

UKUROVA NİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

DOKTORA TEZİ

Orkun COŞKUNTUNCEL

**KARMA DENEMELERDE VE MODELLERDE ROBUST İSTATİSTİKSEL
ANALİZLER**

MATEMATİK ANABİLİM DALI

ADANA, 2005

ÇUKUROVA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

KARMA DENEMELERDE VE MODELLERDE ROBUST
İSTATİSTİKSEL ANALİZLER

Orkun COŞKUNTUNCEL
DOKTORA TEZİ
MATEMATİK ANABİLİM DALI

Bu tez 06 / 07 / 2005 Tarihinde Aşağıdaki Jüri Üyeleri Tarafından Oybirliği ile Kabul Edilmiştir.

İmza.....

İmza.....

İmza.....

Prof.Dr. Olcay ARSLAN
SAKALLIOĞLU Doç.Dr. Rızvan EROL
DANIŞMAN ÜYE

Prof.Dr. Sadullah
ÜYE

İmza.....
Prof.Dr. Naime EKİCİ
ÜYE

İmza.....
Yrd.Doç.Dr. Gökhan ÇUVALCIOĞLU
ÜYE

Bu tez Enstitümüz Matematik Anabilim Dalında hazırlanmıştır.

Kod No:

Prof.Dr Aziz ERTUNÇ
Enstitü Müdürü
İmza ve Mühür

Bu çalışma Çukurova Üniversitesi Araştırma Projesi Birimi tarafından desteklenmiştir.

Proje No: FBE2004D3

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

İÇİNDEKİLER

	<u>Sayfa</u>
ÖZ.....	I
ABSTRACT.....	II
TEŞEKKÜR.....	III
TABLO DİZİNİ.....	IV
ŞEKİL DİZİNİ.....	VII
SİMGE VE KISALTMALAR.....	VIII
1. GİRİŞ.....	1
1.1. Karma Denmelere Bir Bakış.....	1
1.2. Robust İstatistiğe Giriş.....	12
1.2.1. Tanım ve Özellikler.....	13
1.3. Karma Denemelerde Simpleks Latis Tasarımı ve Kanonik Polinom.....	15
1.3.1. $\{q, m\}$ Simpleks Latisin Tasarım Noktalarında Amaç (Yanıt) Fonksiyonu Olarak Kanonik Polinom Katsayıları ve Parametrelerin Tahmini.....	17
1.3.2. $\{3, 2\}$ Simpleks Latis Tasarımına Bir Örnek.....	21
1.4. Konum ve Ölçek Parametreleri için M Tahmin Edicisi.....	24
1.4.1. M tahmin Edicisi İçin Bir Örnek.....	30
2. ROBUST REGRESYON VE KATSAYI TAHMİNİ.....	34
2.1. Robust M Regresyon Tahmin Edicileri.....	35
2.1.1. Hataların Dağılım Formunun Bilinmemesi Durumu.....	37
2.1.2. Hataların Dağılım Formunun Bilinmesi Durumu.....	41
2.1.2.1. Hataların t Dağılımından Geldiğinin Varsayılması.....	43
2.1.2.2. Hataların GT Dağılımından Geldiğinin Varsayılması.....	46
2.1.2.3. Hataların SGT Dağılımından Geldiğinin Varsayılması.....	48
2.2. Örnekler.....	51
3. ROBUST RIDGE VE ROBUST LIU TAHMİN EDİCİLERİ.....	62
3.1. Kötü Koşulluluğu Belirleme Yöntemleri ve Standartlaştırma.....	63
3.2. Alışılmış Ridge ve Alışılmış Liu Tahmin Edicileri.....	68
3.3. Robust Ridge ve Robust Liu Tahmin Edicileri.....	71

3.4. Şiddetli Çoklu İç İlişki ve Sapan Değer Problemlerinin Birlikte Olması	
Durumu için Bir Örnek: Motor Yağı Karışımı Verisi.....	79
4. GENELLEŞTİRİLMİŞ M (GM) TAHMİN EDİCİLERİ.....	90
4.1. X Yönündeki Sapan Değerleri Tanımlama Yöntemleri.....	90
4.2. Genelleştirilmiş M (GM) Tahmin Edici.....	99
4.3. GM Tahminine Dayalı Ridge ve Liu Tahmin Edicileri.....	109
5. KARMA DENEMELERDE PSUDOBİLEŞEN DÖNÜŞÜMÜ VE ROBUST TAHMİN EDİCİLER.....	114
5.1. Sınırların Tutarlılığı ve Psudobileşen Dönüşümü.....	114
5.2. Robust Tahmin Edicilerin Psudobileşen Dönüşümü Yapılmış Karma Veriye Etkilerinin İncelemesi.....	118
5.3. Motor Yağı Karışımı Verisi.....	129
6. SONUÇLAR VE ÖNERİLER.....	141
KAYNAKLAR.....	144
ÖZGEÇMİŞ.....	152

ÖZ

DOKTORA TEZİ

**KARMA DENEMELERDE VE MODELLERDE ROBUST
İSTATİSTİKSEL ANALİZLER**

Orkun COŞKUNTUNCEL

ÇUKUROVA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
MATEMATİK ANABİLİM DALI

Danışman: Prof.Dr. Olcay ARSLAN

Yıl: 2005, Sayfa: 165

Jüri : Prof.Dr. Olcay ARSLAN

Prof.Dr. Sadullah SAKALLIOĞLU

Doç.Dr. Rızvan EROL

Prof.Dr. Naime EKİCİ

Yrd.DoçDr. Gökhan ÇUVALCIOĞLU

Karma denemeler kanonik polinom olarak adlandırılan polinom modelinin özel bir haline gereksinim duyarlar. Genellikle, karma denemelerde bileşenler üzerinde, fiziksel, kimyasal ve ekonomik nedenlerden kaynaklanan, alt ve üst sınırlar şeklinde ek kısıtlamalar vardır. Bu ek kısıtlamalar çoğu zaman kötü koşulluluk probleminin oluşmasına neden olur. Bunun yanı sıra karma veride x ve/veya y yönünde sapan değer problemi de bulunabilir. Bu durumda en sık kullanılan en küçük kareler tahmin edicisi güvenilir sonuçlar vermeyecektir. Bu çalışmanın amacı karma verinin kötü koşullulukla birlikte sapan değere sahip olması durumunda en küçük karelere alternatif daha güvenilir regresyon katsayı tahmini elde edebilmektir. Bu tip karma veriler için sapan değerlere karşı dayanıklı olan robust tahmin ediciler ile Ridge ve Liu tahmin edicileri birlikte kullanılarak elde edilen robust Ridge ve robust Liu tahmin edicilerinin performansları incelenmiştir.

Anahtar Kelimeler: Karma denemeler, Liu tahmin edicisi, Ridge regresyon, M tahmin edici, GM tahin edici.

ABSTRACT

PhD. THESIS

**ROBUST STATISTICAL ANALYSIS FOR MIXTURE EXPERIMENT
AND MODELS**

Orkun COŞKUNTUNCEL

DEPARTMENT OF MATHEMATICS
INSTITUTE OF NATURAL AND APPLIED SCIENCES
UNIVERSITY OF ÇUKUROVA

Supervisor: Prof.Dr. Olcay ARSLAN

Year: 2005, Pages: 165

Jury : Prof.Dr. Olcay ARSLAN

Prof.Dr. Sadullah SAKALLIOĞLU

Doç.Dr. Rızzvan EROL

Prof.Dr. Naime EKİCİ

Yrd.Doç.Dr. Gökhan ÇUVALCIOĞLU

Experiments with mixtures require a special form of polynomial model called canonical polynomial model. In general, additional constraints in the form of lower and upper bounds are placed on the component because of physical, chemical and economic reasons. These additional constraints often cause multicollinearity. One of the estimators that are used to estimate the regression parameters for the mixture data is the ordinary least square estimator. When the mixture data have outliers and/or multicollinearity problem the performance of the LS estimator will be poor. The aim of this study is to propose estimators for the model parameter of a mixture models that can combat with the multicollinearity and the outliers simultaneously. We will combine the robust M and GM estimators with a biased estimator such as Ridge. Since the combined estimator (the resulting hybrid estimator) will be robust against the outlier in y and/or x direction and will deal with the multicollinearity in the design matrix X it will be more stable estimator than the OLS. Examples are given to show the performance of combined estimator when the outliers and multicollinearity are present for experiment with mixture data.

Key Words: Experiments with mixtures, Liu estimator, Ridge regression, M estimator, GM estimator

TEŞEKKÜR

Bu çalışmanın hazırlanmasında değerli bilgileri ile bana ışık tutan, manevi desteğini hiçbir koşulda esirgemeyen çok değerli hocam Prof.Dr. Olcay ARSLAN'a teşekkürlerimi bir borç bilirim. Çalışmalarım sırasında manevi desteklerini eksik etmeyen saygıdeğer hocalarım Prof.Dr. Naime EKİCİ'ye ve Yrd.Doç.Dr. Zerrin ESMERLİĞİL'e ve tüm çalışma arkadaşlarıma teşekkürler.

İşlerimin yoğun olduğu günlerde bana büyük bir sabır gösteren, her sabah uykusunu bölerek beni işime öperek uğurlayan canım oğlum Spiderman Sabit'e ve Ülkü'me, her konuda bana destek veren Babama, Anneme ve kardeşlerime, minicik yüreklerinde benim için büyük bir yer ayıran Ersoy ve İlayda'ya, beni her zaman bağırlarına basan Matematik Bölümü öğrencilerine ve herkese teşekkürler....

TABLO DİZİNİ

	<u>Sayfa</u>
Tablo 1.1. Beş karma deneme sonunda elde edilen sonuçlar.....	3
Tablo 1.2. Meyve suyu karışımı verileri.....	21
Tablo 1.3. Meyve suyu karışımı verisi. Verilere yeni eklenen gözlemler.....	22
Tablo 1.4. Yeni veriler için robust tahmin ediciler ile elde edilen sonuçlar.....	30
Tablo 1.5. Bazı dağılımlar ve ρ fonksiyonları için etki ve ağırlık fonksiyonları.....	32
Tablo 2.1. 20 gözlemlilik karma veri seti.....	52
Tablo 2.2. σ verilerden hesaplandığında katsayı tahminleri.....	54
Tablo 2.3. σ verilerden hesaplandığında elde edilen ağırlıklar.....	55
Tablo 2.4. β ve σ eşanlı olarak tahmin ediliyorken katsayı tahminleri.....	56
Tablo 2.5. σ eşanlı hesaplanıyorken elde edilen ağırlıklar.....	56
Tablo 2.6. 20 gözlemlilik karma veri seti.....	58
Tablo 2.7. SGT için katsayı ve parametre tahminleri.....	60
Tablo 3.1. Tablo 2.1’de verilen karma verinin standartlaştırılmış şekli.....	67
Tablo 3.2. 20 gözlemlilik orijinal veriler için sonuçlar.....	68
Tablo 3.3. 2 sapan değer çıkarılmış 18 gözlemlilik veri için elde edilen sonuçlar.....	70
Tablo 3.4. 18 ve 20 gözlemlilik veriler için Ridge ve Liu Tahminleri.....	71
Tablo 3.5. 20 gözlemlilik veri için robust M regresyon tahmin edicileri.....	77
Tablo 3.6. Robust tahminlerin katsayıları için standart hatalar.....	78
Tablo 3.7. 20 gözlemlilik veri için robust Ridge ve robust Liu tahminleri.....	78
Tablo 3.8. Snee(1975)’in motor yağı karışımı verisi.....	80
Tablo 3.9. Motor yağı karışımı verisi sonuçları.....	81
Tablo 3.10. Bozulmuş gözlemler için elde edilen sonuçlar.....	82
Tablo 3.11. 1 sapan değeri atılmış 17 gözlemlilik veriler için en küçük kareler.....	83
Tablo 3.12. 18 gözlemlilik bozulmuş veriler için M tahmin edicileri.....	83
Tablo 3.13. Robust tahmin edicileri için standart hatalar.....	84
Tablo 3.14. Orijinal veriler için EKK ve Ridge regresyon sonuçları.....	85
Tablo 3.15. 17 ve 18 gözlemlilik veriler için Ridge regresyon sonuçları.....	86
Tablo 3.16. 18 gözlemlilik veriler için robust Ridge regresyon sonuçları.....	87
Tablo 3.17. Orijinal veriler için EKK ve Liu regresyon sonuçları.....	88

Tablo 3.18. 17 ve 18 gözlemleri veriler için Liu regresyon sonuçları.....	88
Tablo 3.19. 18 gözlemleri veriler için robust Liu regresyon sonuçları.....	89
Tablo 4.1. Motor yağı karışımı verisi için teşhis sonuçları.....	98
Tablo 4.2. 16 gözlemleri motor yağı karışımı verisi en küçük kareler sonuçları.....	99
Tablo 4.3. Motor yağı karışımı verisi için M tahmin edicileri.....	105
Tablo 4.4. M tahmin edicileri için katsayıların standart hataları.....	106
Tablo 4.5. Robust tahmin ediciler için ağırlıklar.....	107
Tablo 4.6. Robust M tahmin edicileri için MAD ile standartlaştırılmış hatalar.....	107
Tablo 4.7. Mallows tipi GM tahmini için elde edilen sonuçlar.....	108
Tablo 4.8. Schweppe tipi GM tahmini için elde edilen sonuçlar.....	109
Tablo 4.9. Motor yağı karışımı verisi için en küçük kareler ve alışılmış Ridge tahmin edicisi ile elde edilen sonuçlar.....	110
Tablo 4.10. Huber M tahminine dayalı robust Ridge tahmini sonuçlar.....	111
Tablo 4.11. Motor yağı karışımı verisi için GM tahminlerine dayalı robust Ridge tahmin edicisi sonuçları.....	111
Tablo 4.12. Motor yağı karışımı verisi için en küçük kareler ve alışılmış Liu tahmin edicisi ile elde edilen sonuçlar.....	112
Tablo 4.13. Huber M tahminine dayalı robust Liu tahmini sonuçları.....	112
Tablo 4.14. Motor yağı karışımı verisi için GM tahminlerine dayalı robust Liu tahmin edicisi sonuçları.....	113
Tablo 5.1. Parlayarak yanma verisi.....	119
Tablo 5.2. Standartlaştırılmış orijinal parlayarak yanma verisi için en küçük kareler sonuçları.....	120
Tablo 5.3. Psudobileşen dönüşümü ve standartlaştırma işlemi uygulanmış parlayarak yanma verisi için en küçük kareler sonuçları.....	120
Tablo 5.4. Parlayarak yanma verisi için teşhis sonuçları.....	121
Tablo 5.5. 14 gözlemleri parlayarak yanma verisi sonuçları.....	123
Tablo 5.6. Parlayarak yanma verisi için robust M tahminleri.....	124
Tablo 5.7. Parlayarak yanma verisi katsayıların standart hataları ve $\hat{A}$ değerleri...	125
Tablo 5.8. Robust M tahmin edicileri için ağırlıklar ve hatalar.....	125
Tablo 5.9. Parlayarak yanma verisi için GM tahmin edicileri.....	126

Tablo 5.10. Parlayarak yanma verisi için Ridge tahmin edicisi sonuçları.....	127
Tablo 5.11. Parlayarak yanma verisi için robust Liu tahmin edicisi sonuçları.....	128
Tablo 5.12. Parlayarak yanma verisi için GM tahminine dayalı Ridge ve Liu.....	129
Tablo 5.13. Psüdobileşen dönüşüm uygulanmış motor yağı karışımı verisi.....	130
Tablo 5.14. Psüdobileşen dönüşümü ve standartlaştırma işlemi uygulanmış motor yağı karışımı verisi için en küçük kareler sonuçları.....	130
Tablo 5.15. Motor yağı karışımı verisi için teşhis sonuçları.....	131
Tablo 5.16. Motor yağı karışımı verisi için robust M tahmin edicileri.....	132
Tablo 5.17. Katsayıların standart hataları ve $\hat{\sigma}$ değerleri.....	132
Tablo 5.18. Robust M tahmin edicileri için hatalar ve ağırlıklar.....	133
Tablo 5.19. Motor yağı karışımı verisi için GM tahmin edicileri.....	134
Tablo 5.20. Motor yağı karışımı verisi için alışılmış Ridge ve Liu tahminleri.....	135
Tablo 5.21. Motor yağı karışımı verisi için robust Ridge ve Liu tahminleri.....	135
Tablo 5.22. Mallows tipi GM tahminine dayalı robust Ridge ve Liu tahminleri.....	136
Tablo 5.23. Schweppe tipi GM tahminine dayalı robust Ridge ve Liu tahminleri...	136
Tablo 6.1. Psüdobileşen dönüşümü yapılmış yakıt karışımı verisi.....	140
Tablo 6.2. Yakıt karışımı verisi için ANOVA ve Model uygunluğu testi.....	141
Tablo 6.3. Özel kübik ve karesel modeller için en küçük kareler sonuçları.....	142

ŞEKİLLER DİZİNİ

Sayfa

Şekil 1.1. %50A-%50B benzin karışımının yol grafiği.....	4
Şekil 1.2. $q = 2$, $x_1 + x_2 = 1$ için faktör uzayı bir boyutlu simplekstir. Tüm deneysel noktalar doğru üzerinde olmalıdır.....	5
Şekil 1.3. $q = 3$, $x_1 + x_2 + x_3 = 1$ için faktör uzayı iki boyutlu simplekstir. Tüm deneysel noktalar eşkenar üçgenin sınırları üstünde veya içinde olmalıdır...	6
Şekil 1.4. Karma denemelerde üçgensel koordinatlar.....	6
Şekil 1.5. Üçgensel koordinatlarda doğru ve noktanın gösterilmesi.....	7
Şekil 1.6. En küçük kareler ve robust yöntemlerle uydurulan doğru.....	13
Şekil 1.7. $\{3, 2\}$ ve $\{3, 3\}$ simpleks latislerindeki tasarım noktalarında amaç terminolojisi.....	18
Şekil 1.8. Uydurulan modeller için yüzey ve kontur grafikleri.....	28
Şekil 1.9. Huber, Tukey ve Welsch ρ , ψ ve ağırlık fonksiyonlarının grafikleri.....	33
Şekil 2.1. Serbestlik derecesi $v = 3$ ve $v = 1$ için t dağılımı.....	44
Şekil 2.2. Şekil parametreleri p ve q 'nun değişik değerleri için GT dağılımı.....	46
Şekil 2.3. Çeşitli parametre değerleri için SGT dağılımı.....	48
Şekil 2.4. Tablo 2.5'teki tasarım noktalarının tekrar sayısı, üçgensel koordinatlardaki yerleşimi ve y_i gözlemleri için histogram.....	52
Şekil 2.5. (2.52)'deki model için normal olasılık grafiği (Design Expert 6).....	53
Şekil 2.6. Standartlaştırılmış rezidüleri için indeks grafiği.....	54
Şekil 2.7. Yüzey Grafikleri (a) 20 gözlem (b) 18 gözlem (c) GT.....	57
Şekil 2.8. Tablo 2.6'daki tasarım noktalarının tekrar sayısı, üçgensel koordinatlardaki yerleşimi ve y_i gözlemleri için histogram.....	58
Şekil 2.9. (2.54)'deki model için normal olasılık grafiği (Design Expert 6).....	59
Şekil 2.10. Standartlaştırılmış rezidüleri için indeks grafiği.....	59
Şekil 2.11. Yüzey Grafikleri (a) en küçük kareler (b) GT (c) SGT.....	61
Şekil 4.1. a) iyi leverage nokta b) dikey sapan değer c) kötü leverage nokta.....	91
Şekil 4.2. M ve GM tahmin edicilerinin aykırı gözlemlere karşı davranışları (Hampel, F.R. ve ark., 1986).....	103
Şekil 5.1. Üç bileşenli karma sistem için deneysel bölge	116
Şekil 6.1. Tasarım noktalarının üçgensel koordinatlardaki yerleşimleri.....	141

SİMGE VE KISALTMALAR

$\hat{\beta}_{LS}$	β regresyon parametresinin en küçük kareler tahmini
$\hat{\beta}_R$	Ridge regresyon tahmini
$\hat{\beta}_L$	Liu tahmin edicisi
$\hat{\beta}_M$	β regresyon parametresinin M tahmin edici
$\hat{\beta}_{GMM}$	β regresyon parametresinin Mallows tipi GM tahmin edici
$\hat{\beta}_{GMS}$	β regresyon parametresinin Schweppe tipi GM tahmin edici
$\hat{\beta}_{RM}$	M tahminine dayalı Ridge tahmini
$\hat{\beta}_{LM}$	M tahminine dayalı Liu tahmini
$\hat{\beta}_{RGMM}$	Mallows tipi GM tahminine dayalı Ridge tahmini
$\hat{\beta}_{RGMS}$	Schweppe tipi GM tahminine dayalı Ridge tahmini
$\hat{\beta}_{LGMM}$	Mallows tipi GM tahminine dayalı Liu tahmini
$\hat{\beta}_{LGMS}$	Schweppe tipi GM tahminine dayalı Liu tahmini
MD	Mahalanobis Uzaklığı
RMD	Robust Mahalanobis Uzaklığı
Huber	β regresyon parametresinin Huber M tahmin edicisi
Tukey	β regresyon parametresinin Tukey M tahmin edicisi
t	β regresyon parametresinin t dağılımına dayalı M tahmin edicisi
GT	β regresyon parametresinin GT dağılımına dayalı M tahmin edicisi
SGT	β regresyon parametresinin SGT dağılımına dayalı M tahmin edicisi
$\hat{\sigma}$	σ ölçek parametresinin tahmini

1. BÖLÜM. GİRİŞ

1.1. Karma Denmelere Bir Bakış

Karma denemeler ile ilgili istatistiksel araştırmalar Quenouille'nin 1953 yılında çıkan "The Design and Analysis of Experiment" adlı kitabındaki kısa öneri notu ile başlamıştır. Ardından 1955 yılında Claringbold "Use of the simplex design in the study of the joint action of related hormones" adlı makalesinde karma denemeleri ele almıştır. Ancak karma denemelerle ilgili olarak yapılan ilk kapsamlı çalışma 1958 yılında Henry Scheffé'nin "Experiments with mixtures" adlı makalesidir. Bu çalışmada Scheffé, simpleks latis tasarımı ve bu tasarıma uygun polinom modelini vermiştir. Bu makale ile birlikte karma denemelerdeki araştırmalar hızlı bir gelişim sürecine girmiştir. Bunu takiben 1963 yılında Scheffé, simpleks sentroid tasarım ve uygun polinom modelini vermiştir. Bu tasarımlar günümüzde halen sıkça kullanılmaktadır.

Karma denemeler amaç yüzey (response surface) denemelerinin özel bir halidir. Amaç yüzey çalışmalarında deneysel tasarımdaki herhangi bir faktör için seçilen düzey, diğer faktörlerden bağımsızdır. Örneğin, karışma oranı, reaksiyon zamanı ve sıcaklık değişkenlerinden oluşan üç bileşenli bir kimyasal deneyi düşünelim. Bu deneydeki sıcaklık düzeyi reaksiyon zamanı ve karışma oranından bağımsız olarak seçilebilir. Karma denemelerde ise faktörler karışımı oluşturan bileşenlerin oranlarıdır ve amaç bileşenlerin miktarlarının değil, her birinin oranının fonksiyonudur.

Günlük hayatımızda kullandığımız birçok ürün iki veya daha fazla malzemenin ya da maddenin değişik oranlarda karıştırılmasıyla elde edilir. Örneğin, Un, yağ, şeker, kabartma tozu ve su kullanılarak yapılan kek karışımı. Kum, su ve bir ya da daha fazla çeşit çimentonun karıştırılmasıyla yapılan beton. Portakal, limon ve havuç sularının karıştırılmasıyla elde edilen meyve suyu. Gümüş tozu, bağlayıcı, çözücü ve dengeleyici maddelerin karıştırılmasıyla elde edilen fotoğraf film tabakası. Tat ve koku veren katkı maddelerinin bulunduğu tütün karışımı. Verdiğimiz örneklerde karışımın oluşturulmasında her ürünün en az bir ya da daha fazla

özelliğiyle ilgilenilmektedir. Bütün örneklerde oluşturulan ürünün ölçülecek özelliği formülde görülen bileşenlerin oranına veya yüzdesine bağlıdır.

Karma denemelerde iki veya daha fazla maddenin değişik oranlarda karıştırılmasının nedeni, elde edilen ürünün karışımı oluşturan maddelerin tek tek kullanılmasından daha iyi sonuçlar verebilmesidir. Örneğin, A, B ve C olarak adlandırılan 3 farklı rafineriden alınan benzin stokumuz olsun ve bunların tek tek veya çeşitli oranlardaki karışımlarıyla ilgilenelim. Bu durumda; %50A-%50B veya %50B-%50C veya %50A-%50C, %33A-%33B-%33C veya %25A-%75B veya %25B-%75C veya %25A-%75C gibi çeşitli oranlardaki karışımlarının tek başına A, B ve C benzinlerinden daha iyi sonuç verip vermediği araştırılmaktadır. Araştırma sonunda en iyi sonucu veren karışım belirlenir. Ayrıca bu belirleme ya da seçim yapılırken maliyet, kullanılabilirlik gibi etmenler de göz önünde bulundurulur. Burada yanıtlanması gereken soru: “Eğer A, B ve C benzinlerinin farklı oranlardaki karışımları, fiziksel, kimyasal ve ekonomik şartlar göz önünde bulundurularak, sadece A, B ve C benzinlerinin kullanılmasından daha iyi sonuç verebilir mi?” (Cornell, 1990).

Karma denemelerde bileşenlerin oranları değiştirilerek ürünün ilgilenilen özelliği değiştirilebilir. Deneysel bakımdan, genellikle, ölçülebilir özellik veya ölçülebilir amaç (beton karışımının sertliği gibi) ile kontrol edilebilir değişkenler (beton karışımında bulunan kum, su, çimento bileşenlerinin oranları gibi) arasındaki fonksiyonel ilişki incelenir. Bu incelemenin iki nedeni vardır. Bunlar,

- i) Karışımı oluşturan maddelerin hangi oranlardaki karışımlarının en iyi olduğunu belirlemektir.
- ii) Karışımı oluşturan bileşen oranlarını değiştirerek tüm sistem için en iyi ya da optimum durumu belirlemektir.

Karma denemelerle ilgili söylediklerimizi matematiksel olarak ifade etmek için verdiğimiz benzin karışımı deneyini tekrar ele alalım. A ve B olarak adlandırılan iki farklı rafineriden alınan benzin stokumuzun olduğunu düşünelim. Burada ilgileneceğimiz özellik aynı test aracı ile aynı miktar A veya B veya bunların

değişik oranlardaki karışımlarının yaptıkları yol miktarıdır. A benzininin 1 galonu (1 galon = 3,79 litre) ile ortalama 13 mil (1 mil = 1,61 kilometre) yol yapıldığı ve B benzininin 1 galonu ile ortalama 7 mil yol yapıldığı bilinmektedir. Test aracının deposuna 1 galon A ve 1 galon B benzini koyulduğunda $13 + 7 = 20$ mil yol yapılabileceğini umarız. Yani, %50A-%50B benzin karışımının 1 galonu ile ortalama $20/2 = 10$ mil yol yapılacağını bekleriz. Yanıtlanması gereken soru, “A ve B benzinlerinin %50A-%50B veya %33A-%67B gibi değişik oranlardaki karışımlarını kullanarak elde ettiğimiz ortalama yol miktarının, A ve B benzinlerini tek tek kullandığımızda ulaştığımız ortalama yol miktarından daha fazla olup olmadığıdır”.

