve hataların varyansları farklı ise hatalar ağırlıklandırılarak parametrelerin en çok olabilirlik tahmini aşağıdaki gibi bulunur. ε vektörünün $E(\varepsilon) = 0, V(\varepsilon) = \sigma^2 V$ özellikleri bulunduğu olabilirlik fonksiyonu

$$p(y | \theta, \sigma^2 V) = (2\pi\sigma^2)^{-n/2} |V|^{-1/2} \exp\left(-\frac{1}{2\sigma^2} [y - f(\theta)]^t V^{-1} [y - f(\theta)]\right) \quad (3.11)$$

şeklinde verilir. Parametrelere ve hataların varyans-kovaryans matrisine göre olabilirlik fonksiyonu maksimize edilerek, bu bilinmeyenlerin en çok olabilirlik tahmin edicileri bulunur. Olabilirlik fonksiyonunun logaritması alındığında

$$\begin{aligned} L(\theta, \sigma^2 V) &= -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln(\sigma^2) - \frac{1}{2} \ln|V| - \frac{1}{2\sigma^2} [y - f(\theta)]^t V^{-1} [y - f(\theta)] \\ &= -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln(\sigma^2) - \frac{1}{2} \ln|V| - \frac{1}{2\sigma^2} S(\theta) \end{aligned} \quad (3.12)$$

fonksiyonu elde edilir. Doğrusal olmayan modelin θ parametre vektörü ile V matrisi olabilirlik fonksiyonunun bilinmeyenlerini oluşturur. Bir oran görünümünde olan σ^2 parametresi kolaylık için bir değerinde alınabilir: $\sigma^2 = 1$. Bu şekilde bulunması gereken V matrisinin tahmini üzerinde durulabilir. Fonksiyonun yapısı dikkate alındığında optimal değerleri veren θ parametre vektörünün ve V matrisinin birlikte bulunması zorlu bir çözümü gerektirir. Bağımlı değişkenin dağılımının bilinmediği durumlarda yukarıdaki problemin tam çözümü gerekli

değildir, problemin görünümü değiştirilerek bir çözüme ulaşılabilir. Böyle durumlar için “yarı olabilirlik tahmin” yaklaşımı geliştirilmiştir. Konuya şu şekilde yaklaşılabılır; yukarıdaki logaritma fonksiyonunun $f(\theta)$ vektörüne göre türevinin

$$\frac{\partial L}{\partial f(\theta)} = \frac{1}{\sigma^2} V^{-1} [y - f(\theta)] \quad (3.13)$$

olduğu gözönünde bulundurularak $l(\mu)$ gibi bir “log yarı-olabilirlik fonksiyonunun” olduğu varsayılır. Ayrıca $E(y) = f(\theta) = \mu$ ve $V(y) = \sigma^2 V(\mu)$ olmak üzere log yarı olabilirlik fonksiyonu

$$\frac{\partial l(\mu)}{\partial \mu} = [V(\mu)]^{-1} (y - \mu) \quad (3.14)$$

kısmi diferansiyel denklemler sistemi ile tanımlanır. Burada $l(\mu)$ log yarı olabilirlik fonksiyonunun μ nun bir fonksiyonu olduğu varsayılır. Bu tanımlar McCullagh (1983) tarafından verilmiştir. Kolaylık için $l[f(\theta)]$ ve $V[f(\theta)]$ gösterimleri sırası ile $l(\theta)$ ve $V(\theta)$ şeklinde ifade edilerek $\hat{\theta}$ en çok yarı-olabilirlik tahminleri

$$\begin{aligned} \frac{\partial l}{\partial \theta} &= 0 \\ \frac{\partial l}{\partial \theta} &= \left(\frac{\partial f}{\partial \theta^t} \right)^t \frac{\partial l}{\partial f} = 0 \\ &= [Z(\theta)]^t [V(\theta)]^{-1} [y - f(\theta)] = 0 \end{aligned} \quad (3.15)$$

denklemlerinin çözümü ile elde edilir. $\hat{\theta}$ tahminine yakın θ_k civarında

$$f(\theta) = f(\theta_k) + Z(\theta_k)(\theta - \theta_k) \quad (3.16)$$

Taylor açılımı yardımıyla normal denklemler Gauss-Newton yaklaşımıyla iteratif olarak çözülür. $Z(\theta)$ ve $V(\theta)$, θ_k noktasına doğru $Z_k = Z(\theta_k)$ ve $V_k = V(\theta_k)$ olarak yaklaştırılarak denklem

$$Z_k^t V_k^{-1} [y - f(\theta_k) - Z_k(\theta - \theta_k)] = 0 \quad (3.17)$$

şekline getirilir. Bu normal denklemlerin çözümüyle

$$\theta_{k+1} - \theta_k = \delta_k = \left(Z_k^t V_k^{-1} Z_k \right)^{-1} Z_k^t V_k^{-1} [y - f(\theta_k)] \quad (3.18)$$

iteratif işlemler için kullanılan adım elde edilir. Burada k. adımda θ_k noktasındaki değerler kullanılarak bir sonraki aşamadaki $\hat{\theta}$ vektörünün yaklaşıklığı olan $\theta_{k+1} = \theta_k + \delta_k$ vektörü bulunur.

$$\begin{aligned} \theta_{k+1} &= \theta_k + \left(Z_k^t V_k^{-1} Z_k \right)^{-1} Z_k^t V_k^{-1} [y - f(\theta_k)] \\ \lim_{k \rightarrow \infty} \theta_k &= \hat{\theta} \end{aligned} \quad (3.19)$$

İteratif işlemlerde ağırlığı ifade eden V matrisi de işlemlerden etkilenir, bu durum yarı-olabilirlik yaklaşımının bir özelliğidir. Bir yaklaşım şekline simülatif uygulamalarda değinilmiştir. Yarı olabilirlik yöntemi, en çok olabilirlik yöntemi üzerinde yapılan bir uyarlamadır. Bununla birlikte ağırlıklandırılmış kareler toplamı fonksiyonu

$$[y - f(\theta_k) - Z_k(\theta - \theta_k)]^t V_k^{-1} [y - f(\theta_k) - Z_k(\theta - \theta_k)] \quad (3.20)$$

θ vektörüne göre minimize edilerek θ_{k+1} elde edilir [Seber ve Wild, 1989].

Bilgi matrisi ve beklenen bilgi matrisi

$$I(\theta) = -\frac{\partial^2 l}{\partial \theta \partial \theta^t} \quad , \quad I(\theta_*) = -E \left(\frac{\partial^2 l}{\partial \theta \partial \theta^t} \right) = Z(\theta_*)^t V^{-1} Z(\theta_*) \quad (3.21)$$

aracılığı ile parametre tahminlerinin varyans-kovaryans matrisi

$$V(\hat{\theta}) = I(\theta_*)^{-1} = \left[Z(\theta_*)^t V^{-1} Z(\theta_*) \right]^{-1} \quad (3.22)$$

ve parametre tahminlerinin varyans-kovaryans matrisi tahmini

$$\hat{V}(\hat{\theta}) = \hat{\sigma}^2 \left[Z(\hat{\theta})^t \hat{V}^{-1} Z(\hat{\theta}) \right]^{-1} \quad (3.23)$$

$$\lim_{k \rightarrow \infty} \theta_k = \hat{\theta} \quad , \quad \lim_{k \rightarrow \infty} V_k = \hat{V}$$

olarak elde edilir. Son olarak beklenen değeri bir olan $\hat{\sigma}^2$ tahmini

$$\hat{\sigma}^2 = \frac{\left[y - f(\hat{\theta}) \right]^t \hat{V}^{-1} \left[y - f(\hat{\theta}) \right]}{n - p} \quad (3.24)$$

yardımıyla bulunur.

3.3. L1-Norm Minimizasyonu

Artıkların mutlak değerleri toplamı fonksiyonunu minimize eden değerler parametrelerin bu yaklaşımdaki tahminleridir. Hataların çift üstel dağılımdan gelmesi durumunda en çok olabilirlik tahminleri ile benzer sonuçları bulunur. Bunun dışında hataların dağılım fonksiyonları simetrik ve hataların varyansları aynı ise en küçük kareler tahmini ile bulunan sonuçlarla benzer sonuçlar elde edilir [Oberhofer, 1982].

Kareler toplamı $S(\theta) = \sum_{i=1}^n [y_i - f(x_i, \theta)]^2$ fonksiyonu Öklit uzayı tabanlıdır ve L2-norm olarak adlandırılır. Bir alternatif yaklaşım olarak L1-normu olan $\sum_{i=1}^n |y_i - f(x_i, \theta)|$ fonksiyonu θ vektörüne göre minimize edilmelidir. Çözümün zayıf tutarlılığı için yeterli koşullar Oberhofer (1982) tarafından verilmiştir. Hataların

$$f(\varepsilon_i) = \frac{1}{2\sigma} e^{-\frac{|\varepsilon_i|}{\sigma}}, \quad -\infty < \varepsilon_i < \infty \quad (3.25)$$

ile çift üstel dağılımından gelmesi durumunda L1-norm yaklaşımı en çok olabilirlik tahminini verir. Normal dağılım ile karşılaştırıldığında çift üstel dağılımının daha kalın kuyrukları bulunur. Burada olduğu gibi en küçük kareler ile karşılaştırıldığında L1-normu büyük saçılmalara daha az bir ağırlık verir. Bu yaklaşım mutlak değerce büyük artıkların (aykırı değer) etkilerine daha dirençli olduğundan çoğu zaman sağlam (robust) regresyon tekniği gibi kullanılır.

3.4. Sağlam (Robust) Tahmin Ediciler

En küçük kareler yaklaşımı kullanıldığında bazı hataların aşırı saçılması durumunda parametre tahmini yansız olmakla birlikte tahmin en küçük varyanslı değildir. Büyük mutlak değerli hatalara düşük ağırlıklar verdiren özel ağırlık fonksiyonları aracılığı ile ağırlıklandırma ile en küçük kareler yaklaşımıyla parametre tahminleri bu yöntemle bulunur [Huber, 1977; Seber ve Wild, 1989].

$y_i = f(x_i, \theta_*) + \varepsilon_i, i = 1, 2, \dots, n$ modeline uygulanan $\hat{\theta}$ en küçük kareler tahminine göre normal denklemler

$$\sum_{i=1}^n \frac{\partial f(x_i, \theta)}{\partial \theta_j} \Big|_{\hat{\theta}} [y_i - f(x_i, \hat{\theta})] = 0, \quad j = 1, 2, \dots, p \quad (3.26)$$

şeklinde verilir. Doğrusal regresyona uygulanan sağlam yöntemi buraya uyarlanarak $\tilde{\theta}$ sağlam tahmini

$$\sum_{i=1}^n \frac{\partial f(x_i, \theta)}{\partial \theta_j} \Big|_{\tilde{\theta}} \psi \left(\frac{y_i - f(x_i, \tilde{\theta})}{\tilde{\sigma}} \right) = 0, \quad j = 1, 2, \dots, p \quad (3.27)$$

denklemlerinden çözülerek bulunur [Huber, 1981]. Burada ψ fonksiyonu aşırı değerleri azaltan yada ihmal eden uygun bir ağırlık fonksiyonudur. Ayrıca $\tilde{\sigma}$ yığının standart sapmasının sağlam bir tahminidir. Birçok yaklaşımdan en çok olabilirlik fonksiyonlarının karşılaştırıldığı ve yaklaşıklıklarının incelendiği, bunların üzerine sağlam en çok olabilirlik fonksiyonlarının geliştirildiği Tsou ve Royal (1995) tarafından yapılan teorik çalışmaya başvurulabilir.

3.4.1. Sağlam (Lp) regresyon

Gözlemlerin homojen olmayan bir dağılımdan gelmesi durumunda özellikle normal dağılımla karşılaştırıldığında dağılımın kalın kuyruklu olması dolayısı ile mutlak değerce büyük hatalar (aykırı değerler) üretilir ve en küçük kareler tahminleri üzerinde büyük etkileri bulunur. Bu büyük hatalar en küçük kareler eğrisini kendi üzerlerine çektiğinden ve bunlara karşı gelen artıkları yapay olarak küçülttüğünden bu büyük hataların belirlenmesini de zorlaştırır. En küçük karelerin yüksek düzeyde etkilendiği gözlem etkilerini azaltan sağlam regresyon yaklaşımı konusundaki temel kavramlar Andrews ve dig. (1972), Hill ve Holland (1977), Hogg (1974) ve Hogg'da (1979) bulunabilir.

Çift üstel dağılımından gelen bağımsız ε_i rassal değişkenlerinin olduğu aşağıdaki gibi bir doğrusal regresyon modeli göz önünde bulundursun.

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad i = 1, 2, \dots, n \quad (3.28)$$

Parametrelerin tahmini için en çok olabilirlik fonksiyonunun

$$L(\beta_0, \beta_1) = \prod_{i=1}^n \frac{1}{2\sigma} e^{-\frac{|e_i|}{\sigma}} = \frac{1}{(2\sigma)^n} \exp\left(-\sum_{i=1}^n \frac{|e_i|}{\sigma}\right) \quad (3.29)$$

maksimize edilmesi mutlak (hataların temsili) e_i artıkların toplamının $\sum_{i=1}^n |e_i|$ minimize edilmesi anlamına gelir. Mutlak artıkların toplamının minimizasyonuna çoğu zaman L1-norm regresyon problemi denir. Çift üstel dağılımdan gelen hatalar olduğunda en çok olabilirlik fonksiyonu ile L1-norm regresyon problemi doğal olarak belirlemektedir. Eğer regresyon modeli parametrelere göre doğrusal ise L1-norm regresyon problemi bir doğrusal programlama modeli olarak formüle edilir. Sırasıyla u_i ve v_i ($i=1,2,\dots,n$) uydurulmuş doğrusal yüzeyden olan pozitif ve negatif sapmaları gösterirse parametre tahminleri mutlak artıklar toplamını minimize eder.

$$\begin{aligned} \min Z &= \sum_{i=1}^n (u_i + v_i) \\ \beta_0 + \beta_1 x_i + u_i - v_i &= y_i \quad , \quad i = 1, 2, \dots, n \\ \beta_0 &= \beta'_0 - \beta''_0 \\ \beta_1 &= \beta'_1 - \beta''_1 \\ u_i, v_i, \beta'_0, \beta''_0, \beta'_1, \beta''_1 &\geq 0 \end{aligned} \quad (3.30)$$

β_0 ve β_1 parametrelerinin işaretçe bir kısıtı bulunmaz. Ancak standart doğrusal programlama algoritmalarıyla elde edilen çözümler yansız parametre tahminleri değildir. Silken ve Hartley (1973) yansız çözümü bulan bir doğrusal programlama algoritması geliştirmişlerdir, ayrıca yansız çözüm veren bir yaklaşıma Tarakçı (1992) tarafından değinilmiştir. Doğrusal programlama yaklaşımındaki diğer referanslar Barrodale (1968), Barrodale ve Roberts (1973) ve Gentle ve dig. (1977) şeklinde verilebilir.

L1-norm ve L2-norm regresyon modelleri, model parametrelerini $\sum_{i=1}^n |e_i|^p, 1 \leq p \leq 2$ ifadesine göre minimize ederek bulduran Lp-norm regresyonunun özel durumlarıdır. Genel $1 < p < 2$ durumu için bu bir doğrusal olmayan programlama problemidir. Eğer hatalar kalın kuyruklu (normal olmayan) bir dağılımdan geliyorsa p nin 1.5 civarında seçilmesi ile makul sonuçların alındığı ifade edilmiştir [Hill ve Holland, 1977; Hogg, 1979].

3.4.2. M-tahmin ediciler

Genel olarak artıkların ρ fonksiyonunu minimize eden sağlam tahmin edicileri sınıflandırılabilir. Örneğin doğrusal bir regresyon modeline yönelik

$$\min_{\beta} \sum_{i=1}^n \rho(e_i) = \min_{\beta} \sum_{i=1}^n \rho(y_i - x_i^t \hat{\beta}) \quad (3.31)$$

incelemesinde X matrisinin i. satırı x_i^t vektörü ve β parametre vektörüdür. Bu tür bir tahmin edici en çok olabilirlik (maximum likelihood) dolayısıyla M-tahmin edicisi olarak adlandırılır. Bu şekilde hata dağılımının seçimine bağlı olarak ρ fonksiyonu en çok olabilirlik fonksiyonu ile ilgilidir. Hata dağılımının normal olmasına bağlı olarak en küçük kareler yöntemi kullanılmış ise ρ fonksiyonu $\rho(z) = \frac{1}{2} z^2$ şeklindedir. Eğer seçilen $\rho(z)$ fonksiyonu $\frac{1}{2} z^2$ yapısından farklı ise veya regresyon modeli doğrusal değilse artıklar sabit bir sayı ile çarpıldığında Eş. 3.31 ifadesinin çözümü farklı bulunur. Bundan dolayı M-tahmin edicisi ölçek açısından invariant (değişmez) değildir. Ölçek açısından invariant olmasının sağlanması için

$$\min_{\beta} \sum_{i=1}^n \rho\left(\frac{e_i}{s}\right) = \min_{\beta} \sum_{i=1}^n \rho\left(\frac{y_i - x_i^t \hat{\beta}}{s}\right) \quad (3.32)$$

ifadesi çözümlmelidir. Burada s ölçeğin sağlam tahmininin seçimi çoğunlukla

$$s = \text{medyan}|e_i - \text{medyan}(e_i)|/0.6745 \quad (3.33)$$

şeklinde alınır. Eğer n yeteri kadar büyük ise ve hata dağılımı normal ise 0.6745 sabiti s ölçeğini yaklaşık olarak σ standart sapmasının yansız tahmin edicisi haline getirir. Yukarıda sağlam tahmin $\tilde{\sigma}$ şeklinde gösterilmekle birlikte burada kolaylık için sağlam tahmin s olarak, parametrelerin sağlam tahmini ise $\hat{\beta}$ şeklinde gösterilmiştir. ρ fonksiyonunun β_j parametresine göre birinci mertebeden kısmi türevlerinin sıfıra eşitlenmesiyle Eş. 3.32 ifadesi minimize edilmiş olunur. Bu şekilde p denklemden oluşan

$$\sum_{i=1}^n x_{ij} \psi \left(\frac{y_i - x_i^t \hat{\beta}}{s} \right) = 0 \quad , \quad j = 1, 2, \dots, p \quad (3.34)$$

sisteminde $\psi = \rho'$ ile ρ fonksiyonunun türevidir. Burada doğrusal olmayan ψ fonksiyonu iteratif yöntemlerle çözülmelidir. Bunun için çoğu zaman iteratif tekrar ağırlıklandırılmalı en küçük kareler yöntemi kullanılır [Huber, 1981]. İteratif tekrar ağırlıklandırılmalı en küçük kareler yöntemi için ilk $\hat{\beta}_0$ tahmininin ve s ölçek tahmininin mevcut olduğu varsayılır. Eş. 3.34 ifadesindeki p denklem

$$\sum_{i=1}^n x_{ij} \left[\frac{\psi \left(\frac{y_i - x_i^t \hat{\beta}}{s} \right)}{\frac{y_i - x_i^t \hat{\beta}}{s}} \right] (y_i - x_i^t \hat{\beta}) = 0 \quad , \quad j = 1, 2, \dots, p \quad (3.35)$$

şeklinde yazılırsa

$$\sum_{i=1}^n x_{ij} w_{i0} (y_i - x_i^t \hat{\beta}) = 0 \quad , \quad j = 1, 2, \dots, p \quad (3.36)$$

yardımla ağırlıklar

$$w_{i0} = \frac{\psi\left(\frac{y_i - x_i^t \hat{\beta}_0}{s}\right)}{\frac{y_i - x_i^t \hat{\beta}_0}{s}} \quad (3.37)$$

olarak belirlenir. Matris notasyonu ile Eş. 3.36 ifadesi

$$X^t W_0 X \hat{\beta} = X^t W_0 y \quad (3.38)$$

şeklinde yazılır. Burada elemanları Eş. 3.37 ifadesinden $w_{10}, w_{20}, \dots, w_{n0}$ şeklinde bulunan n boyutlu köşegen W_0 ağırlık matrisidir. Aynı zamanda Eş. 3.38 ifadesi ağırlıklandırılmış en küçük kareler normal denklemleridir. Sırasıyla bir sonraki adımdaki tahmin edici

$$\hat{\beta}_1 = (X^t W_0 X)^{-1} X^t W_0 y \quad (3.39)$$

yolu ile bulunur. Sonraki adımda Eş. 3.37 ifadesinde $\hat{\beta}_0$ yerine $\hat{\beta}_1$ kullanılarak ağırlıklar tekrar hesaplanır. Bu şekilde iteratif tekrar ağırlıklandırılmalı en küçük kareler yöntemi, ağırlıklandırılmış en küçük kareler yönteminin arka arkaya uygulamaları ile gerçekleştirilir.

Sağlam tahminde $\hat{\beta}_0$ başlangıç değerlerinin seçimi için en küçük kareler yönteminin kullanılması aşırı sapmalar dolayısıyla yanıltıcı olabilir. Bunun yerine L1-norm tahminleri başlangıç değerlerinin seçimi için uygun bulunmaktadır.

Parametrelerin güven aralıklarının oluşturulabilmesi ve modele yönelik araştırmaların yapılabilmesi için $\hat{\beta}$ vektörünün varyans-kovaryans matrisinin

belirlenmesi gerekmektedir. Ağırlıklandırılmış en küçük kareler yöntemine göre varyans-kovaryans matrisi tahmini

$$\frac{\sum_{i=1}^n w_i (y_i - x_i^t \hat{\beta})^2}{n-p} (X^t W X)^{-1} \quad (3.40)$$

şeklinde bulunur. Burada $\hat{\beta}$ tahmini ve W matrisi en son iterasyonda bulunan çözümlerdir. Aykırı değerler mevcutsa $\hat{\beta}$ parametre tahmininin varyans-kovaryans matrisi konusunda birçok öneri bulunmakla birlikte hangi yaklaşımın daha iyi olduğu yönünde genel bir kanı bulunmamaktadır. Varyans-kovaryans matrisi konusunda diğer öneriler için Hill (1979) ve Huber'e (1981) bakılabilir.

Sağlam regresyon aykırı değerlerin belirlenmesi açısından oldukça yardımcıdır. Sağlam yaklaşımında ağırlıklandırılan gözlemler dikkatli bir şekilde incelenmeli ve ağırlığı oldukça azaltılan gözlemler araştırılmalıdır [Huber, 1981].

Çizelge 3.1. $z_i = e_i/s$ olmak üzere en çok kullanılan $\rho(z)$ sağlam fonksiyonları

Kriter	$\rho(z)$	$\psi(z)$	$w(z)$	Aralık
En küçük kareler	$\frac{1}{2} z^2$	z	1	$ z < \infty$
Huber t fonksiyonu $t = 2$	$\frac{1}{2} z^2$ $ z t - \frac{1}{2} t^2$	z $t \cdot \text{sign}(z)$	1 $t/ z $	$ z \leq t$ $ z > t$
Ramsay E_a fonksiyonu $a = 0.3$	$\frac{1 - (1 + a z) \exp(-a z)}{a^2}$	$z \exp(-a z)$	$\exp(-a z)$	$ z < \infty$
Andrew dalga fonksiyonu $a = 1.339$	$a \left[1 - \cos\left(\frac{z}{a}\right) \right]$ $2a$	$\sin\left(\frac{z}{a}\right)$ 0	$\frac{\sin(z/a)}{z/a}$ 0	$ z \leq a\pi$ $ z > a\pi$
Hampel fonksiyonu $a = 1.7$ $b = 3.4$ $c = 8.5$	$\frac{1}{2} z^2$ $a z - \frac{1}{2} a^2$ $\frac{a(c z - \frac{1}{2} z^2)}{c-b} - \frac{7}{6} a^2$ $a(b+c-a)$	z $a \cdot \text{sign}(z)$ $\frac{a \cdot \text{sign}(z)(c - z)}{c-b}$ 0	1 $a/ z $ $\frac{a(c - z)}{(c-b) z }$ 0	$ z \leq a$ $a < z \leq b$ $b < z \leq c$ $ z > c$

Sağlam fonksiyonları ψ fonksiyonlarının davranışlarına göre betimlenebilir. Artıklara verilen ağırlıkları belirleyen ψ fonksiyonlarına etki fonksiyonları da denilmektedir. En küçük kareler yaklaşımında ψ fonksiyonu sınırlanmadığından kalın kuyruklu dağılımlardan doğrudan etkilenir. Huber t fonksiyonu sınırlandırılmış olduğundan büyük artıklardan en küçük kareler kadar etkilenmez. Ramsay'in, Andrew'in ve Hampel'in ψ fonksiyonları artıklar büyüdükçe daha küçük ağırlıklar verir; bu açıdan bu fonksiyonlar oldukça dirençlidir. Ramsay ψ fonksiyonu uyumlu indirgeyicidir. Andrew dalga fonksiyonu ve Hampel fonksiyonu sert düşürücüdür, $|z|$ yeteri kadar büyükse ψ fonksiyonları sıfır değerini vermektedir.

Sağlam tahmin, regresyonda mutlak değerce büyük artıkların etkilerinin azaltılması çabaları üzerine geliştirilmiştir. Bundan dolayı doğrusal regresyon modellerine uygulanan sağlam yaklaşımları doğrusal olmayan regresyon modellerine de uygulanmaktadır [Bunke ve Bunke, 1989]. Sağlam yaklaşımının tıp alanındaki uygulamalarının bulunduğu Cook ve diğerlerinin (2009) makalesine bakılabilir.

4. PARAMETRE TAHMİNİ İÇİN TVRDİK YÖNTEMİ

Tvrdik ve dig. (2006) doğrusal olmayan regresyon modeli parametrelerinin tahmini amacıyla sezgisel bir yöntem önermişlerdir. Bu yaklaşımda parametrelerin uygun olmayan başlangıç değerlerinde başlayan iterasyonla makul bir süre içerisinde güvenilir parametre tahminlerine ulaşıldığı ifade edilmiştir. Gauss-Newton tabanlı Levenberg-Marquardt algoritmalarına dayanan standart istatistik paket programları parametrelerin güvenilir tahminlerinin araştırmasında bazen yada çoğu zaman başarısız olmaktadır. Tvrdik ve dig. (2006) birçok istatistik paket programlarını çeşitli zorluk seviyelerindeki modellere denemişler ve denemelerin yaklaşık yarısında tamamı ile başarısız olunulmuş yada parametrelerin tahminleri olması gereken değerlerden belirgin şekilde farklı bulunulduğu ifade edilmiştir. Bu yaklaşım ile sürekliliğin olduğu problemlerde kutu kısıtları aracılığı ile global optimizasyonunun koşullarının oluşması sağlanmaktadır. Levenberg-Marquardt veya Gauss-Newton deterministik algoritmaları ile karşılaştırıldığında önerilen stokastik araştırma yaklaşımlarının tahmini yapılacak parametrelerin başlangıç değerlerine gereksinim duymadığı ifade edilmektedir. Buradaki global optimizasyonu incelemesi 1970'li yıllardan beri üzerinde düşünülen çözümü zor olan bir problem olmakla birlikte problemin çözümü için temeli simülasyona dayanan ve bununla birlikte sezgisel yaklaşımlar da dahil birçok öneri ileri sürülmüştür [Spall, 2003; Back, 1996; Deb, 2005; Winter ve dig., 2005; Wolpert ve Macready, 1997; Price, 1977; Nelder ve Mead, 1964; Ali ve Torn, 2004; Storn ve Price, 1997; Solis ve Wets, 1981].

Bu yöntem için Nelder-Mead (1964) yönteminin anlaşılabilirliği önemli olduğundan ilk önce bu konuya değinilmelidir. İlk olarak iki boyutlu uzaydaki yaklaşım anlatılmaya çalışılmış daha sonra bu yaklaşım yüksek boyutlu uzaylara genelleştirilmiştir [Brent, 1973; Nash ve Walker-Smith, 1987].

4.1. Nelder-Mead Yöntemi

Çokdeğişkenli bir fonksiyonun bir yerel minimumunu bulduran simpleks algoritması Nelder ve Mead tarafından geliştirilmiştir. Tepe noktalarıyla ortaya çıkan basit

geometrik şekil simpleks olarak adlandırılır. Bundan dolayı bir fonksiyonun minimizasyonunu araştıran bu metoda “Nelder-Mead Simpleks Algoritması” yada kısaca “simpleks algoritması” da denir. Ancak bu adlandırma doğrusal programlama problemlerinin çözümünde kullanılan Dantzig ve diğerlerinin metoduyla karıştırılabildiğinden bu yönüme “Nelder-Mead Politop Algoritması” da denmektedir. Kelime olarak “politop” terimi tepe noktaları aracılığı ile oluşan geometrik şekli vurgular [Nash ve Walker-Smith, 1987]. Simpleks algoritmasında iki değişken için simpleks bir üçgen olmak üzere üçgenin üç köşe noktasındaki fonksiyonun değerlerinin karşılaştırılması temeline dayanır. İki değişkenli $f(x, y)$ fonksiyonunun en büyük değerli olduğu en kötü köşe noktası red edilerek yeni bir köşe noktası ile değiştirilir. Bu şekilde elde edilen üçgenlerin büyüklükleri azalırken minimum noktanın koordinatlarına yaklaşılr. Genel olarak N boyutlu uzayda genelleştirilmiş politop N+1 köşe noktası ile ifade edilir. Bu durumda algoritma ile N değişkenli fonksiyonun minimumu bulunur. Hesaplamaları yoğun ancak temel mantığı sade ve gradyent temelli algoritmalarla karşılaştırıldığında özellikle boyutu oldukça yüksek fonksiyonların çözümünde daha etkin olduğu ifade edilmektedir.

İki boyut için iterasyon şu şekilde ilerletilir.

-LGH Başlangıç Üçgeni

Minimize edilecek $f(x, y)$ fonksiyonu olmak üzere, başlangıçta belirlenen üçgenin üç köşe/tepe (vertex) noktası $V_k = (x_k, y_k)$, $k = 1, 2, 3$ olsun. Buna bağlı olarak $f(x, y)$ fonksiyonunun üç noktasındaki $z_k = f(x_k, y_k)$, $k = 1, 2, 3$ değerleri teker teker bulunur. Sonra her bir iterasyon için alt indisler $z_1 < z_2 < z_3$ şeklinde düzenlenir. İki boyutlu düzlemde köşe noktalarının en iyi veya en düşük (lowest) $L = (x_1, y_1)$, iyi (good) $G = (x_2, y_2)$ ve en kötü veya en yüksek (highest) $H = (x_3, y_3)$ şeklindeki adlandırması uygundur. Yüksek boyutlar söz konusu olduğunda G en iyiden sonraki ikinci iyi noktadır.

-Uygun Kenarın Orta Noktası

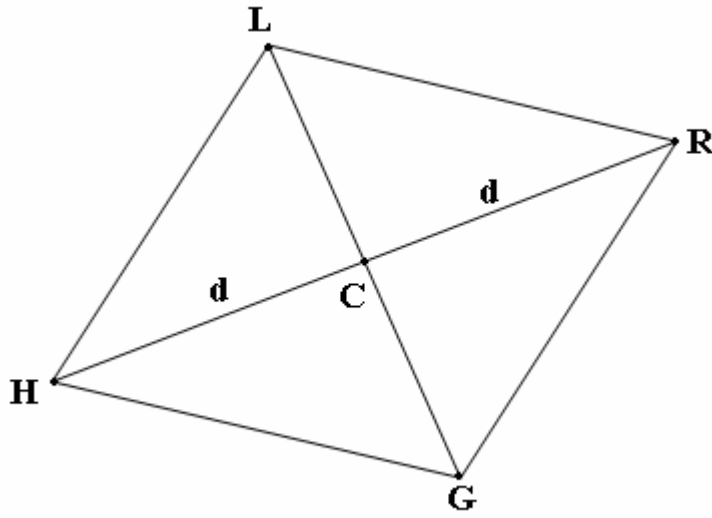
L ve G noktalarını birleştiren doğru parçasının orta noktası belirlenir. Bu köşe noktalarının koordinatlarının ortalaması alınarak C orta noktası bulunur.

$$C = \frac{L + G}{2} = \left(\frac{x_1 + x_2}{2}, \frac{y_1 + y_2}{2} \right) \quad (4.1)$$

-Yansımayla R Noktasının Bulunması

HL doğru parçası üzerinde H den L ye hareket ettikçe fonksiyonun değeri azalır. Benzer şekilde HG kenarında H den G ye hareket ettikçe fonksiyonun değeri düşürülür. Bu durumda H den başlayacak şekilde LG kenarı orta noktasına doğru $f(x, y)$ fonksiyonunun değeri düşürülür; daha da öteye gidilerek HC doğrultusunda fonksiyonun değerinin daha da düşürülmesi mümkün olabilir. Bu şekilde yansımayla oluşturulan R test noktası elde edilir. İlk önce LG kenarının C orta noktası bulunur. Sonra H den C noktasına doğru çizilir ve bu doğru parçasının uzunluğu “d” olarak alınır. Bu doğru HC doğrultusunda genişletilerek C noktasından d uzaklığına kadar gidilerek R noktasına varılır. R noktasının koordinatları şu şekilde bulunur:

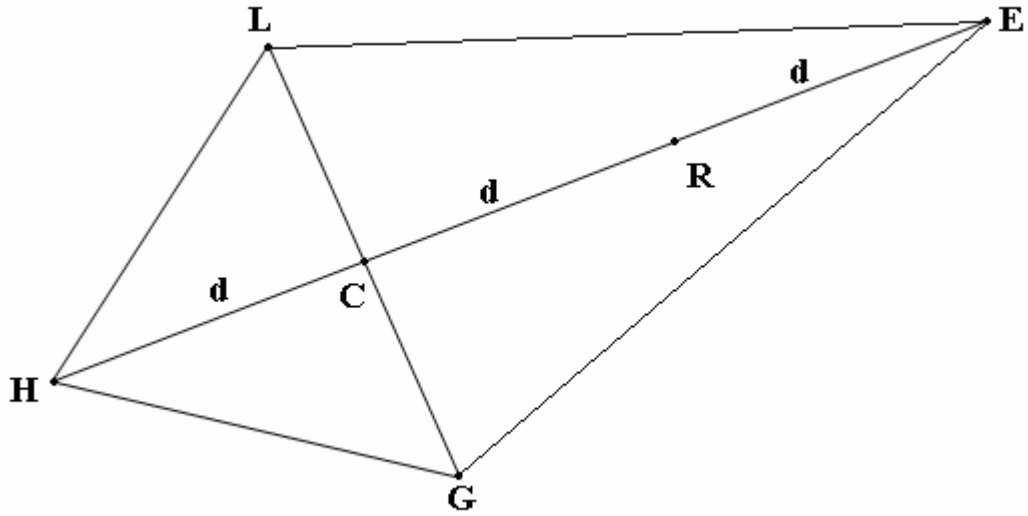
$$R = C + (C - H) = 2C - H \quad (4.2)$$



Şekil 4.1. Nelder-Mead yönteminde LGH üçgeni, C orta noktası ve R yansıma noktası

-Genişleme ile E Noktası

Eğer R noktasındaki fonksiyonun değeri H dan daha küçük ise doğrultu devam ettirilerek fonksiyon değeri daha da azaltılabilir. Bu durumda minimum nokta R noktasından biraz daha uzakta bulunuyor olabilir. Benzer şekilde doğru parçası CR doğrultusunda E noktasına kadar genişletilir. Bu yaklaşımla genişletilmiş LGE üçgeni oluşturulmuş olunur. CR doğrultusunda bir d mesafesi kadar daha gidilerek E noktasına varılır.