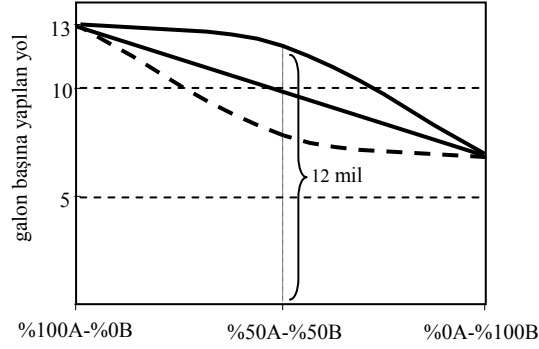
Bu soruyu yanıtlamak için test aracının deposuna %50A-%50B oranlarında benzin karışımından oluşan 2 galon benzin koyulmuş ve benzin bitene kadar gidilmiştir. Bu şekilde 5 deneme yapılmıştır, deney 5 kez tekrarlanmıştır. Denemelerden elde edilen sonuçlar Tablo 1.1’de verilmiştir. Beş deneme sonucunda yapılan ortalama yol, 12 mildir.

Tablo 1.1. Beş karma deneme sonunda elde edilen sonuçlar.

Deneme	%50-%50 benzin karışımının 2 galonu ile yapılan yol (mil)	1 galon karışım için ortalama yol (mil)
1	24,6	12,30
2	23,3	11,65
3	24,3	12,15
4	23,1	11,55
5	24,7	12,35
Ortalama yol = 12.00		

Elde edilen 12 millik ortalama yol iki benzin çeşidinin basit ortalamasından elde edilen 10 milden daha yüksektir. Böylece A ve B benzinlerinden bir karışım yapıldığında birbirlerini tamamladıklarını söyleyebiliriz. Eğer A ve B benzinlerinin her karışımı için elde edilen ortalama yol miktarı basit ortalamadan daha büyükse bu durum Şekil 1.1’deki düz çizgi ile gösterilen eğri ile verilir. Eğer elde ettiğimiz ortalama yol miktarı ile basit ortalama eşit ise, yani %50A-%50B benzin karışımında 10 millik bir ortalama, %33A-%67B benzin karışımından $[(13 \times \%33) + (7 \times \%67)]/100 = 9$ millik bir ortalama elde ediliyorsa bu durum Şekil 1.1’deki doğru ile

verilir. Eğer ortalama yol basit ortalamadan küçükse bu durum Şekil 1.1’de noktalı çizgi ile gösterilen eğri ile verilir (Cornell, 1990).



Şekil 1.1. %50A-%50B benzin karışımının yol grafiği.

Genel olarak karma problemde, ölçülebilir amaç sadece karışımı oluşturan bileşenlerin oranına bağlıdır, karışımın miktarına bağlı değildir. Benzin örneğinde ölçülebilir amaç her galon benzin için yapılan yol miktarıdır. Deneyde yapılan yol miktarı iki benzin çeşidinin oranına bağlıdır, kullanılan yakıtın miktarına bağlı değildir. Karma problemin önemli bir özelliği, bağımsız veya kontrol edilebilir faktörlerin (A ve B benzinleri gibi), hacim, ağırlık veya mol sayısı gibi kontrol edilemeyen oranlar olmamasıdır.

Çalışılan sistemdeki mümkün bileşenlerin sayısını q olarak alırsak ve karmadaki i -inci bileşenin oranını da x_i ile gösterirsek yukarıda anlattıklarımızın ışığında,

$$0 \leq x_i \leq 1, \quad i = 1, 2, \dots, q \quad (1.1)$$

ve

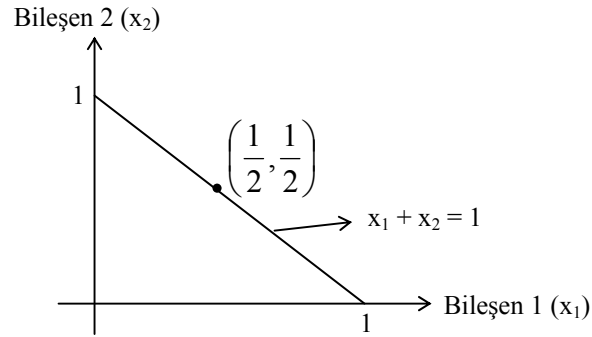
$$\sum_{i=1}^q x_i = x_1 + x_2 + \dots + x_q = 1.0 \quad (1.2)$$

koşulları ortaya çıkar. (1.2)’deki koşula göre negatif olamayan x_i bileşen oranlarının toplamı 1’i verecektir. (1.2)’deki koşul, karma denemelerde oranlar üzerindeki temel kısıtlamadır. (1.1) ve (1.2) ile verilen koşullara karma denemelerin doğal koşulları (doğal kısıtlamaları) denir (Crosier, 1984).

Bundan sonra x_i , $i = 1, 2, \dots, q$ 'ye karmanın bileşenleri diyeceğiz. Karmanın bileşenleri (1.1) ve (1.2)'deki koşulları sağlar. Bir karmada mümkün bileşenlerin sayısı 1 ise (örneğin $x_1 = 1$, $x_2 = 0$, $x_3 = 0$ gibi) böyle karmaya “saf karma” ya da “tek bileşenli karma” denir. Tek bileşenli karmalar genel olarak çok bileşenli karmalara karşı bir ölçüt olarak kullanılır.

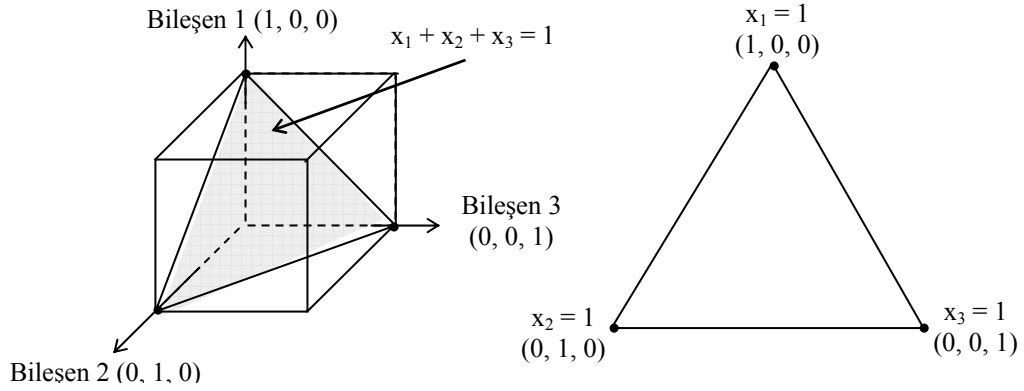
x_i , $i = 1, 2, \dots, q$ 'nın değerleri için kullanılan koordinat sistemine, simpleks koordinat sistemi denir. x_i , $i = 1, 2, \dots, q$ karma bileşenleri üzerindeki (1.1) ve (1.2)'deki doğal kısıtlamalar nedeniyle q bileşenli ve simpleksin sınırları üzerindeki veya içindeki noktalardan oluşan faktör uzayının geometrik tanımı $(q - 1)$ boyutlu simplekstir.

İki bileşen ($q=2$) için faktör uzayı bir doğrudur ve Şekil 1.2'de gösterilmiştir. Bu doğru parçası üzerindeki tüm noktalar (1.1) ve (1.2)'deki doğal koşullardan ötürü $(x_1, 1-x_1)$ veya $(1-x_2, x_2)$ formundadır. Eğer x_1 ve x_2 bir benzin karışımındaki sırasıyla A ve B maddelerinin oranları ise tüm olanaklı A ve B maddelerinin karışımı bu doğru parçası üzerindedir.

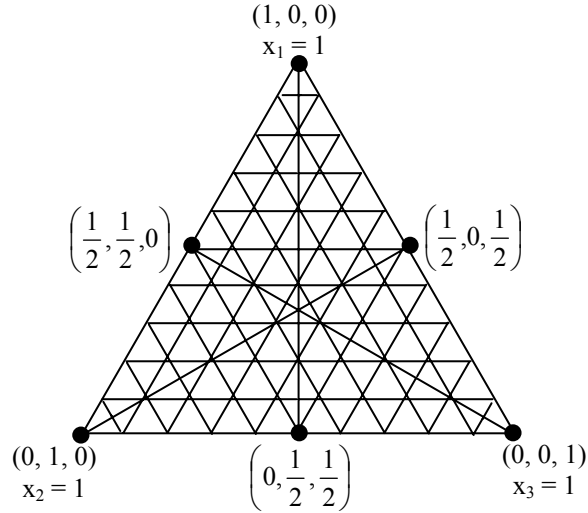


Şekil 1.2. $q = 2$, $x_1 + x_2 = 1$ için faktör uzayı bir boyutlu simplekstir. Tüm deneysel noktalar doğru üzerinde olmalıdır.

Üç bileşen ($q = 3$) için faktör uzayı bir eşkenar üçgendir ve Şekil 1.3'te gösterilmiştir. Şekil 1.3'teki eşkenar üçgenin köşe noktaları tek bileşenli karmaları gösterirler ve $x_i = 1$, $i \neq j$ olmak üzere $x_j = 0$, $i, j = 1, 2, 3$ şeklinde ifade edilir. Üçgenin merkez noktası üç bileşeninde eşit orana sahip olduğu $(1/3, 1/3, 1/3)$ karışımını temsil eder. Şekil 1.4'te üç bileşenli bir sistemdeki koordinatların çizilebileceği üçgensel koordinatlar verilmiştir.

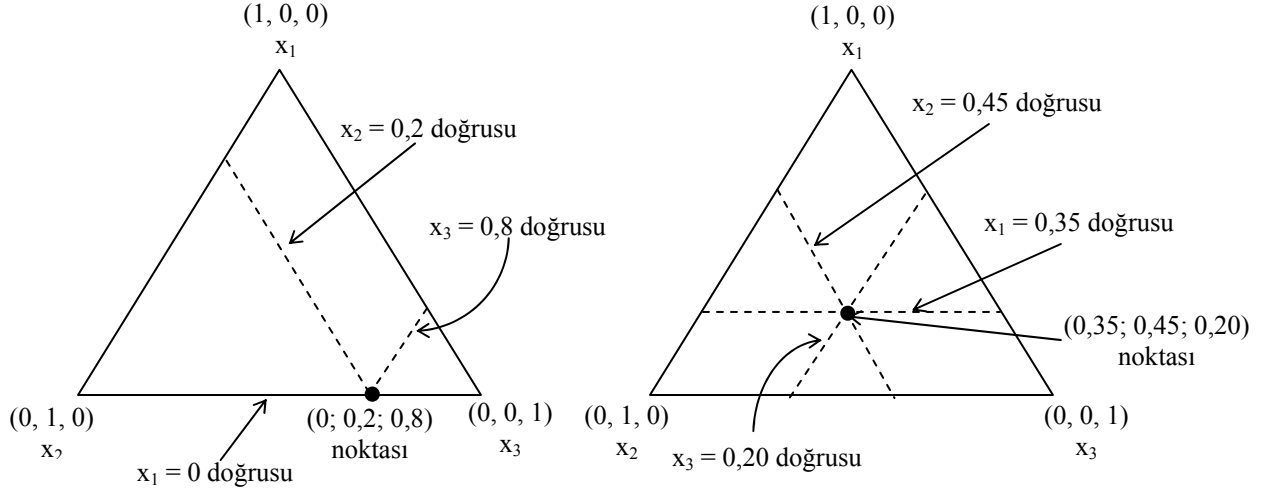


Şekil 1.3. $q = 3$, $x_1 + x_2 + x_3 = 1$ için faktör uzayı iki boyutlu simplekstir. Tüm deneysel noktalar eşkenar üçgenin sınırları üstünde veya içinde olmalıdır.



Şekil 1.4. Karma denemelerde üçgensel koordinatlar.

Şekil 1.5'te $q=3$, $x_1 + x_2 + x_3 = 1$ faktör uzayındaki (iki boyutlu simpleksteki) $x_1 = 0$, $x_2 = 0,2$ ve $x_3 = 0,8$ doğruları ve $(0; 0,2; 0,8)$ noktası; ayrıca $(0,35; 0,45; 0,20)$ noktası ve $x_1 = 0,35$, $x_2 = 0,45$ doğruları gösterilmiştir. Burada $x_1 + x_2 + x_3 = 1$ ve $x_3 = 1 - x_1 - x_2$ olduğu unutulmamalıdır.



Şekil 1.5. Üçgensel koordinatlarda doğru ve noktanın gösterilmesi.

En iyi karışımlar, ürünün kalitesini azaltmayan ve aynı zamanda maliyeti arttırıcı etkisi olmayan karışımlardır. Çoğu zaman karmadaki bileşenlerin en az birinde veya tümünde, bileşen değişim aralığı 0 ile 1 arasında olmayabilir. Bazı veya tüm bileşenler üzerinde alt ve üst sınırlar şeklinde kısıtlamalar olabilir. i -inci bileşen üzerinde böyle bir kısıtlama söz konusu ise, q bileşenli bir sistemde, L_i , i -inci bileşen üzerindeki alt sınır ve U_i , i -inci bileşen üzerindeki üst sınır olmak üzere koşul (ek kısıtlamalar),

$$0 \leq L_i \leq x_i \leq U_i \leq 1 \quad (1.3)$$

şeklinde yazılabilir. Bu kısıtlamalar ayrıca sistemi oluşturan bileşenlerin lineer kombinasyonları üzerinde de olabilir. c_j , lineer kombinasyonlar üzerindeki kısıtlamanın alt sınırı ve d_j üst sınırı olmak üzere,

$$c_j \leq A_{1j}x_1 + A_{2j}x_2 + \dots + A_{qj}x_q \leq d_j \quad (1.4)$$

şeklinde yazılabilir (Piepel, 1983).

Teoride amaç yüzeyi verecek bir fonksiyonel ilişki vardır. Bu ilişkinin

$$\eta = \phi(x_1, x_2, \dots, x_q) \quad (1.5)$$

olduğunu düşünelim. η , bileşenlerin $x_1, x_2, \dots, x_q$ oranlarına bağlı amaç değerlerini göstermektedir. ϕ fonksiyonu ile gösterilen amaç yüzeyin $x_i, i = 1, \dots, q$ 'de sürekli bir fonksiyon olduğu kabul edilecektir. Genelde $\phi(x_1, x_2, \dots, x_q)$ 'i göstermek için polinom fonksiyonlar kullanılır. Normal olarak, düşük dereceli polinomlar, örneğin birinci dereceden polinom,

$$\eta = \beta_0 + \sum_{i=1}^q \beta_i x_i \quad (1.6)$$

ve ikinci dereceden polinom,

$$\eta = \beta_0 + \sum_{i=1}^q \beta_i x_i + \sum_{i \leq j}^q \sum_{j}^q \beta_{ij} x_i x_j \quad (1.7)$$

dir. İki polinomda da q bileşen vardır. Küçük dereceli (birinci ve ikinci dereceden) polinomlar, büyük dereceli (üçüncü ve daha yüksek dereceden) polinomlara göre daha az terim içerirler. Bu nedenle denklemdeki parametrelerin kestirimi için daha az gözlem değerine ihtiyaç duyarlar. Ayrıca çoğu zaman ikinci dereceden polinomlar amaç fonksiyonu oluşturmak için yeterli olmaktadır.

n denemeden oluşan bir karma deneyde η amacı gözlenirken, i -inci denemede $y_i, i = 1, 2, \dots, n$ ile gösterilen gözlenen amaç değerleri için η_i ortalamalı ve σ^2 ortak varyanslı olduğu varsayımı yapılır. ε_i deneysel hatalarını içeren gözlenen amaç fonksiyon değerleri,

$$y_i = \eta_i + \varepsilon_i \quad 1 \leq i \leq n \quad (1.8)$$

şeklindedir. ε_i deneysel hatalarının birbirleriyle ilişkisiz olduğu ve sıfır ortalamalı, σ^2 ortak varyanslı dağıldığı varsayılır. Yani $i, i' = 1, 2, \dots, n$ ve $i \neq i'$ için $E(\varepsilon_i) = 0$,

$E(\varepsilon_i^2) = \sigma^2$ ve $E(\varepsilon_i \varepsilon_{i'}) = 0$ 'dır. Böylece y_i için beklenen değer $i = 1, 2, \dots, n$ olmak üzere $E(y_i) = \eta_i$ 'dir.

$\eta = \phi(x_1, x_2, \dots, x_q)$ ile polinom biçimindeki modeli oluşturmak amacıyla denklemin fonksiyonel ilişkisini yaklaşık olarak değerlendirmek için q bileşenin değişik oranlarıyla ve belirlenmiş sayıda denemeler yapılır. Değişik oranların kümesini oluşturmak deneysel tasarımların ilk adımıdır. Daha sonra modeldeki parametreler tahmin edilir.

Amaç fonksiyonu için; y_i , $i = 1, \dots, n$ i-inci gözlem değerini, x_i , sütunlarında bileşen oranlarının çarpımını içeren $n \times p$ tipindeki X matrisinin i -inci satırını (notasyonları sadeleştirmek için bundan sonra x_i yerine x'_i kullanacağız), β , $p \times 1$ tipinde bilinmeyen parametre vektörünü ve ε_i , sıfır ortalamalı σ^2 ($\sigma^2 < \infty$) varyanslı normal dağılımdan gelen i -inci gözlemin hatasını göstermek üzere n gözlem ve p parametrelili

$$y_i = x'_i \beta + \varepsilon_i, \quad i = 1, \dots, n \quad (1.9)$$

modelini düşünelim. En genel halde regresyon tahmin edicisini veriye uyguladığımız zaman β için regresyon katsayıları dediğimiz $\hat{\beta}$ elde edilir ve y_i 'lerin tahmin edilen (uydurulmuş) değerleri $\hat{y}_i = E(\hat{y}_i) = x'_i \hat{\beta}$ 'dir. Ayrıca rezidüler $e_i = y_i - \hat{y}_i$ olarak verilir. (1.9)'de verdiğimiz model için en çok bilinen klasikleşmiş tahmin edici en küçük kareler tahmin edicisidir. σ 'nın sabit olduğunu düşünelim. Bu durumda β 'nın en küçük kareler tahmini,

$$S(\beta) = \sum_{i=1}^n (y_i - x'_i \beta)^2 \quad (1.10)$$

fonksiyonunu minimize eder. X , $n \times p$ ($p \geq q$) tipinde ve i -inci satırı, i -inci gözleme karşılık gelen sütunlarında p tane x_i değişkeninin değerlerini içeren p tam ranklı katsayı matrisi ve y , $n \times 1$ tipinde amaç değişken üzerinde gözlemlerin bir sütun vektörü olmak üzere β 'nın en küçük kareler çözümü matris formunda,

$$\hat{\beta} = (X'X)^{-1}X'y \quad (1.11)$$

dir ve $\hat{y}$ uydurulmuş değerleri,

$$\hat{y} = X(X'X)^{-1}X'y \quad (1.12)$$

şeklindedir. En küçük kareler yönteminin bu kadar popüler ve klasikleşmiş olmasını temel nedeni hesaplanmasının kolaylığıdır. Tahmin veriler yardımıyla herhangi bir iteratif yöntemle ihtiyaç olmaksızın direk ve kolaylıkla hesaplanmaktadır. Ayrıca diğer tüm yansız tahmin ediciler içinde en iyi lineer yansız tahmin edicidir ve eğer hatalar normal dağılıyorsa, Maksimum Likelihood tahmin edicisine benzediği gibi bu durumda diğer yansız tahmin edicilere göre minimum varyanslı tahmin verir.

Regresyon analizinde karşılaşılan en önemli iki problem kötü koşulluluk yada iç ilişki ve sapan değer problemidir. Karma denemelerde genellikle karışımı oluşturan bileşenler üzerinde fiziksel, kimyasal ve ekonomik nedenlerden kaynaklanan (1.3)'te verilen formda alt ve üst sınırlar şeklinde ve/veya (1.4)'te verilen formda bileşenlerin lineer kombinasyonları üzerinde ek kısıtlamalar bulunur. Dolayısıyla bileşen değişim aralıkları 0 ile 1 aralığında değildir ve buda kötü koşulluluk yada iç ilişki probleminin ortaya çıkmasına neden olur. Ayrıca y_i gözlem değerleri sapan değerlere sahip olabilirler ve böylece ε_i deneysel hataları normal dağılmıyor olabilirler. Bu durumda en küçük kareler ile elde edilen istatistikler tutarlı olmaz. Ayrıca Montgomery ve Voth (1994) Şekil 1.2 ve Şekil 1.3'te verdiğimiz kötü koşulluluk problemi olmayan Simpleks Latis Tasarımı olarak bilinen standart tasarımların, en küçük kareler tahmininin tutarlılığını etkileyen x yönünde sapan değer problemine sahip olduklarını göstermişlerdir.

Kötü koşulluluk problemi için çeşitli teknikler önerilmiştir fakat bunlar arasında en sık kullanılanları Ridge ve Liu regresyon tahmin edicileridir (Hoerl ve Kennard, 1970; Liu, 1993). Ridge tahmin edici, $\hat{\beta}_{EKK}$, β 'nın en küçük kareler tahmin edicisi ve k , ($k > 0$) yanalılık çarpanı olmak üzere,

$$\hat{\beta}_R = (X'X + kI)^{-1}X'X\hat{\beta}_{EKK} \quad (1.13)$$

ve Liu tahmin edici, $0 < d < 1$ olmak üzere,

$$\hat{\beta}_L = (X'X + I)^{-1}(X'X + dI)\hat{\beta}_{EKK} \quad (1.14)$$

dir. Liu tahmin edicisinin Ridge tahmin edicisine göre bir avantajı, Liu tahmin edicisi d parametresinin bir lineer fonksiyonudur ve Ridge tahmin edicisindeki k yanlılık çarpanına göre seçimi daha kolaydır. Dikkat edilirse her iki tahmin edici β 'nin en küçük kareler tahmin edicisini kullanmaktadır ve bu tahmin edici y yönündeki sapan değerlere karşı hassastır. Bu problemi ortadan kaldırmanın yolu, bu tahmin edicilerin Robust şekillerinin tanımlanmasıdır.

Eğer y yönünde sapan değerlere sahip bir veri seti ile çalışıyorsak, bu yöndeki sapan değerlere karşı hassas olan klasik yöntemler (en küçük kareler, Ridge, Liu) tutarlı tahminler vermeyecektir. (1.9)'da verilen modelde y_i gözlemleri normal dağılıyorsa en küçük kareler yöntemi uygun bir yöntemdir ve β 'nin tahmini iyi istatistiksel özelliklere sahip olacaktır. Ancak, y_i gözlemleri normal olmayan, özellikle normalden daha kalın kuyruklu, sapan değer üreten, dağılımlardan geliyorsa en küçük kareler uygun tahmin edici olmayacaktır çünkü bilindiği gibi sapan değerlerin en küçük kareler üzerinde büyük etkileri vardır. Bu durumlarda robust tahmin yöntemlerini kullanmak önerilmiştir (Huber, 1972, 1973, 1981). Eğer veride sapan değerle beraber kötü koşulluluk problemi varsa robust Ridge ve robust Liu tahmin edicilerinin kullanılması önerilmiştir (Silvapulle, 1991; Arslan ve Billor, 1996, 2000). Robust Ridge tahmin edici, $\hat{\beta}^*$, β 'nin bir robust tahmin edicisi ve k , ($k > 0$) yanlılık çarpanı olmak üzere,

$$\hat{\beta}_{RR} = (X'X + kI)^{-1}X'X\hat{\beta}^* \quad (1.15)$$

ve Liu tahmin edici, $0 < d < 1$ olmak üzere,

$$\hat{\beta}_{RL} = (X'X + I)^{-1}(X'X + dI)\hat{\beta}^* \quad (1.16)$$

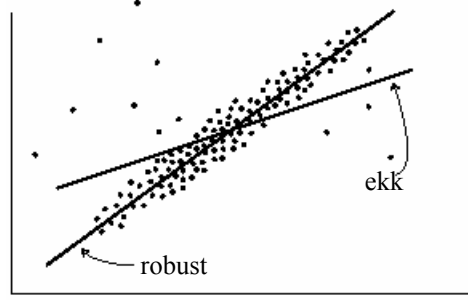
dir.

1.2. Robust İstatistiğe Giriş

İstatistik literatüründe “Robust” kelimesi “Güçlü, Dayanaklı” kelimeleri ile eşanlamlıdır. Robust istatistiğin asıl amacı model hataları için varsayılan dağılımların yanlış olması ve/veya sapan değerlerin bulunması durumunda güvenilebilir sonuçlar verebilmektir.

(1.9)’da verilen model için en küçük kareler ile parametre tahmini, ε hata teriminin normal dağılımından geldiği varsayımı altında yapılır. Normallik varsayımı hipotez testleri ve güven aralıklarının oluşturulması için gerekli olup β katsayılarının tahmini için gerekli değildir. Bu durumda elde ettiğimiz β tahminleri iyi istatistiksel özelliklere sahiptir. Ancak ε hatalarının normal olmayan, kalın kuyruklu, ince kuyruklu ve/veya çarpık dağılımlardan geliyorsa en küçük kareler tahmin edicileri iyi tahmin ediciler olmaktan çıkacaktır.

En çok kullanılan teknik olan en küçük kareler tahmin edicileri $\varepsilon_i = y_i - x_i'\beta$ olmak üzere rezidü kareler toplamını yani $\sum_{i=1}^n \varepsilon_i^2$ ’yi minimum yapma düşüncesi ile çalışır. Verilerde sapan değerler varsa bu durumda ε_i artacak ve kareler toplamını minimum yapmaya çalışan en küçük kareler iyi istatistiksel özelliklere sahip olmayacaktır. Şekil 1.6’da verilen grafikte sapan değerlere sahip bir veri için en küçük kareler ve robust yöntemlerle uydurulmuş doğru verilmiştir. En küçük kareler ile uydurulan doğrunun verileri çok iyi açıklayamadığı buna karşılık robust yöntemle elde edilen doğrunun sapan değerlerden hiç etkilenmediği açıkça görülmektedir.



Şekil 1.6. En küçük kareler ve robust yöntemlerle uydurulan doğru.

Robust istatistiksel analiz yöntemleri anlatılmadan önce gerekli bazı tanım ve özellikler Bölüm 1.2.1’de verilecektir.

1.2.1. Tanım ve Özellikler

X , F dağılımına sahip ve her reel μ için $X + \mu$ ’nün dağılımı F_μ ile gösterilsin. O zaman $\{F_\mu; -\infty < \mu < \infty\}$ ’a F tarafından üretilen konum parametre ailesi denir. μ ’yede konum parametresi (location parameter) denir.

F bir dağılım ve $\theta > 0$ olsun. $F_\theta = F(x/\theta)$, $x > 0$ olarak tanımlansın. $\{F_\theta; \theta > 0\}$ ’a F tarafından genelleştirilen ölçek parametre ailesi denir ve θ ’yada ölçek parametresi (scale parameter) denir.

X , F dağılımına sahip olsun ve $\sigma X + \mu$ ’nün dağılımı $F_{\mu,\sigma}$ ile gösterilsin. Böylece $\{F_{\mu,\sigma}; -\infty < \mu < \infty, \sigma > 0\}$ ’a F tarafından üretilen konum ölçek ailesi denir ve (μ, σ) ’ya konum-ölçek parametresi (location-scale parameter) denir.