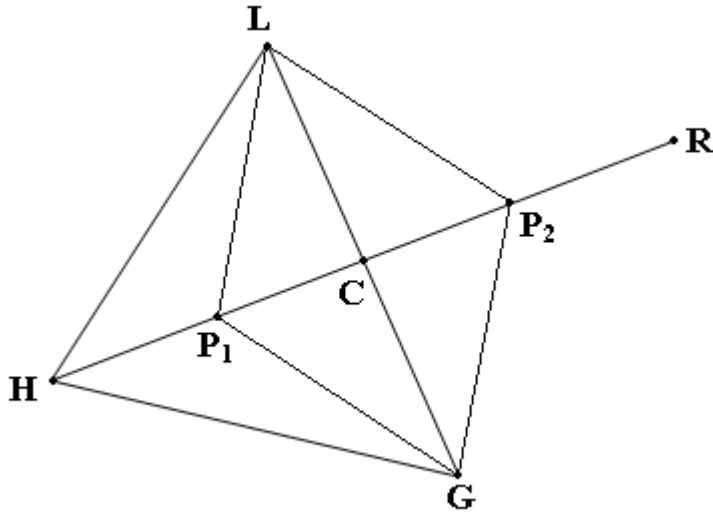
Şekil 4.2. LGH üçgeni, R noktası ve genişletmeyle elde edilen E noktası

Eğer E noktasındaki fonksiyon değeri R den daha küçük ise R den daha iyi bir köşe noktası bulunmuş olunur. E noktasının koordinatları şu şekilde bulunur:

$$E = R + (R - C) = 2R - C \quad (4.3)$$

-Büzülme ile P Noktası

Eğer R ve H noktalarındaki fonksiyon değerleri hemen hemen aynı ise (veya $f(R) > f(G)$ yada $f(R) > f(H)$ ise) şu şekilde davranılmalıdır. C noktasında fonksiyon daha küçük olabilir. Ancak bir üçgenin mevcut olabilmesi için H noktası C noktasının yerine geçemez, nedeni ise LGC bir üçgeni oluşturamaz olmasıdır. HC doğru parçasının orta noktası P_1 ve CR doğru parçasının orta noktası P_2 olsun.

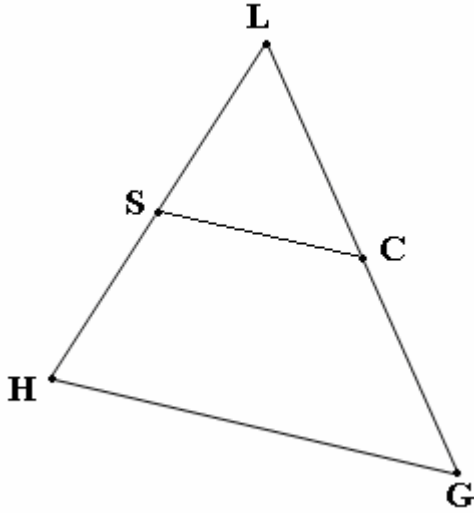


Şekil 4.3. Nelder-Mead yönteminde P_1 ve P_2 karşılaştırılarak bulunan P büzülme noktası

P_1 ve P_2 karşılaştırılarak en küçük fonksiyon değerli nokta P büzülme noktası olarak alınır. Bu durumda yeni üçgen LGP üçgenidir. İki boyutlu durumda P_1 ve P_2 arasındaki seçim önemli görünmeyebilir, fakat yüksek boyutlarda seçim önemlidir.

-L Noktasına Doğru Küçülme

Eğer P noktasındaki fonksiyon değeri H den daha küçük değilse G ve H noktaları L noktasına doğru çekilir. Bu durumda G noktası C ile değiştirilir ve H noktası da S ile değiştirilir. S noktası HL kenarının orta noktasıdır.



Şekil 4.4. L noktasına doğru çekilen üçgen

-Her Adımda Karar

Her adımda H noktasının yerine geçecek olan yeni bir köşe noktası bulunur. Bunun için gerekiyorsa fonksiyon değerleri hesaplanır. H noktasının yerine geçecek yeni köşe noktası bulunmuş ise daha ileri düzeyde sorgulamalar ve hesaplamalar gereksizdir. Bu şekilde iterasyon tamamlanmış olunur. Bir sonraki iterasyon için eldeki üç köşe noktası $z_1 < z_2 < z_3$ şeklinde sıralanarak yeniden işlemlere başlanır.

İterasyonlar ilerledikçe minimum noktaya doğru yaklaşılır ve minimum nokta civarında köşe noktalarının birbirine yaklaşması sonucu küçülen üçgensel bölgede fonksiyon düzleşeceğinden köşe noktalarındaki fonksiyon değerleri de birbirine yakınsar.

4.2. Nelder-Mead Yönteminin Genelleştirilmesi

Nelder-Mead metodu yüksek boyutlu problemlerin çözümü için de genelleştirilmiştir. İki boyut için simpleks veya politop bir üçgendir, bu üçgen üç köşesinin koordinatları şeklinde ifade edilebilir. N boyutlu bir uzayda politop, N+1 köşeden oluşur.

Politopun köşelerindeki fonksiyon değerleri incelendiğinde özel olarak üç köşe şu şekilde seviyelendirilir:

H-en yüksek (highest), en yüksek fonksiyon değerine sahiptir;

L-en düşük (lowest) veya uzaktaki minimumun en iyi tahminidir;

N-yüksekten sonraki ikinci yüksek olan (next to highest), ikinci en büyük fonksiyon değerine sahiptir;

G-en düşükten veya en iyiden sonraki ikinci iyi noktadır.

Farklı ailelerden gelen politop algoritmaları fonksiyon yüzeyi üzerinde biraz farklı sezgisel yollarla daha düşük olan noktaları bulmaya çalışır. Bu yaklaşımlara Nash (1979) ve Parkinson ve Hutchinson (1972) tarafından değinilmiştir. Politop metodlarında kullanılan çeşitli taktiksel yaklaşımları ve politoplara uygulanan geometrik işlemleri tarif eden adlandırmalar bulunur.

1.- Yansıma: H nin dışında kalan diğer noktaların hepsinin ağırlık merkezi/ortalaması bulunur. N boyutlu uzayda bu nokta C olarak adlandırılır. Sonra H den C ye hareket edilerek HC doğrultusunda iki katı yol alınarak yansıma gerçekleştirilir ve yeni R noktasına varılır.

$$R = C + (C - H) = 2C - H \quad (4.4)$$

2.- Genişleme: R noktası özel bir nokta olmayabilir. Eğer en düşük L noktasından daha düşük fonksiyon değerli R noktası bulunmuş ise HC hattı yeni bir E noktasına kadar genişletilir, E noktasında fonksiyon değeri bulunur.

$$E = R + t(R - C) \quad (4.5)$$

(Özel olarak işlemlerden kaçınmak için $t=1$ alınabilir. Ancak CR doğrultusunda amaç fonksiyonunu minimize eden t değeri

$$f(E) = f(R + t(R - C)) = \phi(t) \rightarrow \min \phi(t) \quad (4.6)$$

şeklinde de aranabilir.)

3.- Büzülme: R noktası L noktasından daha düşük değilse HCR hattı üzerindeki diğer noktalar denenebilir.

3.a.- Eğer R deki fonksiyon değeri H den küçük ise “düşük-kenar” büzülmesi C-R aralığındaki bir noktayı dener.

3.b.- Eğer R noktası H den daha büyük ise “yüksek-kenar” büzülmesi H-C aralığındaki bir noktayı dener.

4.- Genel küçülme: L noktası akılda tutulur. Diğer bütün köşeler politopun kenarları boyunca L noktasına doğru hareket ettirilir. Nelder-Mead yaklaşımına göre kullanılan yöntemler aşağıdaki hesaplamalara yönlendirir:

4.1.- C ve R noktaları bulunur ve R noktasındaki fonksiyon hesaplanır.

4.2.- Hesaplamalar aşağıdaki durumlardan birine yönlendirilir:

4.2.a.- R noktası H den daha yüksektir. CH aralığında bir P1 noktası bulunur ve fonksiyon hesaplanır (yukarı-kenar indirilmesi). Eğer P1 noktası H den daha aşağıda ise P1 noktası H nin yerine geçer, eğer değilse genel küçülme gerçekleştirilir.

4.2.b.- R noktası H nin aşağısındadır fakat N nin yukarısındadır. CR aralığında bir P2 noktası bulunup fonksiyon hesaplanır (aşağı-kenar indirilmesi). Eğer P2 noktası R nin aşağısında ise P2 noktası H nin yerine geçer, aksi halde R noktası H nin yerine geçer.

4.2.c.- R noktası N nin aşağısındadır fakat L nin yukarısındadır, bu durumda R noktası H nin yerine geçer.

4.2.d.- R noktası L nin aşağısındadır. E noktası bulunup fonksiyon hesaplanır. Eğer E noktası R den daha düşük ise (bu durumda L nin de aşağısındadır) E noktası H nin yerine geçer, aksi halde politopta R noktası H nin yerine geçer.

Yukarıda verilen açıklamalarla olabilecek karışıklıklar önlenir. Politop algoritmasının çalıştırılmasında fonksiyonun eşit değerli olması durumu da dikkate alınmalıdır, buna bağlı olarak açık ve seçik bir yaklaşımla ne yapılacağının karar

verilmesi amaçlanmalıdır. Politopun köşelerindeki fonksiyon değerlerinin büyüklük sıralamasında da benzer bir karar mekanizmasına gereksinim duyulur.

Politopa uygulanan işlemlerin sonucunda yeni bir politop ortaya çıkar ve elde edilen yeni politop üzerinden aynı yaklaşımlarla işlemler yapılmaya devam edilir.

4.3. Tvrdik Yöntemi

Doğrusal olmayan regresyon modelinin parametrelerinin tahmini için Tvrdik yöntemi ile yukarıda bahsi geçen Nelder-Mead metodu temeline dayanan algoritmalar sınıfı oluşturulmuştur. Birer değişken olarak alınan parametrelerin uygun başlangıç değerleri bulunamadığında makul bir süre içinde güvenilir tahminlerinin bulunmasını sağlayan Tvrdik yöntemi bu başlık altında ele alınmıştır.

Toplamsal ve doğrusal olmayan regresyon modelinde rassal y vektörünün elemanları şu şekilde ifade edilir:

$$y_i = f(x_i, \theta_*) + \varepsilon_i, \quad i = 1, 2, \dots, n \quad (4.7)$$

Burada x_i bağımsız değişken vektörü $x_i^t = (x_{1i}, x_{2i}, \dots, x_{ki})$, θ parametre vektörü, parameteler açısından doğrusal olmayan f fonksiyonu, sıfır ortalamalı aynı dağılımlı bağımsız ε_i rassal değişkenleridir. Artık kareler toplamı $S(\theta)$ fonksiyonu

$$S(\theta) = \sum_{i=1}^n [y_i - f(x_i, \theta)]^2 \quad (4.8)$$

şeklinde verilir. En küçük kareler metoduyla bulunan parametre tahminleri $S(\theta)$ kareler toplamı fonksiyonunu minimize eder. Burada $S(\theta)$ fonksiyonu birden fazla yerel minimuma sahip olabilir. Bu aşamada $\hat{\theta}$ tahmini global optimizasyon problemidir.

4.3.1. Global optimizasyon

Kutu kısıtları ile global optimizasyonu probleminin olduğu sürekli bir fonksiyon mevcut olsun. Amaç fonksiyonunun

$$f : D \rightarrow R \quad , \quad D \subset R^d \quad (4.9)$$

şeklinde verildiği x^* noktası $x^* = \arg \min f(x)$, $x \in D$ olacak şekilde bulunmalıdır. x^* noktası global minimum, D uzayı $D = \prod_{i=1}^d [a_i, b_i]$, $a_i < b_i$, $i = 1, 2, \dots, d$ şeklinde tanımlanan araştırma uzayı ve $f(x)$ amaç fonksiyonunun hesaplanabilir olduğu varsayılmıştır. [Burada kolaylık için θ parametreleri x değişken vektörü şeklinde, $S(\theta)$ kareler toplamı fonksiyonu da $f(x)$ fonksiyonu şeklinde ifade edilmiştir.]

Yukarıda ifade edilen global optimizasyonu probleminin çözümü zor olmakla birlikte problemin çözümü için stokastik araştırma [Spall, 2003] ve evrimsel algoritmalar [Back, 1996] gibi sezgisel yaklaşımlar denenmiştir. Global optimizasyon problemlerinin çözümünde bu yaklaşımların her geçen gün önemi artmaktadır. Her iki yaklaşım, stokastik araştırma ve evrimsel algoritmalar şu anda kullanılan mevcut araştırma stratejisini uyarlamıştır. Evrimsel algoritmalarındaki uyarlama, temel olarak D uzayındaki noktalar popülasyonu üzerinde odaklanır. Bu algoritmaların temelini yönlendiren otomatik uyarlamalı mekanizmaları da kapsayacak şekilde bu girişimler son zamanlarda tanıtılmıştır. Bu mekanizmalar sezgisel evrimsel işlemlerin hesaplamalarına ve birlikte çalışabilirliğine dayanır [Deb, 2005; Winter ve dig., 2005; Tvrdik ve dig., 2001, 2002]. Olası optimizasyon problemlerinde kullanılan bütün araştırma algoritmalarının benzer düzeydeki başarısına sahip teoremlere rağmen, birçok optimizasyon problemlerinde deneysel karşılaştırmalar otomatik uyarlamalı algoritmaların diğerleri karşısında oldukça başarılı olduğu ifade edilmiştir [Wolpert ve Macready, 1997].

4.3.2. Kontrollü rassal araştırma (Controlled random search, CRS)

Yukarıda tanımlanan problemin global minimumunun bulunması için kullanılan kontrollü rassal araştırma (CRS) basit stokastik araştırma algoritmasıdır. Price (1977) tarafından önerilen bu algoritma, Nelder-Mead simpleks [Nelder ve Mead, 1964] metodundan “yansıma” noktası olan yeni deneme noktasının oluşturulması için kullanılmıştır. Global optimizasyon problemlerinin çözümünde başarıyla kullanılan CRS algoritmasının birçok modifikasyonu mevcuttur [Ali ve Törn, 2004]. CRS algoritmasının kesin olmayan (pseudo) kodu şu şekilde yazılabilir:

Algoritma. Genelleştirilmiş kontrollü rassal araştırma

- 1 D uzayında N noktadan oluşan P popülasyonunu rassal olarak üretiniz.
- 2 P deki en yüksek fonksiyon değerli x_{maks} noktasını bulunuz.
- 3 8.adımda sonlandırma olmamış ise 4-8 adımlarını tekrarlayınız.
- 4 Sezgisel yol ile $y \in D$ koşulunda bir deneme noktası üretiniz.
- 5 Eğer $f(y) < f(x_{maks})$ ise
- 6 $x_{maks} := y$;
- 7 yeni bir x_{maks} bulunuz;
- 8 sonlandırınız eğer:
- 9 Durdurma koşuluna kadar.

[Sadece mantıksal akış şemasının verildiği buradaki kodun ifade ettiği işlemler şunlardır:

-Birinci adımda kısıtları belirlenmiş olan D uzayından d boyutlu N tane nokta rassal olarak üretilir ve üretilen bu noktalara P popülasyonu denir. Buradaki d boyutu genel bir adlandırma olmakla birlikte regresyon modelinin p tane parametresi bulunuyorsa bu durumda araştırılan parametrelere karşılık gelen boyut p olur. Her bir parametre bir değişken gibi değerlendirilir.

-İkinci adımda P deki her noktadan f fonksiyonu teker teker hesaplanır ve en yüksek fonksiyon değerli nokta x_{maks} olarak adlandırılır.

- Üçüncü adımda dört ile sekizinci adımlar arasındaki adımların tekrarı sağlanır.
- Dördüncü adımda P noktaları ve elbette x_{maks} göz önünde bulundurularak Nelder-Mead algoritmasına göre y deneme noktası olarak adlandırılacak olan R yansıma noktası veya E genleşme noktası veya bir büzülme noktası bulunur. Metin içinde ifade edildiği gibi deneme noktasının bulunması için Nelder-Mead algoritmasının sezgisel versiyonları kullanılabilir. Yukarıda yazılmış olunan kod deneme noktasının belirlenmesi üzerinedir.
- Beşinci adımda Nelder-Mead algoritmasına göre x_{maks} ın yerine geçecek olan y noktasının $f(y) < f(x_{maks})$ koşulunu sağlayıp sağlamadığı kontrol edilir.
- Altıncı adımda bulunan y noktasının fonksiyon değeri x_{maks} ın fonksiyon değerinden daha küçük ise (en kötü) x_{maks} noktası y noktası olarak alınır. Burada nokta isimleri korunarak yeni y noktası sadece x_{maks} olarak adlandırılır; bu nokta politopun köşe noktaları içerisinde en büyük fonksiyon değerli nokta olmayabilir.
- Yedinci adım ile Nelder-Mead algoritmasına göre köşe noktaları fonksiyon değerlerine göre sıralamaya konarak en büyük fonksiyon değerli köşe noktası x_{maks} olarak adlandırılır.
- Sekizinci adım dokuzuncu adımdaki durdurma koşulu gerçekleşmiş ise aramayı sonlandırır. Eğer durdurma koşulu gerçekleşmemiş ise 4-8 adımları sırasıyla tekrar yerine getirilir.
- Dokuzuncu adım durdurma koşulunu kapsar.]

Dördüncü adımda bahsi geçen sezgisel yaklaşım ile herhangi bir deterministik olmayan kural yerine getirilir. Bu şekilde $y \in D$ koşulunda bir deneme noktası oluşturulur. Kullanılan birçok farklı sezgisel yol mevcuttur, bundan başka sezgisel yollar araştırma süresi boyunca biri diğerinin yerine geçerek değişebilir. Her iterasyonda $q_i, i = 1, 2, \dots, h$ olasılığında rassal olarak seçilebilen h tane sezgisel yol mevcut olsun. q_i olasılıkları önceki başarı oranlarıyla orantılıdır. Eğer sezgisel yol $f(y) < f_{maks}, f_{maks} = f(x_{maks})$ koşulunda “y” gibi bir deneme noktası oluşturmuş

ise şu anki mevcut adımda sezgisel yol başarılıdır. Başarı amaç fonksiyonunun değerindeki bağıl değişimle orantılı olarak ağırlıklandırılır.

$$w_i = \frac{f_{maks} - maks(f(y), f_{min})}{f_{maks} - f_{min}} \quad (4.10)$$

Böylece $w_i \in (0,1]$ ve karşılık gelen q_i olasılığı

$$q_i = \frac{W_i + w_0}{\sum_{j=1}^h (W_j + w_0)} \quad (4.11)$$

şeklinde değerlendirilir; burada W_i önceki araştırma adımlarındaki w_i ağırlıklarının toplamı ve w_0 algoritmanın başlangıç parametresidir. Herhangi bir q_i olasılığı verilen bir $\delta > 0$ limitinin altına düştüğünde evrimsel sürecin bozulmasını engellemek için şimdiki q_i değerleri başlangıç $q_i = 1/h$ değerlerine eşitlenir. Tvrdik ve dig. (2002) gösterdiği şekilde hesaplamaları yapılan sezgisel yollarla birlikte CRS algoritması evrimsel algoritmalar sınıfına dahildir.

Deney fonksiyonlarına ve doğrusal olmayan regresyon modellerine [Tvrdik ve Krivy, 2004] bağlı olarak birçok sezgisel yollar deneysel olarak kontrol edilmiştir. Doğrusal olmayan regresyon parametre tahmini için deneysel sonuçlarla birlikte dört sezgisel yol CRS algoritmasının işletilmesinde kullanılmıştır. Bunlardan üç tanesi Krivy ve Tvrdik (1995) tarafından önerilen (P populasyonundan $d+1$ noktanın seçimiyle) simpleks S deki rassallaştırılmış yansıma tekniğine dayanır. Yeni bir y deneme noktası

$$y = g + U(g - x_H) \quad (4.12)$$

ile simpleks yöntemden üretilir. Burada $x_H = \arg maks_{x \in S} f(x)$ ve S simpleksindeki geri kalan d noktanın ağırlık merkezi g noktasıdır. U çarpan faktörü $\alpha > 0$ ve

$0 < s < \alpha/2$ şeklinde parametreler olmak üzere $[s, \alpha - s)$ bağılı olarak tekdüze dağılımlı rassal bir değişkendir. U çarpanının ortalaması ve varyansı sırasıyla $EU = \alpha/2$ ve $\text{var}(U) = (\alpha - 2s)^2/12$ şeklindedir. İki sezgisel yola bağılı olarak simpleksin bütün $d + 1$ noktası P popülasyonundan rassal olarak seçilir. $\alpha = 2$ ve $s = 0.5$ olarak seçilmiş bir sezgisel yol, $\alpha = 5$ ve $s = 1.5$ ile diğer bir sezgisel yoldur. Üçüncü sezgisel yol dikkate alındığında, simpleksin bir noktası P nin minimum amaç fonksiyonu değerli noktasıdır ve S simpleksinin geri kalan d noktası P nin geri kalan noktalarından rassal olarak seçilir. Üçüncü yolun parametreleri $\alpha = 2$ ve $s = 0.5$ olarak alınmıştır.

Dördüncü sezgisel yol diferansiyel evrimsel (DE) algoritmaya dayanır [Storn ve Price, 1997].

$$u = r_1 + F(r_2 - r_3) \quad (4.13)$$

ifadesine göre u noktası üretilir. Burada r_1 , r_2 ve r_3 üç ayrı nokta olmak üzere P den rassal olarak alınmıştır ve $F > 0$ bir parametredir. y deneme noktasının y_j , $j = 1, 2, \dots, d$ elemanları rassal olarak alınan x ile oluşturulur, buradaki x o anki r_1 , r_2 ve r_3 değerlerinden farklıdır. Buradaki u aşağıdaki koşulu sağlar.

$$y_j = \begin{cases} u_j & , \quad U_j < C \vee j = l \\ x_j & , \quad U_j > C \wedge j \neq l \end{cases} \quad (4.14)$$

burada l tamsayı olmak üzere $\{1, 2, \dots, d\}$ kümesinden rassal olarak seçilir. $U_1, U_2, \dots, U_d$ bağımsız rassal değişkenleri tekdüze $[0, 1]$ dağılımından gelir. $C \in [0, 1]$, y nin oluşturulması aşamasında karşılıklı değiştirilen eleman sayısına etki eden bir parametredir. Araştırma süreci boyunca Ali ve Torn (2004) tarafından F çarpanının

$$F = \begin{cases} maks \left(F_{\min}, 1 - \left| \frac{f_{maks}}{f_{\min}} \right| \right) & , \quad \left| \frac{f_{maks}}{f_{\min}} \right| < 1 \\ maks \left(F_{\min}, 1 - \left| \frac{f_{\min}}{f_{maks}} \right| \right) & , \quad d.d. \end{cases} \quad (4.15)$$

denkleminde göre bulunması önerilmiştir. Burada $f_{\min}$, f_{maks} popülasyon içindeki sırasıyla minimum ve maksimum amaç fonksiyonu değerleridir. $F \in [F_{\min}, 1)$ koşulunu sağlayacak şekilde $F_{\min}$ bir parametredir. Bu yaklaşımın parametreleri $F_{\min} = 0.4$ ve $C=0.9$ olarak alınmıştır.

CRS algoritmasının dokuzuncu adımındaki durdurma koşulu yapılan uygulamalar için $f(x)$ amaç fonksiyonunun indirgemeler arası farkların değişim oranının bir katsayıdan düşük gerçekleşmesi şekli olarak kabul edilmiştir. Yapılan bir iterasyonla amaç fonksiyonunun değerinde bir düşüş gerçekleştirilmiş ise amaç fonksiyonunun değeri bu düşüşle f_k şeklinde indisenerek düşüşlerin gerçekleştirildiği farklar $f_k - f_{k-1} = z_k$ negatif değerler olarak bulunur. Bu durumda durdurma koşulu

$$\delta = 10^{-2} \text{ ile } \frac{z_k}{z_{k-1}} = \frac{f_{k-1} - f_k}{f_{k-2} - f_{k-1}} < \delta \text{ olarak alınmıştır.}$$

Hataların homojen olmaması durumunda ve aykırı değerlerin bulunması durumunda ağırlıklandırmayla birlikte sağlam (robust) fonksiyonlarıyla Tvrdik yöntemi aşamalı olarak kullanılabilir.

5. YAKLAŞIM ÖNERİLERİ

Bu tezde yapılan çalışmalar doğrultusunda birçok nokta farkedilmiş ve burada tanıtılmış yöntemlerin ve yaklaşımların üzerine bazı fikirler geliştirilmiştir. Bu bölümde bu noktalar üzerinde durulmuştur.

5.1. Bir Stokastik Yaklaşım Önerisi

Tvrđik yönteminin yapısı ve davranışı incelendikten sonra şu şekilde bir yaklaşım önerilmiştir. Yapısal olarak temelde bir stokastik yaklaşım stratejisidir.

Krviy ve Tvrđik (1995) tarafından önerilen $y = g + U(g - x_H)$ stokastik araştırması temel alındığında (U katsayısı açısından) Nelder-Mead metodunda R yansıma noktasının ve E genişleme noktasının bulunmasını sağlayan katsayılar 1 ve 2 katsayılarıdır: $R = C + (C - H)$ ve $E = R + (R - C)$ yada $E = C + 2(C - H)$. Eğer R yansıma noktası H noktasından daha düşük değerli değilse HR aralığında karşılaştırma yapılan deneme noktaları aranır. Bu durumda HR aralığı katsayılar anlamında $(-1, +1)$ aralığına denk gelir: $H = C - (C - H)$ ve $R = C + (C - H)$. Katsayılar $(C - H)$ teriminin katsayısından görülebilir. Eğer HR aralığında da bir indirgeme sağlanamıyorsa simpleks L noktasına doğru çekilir. Nelder-Mead metodu temelli araştırmalar için birçok strateji ve yaklaşım önerilmiştir. Tüm bu yaklaşımlar ve stratejiler stokastik bir araştırma ile kısmen simüle edilebilir. Krviy ve Tvrđik (1995) tarafından önerilen yaklaşım, y deneme noktası için $y = x_C + U(x_C - x_H)$ ifadesindeki rassal U çarpanının belirlendiği aralık üzerinde durur. HR doğrultusunda en fazla genişleme E noktası ile 2 katsayısı, eğer R yansıma noktası H noktasından daha düşük değerli değilse HR aralığında H noktasından başlayacak şekilde bir arama stratejisi uygulanırsa H noktasına karşılık gelen katsayı -1 olur. Bu durumda rassal U katsayısının $(-1, +2)$ aralığından gelmesi sağlanarak ifade edilen bütün farklı stratejiler bir şekilde simüle edilmiş olunur. Burada önerilen yaklaşımın temeli $y = x_C + U(x_C - x_H)$ ifadesinden rassal U katsayısının $(-1, +2)$ aralığından

tekdüze gelerek y deneme noktasının elde edilmesine dayanır. Doğrudan bu ifade üzerinden yapılan iterasyonlarda simpleksin tüm köşe noktaları Krviy ve Tvrdik yaklaşımında olduğu gibi sıklıkla bir düzlem içinde sıkışıp kalmaktadır. Tvrdik yönteminde bu sorundan Storn ve Price yaklaşımı olan diferansiyel evrimsel algoritması kullanılarak aşılmıştır. Buradaki asıl amaç, beklenmedik bir nokta üretilerek bulunulan konumdan kurtulmayı sağlamaktır. Bunun için deneme noktasının bir şekilde hatalandırılarak düzlemin dışına çıkılması sağlanır. Deneme noktası büyüklüğü ile orantılı olarak $y = [x_C + U(x_C - x_H)](1 + u)$ ifadesine göre her bileşen için ayrı u_j rassal katsayısı $(-0.1, +0.1)$ aralığından tekdüze dağılımdan gelerek hatalandırılır.

Ancak $y = [x_C + U(x_C - x_H)](1 + u)$ ifadesi gözönünde bulundurulduğunda y deneme noktasının değişkenliğinin kendi büyüklüğü ile orantılı yani orijine göre simetrik olması bazı sorunlara neden olabilir. Genel olarak herhangi bir fonksiyon için optimal değerler orijinden oldukça uzak noktalarda bulunabilir. Eğer optimal noktalar orijinden oldukça uzaklarda bulunuyorsa bir değişkenlik ölçüsü olan u rassal sayısı dolayısıyla çok geniş bir bölgenin taranması yakınsama sorunlarına neden olur. Bundan dolayı bu yaklaşımın bir modifikasyonu üzerinde durulmuştur.

Deneme noktasının bir şekilde hatalandırılması ile $y = x_C + U(x_C - x_H)$ deneme noktasının merkezde olduğu bir bölge oluşturulmalıdır. Bu bölgeyi en iyi şekilde simpleksin veya politopun kapladığı alan ve bu alanın yakın çevresi temsil edebilir. Bu şekilde politopun L noktasına büzülme stratejileri de stokastik yaklaşım içine dahil edilmiş olunur. Bu bölgenin yarıçapı H noktası haricinde C ağırlık merkezinin hesabında kullanılan politopun köşe noktalarının yayılımıyla ilişkilendirilebilir. C ağırlık merkezinin hesabında kullanılan köşe noktalarının karşılıklı bileşenlerinin standart sapması bu yarıçapın büyüklüğü ile orantılıdır; bu durumda N büyük bir sayı ise N boyutlu uzay için iki standart sapmalı bir yarıçap politopun tüm köşelerinin %95 ini (en azından büyük bir bölümünü) kapsar. Bu yarıçap bir sezgisel yol için politopun yakın çevresini de içine alacak şekilde dört standart sapmaya çıkarılmıştır.

Bu seviye bir ölçü olarak alınır. j.ninci bileşen için y_2 deneme noktasının bileşenlerinin $y_{2j} = C_j + U_2(C_j - H_j) + u_{2j}s_j$ şeklinde hesaplandığı bir yaklaşım belirir. Burada u_{2j} rassal sayısı $(-4, +4)$ aralığından tekdüze dağılımdan gelir. C ağırlık merkezinin hesabına dahil edilen H harici P_i köşe noktalarının karşılıklı bileşenlerinden j.ninci bileşen için standart sapması

$$s_j = \left[\frac{1}{N-1} \sum_{i=1}^N (P_{ij} - C_j)^2 \right]^{1/2}, \quad C_j = \frac{1}{N} \sum_{i=1}^N P_{ij}, \quad j = 1, 2, \dots, N \quad (5.1)$$

şeklinde bulunur. Böylece politopun büyüklüğü ölçüsünde izin verilen bir uzay içinde minimum nokta stokastik araştırma ile bulunabilir. Bu şekilde Nelder-Mead yöntemindeki ve buna bağlı olarak simpleks algoritmasındaki tüm stratejiler ve yaklaşımlar burada önerilen yaklaşımla simüle edilmiş olunur.

Bazı durumlarda ara aşamalarda politopun boyutları son derece küçülebileceğinden veya politop yassılaşabileceğinden bazı muhtemel tıkanıklıklardan oransal hatalandırma yaklaşımı aracılığı ile çıkılabilir.

Son olarak politopun en düşük noktası olan L köşesi bir araştırma alanının merkezi olarak kabul edilip büyük bir yarıçap aracılığı ile araştırma alanının sınırları çizilir. Söz gelimi bu yarıçap her bileşen için 100 birim olarak alınabilir. Bir j.ninci bileşen için y_3 deneme noktası $y_{3j} = L_j + u_{3j}$ şeklinde oluşturulur. Burada her bileşen için u_{3j} rassal sayısı $(-100, +100)$ aralığından tekdüze dağılımdan gelir.

Bu başlık altında oluşturulan üç sezgisel yolun her biri stokastik araştırma stratejisine dayanır. Her iterasyon için mevcut politopun köşe noktaları kullanılarak her üç yol ile deneme noktaları bulunur. Bulunan deneme noktalarından en küçük fonksiyon değerli olanı mevcut iterasyon için incelemeye alınır. İncelemeye alınan deneme noktası politopun en büyük değerli H noktasından düşük değerli ise politopun yeni

köşesi olarak kabul edilir. Eğer H noktasından küçük değilse politop değiştirilmeden bırakılır. İterasyon bu şekilde tamamlanır.

Birinci sezgisel yol y_1 deneme noktasının j.ninci bileşeni $y_{1j} = [C_j + U_1(C_j - H_j)](1 + u_{1j})$ şeklindeki oransal hatalandırmadır; herbir u_{1j} rassal sayısı tekdüze(-0.1,+0.1) dağılımından bağımsız olarak gelir. İkinci sezgisel yol y_2 deneme noktasının j.ninci bileşeni s_j standart sapma ölçüsünde $y_{2j} = C_j + U_2(C_j - H_j) + u_{2j}s_j$ şeklindeki rassallaştırmadır; herbir u_{2j} rassal sayısı tekdüze(-4,+4) dağılımından bağımsız olarak gelir. Üçüncü sezgisel yol politopun en düşük L köşesinin merkez olarak alındığı araştırma alanından y_3 deneme noktasının j.ninci bileşeni $y_{3j} = L_j + u_{3j}$ şeklinde oluşturulur; herbir u_{3j} rassal sayısı tekdüze(-100, +100) dağılımından bağımsız olarak gelir. Buradaki ilk iki yol için rassal U_1 ve U_2 katsayıları bağımsız olmak üzere herbiri tekdüze(-1, +2) dağılımından gelir.

Boyut sayısı arttıkça incelenmesi gereken uzaysal alan çok daha hızla büyüyeceğinden indirgemelerin sağlandığı iterasyonların sayısı hızla düşer. Buna bağlı olarak yakınsama için daha fazla iterasyon gerekir.

Burada önerilen yaklaşımda kullanılan üçüncü yolda iterasyonlarda indirgemenin azalmasına bağlı olarak araştırma alanının büyüklüğünü belirleyen yarıçapın uygun bir artışını sağlayan bir bağıntı geliştirilmeye çalışılmıştır. Ancak basit ve fonksiyonel bir ifade bulunamamıştır. İşlemlerin basitliği çalışma performansını kesinlikle arttırdığından bu yapıyı bozacak bir ifadenin yerleştirilmemesine özen gösterilmiştir.

Burada önerilen yaklaşımla hata varyansları farklı olan ve aykırı değerler barındıran gözlemler üzerinden sağlam fonksiyonlar kullanılarak ağırlıklandırma ile birlikte parametre tahminlerinin yapılması mümkündür.

5.2. Yöntemlerin Genelleştirilmesi

Genelleştirilmiş en küçük kareler yöntemini, yarı olabilirlik tahminini ve sağlam yaklaşımlarını kapsayacak şekilde bir genelleme yapılabilir.

Hata varyanslarının değişkenlikleri dolayısıyla gözlemler yada hatalar ağırlıklandırılarak dönüştürülmüş hataların homojen olmaları sağlanır. Hata varyanslarının tahmini için ağırlıklandırılmış en küçük kareler yönteminin arka arkaya kullanımı sağlam ρ (robust) fonksiyonlarının kullanıldığı sağlam tahminlerinin geliştirildiği iteratif tekrar ağırlıklandırılmalı en küçük kareler yöntemini anımsatır. Veri içinde aykırı değerler mevcutsa bu aykırı değerlerin ağırlıkları sağlam ρ fonksiyonları aracılığı ile düşürülerek parametre tahmininin daha hatasız yapılması sağlanır. Bu yaklaşımda da ağırlıklandırmanın merkezi rolü bulunur. Buradaki yaklaşımlar arasındaki asıl bağlantıyı yarı olabilirlik tahmininin geliştirildiği matematiksel yapılanma sağlar. Yarı olabilirlik tahmin yaklaşımıyla bilinmeyen parametreler ve bilinmeyen hata varyansları iteratif yöntemlerle çözülerek bulunur. Hata varyanslarının iteratif yöntemlerle bulunması yaklaşımıyla ağırlıklandırılmış en küçük kareler yönteminin arka arkaya kullanımı ve dolayısıyla iteratif tekrar ağırlıklandırılmalı en küçük kareler yöntemi arasında işlemsel benzerlikler kurulabilmektedir.

Bir karışıklığın önlenmesi için şu şekilde bir adlandırma kullanılmaktadır. $\hat{\theta}$ tahmininin bir vektör olduğu bilinerek tahmin vektörü bu şekilde gösterilmektedir. Parametre vektörünün gerçek değeri θ_* şeklinde ve herhangi bir değer θ olarak ifade edilmiştir. Doğrusal olmayan regresyon modellerinde gerçek parametre değeri, parametre tahmini ve parametrenin herhangi bir değeri (bir değişken şeklinde) kavramları kesin bir şekilde birbirlerinden ayrılmalıdır. Doğrusal regresyon modellerinde parametre tahmininin tek bir işlemle bulunmasından dolayı özellikle parametrenin herhangi bir değeri (bir değişken şekli) kavramına pek gerek duyulmaz. Aşağıdaki yazım şekline bağlı kalınmıştır.

$y = f(\theta_*) + \varepsilon$: Gerçek parametre değeri ve hatalar ile teorik model

$y - f(\hat{\theta}) = e$: Parametre tahmini ve artıklar ile tahmin modeli

$y - f(\theta) = r$: Parametre tahmininin geliştirildiği ara aşamalarda model, herhangi bir θ değeri ve buna karşılık gelen artıklar

Hataların varyansı $V(\varepsilon) = V$ köşegen matrisi ise $V = K^2$ olmak üzere hatalar K^{-1} ağırlığı ile $K^{-1}\varepsilon$ şeklinde ağırlıklandırılarak hatalar üzerindeki aşırılıklar yok edilir. Bu durumda ağırlık matrisi $V^{-1} = W$ şeklinde ifade edilir. Eğer hata varyansı büyük ise dengeleme için w ağırlığı varyansın tersi şeklinde alınarak hatanın dönüştürülmüş büyüklüğü düşürülmüş olur. Örneğin ε_i hatasının varyansı s_i^2 şeklinde tahmin edilmiş ise ağırlığı $w_i = 1/s_i^2$ şeklinde alınır. Genel olarak köşegen ağırlık matrisi $W = \text{diag}\{w_i\}$ şeklinde oluşturulur. Modelin dönüştürülmesi K matrisi aracılığı ile yapılıyorsa köşegen K matrisinin elemanları s_i standart sapma tahminlerinden oluşur: $K = \text{diag}\{s_i\}$.

$$V = K^2 \rightarrow \begin{cases} V = \text{diag}\{s_i^2\} \\ K = \text{diag}\{s_i\} \end{cases} \quad (5.2)$$

$$V^{-1} = W \rightarrow W = \text{diag}\{1/s_i^2\}$$

Hataların varyanslarının tahmini yaklaşımlarıyla aykırı değerlerin incelemelerinin aynı anda yapılması, çalışmaları bazen çıkmaza götürebilmektedir. Çalışmalarda ve analizlerde parametre tahminindeki bu etkilerin azaltılması girişimlerinin ayrı ayrı ele alınması ortak analizinin zor ve karmaşık olduğunun bir göstergesidir. Burada en azından matematiksel olarak bu iki yaklaşımın ortak olarak ele alınabileceğinin olabilirliği gösterilmiştir. Sağlam yaklaşımında kullanılan ρ fonksiyonu işlemlerin akışı dolayısıyla ağırlık fonksiyonu şeklinde ele alınabildiği bir yapıya dönüşmektedir. Özel olarak sabit varyans varsayımında aykırı değerlerin ağırlığını belirleyen ifade

$$w_i = \frac{\psi\left(\frac{r_i}{s}\right)}{\frac{r_i}{s}}, \quad r_i = y_i - f_i(\theta) \quad (5.3)$$

şeklinde belirir. Sağlam ρ fonksiyonunun, r_i artığının s standart sapması tahminine bölünerek elde edilen ölçeklendirilmiş argumanı bulunur. Bu şekilde sağlam fonksiyonunun artıkların mutlak büyüklükleri karşısında değişmez (invariant) olmaları sağlanır. Verinin üzerine kurulduğu model dolayısıyla teorik hatalarının varyansları farklı ise ve aykırı değerler mevcut ise bir sezgi w ağırlığının

$$\left. \begin{array}{l} w_i = 1/s_i^2 \\ w_i = \psi\left(\frac{r_i}{s_i}\right) / \left(\frac{r_i}{s_i}\right) \end{array} \right\} \rightarrow w_i = \frac{\psi\left(\frac{r_i}{s_i}\right)}{s_i^2 \left(\frac{r_i}{s_i}\right)} \quad (5.4)$$

şeklinde olduğunu söyler. Bu aşamada i.ninci gözleme bağlı olarak s_i standart sapması tahmini mevcuttur. Bu şekilde her iki özellik w ağırlığının içinde toparlanır. Örneğin i.ninci gözlem bir aykırı gözlem değilse $\psi(z)/z$ bir değerini verirken i.ninci gözlemin s_i^2 varyansı dolayısıyla ağırlık $w_i = 1/s_i^2$ büyüklüğünde bulunur; eğer i.ninci gözlem bir aykırı değer ise bu durumda ψ fonksiyonu ağırlık yükünü düşürerek bu aykırı değer parametre tahmini üzerindeki etkisini azaltır. Bu sezginin doğru olduğu aşağıda gösterilmiştir.

Bir modelin oluşturulması aşamasında elde edilen r_i artıklarının standart sapmalarına bölümüyle normalleştirildikleri varsayılın. Sağlam regresyon yaklaşımından bir alıntı yapılarak normalleştirilen artıklara bir işlem yapan özel bir ρ fonksiyonu tasarlansın. Buradaki özel ρ fonksiyonu artıkların varyansını dikkate alarak otomatik olarak ağırlıklandırır ve gerektiğinde aykırı değerlerin önemini düşüren bir fonksiyon olsun. $r = y - f(\theta)$ ile en küçük kareler tahmininin ele alındığı kareler

toplamı $S(\theta) = r^t r = (y - f(\theta))^t (y - f(\theta))$ fonksiyonu karesel bir fonksiyondur. Bu durumda ρ fonksiyonu kareseldir, ancak bu fonksiyon artıklara gerektiği şekilde otomatik olarak ağırlık verememektedir. Bundan farklı olarak kısmen kareler toplamı fonksiyonunun özelliklerine, gerektiğinde ağırlıklandırmalı kareler toplamı fonksiyonunun özelliklerine ve bunların yanında kendi içinde istenen her türlü özelliğe sahip bir amaç fonksiyonu tasarlanmalıdır. Bu amaç fonksiyonu kareler toplamı fonksiyonundan farklı olmak üzere

$$T = \sum_i \rho \left(\frac{r_i}{s_i} \right) \quad (5.5)$$

şeklinde tarif edilsin. Amaç fonksiyonunun tüm işlevselliği içindeki ρ fonksiyonundan gelmektedir. Bu aşamada bir deneme olarak sağlam yaklaşımındaki ρ sağlam fonksiyonlarından biri alınabilir, ancak buradaki ρ fonksiyonunun daha kapsamlı özellikleri mevcut olabilir. Eğer amaç fonksiyonunun kayıpların en aza indirilmesi gerektiği şeklinde kurulmuş ise T amaç fonksiyonunun minimize edilmesiyle $\hat{\theta}$ parametre tahminine ulaşılır; diğer bir ifade ile $\hat{\theta}$ parametre tahmini T amaç fonksiyonunu minimize eder. Herhangi bir θ değeri için artıklar $r_i = y_i - f_i(\theta)$ şeklinde tanımlanarak T amaç fonksiyonunun θ_j parametresine göre kısmi türevi

$$T = \sum_i \rho \left(\frac{y_i - f_i(\theta)}{s_i} \right) \rightarrow \frac{\partial T}{\partial \theta_j} = - \sum_i \frac{1}{s_i} \frac{\partial f_i}{\partial \theta_j} \psi \left(\frac{y_i - f_i(\theta)}{s_i} \right), \quad \rho' = \psi$$

$$\frac{\partial T}{\partial \theta_j} = - \sum_i \frac{1}{s_i} \frac{\partial f_i}{\partial \theta_j} \psi \left(\frac{y_i - f_i(\theta)}{s_i} \right) = 0 \rightarrow \hat{\theta} \quad (5.6)$$

sıfıra eşitlenirse çözümle elde edilen sonuç $\hat{\theta}$ parametre tahminidir. Burada ψ fonksiyonu ρ fonksiyonunun türevidir. Normal denklemlerin benzer bir ifadesi

$$\sum_i \frac{1}{s_i} \frac{\partial f_i}{\partial \theta_j} \psi \left(\frac{y_i - f_i(\theta)}{s_i} \right) = 0 \quad (5.7)$$

şeklinde verilir. Normal denklemlerin doğrudan çözümü yerine iteratif yöntemlerle çözüme ulaşılabilmektedir.

$f_i(\theta)$ fonksiyonu bir θ_0 civarında Taylor açılımı aracılığı ile elde edilen

$$f_i(\theta) = f_i(\theta_0) + \nabla f_i(\theta_0)^t (\theta - \theta_0) \quad , \quad \nabla f_i(\theta_0)^t = \frac{\partial f_i}{\partial \theta^t} \Big|_{\theta_0} \quad (5.8)$$

yaklaşık açılımı kullanılarak herhangi bir θ değeri için artıklar $r_i = y_i - f_i(\theta)$ ile T amaç fonksiyonu yaklaşık olarak

$$T = \sum_i \rho \left(\frac{r_i}{s_i} \right) = \sum_i \rho \left(\frac{y_i - f_i(\theta_0) - \nabla f_i(\theta_0)^t (\theta - \theta_0)}{s_i} \right) \quad (5.9)$$

şeklinde alınabilir. Şimdi T fonksiyonunun θ_j parametresine göre kısmi türevi

$$\frac{\partial T}{\partial \theta_j} = - \sum_i \frac{1}{s_i} \frac{\partial f_i}{\partial \theta_j} \Big|_{\theta_0} \psi \left(\frac{y_i - f_i(\theta_0) - \nabla f_i(\theta_0)^t (\theta - \theta_0)}{s_i} \right) = 0 \quad , \quad \rho' = \psi \quad (5.10)$$

sıfıra eşitlenerek $r_i = y_i - f_i(\theta_0) - \nabla f_i(\theta_0)^t (\theta - \theta_0)$ ile

$$\sum_i \frac{1}{s_i} \frac{\partial f_i}{\partial \theta_j} \Big|_{\theta_0} \psi \left(\frac{r_i}{s_i} \right) = 0 \quad (5.11)$$

ifadesi yazılır. ψ fonksiyonunun bulunduğu yere altlı-üstlü $\left(\frac{r_i}{s_i} \right)$ çarpanı eklenirse

yukarıdaki ifade ile eşdeğer

$$\sum_i \frac{1}{s_i} \frac{\partial f_i}{\partial \theta_j} \Big|_{\theta_0} \frac{\psi\left(\frac{r_i}{s_i}\right)}{\left(\frac{r_i}{s_i}\right)} = 0 \rightarrow \sum_i \frac{1}{s_i^2} \frac{\partial f_i}{\partial \theta_j} \Big|_{\theta_0} \frac{\psi\left(\frac{r_i}{s_i}\right)}{\left(\frac{r_i}{s_i}\right)} r_i = 0 \quad (5.12)$$

ifadesi elde edilir. Bu noktada $\psi\left(\frac{r_i}{s_i}\right) / \left(\frac{r_i}{s_i}\right)$ bir ağırlık olarak değerlendirilip soldaki tüm çarpanların θ_0 noktasındaki değerleri hesaplanırsa

$$\sum_i \frac{\partial f_i}{\partial \theta_j} \Big|_{\theta_0} \left[\frac{\psi\left(\frac{r_i}{s_i}\right)}{s_i^2 \left(\frac{r_i}{s_i}\right)} \right] \Big|_{\theta_0} r_i = 0 \quad (5.13)$$

ifadesi elde edilmiş olur. Yarı olabilirlik yaklaşımında normal denklemlerin kurulması için uygulanan yaklaştırma işlemi kullanılmıştır. Burada θ_0 noktasına bağlı olarak

$$r_i = y_i - f_i(\theta) = y_i - f_i(\theta_0) - \nabla f_i(\theta_0)^t (\theta - \theta_0) \rightarrow r_i \Big|_{\theta_0} = y_i - f_i(\theta_0) \quad (5.14)$$

ile

$$\sum_i \frac{\partial f_i}{\partial \theta_j} \Big|_{\theta_0} \left[\frac{\psi\left(\frac{y_i - f_i(\theta_0)}{s_{i0}}\right)}{s_{i0}^2 \left(\frac{y_i - f_i(\theta_0)}{s_{i0}}\right)} \right] \left[y_i - f_i(\theta_0) - \nabla f_i(\theta_0)^t (\theta - \theta_0) \right] = 0 \quad (5.15)$$

şekli elde edilir. Buradaki son eşitlik aranan bütün özellikleri sağlar ve bundan sonra yazılan ifadelerin tümü bu eşitlik ile özdeştir. j.ninci parametre için yazılan yukarıdaki eşitlik matrislerin kullanımıyla daha kısa bir şekilde yazılabilir.

$$\sum_i \frac{\partial f_i}{\partial \theta_j} \Big|_{\theta_0} w_{i0} r_i = 0 \rightarrow z_{j0}^t W_0 \left[y - f(\theta_0) - \frac{\partial f}{\partial \theta^t} \Big|_{\theta_0} (\theta - \theta_0) \right] = 0 \quad (5.16)$$

Burada z_{j0} kolon vektörü Z_0 matrisinin j.ninci kolonudur, W_0 köşegen matrisinin elemanları w_{i0} ağırlıklarından oluşur.

$$w_{i0} = \frac{\psi \left(\frac{y_i - f_i(\theta_0)}{s_{i0}} \right)}{s_{i0}^2 \left(\frac{y_i - f_i(\theta_0)}{s_{i0}} \right)}, \quad W_0 = \text{diag} \{ w_{i0} \}_{i=1}^n \quad (5.17)$$

$$z_{j0}^t = \left\{ \frac{\partial f_i}{\partial \theta_j} \Big|_{\theta_0} \right\}_{i=1}^n, \quad Z_0 = \{ z_{j0} \}_{j=1}^p = \frac{\partial f}{\partial \theta^t} \Big|_{\theta_0} = \begin{bmatrix} \nabla f_1(\theta_0)^t \\ \nabla f_2(\theta_0)^t \\ \vdots \\ \nabla f_n(\theta_0)^t \end{bmatrix} = \left[\left(\frac{\partial f_i}{\partial \theta_j} \Big|_{\theta_0} \right)_{ij} \right] \quad (5.18)$$

Tüm parametreler için Z_0 matrisi tamamlanırsa

$$Z_0^t W_0 [y - f(\theta_0) - Z_0(\theta - \theta_0)] = 0 \quad (5.19)$$

ağırlıklandırılmış en küçük kareler yöntemindeki normal denklemler elde edilir. Buradan θ çözülürse ara iterasyondaki bir tahmin elde edilir; bu θ_1 olarak adlandırılarak bir sonraki iterasyonda θ_0 yerine kullanılır. Bu şekilde θ_2 elde edilir.

$$\theta_1 = \theta_0 + (Z_0^t W_0 Z_0)^{-1} Z_0^t W_0 [y - f(\theta_0)] \quad (5.20)$$

Sonraki iterasyonlar benzer şekilde devam ettirilir.

$$\theta_{k+1} = \theta_k + (Z_k^t W_k Z_k)^{-1} Z_k^t W_k [y - f(\theta_k)] \quad (5.21)$$

$$Z_k = \frac{\partial f}{\partial \theta^t} \Big|_{\theta_k} \quad , \quad W_k = \text{diag} \{ w_{ik} \}_{i=1}^n \quad , \quad w_{ik} = \frac{\psi \left(\frac{y_i - f_i(\theta_k)}{s_{ik}} \right)}{s_{ik}^2 \left(\frac{y_i - f_i(\theta_k)}{s_{ik}} \right)} \quad (5.22)$$

Yeterli sayıdaki iterasyonla $\hat{\theta}$ parametre tahminine ulaşılır.

$$\lim_{k \rightarrow \infty} \theta_k = \hat{\theta} \quad (5.23)$$

Normal denklemlerin çözümünün geliştirildiği Ek 1 deki çalışmaya bakılabilir.

Genel olarak W ağırlık matrisi

$$W = \text{diag} \left\{ \frac{1}{s_i^2} \frac{\psi \left(\frac{r_i}{s_i} \right)}{\frac{r_i}{s_i}} \right\} \quad (5.24)$$

şeklinde ele alınarak

$$w_{i0} = \frac{\psi \left(\frac{r_i}{s_i} \right)}{s_i^2 \left(\frac{r_i}{s_i} \right)} \Big|_{\theta_0} \quad , \quad w_{ik} = \frac{\psi \left(\frac{r_i}{s_i} \right)}{s_i^2 \left(\frac{r_i}{s_i} \right)} \Big|_{\theta_k} \quad (5.25)$$

ağırlıkları θ_0 ve θ_k noktaları için hesaplanabilir olduğu kabul edilebilir.

Eğer ρ fonksiyonu sağlam yaklaşımındaki sağlam ρ fonksiyonlarından biri olarak alınırsa hata varyanslarının farklılıklarından dolayı bir ağırlıklandırma kendiliğinden oluşmaktadır.

T amaç fonksiyonu ile bulunan parametre tahminlerinin varyans-kovaryans matrisi istatistiksel testler için gerekli olduğundan bulunmalıdır. Parametre tahminlerinin varyans-kovaryans matrisi tahminine şu şekilde ulaşılır. Herhangi bir θ için $r_i = y_i - f_i(\theta)$ olmak üzere $T = \sum_i \rho\left(\frac{r_i}{s_i}\right)$ şeklinde verilen bir amaç fonksiyonunun bir ekstremumuna

$$\begin{aligned}\theta_{k+1} &= \theta_k + (Z_k^t W_k Z_k)^{-1} Z_k^t W_k [y - f(\theta_k)] \\ \lim_{k \rightarrow \infty} \theta_k &= \hat{\theta}\end{aligned}\tag{5.26}$$

iterasyonları ile varıldıktan sonra $\hat{\theta}$ parametre tahmini ile

$$\lim_{k \rightarrow \infty} Z_k = \hat{Z} = Z(\hat{\theta}) \quad , \quad \lim_{k \rightarrow \infty} W_k = \hat{W} = W(\hat{\theta})\tag{5.27}$$

tahminleri de bulunmuş olur. Bir σ^2 ölçeği ile eğer hataların varyansı $V(\varepsilon) = \sigma^2 V$ köşegen bir matris ise V matrisinin bir tahmini $\hat{V} = \hat{W}^{-1}$ şeklinde bulunur. Bu aşamada $\hat{W}$ ağırlık tahminlerinin bulunmasından sonra iterasyonlarla elde edilen $\hat{\theta}$ parametre tahminlerini bulan ağırlıklandırılmış kareler toplamı fonksiyonu

$$S = [y - f(\theta)]^t \hat{W} [y - f(\theta)]\tag{5.28}$$

şeklinde verilir. Buradaki ağırlıklandırılmış kareler toplamı fonksiyonunun optimal değerini yukarıda bulunan $\hat{\theta}$ parametre tahminleri verir. Genel olarak V köşegen olmayan simetrik bir matris ve σ^2 bir ölçek olmak üzere $y = f(\theta_*) + \varepsilon$ regresyon modelinin hatalarının $E(\varepsilon) = 0$, $V(\varepsilon) = \sigma^2 V$ özellikleri mevcutsa $V = K^2$ yardımıyla hatalar $K^{-1}\varepsilon = \iota$ şeklinde ağırlıklandırılarak dönüştürülmüş hataların $E(\iota) = 0$, $V(\iota) = \sigma^2 I$ şeklinde homojen olmaları sağlanır. $K^{-1}\varepsilon = \iota$ dikkate alınarak $\hat{\theta}$

parametre tahminlerini bağlayan dönüştürülmüş artıklar üzerinden σ^2 varyansının tahmini

$$s^2 = \frac{\|K^{-1}[y - f(\hat{\theta})]\|^2}{n - p} = \frac{[y - f(\hat{\theta})]^t V^{-1}[y - f(\hat{\theta})]}{n - p} \quad (5.29)$$

şeklinde bulunur. Ağırlıklandırılmış $S = [y - f(\theta)]^t W [y - f(\theta)]$ kareler toplamı fonksiyonu θ_* gerçek parametre değeri civarında

$$S = [y - f(\theta_*) - Z_*(\theta - \theta_*)]^t W [y - f(\theta_*) - Z_*(\theta - \theta_*)] \quad (5.30)$$

şeklinde açılarak minimize edildiğinde

$$\hat{\theta} = \theta_* + (Z_*^t W Z_*)^{-1} Z_*^t W \varepsilon \quad (5.31)$$

$\hat{\theta}$ parametre tahmininin θ_* gerçek parametre değeri civarındaki teorik dağılımının bulunabileceği ifade elde edilir. $V(\varepsilon) = \sigma^2 V$ özelliği göz önünde bulundurularak $\hat{\theta}$ parametre tahminlerinin varyans-kovaryans matrisi

$$V(\hat{\theta}) = \sigma^2 (Z_*^t W Z_*)^{-1} Z_*^t W V W Z_* (Z_*^t W Z_*)^{-1} \quad (5.32)$$

ve asimptotik sonuçlar doğrultusunda varyans-kovaryans matrisi tahmini

$$\hat{V}(\hat{\theta}) = s^2 (\hat{Z}^t \hat{W} \hat{Z})^{-1} \hat{Z}^t \hat{W} V \hat{W} \hat{Z} (\hat{Z}^t \hat{W} \hat{Z})^{-1} \quad (5.33)$$

şeklinde bulunur. Yukarıda T amaç fonksiyonunda köşegen W ağırlık matrisinin iterasyonlarla geliştirilmesindeki nedenlerden biri $y = f(\theta_*) + \varepsilon$ regresyon modelinin hataları varyanslarının farklı olmasıdır. Eğer hataların varyansı $V(\varepsilon) = \sigma^2 V$ köşegen

bir matris ise V matrisinin bir tahmini $\hat{V}^{-1} = \hat{W}$ ifadesinden elde edilebileceğinden $\hat{\theta}$ parametre tahminlerinin varyans-kovaryans matrisi $V(\hat{\theta}) = \sigma^2 (Z_*^t W Z_*)^{-1}$ biçiminde ve parametre tahminlerinin varyans-kovaryans matrisi tahmini ise $\hat{V}(\hat{\theta}) = s^2 (\hat{Z}^t \hat{W} \hat{Z})^{-1}$ biçiminde kısaltılabilir. Aykırı değerler barındıran model artıklarının ağırlıkları köşegen W matrisi aracılığı ile çıkarıldığından bunlara karşı gelen varyans tahminlerinin ölçeklenmiş büyüklüğü $s^2 \hat{V}$ ifadesinden bulunabilir.

Herhangi bir yöntemle bulunan parametre tahminlerini sağlayacak şekilde bir ağırlıklandırılmış $S = [y - f(\theta)]^t W [y - f(\theta)]$ kareler toplamı fonksiyonu oluşturulabilir. Genel olarak $y = f(\theta_*) + \varepsilon$ regresyon modelinin hatalarının $E(\varepsilon) = 0$, $V(\varepsilon) = \sigma^2 V$ özellikleri bulunuyorsa bu yöntemle elde edilen parametre tahminlerinin varyans-kovaryans matrisi $V(\hat{\theta}) = \sigma^2 (Z_*^t W Z_*)^{-1} Z_*^t W V W Z_* (Z_*^t W Z_*)^{-1}$ şeklinde bulunur.

5.3. Hata Varyans-Kovaryans Matrisi ve Optimal Sağlam Fonksiyonu

Eğer ε hataları $(0, V)$ şeklinde dağılıyorsa $\text{cov}(\varepsilon_i, \varepsilon_j) = E(\varepsilon_i \varepsilon_j)$ özelliği gözönünde bulundurularak herhangi bir θ ile $r_i = y_i - f_i(\theta)$ artıkları üzerinden hata varyans-kovaryans matrisinin bir tahmininin çıkarımı için ilk olarak $V = [(r_i r_j)_{ij}]$ şekli ele alınabilir. Eğer tüm matris $V = [(r_i r_j)_{ij}]$ şeklinde kurulursa matrisin satırları veya sütunları arasında doğrusal bağımlılık oluşur, buna bağlı olarak matris tersinir değildir. Kovaryans kavramından çıkarılabilecek $r_i r_j$ ifadesi üzerinde şu şekilde bir

modifikasyon yapılabilir: $r_i r_j = s_i \frac{r_i}{s_i} s_j \frac{r_j}{s_j}$. Burada r_i/s_i ifadesinde artık karşılık

gelen standart sapmaya bölünerek bir anlamda standardize edilir. Eğer r_i/s_i aşırı büyük bir değeri veriyorsa doğal olarak bu büyüklüğü dengeleyebilen sağlam fonksiyonlarının kullanımı düşünülebilir. Amaç hata varyans-kovaryans matrisinin tahmini için matris $r_i r_j$ elemanlarının yerine geçebilecek uyumlu değerler

bulabilmektedir. s_i standart sapmasının tahmininde aykırı değer etkilerinden korunmak için medyan kullanılabilir. Dolayısıyla $\rho \geq 0$ olmak üzere r_i/s_i ifadesini dengeleyen ve uyumlu hale getiren $\rho(r_i/s_i)$ sağlam fonksiyonu bulunmalıdır. Hata varyans-kovaryans matrisinin köşegeni hata varyanslarına karşılık gelirken köşegen dışındaki elemanlar kovaryanslara karşılık gelir. Bu durum gözönünde bulundurulursa $r_i r_j$

ifadesi yerine $s_i s_j \left[\rho \left(\frac{r_i}{s_i} \right) \right]^{1/2} \left[\rho \left(\frac{r_j}{s_j} \right) \right]^{1/2} \text{sign}(r_i r_j)$ ifadesinin konması gerekir. Buna

bağlı olarak hata varyans-kovaryans matrisinin bir tahmini

$$V = \left[\left(s_i s_j \left[\rho \left(\frac{r_i}{s_i} \right) \right]^{1/2} \left[\rho \left(\frac{r_j}{s_j} \right) \right]^{1/2} \text{sign}(r_i r_j) \right)_{ij} \right] \text{ şeklinde ele alınır. Ancak bu matris de}$$

$V = [(r_i r_j)_{ij}]$ matrisi gibi tersinir değildir. Doğal olarak yakın olan gözlemler arasında, örneğin $|i - j| < b$ ile bir b uzaklığına kadar olan gözlemler arasında korelasyon olabilir; uzak gözlemler arasında korelasyonların bulunması (periyodik bir yapı yoksa) pek anlamlı değildir. $r_i/s_i = z_i$ gösterimi ile bir genelleme yapılırsa

$$\rho(z_i, z_j) = \begin{cases} \sqrt{\rho(z_i) \rho(z_j)} \text{sign}(z_i z_j) & , |i - j| \leq b \\ 0 & , |i - j| > b \end{cases} \text{ ifadesi elde edilir. Bu durumda}$$

hata varyans-kovaryans matrisi $V = \left[\left(s_i s_j \rho(z_i, z_j) \right)_{ij} \right]$ şekline girer; bu şekil ile daha

genel yapılar tasarlanabilir, özel bir durum olarak belirli koşullarda

$$\rho(z_i, z_j) = \sqrt{\rho(z_i) \rho(z_j)} \text{sign}(z_i z_j) \text{ gibi bir ifade ele alınmış olunur. Eğer hata}$$

varyans-kovaryans matrisi köşegen bir matris ise optimal sağlam fonksiyonunun çıkarımı kolaylaşır. V hata varyans-kovaryans matrisi köşegen ise köşegen elemanları

$$v_{ii} = s_i^2 \rho(z_i) \text{ olarak hesaplanır. Yukarıda ele alınmış olan } w_i \text{ ağırlık elemanları, } s_i$$

standart sapması ve aykırı değer etkisini azaltan $\psi(z_i)/z_i$ ifadesi ile

$$w_{ii} = \psi(z_i)/s_i^2 z_i \text{ şeklinde kurulmuştur. Bu durumda eşitlik kurulabilir.}$$

$$\begin{aligned}
w = \frac{1}{v} &\rightarrow \frac{\rho'}{s^2 z} = \frac{1}{s^2 \rho} \rightarrow \frac{\rho'}{z} = \frac{1}{\rho} \rightarrow \rho \rho' = z \\
&\rightarrow \rho \frac{d\rho}{dz} = z \rightarrow \rho d\rho = z dz
\end{aligned} \tag{5.34}$$

$\rho' = \psi$ olduğu gözönünde bulundurularak diferansiyel denklem çözülür.

$$\rho^2 = z^2 + C, \quad (z=0, \rho=0) \rightarrow \rho^2 = z^2 \rightarrow \rho = |z| \tag{5.35}$$

Buradan optimal sağlam fonksiyonu $T = \sum_i \rho \left(\frac{r_i}{s_i} \right)$ amaç fonksiyonundan

$$\rho(z) = |z| = \left| \frac{r}{s} \right| \rightarrow \rho' = \psi = \text{sign}(z) = \text{sign}(r) \text{ şeklinde bulunmuş olur. Eğer } V$$

hata varyans-kovaryans matrisi köşegen değilse bu fonksiyonun yapısı oldukça karmaşılaşır ve buradaki gibi basit bir ifade bulunamaz.

Sağlam fonksiyonunun $\rho = |z|$, $\psi = \rho' = \text{sign}(z)$, $w = \frac{\psi}{z} = \frac{1}{|z|}$ olması ile bir L1-norm

regresyonu söz konusu olur; bu yaklaşım ölçeklenmiş mutlaklar toplamı olarak adlandırılabilir. Yapılan uygulamalarda varyansları durağan olmayan ancak ortalamada en küçük varyanslı tahminlerin L1-norm regresyonu ile bulunduğu görülmekle birlikte parametre tahminleri gerçek değerleri civarında durağanlaşmadan salınmaktadır. Bu durum amaç fonksiyonunun yapısından kaynaklanır. Ancak bu yaklaşımla bulunan parametre tahminlerinin bir açıdan dikkat çekici olduğu ifade edilmelidir.

Burada tanımı verilmeye çalışılan $\rho(z_i, z_j)$ fonksiyonu ile bir hata varyans-kovaryans matrisi tahmini $V = \left[\left(s_i s_j \rho(z_i, z_j) \right)_{ij} \right]$ şeklinde yapılandırılabilir.

6. PARAMETRE TAHMİN YÖNTEMLERİNİN KARŞILAŞTIRILMASI İÇİN SİMÜLASYON ÇALIŞMASI

Bu bölümde parametreleri bilinen bazı doğrusal olmayan modeller için simülasyon yoluyla üretilen veri ile parametre tahmin yöntemlerinin çalışmasına ve bu sürecin 1000 kez tekrarı ile elde edilen tahminlerin ortalama değerleri ve standart hatalarına ilişkin sonuçlara yer verilmiştir. Bu amaçla tez çalışmasında bahsedilen parametre tahmin yöntemleri yanlılık ve etkinlik bakımından karşılaştırılmıştır.

Bu bölümde parametreleri $\theta_1 = 2$ ve $\theta_2 = 3$ olan *Model 1*

$$y = \ln\left(\frac{\theta_1 x}{\theta_2 + x}\right) + \varepsilon \quad (6.1)$$

modelinde simülasyon yoluyla üretilen 1a, 1b, 1c ve 1d verileri, parametreleri $\theta_1 = 1000$ ve $\theta_2 = 500$ olan *Model 2*

$$y = \frac{\theta_1}{\theta_2 + x^2} + \varepsilon \quad (6.2)$$

modelinden simülasyon yoluyla üretilen 2a, 2b ve 2c verileri ile parametreleri $\theta_1 = 100$, $\theta_2 = 99$ ve $\theta_3 = 0.2$ olan *Model 3*

$$y = \frac{\theta_1}{1 + \theta_2 e^{-\theta_3 x}} + \varepsilon \quad (6.3)$$

modelinden simülasyon yoluyla üretilen 3a verileri kullanılmıştır. Parametrelerine göre doğrusallaştırılamayan Model 3 aynı zamanda lojistik büyüme modelidir.

İlk olarak verilen modeller aracılığı ile veri üretimi aşağıda açıklandığı şekilde yapılmıştır.

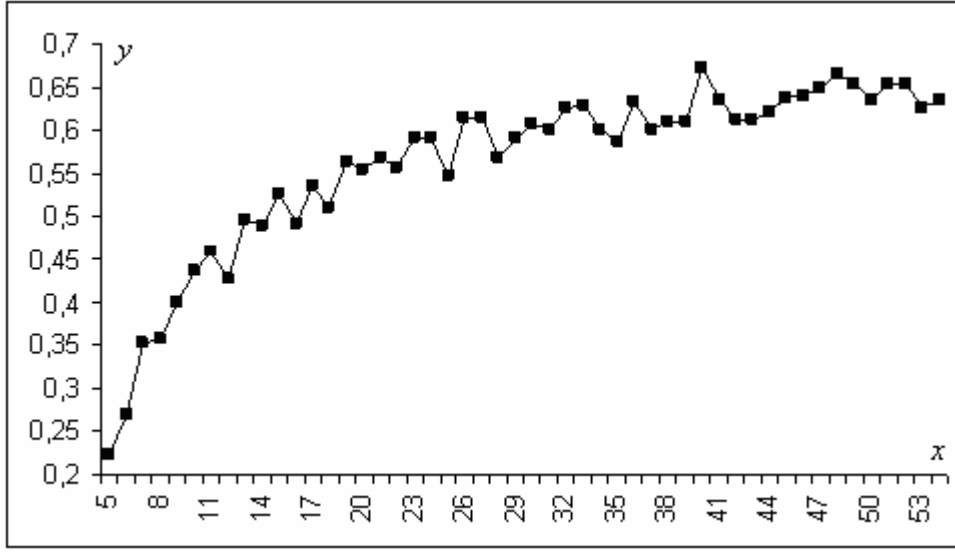
6.1. Model 1a Verisi

$$\text{Model 1: } y = \ln\left(\frac{\theta_1 x}{\theta_2 + x}\right) + \varepsilon, \quad \begin{cases} \theta_1 = 2 \\ \theta_2 = 3 \end{cases} \quad (6.4)$$

Model 1 aracılığı ile bağımsız x değişkeninin 5-54 aralığında birer artarak oluşturulan 50 gözlemlik çıktısına beklenen değerleri sıfır ve varyansı $(0.02)^2 = 0.0004$ olan normal dağılımdan bağımsız hatalar üretilerek eklenmiş ve bu şekilde bir Model 1a verisi elde edilmiştir. Simülasyon ile üretilen Model 1a verisi aşağıda sunulmuş ve grafiği çizilmiştir.

Çizelge 6.1. Model 1 aracılığı ile üretilen Model 1a verisi

x	y	x	y	x	y
5	0,223756	22	0,556189	39	0,609637
6	0,269309	23	0,592260	40	0,671750
7	0,352889	24	0,592054	41	0,635043
8	0,358179	25	0,548095	42	0,611045
9	0,400148	26	0,614315	43	0,611820
10	0,435260	27	0,614454	44	0,621665
11	0,459469	28	0,568256	45	0,637742
12	0,425824	29	0,591174	46	0,639256
13	0,495906	30	0,606713	47	0,648211
14	0,489976	31	0,601323	48	0,665923
15	0,525646	32	0,624910	49	0,653216
16	0,491969	33	0,627390	50	0,634116
17	0,535877	34	0,600888	51	0,653711
18	0,509804	35	0,586957	52	0,654089
19	0,562969	36	0,633695	53	0,626761
20	0,553430	37	0,599687	54	0,635392
21	0,568967	38	0,609528		



Şekil 6.1. Model 1a verisinin diyagramı

6.1.1. Model 1a verisi ve en küçük kareler yöntemi

Model 1 ile oluşturulan veri üzerinden önce Gauss-Newton iterasyonu yöntemi ile parametre tahmini yapılmıştır. İterasyon için parametrelerin başlangıç değerlerine gereksinim duyulduğundan modelin f beklenen fonksiyonu göz önünde bulundurularak parametrelerine göre doğrusal bir yapı elde edilmiştir.

$$y = \ln\left(\frac{\theta_1 x}{\theta_2 + x}\right) \rightarrow e^y = \frac{\theta_1 x}{\theta_2 + x} \rightarrow e^{-y} = \frac{1}{\theta_1} + \frac{\theta_2}{\theta_1} \frac{1}{x} \rightarrow v = \beta_0 + \beta_1 u \quad (6.5)$$

$$\theta_1 = \frac{1}{\beta_0}, \quad \theta_2 = \frac{\beta_1}{\beta_0}$$

$v = e^{-y}$, $u = 1/x$ dönüşümüyle elde edilen veri üzerinden doğrusal modelin parametre tahminleri olağan en küçük kareler yaklaşımıyla $\hat{\beta}_0 = 0.496333$, $\hat{\beta}_1 = 1.555063$ olarak bulunduktan sonra orijinal modelin parametre başlangıç değerleri $\theta_{10} = 2.014778$, $\theta_{20} = 3.133106$ olarak bulunmuştur. Gauss-Newton iterasyonunun işleyebilmesi için gerekli olan Z matrisinin satırı şu şekilde kurulur:

$$\frac{\partial f}{\partial \theta_1} = \frac{1}{\theta_1}, \quad \frac{\partial f}{\partial \theta_2} = \frac{-1}{\theta_2 + x} \quad (6.6)$$

Aşağıda Gauss-Newton iterasyonu sonuçları verilmiştir.