$X = (x_1, x_2, \dots, x_n)$ rasgele örneklem olsun. θ tahmin edicisi her a için,

$$\theta(a + X) = \theta(a + x_1, \dots, a + x_n) = a + \theta(x_1, \dots, x_n) = a + \theta(X) \quad (1.17)$$

oluyorsa θ tahmin edicisine konum eşdeğerli tahmin edici denir. Eğer θ tahmin edicisi her $a > 0$ için,

$$\theta(aX) = \theta(ax_1, \dots, ax_n) = |a| \cdot \theta(X) \quad (1.18)$$

oluyorsa θ 'ya ölçek eşdeğerli tahmin edici denir. Ayrıca eğer θ tahmin edicisi her $a > 0$ ve her b için,

$$\theta(aX + b) = \theta(ax_1 + b, \dots, ax_n + b) = a.\theta(X) + b \quad (1.19)$$

oluyorsa θ 'ya konum-ölçek eşdeğerli tahmin edici denir.

Örneklem ortalaması,

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} \quad (1.20)$$

olarak verilir. Örneklem ortalaması konumun bir ölçüsüdür ve kitle ortalamasının bir tahminidir. Örneklem Medyanı ise, $x_{(i)}$ 'ler $x_1, \dots, x_n$ 'lerin sıra istatistikleri ve $x_{(1)} \leq \dots \leq x_{(n)}$ olmak üzere,

$$\text{Medyan} = \begin{cases} x_{\frac{n+1}{2}}, & n \text{ tek} \\ \frac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2}, & n \text{ çift} \end{cases} \quad (1.21)$$

şeklinde tanımlanır. Dikkat edilirse örneklem ortalaması sapan değerlerden çok aşırı etkilenecek bir yapıya sahiptir ve bu yüzden robust değildir. Ancak örneklem medyanı ise verilerin merkezindeki değere eşit olup sapan değerlere karşı hassas olmayan bir yapıdadır ve robust bir tahmindir. Örneğin Rosner (1975)'in verdiği 10 adet aylık kan basıncı ölçüsünden oluşan $x_i = \{90, 93, 86, 92, 95, 83, 75, 40, 88, 80\}$ verisi için örneklem ortalaması 82.2 iken örneklem medyanı 87 olarak elde edilir.

Örneklem Varyansı,

$$\text{Var}(X) = S_n^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1} \quad (1.22)$$

olarak verilir ve örneklem standart sapması $S_n = \sqrt{S_n^2}$ 'dir.

Örneklem varyansı örneklem ortalamasından hesaplanıyor olup sapan değerlere karşı çok hassastır ve robust olmayan bir ölçek tahmini verir. Bilindiği gibi, medyan küçükten büyüğe sıralanmış bir seride en ortadaki terim olduğundan medyandan hareketle hesaplanacak bir ölçek tahmincisi de tıpkı medyan gibi sapan değerlerden etkilenmeyecektir. Medyandan hareketle hesaplanacak olan bu ölçek tahmincisi Medyan Mutlak Sapma (Median Absolute Deviations-MAD) olarak adlandırılmaktadır. Medyan mutlak sapma,

$$\text{MAD} = \text{medyan}|x_i - \text{medyan}(x_i)| \quad (1.23)$$

formülü ile hesaplanmaktadır. Yani seri terimlerinin medyandan mutlak farklarının medyanı alınarak elde edilmektedir. Bu ölçek tahmin edicisi sapan değerlerden etkilenmediğinden robust bir tahmindir.

1.3. Karma Denemelerde Simpleks Latis Tasarımı ve Kanonik Polinom

Simpleks (kısıtlanmış bölge) üzerinde düzgün aralıklarla dağılmış noktaların yerleşim düzeni latis (kafes) olarak bilinir. Latis adı noktaların bir sıralanışını tanımlamak için kullanılır. Örneğin simpleks üzerinde q bileşenli m-inci dereceden bir polinomu tanımlamak, temsil etmek için $\{q, m\}$ simpleks latis kullanılır. Simpleks latis 0 ile 1 arasında $m + 1$ eşit aralıklı değerlerden oluşur.

$$x_i = 0, \frac{1}{m}, \frac{2}{m}, \dots, 1 \quad (1.24)$$

böylece $\{q, m\}$ simpleks latis tasarımı bileşenlerin tüm olanaklı değerlerini içerir (Scheffé, 1958).

$\{q, m\}$ simpleks latisindeki tasarım noktalarının sayısı,

$$\binom{q+m-1}{m} = \frac{(q+m-1)!}{m!(q-1)!} \quad (1.25)$$

dir. Bu birbirinden farklı ve sınırsız olarak tekrarlanabilen q öğenin sıra göz önüne alınmaksızın, yani 123 veya 132 gibi, m -li kombinasyonlarının sayısıdır. Örneğin,

$\{3, 2\}$ simpleks latis için tasarım noktalarının sayısı $\binom{3+2-1}{2} = 6$ ve $\{3, 3\}$

simpleks latis için tasarım noktalarının sayısı $\binom{3+3-1}{3} = 10$ 'dur.

$\{q, m\}$ simpleks latis için tasarım noktalarında toplanmış veriler için genel regresyon fonksiyonu aşağıdaki formdadır. m -inci dereceden polinom

$$\eta = \beta_0 + \sum_{i=1}^q \beta_i x_i + \sum_{i \leq j}^q \beta_{ij} x_i x_j + \sum_{i \leq j \leq k}^q \beta_{ijk} x_i x_j x_k + \dots \quad (1.26)$$

şeklinde verilir (Cornell, 1990). (1.26)'daki polinomda terim sayısı $\binom{q+m}{m}$ 'dir ve

bu polinom, $x_1 + x_2 + \dots + x_q = 1$ kısıtlaması durumunda karma modeller için anlamlıdır. Kısıtlama nedeniyle,

$$x_q = 1 - \sum_{i=1}^{q-1} x_i \quad (1.27)$$

alınır ve polinomdaki terimlerden biri sadeleştirilir ya da kaldırılır. Bu işlem polinomun derecesini etkilemez. Bu durumda (1.26)'daki polinom derecesi m olan

$q - 1$ bileşenli ve $\binom{q+m-1}{m}$ terimli polinoma dönüşür. Bu polinom daha az bileşen ve terim içerdiği için daha uygundur. (1.26)'daki polinoma alternatif olan m -inci dereceden q bileşenli polinom, (1.2)'deki kısıtlamanın (1.26)'daki polinomun bazı terimleri ile çarpılmasıyla elde edilir. Sonuçta elde edilen polinoma “kanonik polinom” ya da “polinomun kanonik formu” denir. Simpleks üzerinde derecesi m olan q bileşenli belirli bir kanonik polinom modele $\{q, m\}$ simpleks latis karşılık gelir. Bunun terside doğrudur (Cornell, 1990).

Kanonik polinomda karmaların doğal kısıtlamaları kullanılarak dönüşüm yapılmasıyla aşağıdaki karma modellerin kanonik formları elde edilir.

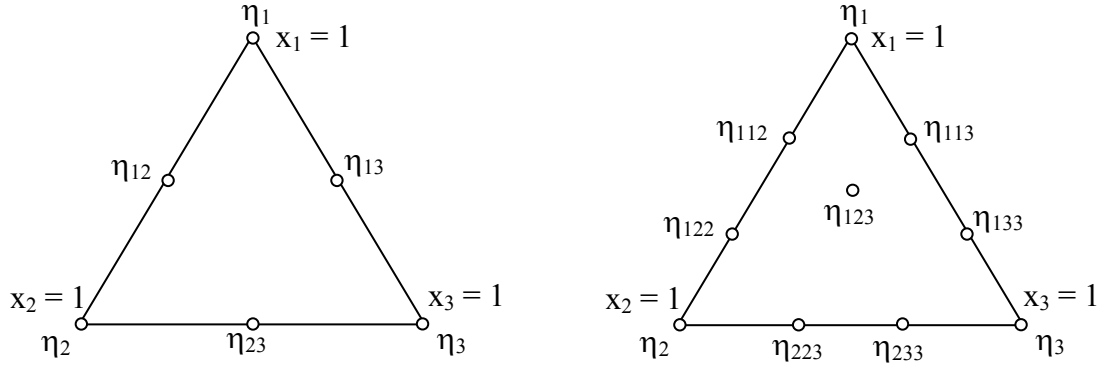
$$\eta = \sum_{i=1}^q \beta_i x_i \quad (\text{Doğrusal model}) \quad (1.28)$$

$$\eta = \sum_{i=1}^q \beta_i x_i + \sum_{i \leq j} \sum_{j=1}^q \beta_{ij} x_i x_j \quad (\text{Karesel model}) \quad (1.29)$$

1.3.1. $\{q, m\}$ Simpleks Latisin Tasarım Noktalarında Amaç (Yanıt) Fonksiyonu Olarak Kanonik Polinom Katsayıları ve Parametrelerin Tahmini

$\{q, m\}$ simpleks latisi ile $\{q, m\}$ polinomu arasında özel bir ilişki vardır. Bu ilişki latisteki tasarım noktalarının sayısı ile polinomdaki terimlerin sayısı arasında bire bir ilişkidir. Bu ilişkinin bir sonucu olarak polinomdaki parametreler, $\{q, m\}$ simpleks latisinin noktalarında beklenen amacın fonksiyonu olarak ifade edilebilir.

Amaç fonksiyondaki i -inci bileşen η_i ile gösterilsin. Eşit oranlı (%50-%50) i ve j -inci bileşen η_{ij} ve yine eşit oranlı i, j ve k -inci bileşen η_{ijk} ile gösterilsin. Şekil 1.7'de $\{3, 2\}$ ve $\{3, 3\}$ simpleks latislerindeki tasarım noktalarında amaç fonksiyonunun bu terminolojisi gösterilmiştir. Bu terminolojiyi ilk kez Scheffé (1958) vermiştir.



Şekil 1.7. $\{3, 2\}$ ve $\{3, 3\}$ simpleks latislerindeki tasarım noktalarında amaç terminolojisi.

Amaç fonksiyonunun terminolojisinin üç özelliği vardır:

1. Alt simge sayısı karmada kullanılan bölümdeki bölene eşittir (η_{ij} 'deki iki alt simge bileşen oranları $1/2$ olan bir karışımı gösterir, yani $x_i = 1/2$ ve $x_j = 1/2$ 'dir.)
2. Alt simgedeki farklı rakamların sayısı karmadaki sıfırdan farklı orana sahip bileşen sayısını verir.
3. Alt simgede yer alan rakamlar karmadaki ilgili bileşenin oranını verir. Örneğin, η_{112} üç alt simgeye sahiptir. O zaman bölüm 3'tür. Alt simgede iki tane 1 ve bir tane 2 vardır. Farklı rakam sayısı iki olup 2 bileşen sıfırdan farklıdır. 1 rakamı iki kez görünmek üzere birinci bileşen için oran $2/3$ ve 2 rakamı bir kez görünmek üzere ikinci bileşen için oran $1/3$ 'tür.

Polinom modeldeki β_i , β_{ij} ve β_{ijk} parametrelerinin η_i , η_{ij} ve η_{ijk} beklenen amaçlar cinsinden ifade edilebilmesi için $\binom{q+m-1}{m}$ tane denklemin çözülmesi gerekir. Bu sayı sadece $\{q, m\}$ polinomundaki parametre sayısı değil ayrıca $\{q, m\}$ simpleks latisindeki tasarım noktalarının da sayısıdır. Bu yüzden $\{q, m\}$

simpleks latis tasarım noktalarında ölçülen η_i , η_{ij} ve η_{ijk} beklenen amaçlarının da sayısıdır. Örneğin,

$$\eta = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{12} x_1 x_2 + \beta_{13} x_1 x_3 + \beta_{23} x_2 x_3 \quad (1.30)$$

ikinci dereceden polinom modeli 3 bileşenli karma sistem için kullanılsın. İkinci dereceden polinom modelindeki β_i ve β_{ij} parametreleri, $\{3, 2\}$ simpleks latis tasarım noktalarındaki η_i ve η_{ij} beklenen amaçlar cinsinden ifade edilecektir..

Köşe noktalarında η_i : $x_i = 1$, $x_j = 0$, $i, j = 1, 2, 3$ $j \neq i$

Kenar orta noktalarda η_{ij} : $x_i = 1/2$, $x_j = 1/2$, $x_k = 0$, $i < j$, $k \neq i, j$

dönüşümü yapılırsa;

$$\begin{aligned} \eta_1 &= \beta_1 & \eta_2 &= \beta_2 & \eta_3 &= \beta_3 \\ \eta_{12} &= \beta_1 \left(\frac{1}{2} \right) + \beta_2 \left(\frac{1}{2} \right) + \beta_{12} \left(\frac{1}{2} \cdot \frac{1}{2} \right) \\ \eta_{13} &= \beta_1 \left(\frac{1}{2} \right) + \beta_3 \left(\frac{1}{2} \right) + \beta_{13} \left(\frac{1}{2} \cdot \frac{1}{2} \right) \\ \eta_{23} &= \beta_2 \left(\frac{1}{2} \right) + \beta_3 \left(\frac{1}{2} \right) + \beta_{23} \left(\frac{1}{2} \cdot \frac{1}{2} \right) \end{aligned} \quad (1.31)$$

6 bilinmeyenli 6 denklem sistemi elde edilir. Bu sistem β_i ve β_{ij} için çözüldüğünde,

$$\begin{aligned} \beta_{12} &= 4\eta_{12} - 2\eta_1 - 2\eta_2 \\ \beta_{13} &= 4\eta_{13} - 2\eta_1 - 2\eta_3 \\ \beta_{23} &= 4\eta_{23} - 2\eta_2 - 2\eta_3 \end{aligned} \quad (1.32)$$

elde edilir. İkinci dereceden q bileşenli bir polinom model $q(q+1)/2$ terim içerir ve beklenen amaçlar $\{q, 2\}$ simpleks latis tasarım noktalarıdır. İkinci dereceden q

bileşenli polinom modelindeki β_i ve β_{ij} , $i, j = 1, 2, \dots, q$, $i < j$ parametreleri $\{q, 2\}$ simpleks latis tasarım noktalarındaki η_i ve η_{ij} beklenen amaçları cinsinden yukarıdaki şekilde ifade edilir. Genel olarak q bileşen için,

$$\begin{aligned}\beta_i &= \eta_i \\ \beta_{ij} &= 4\eta_{ij} - 2(\eta_i + \eta_j)\end{aligned}\quad (1.33)$$

dir. Daha yüksek dereceli ($m > 2$) polinomlar için işlemler, ikinci dereceden polinomdaki işlemlerle benzer şekilde yapılır.

β_i ve β_{ij} parametreleri için oluşturulan denklemlerden, β_i ve β_{ij} parametrelerinin b_i ve b_{ij} tahminlerini bulabiliriz. Bu amaçla η_i ve η_{ij} beklenen amaçları için gözlenmiş değerleri yani y_i ve y_{ij} değerlerini kullanabiliriz.

Köşe noktalarında $x_i = 1$, $x_j = 0$, $j \neq i$ 'deki her bir bileşene karşılık gelen gözlenen amaç değerlerini y_i ve kenar orta noktalarda $x_i = 1/2$, $x_j = 1/2$, $x_k = 0$, $i < j \neq k$ 'daki ikili terimlere karşılık gelen gözlenen amaç değerlerini y_{ij} olarak alalım. $\beta_i = \eta_i$ ve $\beta_{ij} = 4\eta_{ij} - 2(\eta_i + \eta_j)$ 'de y_i ve y_{ij} 'leri η_i ve η_{ij} 'lerle yer değiştirelim. b_i ve b_{ij} , sırasıyla β_i ve β_{ij} parametrelerinin tahminleri olsun. Bu durumda,

$$\begin{aligned}b_i &= y_i, \quad i = 1, \dots, q \\ b_{ij} &= 4y_{ij} - 2(y_i + y_j), \quad i, j = 1, 2, \dots, q; \quad i < j \quad \left(\frac{b_{ij}}{4} = y_{ij} - \frac{y_i + y_j}{2} \right)\end{aligned}\quad (1.34)$$

olacaktır. r_i , r_j ve r_{ij} sırasıyla $x_i = 1$, $x_j = 0$; $x_i = 0$, $x_j = 1$ ve $x_i = x_j = 1/2$, $x_k = 0$ ($i < j$, $k \neq i, j$) noktalarındaki gözlemlerin tekrar sayısını gösterebilir. $\bar{y}_i$, $\bar{y}_j$, $\bar{y}_{ij}$ ortalama değerleri bu gözlemlerden elde edilsin. Yukarıdaki denklemlerde y_i , y_j , ve y_{ij} 'lerin yerine sırasıyla $\bar{y}_i$, $\bar{y}_j$, $\bar{y}_{ij}$ 'leri kullanılırsa β_i ve β_{ij} parametrelerinin en küçük kareler tahminleri,

$$\begin{aligned}\hat{\beta}_i &= \bar{y}_i \\ \hat{\beta}_{ij} &= 4\bar{y}_{ij} - 2(\bar{y}_i + \bar{y}_j)\end{aligned}\quad (1.35)$$

biçiminde elde edilir. Dikkat edilirse β_i ve β_{ij} parametrelerin tahminleri köşe ve orta noktalarda yapılan tekrarlı gözlemlerin örneklem ortalamasına dayalıdır. Ancak örneklem ortalaması çok hassas bir istatistiktir. Alınan ölçümler arasında dış koşullardan kaynaklanan hatalı gözlemler varsa robust olmayan örneklem ortalaması yanlış tahminler elde edilmesine neden olacaktır. Bu yüzden örneklem ortalaması yerine bunun robust tahminini kullanabiliriz.

1.3.2. {3, 2} Simpleks Latis Tasarımına Bir Örnek

{3, 2} simpleks latis tasarımının her noktasında 3 tekrarlı olarak yapılan meyve suyu karışımını ele alalım (Cornell, 1990). Kavun (x_1), ananas (x_2) ve portakal (x_3) meyvelerinden oluşan meyve suyu karışımı elde edilmek isteniyor. Elde edilen karışımın Lezzetli, dayanıklı ve tortusuz olması bekleniyor. Yapılan ölçümlerden sonra elde edilen veriler aşağıdaki tabloda verilmiştir.

Tablo 1.2. Meyve suyu karışımı verileri

Kavun (x_1)	Ananas (x_2)	Portakal (x_3)	Uygunluk (y_i)	Ortalama ($\bar{y}$)
1	0	0	4.8 4.3 4.7	4.60
0	1	0	6.3 6.5 6.2	6.33
0	0	1	7.4 6.9 7.0	7.10
0.5	0.5	0	6.1 6.3 5.8	6.07
0.5	0	0.5	5.9 6.1 6.5	6.17
0	0.5	0.5	6.1 6.2 6.2	6.17

Tablo (1.2)'deki verilere dikkat edilirse, tekrarlı ölçümlerde herhangi bir uyumsuzluk görünmemektedir. Ortalamalarda alınan ölçümlerden aykırı olmayan bir seviyede yani sapan değer ve x 'in değişim aralığı 0 ve 1 olduğundan kötü koşulluluk problemleri yoktur.

(1.35)'de elde edilen ikinci dereceden modelin katsayı tahminlerini kullanarak bilinmeyen β katsayılarını hesaplarsak:

$$b_1 = \bar{y}_1 = 4.60$$

$$b_2 = \bar{y}_2 = 6.33$$

$$b_3 = \bar{y}_3 = 7.10$$

$$b_{12} = 4 \bar{y}_{12} - 2(\bar{y}_1 + \bar{y}_2) = 4.6.07 - 2.(4.60 + 6.33) = 2.42$$

$$b_{13} = 4 \bar{y}_{13} - 2(\bar{y}_1 + \bar{y}_3) = 4.6.17 - 2.(4.60 + 7.10) = 1.28$$

$$b_{23} = 4 \bar{y}_{23} - 2(\bar{y}_2 + \bar{y}_3) = 4.6.17 - 2.(6.33 + 7.10) = -2.18$$

elde edilir. O halde modelimiz

$$\hat{y} = 4.60x_1 + 6.33x_2 + 7.10x_3 + 2.42x_1x_2 + 1.28x_1x_3 - 2.18x_2x_3 \quad (1.36)$$

olarak elde edilir.

Şimdi Tablo 1.2’de verilen meyve suyu karışımı verisinde 3 tekrarlı gözlemler yerine 4 tekrarlı gözlemler alındığını düşünelim. Yeni eklenen gözlemler tezin amacını daha iyi anlatabilmek için eklenmiş hayali gözlemlerdir. Ancak gerçekte bu tip veri setleriyle çalışma ihtimali yüksektir.

Tablo 1.3. Meyve suyu karışımı verisi. Verilere yeni eklenen gözlemler

Kavun (x_1)	Ananas (x_2)	Portakal (x_3)	Uygunluk (y_i)	Ortalama ($\bar{y}$)
1	0	0	4.8 4.3 4.7 10.5	6.075
0	1	0	6.3 6.5 6.2 3.0	5.5
0	0	1	7.4 6.9 7.0 1.4	5.675
0.5	0.5	0	6.1 6.3 5.8 0.8	4.75
0.5	0	0.5	5.9 6.1 6.5 0.3	4.7
0	0.5	0.5	6.1 6.2 6.2 18.4	9.225

Tablo 1.3’teki verilere dikkat edersek yeni eklenen bu gözlemler ilk üç gözlemin ortalamasından çok farklı, örneklem ortalamasını büyüten, sapan değer oluşturan gözlemlerdir. Bu gözlem ve ortalama için katsayıları hesaplarsak

$$b_1 = \bar{y}_1 = 6.075$$

$$b_2 = \bar{y}_2 = 5.5$$

$$b_3 = \bar{y}_3 = 5.675$$

$$b_{12} = 4\bar{y}_{12} - 2(\bar{y}_1 + \bar{y}_2) = -4.15$$

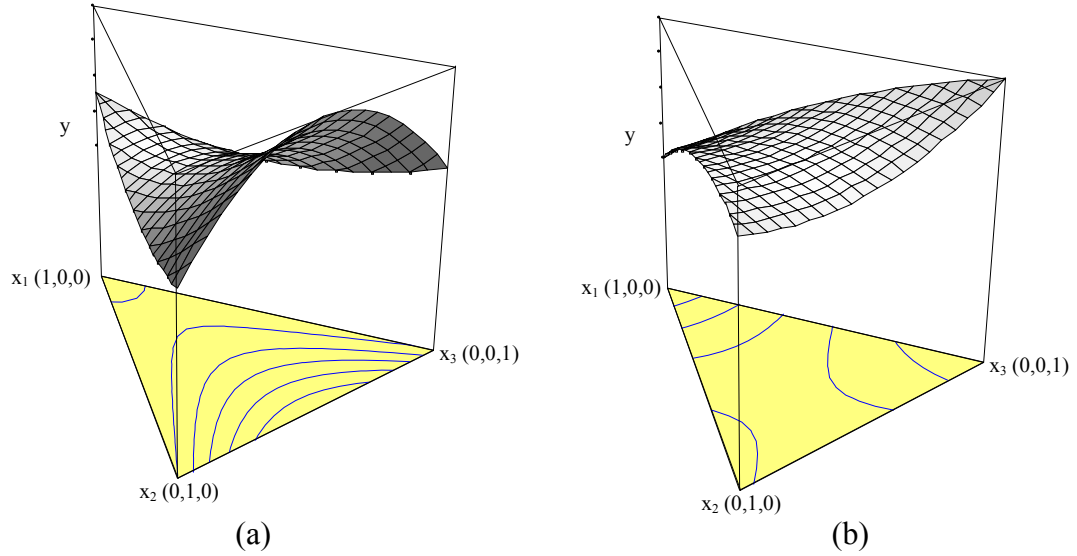
$$b_{13} = 4\bar{y}_{13} - 2(\bar{y}_1 + \bar{y}_3) = -4.7$$

$$b_{23} = 4\bar{y}_{23} - 2(\bar{y}_2 + \bar{y}_3) = 14.55$$

sonuçlarını elde ederiz. Ancak eklediğimiz değerlere dikkat edilirse önceki 3 gözlemden çok farklı gözlemler. Yani bu gözlemler yüzünden katsayı tahminlerimiz güvenilirlikten oldukça uzaklaşıyor. $(1/2, 1/2, 0)$ noktasındaki 18,4 olarak alınmış 4. gözlemin b_{12} katsayısına etkisinin çok fazla olduğu görülmektedir. Uydurulan model,

$$\hat{y} = 6.075x_1 + 5.5x_2 + 5.675x_3 - 4.15x_1x_2 - 4.7x_1x_3 + 14.55x_2x_3 \quad (1.37)$$

şeklindedir. İki veri grubu için elde edilen modellerdeki katsayı tahminleri birbirinden çok farklıdır ve sapan değerlerin modele etkisi açıkça görülmektedir. Şekil 1.8’te her iki model için yüzey ve kontur grafikleri verilmiştir. Aslında örneklem ortalamasına bağlı olan bir tahmin yaptığımız için normal olarak problemsiz veriler için grafiklerin birbirine benzer olmasını bekleriz ancak burada özellikle köşe noktalarda örneklem ortalamasının sapan değerlerden ne kadar fazla etkilendiğini görebiliyoruz.



Şekil 1.8. Uydurulan modeller için yüzey ve kontur grafikleri

Örneklerde elde edilen modellerden ve grafiklerden anlaşıldığı gibi (1.35)'te verilen tasarım tahmin yöntemi örneklem ortalamasına bağlı olduğundan ve sapan değerlere sahip bir veri grubu ile çalıştığımızda, örneklem ortalamasının robust olmamasından dolayı büyük problemler çıkarıyor. İşte veride bulunması muhtemel olan bu tip gözlemlerden etkilenmeyecek bir tahmin edici olarak robust tahmin edicileri düşünüyoruz.

1.4. Konum ve Ölçek Parametreleri için M Tahmin Edicisi

Maksimum Likelihood tipi tahmin edici veya M tahmin edici en çok kullanılan robust tekniklerden biridir. M tahmin edicisi sapan değerlerin etkilerini azaltmak için hataların kareleri toplamı yerine hataların karesel fonksiyonundan daha yavaş artan bir fonksiyon kullanır. Bu fonksiyon, ρ simetrik, tek minimumu sıfır olan pozitif tanımlı ve karesel bir fonksiyondan daha yavaş artacak şekilde seçilmiş bir fonksiyon olmak üzere,

$$\sum \rho(x_i, \theta) = \min \quad (1.38)$$

şeklindedir. (1.38) denkleminin türevini alır sıfıra eşitlersek $\rho' = \psi$ olmak üzere normal denklemler dediğimiz

$$\sum \rho'(x_i, \theta) = \sum \psi(x_i, \theta) = 0 \quad (1.39)$$

denklemleri elde ederiz. Buna θ 'nın M tahmin edicisi veya Maksimum Likelihood tipi tahmin edicisi denir ve eğer $f(x)$, x_i 'lerin geldiği dağılımın olasılık yoğunluk fonksiyonu iken $\rho = -\log f(x_i, \theta)$ olarak seçilirse alışılmış maksimum likelihood tahmini elde edilir. Yani, $f(x)$ olasılık yoğunluk fonksiyonu ile verilen bir dağılımda bilinmeyen konum parametresini θ ile verelim. $f(x - \theta)$, $-\infty \leq \theta \leq \infty$ fonksiyonunu tanımlayalım. $x_1, x_2, \dots, x_n$ bu dağılımdan gelen bağımsız, özdeş dağılımlı bir rasgele örneklem olsun. ML fonksiyonu,

$$L(x, \theta) = \prod_{i=1}^n f(x_i, \theta) \quad (1.40)$$

dır. Bu fonksiyonun $f(x - \theta)$ için logaritması $\rho(x) = -\ln f(x)$ olmak üzere,

$$\ln L(\theta) = \sum_{i=1}^n \ln f(x_i - \theta) = - \sum_{i=1}^n \rho(x_i - \theta) \quad (1.41)$$

dır. Eğer türev olarak maksimum yapabiliyorsak, $\rho'(x) = \psi(x)$ olmak üzere,

$$\frac{d(\ln L(\theta))}{d\theta} = - \sum_{i=1}^n \frac{f'(x_i - \theta)}{f(x_i - \theta)} = \sum_{i=1}^n \psi(x_i - \theta) \quad (1.42)$$

dır. $L(\theta)$ 'yı maksimize eden

$$\sum_{i=1}^n \psi(x_i - \theta) = 0 \quad (1.43)$$

denkleminin çözülmesiyle θ 'nın maksimum likelihood tipi tahmin edicisi veya M-tahmin edicisi elde edilir ve genellikle $\hat{\theta}$ ile gösterilir. M, maksimum likelihood'dan dolaydır (Hogg, R.V., 1979).