Çizelge 6.2. Model 1a verisi için Gauss-Newton iterasyonu sonuçları

k	θ_1	θ_2	S	$\Delta S/S$	$\Delta \theta/\theta$
0	2,014777715	3,133105644	0,016745780		
1	2,016986298	3,155360770	0,016728059	-0,001058247	0,007103216
2	2,016993587	3,155449323	0,016728059	-1,48670E-08	2,80643E-05
3	2,016993606	3,155449558	0,016728059	-1,05582E-13	7,42385E-08
4	2,016993606	3,155449558	0,016728059	0	1,95995E-10

Tabloda S kareler toplamına bağlı olarak

$$\frac{\Delta S}{S} = \frac{S(\theta_k) - S(\theta_{k-1})}{S(\theta_{k-1})} \quad (6.7)$$

iterasyonlar ilerlerken kareler toplamı fonksiyonundaki oransal küçülmeyi gösterir; $\Delta S/S$ oranı negatif ise küçülme gerçekleşir. Burada $\Delta \theta/\theta$ ise

$$\max_j \left| \frac{\theta_{j,k} - \theta_{j,k-1}}{\theta_{j,k-1}} \right|, \quad j = 1, 2, \dots, p \quad (6.8)$$

ile p parametre içerisinde en çok oransal değişim gösteren oranı ifade eder; örneğin

$\delta = 10^{-6}$ için $\max_j \left| \frac{\theta_{j,k} - \theta_{j,k-1}}{\theta_{j,k-1}} \right| < \delta$ ile durağanlaşmanın gerçekleştiği kabul edilebilir.

Verilen bilgiler ışığında $k = 3$ iterasyonunda optimal parametre tahminlerine ulaşılmıştır.

Model 1a verisi üzerinden elde edilen s^2 varyans tahmini, σ^2 teorik varyans, $\hat{\theta}$ parametre tahminleri, θ_* parametrelerinin gerçek değeri, $\hat{\theta} - \theta_*$ parametre tahminleri ile gerçek değerleri arasındaki farkları, $\hat{\theta}$ parametre tahminleri ve s^2 varyans tahmini dolayısıyla $s(\hat{\theta})$ parametre tahminlerinin standart hata tahminleri, parametrelerin ve hata varyansının gerçek değerleri bilindiğinden $\sigma(\hat{\theta})$ parametre tahminlerinin teorik standart hataları sırası ile şöyledir:

$$\begin{aligned} s^2 &= 0.000348501, \quad \sigma^2 = 0.0004 \\ \hat{\theta} &= \begin{pmatrix} 2.016993 \\ 3.155449 \end{pmatrix}, \quad \theta_* = \begin{pmatrix} 2 \\ 3 \end{pmatrix}, \quad \hat{\theta} - \theta_* = \begin{pmatrix} -0.016993606 \\ -0.155449558 \end{pmatrix} \\ s(\hat{\theta}) &= \begin{pmatrix} 0.010070727 \\ 0.104982096 \end{pmatrix}, \quad \sigma(\hat{\theta}) = \begin{pmatrix} 0.010622072 \\ 0.110385232 \end{pmatrix} \end{aligned} \quad (6.9)$$

6.1.2. Model 1a süreci için simülasyon çalışması

Model 1a verisi oluşturma süreci 1000 kez tekrarlanarak herbiri 50 şer gözlem den oluşan 1000 örnek oluşturulmuştur. Elde edilen 1000 örneğin herbiri için ayrı ayrı en küçük kareler (EKK) yöntemi ve L1-norm regresyonu (L1) yöntemi uygulanarak parametre tahminleri yapılmıştır. Parametre tahminleri ortalamalarında ve özellikle standart hata tahminleri ortalamalarında 20 tekrar civarında durağanlaşma gözlenmiştir. Simülasyonla Model 1a süreci ile üretilen 1000 örneğin $\bar{\hat{\theta}}$ parametre tahminleri ortalamaları ve $\bar{s}(\hat{\theta})$ parametre tahminlerinin standart hataları tahminleri ortalamaları aşağıdaki çizelgede sunulmuştur. Çizelgede “Tekrar” olarak gösterilen sütun simülasyon yapılarak üretilen tekrarların sayısıdır.

Çizelge 6.3. Model 1a süreci örneklerine uygulanan yöntemler ve sonuçları

Yöntem	Tekrar	$\bar{\hat{\theta}}_1$	$\bar{s}(\hat{\theta}_1)$	$\bar{\hat{\theta}}_2$	$\bar{s}(\hat{\theta}_2)$
EKK	1000	1.999605	0.010607	2.991848	0.110179
L1	1000	1.998855	0.000843	3.006192	0.008633

En küçük kareler yöntemi ile bulunan parametre tahminleri ve standart hataları tahminleri ortalamaları simülasyonla üretilen tekrar sayıları arttıkça bir limite yakınsama eğilimindedir. Parametre tahminlerinin ortalamaları gerçek parametre değerlerine, parametre tahminlerinin standart hata tahminleri ortalamaları da parametre tahminlerinin teorik standart hatalarına doğru yakınsamaktadır. L1-norm regresyonu yöntemi ile bulunan parametre tahminlerinin standart hataları tahminleri ortalamaları daha küçük bulunmuştur; bu noktaya bölümün sonunda değinilmiştir.

6.2. Model 1b Verisi

Parametre değerleri $\theta_1 = 2, \theta_2 = 3$ olan Model 1 aracılığı ile bağımsız x değişkeninin 5-54 aralığında birer artarak oluşturulan 50 gözlemlik çıktısına beklenen değerleri sıfır ve varyansı $(0.02)^2 = 0.0004$ olan normal dağılımdan bağımsız hatalar üretilerek eklenmiş ancak 25. ve 53. gözlemlerinin aykırı değerler olması için normal hatalar yerine büyük mutlak değerli -0.3 hataları eklenerek Model 1b verisi oluşturulmuştur. Bu şekilde üretilen Model 1b verisi aşağıda sunulmuştur.

Çizelge 6.4. Model 1b verisi

x	y	x	y	x	y
5	0,223756	22	0,556189	39	0,609637
6	0,269309	23	0,592260	40	0,671750
7	0,352889	24	0,592054	41	0,635043
8	0,358179	25	0,279818	42	0,611045
9	0,400148	26	0,614315	43	0,611820
10	0,435260	27	0,614454	44	0,621665
11	0,459469	28	0,568256	45	0,637742
12	0,425824	29	0,591174	46	0,639256
13	0,495906	30	0,606713	47	0,648211
14	0,489976	31	0,601323	48	0,665923
15	0,525646	32	0,624910	49	0,653216
16	0,491969	33	0,627390	50	0,634116
17	0,535877	34	0,600888	51	0,653711
18	0,509804	35	0,586957	52	0,654089
19	0,562969	36	0,633695	53	0,338087
20	0,553430	37	0,599687	54	0,635392
21	0,568967	38	0,609528		



Şekil 6.2. Model 1b verisi için diyagram

6.2.1. Model 1b verisi ve en küçük kareler yöntemi

Modelin f beklenen fonksiyonundan elde edilen doğrusal yapının parametre tahminleri $\hat{\beta}_0 = 0.508078949$, $\hat{\beta}_1 = 1.462518241$ şeklinde bulunduktan sonra orijinal modelin parametre başlangıç değerleri $\theta_{10} = 1.968198$, $\theta_{20} = 2.878526$ olarak elde edilmiştir. Bu vektör Gauss-Newton iterasyonu için başlangıç değeri olarak kullanılmış ve aşağıdaki iterasyon sonuçları bulunmuştur.

Çizelge 6.5. Model 1b verisi için Gauss-Newton iterasyonu sonuçları

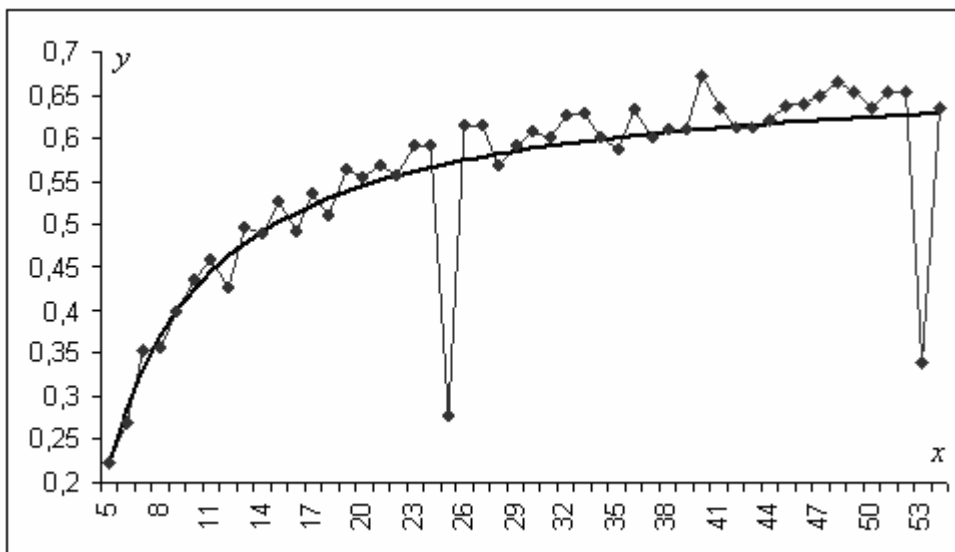
k	θ_1	θ_2	S	$\Delta S/S$	$\Delta \theta/\theta$
0	1,968198058	2,878525562	0,192568805		
1	1,975006883	2,912066372	0,192313633	-0,001325097	0,011652080
2	1,975019848	2,912113259	0,192313632	-5,99512E-09	1,61010E-05
3	1,975019846	2,912113228	0,192313632	0	1,06794E-08

Veri üzerinden bulunan tahminler ile birlikte parametrelerin ve hata varyansının gerçek değerleri gözönünde bulundurularak yapılan hesaplamalar aşağıda sunulmuştur.

$$\begin{aligned}
s^2 &= 0.004006534, & \sigma^2 &= 0.0004 \\
\hat{\theta} &= \begin{pmatrix} 1.975019 \\ 2.912113 \end{pmatrix}, & \theta_* &= \begin{pmatrix} 2 \\ 3 \end{pmatrix}, & \hat{\theta} - \theta_* &= \begin{pmatrix} -0.024980154 \\ -0.087886772 \end{pmatrix} \\
s(\hat{\theta}) &= \begin{pmatrix} 0.033062426 \\ 0.345629960 \end{pmatrix}, & \sigma(\hat{\theta}) &= \begin{pmatrix} 0.010622072 \\ 0.110385232 \end{pmatrix}
\end{aligned} \quad (6.10)$$

Burada $\hat{\theta}$ parametre tahminleri ile varyans tahmininin s^2 olduğu $s(\hat{\theta})$ parametre tahminlerinin standart hata tahminleri ile $\sigma(\hat{\theta})$ teorik standart hataları arasındaki fark dikkati çeker. Bu farkın nedeni iki aykırı değer parametre tahminlerini etkilemesinden dolayıdır. Teorik $\sigma(\hat{\theta})$ standart hataları iki aykırı değer varlığı ihmal edilerek geri kalan hatalara verilen teorik $\sigma^2 = 0.0004$ hata varyansı ve gerçek θ_* parametre değerleri kullanılarak hesaplanmıştır. Hata varyansı iki aykırı değeri kapsamamakla birlikte s^2 hata varyansı tahmini aykırı değerleri de içine alacak şekilde bulunmuştur.

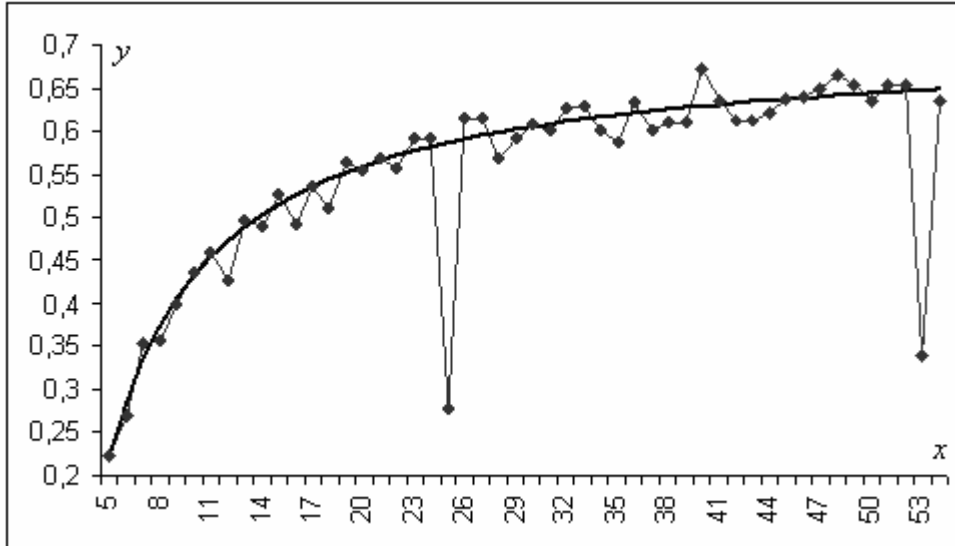
Elde edilen sonuçlar doğrultusunda aşağıda veri ve regresyon eğrisi birlikte çizilmiştir.



Şekil 6.3. Model 1b verisi ile en küçük kareler eğrisi

6.2.2. Model 1b verisi ve mutlak değerler toplamı

Model 1b verisi kullanılarak büyük hatalara karşı daha az duyarlı olan mutlak artıklar toplamı üzerinden L1-norm regresyonu uygulanarak parametre tahmini yapılmıştır. Yukarıda bulunmuş olan doğrusallaştırılmış model yardımıyla elde edilen orjinal modelin parametre başlangıç değerleri (1.968198058, 2.878525562) L1-norm regresyon problemi için başlangıç değerleri olarak kullanılmıştır. Bu yaklaşımla elde edilen L1-norm regresyon parametre tahminleri (2.022004536, 3.114189142) olarak bulunmuştur. Aşağıda veri ve uydurulan eğri birlikte çizilmiştir.



Şekil 6.4. Model 1b verisi için L1-norm regresyon yöntemi ile bulunan eğri

6.2.3. Model 1b verisi ve M-tahmin edici

Model 1b verisi üzerinden Andrew dalga fonksiyonunun kullanıldığı M-tahmin edici ile parametre tahmini yapılmıştır. İterasyonlar için parametre başlangıç değerleri olarak L1-norm regresyon parametre tahminleri (2.022004536, 3.114189142) alınmıştır. Aşağıda M-tahmin edici için yapılan iteratif tekrar ağırlıklandırılmalı en küçük kareler yöntemi iterasyonu sonuçları verilmiştir.

Çizelge 6.6. Andrew dalga fonksiyonunun kullanıldığı M-tahmin edici ile yapılan iterasyon sonuçları

k	θ_1	θ_2	S	$\Delta S/S$	$\Delta \theta/\theta$
0	2,022004536	3,114189142	0,205336689		
1	2,019977775	3,149062388	0,202433231	-0,014139989	0,011198178
2	2,019521508	3,151830139	0,202046584	-0,001909997	0,000878913
3	2,019443407	3,152093420	0,201988948	-0,000285260	8,35325E-05
4	2,019430541	3,152112543	0,201980324	-4,26942E-05	6,37150E-06
5	2,019428490	3,152112999	0,201979041	-6,35466E-06	1,01563E-06
6	2,019428170	3,152112784	0,201978851	-9,40639E-07	1,58278E-07

Altıncı iterasyonda parametre tahminlerine ulaşıldığı varsayılabilir.

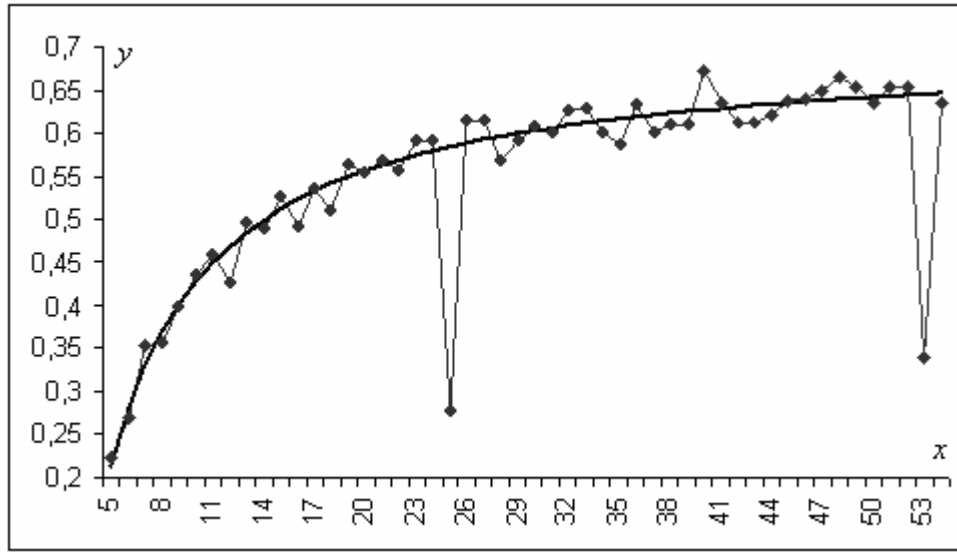
Andrew dalga fonksiyonu kullanılarak M-tahmin edici aracılığı ile elde edilen ağırlıklandırılmış en küçük kareler parametre tahminleri ile birlikte bulunan diğer tahminler toplu olarak aşağıda sunulmuştur.

$$s^2 = 0.000254595$$

$$\hat{\theta} = \begin{pmatrix} 2.019428 \\ 3.152112 \end{pmatrix}, \quad \theta = \begin{pmatrix} 2 \\ 3 \end{pmatrix}, \quad \hat{\theta} - \theta_* = \begin{pmatrix} 0.019428112 \\ 0.152112701 \end{pmatrix} \quad (6.11)$$

$$s(\hat{\theta}) = \begin{pmatrix} 0.009089398 \\ 0.092964348 \end{pmatrix}$$

Ağırlıklar ayırt edilmeksizin bütün gözlemler kullanılarak parametre tahminleri yapılmıştır. Aşağıda veri ve uydurulan eğri birlikte çizilmiştir.



Şekil 6.5. Model 1b verisi için Andrew dalga fonksiyonu kullanılarak M-tahmin edici aracılığı ile bulunan eğri

İteratif tekrar ağırlıklandırılmalı en küçük kareler yöntemi ile elde edilen ağırlıklar aşağıda verilmiştir.

Çizelge 6.7. Andrew dalga fonksiyonu kullanılarak M-tahmin edici ile elde edilen ağırlıklar

x	w	x	w	x	w
5	0,977101	22	0,961422	39	0,943424
6	0,969614	23	0,924704	40	0,582009
7	0,888878	24	0,961934	41	0,990470
8	0,963067	25	0	42	0,911713
9	0,998626	26	0,845640	43	0,904039
10	0,990024	27	0,886795	44	0,965878
11	0,982177	28	0,822903	45	0,998334
12	0,598372	29	0,982846	46	0,998218
13	0,975147	30	0,996510	47	0,974578
14	0,977133	31	0,994788	48	0,836691
15	0,956214	32	0,939102	49	0,961300
16	0,783678	33	0,940889	50	0,986269
17	0,997598	34	0,958119	51	0,971777
18	0,775567	35	0,801508	52	0,975473
19	0,955702	36	0,947926	53	0
20	0,997787	37	0,893511	54	0,972683
21	0,991342	38	0,956028		

25. ve 53. gözlemlere sıfır ağırlık verilmekle birlikte 12. ve 40. gözlemlere verilen düşük ağırlıklar dikkati çeker.

Andrew dalga fonksiyonu ile ağırlıklarının sıfıra düşürüldüğü verilerin parametre tahmininin bulunmasındaki etkileri de yok edilmiştir. Bu durumda artıklar üzerinden tahmin edilen varyans tahmini için serbestlik derecesinin seçimi bir sorun olabilir. Modelin kurulması aşamasında mevcut olan gözlemlerin ağırlıklarının sıfırlanması ile bu gözlemlerin ihmal edilip edilmediği sorunu mevcuttur. Eğer ihmal edilirlerse ve serbestlik derecesi buna bağlı olarak düşürülürse varyans tahmini artar.

Ayrıca ağırlıkları sıfırlamayan ve anlaşılması kolay olan Huber t fonksiyonu parametre tahmini için kullanılmıştır. Huber t fonksiyonu için $t = 2$ olarak alınmasıyla ikiden küçük (standartlaştırılmış) artıkların en küçük kareler yaklaşımındaki gibi ağırlıkları bir olarak alınır. Ancak ikiden büyük ise bir aykırı değer olarak değerlendirilip mutlak büyüklüğü ile ters orantılı olarak ağırlıklandırılır. Bu şekildeki bir ağırlıklandırmanın mutlaklar toplamından elde edilen ağırlıklar olmasından dolayı parametre tahmini açısından avantajları bulunur. Huber t fonksiyonu en küçük kareler yaklaşımı ile mutlaklar toplamı yaklaşımını iki ayrı koşul olarak içinde barındırır.

Model 1b verisi üzerinden Huber t fonksiyonunun kullanıldığı M-tahmin edici ile parametre tahmini yapılmıştır. Parametre başlangıç değerleri için L1-norm regresyon parametre tahminleri (2.022004536, 3.114189142) alındıktan sonra aşağıda Huber t fonksiyonu ile M-tahmin edici için yapılan iteratif tekrar ağırlıklandırmalı en küçük kareler iterasyonu sonuçları verilmiştir.

Çizelge 6.8. Huber t fonksiyonunun kullanıldığı M-tahmin edici ile yapılan iterasyon sonuçları

k	θ_1	θ_2	S	$\Delta S/S$	$\Delta \theta/\theta$
0	2,022004536	3,114189142	0,205336689		
1	2,013980863	3,130522193	0,199545077	-0,028205444	0,00524472
2	2,013671652	3,130996068	0,199369306	-0,000880856	0,000153532
3	2,013660551	3,131035234	0,199362426	-3,45091 E-05	1,25090 E-05
4	2,013660149	3,131037425	0,199362156	-1,35622 E-06	6,99902 E-07
5	2,013660135	3,131037531	0,199362145	-5,30937 E-08	3,37564 E-08

Dördüncü iterasyonda parametre tahminlerine ulaşıldığı varsayılabilir.

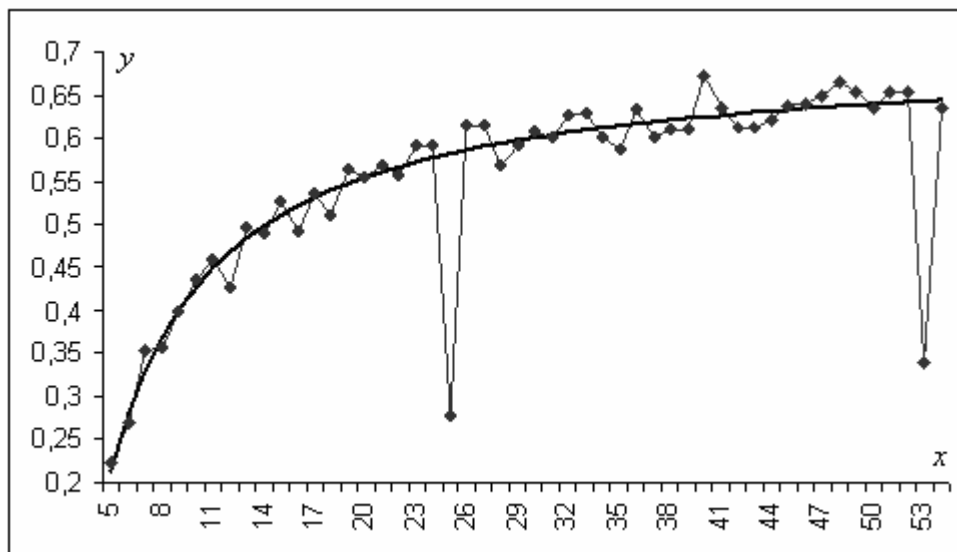
Huber t fonksiyonu yardımıyla bulunan tahminler toplu olarak aşağıda sunulmuştur.

$$s^2 = 0.000803552$$

$$\hat{\theta} = \begin{pmatrix} 2.013660135 \\ 3.131037531 \end{pmatrix}, \quad \theta = \begin{pmatrix} 2 \\ 3 \end{pmatrix}, \quad \hat{\theta} - \theta_* = \begin{pmatrix} 0.013660135 \\ 0.131037531 \end{pmatrix} \quad (6.12)$$

$$s(\hat{\theta}) = \begin{pmatrix} 0.015581229 \\ 0.160415368 \end{pmatrix}$$

Aşağıda veri ve uydurulan eğri birlikte çizilmiştir.



Şekil 6.6. Model 1b verisi için Huber t fonksiyonu kullanılarak M-tahmin edici aracılığı ile bulunan eğri

Huber t fonksiyonunun kullanımıyla elde edilen ağırlıklar aşağıdaki tabloda verilmiştir.

Çizelge 6.9. Huber t fonksiyonu kullanılarak M-tahmin edici ile elde edilen ağırlıklar

x	w	x	w	x	w
5	1	22	1	39	1
6	1	23	1	40	0,828912
7	1	24	1	41	1
8	1	25	0,129382	42	1
9	1	26	1	43	1
10	1	27	1	44	1
11	1	28	1	45	1
12	0,924402	29	1	46	1
13	1	30	1	47	1
14	1	31	1	48	1
15	1	32	1	49	1
16	1	33	1	50	1
17	1	34	1	51	1
18	1	35	1	52	1
19	1	36	1	53	0,128391
20	1	37	1	54	1
21	1	38	1		

25. ve 53. gözlemlere oldukça düşük ağırlıklar verilmiş olmakla birlikte Andrew dalga fonksiyonunun kullanımındaki benzer şekilde 12. ve 40. gözlemlere verilen düşük ağırlıklar dikkat çekicidir.

6.2.4. Model 1b süreci için simülasyon çalışması

Model 1b verisi oluşturma süreci 1000 kez tekrarlanarak herbiri 50 şer gözlemden oluşan 1000 örnek oluşturulmuştur. Elde edilen 1000 örneğin herbiri için en küçük kareler (EKK) yöntemi, Andrew dalga fonksiyonunun kullanıldığı M-tahmin edicisi (MA), Huber t fonksiyonunun kullanıldığı M-tahmin edicisi (MH) ve L1-norm regresyonu (L1) ile parametre tahminleri yapılmıştır. Bu yöntemler için $\bar{\hat{\theta}}$ parametre tahminleri ortalamaları ve $\bar{s}(\hat{\theta})$ standart hataları tahminleri ortalamaları aşağıdaki çizelgede sunulmuştur.

Çizelge 6.10. Model 1b süreci örneklerine uygulanan yöntemler ve sonuçları

Yöntem	Tekrar	$\bar{\hat{\theta}}_1$	$\bar{s}(\hat{\theta}_1)$	$\bar{\hat{\theta}}_2$	$\bar{s}(\hat{\theta}_2)$
EKK	1000	1.956184	0.032222	2.755891	0.336032
MA	1000	1.999505	0.009621	2.999370	0.098436
MH	1000	1.993663	0.015927	2.965328	0.163662
L1	1000	1.997093	0.000948	2.984300	0.009682

Aykırı değerler barındıran Model 1b verisinin en küçük kareler yöntemi ile bulunan parametre tahminlerinin yanlı olduğu görülür. Parametre tahminlerinin standart hataları tahminlerindeki artış dolayısıyla parametre tahminleri etkin değildir. M-tahmin edicileri için uygun başlangıç değerleri L1-norm regresyonu ile bulunabildiğinden bu yöntemle bulunan parametre tahminleri de dahil edilmiştir. L1-norm regresyonu tahminlerindeki sapmaların küçük olması dolayısıyla yansızlık, bu yaklaşımla bulunan parametre tahminlerinin standart hataları tahminlerinin en küçük olması dolayısıyla etkinlik koşulları sağlanır. Simülasyon tekrarının artmasına bağlı olarak M-tahmin edicilerin parametre tahminlerinin gerçek değerleri civarına yakınsadıkları görülür. Aykırı değerler barındıran örnekler için özellikle en küçük kareler yöntemi ile karşılaştırıldığında M-tahmin edicilerinin parametre tahminlerinin daha küçük varyanslı oldukları görülür. Huber t fonksiyonu ile karşılaştırıldığında mutlak değerce büyük artıklara Andrew dalga fonksiyonunun kullanıldığı M-tahmin edicisi çok daha küçük ve hatta sıfır ağırlıklar verdiğinden parametre tahminleri üzerindeki aykırı değer etkileri hemen hemen sıfırlanmaktadır. Bundan dolayı Huber t fonksiyonu ile karşılaştırıldığında Andrew dalga fonksiyonunun kullanıldığı M-tahmin edicisinin parametre tahminleri parametrelerin gerçek değerlerine daha çok yakınsadığı ve parametre tahminlerinin standart hata tahminlerinin daha küçük olduğu açıkça görülebilir.

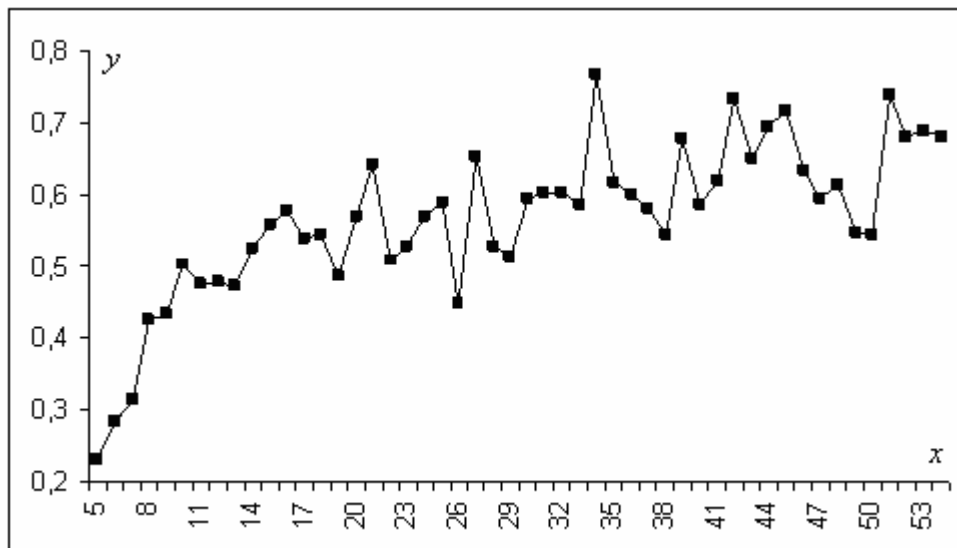
6.3. Model 1c Verisi

Bir uygulama olarak köşegen elemanlardan oluşan hata varyans-kovaryans matrisi aracılığı ile hatalar üretilip veri oluşturulmuş ve bu veri üzerinden parametre tahmini yapılmıştır. Bağımsız, beklenen değerleri sıfır ve varyansı $10^{-4} \times$ olan normal

dağılımdan hatalar üretilip gerçek parametre değerleri $\theta_1 = 2, \theta_2 = 3$ olan Model 1 çıktısına eklenerek Model 1c verisi elde edilmiştir. Bu şekilde üretilen Model 1c verisi aşağıda sunulmuş ve grafiği çizilmiştir.

Çizelge 6.11. Model 1 aracılığı ile üretilen Model 1c verisi

x	y	x	y	x	y
5	0,231665	22	0,507396	39	0,678113
6	0,282585	23	0,529021	40	0,587413
7	0,314460	24	0,570213	41	0,620234
8	0,425916	25	0,589240	42	0,733862
9	0,432059	26	0,447885	43	0,649994
10	0,503873	27	0,653050	44	0,695298
11	0,476142	28	0,528605	45	0,717103
12	0,476973	29	0,513542	46	0,633819
13	0,472769	30	0,594837	47	0,595681
14	0,525354	31	0,603346	48	0,614273
15	0,559449	32	0,602542	49	0,546608
16	0,577066	33	0,586645	50	0,544558
17	0,538230	34	0,767030	51	0,739011
18	0,544277	35	0,615308	52	0,680712
19	0,485957	36	0,601274	53	0,690244
20	0,570084	37	0,581470	54	0,679183
21	0,640339	38	0,544590		



Şekil 6.7. Model 1c verisinin diyagramı

6.3.1. Model 1c verisi ve genelleştirilmiş en küçük kareler

Doğrusallaştırılmış modelin parametre tahminleri olağan en küçük kareler yöntemiyle $\hat{\beta}_0 = 0.499284$, $\hat{\beta}_1 = 1.443873$ şeklinde bulunduktan sonra orijinal parametrelerin başlangıç değerleri $\theta_{10} = 2.002867$, $\theta_{20} = 2.891885$ olarak hesaplanır. Bu değerler genelleştirilmiş en küçük kareler yöntemi için kullanılan Gauss-Newton iterasyonu başlangıç değeri olarak alınmıştır.

Hata varyansları başlangıçtan bilindiği varsayılarak hata varyansları $10^{-4}x$ olarak alınmıştır. Buna göre aşağıda Gauss-Newton iterasyonu sonuçları verilmiştir.

Çizelge 6.12. Model 1c verisi için Gauss-Newton iterasyonu sonuçları

k	θ_1	θ_2	S	$\Delta S/S$	$\Delta \theta/\theta$
0	2,002866871	2,891884812	55,62176886		
1	2,008335168	2,916235800	55,56684784	-0,000987402	0,008420456
2	2,008328444	2,916144520	55,56684747	-6,57178E-09	3,13006E-05
3	2,008328505	2,916145018	55,56684747	-1,27232E-13	1,70688E-07
4	2,008328505	2,916145015	55,56684747	0	9,29672E-10

Üçüncü iterasyonda parametre tahminlerine varıldığı görülür. Hata varyanslarının başlangıçtan bilindiği varsayılarak hata varyansları $10^{-4}x$ olarak alınmıştır. s^2 varyans tahmini, σ^2 teorik varyans, $\hat{\theta}$ parametre tahminleri, θ_* parametrelerinin gerçek değeri, $\hat{\theta} - \theta_*$ parametre tahminleri ile gerçek değerleri arasındaki farkları, $s(\hat{\theta})$ parametre tahminlerinin standart hata tahminleri, parametrelerin ve teorik varyansın gerçek değerleri bilindiğinden $\sigma(\hat{\theta})$ parametre tahminlerinin teorik standart hataları sırası ile aşağıda verilmiştir.

$$\begin{aligned}
 s^2 &= 1.157642657, \quad \sigma^2 = 1 \\
 \hat{\theta} &= \begin{pmatrix} 2.008328505 \\ 2.916145015 \end{pmatrix}, \quad \theta_* = \begin{pmatrix} 2 \\ 3 \end{pmatrix}, \quad \hat{\theta} - \theta_* = \begin{pmatrix} 0.008328505 \\ -0.083854985 \end{pmatrix} \\
 s(\hat{\theta}) &= \begin{pmatrix} 0.028272017 \\ 0.200628274 \end{pmatrix}, \quad \sigma(\hat{\theta}) = \begin{pmatrix} 0.026268031 \\ 0.188660686 \end{pmatrix}
 \end{aligned} \tag{6.13}$$

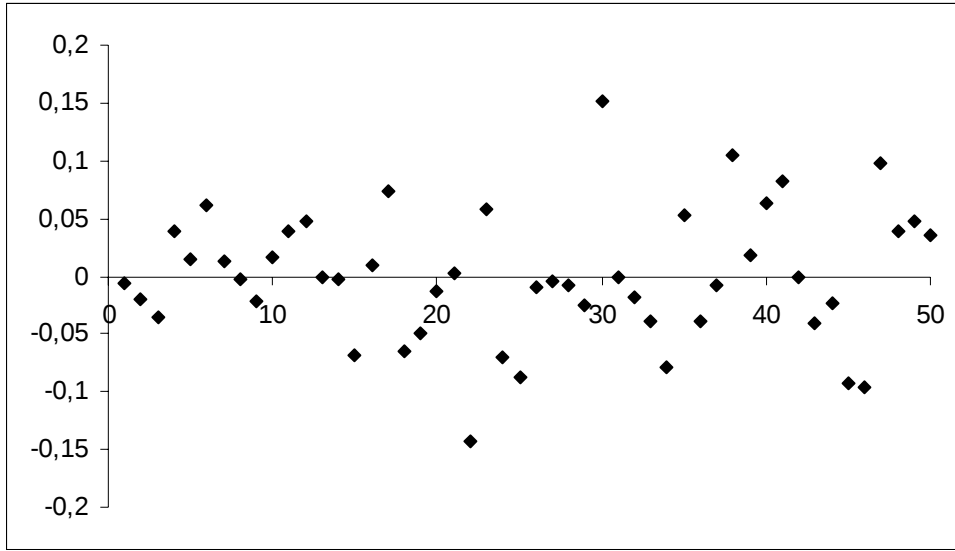
Hata varyanslarının $10^{-4}x$ olduğu bilinerek hatalar kendi standart sapmalarına bölündüğünde hatalar standart normal dağılımlı hatalara dönüştüğünden dolayı bir kolaylık için teorik varyans bir olarak ölçeklenir. Bu durumda $\sigma(\hat{\theta})$ parametre tahminlerinin teorik standart hataları, θ_* parametrelerin gerçek değerleri ve σ^2 teorik hata varyansı ile bulunur. $\hat{\theta}$ parametre tahminleri kullanılarak elde edilen model artıklarının bilinen hata varyansları aracılığı ile ağırlıklandırılarak bulunan s^2 hata varyansı tahmini ile $\hat{\theta}$ parametre tahminleri aracılığı ile $s(\hat{\theta})$ parametre tahminlerinin standart hata tahminleri bu yol ile elde edilmiştir.