Şimdilik ölçek tahmin edicisinin verilerden elde edilebildiğini düşünerek sadece konum tahmin edicisinin elde edilmesiyle ilgileneceğiz. Yani, $\sum \rho(x_i - \theta) = \min$ problemi veya $\sum \psi(x_i - \theta) = 0$ denklem sisteminin çözümünü araştıracağız. (1.39) denklemini w_i ağırlıkları

$$w_i = \frac{\psi(x_i - \theta)}{x_i - \theta}, \quad (x_i - \theta \neq 0) \quad (1.44)$$

olmak üzere

$$\sum w_i(x_i - \theta) = 0 \quad (1.45)$$

şeklinde ağırlıklı formda yazabiliriz. Buradan

$$\sum w_i x_i = \sum w_i \theta \quad (1.46)$$

elde edilir ve son yazılan eşitlik bize θ 'nın

$$\hat{\theta} = \frac{\sum w_i x_i}{\sum w_i} \quad (1.47)$$

biçimindeki örnekleme bağlı ağırlıklar ile elde edilen tahminini verir. Robust istatistiksel analizde kullanılan ρ fonksiyonlarının yapısından dolayı (1.38)'deki fonksiyonun θ 'ya göre minimumunu analitik yöntemlerle bulmak mümkün değildir. Bu yüzden sayısal yöntemler kullanarak çözüm elde edilmeye çalışılır. Herhangi bir

sayısal yöntem (örneğin Newton-Rapson yöntemi) kullanılarak bu çözümün bulunabilmesine rağmen en sık kullanılan yöntem tekrar ağırlıklandırılmış en küçük kareler (reweighted least squares) yöntemidir.

Bir kestiricinin robustlığını (dayanıklılığını) ölçmek için kullanılan iki yöntem vardır. Bu yöntemlerden birincisi etki fonksiyonu ve diğeri ise bozulma noktasıdır (breakdown point). (1.39)'da verilen ψ fonksiyonuna Etki Fonksiyonu denir. Etki fonksiyonu genel olarak yerel dayanıklılığın ölçüsü olarak adlandırılır ve bir tek gözlemin çok büyük veya çok küçük olması durumunun tahmin ediciye etkisini ölçer. Örneğin $\rho(x) = x^2/2$ seçilirse en küçük kareler tahmin edicilerine ulaşılır ve en küçük kareler için etki fonksiyonu $\psi(x) = x$ 'tir. Dikkat edilirse $\psi(x)$ sınırlı bir fonksiyon değildir ve $x \rightarrow \infty$ için $\psi(x)$ 'de sonsuza gider. Bu en küçük kareler yönteminin, tahminler üzerinde verilerin etkisinin, hataların büyüklüğü ile lineer arttığı ve robust olmadığı anlamına gelmektedir. Bu yüzden robust istatistikte seçilen ρ fonksiyonları türevleri sınırlı olacak şekilde seçilirler. Örneğin literatürde kullanılan ve bu tezde de kullanılacak Tablo 1.5'te verilen Huber ve Tukey ρ fonksiyonları gibi. Bozulma noktası ise, genel (global) robustluk ölçüsü olarak tanımlanır ve tahmin ediciyi bozan en küçük sayıdaki sapan değerlerin, örneklem sayısına oranı olarak verilir. Örneklem ortalamasının bozulma noktasının $1/n$ olduğu ve $n \rightarrow \infty$ için sıfır olduğu bilinmektedir. Yani bir tek bozuk gözlem bile örneklem ortalamasının bozulmasına yetecektir. Diğer yandan örneklem medyanının bozulma noktasının %50 olduğu bilinmektedir. Yani verilerin yarısına yakını bozuk olsa bile örneklem medyanı tutarlı bir tahmin edicidir. En küçük kareler için bozulma noktası $1/n$ olup n 'in artan değerleri için sıfıra yaklaşır. Yani en küçük kareler %0'lık bozulma noktasına sahiptir. Buda bize en küçük karelerin sapan değerlere karşı ne kadar hassas olduğunu gösterir.

Bir başka robust yaklaşım ise ε_i hatalarının normal olmayan daha kalın kuyruklu veya çarpık dağılımlardan geldiği varsayımı ile elde edilebilir. Daha öncede belirttiğimiz gibi, hataların normal dağılımdan gelmediği durumlarda normallik varsayımı ile elde edilen tahmin ediciler yanlış sonuçlar verecektir. Örneğin hataların t dağılımından geldiği varsayılırsa maksimum likelihood yöntemiyle konum

(location) parametresi tahmin edebiliriz. Buna göre, $f(x)$, t dağılımının olasılık yoğunluk fonksiyonu olmak üzere, $\rho = -\log f(x)$ olarak seçersek,

$$\rho(x) = \frac{v+1}{2} \cdot \log \left(1 + \frac{x^2}{v} \right) \quad (1.48)$$

olup ψ etki fonksiyonu $\psi(x) = \rho'(x)$ 'ten

$$\psi(x) = \frac{(v+1)x}{v+x^2} \quad (1.49)$$

olup ağırlıklar,

$$w_i = \frac{v+1}{v+(x_i - \theta)^2} \quad (1.50)$$

olarak elde edilir ve $\hat{\theta}$ konum tahmini elde edilir. Benzer şekilde hataların t dağılımından daha esnek bir yapıya sahip olan GT dağılımlarından geldiğini veya hataların çarpık bir dağılımdan geldiği düşünülüyorsa SGT dağılımından geldiği varsayımı daha iyi sonuçlar verebilir. Bu tezde kullanacağımız dağılımların olasılık yoğunluk fonksiyonları ile ρ , ψ ve ağırlıkları Tablo 1.5'te verilmiştir.

GT ve t sapan değerlerden çok fazla etkilenmez diğer yandan normal dağılımın en küçük kareler kestiricisi $\bar{x}$ sapan değerlerden çok etkilenir. Yani $\bar{x}$ kalın kuyruklu dağılımlar için iyi bir tahmin edici değildir. Robust tahmin edicisini verilerin merkezinde normal dağılım gibi kuyruklarda ise başka dağılımlara ait gibi görülen dağılımlar için kullanırız.

θ 'nın $x_1, \dots, x_n$ örnekleme için bulunduğunu ve daha sonrada örneklemden elde edilen $\hat{\theta}$ 'dan sapmaların örneğin üç katı ile değiştirildiğini düşünelim. Bu yeni düzenlenmiş örnekleme için elde edilecek yeni $\hat{\theta}$ çözümü aynı olmayacaktır. Yani tahmin edici (1.18)'de verildiği gibi ölçek eşdeğerli değildir ve bu problemi çözmek

için Huber (1981) konum ve ölçek parametrelerinin (1.51) yardımıyla eşanlı olarak çözülmesini önermiştir. Yani ölçek parametresinin de M tahmin edicisini bulabiliriz. Fakat bu bölümde bizim ilgilendiğimiz konum parametresi olduğundan şimdilik ölçek parametresinin verilerden hesaplanabildiğini, yani bilindiğini varsayarak sadece konum parametresinin tahminini elde etmeye çalışacağız. O halde θ 'nın $\hat{\theta}$ tahmini için ölçek eşdeğerli olanını bulmamız gerekir. Bunun için $\hat{\sigma}$ ölçeğin (σ 'nın) robust tahmini olmak üzere, $\hat{\sigma}$ 'yı bir defa tahmin edip kullanacağız. Bu durumda,

$$\sum_{i=1}^n \psi\left(\frac{x_i - \theta}{\hat{\sigma}}\right) = 0 \quad (1.51)$$

denklemini çözebiliriz. Çözümde kullanılacak ölçek tahmini,

$$\hat{\sigma} = \frac{\text{med} |x_i - \text{med}(x_i)|}{0,6745} \quad (1.52)$$

dir. Paydadaki 0,6745 değeri, eğer n büyük ve örneklem normal dağılımdan geliyorsa $\hat{\sigma}$ 'yı, σ 'nın yansız tahmin edicisi yapar ($\hat{\sigma} \sim \sigma$). Örneklem standart sapması sapan değerlerden çok etkilenen ve robust olmayan bir tahmin edici olduğundan $\hat{\sigma}$ yerine kullanılamaz. Örneklerimizde hataları $\hat{\sigma}$ ile standartlaştırarak iterasyonları yapacağız.

(1.35)'te verdiğimiz simpleks-latis tasarımı için parametre tahmini, tasarım noktalarında yapılan tekrarlı gözlemlerin ortalamasına daylıdır. Ortalamanın sapan değerlere karşı çok hassas olmasından dolayı, sapan değerlere karşı dirençli olan tahmin ediciler olarak robust M tahmin edicileri tasarım noktalarındaki çok tekrarlı gözlemlere uygularsak, $\hat{\beta}^*$, bilinmeyen β parametresinin robust M tahmin edicisi ve $\bar{y}^*$, tasarım noktalarındaki tekrarlı gözlemlerinin ortalamasının robust M tahmin edicisi olmak üzere,

$$\begin{aligned}\hat{\beta}_i^* &= \bar{y}_i^* \\ \hat{\beta}_{ij}^* &= 4\bar{y}_{ij}^* - 2(\bar{y}_i^* + \bar{y}_j^*)\end{aligned}\quad (1.53)$$

robust tahminlerini elde ederiz.

1.4.1. M tahmin Edicisi İçin Bir Örnek

Bu kısımda Tablo 1.3'te verilen tasarım noktalarındaki tekrar sayıları birer arttırılarak oluşturulmuş meyve suyu karışımı verisi ele alınacaktır. Tablo 1.2'deki orijinal ve değiştirilmiş veriler için (1.36) ve (1.37)'de elde edilen katsayı tahminleri ile M tahmin edicilerinden elde edilen katsayı tahminleri karşılaştırılacaktır.

İlk olarak hataların hangi dağılımdan geldiğinin bilinmemesi durumunda Huber ve Tukey etki fonksiyonlarını kullanarak daha sonra da hataların t, GT ve SGT dağılımından geldiğini varsayarak iterasyonları yaptırdığımızda elde edilen tasarım noktalarındaki ortalama sonuçları Tablo 1.4'te verilmiştir. Huber'in ρ fonksiyonu tek minimuma sahiptir. Bu yüzden minimum probleminde çözüme yakınsaması daha kolaydır. Ancak Tukey'in ρ fonksiyonunun minimumu tek değildir. Bu yüzden iterasyonlarda Tukey ρ fonksiyonu için birden fazla başlangıç değeri seçmek daha uygun olacaktır.

Tablo 1.4. Yeni veriler için robust tahmin ediciler ile elde edilen tasarım noktalarındaki ortalama sonuçları

Tasarım Noktası	veri eklenmeden önceki/sonraki ort.	Huber	Tukey	t	GT	SGT
x_1	4.60 / 6.075	4.85	4.60	4.72	4.61	4.40
x_2	6.33 / 5.5	6.48	6.33	6.14	6.34	6.22
x_3	7.10 / 5.675	6.85	7.09	6.97	7.08	6.94
x_1x_2	6.07 / 4.75	5.81	6.06	5.93	6.02	5.85
x_1x_3	6.17 / 4.7	5.86	6.16	6.04	6.13	5.95
x_2x_3	6.17 / 9.225	6.22	6.16	6.22	6.16	6.11

Öncelikle orijinal ve yeni verideki ortalamalara dikkat edersek eklenen verilerin ortalamayı çok etkilediğini görürüz. İterasyonlar sonucunda elde edilen sonuçlara dikkat edilirse yeni veriler için örneklem ortalaması ile robust yöntemlerle

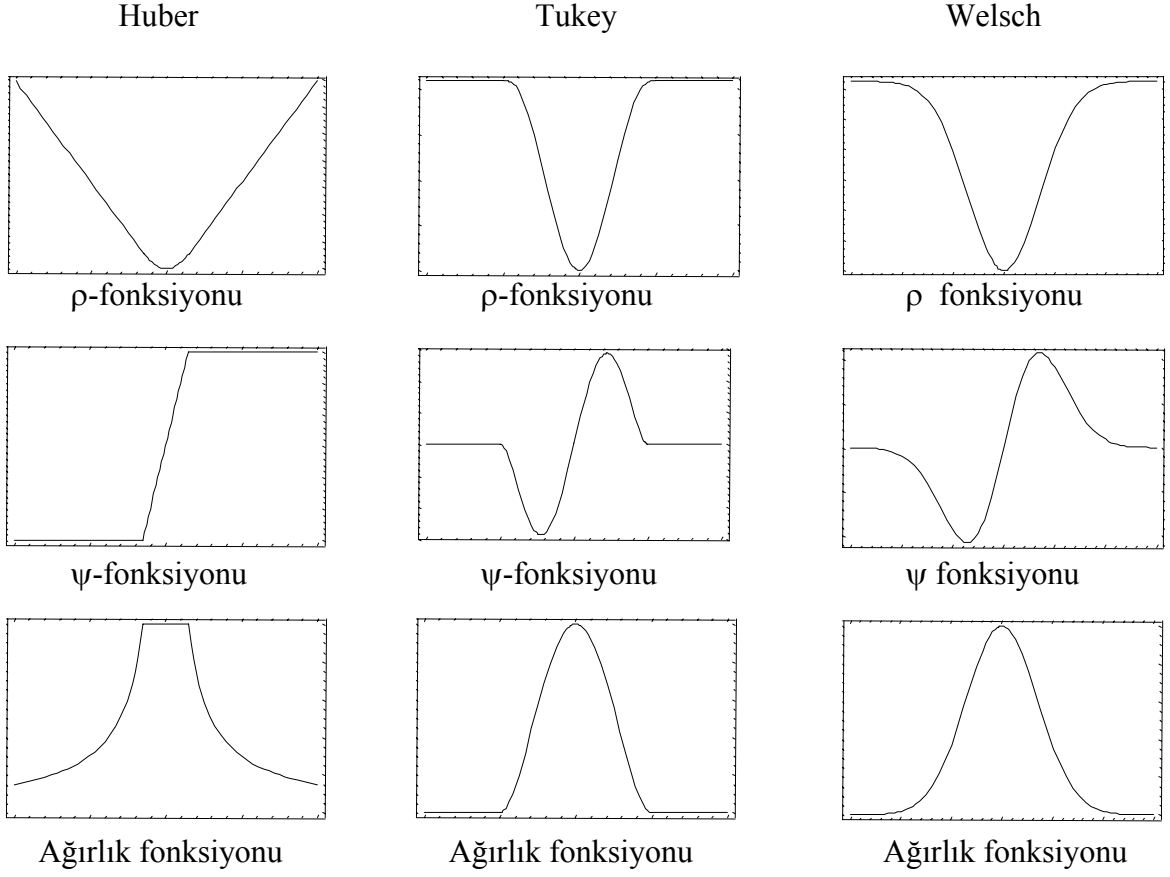
bulunan ortalama arasındaki fark hemen dikkat çekmektedir. Robust yöntemler yeni eklenmiş sapan değerlerden hemen hemen hiç etkilenmemiştir. (1.37)’te klasik yöntem için uydurulan model elde edilmişti. Robust yöntemle (örneğin GT) elde ettiğimiz sonuç için uydurulan model

$$\hat{y} = 4.61x_1 + 6.34x_2 + 7.08x_3 + 2.18x_1x_2 - 1.14x_1x_3 - 2.2x_2x_3 \quad (1.54)$$

olarak elde edilir. Tahmin edilen β katsayılarının sapan değerlerden nasıl etkilendiğini açıkça görmekteyiz. (1.36)’da 3 gözlemlili veri için elde ettiğimiz klasik model ile (1.53)’te elde ettiğimiz 4 gözlemlili sapan değerli veri için robust modele dikkat edilirse birbirlerine yakın olduğu görülür. Yani sapan değerler olmasına rağmen simpleks latis tasarımında robust yöntemler kullanıldığında çok iyi sonuçlar elde edilmektedir.

Tablo 1.5. Bazı dağılımlar ve ρ fonksiyonları için etki ve ağırlık fonksiyonları

t	$F(x) = \frac{\Gamma\left(\frac{v+1}{2}\right)}{\sqrt{\pi v} \Gamma(v/2)} \left(1 + \frac{x^2}{v}\right)^{-\left(\frac{v+1}{2}\right)}, v \text{ serbestlik dereceli}$		
	$\rho(x) = \frac{v+1}{2} \ln\left(1 + \frac{x^2}{v}\right)$	$\psi(x) = \frac{(v+1)x}{v + x^2}$	$w(x) = \frac{v+1}{v + x_i^2}$
GT	$f(x, \mu, \sigma, p, q) = \frac{p}{2\sigma q^{1/p} B(1/p, q)} \left(1 + \frac{ x ^p}{q\sigma^p}\right)^{-q-\frac{1}{p}} \quad \begin{matrix} -\infty < x < \infty \\ \sigma > 0, p > 0, q > 0 \end{matrix}$		
	$\rho(x) = n \left[\ln p + \ln \Gamma\left(q + \frac{1}{p}\right) - \ln \sigma - \ln 2 - \left(\frac{1}{p}\right) \ln q - \ln \Gamma\left(\frac{1}{p}\right) - \ln \Gamma(q) \right] - \left(q + \frac{1}{p}\right) \sum_{i=1}^n \ln\left(1 + \frac{ x_i ^p}{q\sigma^p}\right)$		
	$\psi(x) = \left(\frac{pq+1}{p}\right) \frac{ x ^{p-1}}{1 + \frac{ x ^p}{q\sigma^p}}$	$w(x) = \frac{(pq+1) x ^{p-2}}{q + x ^p}$	
SGT	$f(x, \mu, \sigma, \lambda, p, q) = \frac{p}{2B(1/p, q)q^{1/p}\sigma} \left[1 + \frac{ x ^p}{[1 + \text{sgn}(x)\lambda]^p q\sigma^p}\right]^{-\frac{pq+1}{p}} \quad \begin{matrix} \sigma > 0, p > 0, q > 0, \\ \lambda < 1 \text{ \u00e7arpıklık} \\ \text{parametreli} \end{matrix}$		
	$\rho(x) = n \ln\left(p/2B\left(\frac{1}{p}, q\right)q^{\frac{1}{p}}\right) - n \ln \sigma - \left(\frac{pq+1}{p}\right) \sum_{i=1}^n \ln\left(1 + \frac{ x ^p}{[1 + \text{sgn}(x)\lambda]^p q\sigma^p}\right)$		
	$\psi(x) = \left(\frac{pq+1}{q\sigma^p}\right) \sum_{i=1}^n \frac{[1 + \text{sgn}(x_i)\lambda]^{-p} x_i ^{p-1} \text{sgn}(x_i)}{1 + \frac{ x_i ^p}{[1 + \text{sgn}(x_i)\lambda]^p q\sigma^p}}$	$w(x) = \frac{(pq+1) \frac{ x_i ^{p-2}}{\sigma^{p-2}}}{[1 + \text{sgn}(x_i)\lambda]^p q + \frac{ x_i ^p}{\sigma^p}}$	
Huber	$\rho(x) = \begin{cases} \frac{x^2}{2}, & x \leq k \\ k x - \frac{k^2}{2}, & x > k \end{cases}$	$\psi(x) = \begin{cases} x, & x \leq k \\ k \text{sgn}(x), & x > k \end{cases}$	$w(x) = \begin{cases} 1, & x \leq k \\ \frac{k}{ x - \theta }, & x > k \end{cases}$
Tukey	$\rho(x) = \begin{cases} \frac{c^2}{6} \left(1 - \left[1 - \left(\frac{x}{c}\right)^2\right]^3\right), & x \leq c \\ \frac{c^2}{6}, & x > c \end{cases}$		
	$\psi(x) = \begin{cases} x \left[1 - \left(\frac{x}{c}\right)^2\right]^2, & x \leq c \\ 0, & x > c \end{cases}$	$w(x) = \begin{cases} \left[1 - \left(\frac{x}{c}\right)^2\right]^2, & x \leq c \\ 0, & x > c \end{cases}$	
Welsch	$\rho(x) = \frac{c^2}{2} (1 - e^{\left(-\left(\frac{x}{c}\right)^2\right)})$	$\psi(x) = x e^{\left(-\left(\frac{x}{c}\right)^2\right)}$	$w(x) = e^{\left(-\left(\frac{x}{c}\right)^2\right)}$ (Bu tezde ele alınmayacak)

Şekil 1.9. Huber, Tukey ve Welsch ρ , ψ ve ağırlık fonksiyonlarının grafikleri.

2. BÖLÜM. ROBUST REGRESYON VE KATSAYI TAHMİNİ

Robust istatistik, istatistiksel yöntemlerin sağlamlık teorisidir. Robust istatistik verilerin modelden sapmalarının klasik yöntemlere etkilerini inceler ve gerekiyorsa daha uygun bir yöntem geliştirir. En küçük kareler, regresyon modeli normal dağılıma sahipken çoğu zaman uygun yöntemdir ve iyi istatistiksel özelliklere sahip tahminler verir. Ancak normallikten sapmalar olduğu durumlarda uygun yöntem olmaktan çok uzaktadır. Bu tip durumlarda alternatif tahmin edici olarak robust tahmin edicileri düşünürüz. İstatistiksel çalışmalarda, regresyon tahmin edicisi için başlangıçta kabul edilen istatistiksel model doğru olmazsa bile iyi sonuçlar verebilen yöntemler elde edebilme uğruna hatırı sayılır çabalar sarf edilmiştir. En küçük karelerin normallik varsayımı altında bile en uygun tahmin edici olduğu düşüncesi Tukey (1960)'ın "A survey of sampling from contaminated distributions" adlı çalışması ile son bulmuştur. Daha sonra bu çalışmadan esinlenerek birbirine paralel 4 önemli robust teori, Huber (1964), Huber (1965), Hampel (1968), Rousseeuw (1987) ortaya atılmıştır.

Bu bölümde (1.9)'da verilen n gözlemlili, p parametrelili regresyon modelinin parametre tahminlerini inceleyeceğiz. Birçok istatistiksel yöntem bazı kabullenmeler ile kısıtlanmıştır. Örneğin verilerin dağılımının normal dağılımdan geldiğinin kabul edilmesi gibi. Regresyon analizinde karşılaşılan ve bu kabullenmelerin yanlış olmasına neden olan en büyük problem veri setindeki bazı gözlemlerin diğer gözlemlerden farklı davranması yani sapan değer problemidir. Robust istatistiksel yöntemlerin esas hedefi bu tip sapan değer içeren veri setleri için kullanılabilir, tutarlı sonuçlar veren yöntemler geliştirmektir. Robust kelimesi ilk olarak 1953'te G.E.P. Box tarafından kullanılmıştır. Birinci bölümde konum parametresi için robust tahminler elde etmiştik. Bu (1.9)'da verilen modelde $p = 0$ alınması ile eşdeğerlidir. Yani, $y_i = \beta_0 + \varepsilon_i$ modelindeki β_0 parametresinin tahmin edilmesidir. Klasik yöntemler kullanıldığında β_0 'ın tahmini $\hat{\beta}_0$, $\bar{y}$ örneklem ortalamasına eşit olur ve daha öncede belirttiğimiz gibi bu tahmin normal olmayan dağılımların ürettiği sapan değerlerden çok etkilenir.

2.1. Robust M Regresyon Tahmin Edicileri

Robust yaklaşımlar sapan değerlerin ortaya çıktığı bir çok durumda uygulama alanı bulurlar. Çünkü robust istatistiksel yöntemler başlangıç kabullenmelerinin yanlış olması durumunda veya sapan değer olması durumunda bile tutarlı sonuçlar vermek zorundadırlar. Birçok robust tahmin yöntemi vardır. Bunlar arasında en çok kullanılanlar M-tahmin edicileri (maksimum likelihood tipi tahmin edici), L-tahmin edicileri (sıra istatistiklerinin lineer kombinasyonları), R-tahmin edicileri (ranka dayalı veya rank dönüşümüne dayalı tahmin edici) (Huber, 1981), RM-tahmin edicileri (repeated median-tekrarlı medyan tahmin edicileri) (Siegel, 1982), LMS-tahmin edicileri (medyan karelerinin en küçüğünü kullanan tahmin edici) (Rousseeuw, 1984)'dir. Bu tezde en çok kullanılan robust istatistiksel yöntem olan M-tahmin edicileri ile ilgileneceğiz.

(1.9)'da verilen model için en küçük kareler tahmini (1.11)'de verilmişti. Ancak daha öncede söylediğimiz gibi hataların normal olmadığı veya sapan değerlerin olması durumunda en küçük kareler ile elde edilen istatistiksel analiz sonuçları güvenilir olmayacaktır. Böyle durumlar için robust istatistiksel analiz yöntemleri uygun olacaktır.

En genel halde (1.9)'daki β katsayısı için M tahmin edicisi, $\rho(e)$

$$i) \rho(e) \geq 0 \quad ii) \rho(0) = 0 \quad iii) \rho(e) = \rho(-e) \quad iv) |e_i| > |e_j|, i \neq j \text{ iken } \rho(e_i) \geq \rho(e_j)$$

koşullarını sağlayan bilinen bir fonksiyon olmak üzere

$$\sum_{i=1}^n \rho(e_i) = \sum_{i=1}^n \rho(y_i - x_i' \beta) \quad (2.1)$$

fonksiyonunu minimum yapar. Fonksiyonu minimum yapacak β değerini elde etmek için (2.1)'deki fonksiyonun β 'ya göre türevini alıp sıfıra eşitlenir. O zaman,

$$\sum_{i=1}^n \rho'(y_i - x_i' \beta) x_i' = 0 \quad (2.2)$$

elde edilir. (2.2)'teki fonksiyonu

$$\sum_{i=1}^n \frac{\rho'(e_i)}{e_i} \cdot e_i \cdot x_i = 0, e_i \neq 0 \quad (2.3)$$

olarak yazabiliriz. (2.3)'teki fonksiyonda $w_i = \rho'(e_i)/e_i$ dersek fonksiyon,

$$\sum_{i=1}^n w_i x_i e_i = \sum_{i=1}^n w_i x_i (y_i - x_i' \beta) = 0 \quad (2.4)$$

şeklinde yazılır ve buradan

$$\sum_{i=1}^n w_i x_i y_i = \sum_{i=1}^n w_i x_i x_i' \beta \quad (2.5)$$

normal denklemleri elde edilir. Buna göre β için M tahmin edicisi,

$$\hat{\beta} = \left(\sum_{i=1}^n w_i x_i x_i' \right)^{-1} \sum_{i=1}^n w_i x_i y_i \quad (2.6)$$

dir. Matris formunda (2.5)'teki normal denklemler, W matrisi köşegeninde w_i ağırlıkları bulunan köşegen bir matris olmak üzere,

$$X'WX\beta = X'Wy \quad (2.7)$$

ve $\hat{\beta}$ tahmini

$$\hat{\beta} = (X'WX)^{-1} X'Wy \quad (2.8)$$

olarak elde edilir. ε_i hataları için iki durum ele alınabilir. Birincisi hataların bilinmeyen bir F dağılımından geldiğinin varsayılması ve ikincisi hataların geldiği dağılım formunun bilinmesi durumlarıdır. Bölüm 2.1.1 ve 2.1.2’de bu iki durum ele alınacaktır.

2.1.1. Hataların Dağılım Formunun Bilinmemesi Durumu

M tahmin edicisi hesaplanırken birkaç durum ele alınabilir. İlk olarak hataların geldiği dağılımın bilinmediğini ve σ ölçek parametresinin önceden hesaplanabildiğini düşünelim. σ ölçek parametresinin bir tahmini olarak, $e_i = y_i - \hat{y}_i$ hatalarının MAD’ını yani,

$$\hat{\sigma} = \frac{1}{0.6745} \cdot \text{medyan}|e_i - \text{medyan}(e_i)| = \frac{1}{0.6745} \cdot \text{mad}(e_i) = 1.483 \cdot \text{mad}(e_i) \quad (2.9)$$

kullanabiliriz. Daha sonra e_i hataları $\hat{\sigma}$ ile standartlaştırılır ve β için (2.8)’deki tahmin elde edilir. ρ fonksiyonu olarak Tablo 1.5’te verilen Huber veya Tukey ρ fonksiyonları seçilebilir. İteratif tekrar ağırlıklandırılmış en küçük kareler yöntemi ile M tahmini elde edilir. Ancak Tukey ρ fonksiyonu Huber ρ fonksiyonu gibi minimumu tek olan bir fonksiyon olmadığından iterasyon için seçilen başlangıç değerine dikkat edilmelidir. Bir yakınsama olmuyorsa başka bir başlangıç değeri seçilerek işlemler tekrar yapılmalıdır.

Huber ρ fonksiyonu,

$$\rho(e) = \begin{cases} \frac{e^2}{2} & , -k \leq e \leq k \\ k|e| - \frac{k^2}{2} & , \text{d.y.} \end{cases} \quad (2.10)$$

şeklindedir. i -inci gözlem için $e_i = y_i - x_i'\beta$ olup (2.1)’deki fonksiyonu minimize edeceğiz. Bunun için fonksiyonun β ’ya göre türevini alıp sıfıra eşitlersek

$$\sum_{i=1}^n x_i \rho'(y_i - x_i' \beta) = 0, x_i = (x_{i1}, \dots, x_{ip}), i = 1, \dots, n, 0 = (0, \dots, 0) \quad (2.11)$$

tahmin edici denklem sistemini elde ederiz. Bu p bilinmeyenli denklem sisteminin çözümü iteratif yöntemle elde edilir. En yaygın yöntem tekrar ağırlıklandırılmış en küçük kareler yöntemidir. Bu çalışmada bu yöntem kullanılacaktır. İteratif yönteme ihtiyaç duymamızın nedeni ağırlıkların rezidüleri, rezidülerin tahmin edilen katsayılar ve katsayıların rezidüleri bağlı olmasından kaynaklanır. Dolayısıyla bu iteratif bir çözüm gerektir. Huber ρ fonksiyonu için w_i fonksiyonu,

$$\rho'(e_i) = \begin{cases} e_i & , -k \leq e_i \leq k \\ 2k & , \text{d.y.} \end{cases} \quad (2.12)$$

olmak üzere,

$$w_i = \frac{\rho'(e_i)}{e_i} = \begin{cases} 1 & , -k \leq e_i \leq k \\ \frac{2k}{|e_i|} & , \text{d.y.} \end{cases} \quad (2.13)$$

olarak elde edilir.

Tukey'in ρ fonksiyonu,

$$\rho(e) = \begin{cases} \frac{c^2}{6} \left(1 - \left[1 - \left(\frac{e}{c} \right)^2 \right]^3 \right), & |e| \leq c \\ \frac{c^2}{6}, & |e| > c \end{cases} \quad (2.14)$$

şeklindedir. $c = 5$ veya $c = 6$ seçilebilir. w_i ağırlık fonksiyonu $\rho'(e_i)$,

$$\rho'(e_i) = \begin{cases} e_i \left(1 - \left(\frac{e_i}{c} \right)^2 \right)^2, & |e_i| \leq c \\ 0, & |e_i| > c \end{cases} \quad (2.15)$$

olmak üzere,

$$w_i = \frac{\rho'(e_i)}{e_i} = \begin{cases} \left[1 - \left(\frac{e_i}{c} \right)^2 \right]^2, & |e_i| \leq c \\ 0, & |e_i| > c \end{cases} \quad (2.16)$$

olarak elde edilir ve ağırlıklı en küçük kareler için iteratif yöntem uygulanır.

Şekil 1.9’da verilen Huber ve Tukey ρ fonksiyonlarının grafiklerine dikkat edilirse Tukey ρ fonksiyonu bir süre arttıktan sonra sabitleşirken Huber ρ fonksiyonu artmaya devam etmektedir. Yani Tukey ρ fonksiyonu kullanılarak çözüme daha hızlı ulaşılmasına rağmen daha öncede belirtildiği gibi iterasyonlarda seçilen başlangıç değeri çok önemlidir.

İkinci olarak gene hataların hangi dağılımdan geldiğinin bilinmediğini ve σ ölçek parametresinin de bilinmediğini düşünelim. σ ölçek parametresini β ile eşanlı olarak tahmin edebiliriz. Bu durumda Huber (1981), $\rho_0(0) = 0$ ve $a > 0$ olmak üzere,

$$L(\beta, \sigma) = \frac{1}{n} \sum_{i=1}^n \left[\rho_0 \left(\frac{y_i - x_i' \beta}{\sigma} \right) + a \right] \sigma \quad (2.17)$$

fonksiyonunu β ve σ ’ya göre minimize etmeyi önermiştir (Literatürde bu yaklaşım *Huber’in 2. Yaklaşımı* olarak bilinir). Bunun için (2.17)’deki fonksiyonun β ve σ ’ya göre türevlerini alıp sıfıra eşitlersek,

$$\begin{aligned} \frac{\partial L}{\partial \beta} &= \frac{1}{n} \sum_{i=1}^n \rho'_0 \left(\frac{y_i - x_i' \beta}{\sigma} \right) x_i' = 0 \\ \frac{\partial L}{\partial \sigma} &= \frac{1}{n} \sum_{i=1}^n \rho'_0 \left(\frac{y_i - x_i' \beta}{\sigma} \right) \left(\frac{-(y_i - x_i' \beta)}{\sigma^2} \right) \sigma + \rho_0 \left(\frac{y_i - x_i' \beta}{\sigma} \right) + a = 0 \end{aligned}$$

elde ederiz. Son elde ettiğimiz denklemlerde $\psi_0(x) = \rho'_0(x)$ ve $\chi_0(x) = x\psi_0(x) - \rho'_0$ alırsak,

$$\frac{1}{n} \sum_{i=1}^n \psi_0 \left(\frac{y_i - x_i' \beta}{\sigma} \right) x_i = 0 \quad (2.18)$$

ve

$$\frac{1}{n} \sum_{i=1}^n \chi_0 \left(\frac{y_i - x_i' \beta}{\sigma} \right) = a \quad (2.19)$$

minimizasyon problemi elde edilir. $\rho = (1/2)x^2$ seçilirse en küçük kareler tahmin edicisi elde edilir ve ρ 'nun bu klasik seçimi için normal model varsayımı altında $\hat{\sigma}$ 'nın σ 'nın yansız tahmin edicisi olabilmesi için Huber a parametresinin,

$$a = \frac{n-p}{n} E(\chi_0) \quad (2.20)$$

olarak seçilmesini önermiştir.

Aşağıda vereceğimiz algoritma yardımıyla β ve σ için ağırlıklı en küçük kareler katsayı tahminleri Huber ve Tukey ρ fonksiyonları için yapılır. Algoritma σ 'nın eşanlı tahmin edilip edilmeyeceğine göre 4 veya 6 adımdan oluşabilir.

Algoritma I (Huber ve Tukey için katsayı tahminleri)

Ağırlıklı en küçük kareler için bir algoritma $e_i = y_i - x_i' \beta$ olmak üzere aşağıdaki gibidir.

- i. β_{ilk} ve σ_{ilk} başlangıç değerleri seçilir. β için başlangıç değeri en küçük kareler veya LMS olabilir. σ için ise başlangıç değeri (2.8)'den hesaplanabilir.
- ii. Verilerden e_i ve σ_{ilk} hesaplanır. Rezidüel e_i/σ_{ilk} şeklinde standartlaştırılır.
- iii. w_i ağırlıkları belirlenerek köşegeninde w_i 'ler olan W köşegen matrisi oluşturulur.

- iv. $\mathbf{e}$, $\mathbf{e}_i$ 'lerin bir vektörü olmak üzere $\hat{\boldsymbol{\tau}} = (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{e}$ hesaplanır ve $\beta^{(1)} = \beta_{ilk} + \hat{\boldsymbol{\tau}}$ elde edilir. β_{ilk} yerine $\beta^{(1)}$ alınır.
- v. Buradan $\beta^{(1)}$ ve σ_{ilk} yardımıyla ve $\alpha = E(\psi^2) = 0.7785$ olmak üzere,

$$\sigma^{(1)} = \frac{1}{(n-p)\alpha} \sum_{i=1}^n \psi \left(\frac{y_i - \mathbf{x}_i' \hat{\boldsymbol{\beta}}^{(1)}}{\sigma_{ilk}} \right)^2 (\sigma_{ilk})^2 \quad (2.21)$$

olarak elde edilir.

- vi. σ_{ilk} yerinde $\sigma^{(1)}$ alınarak aynı işlemler tekrar edilir. İterasyona $\hat{\boldsymbol{\tau}}$ sıfıra yaklaşıncaya kadar, örneğin $\hat{\boldsymbol{\tau}}$, 0.00001 olduğunda, son verilir.

Eğer σ eşanlı olarak tahmin edilmek istenmiyorsa algoritmanın v. ve vi. adımları çıkarılır ve algoritmanın başında seçilen σ_{ilk} , σ 'nın bir tahmini olarak alınır ve iterasyon içerisinde yeniden hesaplanmadan kullanılır.

2.1.2. Hataların Dağılım Formunun Bilinmesi Durumu

Bu kısımda hataların geldiği dağılımın simetrik olduğu düşünülüyorsa normal dağılıma alternatif kalın kuyruklu t veya GT dağılımından geldiği varsayılacak ve hataların simetrik olmayan çarpık bir dağılımdan geldiği düşünülüyorsa SGT dağılımından geldiği varsayılacaktır. İlk olarak (1.9)'da verdiğimiz modelde

hataların $\frac{1}{\sigma} f \left(\frac{(y_i - \mathbf{x}_i' \boldsymbol{\beta})^2}{\sigma^2} \right)$ olasılık yoğunluk fonksiyonlarının ailesine sahip bir

dağılımdan geldiği ve σ ölçek parametresinin bilinmediğini düşünelim. Buna göre β ve σ için eşanlı M tahmin edicisini Maksimum Likelihood yöntemiyle bulabiliriz. İlk olarak Maksimum Likelihood fonksiyonunu oluşturalım.

$$L(\beta, \sigma) = \prod_{i=1}^n \frac{1}{\sigma} f \left(\frac{(y_i - \mathbf{x}_i' \boldsymbol{\beta})^2}{\sigma^2} \right) = \frac{1}{\sigma^n} \prod_{i=1}^n f \left(\frac{(y_i - \mathbf{x}_i' \boldsymbol{\beta})^2}{\sigma^2} \right) \quad (2.22)$$

elde ederiz. İşlemleri kolaylaştırması açısından (2.22)'deki fonksiyonun logaritmasını alıp log Likelihood fonksiyonunu oluşturursak,

$$L = \log L(\beta, \sigma) = -n \log \sigma + \sum_{i=1}^n \log \left[f \left(\frac{(y_i - x_i' \beta)^2}{\sigma^2} \right) \right] \quad (2.23)$$

log Likelihood fonksiyonunu β ve σ 'ya göre maksimum yapmak istiyoruz. Bunun için (2.23)'teki L fonksiyonunun β ve σ 'ya göre türevlerini alıp sıfıra eşitlersek,

$$\frac{\partial L}{\partial \beta} = \sum_{i=1}^n \left[\frac{\frac{2(y_i - x_i' \beta)x_i'}{\sigma^2}}{f \left(\frac{(y_i - x_i' \beta)^2}{\sigma^2} \right)} f' \left(\frac{(y_i - x_i' \beta)^2}{\sigma^2} \right) \right] = 0 \quad (2.24)$$

$$\frac{\partial L}{\partial \sigma} = -\frac{n}{\sigma} + \sum_{i=1}^n \left[\frac{\frac{-2\sigma(y_i - x_i' \beta)^2}{\sigma^4}}{f \left(\frac{(y_i - x_i' \beta)^2}{\sigma^2} \right)} f' \left(\frac{(y_i - x_i' \beta)^2}{\sigma^2} \right) \right] = 0 \quad (2.25)$$

tahmin edici denklemleri elde edilir. Bu denklemleri,

$$w_i = \frac{f' \left(\frac{(y_i - x_i' \beta)^2}{\sigma^2} \right)}{f \left(\frac{(y_i - x_i' \beta)^2}{\sigma^2} \right)} \quad (2.26)$$

olmak üzere,

$$\frac{\partial L}{\partial \beta} = \sum_{i=1}^n w_i \cdot \frac{2(y_i - x_i' \beta)x_i'}{\sigma^2} = 0 \quad (2.27)$$

$$\frac{\partial L}{\partial \sigma} = -\frac{n}{\sigma} + \sum_{i=1}^n w_i \cdot \frac{-2(y_i - x_i' \beta)^2}{\sigma^3} = 0 \quad (2.28)$$

olarak yazabiliriz. Buradan (2.27) ve (2.28) denklemlerinden $\hat{\beta}$ için (2.6)'de elde edildiği gibi

$$\hat{\beta} = \left(\sum_{i=1}^n w_i x_i x_i' \right)^{-1} \sum_{i=1}^n w_i x_i y_i \quad (2.29)$$

ve

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n w_i (y_i - x_i' \hat{\beta})^2 \quad (2.30)$$

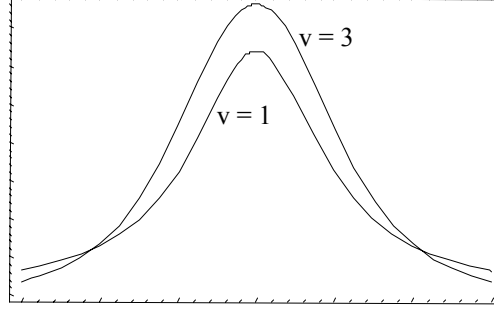
M tahminleri elde edilir. Hataların t, GT ve SGT dağılımlarından geldiğini varsayarak (2.29) ve (2.30)'de bulduğumuz tahminleri benzer şekilde elde ederiz.

2.1.2.1. Hataların t Dağılımından Geldiğinin Varsayılması

(1.9)'daki modelde e_i rasgele hatalarının bağımsız özdeş dağılımlı, sıfır ortalamalı, σ bilinmeyen ölçek parametrelili v serbestlik dereceli t dağılımından ($e_i \sim t(0, \sigma; v)$) geldiğini varsayalım. Öyleyse i-inci gözlem y_i 'nin yoğunluk fonksiyonu

$$f(y_i) = \frac{\Gamma\left(\frac{v+1}{2}\right)}{\Gamma\left(\frac{v}{2}\right) \sqrt{v\pi}} \left(1 + \frac{(y_i - x_i' \beta)^2}{v} \right)^{-\frac{v+1}{2}}, \quad -\infty < y_i < \infty, \quad i = 1, \dots, n \quad (2.31)$$

şeklindedir. Şekil 2.2'de t dağılımının olasılık yoğunluk fonksiyonun grafiği çeşitli v serbestlik derecesi için verilmiştir.



Şekil 2.1. Serbestlik derecesi $v = 3$ ve $v = 1$ için t dağılımı.

Buna göre log likelihood fonksiyonu,

$$l(\beta, \sigma) = \frac{n}{2} \log\left(\frac{1}{\sigma}\right) - \frac{v+1}{2} \sum_{i=1}^n \log\left[v + \frac{(y_i - x_i' \beta)^2}{\sigma^2}\right] \quad (2.32)$$

olur. $l(\beta, \sigma)$ fonksiyonunu β ve σ 'ya göre maksimize etmek için β ve σ 'ya göre türevlerini alıp sıfıra eşitlersek,

$$\frac{\partial l}{\partial \beta} = \frac{v+1}{2} \sum_{i=1}^n \frac{(2(y_i - x_i' \beta)x_i)/\sigma^2}{v + \frac{(y_i - x_i' \beta)^2}{\sigma^2}} = \sum_{i=1}^n \frac{(v+1) \cdot \frac{(y_i - x_i' \beta)}{\sigma^2} \cdot x_i}{v + \frac{(y_i - x_i' \beta)^2}{\sigma^2}} = 0 \quad (2.33)$$

$$\begin{aligned} \frac{\partial l}{\partial \sigma} &= -\frac{n}{2\sigma} - \frac{v+1}{2} \sum_{i=1}^n \frac{(-2\sigma(y_i - x_i' \beta)^2)/\sigma^4}{v + \frac{(y_i - x_i' \beta)^2}{\sigma^2}} = -\frac{n}{2\sigma} + \frac{1}{\sigma^3} \sum_{i=1}^n \frac{(v+1) \cdot (y_i - x_i' \beta)}{v + \frac{(y_i - x_i' \beta)^2}{\sigma^2}} \\ &= \frac{1}{\sigma^3} \left(-\frac{n\sigma^2}{2} + \sum_{i=1}^n \frac{(v+1)(y_i - x_i' \beta)^2}{v + \frac{(y_i - x_i' \beta)^2}{\sigma^2}} \right) = 0 \end{aligned} \quad (2.34)$$

elde edilir. Eğer,

$$w_i = \frac{v+1}{v + \frac{(y_i - x_i' \beta)^2}{\sigma^2}} \quad (2.35)$$

olarak alırsak tahmin edici denklemler,

$$\frac{\partial l}{\partial \beta} = \sum_{i=1}^n w_i x_i' \left(\frac{y_i - x_i' \beta}{\sigma} \right) = 0 \quad (2.36)$$

$$\frac{\partial l}{\partial \sigma} = -n\sigma^2 \sum_{i=1}^n w_i \left(\frac{y_i - x_i' \beta}{\sigma} \right)^2 = 0 \quad (2.37)$$

şeklini alır. Buradan,

$$\hat{\beta} = \left(\sum_{i=1}^n w_i x_i x_i' \right)^{-1} \left(\sum_{i=1}^n w_i x_i y_i \right)$$

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n w_i \left(\frac{y_i - x_i' \beta}{\sigma} \right)^2$$

elde edilir. Ağırlıklı en küçük kareler için algoritma aşağıdaki gibi verilir.

Algoritma II

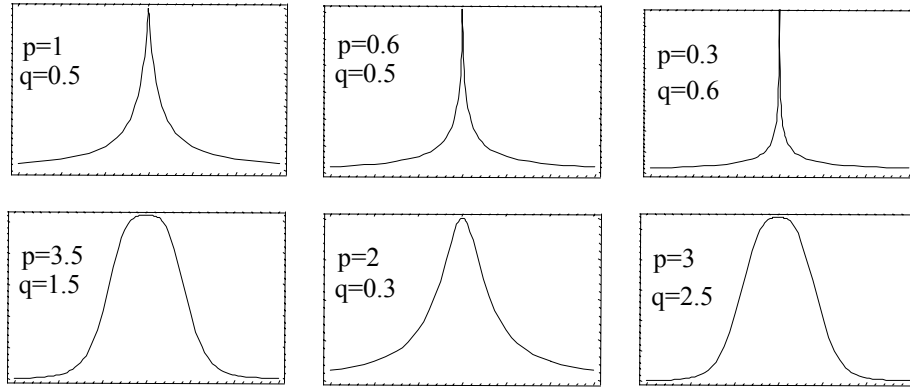
- i. v serbestlik derecesi ($v > 0$) belirlenir ve β ve σ için $\beta^{(0)}$ ve $\sigma^{(0)}$ başlangıç değerleri belirlenir. $x_i' \beta$ ve e_i hesaplanıp $\sigma^{(0)}$ 'a bölünür, v ile toplanır.
- ii. w_i fonksiyonu yardımıyla $w_i^{(0)}$ ağırlıklar belirlenir ve W köşegen matrisi oluşturulur.
- iii. $\beta^{(1)} = (X'WX)^{-1}X'Wy$ hesaplanır.
- iv. $\hat{\sigma}^{(1)}$ için $w_i^{(0)}$ 'lar yardımıyla $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n w_i (y_i - x_i' \beta^{(1)})^2$ 'den $\hat{\sigma}^{(1)}$ bulunur.
- v. $\beta^{(0)}$ yerine $\beta^{(1)}$ ve $\hat{\sigma}^{(0)}$ yerine $\hat{\sigma}^{(1)}$ alınarak aynı işlemler tekrar edilir.

2.1.2.2. Hataların GT Dağılımından Geldiğinin Varsayılması

(1.9)'daki ε_i rasgele hatalarının bağımsız özdeş dağılımlı bilinmeyen σ ölçek parametrelili, sıfır ortalamalı bilinen p ve q şekil parametresine sahip GT dağılımından geldiğini varsayalım. i -inci gözlem y_i 'nin yoğunluk fonksiyonu,

$$f(y_i) = \frac{pq^q}{2\sigma B(1/p, q)} \left(q + \frac{|y_i - x_i' \beta|^p}{\sigma^p} \right)^{-q - \frac{1}{p}}, -\infty < y_i < \infty, i = 1, \dots, n \quad (2.38)$$

dir. GT dağılımı (McDonald and Newey (1988)), regresyondaki hataların modellenmesinde t dağılımına alternatif olarak verilmiştir. p ve q şekil parametrelerinin büyük değerleri için yoğunluk fonksiyonu daha ince kuyruklu ve benzer şekilde küçük değerleri için kalın kuyruklu. Şekil 2.3'te çeşitli p ve q değerleri için GT dağılımının olasılık yoğunluk fonksiyonu verilmiştir. p 'nin 1, 0.6, 0.3 olması durumunda dağılım simetri merkezinde sivridir.



Şekil 2.2. Şekil parametreleri p ve q 'nin değişik değerleri için GT dağılımı.

Buna göre log likelihood fonksiyonu,

$$l(\beta, \sigma) = -n \log \sigma - \left(q + \frac{1}{p} \right) \sum_{i=1}^n \log \left(q + \frac{|y_i - x_i' \beta|^p}{\sigma^p} \right) \quad (2.39)$$

olur. Bunu β ve σ 'ya göre maksimize etmek için β ve σ 'ya göre türevlerini alıp sıfıra eşitlersek,

$$\frac{\partial l}{\partial \beta} = \left(q + \frac{1}{p} \right) \sum_{i=1}^n \frac{\sigma^{-p} \cdot p \cdot x_i' \cdot |y_i - x_i' \beta|^{p-1} \cdot \text{sgn}(y_i - x_i' \beta)}{q + \frac{|y_i - x_i' \beta|^p}{\sigma^p}} = 0, p > 1 \quad (2.40)$$

$$\frac{\partial l}{\partial \sigma} = -\frac{n}{\sigma} + \left(q + \frac{1}{p} \right) \sum_{i=1}^n \frac{p \cdot \sigma^{-p-1} \cdot |y_i - x_i' \beta|^p}{q + \frac{|y_i - x_i' \beta|^p}{\sigma^p}} = 0 \quad (2.41)$$

elde edilir. Elde ettiğimiz bu denklemleri,

$$w_i = \frac{(pq+1) |y_i - x_i' \beta|^{p-2} \hat{\sigma}^{2-p}}{q + |y_i - x_i' \beta|^p \hat{\sigma}^{-p}} \quad (2.42)$$

olmak üzere tahmin edici denklemleri,

$$\sum_{i=1}^n w_i x_i' (y_i - x_i' \beta) = 0 \quad (2.43)$$

$$\sum_{i=1}^n w_i (y_i - x_i' \beta)^2 - n \hat{\sigma}^2 = 0 \quad (2.44)$$

şeklinde yazabiliriz. Buradan,

$$\hat{\beta} = \left(\sum_{i=1}^n w_i x_i x_i' \right)^{-1} \left(\sum_{i=1}^n w_i x_i y_i \right)$$

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n w_i (y_i - x_i' \hat{\beta})^2$$

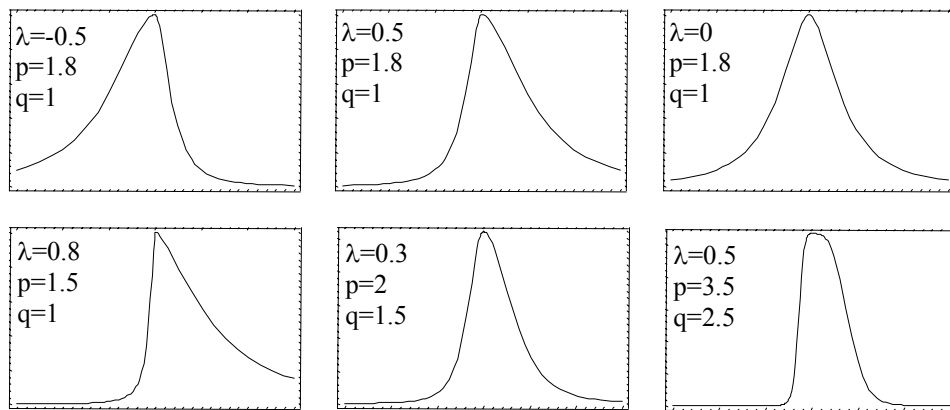
tahminlerini elde ederiz. Bölüm 2.1.2.1'de t dağılımı için verilen algoritma II ile iteratif çözüm ağırlıklı en küçük kareler yardımıyla elde edilir.

2.1.2.3. Hataların SGT Dağılımından Geldiğinin Varsayılması

Son olarak (1.9)'daki ε_i rasgele hatalarının sıfır ortalamalı, bilinmeyen σ ölçek parametrelili, $|\lambda| < 1$ olacak şekilde çarpıklık parametrelili, $p, q > 0$ olacak şekilde bilinen şekil parametrelili SGT dağılımına sahip olduğunu varsayalım. i -inci gözlem y_i 'nin yoğunluk fonksiyonu

$$f(y_i) = \frac{p}{2B(1/p, q)q^{1/p}} \left[1 + \frac{\left| \frac{y_i - x_i' \beta}{\sigma} \right|^p}{\left[1 + \operatorname{sgn} \left(\frac{y_i - x_i' \beta}{\sigma} \right) \lambda \right]^p} q \sigma^p \right]^{\frac{pq+1}{p}} \quad (2.45)$$

dir. SGT dağılımı (Theodossiou, 1998) tarafından verilmiştir. SGT dağılımının parametreleriyle oynayarak birçok dağılıma ulaşabiliriz. Örneğin, $\lambda = 0$ için GT dağılımını, $\lambda = 0$ ve $p = 2$ için t dağılımı, $\lambda = 0$, $p = 2$ ve $q \rightarrow \infty$ için normal dağılımı verir. Şekil 2.4'te SGT dağılımının parametrelerinin çeşitli değerleri için olasılık yoğunluk fonksiyonunun grafiği verilmiştir.