Genelleştirilmiş en küçük kareler yöntemin şimdiki uygulaması için verilen hata varyans-kovaryans matrisinin bilinmediği varsayılarak parametre tahminleri yapılmıştır. Hata korelasyonlarının sıfır olduğu varsayılarak hata varyansları artıklar üzerinden tahmin edilmeye çalışılmıştır. İlk önce ağırlıklandırma yapılmadan (tüm ağırlıkların bir olduğu varsayılarak) parametre tahmini yapılmıştır. Yukarıda doğrusallaştırılmış model parametre tahminleri aracılığı ile elde edilen $\theta_{10} = 2.002867$, $\theta_{20} = 2.891885$ parametre başlangıç değerleri en küçük kareler yöntemine uygulanan Gauss-Newton iterasyonları için kullanılmıştır. İterasyonlar aşağıda sunulmuştur.

Çizelge 6.13. Gauss-Newton iterasyonları

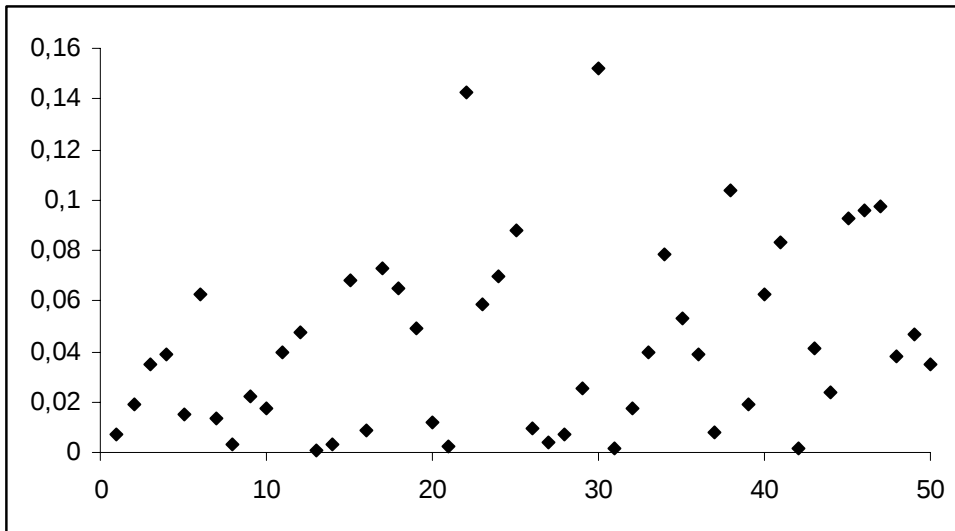
k	θ_1	θ_2	S	$\Delta S/S$	$\Delta \theta/\theta$
0	2,002867	2,891885	0,158332		
1	2,007147	2,905908	0,158203	-0,00081	0,004849
2	2,007153	2,905930	0,158203	-1,4E-09	7,47E-06
3	2,007153	2,905930	0,158203	0	5,17E-09
4	2,007153	2,905930	0,158203	0	3,57E-12
5	2,007153	2,905930	0,158203	0	2,22E-15

Parametre tahminlerine üçüncü iterasyonda ulaşılmıştır. Elde edilen artıkların saçılma grafiği aşağıda verilmiştir.



Şekil 6.8. Ağırlıklandırma yapılmadan elde edilen artıklar için serpmeye diyagramı

Genel olarak artıkların saçılma şekli varyansın kare kökü olan standart sapmaya karşılık gelir; bu şekilde oransal bir bağ kurulabilir. Standart sapmalarla iyi bir karşılaştırma kurabilmek için artıkların mutlak değerleri gözlem sırasına konup saçılma grafiği çizilmiştir.

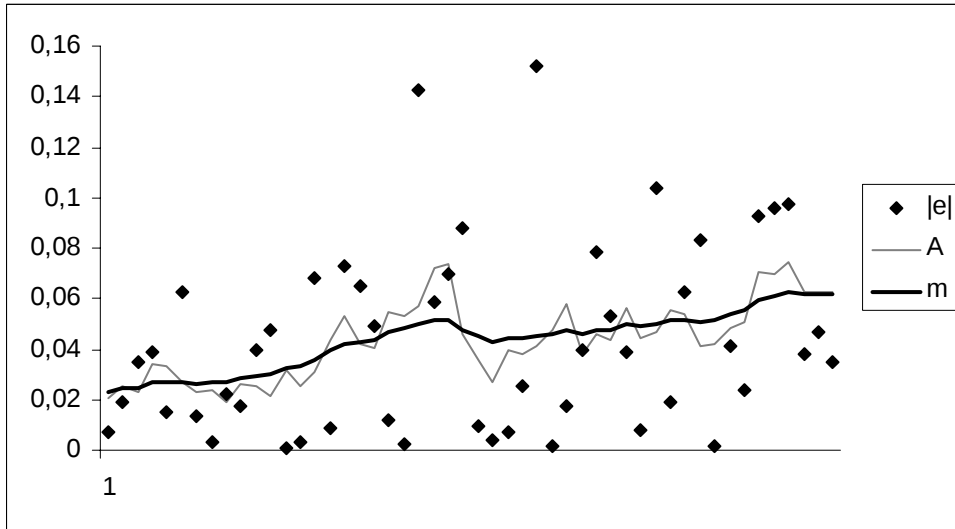


Şekil 6.9. Artıkların mutlak değerleri saçılma grafiği

Standart sapmaların gözlem seviyesine bağlı olarak sürekli bir fonksiyon olduğu ve yavaşça değiştiği varsayılmıştır. Aksi halde standart sapmanın ani noktasal değişimi

aykırı değerlerin belirmesine neden olabilir. Aykırı değerleri belirleyen bir yaklaşım sağlam regresyon yaklaşımı olabilir. Standart sapmanın bölgesel olarak yavaşça değişimi bölgesel standart sapmalarının farklılıklarının gözlenmesiyle belirlenebilir. Bölgesel farklılıkların gözlenebilmesi için artıkların mutlak değerlerinin beş birimlik merkezi hareketli ortalamaları alınıp grafiğe eklenmiştir: $A_t = \frac{1}{5}(|e_{t-2}| + |e_{t-1}| + |e_t| + |e_{t+1}| + |e_{t+2}|)$. Başlangıç ve bitiş taraflarında merkezi hareketli ortalamalar hesaplanamayacağından eldeki artıklar kullanılarak merkezi olmayan ve daralan hareketli ortalamalar hesaplanmıştır: $A_1 = \frac{1}{3}(|e_1| + |e_2| + |e_3|)$, $A_2 = \frac{1}{4}(|e_1| + |e_2| + |e_3| + |e_4|)$ gibi. Eğer gözlemler eşit aralıkta toplanmamış ise çekirdek (kernel) düzeltme teknikleri kullanılabilir. Bir diğer yaklaşım beş birimlik merkezleştirilmiş medyanları bulmak olabilir: $A_t = \text{medyan}(|e_{t-2}|, |e_{t-1}|, |e_t|, |e_{t+1}|, |e_{t+2}|)$. Eğer veri içinde aykırı değerler bulunuyorsa medyan yaklaşımı daha avantajlıdır. Aykırı değerlerin belirlenmesi amacıyla medyanın ve sağlam regresyonun kullanımına yönelik Fried, Einbeck ve Gather (2007) tarafından yazılan makaleye bakılabilir. Aykırı değerlerin eklenmediği simülasyon çalışmalarında medyan ile hareketli ortalamanın karşılaştırmalarında hareketli ortalamaların daha durağan olduğu görülmüştür. Hareket aralığının beş birim olarak seçilmesinin nedeni bölgeyi temsil edecek şekilde yeterince gözlemi dikkate alması ve merkezileştirme hesabının kolay olması içindir. Ayrıca bir bölgede standart sapmanın pek değişmediği varsayımında hatalar bir dağılımdan gelirken dağılım örneğinin olasılık olarak iki eşit bölgeye ayrıldığında tesadüfen beş hatanın hepsinin dağılımın bir yarısından gelme olasılığı $(\frac{1}{2})^5 = \frac{1}{32}$ olarak bulunur. Bu olasılık bir tıkanıklık olasılığının azlığına işaret eder. En azından merkezi hareketli ortalamaları veya medyanı etkileyecek hatalı dalgalanmalar bu şekilde önlenmiş olur. Yapılan simülasyon çalışmaları, mutlak artıkların merkezi hareketli ortalamalarının yeterli bir durağanlığı vermediğini göstermektedir. Bunun üzerine merkezi hareketli ortalama serisi üzerine parametre ile ayarlamalara imkan veren üstel düzeltme tekniği uygulanmıştır: $S_t = \alpha A_t + (1 - \alpha)S_{t-1}$. Tıp alanında risk faktörlerinin kontrolü için yavaş değişen dinamik süreçlerden elde edilen göstergelere uygulanan üstel hareketli ortalamalarının konu edildiği Grigg ve Spiegelhalter (2007) tarafından yayınlanan bir

makaleye bakılabilir. Ancak üstel düzeltme hesaplandığı seriyi geriden takip ettiğinden bu açıdan iyi bir gösterge değildir. Bunun üzerine geriye doğru üstel düzeltme serisi de hesaplanmıştır: $R_t = \alpha A_t + (1 - \alpha)R_{t+1}$. İleri üstel düzeltme için $S_1 = A_1$ ile A_t serisinin ilk değeri, geri üstel düzeltme için $R_T = A_T$ ile A_t serisinin son değeri alınmıştır. Dengeleme için $m_t = \frac{1}{2}(S_t + R_t)$ ortalama çizgisi bulunmuştur. Mutlak artıklar gözönünde bulundurularak α parametresi ayarlanarak m_t serisinin uygunluğu kontrol edilebilir. Simülasyon çalışmaları $0.1 < \alpha < 0.2$ için yeterli bir uygunluğun sağlandığını gösterir. Bu şekilde m_t serisi standart sapmaların oransal büyüklükleri arasında doğrudan bağlantı kurar. Bu durumda m_t serisi elemanlarının K matrisinin köşegen elemanları olarak kullanılması uygundur. Mutlak artıklar üzerine A_t merkezi hareketli ortalamalarının ve $\alpha = 0.2$ için m_t serisinin eklendiği grafik aşağıda çizilmiştir.



Şekil 6.10. Mutlak artıklar, A_t merkezi hareketli ortalamaları ve m_t serisi

Buna göre K matrisinin köşegen elemanları m_t serisi olarak seçilirse tahmini standart sapmaların oransal büyüklükleri aşağıdaki tablodadır.

Çizelge 6.14. i. gözlem düzeyine bağlı olarak K matrisinin i. köşegen elemanı

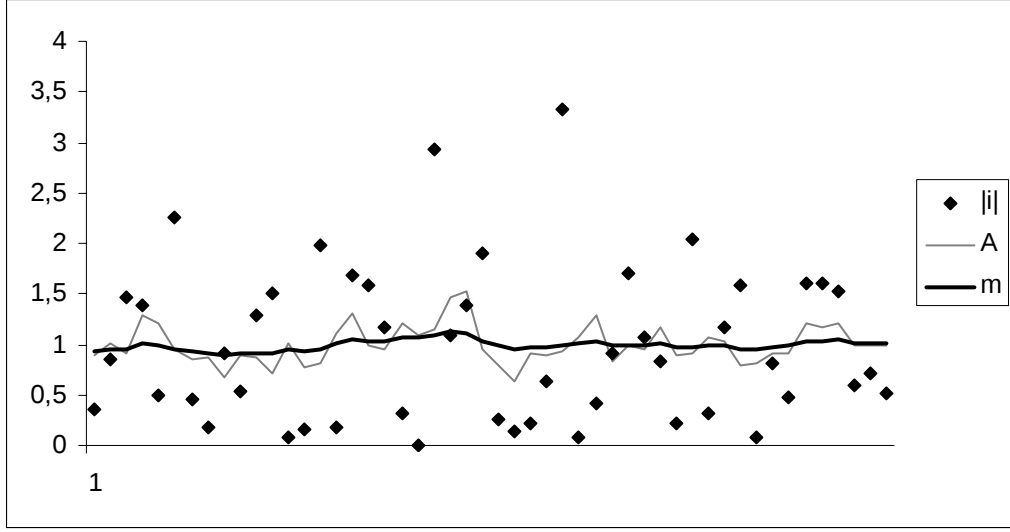
i	mi	i	mi	i	mi
1	0,023256	18	0,042531	35	0,047570
2	0,024447	19	0,043884	36	0,049599
3	0,024943	20	0,046936	37	0,048795
4	0,026839	21	0,048093	38	0,049581
5	0,027156	22	0,049552	39	0,051211
6	0,026640	23	0,051731	40	0,051637
7	0,026430	24	0,051516	41	0,050814
8	0,026757	25	0,047276	42	0,051822
9	0,026747	26	0,044869	43	0,053802
10	0,028231	27	0,043082	44	0,055710
11	0,029121	28	0,044278	45	0,059525
12	0,029850	29	0,044255	46	0,060835
13	0,032407	30	0,044975	47	0,062227
14	0,033356	31	0,046263	48	0,061401
15	0,035825	32	0,047900	49	0,061662
16	0,039316	33	0,045997	50	0,061871
17	0,042213	34	0,047218		

K matrisi sabit olarak alındığında ve parametre başlangıç değerleri ile birlikte ağırlıklandırılmış en küçük kareler yöntemi için hesaplanan iterasyonlar aşağıdaki çizelgede verilmiştir. Doğrusallaştırılmış modelin parametre tahminlerinden elde edilen değerler orijinal modelin parametrelerine çevrilerek $\theta_{10} = 2.002867$, $\theta_{20} = 2.891885$ parametre başlangıç değerleri yukarıda bulunmuştur.

Çizelge 6.15. Gauss-Newton iterasyonu sonuçları

k	θ_1	θ_2	S	$\Delta S/S$	$\Delta \theta \theta$
0	2,002866871	2,891884812	77,28750391		
1	2,012521667	2,916260712	76,86410454	-0,005478238	0,008429070
2	2,012531629	2,916166542	76,86410059	-5,14081E-08	3,22912E-05
3	2,012531688	2,916167061	76,86410059	-1,19571E-13	1,77843E-07
4	2,012531688	2,916167058	76,86410059	0	9,78305E-10
5	2,012531688	2,916167058	76,86410059	0	5,38170E-12

Üçüncü iterasyonda parametre tahminlerine ulaşıldığı sonucu çıkarılabilir. Parametrelerin genelleştirilmiş en küçük kareler tahmini bulunduktan sonra mutlak ağırlıklandırılmış artıkları, mutlak ağırlıklandırılmış artıklar üzerine kurulan A_t merkezi hareketli ortalamaları ve $\alpha = 0.2$ için m_t serisi ortak grafiğinden görüldüğü gibi ağırlıklandırılmış artıkların varyans durağanlığı sağlanmıştır.



Şekil 6.11. Mutlak ağırlıklandırılmış artıklar, A_t merkezi hareketli ortalamaları ve m_t serisi

Hata varyanslarının tahmini ile ağırlıklandırılmış en küçük kareler yöntemiyle parametre tahminleri bulunmakla birlikte diğer tahminler ve teorik değerler aşağıda sunulmuştur.

$$\begin{aligned}
 s^2 &= 1.601335429, \quad \sigma^2 = 1 \\
 \hat{\theta} &= \begin{pmatrix} 2.012531688 \\ 2.916167058 \end{pmatrix}, \quad \theta_* = \begin{pmatrix} 2 \\ 3 \end{pmatrix}, \quad \hat{\theta} - \theta_* = \begin{pmatrix} 0.012531688 \\ -0.083832942 \end{pmatrix} \\
 s(\hat{\theta}) &= \begin{pmatrix} 0.028443121 \\ 0.217031255 \end{pmatrix}, \quad \sigma(\hat{\theta}) = \begin{pmatrix} 0.026268031 \\ 0.188660686 \end{pmatrix}
 \end{aligned} \tag{6.14}$$

s^2 varyans tahmini hata varyanslarının tahminleri dolayısıyla ağırlıklardan etkilendiğinden σ^2 varyansından oldukça farklı bulunmuştur. Buradaki süreç için beklenen değerleri sıfır ve varyansı $10^{-4}x$ olan normal dağılımdan hatalar üretildiğinden teorik değerlerin hesabı için $1/(10^{-4}x)$ ağırlıkları doğrudan kullanılırsa hatalar standart normal dağılımlı hatalara dönüşerek ölçeklenmiş varyans $\sigma^2 = 1$ olarak bulunur. Buradan $\sigma(\hat{\theta})$ parametre tahminlerinin teorik standart hataları θ_* parametrelerin gerçek değerleri ve σ^2 teorik hata varyansı ile

bulunur. Buradan ayrı olarak $\hat{\theta}$ parametre tahminleri kullanılarak elde edilen model artıkları, tahmini yapılan hata varyansları aracılığı ile ağırlıklandırılarak s^2 hata varyansı tahmini bulunmuş ve $s(\hat{\theta})$ parametre tahminlerinin standart hata tahminleri hesabında parametre tahminleri, ağırlık tahminleri ve hata varyansı tahmini kullanılmıştır.

6.3.2. Model 1c verisi ve yarı olabilirlik yöntemi

Yarı olabilirlik yöntemi genelleştirilmiş en küçük kareler yöntemine benzemekle birlikte parametre tahmininin geliştirildiği her iterasyon içinde hataların varyansları tekrar tahmin edilir.

Her iterasyonda $S(\theta)$ ağırlıklandırılmış kareler toplamı fonksiyonunun izafi olarak değerlerini koruyabilmesi için hesaplanan ağırlıkların bir tür normalizasyonu bulunmuştur. Burada bu işlem şu şekilde yapılmıştır. Eğer ağırlıklandırma kullanılmıyor olsaydı 50 gözlemden oluşan verinin her gözlemine bir birimlik ağırlık verilmiş olurdu. Ağırlıklandırılmış en küçük kareler yönteminde kullanılan artıkların mutlak değerlerinin düzeltilmesi yaklaşımındaki m_t serisinin elemanları toplamı buradaki uygulama için 50 olacak şekilde m_t serisinin elemanları sabit bir sayı ile çarpılmıştır. Bu şekilde ağırlığı birden fazla olan veya birden az olan gözlemler kolayca ayırt edilebilir. Her iterasyonda m_t serisinin elemanları bu şekilde ayarlanarak V köşegen matrisinin kare kökü olan K köşegen matrisinin elemanları m_t serisinin elemanları olarak alınmıştır. Bu değerler standart sapma tahminleri olmayıp bu büyüklüklerin bir katıdır.

$$\{m_t\}_{t=5}^{54} \rightarrow \sum_t cm_t = 50 \quad , \quad c \in R^+ \quad , \quad K = \text{diag}\{cm_t\}_{t=5}^{54} \quad (6.15)$$

Bu şekilde her iterasyonda aşağıdaki fark denkleminde parametre çıktıları elde edilir.

$$\begin{aligned}\theta_{k+1} &= \theta_k + \left(Z_k^t V_k^{-1} Z_k\right)^{-1} Z_k^t V_k^{-1} [y - f(\theta_k)] \\ \lim_{k \rightarrow \infty} \theta_k &= \hat{\theta} \quad , \quad \lim_{k \rightarrow \infty} V_k = \hat{V}\end{aligned}\tag{6.16}$$

Buna bağılı olarak iterasyonlar şu şekilde yerine getirilir. $k=0$ iterasyonu için parametre başlangıç değerleri belirlenmiş olmalıdır. En başta θ_0 başlangıç değerleri kullanılarak $y - f(\theta_0)$ artıkları elde edilir. Bu artıkların mutlak değerlerinin düzgünleştirilmesi işlemiyle K_0 matrisi ve $V_0 = K_0^2$ varyans matrisi elde edilir. Yukarıdaki $\theta_1 = \theta_0 + \left(Z_0^t V_0^{-1} Z_0\right)^{-1} Z_0^t V_0^{-1} [y - f(\theta_0)]$ iterasyonu ile θ_1 çıktısı bulunur. θ_1 kullanılarak $y - f(\theta_1)$ artıkları elde edilip artıklarının mutlak değerlerinin düzgünleştirilmesiyle K_1 matrisi ve $V_1 = K_1^2$ varyans matrisi bulunur. Bulunanların yardımıyla θ_2 çıktısı bulunur. Bu şekilde iterasyonlar devam ettirilerek parametrelerin tahminlerine ulaşılır.

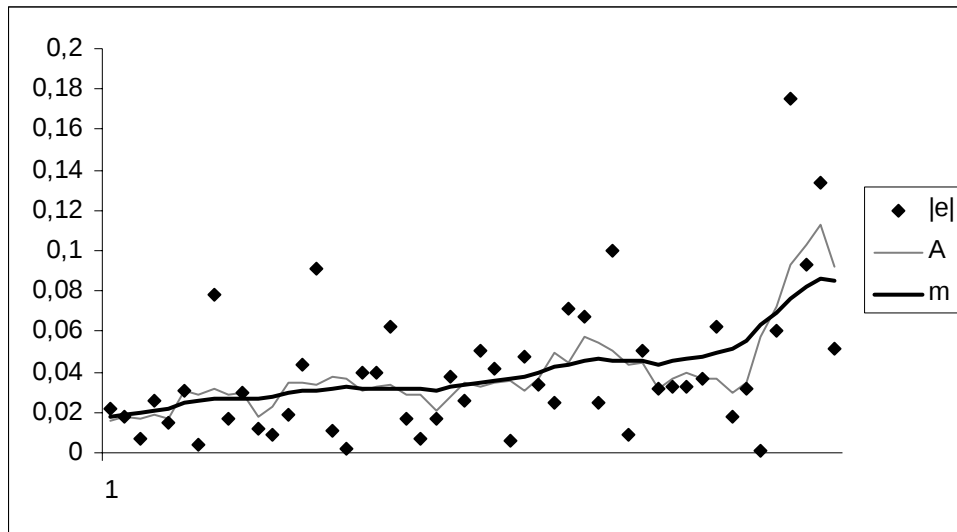
Genelleştirilmiş en küçük kareler yöntemine uygulanan Model 1 aracılığı ile üretilen Model 1c verisi örneği burada tekrar ele alınmıştır. $k=0$ iterasyonu için parametre başlangıç değerleri olarak genelleştirilmiş en küçük kareler yönteminde kullanılan $\theta_{10} = 2.002867$, $\theta_{20} = 2.891885$ başlangıç değerleri alınmıştır. Yarı olabilirlik yaklaşımına göre yapılan iterasyonlar aşağıda verilmiştir.

Çizelge 6.16. Model 1c verisi için yarı olabilirlik yöntemi iterasyonları

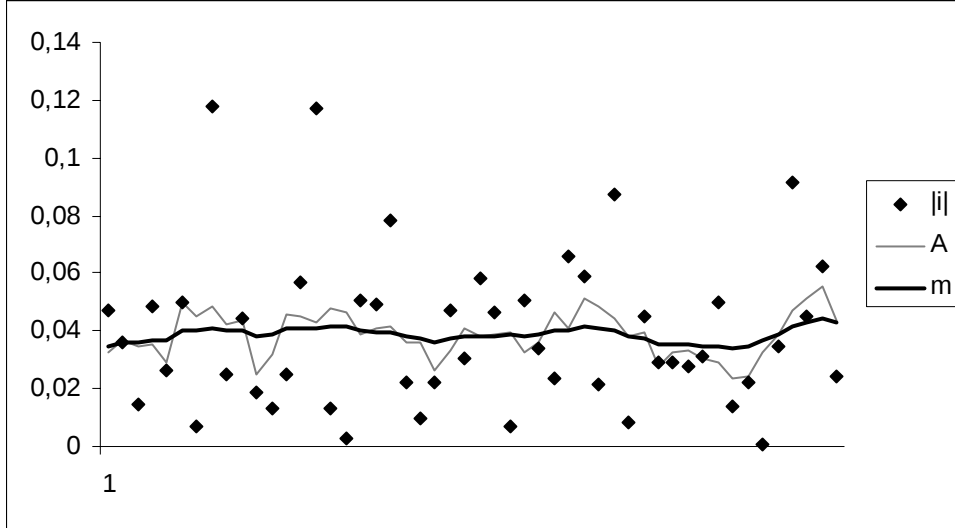
k	θ_1	θ_2	S	$\Delta S/S$	$\Delta \theta/\theta$
0	2,002866871	2,891884812	0,142516473		
1	2,013891677	2,927014599	0,140780233	-0,012182730	0,012147713
2	2,016217320	2,934671499	0,140577282	-0,001441619	0,002615942
3	2,016820909	2,937034782	0,140537736	-0,000281309	0,000805297
4	2,016986103	2,937802995	0,140528298	-6,71599E-05	0,000261561
5	2,017033907	2,938061479	0,140525823	-1,76120E-05	8,79856E-05
6	2,017048505	2,938150568	0,140525129	-4,93338E-06	3,03223E-05
7	2,017053176	2,938181780	0,140524924	-1,46505E-06	1,06231E-05
8	2,017054728	2,938192835	0,140524859	-4,58170E-07	3,76232E-06
9	2,017055258	2,938196778	0,140524838	-1,49523E-07	1,34190E-06
10	2,017055442	2,938198190	0,140524831	-5,04096E-08	4,80774E-07
11	2,017055507	2,938198698	0,140524829	-1,73957E-08	1,72742E-07
12	2,017055530	2,938198880	0,140524828	-6,09954E-09	6,21778E-08
13	2,017055539	2,938198946	0,140524828	-2,16142E-09	2,24058E-08
14	2,017055542	2,938198970	0,140524828	-7,71178E-10	8,07964E-09
15	2,017055543	2,938198979	0,140524827	-2,76357E-10	2,91484E-09

Onuncu iterasyonda parametre tahminlerine ulaşıldığı varsayılabilir.

Yarı olabilirlik tahminine göre son iterasyon model artıkları mutlak değeri, mutlak değer üzerinden hesaplanan A_t merkezi hareketli ortalamaları ve $\alpha = 0.2$ için m_t serisi ortak grafiği aşağıda verilmiştir.

Şekil 6.12. Artıkların mutlak değeri, A_t merkezi hareketli ortalamaları ve m_t serisi

Ayrıca ağırlıklandırılmış artıkların mutlak değerleri üzerinden hesaplanan A_t merkezi hareketli ortalamaları ve $\alpha = 0.2$ için m_t serisi ortak grafiği aşağıdadır.



Şekil 6.13. Ağırlıklandırılmış artıkların mutlak değerleri grafiği

İterasyonlar sonucu ağırlıklandırılmış artıklar sabit varyans koşulunu sağlar.

Yarı olabilirlik tahmini ile parametre tahminleri bulunmakla birlikte diğer tahminler ve teorik değerler aşağıda sunulmuştur.

$$\begin{aligned}
 s^2 &= 0.002927601, & \sigma^2 &= 1 \\
 \hat{\theta} &= \begin{pmatrix} 2.017055543 \\ 2.938198979 \end{pmatrix}, & \theta_* &= \begin{pmatrix} 2 \\ 3 \end{pmatrix}, & \hat{\theta} - \theta_* &= \begin{pmatrix} 0.017055543 \\ -0.061801021 \end{pmatrix} \\
 s(\hat{\theta}) &= \begin{pmatrix} 0.028208300 \\ 0.217300002 \end{pmatrix}, & \sigma(\hat{\theta}) &= \begin{pmatrix} 0.026268031 \\ 0.188660686 \end{pmatrix}
 \end{aligned} \tag{6.17}$$

Hata varyansları ile ilgili oransal tahminlerin bir sabitle çarpılması ve bu tahminlerin dengelenmesi dolayısıyla s^2 varyans tahmini σ^2 varyansından farklı bulunur. Hata varyansları tahminleri doğrudan yapılmamıştır, bunun yerine hata standart sapmaları tahminlerinin bir oranı m_t serisi aracılığı ile tarif edilmiştir. Beklenen değerleri sıfır ve varyansı $10^{-4}x$ olan normal dağılımdan hatalar üretilmiş

olduğundan bu bilgi kullanılarak $1/(10^{-4}x)$ ağırlıkları doğrudan kullanıldığında hatalar standart normal dağılımlı hatalara dönüşerek ölçeklenmiş varyans $\sigma^2 = 1$ olarak bulunur.

Yukarıdaki yarı olabilirlik tahmini için yapılan iterasyonlarda V hata varyans matrisinin oransal bir büyüklüğü kullanılmaktadır. Hata varyanslarının karekökü olan standart sapmalarının büyüklükleri ile hataların mutlak değerlerinin düzgünleştirilmiş büyüklükleri arasında bir oran bulunabilir. Hatalar normal dağılıyorsa hataların mutlak değerlerinin ortalaması

$$m = 2 \int_{x=0}^{\infty} \frac{x}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2} dx = \sqrt{\frac{2}{\pi}}\sigma \rightarrow m = \sqrt{\frac{2}{\pi}}\sigma \quad (6.18)$$

şeklinde bulunur. Artıkların mutlak değerlerinin düzgünleştirilmesiyle bir m_i serisi bulunmuş ise bu değerler hataların mutlak değerlerinin bölgesel ortalamasının bir yaklaşıklığı olarak kabul edilebilir. Böylece standart sapmalar ve varyanslar

$$s_i = \sqrt{\frac{\pi}{2}}m_i, \quad s_i^2 = \frac{\pi}{2}m_i^2 \quad (6.19)$$

şeklinde tahmin edilir. Hatalar normal dağılmamakla birlikte normal dağılıma yakın dağılıyorsa buradaki sonuçlar yaklaşık olarak alınabilir.

Değinilmesi gereken önemli bir nokta mevcuttur. Ağırlıklandırılmış en küçük kareler yaklaşımında V hata varyans-kovaryans matrisi yerine k skalar pozitif bir sayı olmak üzere kV matrisi kullanıldığında s^2 varyans tahmini bu değişiklikten doğrudan etkilenir.

$$\begin{cases} s^2 = \frac{1}{n-p} [y - f(\hat{\theta})]^t V^{-1} [y - f(\hat{\theta})] \\ \hat{V}(\hat{\theta}) = s^2 (Z(\hat{\theta})^t V^{-1} Z(\hat{\theta}))^{-1} \end{cases} \quad (6.20)$$

$$V \rightarrow kV$$

$$\begin{aligned} s_1^2 &= \frac{1}{n-p} [y - f(\hat{\theta})]^t (kV)^{-1} [y - f(\hat{\theta})] = \frac{k^{-1}}{n-p} [y - f(\hat{\theta})]^t V^{-1} [y - f(\hat{\theta})] \\ &= k^{-1} s^2 \end{aligned} \quad (6.21)$$

$$\begin{aligned} \hat{V}_1(\hat{\theta}) &= s_1^2 (Z(\hat{\theta})^t (kV)^{-1} Z(\hat{\theta}))^{-1} = k s^2 (Z(\hat{\theta})^t V^{-1} Z(\hat{\theta}))^{-1} = s^2 (Z(\hat{\theta})^t V^{-1} Z(\hat{\theta}))^{-1} \\ &= \hat{V}(\hat{\theta}) \end{aligned}$$

Ancak V matrisi yerine bu matrisin bir katı veya bir oranı kullanılırsa parametre tahminlerinin $\hat{V}(\hat{\theta})$ varyans-kovaryans matrisi tahmini değişmez. V matrisinin yerine kullanılan kV matrisinin elemanları arasındaki oranların korunuyor olması önemlidir.

6.3.3. Model 1c süreci için simülasyon çalışması

Simülasyon aracılığı ile Model 1c sürecinden 1000 örnek üretilmiş ve her bir örnek verisi ile model parametreleri ayrı ayrı tahmin edilmiştir. İlk önce hata varyanslarının başlangıçta bilinmesi durumunda parametre tahminleri ağırlıklandırılmış en küçük kareler yöntemi uygulanarak bulunmuştur; bu tahminler AEKK0 olarak gösterilmiştir. Sonraki uygulamada hata varyanslarının bilinmediği varsayılarak hata varyansları tahmini yapıldıktan sonra parametre tahminleri ağırlıklandırılmış en küçük kareler (AEKK) yöntemi ile yapılmıştır. Bir ara aşama olarak kullanılan en küçük kareler (EKK) yöntemi ile bulunan parametre tahminleri ek olarak verilmiştir. Parametre tahminlerinin yarı olabilirlik tahminleri YO olarak gösterilmiştir. Simülasyon sonuçları aşağıdaki çizelgede sunulmuştur.

Çizelge 6.17. Model 1c süreci örneklerine uygulanan yöntemler ve sonuçları

Yöntem	Tekrar	$\hat{\theta}_1$	$\bar{s}(\hat{\theta}_1)$	$\hat{\theta}_2$	$\bar{s}(\hat{\theta}_2)$
EKK	1000	1.999007	0.028874	2.997356	0.300140
AEKK0	1000	1.998818	0.026122	2.994613	0.187630
AEKK	1000	1.998940	0.024103	2.997902	0.185377
YO	1000	1.999435	0.024324	3.001155	0.182423
L1	1000	2,000516	0.002798	3,005918	0.023926

Hata varyanslarının farklı olduğu veri üzerinden bulunan en küçük kareler parametre tahminleri, standart hataları tahminlerinin büyük olması dolayısıyla etkin tahmin ediciler değildir. Simülasyon tekrarı arttıkça ağırlıklandırılmış en küçük kareler yöntemi ile karşılaştırıldığında parametrelerin yarı olabilirlik tahminlerinin gerçek parametre değerlerine biraz daha yakınsadığı görülebilir. Simülasyon sonuçlarına göre diğer tahmin yöntemleri ile karşılaştırıldığında L1-norm parametre tahminlerinin standart hataları tahminleri en küçük bulunduğundan bu yöntemle bulunan parametre tahminleri etkin tahmin edicilerdir.

Özel olarak yarı olabilirlik yöntemi ile genelleştirilmiş en küçük kareler yöntemi karşılaştırıldığında işlem azlığı dolayısıyla genelleştirilmiş en küçük kareler yönteminin kullanımı tercih edilebilir. Her iterasyonda (simülasyon tekrarından ayrı olarak) hata varyanslarının tahmininin yapılması ek bir üstünlük getirmemektedir. Yarı olabilirlik yönteminde hata varyanslarının tahmininin ilk bir-iki iterasyonda durağanlaştığı, ancak daha sonraki iterasyonlarda hata varyansları tahminindeki çok küçük değişimlerin parametre tahminlerinin durağanlaşmasını geciktirdiği yapılan denemelerden görülmüştür. Bundan dolayı yarı olabilirlik yönteminde parametre tahminlerinin yakınsayabilmesi için daha fazla iterasyonun yapılması gerekmektedir.

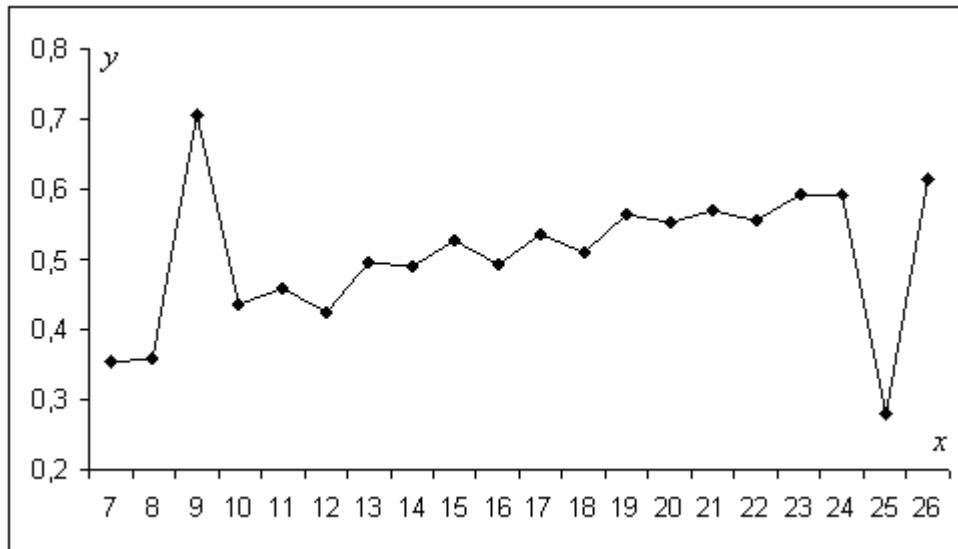
6.4. Model 1d Verisi

Parametre değerleri $\theta_1 = 2, \theta_2 = 3$ olan Model 1 aracılığı ile bağımsız x değişkeninin 7-26 aralığında birer artarak oluşturulan 20 gözlemlik çıktısına beklenen değerleri sıfır ve varyansı $(0.02)^2 = 0.0004$ olan normal dağılımdan

bağımsız hatalar üretilerek eklenmiş olmakla birlikte 9. ve 25. gözlemlerinin aykırı değerler olabilmesi için normal hatalar yerine sırasıyla büyük mutlak değerli 0.3 ve -0.3 hataları konularak Model 1d verisi oluşturulmuştur. Bu şekilde üretilen Model 1d verisi ve grafiği aşağıda gösterilmiştir.

Çizelge 6.18. İki aykırı değer barındıran 20 gözlemlili Model 1d verisi

x	y	x	y
7	0,352889	17	0,535877
8	0,358179	18	0,509804
9	0,705465	19	0,562969
10	0,435260	20	0,553430
11	0,459469	21	0,568967
12	0,425824	22	0,556189
13	0,495906	23	0,592260
14	0,489976	24	0,592054
15	0,525646	25	0,279818
16	0,491969	26	0,614315



Şekil 6.14. Model 1d verisinin diyagramı

6.4.1. Model 1d verisi ve en küçük kareler yöntemi

Doğrusallaştırılan model parametreleri kullanılarak elde edilen parametre tahminleri orijinal parametre değerlerine çevrilerek (1.817135704, 1.44694454) başlangıç değerleri bulunmuştur. Orijinal veri üzerinden L1-norm regresyonu için bu başlangıç

değerleri kullanılmış ve (2.000109755, 2.89638538) parametre tahminleri elde edilmiştir. L1-norm regresyonu parametre tahminleri en küçük kareler yöntemi için başlangıç değerleri olarak kullanılmış ve aşağıdaki iterasyon sonuçları elde edilmiştir:

Çizelge 6.19. Model 1d verisi için en küçük kareler iterasyonu sonuçları

k	θ_1	θ_2	S	$\Delta S/S$	$\Delta \theta/\theta$
0	2,000109755	2,896385380	0,184714410		
1	1,796445561	1,200281142	0,163788738	-0,113286624	0,585593426
2	1,817580411	1,385612130	0,163390924	-0,002428823	0,154406315
3	1,817582904	1,385806276	0,163390922	-1,70261E-08	0,000140116
4	1,817582620	1,385803797	0,163390922	-4,12115E-13	1,78906E-06
5	1,817582623	1,385803828	0,163390922	0	2,28723E-08

Beşinci iterasyonda parametre tahminleri durağanlaşır. En küçük kareler parametre tahmini ve standart hata tahmini aşağıdaki gibi verilir:

$$\hat{\theta} = \begin{pmatrix} 1.817583 \\ 1.385804 \end{pmatrix}, \quad \hat{\theta} - \theta_* = \begin{pmatrix} -0.182417 \\ -1.614196 \end{pmatrix} \quad (6.22)$$

$$s(\hat{\theta}) = \begin{pmatrix} 0.110754828 \\ 0.904435778 \end{pmatrix}$$

6.4.2. Model 1d verisi ve M-tahmin edici

Parametre başlangıç değerleri için L1-norm regresyon parametre tahminleri kullanılarak Huber t fonksiyonu ile parametrelerin M-tahmin edicileri bulunmuştur. İteratif tekrar ağırlıklandırılmalı en küçük kareler iterasyonu sonuçları aşağıda verilmiştir.

Çizelge 6.20. İteratif tekrar ağırlıklandırılmalı en küçük kareler iterasyonu sonuçları

k	θ_1	θ_2	S	$\Delta S/S$	$\Delta \theta/\theta$
0	2,000109755	2,896385380	0,030011725		
1	1,990364538	2,886625629	0,029866274	-0,004846499	0,004872341
2	1,990528124	2,892126591	0,029850253	-0,000536405	0,001905672
3	1,990581191	2,892869197	0,029847646	-8,73441E-05	0,000256768
4	1,990587527	2,892945572	0,029847361	-9,52938E-06	2,64012E-05
5	1,990588155	2,892952747	0,029847334	-9,16483E-07	2,48033E-06
6	1,990588213	2,892953396	0,029847332	-8,37446E-08	2,24281E-07

İteratif tekrar ağırlıklandırılmalı bir mekanizma ile çalışan ağırlıklandırılmış en küçük kareler normal denklemlerinin çözümü iterasyonların gidiş yolunu belirlediğinden $S(\theta)$ ağırlıklandırılmış kareler toplamı fonksiyonunun değerinde zaman zaman artışlar görülebilir. Bu gibi durumlarda iterasyonlar ilerlerken ağırlıklandırılmış kareler toplamı fonksiyonundaki $\Delta S/S$ oransal değişimlerinin $|\Delta S/S|$ mutlak değerinin küçülmesi izlenmelidir.

Huber t fonksiyonuyla parametrelerin M-tahmin edicileri ile elde edilen parametre tahminleri ve standart hata tahminleri aşağıda sunulmuştur:

$$\begin{aligned} \hat{\theta}_{M,H} &= \begin{pmatrix} 1.990588 \\ 2.892953 \end{pmatrix}, \quad \hat{\theta} - \theta_* = \begin{pmatrix} -0.009411 \\ -0.107046 \end{pmatrix} \\ s(\hat{\theta}) &= \begin{pmatrix} 0.061012041 \\ 0.514044133 \end{pmatrix} \end{aligned} \quad (6.23)$$

6.4.3. Model 1d süreci için simülasyon çalışması

Simülasyon aracılığı ile Model 1d sürecinden 1000 örnek üretilmiş ve herbiri için ayrı ayrı en küçük kareler (EKK), Huber t fonksiyonunun kullanıldığı M-tahmin edici (MH) ve L1-norm regresyonu (L1) yöntemi uygulanarak model parametre tahminleri yapılmıştır. Bu yöntemler için $\bar{\hat{\theta}}$ parametre tahminleri ortalamaları ve $\bar{s}(\hat{\theta})$ parametre tahminlerinin standart hataları tahminleri ortalamaları aşağıdaki çizelgede sunulmuştur.

Çizelge 6.21. Model 1d süreci örneklerine uygulanan yöntemler ve sonuçları

Yöntem	Tekrar	$\bar{\hat{\theta}}_1$	$\bar{s}(\hat{\theta}_1)$	$\bar{\hat{\theta}}_2$	$\bar{s}(\hat{\theta}_2)$
EKK	1000	1.800982	0.107114	1.253630	0.873338
MH	1000	1.959468	0.061675	2.640859	0.517921
L1	1000	2.003016	0.008765	2.934311	0.076557

İki aykırı değeri barındıran örneğin çapının küçük olması dolayısıyla en küçük kareler yöntemi ile elde edilen parametre tahminlerinin gerçek değerlerinden sapmalı olduğu açıklıkla görülebilir. Parametre tahminlerinin standart hatalarının büyük değerli oluşları ile bu durum onaylanır. M-tahmin edicileri için uygun başlangıç değerleri L1-norm regresyonu ile sağlandığından bu yöntemle parametre tahminleri de bulunmuştur. L1-norm tahminlerinin yansız ve etkin tahmin ediciler oldukları açıklıkla görülür. En küçük kareler parametre tahminleri ile karşılaştırıldığında Huber t fonksiyonunun kullanıldığı M-tahmin edici parametre tahminlerinin daha az sapmalı olduğu ve parametre tahminlerinin standart hataları tahminlerinin daha küçük olduğu görülür. Bundan dolayı M-tahmin edici parametre tahminleri yansız ve etkin tahmin edicilerdir.

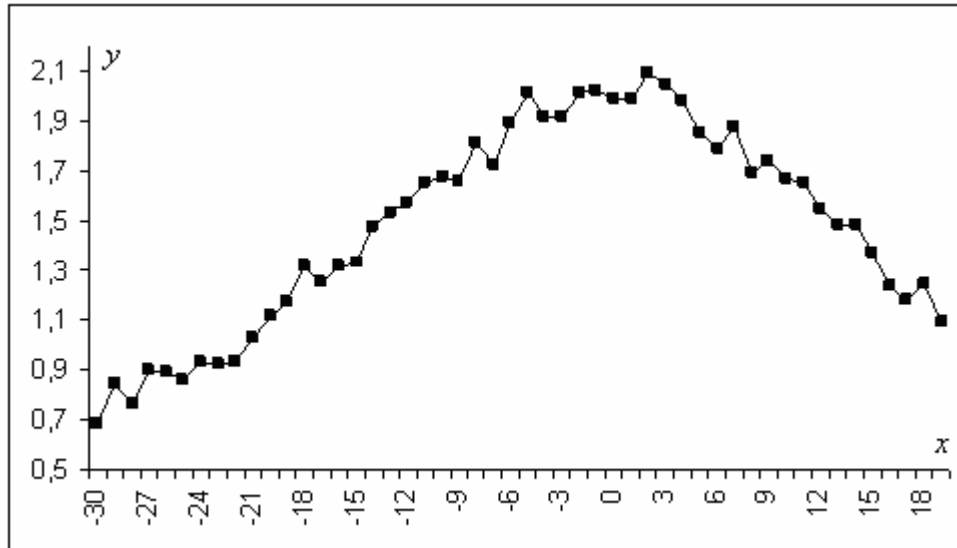
6.5. Model 2a Verisi

$$\text{Model 2: } y = \frac{\theta_1}{\theta_2 + x^2} + \varepsilon, \quad \begin{cases} \theta_1 = 1000 \\ \theta_2 = 500 \end{cases} \quad (6.24)$$

Bağımsız x değişkeni için -30 dan başlayarak 19 a kadar birer aralıklı 50 gözlem oluşturulmuştur. $\theta_1 = 1000$, $\theta_2 = 500$ olan Model 2 aracılığı ile oluşturulan çıktısına beklenen değerleri sıfır ve varyansı $(0.05)^2 = 0.0025$ olan normal dağılımdan bağımsız hatalar üretilerek eklenmiştir. Bu şekilde üretilen Model 2a verisi aşağıda sunulmuştur.

Çizelge 6.22. Model 2 yardımıyla üretilen Model 2a verisi

x	y	x	y	x	y
-30	0,688078	-13	1,530482	4	1,986867
-29	0,844590	-12	1,573714	5	1,859185
-28	0,762801	-11	1,651030	6	1,790134
-27	0,900559	-10	1,677505	7	1,876348
-26	0,890193	-9	1,666213	8	1,692601
-25	0,859339	-8	1,811945	9	1,743632
-24	0,936208	-7	1,728230	10	1,673143
-23	0,927257	-6	1,893119	11	1,654078
-22	0,934450	-5	2,014006	12	1,548116
-21	1,029529	-4	1,922331	13	1,487718
-20	1,118279	-3	1,916111	14	1,490266
-19	1,172173	-2	2,018763	15	1,377717
-18	1,320926	-1	2,023418	16	1,239906
-17	1,250681	0	1,992907	17	1,183938
-16	1,315116	1	1,988107	18	1,249398
-15	1,336593	2	2,098768	19	1,095311
-14	1,475397	3	2,047709		



Şekil 6.15. Model 2a verisinin diyagramı

6.5.1. Model 2a verisi ve en küçük kareler yöntemi

Model 2'den elde edilen veri üzerinden Gauss-Newton iterasyonu yöntemi ile parametre tahmini yapılmıştır. İterasyon için gerekli olan parametrelerin başlangıç

değerleri modelin f beklenen fonksiyonu parametrelerine göre doğrusal bir yapıya getirilerek elde edilmeye çalışılmıştır.

$$y = \frac{\theta_1}{\theta_2 + x^2} \rightarrow \frac{1}{y} = \frac{\theta_2}{\theta_1} + \frac{1}{\theta_1} x^2 \rightarrow v = \beta_0 + \beta_1 u$$

$$\theta_1 = \frac{1}{\beta_1}, \quad \theta_2 = \frac{\beta_0}{\beta_1} \quad (6.25)$$

$v = 1/y, u = x^2$ dönüşümüyle elde edilen veriler üzerinden doğrusal modelin parametre tahminleri olağan en küçük kareler yaklaşımıyla $\hat{\beta}_0 = 0.50304849$, $\hat{\beta}_1 = 0.000980948$ şeklinde bulunmuştur. Buradan orijinal modelin parametre başlangıç değerleri (1019.422365, 512.8188833) olarak elde edilir. Gauss-Newton iterasyonu için gerekli olan Z matrisinin satırı aşağıdaki şekilde verilir:

$$\frac{\partial f}{\partial \theta_1} = \frac{1}{\theta_2 + x^2}, \quad \frac{\partial f}{\partial \theta_2} = \frac{-\theta_1}{(\theta_2 + x^2)^2} \quad (6.26)$$

Aşağıda Gauss-Newton iterasyonu sonuçları verilmiştir.

Çizelge 6.23. Gauss-Newton iterasyonu sonuçları

k	θ_1	θ_2	S	$\Delta S/S$	$\Delta \theta/\theta$
0	1019,422365	512,8188833	0,151881569		
1	984,8533449	488,6447487	0,141586726	-0,067782041	0,047139712
2	985,6477725	489,2868244	0,141571362	-0,000108511	0,001313993
3	985,6592421	489,2938638	0,141571362	-4,73107E-09	1,43870E-05
4	985,6593579	489,2939327	0,141571362	-4,43645E-13	1,40897E-07
5	985,6593590	489,2939334	0,141571362	0	1,37846E-09

Dördüncü iterasyonda optimal parametre tahminlerine ulaşıldığı varsayılabilir. $\hat{\theta}$ parametre tahminleri ile birlikte diğer bulunanlar şöyledir:

$$\begin{aligned}
s^2 &= 0.002949403, \quad \sigma^2 = 0.0025 \\
\hat{\theta} &= \begin{pmatrix} 985.659 \\ 489.293 \end{pmatrix}, \quad \theta_* = \begin{pmatrix} 1000 \\ 500 \end{pmatrix}, \quad \hat{\theta} - \theta_* = \begin{pmatrix} -14.34064099 \\ -10.70606663 \end{pmatrix} \\
s(\hat{\theta}) &= \begin{pmatrix} 25.70722079 \\ 15.01898125 \end{pmatrix}, \quad \sigma(\hat{\theta}) = \begin{pmatrix} 24.31013495 \\ 14.28810448 \end{pmatrix}
\end{aligned} \tag{6.27}$$

6.5.2. Model 2a süreci için simülasyon çalışması

Simülasyon aracılığı ile benzer şekilde 1000 hata vektörü üretilip modele eklenmiş ve Model 2a sürecinden 1000 örnek elde edilmiştir. Elde edilen örneklerin herbiri için ayrı ayrı en küçük kareler (EKK) yöntemi ve L1-norm regresyonu (L1) yöntemi ile model parametreleri tahminleri yapılmıştır. Bu yöntemler için $\bar{\hat{\theta}}$ parametre tahminleri ortalamaları ve $\bar{s}(\hat{\theta})$ parametre tahminlerinin standart hataları tahminleri ortalamaları aşağıdaki çizelgede sunulmuştur.

Çizelge 6.24. Model 2a süreci örneklerine uygulanan yöntemler ve sonuçları

Yöntem	Tekrar	$\bar{\hat{\theta}}_1$	$\bar{s}(\hat{\theta}_1)$	$\bar{\hat{\theta}}_2$	$\bar{s}(\hat{\theta}_2)$
EKK	1000	1000.886	24.19843	500.4253	14.22193
L1	1000	998.9801	2.620815	500.1944	1.635320

Simülasyonla yapılan tekrar sayıları arttıkça en küçük kareler parametre tahminleri ortalamalarının parametrelerin gerçek değerlerine, parametre tahminlerinin standart hataları tahminleri ortalamalarının da parametre tahminlerinin teorik standart hatalarına yakınsadığı görülebilir. L1-norm parametre tahminlerinin standart hatalarının oldukça küçük olduğu görülmesi dolayısıyla bu yaklaşımla bulunan parametre tahminleri etkin tahmin edicilerdir. Ayrıca L1-norm tahminlerinin daha az sapmalı olması dolayısıyla parametre tahminleri yansızdır.

6.6. Model 2b Verisi

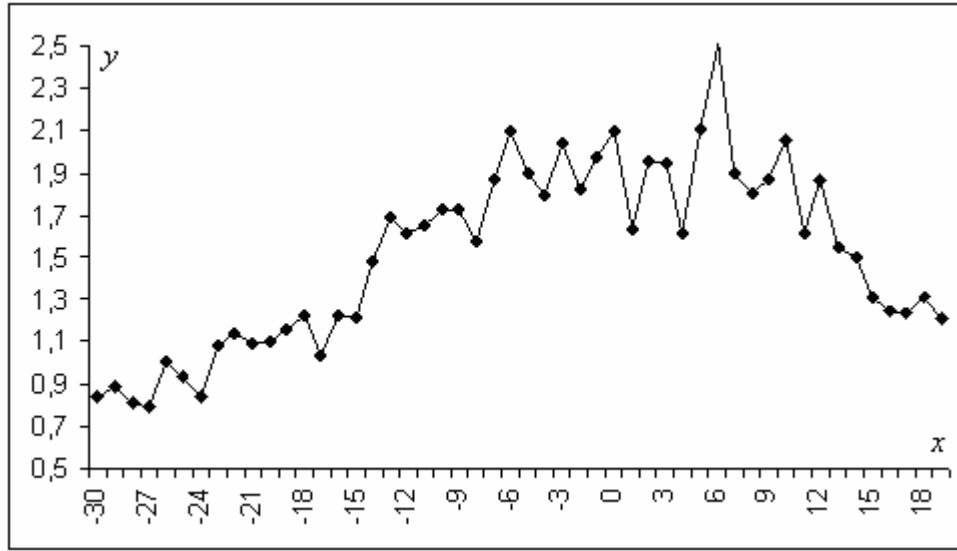
Bu örnek $-30 \leq x \leq 19$ aralığında x bağımsız değişkeninin birer artacak şekilde

$$y = \frac{\theta_1}{\theta_2 + x^2} + \varepsilon \rightarrow f = \frac{\theta_1}{\theta_2 + x^2} \quad \text{ifadesi aracılığı ile Model 2 üzerinden}$$

üretilmiştir. Beklenen değerleri sıfır ve varyansı $(0.1 \times f)^2$ olan normal dağılımdan bağımsız hatalar üretilip gerçek parametre değerleri $\theta_1 = 1000$, $\theta_2 = 500$ olan Model 2 çıktısına eklenerek Model 2b verisi oluşturulmuştur. Bu şekilde üretilen Model 2b verisi aşağıda sunulmuştur.

Çizelge 6.25. Model 2 yardımıyla üretilen Model 2b verisi

x	y	x	y	x	y
-30	0,836171	-13	1,686370	4	1,615333
-29	0,890557	-12	1,610923	5	2,105738
-28	0,814264	-11	1,653976	6	2,513157
-27	0,790827	-10	1,723131	7	1,897754
-26	1,006432	-9	1,727764	8	1,797433
-25	0,931082	-8	1,578836	9	1,867840
-24	0,835760	-7	1,870014	10	2,054548
-23	1,089499	-6	2,096781	11	1,616238
-22	1,142193	-5	1,896936	12	1,866900
-21	1,098671	-4	1,788047	13	1,543746
-20	1,101827	-3	2,039964	14	1,496753
-19	1,161884	-2	1,818243	15	1,315353
-18	1,226609	-1	1,972254	16	1,249971
-17	1,039250	0	2,090657	17	1,239071
-16	1,230307	1	1,631944	18	1,312297
-15	1,213343	2	1,957253	19	1,203734
-14	1,479652	3	1,939622		



Şekil 6.16. Model 2b verisinin diyagramı

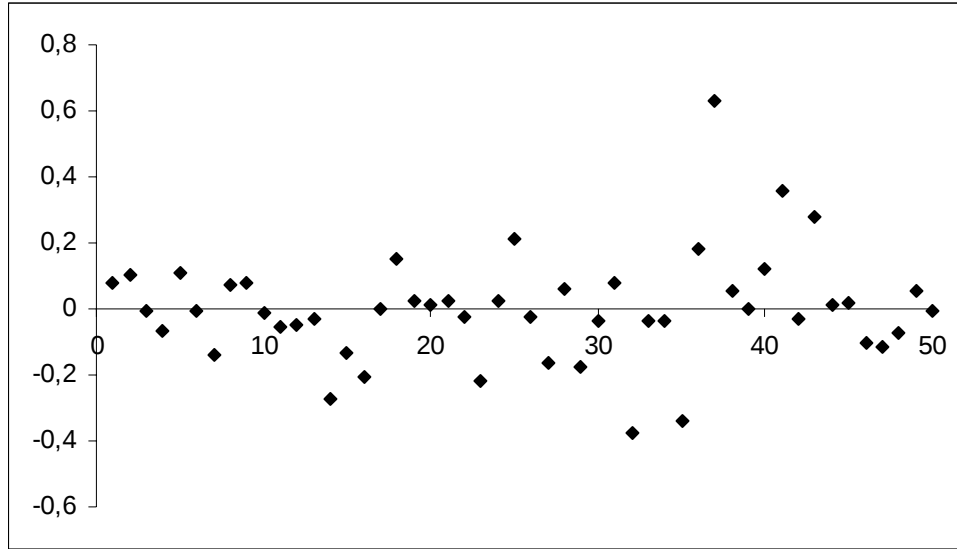
6.6.1. Model 2b verisi ve geliştirilmiş en küçük kareler

Buradaki uygulama için hata varyans-kovaryans matrisinin bilinmediği ve hatalar arası korelasyonlarının sıfır olduğu varsayılarak parametre tahminleri yapılmıştır. Model 2 doğrusallaştırılarak doğrusal modelin parametreleri olağan en küçük kareler yöntemiyle $\hat{\beta}_0 = 0.511908125$, $\hat{\beta}_1 = 0.000875777$ şeklinde tahmin edildikten sonra orijinal modelin parametre başlangıç değerlerine $\theta_{10} = 1141.843734$, $\theta_{20} = 584.5190855$ şeklinde çevrilmiştir. Bu başlangıç değerleri olağan en küçük kareler yöntemine uygulanan Gauss-Newton iterasyonları için kullanılmış ve bulunan parametre tahminleri yardımıyla model artıkları elde edilmiştir.

Çizelge 6.26. Gauss-Newton iterasyonları

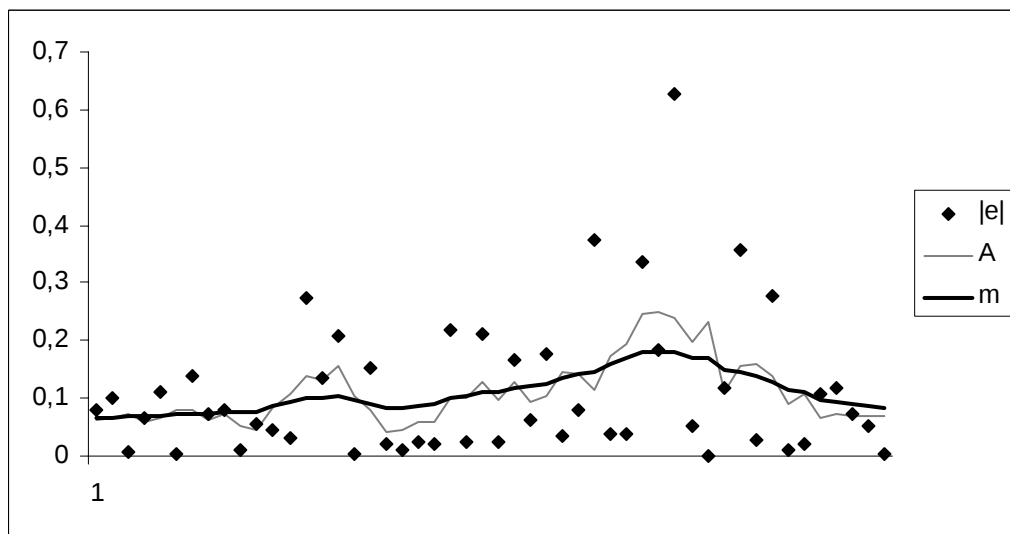
k	θ_1	θ_2	S	$\Delta S/S$	$\Delta \theta/\theta$
0	1141,843734	584,5190855	1,388721175		
1	1086,153410	538,8726791	1,337825500	-0,036649312	0,078092243
2	1091,072524	542,6548487	1,337451675	-0,000279427	0,007018670
3	1090,979301	542,6056164	1,337451640	-2,58410E-08	9,07248E-05
4	1090,980884	542,6065532	1,337451640	-6,55398E-12	1,72648E-06
5	1090,980854	542,6065354	1,337451640	-1,11022E-15	3,27638E-08

Parametre tahminine beşinci iterasyonda ulaşılmıştır. Parametre tahminleriyle model artıklarının saçılma grafiği aşağıda verilmiştir.



Şekil 6.17. Ağırlıklandırma yapılmadan elde edilen model artıkları serpmne diyagramı

Artıkların mutlak değerleri üzerine A_t merkezi hareketli ortalamalarının ve $\alpha = 0.2$ için m_t serisinin eklendiği grafik aşağıda verilmiştir.



Şekil 6.18. Mutlak artıklar, A_t merkezi hareketli ortalamaları ve m_t serisi

Artıkların mutlak değerleri ortalamaları yukarıdaki örnekteki gibi izlenerek standart sapmaların oransal büyüklükleri tahmin edilmiştir. Ağırlık amacıyla kullanılan K matrisinin köşegen elemanları m_i serisi olarak seçilirse standart sapmaların oransal büyüklükleri yaklaşık olarak bulunmuş olur. K matrisinin köşegen elemanları olan m_i serisi aşağıdaki çizelgede verilmiştir.

Çizelge 6.27. Gözlem düzeylerine bağlı olarak K matrisinin köşegen elemanları

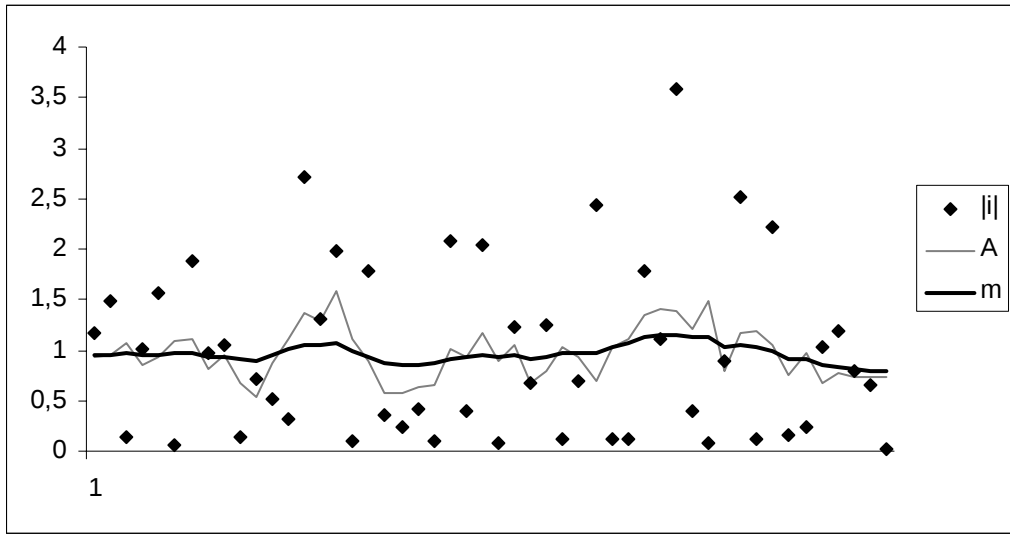
i	m_i	i	m_i	i	m_i
1	0,066434	18	0,090058	35	0,178755
2	0,067296	19	0,084143	36	0,181303
3	0,069151	20	0,084337	37	0,179466
4	0,068309	21	0,087371	38	0,170921
5	0,070249	22	0,090537	39	0,168927
6	0,072924	23	0,099150	40	0,148345
7	0,074301	24	0,103651	41	0,146086
8	0,073083	25	0,111169	42	0,139017
9	0,075536	26	0,111759	43	0,128819
10	0,075274	27	0,119229	44	0,115089
11	0,077250	28	0,119681	45	0,109479
12	0,085501	29	0,126101	46	0,098170
13	0,092505	30	0,136732	47	0,093476
14	0,099256	31	0,142236	48	0,088600
15	0,100124	32	0,145471	49	0,084956
16	0,102763	33	0,159165	50	0,082041
17	0,095385	34	0,168133		

K matrisi sabit olarak alınıp ağırlık amacıyla kullanıldığında $\theta_{10} = 1141.843734$, $\theta_{20} = 584.5190855$ parametre başlangıç değerleri ile birlikte Gauss-Newton iterasyonu sonuçları aşağıda verilmiştir.

Çizelge 6.28. Model 2b verisi için Gauss-Newton iterasyonu sonuçları

k	θ_1	θ_2	S	$\Delta S/S$	$\Delta \theta/\theta$
0	1141,843734	584,5190855	83,31860691		
1	1102,100571	553,1020050	81,94925045	-0,016435182	0,053748597
2	1103,142772	554,0637312	81,94667285	-3,14537E-05	0,001738786
3	1103,160937	554,0760349	81,94667268	-2,03834E-09	2,22063E-05
4	1103,161148	554,0761731	81,94667268	-2,50910E-13	2,49459E-07
5	1103,161151	554,0761747	81,94667268	0	2,79854E-09

Dördüncü iterasyonda parametre tahminlerine ulaşılmıştır. Parametrelerin genelleştirilmiş en küçük kareler tahmini bulunduktan sonra mutlak ağırlıklandırılmış artıkları, mutlak ağırlıklandırılmış artıklar üzerine kurulan A_t merkezi hareketli ortalamaları ve m_t serisi grafiğinden artıkların varyans durağanlığının büyük ölçüde sağlandığı görülür.



Şekil 6.19. Mutlak ağırlıklandırılmış artıklar, A_t ortalamaları ve m_t serisi

Hata varyanslarının tahminleri yardımıyla ağırlıklandırılmış en küçük kareler yöntemiyle parametre tahminleri bulunmakla birlikte diğer tahminler ve teorik değerler aşağıda verilmiştir.

$$\begin{aligned}
 s^2 &= 1.707222348, \quad \sigma^2 = 1 \\
 \hat{\theta} &= \begin{pmatrix} 1103.161151 \\ 554.0761747 \end{pmatrix}, \quad \theta_* = \begin{pmatrix} 1000 \\ 500 \end{pmatrix}, \quad \hat{\theta} - \theta_* = \begin{pmatrix} 103.1611508 \\ 54.07617468 \end{pmatrix} \quad (6.28) \\
 s(\hat{\theta}) &= \begin{pmatrix} 62.95324625 \\ 40.03011500 \end{pmatrix}, \quad \sigma(\hat{\theta}) = \begin{pmatrix} 54.55317074 \\ 35.58454252 \end{pmatrix}
 \end{aligned}$$

s^2 varyans tahmini hata varyanslarının tahminlerinden etkilendiğinden σ^2 varyansından farklı bulunmuştur. Buradaki süreç için beklenen değerleri sıfır ve varyansı $(0.1 \times f)^2$ olan normal dağılımdan hatalar üretildiğinden hatalar

$1/(0.1 \times f)^2$ şeklinde ağırlıklandırılırsa hataların standart normal dağılımlı hatalara dönüşmesi dolayısıyla ölçeklenmiş varyans $\sigma^2 = 1$ olarak bulunur. Buradan $\sigma(\hat{\theta})$ parametre tahminlerinin teorik standart hataları θ_* parametrelerin gerçek değerleri ve σ^2 teorik hata varyansı ile birlikte $1/(0.1 \times f)^2$ ağırlıkları kullanılarak bulunur. Ayrıca $\hat{\theta}$ parametre tahminleri kullanılarak elde edilen model artıkları hata varyansları tahmini ile ağırlıklandırılarak hesaplanan s^2 hata varyansı tahmini ile birlikte $s(\hat{\theta})$ parametre tahminlerinin standart hata tahminleri bulunmuştur.

6.6.2. Model 2b verisi ve yarı olabilirlik yöntemi

Bu modelin hatalarının varyanslarının bilinmediği ve hatalar arası korelasyonlarının sıfır olduğu varsayımıyla parametrelerin tahmini yapılmıştır.

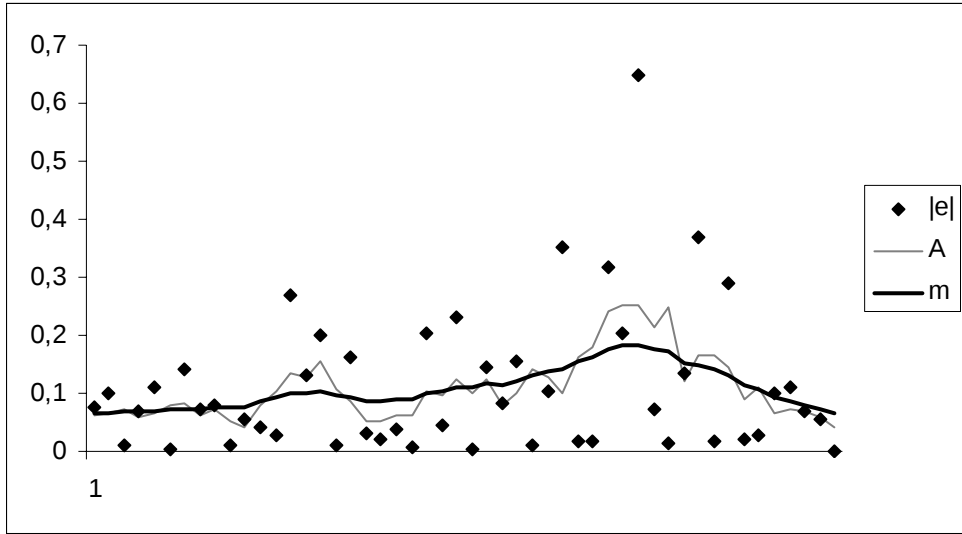
Model 2 doğrusallaştırılarak doğrusal modelin parametreleri olağan en küçük kareler yöntemiyle tahmin edilmiş sonra bu değerler orijinal modelin parametre başlangıç değerlerine $\theta_{10} = 1141.843734$, $\theta_{20} = 584.5190855$ şeklinde dönüştürülmüştür. Bu başlangıç değerleri yardımıyla elde edilen model artıklarının mutlak değerlerinin düzgünleştirilmesi işlemi uygulanarak bir sonraki iterasyon için ağırlıklar tahmin edilmiştir. $\theta_{k+1} = \theta_k + (Z_k^t V_k^{-1} Z_k)^{-1} Z_k^t V_k^{-1} [y - f(\theta_k)]$ iterasyonlarına uygun olarak her iterasyonda güncellenen parametre tahminleri ile güncellenen ağırlıklar eş zamanlı olarak yürütülmüştür. Yarı olabilirlik tahmin yaklaşımına göre yapılan iterasyonlar aşağıda sunulmuştur.