Şekil 2.3. Çeşitli parametre değerleri için SGT dağılımı.

Buna göre log likelihood fonksiyonu,

$$l(\beta, \sigma) = n \log \left(\frac{p}{2B(1/p, q)q^{1/p}} \right) - n \log \sigma - \left(\frac{pq+1}{p} \right) \sum_{i=1}^n \log \left(1 + \frac{\left| \frac{y_i - x_i' \beta}{\sigma} \right|}{\left[1 + \operatorname{sgn} \left(\frac{y_i - x_i' \beta}{\sigma} \right) \lambda \right]^p q \sigma^p} \right) \quad (2.46)$$

olur. $l(\beta, \sigma)$ fonksiyonunu β ve σ 'ya göre maksimize etmek için β ve σ 'ya göre türevlerini alıp sıfıra eşitlersek,

$$\begin{aligned} \frac{\partial l}{\partial \beta} &= \left(\frac{pq+1}{p} \right) \sum_{i=1}^n \left(\frac{p \cdot \left| \frac{y_i - x_i' \beta}{\sigma} \right|^{p-1} x_i \left[1 + \operatorname{sgn} \left(\frac{y_i - x_i' \beta}{\sigma} \right) \lambda \right]^p q \sigma^p}{\left[1 + \operatorname{sgn} \left(\frac{y_i - x_i' \beta}{\sigma} \right) \lambda \right]^{2p} q^2 \sigma^{2p}} \right) = 0 \\ &= \left(\frac{pq+1}{q \sigma^p} \right) \sum_{i=1}^n \left(\frac{\left[1 + \operatorname{sgn} \left(\frac{y_i - x_i' \beta}{\sigma} \right) \lambda \right]^{-p} \left| \frac{y_i - x_i' \beta}{\sigma} \right|^{p-1} x_i}{1 + \frac{\left| \frac{y_i - x_i' \beta}{\sigma} \right|^p}{\left[1 + \operatorname{sgn} \left(\frac{y_i - x_i' \beta}{\sigma} \right) \lambda \right]^p q \sigma^p}} \right) = 0, \quad p > 1 \quad (2.47) \end{aligned}$$

$$\frac{\partial l}{\partial \sigma} = -\frac{n}{\sigma} - \left(\frac{pq+1}{q\sigma^{p+1}} \right) \sum_{i=1}^n \left(\frac{\left[1 + \operatorname{sgn}\left(\frac{y_i - x_i' \beta}{\sigma} \right) \lambda \right]^{-p} \left| \frac{y_i - x_i' \beta}{\sigma} \right|^p}{1 + \frac{\left| \frac{y_i - x_i' \beta}{\sigma} \right|^p}{\left[1 + \operatorname{sgn}\left(\frac{y_i - x_i' \beta}{\sigma} \right) \lambda \right]^p q \sigma^p}} \right) \quad (2.48)$$

$$\frac{\partial L}{\partial \lambda} = \left(\frac{pq+1}{q\sigma^p} \right) \sum_{i=1}^n \left(\frac{\left[1 + \operatorname{sgn}\left(\frac{y_i - x_i' \beta}{\sigma} \right) \lambda \right]^{-p-1} \left| \frac{y_i - x_i' \beta}{\sigma} \right|^p \operatorname{sgn}\left(\frac{y_i - x_i' \beta}{\sigma} \right)}{1 + \frac{\left| \frac{y_i - x_i' \beta}{\sigma} \right|^p}{\left[1 + \operatorname{sgn}\left(\frac{y_i - x_i' \beta}{\sigma} \right) \lambda \right]^p q \sigma^p}} \right) \quad (2.49)$$

elde edilir. Elde ettiğimiz bu denklemleri,

$$w_i = \left(\frac{pq+1}{q\sigma^p} \right) \frac{\left[1 + \operatorname{sgn}\left(\frac{y_i - x_i' \beta}{\sigma} \right) \lambda \right]^{-p} \left| \frac{y_i - x_i' \beta}{\sigma} \right|^{p-2}}{1 + \frac{\left| \frac{y_i - x_i' \beta}{\sigma} \right|^p}{\left[1 + \operatorname{sgn}\left(\frac{y_i - x_i' \beta}{\sigma} \right) \lambda \right]^p q \sigma^p}} \quad (2.50)$$

olmak üzere tahmin edici denklemleri,

$$\sum_{i=1}^n w_i x_i' (y_i - x_i' \beta) = 0 \quad (2.51)$$

$$\sum_{i=1}^n w_i (y_i - x_i' \beta)^2 - \frac{n}{\sigma} = 0 \quad (2.52)$$

şeklinde yazabiliriz. Buradan,

$$\hat{\beta} = \left(\sum_{i=1}^n w_i x_i x_i' \right)^{-1} \left(\sum_{i=1}^n w_i x_i y_i \right) \quad (2.53)$$

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n w_i (y_i - x_i' \hat{\beta})^2 = 0 \quad (2.54)$$

$$\hat{\lambda} = \frac{1}{n\sigma^2} \sum_{i=1}^n w_i (y_i - x_i' \hat{\beta})^2 \operatorname{sgn}(y_i - x_i' \hat{\beta}) \quad (2.55)$$

tahminlerini elde ederiz. Bölüm 2.1.2.1’de t dağılımı için verilen algoritma II yardımıyla ağırlıklı en küçük kareler tahminleri elde edilir.

2.2. Örnekler

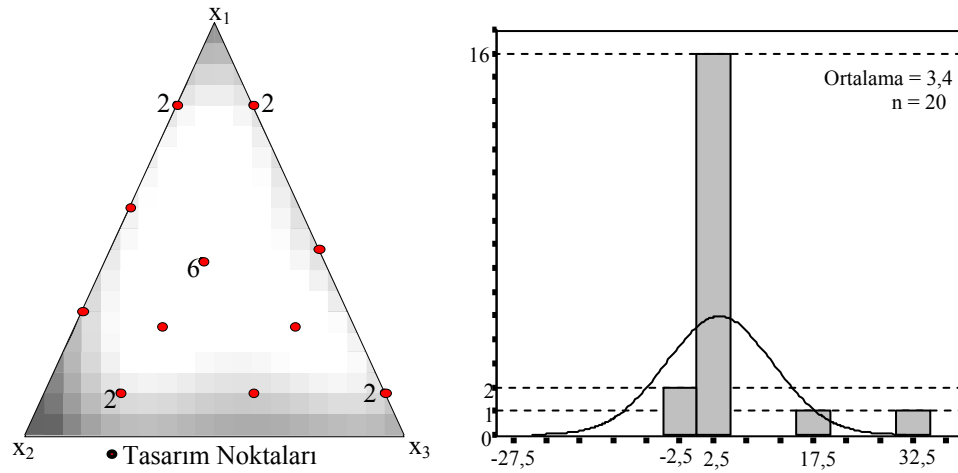
Bu kısımda sapan değerlere sahip ve/veya çarpık olan karma veriler için robust yöntemlerin performansını göstermek için bazı örnekler vereceğiz. Burada ele alacağımız problem sapan değerlere sahip ve/veya çarpık veriler için sapan değerlerden ve/veya normal olmayan çarpık verilerden etkilenmeyen bir model kurabilecek miyiz? Birinci bölümde simpleks latis tasarım noktalarında yapılan tekrarlı gözlemler için bu tezde ele alacağımız robust yöntemlerle sapan değerlerden etkilenmeyen modellerler kurmuştuk. Ancak bu kısımda simpleks latis tasarım noktalarının dışında başka noktalarda da deneyler yapılarak oluşturulmuş veri setlerini ele alacağız. Veriler Design Expert programının simülasyon aracı kullanılarak elde edilmiştir. Örneklerde ilk önce hataların geldiği dağılımın bilinmediğini varsayarak Huber ve Tukey ρ fonksiyonları yardımıyla katsayı tahmini yapıp daha sonra hataların normalden daha kalın kuyruklu ve/veya çarpık bir dağılımdan (t, GT, SGT) geldiğini varsayarak katsayı tahminleri yapacağız ve bunları en küçük kareler yöntemiyle elde edilen tahminler ile karşılaştıracaktır.

Örnek 2.1: Bileşen değişim aralıkları 0 ile 1 arasında olan 20 gözlemlili veri seti Tablo 2.1’de verilmiştir ve simpleks latis tasarım noktalarından başka noktalarda

karışıma sahip bir veri setidir. Tablodaki verilere dikkat edilirse (0.1, 0.7, 0.2), (0.1, 0, 0.9), (0.8, 0, 0.2) ve (0.8, 0.2, 0) noktalarında 2’şer, (0.42, 0.32, 0.26) noktasında 6, ve diğer noktalarda birer tekrardan oluşmaktadır ve 3 ile 18 numaralı denemelerde elde edilen sonuçlar diğerlerinden oldukça farklıdır. Üçgensel koordinatlarda tasarım noktalarının yerleşimi ve ilgili noktadaki tekrar sayısı ile y_i gözlemleri için histogram Şekil 2.5’te verilmiştir.

Tablo 2.1. 20 gözlemlili karma veri seti.

Sıra	x_1	x_2	x_3	Y	Sıra	x_1	x_2	x_3	Y
1	0.30	0.70	0.00	2.6720	11	0.42	0.32	0.26	2.2620
2	0.10	0.70	0.20	-1.0400	12	0.10	0.35	0.55	0.4877
3	0.42	0.32	0.26	16.4400	13	0.80	0.20	0.00	2.6050
4	0.26	0.51	0.23	1.6770	14	0.10	0.70	0.20	1.9210
5	0.42	0.32	0.26	0.7822	15	0.42	0.32	0.26	0.4704
6	0.10	0.00	0.90	0.9480	16	0.42	0.32	0.26	0.4795
7	0.80	0.00	0.20	0.9960	17	0.80	0.00	0.20	2.4210
8	0.42	0.32	0.26	0.3473	18	0.10	0.00	0.90	33.8290
9	0.26	0.16	0.58	-1.3672	19	0.45	0.00	0.55	0.4890
10	0.80	0.20	0.00	1.7030	20	0.55	0.45	0.00	0.5950

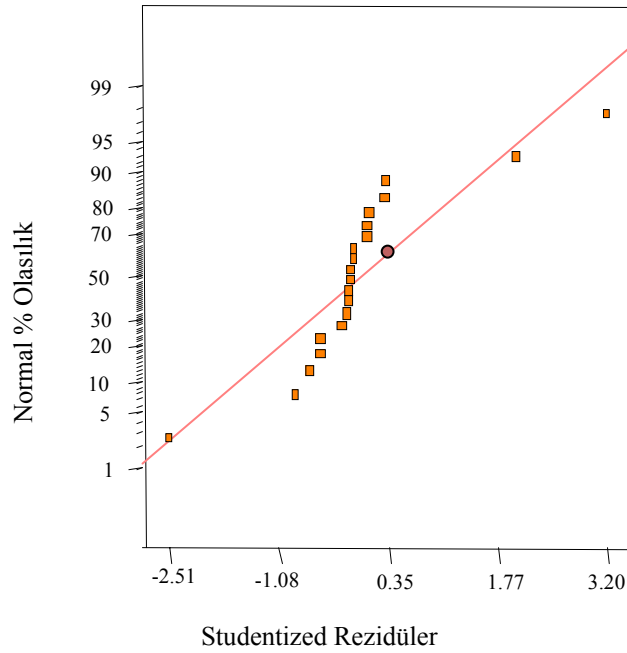


Şekil 2.4. Tablo 2.5’teki tasarım noktalarının tekrar sayısı, üçgensel koordinatlardaki yerleşimi ve y_i gözlemleri için histogram.

Tablo 2.1’deki veriler için en küçük kareler tahmin edicileri ile elde edilen model,

$$\hat{y} = 1.52x_1 + 1.03x_2 + 19.85x_3 + 10.98x_1x_2 - 29x_1x_3 - 35.22x_2x_3 \quad (2.56)$$

şeklindedir. Bu model için $\hat{\sigma} = 7.85$ 'tir. Bu model için normal olasılık grafiği Şekil 2.5'te verilmiştir. Grafiğe dikkat edilirse veriler daha kalın kuyruklu bir dağılımdan gelmektedir ve veri grubundan uzakta bulunan 3 gözlem 3., 6. ve 18. sıradaki gözlemlerdir.

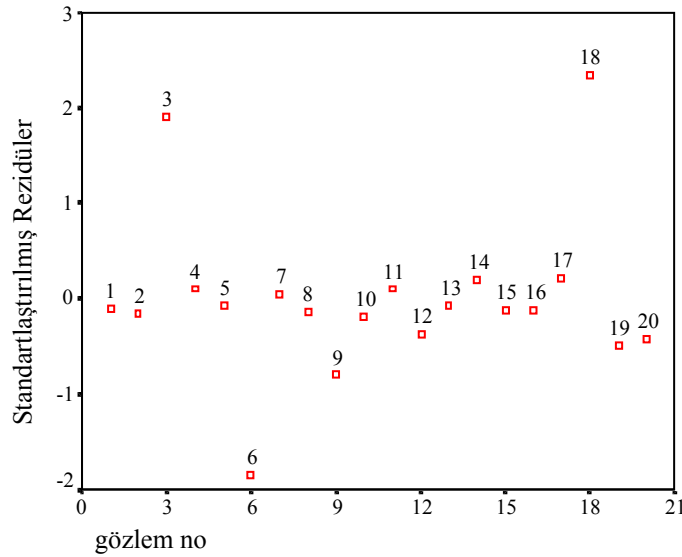


Şekil 2.5. (2.52)'deki model için normal olasılık grafiği (Design Expert 6).

Standartlaştırılmış rezidüler için indeks grafiği Şekil 2.7'de verilmiştir. İndeks grafikleri en küçük kareler ile robust yöntemler ile elde edilen sonuçlar arasında farklılıklar olması durumunda sapan değerleri tespit etmek için kullanılan güvenilir bir araçtır. Standartlaştırma işlemi $e_i = y_i - \hat{y}_i$ rezidülerinin $\hat{\sigma}$ ile bölünmesiyle yapılır. Şekilde 3., 6. ve 18. gözlemler büyük rezidülere sahiptirler.

En küçük kareler ile elde edilen modeli inceledikten sonra şimdide robust yöntemleri ele alalım. Başta hataların geldiği dağılımın bilinmediğini varsayarak Huber ve Tukey ile katsayı tahmin yapacağız ve daha sonra hataların normalden daha kalın kuyruklu t ve GT dağılımlarından geldiğini varsayarak parametre tahmin yapacağız. İlk olarak σ ölçek tahminini verilerden hesaplayıp daha sonrada β ile eşanlı olarak tahmin edeceğiz. σ için bir robust tahmin (2.8)'den elde edilir ve bu

tahmin $\hat{\sigma} = 0.906$ 'dır (bu $\hat{\sigma}$ değerini σ 'nın eşanlı tahmini sırasında başlangıç değer olarak alacağız).



Şekil 2.6. Standartlaştırılmış rezidüleri için indeks grafiği.

Elde ettiğimiz sonuçlar, katsayı tahminlerini karşılaştırmanın daha kolay olması açısından Tablo 2.2 ve Tablo 2.4'te verilmiştir.

Tablo 2.2. σ verilerden hesaplandığında katsayı tahminleri.

Katsayı	$\hat{\beta}_{LS}$ 20 Gözlem	$\hat{\beta}_{LS}$ 18 Gözlem	HUBER	TUKEY	t	GT
	$\hat{\sigma}_{\text{başlangıç}} = 7.85$	$\hat{\sigma}_{\text{başlangıç}} = 1.1$	$\hat{\sigma}_{\text{başlangıç}} = 0.9064$			
β_1	1.524	2.67	2.7025	2.790	3.080	2.621
β_2	1.033	2.47	3.1296	3.005	4.726	4.777
β_3	19.850	0.80	2.2011	0.900	1.283	1.357
$\beta_1\beta_2$	10.982	-3.08	-4.0174	-4.469	-8.550	-7.724
$\beta_1\beta_3$	-29.107	-4.53	-6.4171	-5.039	-6.562	-5.936
$\beta_2\beta_3$	-35.224	-6.40	-8.2392	-6.731	-8.287	-8.809
$\hat{\sigma}_{\text{son}}$			1.879	0.786	0.745	0.840

Elde edilen sonuçlar en küçük karelerden oldukça farklıdır ve en küçük kareler verideki sapan değerlerden çok fazla etkilenmiştir. Ağırlık fonksiyonları rezidülerin azalan fonksiyonlarıdır ve verilerin merkezinden uzaklaşan gözlemler

için (örneğin sapan değerler) küçük ağırlıklar elde edilir. Buna göre σ 'nın başlangıç değeri kullanılıyorken elde edilen ağırlıklar Tablo 2.3'te verilmiştir.

Tablo 2.3. σ verilerden hesaplandığında elde edilen ağırlıklar.

Sıra	Huber	Tukey	t	GT
1	1.000	0.970	1.874	5.460
2	0.512	0.688	0.165	0.193
3	0.078	0.000	0.006	0.008
4	1.000	0.946	1.295	1.808
5	1.000	0.999	1.909	4.545
6	1.000	0.993	1.986	57.66
7	1.000	0.975	1.286	2.776
8	1.000	0.993	1.871	4.589
9	0.626	0.849	0.507	0.581
10	1.000	0.989	1.755	6.768
11	0.824	0.833	0.452	0.512
12	1.000	0.990	1.986	7.274
13	1.000	0.984	1.442	1.479
14	1.000	0.966	1.990	9.253
15	1.000	0.998	1.968	7.568
16	1.000	0.998	1.973	8.005
17	1.000	0.956	1.187	1.099
18	0.037	0.000	0.001	0.001
19	1.000	0.999	1.998	13.22
20	0.934	0.907	0.800	0.990

Tablodaki ağırlık değerlerine dikkat edilirse 3. ve 18. gözlemler en düşük ağırlık değerine sahiptir ve bu değerler sapan değerler olabilir. En küçük kareler ile elde edilen model için ise bu iki gözlem ile beraber 6. gözlemde problemli olarak görünüyordu. Şimdi σ ölçek tahminini β ile eşanlı olarak tahmin ederek elde ettiğimiz sonuçlara bakalım. σ için başlangıç değeri $\hat{\sigma}_{ilk}$ ve iterasyonlar sonucunda elde edilen değerler $\hat{\sigma}_{son}$ ve katsayı tahminleri Tablo 2.4'te verilmiştir.

Tablo 2.4. β ve σ eşanlı olarak tahmin ediliyorken katsayı tahminleri.

Katsayı	$\hat{\beta}_{LS}$ 20 Gözlem	$\hat{\beta}_{LS}$ 18 Gözlem	HUBER	TUKEY	t	GT
β_1	1.524	2.67	2.223	2.936	2.255	2.365
β_2	1.033	2.47	1.839	3.691	4.692	4.753
β_3	19.850	0.80	5.543	0.951	1.426	1.366
$\beta_1\beta_2$	10.982	-3.08	1.626	-6.256	-6.457	-6.989
$\beta_1\beta_3$	-29.107	-4.53	-9.963	-5.544	-5.801	-5.654
$\beta_2\beta_3$	-35.224	-6.40	-12.993	-7.072	-8.867	-8.923
σ tahminleri						
$\hat{\sigma}_{ilk}$	1.1	7.85	0.9064	0.9064	0.9064	0.9064
$\hat{\sigma}_{son}$			3.5063	0.7157	0.2991	0.6150

Burada da elde edilen sonuçlar en küçük karelerden farklı olduğu gibi iterasyonlarda σ 'nın başlangıç tahmini ile sabit tutulduğunda elde edilen sonuçlara yakın sayılabilir. Tablo 2.5'te elde edilen ağırlıklar verilmiştir.

Tablo 2.5. σ eşanlı hesaplanıyorken elde edilen ağırlıklar.

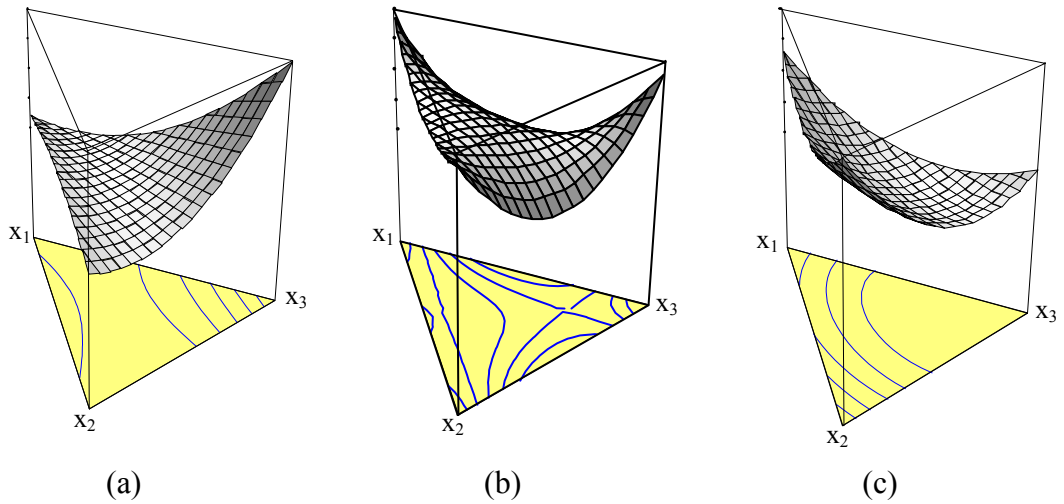
Sıra	Huber	Tukey	t	GT
1	1.000	0.970	1.906	5.728
2	1.000	0.412	0.019	0.094
3	0.306	0.000	0.000	0.003
4	1.000	0.931	0.385	1.200
5	1.000	0.997	1.376	3.547
6	1.000	0.990	1.966	20.08
7	1.000	0.953	1.531	2.968
8	1.000	0.991	1.243	3.403
9	1.000	0.765	0.070	0.310
10	1.000	0.984	1.999	13.19
11	1.000	0.731	0.061	0.271
12	1.000	0.992	1.843	5.521
13	1.000	0.971	0.200	0.770
14	1.000	0.983	1.909	7.449
15	1.000	0.997	1.760	5.808
16	1.000	0.997	1.795	6.141
17	1.000	0.937	0.107	0.507
18	0.160	0.000	0.000	0.000
19	1.000	0.999	1.701	7.006
20	1.000	0.866	0.125	0.539

Elde edilen ağırlıklara dikkat edilirse burada da 3. ve 18. gözlemler en küçük ağırlık değerlerine sahiptir ve bu değerler sapın değerleridir. Şimdi verideki en

küçük ağırlık değerine sahip 2 sapan değeri çıkararak 18 gözlem değeri için en küçük kareler tahmin edicileri kullanarak elde edilen model,

$$\hat{y} = 2.67x_1 + 2.47x_2 + 0.8x_3 - 3.08x_1x_2 - 4.53x_1x_3 - 6.4x_2x_3 \quad (2.57)$$

şeklindedir. Bu model için $\hat{\sigma} = 1.11$ 'dir. Bu model (1.52)'de en küçük kareler ile 20 gözlem kullanarak elde ettiğimiz modelden oldukça farklıdır. Ayrıca robust yöntemlerle elde edilen modeller ile çok benzerdir. Aşağıdaki şekilde 20 ve 18 gözlemlili modeller için ve GT ile elde ettiğimiz model için yüzey grafikleri verilmiştir.

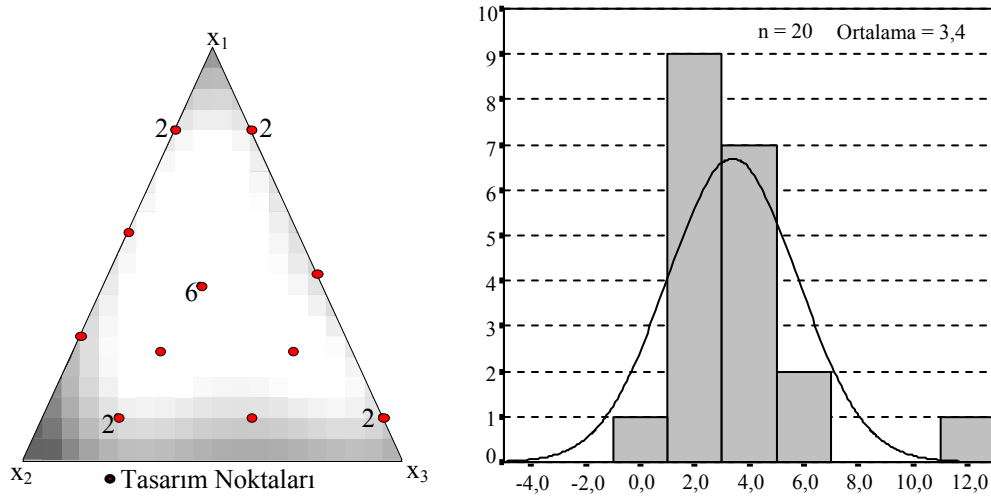


Şekil 2.7. Yüzey Grafikleri (a) 20 gözlem (b) 18 gözlem (c) GT

Örnek 2.2: Bu örneğimizde gene bileşen değişim aralıkları 0 ile 1 arasında olan 20 gözlemlili veri seti kullanacağız. Ancak bu sefer veriler normalden daha kalın kuyruklu değil de normalden daha çarpık bir dağılıma sahiptir. Veriler Tablo 2.6'da verilmiştir. Üçgensel koordinatlarda tasarım noktalarının yerleşimi ve ilgili noktadaki tekrar sayısı ile y_i gözlemleri için histogram Şekil 2.9'da verilmiştir.

Tablo 2.6. 20 gözlemlik karma veri seti.

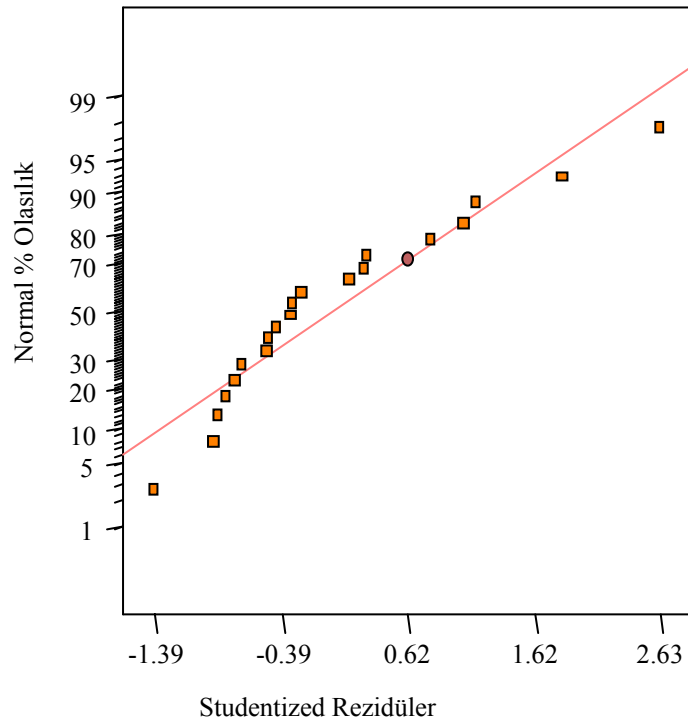
Sıra	x_1	x_2	x_3	y	Sıra	x_1	x_2	x_3	y
1	0.30	0.70	0.00	5.5110	11	0.42	0.32	0.26	1.8900
2	0.10	0.70	0.20	2.7910	12	0.10	0.35	0.55	8.6130
3	0.42	0.32	0.26	3.1630	13	0.80	0.20	0.00	1.5210
4	0.26	0.51	0.23	3.6090	14	0.10	0.70	0.20	2.1200
5	0.42	0.32	0.26	1.9920	15	0.42	0.32	0.26	7.6870
6	0.10	0.00	0.90	2.8040	16	0.42	0.32	0.26	3.7300
7	0.80	0.00	0.20	0.4169	17	0.80	0.00	0.20	1.6783
8	0.42	0.32	0.26	0.5283	18	0.10	0.00	0.90	1.8970
9	0.26	0.16	0.58	6.9420	19	0.45	0.00	0.55	1.5447
10	0.80	0.20	0.00	1.7140	20	0.55	0.45	0.00	12.6720

**Şekil 2.8.** Tablo 2.6'daki tasarım noktalarının tekrar sayısı, üçgensel koordinatlardaki yerleşimi ve y_i gözlemleri için histogram.