Çizelge 6.29. Model 2b verisi için yarı olabilirlik tahmin yöntemi iterasyonları

k	θ_1	θ_2	S	$\Delta S/S$	$\Delta \theta/\theta$
0	1141,843734	584,5190855	0,976212646		
1	1106,802159	557,3894995	0,961062058	-0,015519762	0,046413516
2	1104,690074	555,8849964	0,961253006	0,000198684	0,002699195
3	1104,387096	555,6712326	0,961284659	3,29286E-05	0,000384547
4	1104,343303	555,6406163	0,961289248	4,77360E-06	5,50979E-05
5	1104,336996	555,6362501	0,961289898	6,76149E-07	7,85789E-06
6	1104,336093	555,6356316	0,961289989	9,47819E-08	1,11320E-06
7	1104,335965	555,6355446	0,961290001	1,31582E-08	1,56488E-07
8	1104,335947	555,6355325	0,961290003	1,80664E-09	2,18037E-08
9	1104,335944	555,6355309	0,961290003	2,44805E-10	3,00672E-09
10	1104,335944	555,6355306	0,961290003	3,26374E-11	4,09563E-10
11	1104,335944	555,6355306	0,961290003	4,26237E-12	5,49609E-11
12	1104,335944	555,6355306	0,961290003	5,42233E-13	7,23821E-12
13	1104,335944	555,6355306	0,961290003	6,52811E-14	9,30145E-13
14	1104,335944	555,6355306	0,961290003	7,77156E-15	1,15574E-13
15	1104,335944	555,6355306	0,961290003	0	1,35447E-14
16	1104,335944	555,6355306	0,961290003	0	1,88738E-15

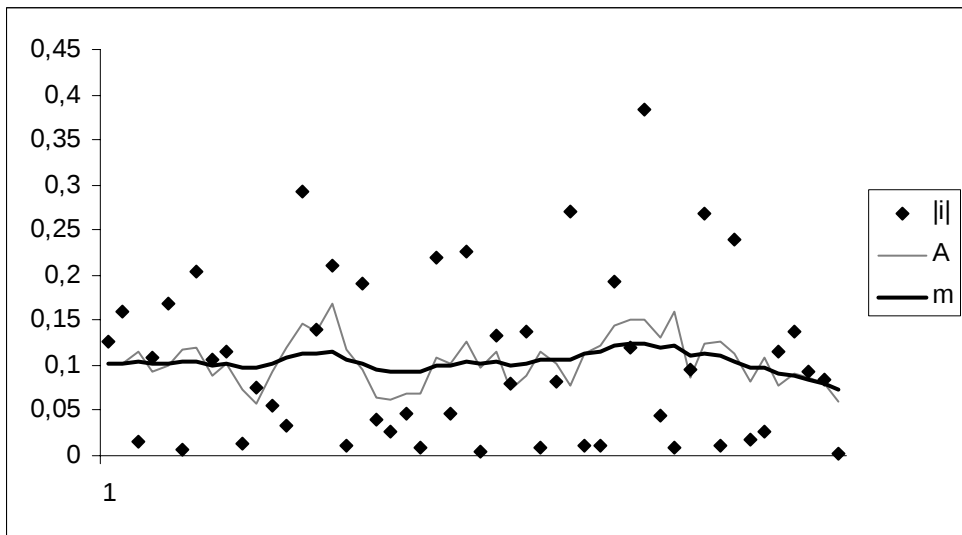
Yedinci iterasyonda parametre tahminlerine ulaşıldığı varsayılabilir. Parametrelerin tahminleri gerçek değerlerinden sapmalı bulunduğu tahminlerde yanlışlık mevcuttur. İterasyonların ilerlemesiyle ağırlıklandırılmış kareler toplamı fonksiyonunun değerinde çok küçük oransal artışların gerçekleştiği dikkati çeker. Ağırlıklandırılmış kareler toplamı fonksiyonu değişmekte ve optimal bir yapıya varmaktadır.

Yarı olabilirlik tahminine göre model artıkları mutlak değerleri, mutlak değerler üzerinden hesaplanan A_i merkezi hareketli ortalamaları ve $\alpha = 0.2$ için m_t serisi ortak grafiği aşağıda verilmiştir.



Şekil 6.20. Artıkların mutlak değeri, A_t merkezi hareketli ortalamaları ve m_t serisi

Ayrıca ağırlıklandırılmış artıkların mutlak değerleri üzerinden hesaplanan A_t merkezi hareketli ortalamaları ve $\alpha = 0.2$ için m_t serisi ortak grafiği aşağıdadır.



Şekil 6.21. Ağırlıklandırılmış artıkların mutlak değeri grafiği

İterasyonlar sonucu ağırlıklandırılmış artıklar sabit varyans koşulunu sağlar. Yarı olabilirlik yöntemiyle parametre tahminleri bulunmakla birlikte diğer tahminler ve teorik değerler aşağıda sunulmuştur.

$$\begin{aligned}
s^2 &= 0.020026875, \quad \sigma^2 = 1 \\
\hat{\theta} &= \begin{pmatrix} 1104.335944 \\ 555.6355306 \end{pmatrix}, \quad \theta_* = \begin{pmatrix} 1000 \\ 500 \end{pmatrix}, \quad \hat{\theta} - \theta_* = \begin{pmatrix} 104.3359437 \\ 55.6355306 \end{pmatrix} \quad (6.29) \\
s(\hat{\theta}) &= \begin{pmatrix} 62.06207437 \\ 39.63905768 \end{pmatrix}, \quad \sigma(\hat{\theta}) = \begin{pmatrix} 54.55317074 \\ 35.58454252 \end{pmatrix}
\end{aligned}$$

Hata varyansları ile ilgili oransal tahminlerin bir sabitle çarpılması ve bu tahminlerin dengelenmesi dolayısıyla s^2 varyans tahmini σ^2 varyansından farklıdır; hata varyansları tahminleri doğrudan bulunmayıp hata standart sapmaları tahminlerinin oranları m_t serisi terimleri olarak alınmıştır. Beklenen değerleri sıfır ve varyansı $(0.1 \times f)^2$ olan normal dağılımdan hatalar üretilmiş olduğundan $1/(0.1 \times f)^2$ ağırlıkları doğrudan kullanıldığında hatalar standart normal dağılımlı hatalara dönüşerek ölçeklenmiş varyans $\sigma^2 = 1$ olarak bulunur.

Tablolarda $S(\theta)$ kareler toplamı veya $S(\theta)$ ağırlıklandırılmış kareler toplamı gösterilmeyebilir. İterasyonlar ağırlıklandırılmış en küçük kareler normal denklemlerinin çözümü üzerinde yürüdüğünden değişen hata varyansları tahminine bağlı olarak ağırlıklandırılmış kareler toplamı fonksiyonunun değerinde artışlar olabilmektedir. Bundan dolayı parametreler içerisinde en çok oransal değişim gösteren $\Delta\theta/\theta$ oranının izlenmesi yeterli görülebilir.

6.6.3. Model 2b süreci için simülasyon çalışması

Simülasyonla Model 2b verisinin hata yapısına benzer şekilde hatalar üretilip Model 2 çıktısına eklenerek çapı 50 birim olan 1000 tane örnek üretilmiştir. Her örnek için en küçük kareler (EKK), ağırlıklandırılmış en küçük kareler (AEKK), yarı olabilirlik (YO) ve L1-norm regresyonu (L1) yöntemi ile model parametre tahminleri yapıp parametre tahminleri ortalaması ve tahminlerin standart hataları tahminleri ortalaması aşağıda verilmiştir.

Çizelge 6.30. Değişken varyanslı Model 2b süreci örneklerine uygulanan yöntemler ve sonuçları

Yöntem	Tekrar	$\hat{\theta}_1$	$\bar{s}(\hat{\theta}_1)$	$\hat{\theta}_2$	$\bar{s}(\hat{\theta}_2)$
EKK	1000	999.3825	73.46792	501.4600	43.34390
AEKK	1000	998.2062	51.92088	500.6847	33.23463
YO	1000	998.4398	51.16950	501.2304	32.94670
L1	1000	1003.413	6.820750	505.1839	4.436187

Değişken varyanslı Model 2b süreci ile üretilen örneklere en küçük kareler yöntemi uygulandığında parametre tahminlerinin ortalamaları parametre gerçek değerlerine yakınsamakla birlikte diğer yöntemlerle bulunan sonuçlarla karşılaştırıldığında parametre tahminlerinin standart hataları tahminleri ortalamaları en büyük olarak bulunmuştur. Ağırlıklandırılmış en küçük kareler yöntemi ve yarı olabilirlik yöntemi ile bulunan parametre tahminlerinin standart hataları tahminleri ortalamaları benzer olmakla birlikte yarı olabilirlik yöntemi ile bulunan parametre tahminlerinin standart hataları tahminleri ortalaması biraz daha küçüktür; ayrıca her iki yöntemle bulunan parametre tahminleri ortalamaları beklendiği gibi parametrelerin gerçek değerlerine yakınsamaktadır.

6.7. Model 2c Verisi

İçinde aykırı değerlerin de bulunduğu ve hatalarının varyanslarının farklı olduğu bir veri üretilip bu veri üzerinden parametre tahmininin yapıldığı bir örnek incelenmiştir.

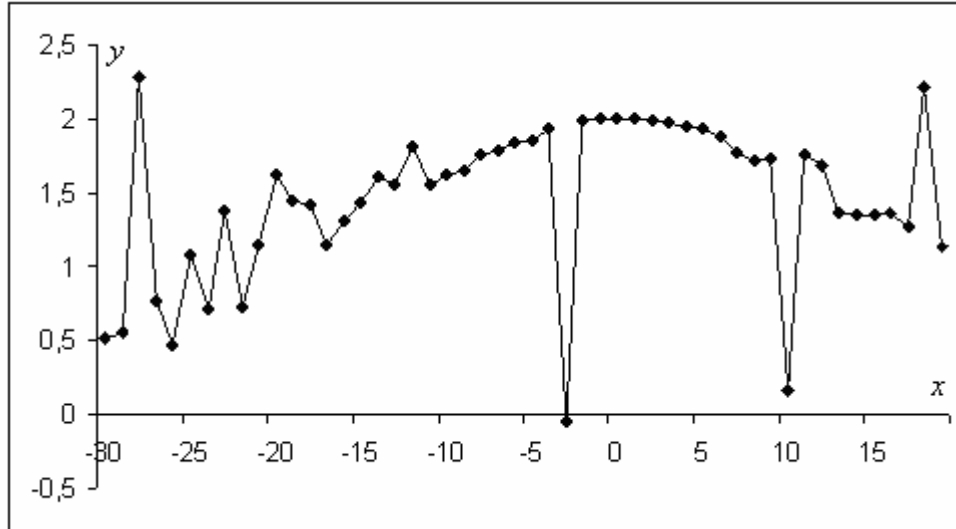
Bu örnek $y = \frac{\theta_1}{\theta_2 + x^2} + \varepsilon \rightarrow f = \frac{\theta_1}{\theta_2 + x^2}$ ile Model 2 yardımıyla x bağımsız

değişkeninin $[-30, 19]$ aralığında birer artarak 50 gözlem olacak şekilde üretilmiştir. Beklenen değerleri sıfır ve $\kappa = 0.0005$ ile varyansı $(\kappa x^2 f)^2$ olan normal dağılımdan simülasyon aracılığı ile bağımsız hatalar üretilip gerçek parametre değerleri $\theta_1 = 1000$, $\theta_2 = 500$ olan Model 2 çıktısına eklenmiştir. Ayrıca veride aykırı değerler olması için $x = -28$, $x = -3$, $x = 10$ ve $x = 18$ gözlemlerine sırasıyla 1.5, -2, -1.5 ve 1 büyük mutlak değerli hataları doğrudan eklenip Model

2c verisi oluşturulmuştur. Bu şekilde üretilen Model 2c verisi aşağıda sunulmuş ve grafiği çizilmiştir.

Çizelge 6.31. Model 2 yardımıyla üretilen Model 2c verisi

x	y	x	y	x	y
-30	0,506867	-13	1,559073	4	1,939361
-29	0,554089	-12	1,810165	5	1,926989
-28	2,278816	-11	1,557846	6	1,879469
-27	0,772036	-10	1,6247	7	1,770252
-26	0,468581	-9	1,652009	8	1,720999
-25	1,084545	-8	1,756038	9	1,735476
-24	0,711821	-7	1,778212	10	0,166667
-23	1,376912	-6	1,841476	11	1,761902
-22	0,729273	-5	1,856198	12	1,692147
-21	1,145512	-4	1,926322	13	1,368336
-20	1,625687	-3	-0,03536	14	1,345163
-19	1,44921	-2	1,986611	15	1,351829
-18	1,420569	-1	1,997419	16	1,359412
-17	1,154871	0	2	17	1,267735
-16	1,310898	1	1,995252	18	2,213592
-15	1,438524	2	1,987579	19	1,133061
-14	1,602636	3	1,972762		



Şekil 6.22. Model 2c verisinin diyagramı

6.7.1. Model 2c verisi ve parametre tahmini

Uyarlanmış bir sağlam yöntemi ve yarı olabilirlik yönteminin her ikisinin birden kapsandığı yöntemlerin genelleştirilmesi yaklaşımında gözlem düzeylerinde değişkenlik gösteren standart sapma tahminleri şu şekilde bulunmuştur. Aykırı değerler de dahil bütün artıkların mutlak değerleri üzerine beş birimlik merkezleştirilmiş $med_t = (|e_{t-2}|, |e_{t-1}|, |e_t|, |e_{t+1}|, |e_{t+2}|)$ medyanları bulunup medyanlar üzerine $\alpha = 0.2$ ile S_t üstel düzeltmesi ve R_t geri üstel düzeltmesi hesaplanarak dengeleme için $m_t = \frac{1}{2}(S_t + R_t)$ ortalaması bulunmuştur. Buradan standart sapma tahminleri $s_t = \sqrt{\pi/2}m_t$ yardımıyla elde edilmiştir.

Aykırı değerlerin bulunduğu değişken varyanslı hatalara sahip Model 2c verisine uyarlanmış bir sağlam yönteminin ve yarı olabilirlik yönteminin her ikisinin birden kapsandığı yöntemlerin genelleştirilmesi yaklaşımı uygulanarak model parametreleri tahmin edilmiştir. İlk olarak model parametrelerine göre doğrusallaştırılarak parametre başlangıç değerleri bulunmuştur. Aykırı değerlerin ve değişken varyanslı hataların varlığı dolayısıyla doğrusallaştırılmış modelin parametre tahminleri L1-norm regresyonu ile $\hat{\beta}_0 = 0.499126116$, $\hat{\beta}_1 = 0.001062122$ olarak bulunmuş ve orijinal model için parametre başlangıç değerleri $\theta_{10} = 941.5116414$, $\theta_{20} = 469.9330484$ olarak elde edilmiştir.

Değişken varyanslar ihmal edildiğinde Huber t fonksiyonundan uyarlamayla ağırlığın

$$w(z) = \begin{cases} 1 & , |z| \leq 2 \\ 16/z^4 & , |z| > 2 \end{cases} \quad (6.30)$$

olarak alındığı bir sağlam yaklaşımı incelenmiştir. Buradaki uyarlanan ağırlık fonksiyonuyla özellikle büyük z değerleri ile ağırlığın hızla düşürülerek parametre tahminlerinde aykırı değer etkileri tam olarak sıfırlanmamakla birlikte son derece

küçültülmektedir. Değişken varyanslar da dahil edildiğinde ρ sağlam fonksiyonu, $\psi = \rho'$ etki fonksiyonu ve $w = \psi/z$ ağırlık fonksiyonu şu şekilde bulunur.

$$\begin{aligned}\rho(z) &= \begin{cases} \frac{z^2}{2s^2} & , \quad |z| \leq 2 \\ \frac{1}{s^2} \left(4 - \frac{8}{z^2} \right) & , \quad |z| > 2 \end{cases} \\ \psi(z) &= \begin{cases} \frac{z}{s^2} & , \quad |z| \leq 2 \\ \frac{16}{s^2 z^3} & , \quad |z| > 2 \end{cases} \\ w(z) &= \begin{cases} \frac{1}{s^2} & , \quad |z| \leq 2 \\ \frac{16}{s^2 z^4} & , \quad |z| > 2 \end{cases}\end{aligned}\tag{6.31}$$

Herhangi bir θ vektörü için artıklar $r_i = y_i - f_i(\theta)$ şeklinde ve bu artıkların herbirine karşı gelen standart sapmalar s_i şeklinde tanımlanarak T amaç fonksiyonu

$$T = \sum_i \rho(z_i) \quad , \quad z_i = \frac{r_i}{s_i}\tag{6.32}$$

şeklinde ifade edilir. Bununla birlikte iterasyonların ilerlemesine bağlı olarak ağırlıklandırılmış $S(\theta) = r^t W r$ kareler toplamı fonksiyonu da izlenebilir. Bunlardan son iterasyon ağırlıklandırılmış kareler toplamı fonksiyonu $\hat{V}(\hat{\theta})$ parametre tahminlerinin varyans-kovaryans matrisinin tahmini için önemlidir.

$\theta_{10} = 941.5116414$, $\theta_{20} = 469.9330484$ parametre başlangıç değerleri yukarıda tarif edilen yöntem ile ilgili Gauss-Newton iterasyonları için kullanılırsa aşağıdaki iterasyonlar elde edilir.

Çizelge 6.32. Yöntemlerin genl. yakl. Gauss-Newton iterasyonları

k	θ_1	θ_2	S	T	$\Delta S/S$	$\Delta T/T$	$\Delta \theta/\theta$
0	941,51164	469,93305	32,426222	6985,5241			
1	978,77673	489,85578	33,430116	7740,1602	0,0309593	0,1080286	0,0423948
2	983,96709	492,51537	33,384068	7580,0306	-0,001377	-0,020688	0,0054293
3	985,85287	493,47138	33,373201	7522,3824	-0,000326	-0,007605	0,0019411
4	986,54460	493,81977	33,370349	7501,1850	-8,54E-05	-0,002818	0,0007060
5	986,80006	493,94793	33,369491	7493,3165	-2,57E-05	-0,001049	0,0002595
6	986,89481	493,99536	33,369208	7490,3856	-8,48E-06	-0,000391	9,602E-05
7	986,93005	494,01298	33,369110	7489,2924	-2,95E-06	-0,000146	3,571E-05
8	986,94317	494,01953	33,369074	7488,8844	-1,06E-06	-5,45E-05	1,330E-05
9	986,94806	494,02198	33,369061	7488,7322	-3,88E-07	-2,03E-05	4,958E-06
10	986,94989	494,02289	33,369057	7488,6753	-1,43E-07	-7,59E-06	1,849E-06
11	986,95057	494,02323	33,369055	7488,6541	-5,30E-08	-2,83E-06	6,900E-07
12	986,95082	494,02335	33,369054	7488,6462	-1,97E-08	-1,06E-06	2,575E-07
13	986,95092	494,02340	33,369054	7488,6432	-7,34E-09	-3,95E-07	9,611E-08
14	986,95095	494,02342	33,369054	7488,6421	-2,74E-09	-1,47E-07	3,587E-08
15	986,95097	494,02343	33,369054	7488,6417	-1,02E-09	-5,50E-08	1,339E-08
16	986,95097	494,02343	33,369054	7488,6416	-3,81E-10	-2,05E-08	4,998E-09

Onbirinci iterasyonda parametre tahminlerine ulaşıldığı varsayılabilir. Parametre tahminleri ve bunlara karşı gelen standart sapma tahminleri sırasıyla şöyledir:

$$\hat{\theta} = \begin{pmatrix} 986.950973 \\ 494.023429 \end{pmatrix}, \quad s(\hat{\theta}) = \begin{pmatrix} 38.59804482 \\ 20.28693643 \end{pmatrix} \quad (6.33)$$

Aykırı değerler ihmal edilerek gerçek parametre değerleri $\theta_{1*} = 1000$, $\theta_{2*} = 500$ ve gerçek ağırlıklar kullanılarak hesaplanan parametre tahminlerinin teorik standart hataları $\sigma(\hat{\theta}_1) = 51.05335558$, $\sigma(\hat{\theta}_2) = 25.60023758$ olarak bulunur.

İterasyonların ilerleme şeklini iteratif tekrar ağırlıklandırmalı en küçük kareler normal denklemlerinin çözümü belirlediğinden iterasyonlar ilerlerken ağırlıklar da değişebildiğinden S ağırlıklandırılmış kareler toplamı fonksiyonunun değerinde veya T amaç fonksiyonunun değerinde bazen artışlar olabilmektedir.

6.7.2. Model 2c süreci için simülasyon çalışması

Simülasyonla Model 2c verisinin hata yapısına benzer şekilde hatalar üretilip Model 2 çıktısına eklenerek çapı 50 olan 1000 örnek oluşturulmuştur. Herbir örnek için ayrı ayrı en küçük kareler (EKK), ağırlıklandırılmış en küçük kareler (AEKK), yarı olabilirlik (YO) ve L1-norm regresyonu (L1) yöntemi ile model parametre tahminleri bulunup parametre tahminleri ortalamaları ve parametre tahminlerinin standart hataları tahminleri ortalamaları aşağıda verilmiştir.

“Yaklaşım Önerileri” bölümünde yöntemlerin genelleştirilmesi yaklaşımları genel olarak tarif edilmiştir. Burada uyarlanmış bir sağlam yönteminin ve yarı olabilirlik yönteminin her ikisinin birden kapsandığı bir yaklaşım (YG) kullanılarak parametre tahminleri de yapılmıştır.

Çizelge 6.33. Aykırı değerler barındıran değişken varyanslı Model 2c süreci örneklerine uygulanan yöntemler ve sonuçları

Yöntem	Tekrar	$\bar{\hat{\theta}}_1$	$\bar{s}(\hat{\theta}_1)$	$\bar{\hat{\theta}}_2$	$\bar{s}(\hat{\theta}_2)$
EKK	1000	1546.650	458.5924	866.8103	292.1995
AEKK	1000	1316.651	307.2773	711.8718	193.1096
YO	1000	1605.977	574.0238	909.2872	368.0901
YG	1000	995.0201	49.79585	497.5240	25.88549
L1	1000	1017.562	9.703812	509.4650	5.177553

Aykırı değerler barındıran değişken varyanslı Model 2c süreci örnekleri üzerinden en küçük kareler, ağırlıklandırılmış en küçük kareler ve yarı olabilirlik yöntemleri ile bulunan model parametre tahminleri ortalamalarının yakınsadığı limitlerin parametrelerin gerçek değerlerinden farklı olduğu görülür. Bu durum parametre tahminlerinin standart hata tahminleri ortalamalarının büyüklükleri ile de desteklenir. Bu yaklaşımlarla bulunan parametre tahminleri yanlıdır ve etkin değildir.

Uyarlanmış bir sağlam yönteminin ve yarı olabilirlik yönteminin her ikisinin birden kapsandığı bir yaklaşımla bulunan parametre tahminleri ortalamaları gerçek parametre değerleri civarına yakınsamakla birlikte parametre tahminlerinin standart

hataları tahminleri ortalamaları da yukarıda ifade edilen yöntemlerin sonuçları ile karşılaştırıldığında en küçüktür. Bu yaklaşımla bulunan parametre tahminlerinin gerçek parametre değerlerine olan yakınlığı ve standart hataları tahminlerinin teorik standart hatalarına olan yakınlığı dolayısıyla önerilen yaklaşımla bulunan tahminler yansız ve etkindir.

L1-norm regresyonu parametre tahminlerinin (YG) yaklaşımından daha sapmalı olduğu, ancak standart hata tahminlerinin daha küçük olduğu görülür. L1-norm regresyonu tahminleri için yapılan simülasyon çalışmaları ile dikkate değer bazı özellikler farkedilmiştir. Bu yöntem ile elde edilen parametre tahminlerinin gerçek değerlerine yakın olduğu görülmekle birlikte bu yöntemle elde edilen parametre tahminlerinin standart hatalarının tahminleri üzerinde bir miktar belirsizlikler bulunmaktadır. Parametre tahminlerindeki çok küçük değişimlerin standart hataları üzerinde çok büyük etkileri olabilmektedir. Parametre uzayında yüzeyin eğrilik yarıçapı küçüldükçe parametre tahminine karşılık gelen standart hata da küçülür. Parametre tahmini, mutlaklar toplamı fonksiyonunu oluşturan mutlak değer fonksiyonlarından birinin dip noktasında bulunuyorsa bu noktaya karşı gelen eğrilik yarıçapı sıfıra yakınsadığından buna karşılık gelen standart hata da sıfıra doğru yakınsar. L1-norm regresyonunda parametre tahminlerinin standart hatalarının aşırı değişkenliklerinin nedeni bu oluşumdur. Ancak L1-norm parametre tahminleri diğer yöntemlerle bulunan tahminlerle karşılaştırıldığında bu yaklaşımla bulunan tahminlerin daha az sapmalı olduğu görülür.

6.8. Model 3a Verisi

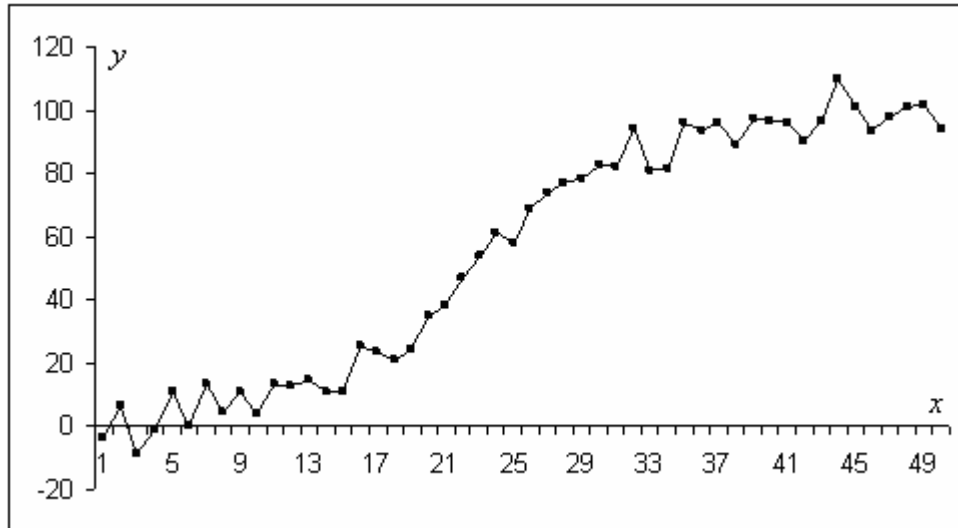
$$\text{Model 3: } y = \frac{\theta_1}{1 + \theta_2 e^{-\theta_3 x}} + \varepsilon, \quad \begin{cases} \theta_1 = 100 \\ \theta_2 = 99 \\ \theta_3 = 0.2 \end{cases} \quad (6.34)$$

Parametre değerleri $\theta_1 = 100, \theta_2 = 99, \theta_3 = 0.2$ olan Model 3 olarak adlandırılan lojistik büyüme modeli aracılığı ile bağımsız x değişkeninin 1-50 aralığında birer

artarak oluşturulan 50 gözlemlik çıktısına beklenen değerleri sıfır ve varyansı $(5)^2 = 25$ olan normal dağılımdan bağımsız hatalar üretilerek eklenmiş ve bu şekilde Model 3a verisi elde edilmiştir. Simülasyon ile üretilen bir Model 3a verisi aşağıda sunulmuş ve grafiği çizilmiştir.

Çizelge 6.34. Model 3 aracılığı ile üretilen Model 3a verisi

x	y	x	y	x	y
1	-3,7687	18	20,74745	35	95,7502
2	6,520639	19	24,02958	36	93,70607
3	-8,85645	20	35,08677	37	96,32561
4	-1,0885	21	37,97518	38	89,30728
5	10,94341	22	46,9103	39	97,06231
6	0,334829	23	53,59911	40	96,5726
7	13,41763	24	61,51299	41	95,98736
8	4,620998	25	58,01403	42	90,67534
9	11,05636	26	69,19285	43	96,47374
10	4,016944	27	73,98284	44	109,7226
11	13,71934	28	77,08399	45	100,8424
12	13,08484	29	78,33789	46	93,46656
13	14,78688	30	82,49062	47	97,89733
14	10,89164	31	82,14708	48	100,9256
15	10,71341	32	94,31115	49	101,5015
16	25,56832	33	80,70995	50	94,11074
17	23,51513	34	81,82813		



Şekil 6.23. Model 3a verisinin diyagramı

6.8.1. Model 3a verisi ve en küçük kareler yöntemi

Gauss-Newton iterasyonunun işleyebilmesi için gerekli olan Z matrisinin satırı şu şekilde kurulur:

$$\frac{\partial f}{\partial \theta_1} = \frac{1}{1 + \theta_2 e^{-\theta_3 x}} \quad , \quad \frac{\partial f}{\partial \theta_2} = \frac{-\theta_1 e^{-\theta_3 x}}{(1 + \theta_2 e^{-\theta_3 x})^2} \quad , \quad \frac{\partial f}{\partial \theta_3} = \frac{\theta_1 \theta_2 x e^{-\theta_3 x}}{(1 + \theta_2 e^{-\theta_3 x})^2} \quad (6.35)$$

Örnek üzerinden büyüme modelini karakterize eden üç noktanın yaklaşık değerleri tahmin edilip bu noktalar yardımıyla üç parametrenin başlangıç değerleri bulunmuştur. İterasyonlar yardımı ile bulunan en küçük kareler tahminleri ile teorik değerler karşılıklı sunulmuştur.

$$\begin{aligned} s^2 &= 23.6373049 \quad , \quad \sigma^2 = 25 \\ \hat{\theta} &= \begin{pmatrix} 100.0790 \\ 93.53276 \\ 0.201658 \end{pmatrix} \quad , \quad \theta_* = \begin{pmatrix} 100 \\ 99 \\ 0.2 \end{pmatrix} \quad , \quad \hat{\theta} - \theta_* = \begin{pmatrix} 0.079015 \\ -5.467241 \\ 0.001658 \end{pmatrix} \\ s(\hat{\theta}) &= \begin{pmatrix} 1.503292 \\ 23.00175 \\ 0.011689 \end{pmatrix} \quad , \quad \sigma(\hat{\theta}) = \begin{pmatrix} 1.586710 \\ 25.37959 \\ 0.011978 \end{pmatrix} \end{aligned} \quad (6.36)$$

Lojistik eğrisi üzerine kurulan lojistik regresyon analizinin ve uygulamalarının incelendiği Başarır (1990) tarafından yapılan bir çalışmadan yararlanılabilir.

6.8.2. Model 3a süreci için simülasyon çalışması

Model 3a verisi oluşturma süreci 1000 kez tekrarlanarak herbiri 50 şer gözlemden oluşan 1000 örnek oluşturulmuştur. Elde edilen 1000 örnek için ayrı ayrı en küçük kareler (EKK) yöntemi ve L1-norm regresyonu (L1) yöntemi ile model parametre tahminleri yapılmıştır. $\bar{\hat{\theta}}$ parametre tahminleri ortalamaları ve $\bar{s}(\hat{\theta})$ parametre

tahminlerinin standart hataları tahminleri ortalamaları aşağıdaki çizelgede sunulmuştur.

Çizelge 6.35. Model 3a süreci örneklerinin parametre tahminleri

Yöntem	$\bar{\hat{\theta}}_1$	$\bar{s}(\hat{\theta}_1)$	$\bar{\hat{\theta}}_2$	$\bar{s}(\hat{\theta}_2)$	$\bar{\hat{\theta}}_3$	$\bar{s}(\hat{\theta}_3)$
EKK	100.0415	1.584781	104.0559	27.27273	0.200572	0.011980
L1	100.0607	0.078143	105.2903	1.693386	0.200392	0.000756

Simülasyon tekrarının artmasına bağlı olarak en küçük kareler yönteminde parametre tahminleri ortalamalarının gerçek parametre değerlerine yakınsadığı, parametre tahminlerinin standart hataları tahminleri ortalamalarının teorik değerlerine yakınsadığı görülür. En küçük kareler yöntemi ve L1-norm yöntemi ile bulunan parametre tahminleri benzer olmakla birlikte L1-norm yöntemi ile bulunan tahminlerin standart hata tahminleri daha küçük bulunmuştur.

6.9. Stokastik Araştırma Stratejilerinin Performansına Yönelik Simülasyon Çalışması

Tvrđik ve diğerlerinin önerdiği yöntem ile bu tezde bir yaklaşım olarak önerilen yöntem parametre tahminleri açısından karşılaştırılmıştır. Yapısal olarak bu iki yöntem yukarıda ele alınmış olunan Gauss-Newton tabanlı algoritmalarından oldukça farklı olduğundan ve özellikle parametre tahminlerinin standart hatalarının doğrudan hesaplanamamalarından dolayı ayrı bir sınıf olarak ele alınmıştır.

6.9.1. Model 1a verisi ve Tvrđik yöntemi

Model 1a verisi üzerine Krviy ve Tvrđik (1995) tarafından önerilen üç sezgisel yolun kullanıldığı ve Storn ve Price (1997) yaklaşımı olan diferansiyel evrimsel algoritmasının da kullanıldığı dört sezgisel yolu kapsayan Tvrđik ve diğerlerinin yöntemi uygulanmıştır. Kutu kısıtlarının (-100, +100) aralığında alınarak yapılan bir uygulamada 4000 iterasyonun sonunda şu sonuca ulaşılmıştır:

Çizelge 6.36. Tvrdik yöntemine göre gerçekleşen yakınsama

	H	N	L
θ1	2,018328	2,018328	2,018328
θ2	3,17014	3,17014	3,17014
S	0,016735	0,016735	0,016735

Bu durumda en iyi yakınsama politopun L köşe noktası (2.02, 3.17) olarak gerçekleşir. Yapılan 4000. iterasyon ile elde edilen tablo şöyledir:

Çizelge 6.37. Tvrdik yönteminin son iterasyon tablosu

	R1	R2	R3	R4	
θ1	2,018328	2,018328	2,018328	27,43509	
θ2	3,17014	3,17014	3,17014	146,0921	
S	0,016735	0,016735	0,016735	42,79646	
w	0	0	0	0	T 1
W	24,28576	8,237436	22,71353	14,32916	4000
q	0,349104	0,118412	0,326504	0,20598	

Burada sırasıyla parametrelerinin $\alpha = 2$ ve $s = 0.5$ olduğu R1 deneme noktasının bulunduğu birinci sezgisel yol, parametrelerinin $\alpha = 5$ ve $s = 1.5$ olduğu R2 deneme noktasının bulunduğu sezgisel yol, simpleksin bir noktası olarak populasyonun minimum değerli noktasının alındığı ve parametrelerinin $\alpha = 2$ ve $s = 0.5$ olduğu R3 deneme noktasının bulunduğu üçüncü sezgisel yol ve parametrelerinin $F_{\min} = 0.4$ ve $C=0.9$ olduğu Storn ve Price yaklaşımı olan diferansiyel evrimsel algoritması ile R4 deneme noktasının bulunduğu dördüncü sezgisel yoldur. W satırı, amaç fonksiyonu değerindeki başarılı bulunan w bağıl değişim oranlarının toplamıdır. Başlangıçta sıfır numaralı iterasyon için w_0 ağırlıkları bir olarak alınmıştır.