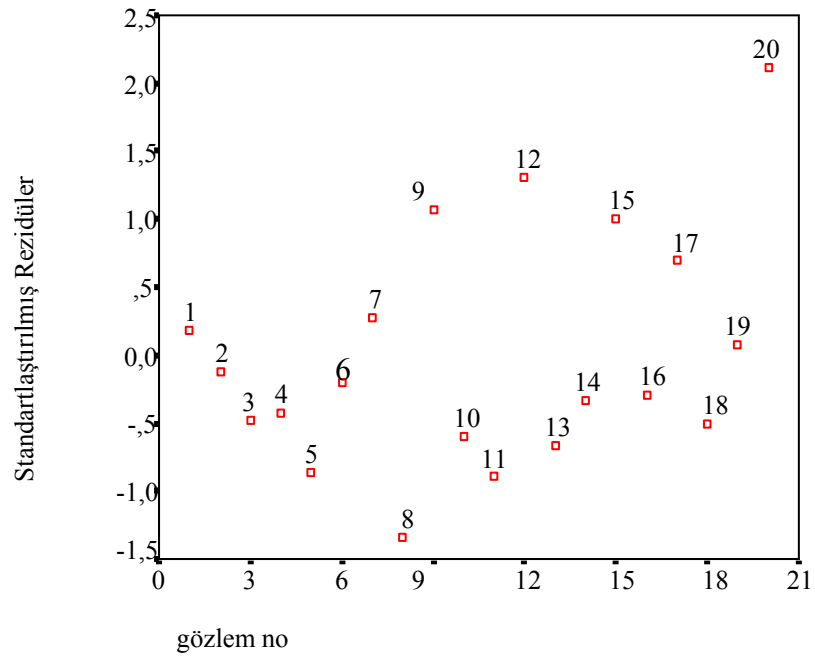
Tablo 2.6'daki veriler için en küçük kareler tahmin edicileri ile elde edilen model,

$$\hat{y} = -0.776x_1 - 0.732x_2 + 4.311x_3 + 27.169x_1x_2 - 3.353x_1x_3 - 8.813x_2x_3 \quad (2.58)$$

şeklindedir. Bu model için $\hat{\sigma} = 9.609$ 'tir. Bu model için normal olasılık grafiği Şekil 2.10'da verilmiştir. Grafiğe dikkat edilirse veriler normalden farklı çarpık (pozitif yönde çarpık) dağılımdan gelmektedir ve veri grubundan uzakta bulunan 3 gözlem 8., 12. ve 20. sıradaki gözlemlerdir. Şekil 2.11'de standartlaştırılmış rezidüleri için indeks grafiği verilmiştir.



Şekil 2.9. (2.54)'deki model için normal olasılık grafiği (Design Expert 6).



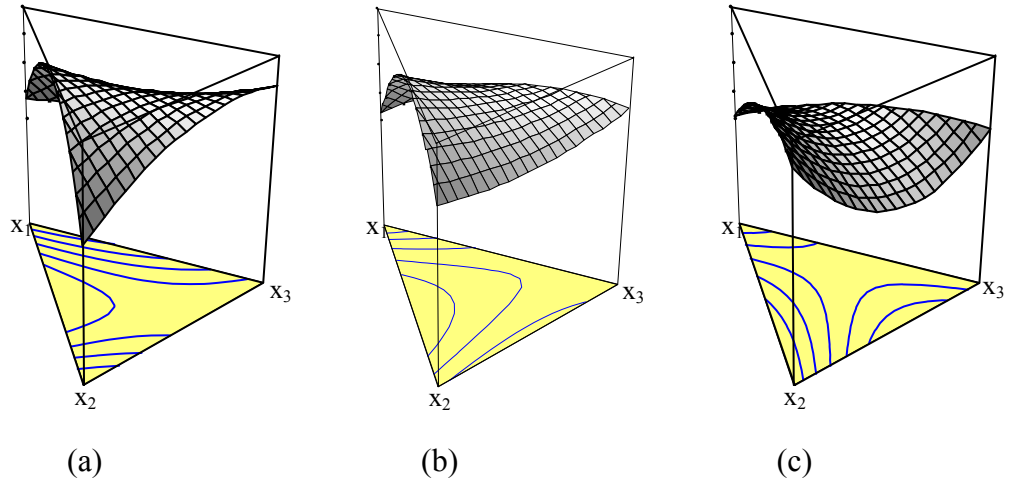
Şekil 2.10. Standartlaştırılmış rezidümler için indeks grafiği.

Şekil 2.10'daki normal olasılık grafiğinden çarpık dağılım problemi olduğunu belirledikten sonra hataların çarpık SGT dağılımından geldiğini varsayarak katsayı tahmini yapabiliriz. Elde ettiğimiz katsayı tahminleri ile λ çarpıklık parametresinin ilk ve son değerleri ile σ ölçek parametresinin ilk ve son değerleri Tablo 2.7'de verilmiştir. $\lambda = 0$ için SGT, GT ile aynı sonuca yakınsar ve $\lambda = 0$ ve $p = 2$ için SGT, t ile aynı sonuca yakınsar.

Tablo 2.7. SGT için katsayı ve parametre tahminleri.

Katsayı	$\hat{\beta}_{LS}$ 20 Gözlem	SGT	t	GT
β_1	-0.776	-1.268	-1.480	-1.768
β_2	-0.732	5.135	4.237	4.146
β_3	4.311	1.791	1.998	1.762
$\beta_1\beta_2$	27.169	9.475	13.399	14.499
$\beta_1\beta_3$	-3.353	4.579	6.630	9.046
$\beta_2\beta_3$	8.813	-17.523	-10.811	-10.592
σ ve λ tahminleri				
$\hat{\sigma}_{ilk}$	9.609	2.173	2.173	2.173
$\hat{\sigma}_{son}$		0.912	0.654	0.615
λ_{ilk}		0.5		
λ_{son}		0.73		

t ve GT için elde edilen sonuçlar en küçük karelerden farklı olduğu gibi katsayıların işaretlerinde de farklılıklar vardır. Aynı şekilde SGT içinde elde edilen sonuçlara en küçük kareler ile elde edilen sonuçlardan farklıdır ve burada da katsayıların işaretleri farklılıklar göstermektedir. Şekil 2.12'de SGT, GT ve en küçük kareler için yüzey grafikleri verilmiştir.



Şekil 2.11. Yüzey Grafikleri (a) en küçük kareler (b) GT (c) SGT

3. BÖLÜM. KARMA DENEMELER İÇİN ROBUST RIDGE VE ROBUST LIU TAHMİN EDİCİLERİ

Karma denemelerde kötü koşulluluk veya iç ilişki problemi araştırmacıların sıkça karşılaştıkları ve önemle üzerinde durulması gereken bir problemdir. Karma denemeler, kanonik polinom olarak adlandırılan özel polinom modeline gereksinim duyarlar. Karma denemelerde polinomlar genelde ek kısıtlamalara maruz kaldıklarından iç ilişki ya da kötü koşulluluk olarak bilinen istenmeyen durumlar ortaya çıkar.

Karma sistemlerde sistemi oluşturan bileşenler üzerinde sırasıyla (1.1) ve (1.2)'deki doğal kısıtlamaların olduğu belirtilmiştir. Çoğu zaman bileşenler üzerinde bu kısıtlamaların yanı sıra bileşenlerden bazıları veya tümü üzerinde alt ve üst sınırlar şeklinde ve/veya bileşenlerin lineer kombinasyonları üzerinde (1.3) ve (1.4)'deki ek kısıtlamalarda bulunabilir. Bileşenler üzerindeki bu ek kısıtlamalar karma problemde çoğu kez kötü koşulluluğa ya da iç ilişkiye neden olur. Kötü koşulluluğun yanı sıra veride y yönünde sapan değerlerin varlığı da söz konusu olabilir. Böyle durumlarda en küçük kareler ile elde edeceğimiz istatistiksel sonuçlar bizi çok yanlış sonuçlara götürebilir.

Bu bölümde, karma sistemlerde sapan değer ve iç ilişki probleminin aynı anda olması durumunda robust tahmin yöntemlerinin etkilerini inceleyeceğiz. Kötü koşulluluk problemi ile mücadelede alternatif yöntemlerden ikisi Ridge ve Liu yöntemleridir. Bu iki tahmin edici en küçük karelerden elde edilen katsayı tahminlerini kullanırlar. Ancak en küçük kareler y yönünde sapan değerlere karşı çok hassastır ve dolayısıyla y yönünde sapan değerler olması durumunda Ridge ve Liu tahminleri de hassas olacaktır. Bunun için y yönündeki sapan değerlerden etkilenmeyen robust yöntemleri, çoklu iç ilişkinin etkisini azaltan Ridge ve Liu ile birlikte kullanarak daha dayanıklı tahminler elde edebiliriz. Bu bölümde, kötü koşulluluk ile sapan değer problemine sahip karma veriler için robust Ridge ve robust Liu tahmin edicilerin etkileri incelenecektir.

3.1. Kötü Koşulluluğu Belirleme Yöntemleri ve Standartlaştırma

(1.9)'da verilen Scheffé kanonik polinom modeli matris formunda,

$$y = X\beta + \varepsilon \quad (3.1)$$

ile gösterilir. (3.1)'deki modelde, X , $n \times p$ ($p \geq q$) tipinde ve i -inci satırı, i -inci gözleme karşılık gelen sütunlarında p tane x_i değişkeninin değerlerini içeren matrisi; y , $n \times 1$ tipinde amaç değişkeni üzerinde gözlemlerin bir sütun vektörünü; β , $p \times 1$ tipinde tahmin edilecek parametrelerinin bir sütun vektörünü; ε , $n \times 1$ tipinde $E(\varepsilon) = 0$ ve $\text{Var}(\varepsilon) = \sigma^2 I$ özelliklerini sağlayan rasgele hataların bir sütun vektörünü göstermektedir. Bu model için en küçük kareler tahmin edicisi

$$\hat{\beta}_{LS} = (X'X)^{-1}X'y \quad (3.2)$$

ve $\hat{\beta}_{LS}$ 'nin varyans-kovaryans matrisi,

$$\text{Var}(\hat{\beta}_{LS}) = \sigma^2(X'X)^{-1} \quad (3.3)$$

şeklindedir. $x_1, x_2, \dots, x_p$ yukarıdaki modeldeki X matrisinin p sütunu olsun. X matrisinin sütunları arasında tam bir lineer bağımlılık varsa; yani,

$$\sum_{j=1}^p c_j x_j = 0 \quad (3.4)$$

olacak şekilde tümü sıfır olmayan c_j 'ler varsa X matrisinin rankı p 'den küçüktür. Bu durumda $X'X$ matrisi singülerdir ve tersi yoktur. Bununla birlikte

$$\sum_{j=1}^p c_j x_j \cong 0 \quad (3.5)$$

ise çoğu kez kötü koşulluluk olarak tanımlanan durum oluşur. Kötü koşulluluk ya da iç ilişkinin istenmeyen etkileri vardır.

Bağımsız değişkenler arasında çoklu iç ilişkinin ya da kötü koşulluluğun görülmesinin temel olarak üç nedeni vardır. Birincisi, deney tasarımında yeterli planlama yapılmamasından veya araştırmanın zayıf gözlemsel verilere dayandırılmasından kaynaklanan iç ilişki problemidir. İkincisi, bağımsız değişkenlerin kuvvetleri veya çarpımları gibi matematiksel işlemler sonucu oluşturulan yeni bir değişken nedeniyle ortaya çıkan yapısal veya model belirlenmesinden kaynaklanan iç ilişkidir. Yapısal çoklu iç ilişki, modeldeki bileşen sayısının gözlem sayısından fazla olduğu durumlarda da görülür. Üçüncüsü ise, bağımsız değişkenler üzerindeki kısıtlamalardan kaynaklanan iç ilişkidir. Özellikle karma denemelerde bileşenler üzerinde genellikle alt ve üst sınırlar şeklinde kısıtlamaların varlığı şiddetli iç ilişki probleminin doğmasına ve en küçük karelerden elde edilen sonuçların tutarsız olmasına neden olacaktır.

Kötü koşulluluğun varlığının belirlenmesinde en çok kullanılan 2 yöntem VIF_i (Varyans şişirme faktörleri) ve koşul sayısı yöntemleridir. (VIF) 'ler, $(X'X)^{-1}$ matrisinin köşegen elemanlarıdır. Brownlee (1965), R_i^2 , i-inci gözlem çıkarıldıktan sonra elde edilen R^2 değeri olmak üzere,

$$VIF_i = \frac{1}{1 - R_i^2} \quad (3.6)$$

olduğunu göstermiştir. Bu, büyük R_i^2 (örneğin $R_i^2 > 0,99$) değerlerinden kaçınmanın büyük VIF_i 'den kaçınmaya denk olduğunu gösterir. Gorman (1970), $R_i^2 > 0,99$ ise kötü koşulluluk problemi olduğunu belirtmiştir. Bu $VIF_i > 100$ olması ile aynı anlama gelmektedir. Marquardt (1970), VIF 'lerin herhangi birinin 10'dan büyük olması durumunda en küçük kareler kestiricilerinin kullanılması ile elde edilen tahminlerin kararlı olmadıklarını bunun yerine alternatif model oluşturulması gerektiğini veya alternatif tahmin edicilerin kullanılması gerektiğini belirtmiştir. Alternatif tahmin edici olarak Ridge Regresyon tahmin edicisi ve Liu tahmin edicisi

kullanılabilir. Kötü koşulluluğun belirlenmesinde uygulanabilecek diğer bir yöntem, $X'X$ matrisinin özdeğerlerine dayalı koşul sayısı yöntemidir. Sıfırdan farklı özdeğerlerin sayısı matrisin rankını verir. $X'X$ matrisinin özdeğerlerinden biri sıfıra eşitse $X'X$ matrisi singülerdir ve böylece X matrisinin sütunları lineer bağımlıdır. Özdeğerler elde edildikten sonra en büyük özdeğer ile en küçük özdeğerin oranı alınarak $X'X$ matrisinin koşul sayısı elde edilir.

$$\text{Koşul sayısı} = \frac{\lambda_{\max}}{\lambda_{\min}} \quad (3.7)$$

Belsley ve ark. (1980), koşul sayısının 25'den büyük olduğu durumların araştırılması gereken durumlar olduğunu belirtmişlerdir. Genel olarak koşul sayısı 100'den küçük ise, çoklu iç ilişki probleminin ciddi boyutlarda olmadığı, 100 ile 1000 arasında ise, güçlü bir iç ilişki problemi olduğu, 1000'den büyük ise, çok ciddi bir iç ilişki problemi olduğu söylenir.

Birçok istatistiksel paket programın oluşturulmuş modelin en küçük kareler kestiricilerini hesaplarken değişkenleri standartlaştırıp daha sonra $X'X$ matrisini oluşturur. Değişkenleri standartlaştırmanın birçok faydası vardır. İlk olarak $X'X$ matrisinin tersi alınırken oluşan yuvarlama hatalarını azaltır. Ayrıca değişkenlerin ve katsayıların yorumlanabilirliğini artırır. Özellikle x_i 'ler farklı birimlerde ölçümler içeriyorsa daha kolay yorumlanabilen sonuçlar verir. Standartlaştırılmış değişkenler yardımıyla farklı değişkenlerin regresyon katsayılarını direk olarak karşılaştırmakta mümkün olmaktadır. x_i 'ler

$$w_{ij} = \frac{x_{ij} - \bar{x}_i}{\sqrt{\sum_{i=1}^n (x_{ij} - \bar{x}_i)^2}} \quad (3.8)$$

şeklinde standartlaştırılırlar. Standartlaştırılmış değerler cinsinden model

$$y = \alpha^* 1 + W\beta^* + \varepsilon \quad (3.9)$$

dur. Karma denemelerde model sabit terim içermediğinden değişkenleri

$$w_{ij} = \frac{x_{ij}}{\sqrt{\sum_{i=1}^n x_{ij}^2}} \quad (3.10)$$

şeklinde standartlaştırırız. Bu durumda model,

$$y = W\beta^* + \varepsilon \quad (3.11)$$

şeklinde olur. Örneğin,

$$y = \sum_{i=1}^n \beta_i x_i + \sum_{i=1}^n \beta_{ij} x_i x_j + \varepsilon \quad (3.12)$$

ikinci dereceden Scheffé karma modelini ele alalım. Bu modeldeki lineer terimler ve karesel terimler sırasıyla,

$$w_{ui} = \frac{x_{ui}}{\sqrt{\sum_{i=1}^n x_{ui}^2}} \quad \text{ve} \quad w_{ui}w_{uj} = \frac{x_{ui}x_{uj}}{\sqrt{\sum_{i=1}^n x_{ui}^2 x_{uj}^2}} \quad (3.13)$$

şeklinde standartlaştırılırlar. Buna göre standartlaştırılmış değişkenler cinsinden 2. dereceden 3 değişkenli Scheffé tipi karma model,

$$y = \beta_1^* w_1 + \beta_2^* w_2 + \beta_3^* w_3 + \beta_{12}^* w_1 w_2 + \beta_{13}^* w_1 w_3 + \beta_{23}^* w_2 w_3 \quad (3.14)$$

olarak elde edilir.

Örnek 3.1. Örnek 2.1’de standartlaştırma işlemi uygulanmadan detaylı olarak incelemesi yapılan Tablo 2.1’de verdiğimiz 3 bileşenli 2 sapan değere sahip karma veriyi ele alarak standartlaştırma işleminin etkilerini inceleyelim. Tablo 3.1’de (3.13)’te verilen bağıntılar kullanılarak standartlaştırılmış veriler verilmiştir.

Tablo 3.1. Tablo 2.1’de verilen karma verinin standartlaştırılmış şekli.

Sıra	x_1	x_2	x_3	x_1x_2	x_1x_3	x_2x_3	y
1	0.1430	0.4202	0.0000	0.3865	0.0000	0.0000	2.6720
2	0.0477	0.4202	0.1122	0.1288	0.0417	0.3740	-1.0400
3	0.2003	0.1921	0.1458	0.2473	0.2275	0.2222	16.4390
4	0.1240	0.3061	0.1290	0.2440	0.1246	0.3133	1.6770
5	0.2003	0.1921	0.1458	0.2473	0.2275	0.2222	0.7822
6	0.0477	0.0000	0.5047	0.0000	0.1875	0.0000	0.9480
7	0.3814	0.0000	0.1122	0.0000	0.3334	0.0000	0.9960
8	0.2003	0.1921	0.1458	0.2473	0.2275	0.2222	0.3473
9	0.1240	0.0960	0.3253	0.0766	0.3142	0.2479	-1.3672
10	0.3814	0.1201	0.0000	0.2944	0.0000	0.0000	1.7030
11	0.2003	0.1921	0.1458	0.2473	0.2275	0.2222	2.2620
12	0.0477	0.2101	0.3084	0.0644	0.1146	0.5142	0.4877
13	0.3814	0.1201	0.0000	0.2944	0.0000	0.0000	2.6050
14	0.0477	0.4202	0.1122	0.1288	0.0417	0.3740	1.9210
15	0.2003	0.1921	0.1458	0.2473	0.2275	0.2222	0.4704
16	0.2003	0.1921	0.1458	0.2473	0.2275	0.2222	0.4795
17	0.3814	0.0000	0.1122	0.0000	0.3334	0.0000	2.4210
18	0.0477	0.0000	0.5047	0.0000	0.1875	0.0000	33.8293
19	0.2146	0.0000	0.3084	0.0000	0.5157	0.0000	0.4890
20	0.2622	0.2701	0.0000	0.4555	0.0000	0.0000	0.5950

Tablo 3.2’de 20 gözlemlili orijinal verilerin standartlaştırılmış ve standartlaştırılmamış şekilleri için elde edilen sonuçlar verilmiştir. Standartlaştırılmış değişkenler cinsinden elde edilen sonuçlara dikkat edilirse MSE ve koşul sayısındaki iyileşme açıkça görülmektedir. Etkileşim terimlerinin standart hatalarında ve VIF’lerindeki iyileşme dikkat çekicidir. Standartlaştırma işlemi göz ardı edilmeyecek kadar iyi sonuçlar elde etmemizi sağlamıştır. Bu yüzde yukarıda bahsettiğimiz avantajları göz önüne alarak bundan sonra tüm hesaplamalarımızı standartlaştırılmış değişkenler cinsinden yapacağız.

Tablo 3.2. 20 gözlemlili orijinal veriler için sonuçlar.

Terim	Standartlaştırılmamış				Standartlaştırılmış			
	$\hat{\beta}_{LS}$	$se(\hat{\beta}_{LS})$	Özdeğer	VIF _i -ler	$\hat{\beta}_{LS}$	$se(\hat{\beta}_{LS})$	Özdeğer	VIF _i -ler
$\hat{\beta}_1$	1.524	8.999	7.259	1.314	3.206	18.876	3.823	5.781
$\hat{\beta}_2$	1.033	13.866	2.073	3.118	1.731	23.098	1.070	8.657
$\hat{\beta}_3$	19.851	7.797	1.561	0.987	35.401	13.904	0.740	3.137
$\hat{\beta}_{12}$	10.986	40.150	0.070	26.154	5.957	21.821	0.222	7.726
$\hat{\beta}_{13}$	-29.108	39.417	0.030	25.278	-13.976	18.946	0.085	5.824
$\hat{\beta}_{23}$	-35.225	43.563	0.026	30.791	-13.192	16.310	0.060	4.316
	$\hat{\sigma}^2 = 61.63191$ MSEOLS = 5401.59 KS = 278.626837				$\hat{\sigma}^2 = 61.63121$ MSEOLS = 2184.258 KS = 63.7166			

3.2. Alışılmış Ridge ve Alışılmış Liu Tahmin Edicileri

Regresyon analizinde en çok karşılaşılan en önemli iki problem kötü koşulluluk ve sapan değer problemleridir. Bu problemler elde edilen tahminlerin doğruluğunu önemli derecede etkiler ve kesinlikle araştırılması gereklidir. Özellikle karma denemelerde en çok kullanılan modeller 2. dereceden olduğundan ve 2. dereceden model için X veri matrisinin sütunları bileşen oranlarının çarpımlarını içerdiğinden ve ayrıca karışımı oluşturan bileşenler üzerinde fiziksel, kimyasal, ekonomik nedenlerden dolayı bulunan ek kısıtlamamalar yüzünden kötü koşulluluk probleminin ortaya çıkması ihtimali çok yüksektir. En küçük kareler en iyi lineer yansız tahmin edicilerdir. Yani yansız tahmin ediciler arasında minimum varyansa sahip katsayı tahminleri üretirler. Ancak bu en küçük karelerin verdiği varyansın küçük olduğu anlamına gelmez. Ridge ve Liu regresyon tahmin edicileri X veri matrisinin kötü koşulluluğa sahip olduğu durumlarda β 'nin yansız tahmini en küçük karelerden daha küçük varyansa sahip daha kararlı yanlı tahmin verirler.

(3.1) modeli için Ridge regresyon tahmin edicisi normal denklemleri

$$(X'X + kI)\hat{\beta}_R = X'y \quad (3.15)$$

ve β 'nin ridge tahmin edicisi $k \geq 0$ yanlılık çarpanı ve $\hat{\beta}_{LS}$, β 'nin en küçük kareler tahmini olmak üzere,

$$\hat{\beta}_R(k) = (X'X + kI)^{-1}X'y = (X'X + kI)^{-1}X'X\hat{\beta}_{LS} \quad (3.16)$$

dir ve varyans kovaryans matrisi,

$$\text{Var}(\hat{\beta}_R) = \sigma^2(X'X + kI)^{-1}X'X(X'X + kI)^{-1} \quad (3.17)$$

dir. Ridge için MSE, λ_i , $X'X$ matrisinin özdeğerleri olmak üzere,

$$\text{MSE}(\hat{\beta}_R) = \sigma^2 \sum_{i=1}^p \frac{\lambda_i}{(\lambda_i + k)^2} + k^2 \beta'_{LS} (X'X + kI)^{-2} \beta_{LS} \quad (3.18)$$

dir. $\hat{\beta}_R$ en küçük kareler tahminine dayalıdır ve eğer $k = 0$ alınırsa en küçük kareler tahmin edicisi elde edilir. Ridge tahmin edicisinde yanlılık parametresi k 'nın seçimi çok önemli bir yer tutar. k 'nın bir uygun seçimi $\hat{\beta}_{LS}$ ve $\hat{\sigma}$ en küçük kareler tahminleri olmak üzere,

$$k_{HK} = \frac{p\hat{\sigma}^2}{\hat{\beta}'_{LS}\hat{\beta}_{LS}} \quad (3.19)$$

dır (Hoerl ve Kennard, 1970).

Liu tahmin edicisi (Liu, 1993), çoklu içi ilişki problemi ile mücadele edebilen, Ridge tahmin edicisine alternatif yanlı tahmin edicidir. (3.1) modeli için alışılmış Liu tahmin edicisi $0 < d < 1$ yanlılık parametresi olmak üzere,

$$\hat{\beta}_L(d) = (X'X + I)^{-1}(X'X + dI)\hat{\beta}_{LS} \quad (3.20)$$

dir ve varyans kovaryans matrisi,

$$\text{Var}(\hat{\beta}_L) = \sigma^2(X'X + I)^{-1}(X'X + dI)(X'X)^{-1}(X'X + dI)(X'X + I)^{-1} \quad (3.21)$$

dir. Liu tahmin edicisinin Ridge tahmin edicisine göre avantajı d yanlılık parametresinin bir lineer fonksiyonu olmasıdır ve böylece d'nin seçimi k'dan daha kolay olacaktır (Liu, 1993).

Örnek 3.2. İlk olarak Örnek 3.1'de ele aldığımız Tablo 3.1'de verilen standartlaştırılmış karma veriyi ele alalım. Örnek 3.1'de standartlaştırılmış veri için en küçük kareler tahmin edicileri elde edilmişti. İlk olarak 2 sapan değeri çıkararak 18 gözlem kullanıldığında elde edilen sonuçları inceleyelim. Elde edilen sonuçlar Tablo 3.3'te verilmiştir.

Tablo 3.3. 2 sapan değeri çıkarılmış 18 gözlemlilik veri için elde edilen sonuçlar.