6.9.2. Model 1a verisi ve önerilen stokastik yaklaşım

Beşinci bölümde önerilen yaklaşım Model 1a verisine uygulandığında, başlangıç için kutu kısıtı (-1000000, +1000000) aralığında alınarak 2500 iterasyon sonunda elde edilen yakınsama şöyledir:

Çizelge 6.38. Beşinci bölümde önerilen yaklaşımla gerçekleşen yakınsama

	H	N	L
θ1	2,016992495	2,01699317	2,016993394
θ2	3,155436936	3,155439778	3,155446532
S	0,016728059	0,016728059	0,016728059

En iyi yakınsama politopun L köşe noktası (2.017, 3.155) olarak gerçekleşir; yakınsamanın beş basamağa kadar gerçekleştiği görülebilir. Aşağıda son iterasyon tablosu verilmektedir.

Çizelge 6.39. Önerilen yaklaşımın son iterasyon tablosu

	R1	R2	R3		
θ1	2,211639	2,016995	-87,1052		2,016995
θ2	3,086542	3,155449	80,50135		3,155449
S	0,467303	0,016728	305,0371		0,016728
	0	1	0	T	
Δn	0	0	0	1	593
n	72	392	129	2500	

Sırasıyla R1, R2 ve R3 kolonları birinci, ikinci ve üçüncü sezgisel yol ve bulunan deneme noktalarını gösterir. n satırı ile R1, R2 ve R3 kolonlarının kesişimi indirgemenin sağlandığı başarılı iterasyon sayılarını gösterir. Δn satırı 2500.üncü iterasyonda indirgemenin gerçekleşmediğini söyler. Son iterasyonda ikinci sezgisel yol ile bulunan deneme noktası en küçük değerli olmakla birlikte bulunan deneme noktası politop için kabul edilmemiştir. Yapılan 2500 iterasyondan 593 ünde indirgeme sağlanmıştır. Yapılan indirgemelerin 72 tanesi birinci yol ile, 392 tanesi ikinci yol ile ve 129 tanesi üçüncü yol ile gerçekleştirilmiştir. Son iterasyonda bulunan politopun H haricinde P_i köşe noktalarının karşılıklı bileşenlerinin standart sapmaları sırasıyla 1.58×10^{-7} ve 4.78×10^{-6} olarak bulunur; bu değerler politopun büyüklüğü hakkında bir fikir verebilir.

6.9.3. Model 1a süreci için stokastik arařtırmalarının performansına yönelik yapılan simülasyon çalışması

Stokastik arařtırma algoritmaları olan Tvrdik yönteminin ve önerilen yöntemin performansları řu řekilde karşılaştırılmıştır. Simülasyon aracılığı ile sabit varyanslı Model 1a sürecinden her biri 50 gözlemden oluşan 1000 örnek üretilip her bir örnek için model parametre tahminleri Tvrdik yöntemi ile ve önerilen yöntem ile yapılmıştır. Bulunan $\bar{\theta}$ parametre tahminlerinin ortalaması ve $s(\bar{\theta})$ parametre tahminleri ortalamasının standart hatası tahmini ařağıdaki tabloda sunulmuştur. Her bir örnek üzerinden her iki yöntem için kutu kısıtları (-100, +100) olarak alınmış, parametre tahmini için Tvrdik yöntemiyle 5000 iterasyon, önerilen yöntemle 300 iterasyon yapılmıştır.

Çizelge 6.40. Tvrdik yöntemi ve önerilen yöntemle elde edilen tahminlerin karşılaştırılması

Yöntem	Tekrar	İterasyon	$\bar{\theta}_1$	$s(\bar{\theta}_1)$	$\bar{\theta}_2$	$s(\bar{\theta}_2)$
Tvrdik	1000	5000	2.014707	0.008569	3.121703	0.061844
Önerilen	1000	300	2.001080	0.000583	3.012634	0.006827

Önerilen yöntemle parametre tahminleri daha az sapmalı ve parametre tahminleri ortalamasının standart hata tahmini daha küçük bulunmuştur. İterasyon sayıları açısından Tvrdik yöntemiyle karşılaştırıldığında önerilen yöntemin daha hızlı çalıştığı açıklıkla görülür. Daha az iterasyon sayısına karşılık parametre tahminleri ortalaması gerçek değerlerine daha yakındır. Her iki stokastik yöntem karşılaştırıldığında önerilen yöntemle bulunan parametre tahminlerindeki yanlılığın daha küçük olduğu görülür.

6.9.4. Model 1c verisi ve Tvrdik yöntemi

Bir uygulama olarak hata varyanslarının farklı olduğu Model 1c verisine Tvrdik yöntemi kullanılarak parametre tahmini yapılmıştır. Veri kullanılarak tahmin edilen model artıklarının homojenliği sağlanması için artıkların mutlak değerlerinin

düzgünleştirilmesi işlemi ile oluşturulan V^{-1} ağırlık matrisi kareler toplamı fonksiyonuna dahil edilmiştir.

$$S(\theta) = (y - f(\theta))^t V^{-1} (y - f(\theta)) \quad (6.37)$$

Bu şekilde tanımlanan kareler toplamı fonksiyonu, kutu kısıtları (-100, +100) olarak alınan Tvrdik yöntemi ile minimize edilmiş ve şu sonuçlar elde edilmiştir.

Çizelge 6.41. Tvrdik yöntemine göre gerçekleşen yakınsama

	H	N	L
θ1	2,010661	1,994549	1,878604
θ2	1,999925	3,523906	2,089078
S	0,247665	0,217956	0,208464

En iyi yakınsama politopun L köşe noktası (1.88, 2.09) olarak gerçekleşir. Yakınsama 10000 iterasyon sonunda gerçekleştirilebilmiştir. 10000. ninci iterasyonda elde edilen tablo şöyledir:

Çizelge 6.42. Tvrdik yönteminin son iterasyon tablosu

	R1	R2	R3	R4	
θ1	1,829093	1,777812	1,838982	-10,4794	
θ2	3,97668	4,534973	3,86901	94,19063	
S	1,061719	1,852695	0,937371	8,218842	
w	0	0	0	0	T
W	10,83974	1	8,185066	6,364452	10000
q	0,410763	0,037894	0,310167	0,241176	

Hata homojenliği olmayan verilerde Tvrdik yöntemi çok ağır çalışmaktadır ve yakınsama hızı çok düşüktür.

6.9.5. Model 1c verisi ve önerilen yöntem

Hata varyanslarının farklı olduğu Model 1c verisi üzerinden artıkların mutlak değerlerinin düzgünleştirilmesi ile elde edilen ağırlıklandırmanın dahil edildiği

ağırlıklandırılmış kareler toplamı fonksiyonu kullanılarak önerilen yöntemle parametre tahmini yapılmıştır. Kutu kısıtının $(-1000000, +1000000)$ aralığında olduğu araştırma ile 6910 iterasyon sonunda elde edilen yakınsama şöyledir:

Çizelge 6.43. Önerilen yöntemle gerçekleşen yakınsama

	H	N	L
θ1	2,02306409	2,02306409	2,02306409
θ2	2,96446223	2,96446219	2,96446218
S	0,14035341	0,14035341	0,14035341

Yakınsama sekiz basamağa kadar gerçekleşir. Son iterasyon tablosu aşağıda sunulmuştur.

Çizelge 6.44. Önerilen yöntemin son iterasyon tablosu

	R1	R2	R3		
θ1	2,005938	2,023064	15,99184		2,023064
θ2	2,869689	2,964462	31,44839		2,964462
S	0,141444	0,140353	90,32373		0,140353
	0	1	0	T	
Δn	0	0	0	1	1116
n	79	868	169	6910	

Yapılan 6910 iterasyondan 1116 sında indirgeme gerçekleşmiştir. Bileşen standart sapmaları sırasıyla 1.78×10^{-9} ve 6.14×10^{-9} olarak bulunur.

6.9.6. Model 3a süreci için stokastik araştırması simülasyonu çalışması

Sabit varyanslı Model 3a verisi oluşturma süreci 1000 kez tekrarlanarak her biri 50 şer gözlemden oluşan 1000 örnek oluşturulmuştur. Her bir örnek için model parametre tahminleri Tvrdik yöntemi ile ve beşinci bölümde önerilen yöntem ile yapılmıştır. Bulunan $\hat{\theta}$ parametre tahminlerinin ortalamaları ve $s(\hat{\theta})$ parametre tahminleri ortalamalarının standart hataları tahminleri aşağıdaki çizelgede verilmiştir. Kutu kısıtları her iki yöntem için $(-100, +100)$ olarak alınarak Tvrdik yönteminde her bir örnek için 5000 iterasyon, önerilen yöntem için 300 iterasyon yapılmıştır.

Çizelge 6.45. Tvrdik yöntemi ve önerilen yöntemle elde edilen tahminlerin karşılaştırılması

Yöntem	$\hat{\theta}_1$	$s(\hat{\theta}_1)$	$\hat{\theta}_2$	$s(\hat{\theta}_2)$	$\hat{\theta}_3$	$s(\hat{\theta}_3)$
Tvrdik	92.22743	5.627434	114.4743	9.695720	0.246208	0.0427520
Önerilen	104.0122	0.622286	108.5084	1.135772	0.208962	0.001321

Tvrdik yöntemi ile karşılaştırıldığında ve gerçek parametre değerleri gözönünde bulundurulduğunda önerilen yaklaşımla parametre tahminleri daha az sapmalı dolayısıyla yanlışlık daha az ve parametre tahminleri ortalamasının standart hata tahmini daha küçük bulunmuştur.

7. SONUÇ

Gerçek parametre değerleri ile üç farklı model çıktısı üzerine simülasyon aracılığı ile hatalar üretilip eklenerek veri oluşturulmuş, sonra bu veri üzerinden parametre tahminleri bulunmuştur. Elde edilen tahminlerin yanlılıkları ve standart hataları açısından tahmin ediciler karşılaştırılmıştır. Veriye aykırı değerler eklenerek en küçük kareler yöntemi ile parametre tahminleri bulunduğu parametre tahmin edicilerinin standart hatalarının arttığı gözlenmiştir. Hata varyanslarının farklı olduğu veri üzerinden parametre tahminleri, ağırlıklandırma yapılmadan ve ağırlıklandırılmış en küçük kareler yöntemi ile bulunmuştur. Parametre tahmin sonuçları karşılaştırıldığında beklenildiği gibi ağırlıklandırılmış en küçük kareler yöntemi ile elde edilen parametre tahminleri gerçek değerlerinden daha az sapmalı ve parametre tahminlerinin standart hataları daha küçük bulunmuştur.

Yarı olabilirlik tahmin yaklaşımında ağırlıklandırılmış en küçük kareler yönteminde kullanılan Gauss-Newton iterasyonları ortaya çıkmaktadır, fakat her iterasyonda parametre tahmini ile hata varyans-kovaryans matrisi tahmininin tekrar güncellenmesi gereklidir. Her iterasyonda bulunması gereken hata varyansı tahminindeki küçük değişiklikler parametre tahminlerinin durağanlaşmasını geciktirdiğinden fazladan birçok iterasyonun yapılması gerektiği çalışmalardan görülmüştür. Yarı olabilirlik yaklaşımının üzerinde kurulduğu varsayımlar ve matematik yapılanma birçok başka tahmin yöntemleri arasında bağlantı kurulabilmesine imkan vermektedir.

Hatalar kalın kuyruklu bir dağılımdan geliyorsa birçok aykırı değer oluşacağından parametre tahminleri etkilenir. Bu gibi durumlar için güvenilir parametre tahminlerinin bulunmasını sağlayan sağlam yaklaşımları geliştirilmiştir. Bunlardan sağlam Lp-norm regresyonu ile aykırı değerlerin önemi küçültülerek parametre tahminlerinin aykırı değerlerden daha az etkilenmesi sağlanır. Diğer bir sağlam yaklaşımı olan M-tahmin edicilerde ağırlık fonksiyonları aracılığı ile aykırı değerlere düşük ağırlıklar verilerek parametre tahmini üzerindeki baskılar azaltılır. Uygulama amacıyla aykırı değerler barındıran bir veri simülasyon aracılığı ile üretilip bu veri

üzerinden M-tahmin edici kullanılarak parametre tahmini yapılmıştır. Diğer tahmin yöntemleri ile karşılaştırıldığında M-tahmin edici ile bulunan parametre tahminlerindeki sapmaların ve parametre tahminleri standart hata tahminlerinin belirgin bir oranda daha küçük olduğu görülmüştür. Ayrıca Andrew dalga fonksiyonunun ve Huber t fonksiyonunun kullanıldığı M-tahmin edicileri karşılaştırıldığında Andrew dalga fonksiyonunun kullanıldığı M-tahmin edici ile bulunan parametre tahminlerinin daha az sapmalı olduğu ve parametre tahminleri standart hata tahminlerinin daha küçük olduğu görülmüştür.

Yukarıda ifade edilen yöntemlerden farklı olarak bir stokastik araştırma algoritması olan Tvrdik ve diğerlerinin yöntemi tanıtılmıştır. Bu yöntem kutu kısıtı içinde rassal olarak oluşturulan bir populasyon üzerinden çalışır; populasyonun minimum ve maksimum değerli noktalarının yöntem için büyük önemi bulunur. Ancak bu yöntem Gauss-Newton yönteminde olduğu gibi parametre başlangıç değerlerine bağımlı değildir. Tvrdik yöntemi kutu içindeki noktalardan başlayarak global minimuma doğru yakınsamayı gerçekleştirebildiğinden en azından başlangıç aşamaları açısından Gauss-Newton yöntemi ve uyarlamalarından daha güvenilirdir. Hata varyanslarının sabit olduğu ve sabit olmadığı veriler üzerine Tvrdik yöntemi kullanılarak parametre tahminleri yapılmıştır. Ancak hata varyansları sabit olmadığından her iterasyonda hata varyanslarının tahmininin yapılması dolayısıyla Tvrdik yöntemi ile yakınsama hızı oldukça düşmüştür. Buna bağlı olarak elde edilen parametre tahminleri hassas olmamakla birlikte makul sayılabilir. Ek olarak Tvrdik yönteminde kullanılan stokastik özellikteki Storn ve Price algoritmasının biraz daha kontrollü bir şekilde yönlendirilebileceği bir yenilik önerilmiştir.

Bu çalışmada bir stokastik araştırma algoritması olan bir yöntem önerilmiştir. Bu yöntem kullanılarak parametre tahmini yapılmış ve Tvrdik ve diğerlerinin yöntemiyle elde edilen parametre tahminleri ile karşılaştırılmıştır. Stokastik araştırma algoritması sınıfından olan Tvrdik yöntemi ile karşılaştırıldığında önerilen yöntemin daha hızlı çalıştığı ve sonuçların daha az sapmalı olduğu görülmekle birlikte parametre tahminleri ortalaması standart hatası daha küçük bulunmuştur. Tvrdik yönteminden farklı olarak önerilen yaklaşımın yapısı incelenirse yapısal olarak

Tvrdik yönteminden daha basit olduğu ve yapısı dolayısıyla daha başarılı sonuçların elde edildiği görülür.

Ağırlıklandırma temelinde çalışan yöntemler ile sağlam yaklaşımları tek bir çatı altında toparlanabilmektedir. Bu bağlantıyı özellikle yarı olabilirlik tahmininin üzerinde inşa edildiği varsayımlar ve matematiksel yapılanma sağlar. Bu yapılar incelendiğinde doğal bir genelleme süreci beraberinde gelmektedir. Bunun bir sonucu olarak daha ileriye gidilerek ağırlık fonksiyonları gözönünde bulundurulduğunda parametre tahmininin bulunmasını sağlayan daha genel bir amaç fonksiyonunun tasarlanabilir olduğu görülebilir. Yaklaşım önerileri bölümünde uyarlanmış bir sağlam yönteminin ve yarı olabilirlik yönteminin her ikisinin birden kapsandığı bir yaklaşıma değinilerek simülasyon çalışmalarında bu yaklaşımla parametre tahminleri yapılmıştır; bulunan tahminlerin az sapmalı ve etkin olduğu görülmüştür.

Yapılan bu çalışmalarla birlikte parametre tahmini için kullanılan yöntemlere değinilmiş ve bunlar üzerine uygulama yapılmıştır. Bunlarla birlikte bazı fikirler geliştirilmiş ve öneri olarak sunulmuştur. Ancak yapılan çalışmalarla buradaki öneri özelliğindeki yaklaşımların bazıları tam olarak sonuçlandırılmamıştır. Bunlar üzerinde çalışılması gerekmektedir.

KAYNAKLAR

- Agarwall, R.P. and O'Regan, D., "Nonlinear Analysis and Applications" (2 cilt), **Kluwer Academic Publishers** (2003).
- Aksoy, S., "Otokorelasyonlu Hata Terimli Doğrusal Olmayan Regresyon Modellerinde Parametre Tahmini", **Gazi Üniversitesi Fen Bilimleri Enstitüsü** Doktora Tezi, İstatistik (1996).
- Ali, M.M. and Torn, A., "Population set based global optimization algorithms: some modifications and numerical studies", **Comput. Oper. Res.** 31: 1703-1725, (2004).
- Andrews, D.F., Bickel, P.J., Hampel, F.R., Huber, P.J., Rogers, W.H. and Tukey, J.W., "Robust Estimates of Location", **Princeton University Press**, Princeton, New Jersey (1972).
- Aydın, C., "Kovaryans matrisi için bilgi verici olmayan önsel dağılım ve doğrusal olmayan modellerde parametre kestirimi", **Gazi Üniversitesi Fen Bilimleri Enstitüsü** Doktora Tezi, İstatistik (1992).
- Back, T., "Evolutionary Algorithms in Theory and Practice", **Oxford University Press**, New York (1996).
- Barrodale, I., "L1 approximations and the analysis of data", **Appl. Statist.**, 17: 51-57 (1968).
- Barrodale, I. and Roberts, F.D.K., "An improved algorithm for discrete L1 linear approximation", **SIAM J. Numer. Anal.**, 10: 839-848 (1973).
- Başarır, G., "Çok Değişkenli Verilerde Ayrımsama Sorunu ve Lojistik Regresyon Analizi", **Hacettepe Üniversitesi Fen Bilimleri Enstitüsü** İstatistik Anabilim Dalı Doktora Tezi, Kasım (1990).
- Bates, D.M. and Watts, D.G., "Nonlinear Regression Analysis and Its Applications", **Wiley**, New York (1988).
- Bates, D.M., Hamilton, D.C. and Watts, D.G., "Calculation of intrinsic and parameter effects curvatures for nonlinear regression models", **Commun. Stat. Simul. Comput. B**, 12: 469-477 (1983).
- Bierens, H.J. and Gallant, R., "Nonlinear Models" (2 cilt), **The international library of critical writings in econometrics 8**, **Edward Elgar Publishing Limited**, Cheltenham, UK (1997).
- Box, M.J., "An experimental design criterion for precise estimation of a subset of the parameters in a nonlinear model", **Biometrika**, 58: 149-153 (1971).

Brent, R.P., “Algorithms of minimization without derivatives”, **Prentice Hall**, Englewood Cliffs, New Jersey (1973).

Bronson, R. and Naadimuthu, G., “Theory and Problems of Operations Research”, second edition, **Schaum’s outline series, McGraw-Hill**, New York (1997).

Bunke, H. and Bunke, O., “Nonlinear Regression, Functional Relations and Robust Methods, Statistical Methods of Model Building”, **John Wiley and Sons** (1989).

Candan, M., “Doğrusal Regresyon Çözümlemesinde Sağlam Kestiriciler”, **Hacettepe Üniversitesi Fen Bilimleri Enstitüsü** İstatistik Anabilim Dalı Bilim Uzmanlığı Tezi (1995).

Chatterjee, S. and Price, B., “Regression Analysis by Example”, Second Edition pp 33-44, **John Wiley** (1991).

Cook, R.J., Lawless, J.F., Lakhal-Chaieb, L. and Lee, K.A., “Robust Estimation of Mean Functions and Treatment Effects for Recurrent Events Under Event-Dependent Censoring and Termination: Application to Skeletal Complications in Cancer Metastatic to Bone”, **Journal of the American Statistical Association**, 104(485): 60-75, Applications and Case Studies (2009).

Çağlar, A., “Doğrusal Olmayan Modellerde Tahminlere İlişkin Güven Bölgelerinin Oluşturulması ve Bir Karşılaştırma”, **Gazi Üniversitesi Fen Bilimleri Enstitüsü** Yüksek Lisans Tezi, İstatistik (1995).

Deb, K., “A population-based algorithm-generator for real parameter optimization”, **Soft Comput.** 9: 236-253 (2005).

Denison, D.D., Hansen, M.H., Holmes, C.C., Mallick, B. and Yu, B., “Nonlinear Estimation and Classification”, **Springer-Verlag New York Inc.** (2003).

Fried, R., Einbeck, J. and Gather, U., “Weighted Repeated Median Smoothing and Filtering”, **Journal of the American Statistical Association**, 102(480): 1300-1308, Theory and Methods (2007).

Gallant, A.R., “Nonlinear Statistical Models”, **Wiley**, New York (1987).

Gentle, J.M., Kennedy, W.J. and Sposito, V.A., “On least absolute deviations estimators”, **Commun. Statist.**, A6: 839-845 (1977).

Grigg, O. and Spiegelhalter, D., “A Simple Risk-Adjusted Exponentially Weighted Moving Average”, **Journal of the American Statistical Association**, 102(477): 140-152, Theory and Methods (2007).

Günay, S., “Kitle Parametrelerini Tahmin Yöntemleri”, **Hacettepe Üniversitesi Fen Fakültesi İstatistik Enstitüsü**, Ankara (1972).

Hamilton, L.C., "Regression with Graphics: Second Courses in Applied Statistics", pp 145-182, **Brooks Cole** (1992).

Hill, R.W., "On estimating the covariance matrix of robust regression M-estimates", **Commun. Statist.**, A8: 1183-1196 (1979).

Hill, R.W. and Holland, P.W., "Two robust alternatives to least squares regression", **J. Amer. Statist. Assoc.**, 72: 828-833 (1977).

Hogg, R.V., "Adaptive robust procedures: A partial review and some suggestions for future applications and theory", **J. Amer. Statist. Assoc.**, 69: 909-925 (1974).

Hogg, R.V., "An introduction to robust estimation", in R.L. Launer and G.N. Wilkinson (Eds.) *Robustness in Statistics*, **Academic**, New York (1979).

Huber, P.J., "Robust statistics: A review", **Ann. Math. Stat.**, 43: 1041-1067 (1972).

Huber, P.J., "Robust methods of estimation of regression coefficients", **Math. Operationsforsch. Stat.**, 8: 141-153 (1977).

Huber, P.J., "Robust Statistics", **Wiley**, New York (1981).

Jureckova, J., "Asymptotic relations of M-estimates and R-estimates in regression models", **Ann. Statist.**, 5: 464-472 (1977).

Krivy, I. and Tvrdik, J., "The Controlled Random Search Algorithm in Optimizing Regression Models", **Comput. Statist. Data Anal.** 20: 229-234 (1995).

Lipsitz, S.R., Fitzmaurice, G.M., Ibrahim, J.G., Sinha, D., Parzen, M. and Lipshultz, S., "Joint Generalized Estimating Equations", **Journal of the Royal Statistical Society**, Series A, 172(1): 3-20 (2009).

Marquardt, D.W., "An algorithm for least squares estimation of nonlinear parameters", **J. Soc. Ind. Appl. Math.**, 2: 431-441 (1963).

McCullagh, P., "Quasi-likelihood functions", **Ann. Stat.**, 11: 59-67 (1983).

McKee, J. and McClendon, J., "Multiple Regression and Causal Analysis", pp 230-271, **Peacock Publishing** (1994).

Miller, R.G.Jr., "Simultaneous Statistical Inference" 2nd ed., **McGraw-Hill**, New York (1981).

Nash, J.C., "Compact Numerical Methods for Computers: Linear Algebra and Function Minimisation", **Adam Hilger: Bristol & Halsted Press**, New York (1979).

Nash, J.C. and Walker-Smith, M., “Nonlinear Parameter Estimation”, **Marcel Dekker Inc.**, New York (1987).

Nelder, J.A. and Mead, R., “A simplex method for function minimization”, **Comput. J.** 7: 308-313 (1964).

Oberhofer, W., “The consistency of nonlinear regression minimizing the L1-norm”, **Ann. Stat.**, 10: 316-319 (1982).

Parkinson, J.M. and Hutchinson, D., “An investigation into the efficiency of variants on the simplex method”, in **Lootsma** pp: 115-135 (1972).

Price, W.L., “A controlled random search procedure for global optimization”, **Comput. J.** 20: 367-370 (1977).

Rardin R., “Optimization in Operations Research” **Prentice-Hall Inc. Simon & Schuster / A Viacom Company Upper Saddle River**, New Jersey (1998).

Seber, G.A.F. and Wild, C.J., “Nonlinear Regression”, **John Wiley and Sons**, New York (1989).

Sielken, R.L. and Hartley, H.U., “Two linear programming algorithms for unbiased estimation of linear models”, **J. Amer. Statist. Assoc.**, 68: 639-641 (1973).

Solis, F.J. and Wets, R.J.B., “Minimization by random search techniques”, **Math. Oper. Res.** 6: 19-30 (1981).

Spall, J.C., “Introduction to Stochastic Search and Optimization: Estimation, Simulation and Control”, **Wiley-Interscience**, New York (2003).

Storn, R. and Price, K., “Differential evolution—a simple and efficient heuristic for global optimization over continuous spaces”, **J. Global Optim.** 11: 341-359 (1997).

Tarakçı, M., “Doğrusal Programlama Tekniklerinin Regresyon Çözümlemesine Uygulaması”, **Hacettepe Üniversitesi Fen Bilimleri Enstitüsü İstatistik Anabilim Dalı Bilim Uzmanlığı Tezi**, Eylül (1992).

Tsou, T.S. and Royall, R.M., “Robust Likelihoods”, **Journal of the American Statistical Association**, 90(429): 316-320, Theory and Methods (1995).

Tvrđik, J. and Krivy, I., “Comparison of algorithms for nonlinear regression estimates”, in: Antoch (J.) (Ed.), COMPSTAT 2004. **Physica-Verlag**, Heidelberg, New York, pp: 1917-1924 (2004).

Tvrđik, J., Krivy, I. and Misik, L., “Evolutionary algorithm with competing heuristics”, in: Osmera (P.) (Ed.), MENDEL 2001, **Seventh International Conference on Soft Computing**, Technical University, Brno, pp:58-64 (2001).

Tvrdik, J., Krivy, I. and Misik, L., “Computing heuristic in evolutionary algorithms”, in: Sincak (P.) et al. (Eds.), *Intelligent Technologies-Theory and Applications*, IOS Press, Amsterdam, pp: 159-165 (2002).

Tvrdik, J., Krivy, I. and Misik, L., “Adaptive population-based search: Application to estimation of nonlinear regression parameters”, Computational Statistics and Data Analysis 52: 713-724 available online 9 november 2006, *Elsevier* (2006).

Wedderburn, R.W.M., “Quasi-likelihood functions, generalized linear models and the Gauss-Newton method”, *Biometrika*, 61: 439-447 (1974).

Winter, G., Galvan, B., Alonso, S., Gonzales, B., Jimenez, J. I. and Greimer, D., “A flexible evolutionary agent: cooperation and competition among real-coded evolutionary operators”, *Soft Comput.* 9: 299-323 (2005).

Wolpert, D.H. and Macready, W.G., “No free lunch theorems for optimization”, IEEE *Trans. Evol. Comput.* 1: 67-82 (1997).

Woo, M.J. and Sriram, T.N., “Robust Estimation of Mixture Complexity”, *Journal of the American Statistical Association*, 101(476): 1475-1486, Theory and Methods (2006).

EKLER

EK-1 Normal Denklemler

Bileşen sayıları n olmak üzere $a = (a_1, a_2, \dots, a_n)$ ve $b = (b_1, b_2, \dots, b_n)$ gibi iki vektörün skalar çarpımları $a \cdot b = a_1 b_1 + a_2 b_2 + \dots + a_n b_n = 0$ ile sıfır ise bu iki vektör ortogonaldır veya normaldir, birbirlerine diktirler. Örneğin n bileşenli bir r vektörünün tüm bileşenleri t değişkeninin bir fonksiyonu ise bu durumda t değişkenine bağımlı vektör fonksiyonu $r = (r_1(t), r_2(t), \dots, r_n(t))$ şeklinde ifade edilir.

W simetrik bir matris olmak üzere bir amaç fonksiyonu $S = r^t W r$ şeklinde kurulup amaç fonksiyonunun ekstremumlarının aranmasıyla elde edilen

$$\frac{dS}{dt} = 2 \left(\frac{dr}{dt} \right)^t W r \rightarrow \left(\frac{dr}{dt} \right)^t W r = 0 \quad \text{normal/ortogonal denklemin çözümüyle}$$

ekstremum noktaları bulunmuş olur. Burada $\left(W \frac{dr}{dt}, r \right)$ veya $\left(\frac{dr}{dt}, W r \right)$ vektör çifleri normal/ortogonaldır. Eğer r vektörünün tüm bileşenleri p farklı değişkene bağımlı ise bileşenler $(t_1, t_2, \dots, t_p)$ vektörüne bağımlı çokdeğişkenli birer fonksiyondur, bu durumda r vektörü fonksiyonu $(t_1, t_2, \dots, t_p)$ vektörüne bağımlıdır. Böylece t

$$\text{değişkenlerinin sayısına bağlı olarak } p \text{ tane } \left(\frac{\partial r}{\partial t_j} \right)^t W r = 0 \text{ normal denklemin}$$

çözümüyle ekstremum noktalar bulunabilir. Normal denklemler, bir sınıf amaç fonksiyonunun ekstremumlarının araştırması şeklinden kaynaklanmayabilir. Birçok olgu normal denklemlerin oluşmasına neden olabilir. Genel olarak bir sistem anlamında p tane denklemden oluşan normal denklemler $\sum_i \omega_{ij}(\theta) \psi_i(\theta) = 0$ şeklinde

ifade edilebilir. Eğer j .ninci normal denklem için ω_{ij} fonksiyonu bir φ_i

fonksiyonunun θ_j değişkenine göre kısmi türevi $\omega_{ij} = \frac{\partial \varphi_i}{\partial \theta_j}$ şeklinde ise normal

denklemleri sağlayan bir yakınsama noktası bulunabilir. Bu durumda $j = 1, 2, \dots, p$

için $\sum_i \frac{\partial \varphi_i}{\partial \theta_j} \psi_i = 0$ normal denklemlerinin çözümüne iterasyonlarla varılabilir. Bir

EK-1 (Devam) Normal Denklemler

dengeleme $\sum_i \frac{\partial \varphi_i}{\partial \theta_j} \frac{\psi_i}{\varphi_i} \varphi_i = 0$ şeklinde yapılip k.nıncı adımda φ_i fonksiyonu

$\theta_k = (\theta_{1k}, \theta_{2k}, \dots, \theta_{pk})$ noktası civarında Taylor serisine açılır ve diğer çarpanlar θ_k

noktasına yaklaştırılırsa $\sum_i \left(\frac{\partial \varphi_i}{\partial \theta_j} \frac{\psi_i}{\varphi_i} \right) \Big|_{\theta_k} [\varphi_i(\theta_k) + \nabla \varphi_i(\theta_k)^t (\theta - \theta_k)] = 0$ denklemleri

ortaya çıkar. Buradan p denklem için $\frac{\partial \varphi}{\partial \theta^t} \Big|_{\theta_k} = -X_k$, $W_k = \text{diag} \{w_{ik}\}_{i=1}^n$ ve

$w_{ik} = \frac{\psi_i(\theta_k)}{\varphi_i(\theta_k)}$ terimleri tamamlanır ve ifadeden θ çekilip bu θ_{k+1} olarak

adlandırılırsa $\theta_{k+1} = \theta_k + (X_k^t W_k X_k)^{-1} X_k^t W_k \varphi(\theta_k)$ Gauss-Newton iterasyonları belirir.

Özel olarak $\psi_i = 1$ olması durumunda $\sum_i \frac{\partial \varphi_i}{\partial \theta_j} = 0$ toplamı $\sum_i \frac{\partial \varphi_i}{\partial \theta_j} \frac{1}{\varphi_i} \varphi_i = 0$ şeklinde

ele alınarak bir yaklaştırmayla $\sum_i \left(\frac{\partial \varphi_i}{\partial \theta_j} \frac{1}{\varphi_i} \right) \Big|_{\theta_k} [\varphi_i(\theta_k) + \nabla \varphi_i(\theta_k)^t (\theta - \theta_k)] = 0$

denklemleri belirir. Benzer şekilde p denklem için $\frac{\partial \varphi}{\partial \theta^t} \Big|_{\theta_k} = -X_k$, $W_k = \text{diag} \{w_{ik}\}_{i=1}^n$

ve $w_{ik} = \frac{1}{\varphi_i(\theta_k)}$ terimleri tamamlanırsa $\theta_{k+1} = \theta_k + (X_k^t W_k X_k)^{-1} X_k^t W_k \varphi(\theta_k)$

iterasyonları ile yakınsama noktasına varılır. Buradaki iki benzer yaklaşımla bütün normal denklemlerin yakınsama noktaları gerekiyorsa uyarlamalar yapılarak bulunabilir.

Özellikle $T = \sum_i \varphi_i \rightarrow \frac{\partial T}{\partial \theta_j} = 0$ kaynağına bağlı olarak ortaya çıkan $\sum_i \frac{\partial \varphi_i}{\partial \theta_j} = 0$

ifadesi sayısal yöntemlerle integrasyonu işaret ederek beklenen fonksiyonun diferansiyel denklemlerin veya diferansiyel denklem sistemlerinin çözümü şeklinde olduğu doğrusal olmayan regresyon modellerinin optimal parametrelerinin bulunabilmesinin bir yolunu açar. Bu yönde her şekilde genelleştirmeler yapılabilir (Parametre tahminleri ve parametre tahminlerinin standart hata tahminleri Gauss-

EK-1 (Devam) Normal Denklemler

Newton iterasyonlarının işlendiği yaklaşımlarda olduğu gibi bulunabilmektedir).

Genel olarak $\sum_i \omega_{ij}(\theta) \psi_i(\theta) = 0$ şeklinde p tane denklemden oluşan normal denklemlerde değişken $\psi_i(\theta)$ konumundaki fonksiyonlar bir θ_k noktası civarında Taylor serisine açılıp katsayı $\omega_{ij}(\theta)$ konumundaki fonksiyonlar bu θ_k noktasına yaklaştırılırsa $\sum_i \omega_{ij}(\theta_k) [\psi_i(\theta_k) + \nabla \psi_i(\theta_k)^t (\theta - \theta_k)] = 0$ denklemleri elde edilir. Eğer normal denklemlerin bir çözümü olan yakınsama noktası θ_k noktasına yakınsa elde edilen denklemlerin θ_{k+1} çözümü bu yakınsama bölgesi içindedir. $\frac{\partial \psi}{\partial \theta^t} |_{\theta_k} = -X_k$ ve $\Omega_k = [\omega_{ij}(\theta_k)]$ gösterimleri ile elde edilen denklemler matris notasyonu $\Omega_k^t [\psi(\theta_k) - X_k(\theta - \theta_k)] = 0$ şeklinde topluca yazılarak θ çekilir ve bu θ_{k+1} olarak gösterilirse $\psi(\theta_k) = \psi_k$ ile $\theta_{k+1} = \theta_k + (\Omega_k^t X_k)^{-1} \Omega_k^t \psi_k$ iterasyonları elde edilir.

ÖZGEÇMİŞ

Kişisel Bilgiler

Soyadı, adı : SERİN, Tarhan
 Uyuğu : T.C.
 Doğum tarihi ve yeri : 20 / 07 / 1968 Eskişehir
 Medeni Hali : Bekar
 Telefon : 0 (312) 426 97 94
 e-mail : tarhanserin@mynet.com

Eğitim Derece

Eğitim Birimi

Mezuniyet tarihi

Doktora	Gazi Üniversitesi / İstatistik Bölümü	2010
Yüksek Lisans	Gazi Üniversitesi / İstatistik Bölümü	2004
Yüksek Lisans	Gazi Üniversitesi / Endüstri Müh. Böl.	1999
Lisans	Gazi Üniversitesi / Endüstri Müh. Böl.	1995

Yabancı Dil

İngilizce