Terim	$\hat{\beta}_{LS}$ (20 gözlem)	18 gözlem için sonuçlar			
		$\hat{\beta}_{LS}$	$se(\hat{\beta}_{LS})$	Özdeğer	VIF _r -ler
$\hat{\beta}_1$	3.206	5.611	2.698	3.493	5.894
$\hat{\beta}_2$	1.731	4.120	3.316	0.937	8.898
$\hat{\beta}_3$	35.401	1.444	2.599	0.730	5.466
$\hat{\beta}_{12}$	5.957	-1.678	3.169	0.150	8.130
$\hat{\beta}_{13}$	-13.976	-2.173	2.840	0.082	6.530
$\hat{\beta}_{23}$	-13.192	-2.398	2.431	0.055	4.783
	$\hat{\sigma}^2 = 61.63121$ MSEOLS = 2184.258 KS = 63.7166	$\hat{\sigma}^2 = 1.23536$ MSEOLS = 49.04484 KS = 63.5090			

En küçük kareler sonuçlarına dikkat edilirse sapan değerlerin katsayı tahminlerine etkileri açıkça görülmektedir. Özellikle varyans ve MSE oldukça düşük elde edilmesine rağmen koşul sayısı 25'ten yüksek olup araştırılması gerekir. 18 ve 20 gözlemlilik veriler için Ridge ve Liu tahminleri Tablo 3.4'te verilmiştir.

Tablo 3.4. 18 ve 20 gözlemlili veriler için Ridge ve Liu Tahminleri

Terim	20 gözlem için Ridge ve Liu Tahminleri				18 gözlem için Ridge ve Liu Tahminleri			
	$\hat{\beta}_R$	R-VIF _i	$\hat{\beta}_L$	L-VIF _i	$\hat{\beta}_R$	R-VIF _i	$\hat{\beta}_L$	L-VIF _i
$\hat{\beta}_1$	0.730	0.742	1.937	0.939	3.371	1.313	3.604	1.836
$\hat{\beta}_2$	0.499	0.672	0.870	1.229	1.831	1.357	2.491	2.594
$\hat{\beta}_3$	19.876	0.817	17.716	0.670	0.322	1.546	0.788	1.754
$\hat{\beta}_{12}$	4.165	0.699	3.148	1.135	0.860	1.399	-0.271	2.401
$\hat{\beta}_{13}$	-1.870	0.746	-3.112	0.942	-0.414	1.338	-0.917	1.998
$\hat{\beta}_{23}$	-6.345	0.797	-5.152	0.792	-1.170	1.337	-1.268	1.558
	k = 0.221252 MSER = 560.455		d = 0.317 MSEL=857.341		k = 0.1161136 MSER = 15.233		d = 0.499 MSEL = 26.944	

Elde edilen sonuçlara dikkat edilirse uygun k ve d yanlılık parametreleri kullanılarak gerek 20 gerekse 18 gözlemlili veriler için oldukça küçük MSE değerleri elde edilmiştir. Koşul sayısı değeri 100'ün altında olmasına rağmen MSE değerlerine etkisi dikkatten kaçmayacak orandadır. Burada uygun k değeri (3.19)'da verilen k_{HK} ile hesaplanmış ve d değeri ise en küçük MSE değeri elde edilence sona eren döngü kullanılarak elde edilmiştir.

Örnek 3.1 ve Örnek 3.2'de sapan değer ve kötü koşulluluk probleminin karma veriye etkileri incelendi. Dikkat edilirse yaptığımız inceleme sapan değerlerden çok fazla etkilenen en küçük kareler tahmin edicisine dayanmaktadır. Çünkü Ridge ve Liu tahmin edicileri en küçük kareler tahminine dayalı sonuç verirler. Bu bölümde amacımız sapan değer ve kötü koşulluluk problemlerinin karma verilere etkilerini en aza indirebilmek için Ridge ve Liu tahmin edicileri ile birlikte Robust M regresyon tahmin edicilerinin birlikte kullanılmasıdır.

3.3. Robust Ridge ve Robust Liu Tahmin Edicileri

(3.1)'deki modelde X değişkenleri standartlaştırılmış olsun. Böylece $X'X$ matrisi korelasyon matrisi formunda elde edilir. $X'X$ matrisinin özdeğerleri $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p > 0$ ve $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_p)$ olsun. λ_i özdeğerlerine karşılık gelen özvektörler q_i 'ler olmak üzere, $P = (q_1, \dots, q_p)$, $X'X = P\Lambda P'$ olacak şekilde ortogonal

bir matris olsun. Bu durumda (3.1)'deki model kanonik formda, $C = XP$, $\alpha = P'\beta$ olmak üzere,

$$y = C\alpha + \varepsilon \quad (3.22)$$

şeklinde gösterilir. Kanonik formda gösterim bazı kolaylıklar sağladığı gibi elde edilen α 'nın herhangi bir tahmin edicisi $\hat{\alpha}$ ile $\hat{\beta}$ tahminlerinin MSE'leri eşit olacaktır. Böylece sadece kanonik model kullanmak yeterlidir.

α 'nın en küçük kareler tahmin edicisi $\hat{\alpha}_{LS} = (C'C)^{-1}C'y$ 'dir. α 'nın alışılmış Ridge tahmin edicisi $\hat{\alpha}_R(k)$,

$$Z(k) = (\Lambda + kI)^{-1}\Lambda = (C'C + kI)^{-1}C'C \quad (3.23)$$

olmak üzere $\hat{\alpha}_R(k) = Z(k)\hat{\alpha}_{LS}$ ' dir. Bu tahmin edici veride sapan değerlerin olması durumunda ve $X'X$ kötü koşulluluğa sahip iken $\hat{\alpha}_{LS}$ en küçük kareler tahmininden daha tutarlı sonuçlar verir. $\hat{\alpha}_R(k)$ 'nin MSE'si, $k > 0$ olmak üzere,

$$MSE\{\hat{\alpha}_R(k)\} = \sigma^2 \sum_{i=1}^p \frac{\lambda_i}{(\lambda_i + k)^2} + \sum_{i=1}^p \left(\frac{k\alpha_i}{\lambda_i + k} \right)^2 \quad (3.24)$$

şeklinde verilir.

Yukarıda bahsettiğimiz nedenlerden dolayı alışılmış Ridge tahmin edicisi y yönündeki sapan değerlere karşı hassastır. Silvapulle (1991), bu problemle mücadele etmede alternatif bir yöntem olarak robust M tahmin yöntemine dayalı Robust Ridge tahmin edicisini (Ridge tipi M tahmin edici) önermiştir. (3.19) modelindeki α 'nın M tahmin edicisi $\hat{\alpha}_M$ olmak üzere,

$$\hat{\alpha}_{RM}(k) = Z(k)\hat{\alpha}_M \quad (3.25)$$

dir. $\hat{\alpha}_R(k)$, $\hat{\alpha}_{LS}$ 'nin kısıtlanmış versiyonudur ve bu kısıtlama $Z(k)$ ile yapılır. Buna göre $\hat{\alpha}_{LS}$ yerine $\hat{\alpha}_M$ ile yapılan kısıtlama ile kalın kuyruklu dağılımlardan gelen hataların $\hat{\alpha}_R(k)$ üzerindeki etkilerinin azalması beklenir. Ω_{ii} , $\Omega = \text{Cov}(\hat{\alpha}_M)$ matrisinin köşegen elemanları olmak üzere, $\hat{\alpha}_{RM}(k)$ için MSE,

$$\text{MSE}\{\hat{\alpha}_{RM}(k)\} = \sum_{i=1}^p \frac{\lambda_i}{(\lambda_i + k)^2} \Omega_{ii} + \sum_{i=1}^p \left(\frac{k\alpha_i}{(\lambda_i + k)} \right)^2 \quad (3.26)$$

dir. Sivapulle (1991), *C varsayımları* olarak adlandırdığı aşağıdaki varsayımları vermiştir.

- i) M tahmin edicisinde kullanılan ψ fonksiyonunun çarpık-simetrik ve azalmayan bir fonksiyondur.
- ii) Hatalar simetriktir.
- iii) $\text{Cov}(\hat{\alpha}_M)$ sonludur.
- iv) $\hat{\alpha}_M$ iteratif yöntemle elde edilmiştir.

Bu varsayımlar altında $0 < k < k_1$ için $\text{MSE}\{\hat{\alpha}_{RM}(k)\} < \text{MSE}\{\hat{\alpha}_M\}$ olacak şekilde pozitif bir k_1 değerinin var olduğunu ve eğer $\Omega_{ii} < \sigma^2 \lambda_i^{-1}$, her i için ise $\text{MSE}\{\hat{\alpha}_{RM}(k)\} < \text{MSE}\{\hat{\alpha}_R(k)\}$ olduğunu göstermiştir.

k değerinin uygun seçimini Hoerl, Kennard ve Baldwin (1975),

$$\sigma^2 = \frac{1}{n-p} \sum_{i=1}^n (y_i - \hat{\beta}_{0LS} - c_i' \hat{\alpha}_{LS})^2 \quad (3.27)$$

olmak üzere

$$k_R = \frac{p\hat{\sigma}^2}{\|\hat{\alpha}_{LS}\|^2} \quad (3.28)$$

olarak vermişlerdir. Ayrıca Lowless ve Wang (1976) k'nın bir başka tahminini

$$k_R = \frac{p\hat{\sigma}^2}{\sum_{i=1}^p \lambda_i \hat{\alpha}_i^2} \quad (3.29)$$

şeklinde vermişlerdir. Benzer şekilde Silvapulle (1991), s^2 , σ^2 'nin bir robust tahmini ve

$$\hat{A}^2 = \frac{s^2(n-p)^{-1} \sum_{i=1}^n [\psi(r_i/s)]^2}{\left[\frac{1}{n} \sum_{i=1}^n \psi'(r_i/s) \right]^2} \quad (3.30)$$

olmak üzere, M tahminine dayalı k'nın robust seçimini,

$$k_{RM} = \frac{p\hat{A}}{\|\hat{\alpha}_M^2\|} \quad (3.31)$$

olarak vermiştir. Dikkat edilirse $\psi(x) = x$ olarak seçilirse $\hat{\alpha}_M = \hat{\alpha}_{LS}$ ve $\hat{A}^2 = \hat{\sigma}^2$ olur. Böylece ψ 'nin bu seçimi için $k_R = k_{RM}$ ve $\hat{\alpha}_R(k_R) = \hat{\alpha}_{RM}(k_{RM})$ olur. Yani alışılmış Ridge ile robust Ridge birbirine eşittir.

(3.22)'deki kanonik model için α 'nın Liu tahmini,

$$\hat{\alpha}_L(d) = (C'C + I)^{-1}(C'C + dI)\hat{\alpha}_{LS} = (\Lambda + I)^{-1}(\Lambda + dI)\hat{\alpha}_{LS} \quad (3.32)$$

dir. α 'nın herhangi bir $\hat{\alpha}$ tahmini için $MSE(\hat{\alpha}) = MSE(\hat{\beta})$ olacak şekilde $\hat{\beta} = P'\hat{\alpha}$ şeklinde bir tahmine karşılık gelir. Bu yüzden sadece kanonik formu ele almak yeterlidir.

Daha öncede bahsettiğimiz gibi Liu tahmini y yönündeki sapan değerlerden çok etkilenen en küçük kareler tahmin edicilerini kullandığından, sapan değer probleminin olduğu durumlarda tutarlı olmayan sonuçlar üretebilir. $\hat{\alpha}_M$, α 'nın M tahmin edicisi olmak üzere α 'nın Liu tipi M tahmin edicisi $\hat{\alpha}_{LM}(d)$,

$$\hat{\alpha}_{LM}(d) = (\Lambda + I)^{-1}(\Lambda + dI)\hat{\alpha}_M \quad (3.33)$$

şeklinde tanımlanır (Arslan ve Billor, 2000). Bu tahmin edici M tahmin edicisine dayalı olup çoklu iç ilişki problemi ve y yönünde sapan değerlerin olması durumunda etkilidir. Ridge tahmininde olduğu gibi Liu tahmin edicisi ile Liu tipi M tahmin edicinin karşılaştırma kriteri MSE olarak alınacaktır. $\hat{\alpha}_{LM}(d)$ 'nin MSE'si λ_i 'ler $X'X$ matrisinin özdeğerleri, $0 < d < 1$ ve Ω_{ii} $\Omega = \text{Cov}(\hat{\alpha}_M)$ matrisinin köşegen elemanları olmak üzere,

$$\text{MSE}\{\hat{\alpha}_{LM}(d)\} = \sum_{i=1}^p \frac{(\lambda_i + d)^2}{\lambda_i(\lambda_i + 1)^2} \Omega_{ii} + (d-1)^2 \sum_{i=1}^p \frac{\alpha_i^2}{(\lambda_i + 1)^2} \quad (3.34)$$

dir.

Arslan ve Billor (2000), C varsayımları altında, $\text{MSE}\{\hat{\alpha}_{LM}(d)\} < \text{MSE}\{\hat{\alpha}_M(d)\}$ olacak şekilde $0 < d < 1$ varlığını ve her i için $\Omega_{ii} < \sigma^2 \lambda_i^{-1}$ ise her $0 < d < 1$ için $\text{MSE}\{\hat{\alpha}_{LM}(d)\} < \text{MSE}\{\hat{\alpha}_L(d)\}$ olduğunu göstermişlerdir. Buna göre (3.34)'deki $\text{MSE}\{\hat{\alpha}_{LM}(d)\}$,

$$d = \frac{\sum_{i=1}^p \frac{\alpha_i^2 - \lambda_i \Omega_{ii}}{(\lambda_i + 1)^2}}{\sum_{i=1}^p \frac{\alpha_i^2 + \Omega_{ii}}{(\lambda_i + 1)^2}} \quad (3.35)$$

de minimize edilmiştir. Ayrıca $MSE\{\hat{\alpha}_{LM}(d)\} < MSE\{\hat{\alpha}_L(d)\}$ 'nin gösterilmesi için iki MSE farkının sıfırdan küçük olduğunu göstermek yeterlidir ve farkın sıfırdan küçük olabilmesi için $\Omega_{ii} < \sigma^2 \lambda_i^{-1}$ olması gerektiğini görmek kolaydır.

d 'nin bir robust tahminini (2.33) denkleminde α_i^2 ve Ω_{ii} 'lerin yerine onların yansız tahminlerini kullanarak elde edebiliriz. $\hat{\alpha}_M$ 'nin α ortalamalı, $A^2 \Lambda^{-1}$ varyanslı, normal dağılıma sahip olduğu bilinmektedir. Yani, s_0 ölçeğin bir tahmini ve

$$A^2 = \frac{\sigma E[\psi^2(\varepsilon/\sigma)]}{[E(\psi'(\varepsilon/\sigma))]^2} \quad (3.36)$$

olmak üzere $n^{1/2}(\hat{\alpha}_M - \alpha) \rightarrow N(0, A^2 \Lambda^{-1})$ 'dir. Böylece α_i^2 'nin yansız tahmini

$\hat{\alpha}_{M_i}^2 - \frac{\hat{A}^2}{\lambda_i}$ ve Ω_{ii} 'nin yansız tahmini (3.30)'da verilen $\hat{A}^2$ kullanılarak elde edilen

$\hat{A}^2 / \lambda_i$ 'dir (Huber, 1981). Bu yansız tahminleri (3.35) denkleminde yerlerine yazıp gerekli düzenlemeler yapılırsa d 'nin bir robust tahmini

$$\hat{d}_M = 1 - \hat{A}^2 \frac{\sum_{i=1}^p \frac{1}{\lambda_i(\lambda_i + 1)}}{\sum_{i=1}^p \frac{\hat{\alpha}_{M_i}^2}{(\lambda_i + 1)^2}} \quad (3.37)$$

elde edilir. d yanlılık parametresi 0 ile 1 aralığında olmak zorundadır. Bu yüzden d 'nin $h > 0$ olacak şekilde

$$\hat{d}_{Mh} = 1 - h \hat{A}^2 \frac{\sum_{i=1}^p \frac{1}{\lambda_i(\lambda_i + 1)}}{\sum_{i=1}^p \frac{\hat{\alpha}_{M_i}^2}{(\lambda_i + 1)^2}} \quad (3.38)$$

şeklinde genelleştirilmiş formunu kullanabiliriz. Pratikte eğer d , 0 ile 1 arasında ise $\hat{d}_M$ kullanılmasında sakınca yoktur. Ancak diğer durumlarda d 'nin genelleştirilmiş formunun kullanılması önerilmiştir (Arslan ve Billor, 2000).

Örnek 3.3. Tablo 3.1'de verdiğimiz standartlaştırılmış karma veriyi tekrar ele alalım. Örnek 3.2'de sapan değerlerin model katsayıları üzerindeki etkileri incelenmişti. Bu veri için robust M tahmin edicileri Tablo 3.5'te verilmiştir.

Tablo 3.5. 20 gözlemlili veri için robust M regresyon tahmin edicileri

Terim	$\hat{\beta}_{LS}$ 20 gözlem	$\hat{\beta}_{LS}$ 18 gözlem	Huber	Tukey	T	GT
$\hat{\beta}_1$	3.206	5.611	4.688	6.162	6.467	5.422
$\hat{\beta}_2$	1.731	4.120	3.070	6.160	7.850	7.794
$\hat{\beta}_3$	35.401	1.444	9.887	1.697	2.279	2.383
$\hat{\beta}_{12}$	5.957	-1.678	0.878	-3.409	-4.638	-4.257
$\hat{\beta}_{13}$	-13.976	-2.173	-4.784	-2.663	-3.145	-2.738
$\hat{\beta}_{23}$	-13.192	-2.398	-4.867	-2.651	-3.093	-3.243
20 gözlem için $\hat{\sigma}^2 = 61.63121$ MSEOLS = 2184.258			$k = 1.345$ $\hat{\sigma} = 3.506$	$c = 6$ $\hat{\sigma} = 0.7153$	$v = 2$ $\hat{\sigma} = 0.659$	$p = 1.2$ $q = 2$ $\hat{\sigma} = 1.237$

Sapan değere sahip 20 gözlemlili standartlaştırılmış veri için elde edilen robust M tahminleri ile en küçük kareler tahminleri arasında farklılıklar olması beklediğimiz ve sürpriz olmayan bir sonuçtur. Daha öncede belirttiğimiz gibi robust yöntemler y yönünde sapan değerlerin olması durumunda çok etkili sonuçlar veren yöntemlerdir ve dikkat edilirse 18 gözlemlili sapan değer bulunmayan veri için elde edilen en küçük kareler katsayı tahminlerine yakın katsayı tahminleri elde edilmiştir. Tablo 3.6'da (3.37)'de verilen $\hat{A}^2$ kullanılarak elde edilen standart hatalar verilmiştir. Robust tahmin edicilerinin standart hataları $\hat{A}^2(X'X)^{-1}$ matrisinin köşegen elemanlarının karekökü hesaplanarak elde edildi.

Tablo 3.6. Robust tahminlerin katsayıları için standart hatalar

Terim	$se(\hat{\beta}_{LS})$	$se(\text{Huber})$	$se(\text{Tukey})$	$se(T)$	$se(GT)$
$\hat{\beta}_1$	18.876	3.719	1.814	0.887	3.025
$\hat{\beta}_2$	23.098	4.552	2.220	1.085	3.701
$\hat{\beta}_3$	13.904	2.740	1.336	0.653	2.228
$\hat{\beta}_{12}$	21.821	4.300	2.097	1.025	3.496
$\hat{\beta}_{13}$	18.946	3.734	1.821	0.890	3.036
$\hat{\beta}_{23}$	16.310	3.214	1.568	0.766	2.613
$\hat{A}^2$		2.393	0.569	0.136	1.582

Robust yöntemlerle elde edilen standart hataların en küçük karelerin standart hatalarından çok daha iyi olduğu görülmektedir.

Şimdi Ridge ve Liu tahminlerinde kullanılan en küçük kareler katsayı tahmini yerine Tukey ρ fonksiyonu ve $\hat{A}^2$ kullanılarak elde edilen katsayı tahminini kullanalım. Elde edilen sonuçlar Tablo 3.7’de verilmiştir.

Tablo 3.7. 20 gözlemlili veri için M tahminine dayalı robust Ridge ($\hat{\beta}_{RM}$) ve robust Liu $\hat{\beta}_{LM}$ tahminleri

Terim	$\hat{\beta}_{LS}$ 20 gözlem	$se(\hat{\beta}_{LS})$	$\hat{\beta}_{RM}$	$se(\hat{\beta}_{RM})$	$\hat{\beta}_{LM}$	$se(\hat{\beta}_{LM})$
$\hat{\beta}_1$	3.206	18.876	2.747	0.650	3.001	0.731
$\hat{\beta}_2$	1.731	23.098	2.067	0.618	2.719	0.836
$\hat{\beta}_3$	35.401	13.904	0.503	0.682	0.741	0.617
$\hat{\beta}_{12}$	5.957	21.821	0.795	0.631	-0.364	0.803
$\hat{\beta}_{13}$	-13.976	18.946	-0.306	0.651	-0.658	0.732
$\hat{\beta}_{23}$	-13.192	16.310	-0.735	0.673	-0.835	0.671
	20 gözlem için $\hat{\sigma}^2 = 61.63121$ MSEOLS = 2184.258		k = 0.2212 MSER = 287.9927	Tukey $\hat{A}^2$ için	d = 0.317 MSEL = 391.026	Tukey $\hat{A}^2$ için

Elde edilen sonuçları Örnek 3.2’de elde ettiğimiz Tablo 3.4’teki sonuçlarla karşılaştırdığımızda MSE değerlerinde %50 oranında bir iyileşme olduğunu görüyoruz. Ayrıca en küçük kareler tahmini yerine robust M tahmin edicisini kullandığımızda sapan değerlerin katsayı tahminleri üzerindeki etkileri daha iyi

görülmektedir. 18 gözlemlili veri için elde edilen Ridge ve Liu tahminleri ile 20 gözlemlili veri için elde edilen robust Ridge ve robust Liu tahminleri birbirlerine yakın elde edilmiştir. Özellikle $\hat{\beta}_3$ katsayı tahmini 20 gözlemlili veri için Ridge ile 19.876 ve Liu ile 17.716 elde edilirken 18 gözlemlili sapan değersiz veri için Ridge ile 0.322 ve Liu ile 0.788 olarak hesaplanmıştır. 20 gözlemlili veri için robust Ridge ile 0.503 ve robust Liu ile 0.741 olarak elde edilmiş ve sapan değerlerin katsayı tahminleri üzerindeki etkilerinin ciddi oranda azalması sağlanmıştır.

3.4. Şiddetli Çoklu İç İlişki ve Sapan Değer Problemlerinin Birlikte Olması Durumu için Bir Örnek: Motor Yağı Karışımı Verisi

Bu bölümde çok iyi bilinen Snee (1975)'in motor yağı karışımı verisi için robust Ridge ve Robust Liu tahmin edicilerinin etkilerini inceleyeceğiz. Çalışmanın amacı üç farklı motor yağının (x_2, x_3, x_4) karıştırılmasıyla elde edilen yağ karışımına çeşitli kritik fiziksel özellikleri (y) istenilen seviyeye getirecek şekilde eklenecek katkı maddesinin (x_1) miktarının belirlenmesidir. Bileşenler karma denemelerin doğal kısıtlamalarının yanı sıra aşağıdaki ek kısıtlamalara sahiptirler.

$$\begin{aligned} 0,07 &\leq x_1 \leq 0,18 \\ 0,00 &\leq x_2 \leq 0,30 \\ 0,37 &\leq x_3 \leq 0,70 \\ 0,00 &\leq x_4 \leq 0,15 \end{aligned} \tag{3.39}$$

Karma denemelerde bu tip ek kısıtlamalar yüzünden çoğunlukla kötü koşulluluk problemi ortaya çıkar. Ayrıca dikkat edilirse birinci ve dördüncü bileşenlerin değişim aralıkları çok dardır ve bu tip dar aralıklı değişim aralığına sahip veri grupları için kötü koşulluluk problemi daha ciddi boyutlardadır ve bu istenmeyen durum yukarıda da bahsettiğimiz gibi çalışmadaki hassasiyeti ciddi boyutlarda etkiler. Çünkü dar değişim aralığına sahip değişkenlerin 2. dereceden model için VIF_i değerlerinin yüksek olması beklenir. Değişkenleri standartlaştırmanın yararlarından

faydalanabilmek için (3.14)'de verilen ikinci dereceden model için değişkenleri (3.13)'te verilen yöntemle standartlaştıracağız. 18 gözlemden oluşan veri orijinal ve standartlaştırılmış olarak Tablo 3.8'de verilmiştir. Ayrıca çalışmanın amacını daha iyi açıklayabilmek için 5. gözlem değeri y_5 , sapan değer olacak şekilde değiştirilecektir ve bu Tablo 3.8'de y_{bozuk} olarak verilmiştir.

Tablo 3.8. Snee(1975)'in motor yağı karışımı verisi.

Sıra	x_1	x_2	x_3	x_4	w_1	w_2	w_3	w_4	y_{orj}	y_{bozuk}
1	0.1500	0.0000	0.7000	0.1500	0.2512	0.0000	0.2681	0.3314	13.89	13.89
2	0.1800	0.3000	0.3700	0.1500	0.3015	0.3517	0.1417	0.3314	13.99	13.99
3	0.0700	0.2300	0.7000	0.0000	0.1172	0.2696	0.2681	0.0000	7.6	7.6
4	0.0700	0.0800	0.7000	0.1500	0.1172	0.0938	0.2681	0.3314	9.45	9.45
5	0.1800	0.1200	0.7000	0.0000	0.3015	0.1407	0.2681	0.0000	12.93	3.417
6	0.0700	0.3000	0.6300	0.0000	0.1172	0.3517	0.2413	0.0000	7.38	7.38
7	0.0700	0.3000	0.4800	0.1500	0.1172	0.3517	0.1839	0.3314	8.58	8.58
8	0.1800	0.0000	0.6700	0.1500	0.3015	0.0000	0.2566	0.3314	15.65	15.65
9	0.1800	0.3000	0.5200	0.0000	0.3015	0.3517	0.1992	0.0000	11.94	11.94
10	0.1800	0.0000	0.7000	0.1200	0.3015	0.0000	0.2681	0.2651	15.24	15.24
11	0.0700	0.2275	0.6275	0.0750	0.1172	0.2667	0.2403	0.1657	8.24	8.24
12	0.1800	0.1440	0.5920	0.0840	0.3015	0.1688	0.2268	0.1856	13.84	13.84
13	0.1250	0.3000	0.5000	0.0750	0.2093	0.3517	0.1915	0.1657	10.08	10.08
14	0.1300	0.0860	0.7000	0.0840	0.2177	0.1008	0.2681	0.1856	11.48	11.48
15	0.1250	0.2375	0.6375	0.0000	0.2093	0.2784	0.2442	0.0000	9.64	9.64
16	0.1300	0.1360	0.5840	0.1500	0.2177	0.1594	0.2237	0.3314	11.94	11.94
17	0.1330	0.1630	0.6170	0.0870	0.2227	0.1911	0.2363	0.1922	11.25	11.25
18	0.1800	0.1500	0.5200	0.1500	0.3015	0.1759	0.1992	0.3314	14.65	14.65

Tablo 3.2'de standartlaştırılmış ve standartlaştırılmamış değişkenler cinsinden en küçük kareler katsayıları, standart hatalar X matrisinin özdeğerleri, koşul sayıları, VIF_i 'ler ve katsayı tahminleri verilmiştir. Bu veri için elde edilen sonuçlara dikkat edilirse Örnek 3.1'de ele aldığımız 20 gözlemlilik karma veriye göre koşul sayısı çok büyüktür ve kötü koşulluluk problemi çok ciddi boyutlardadır.

Tablo 3.9. Motor yağı karışımı verisi sonuçları.

Terim	Standartlaştırılmamış				Standartlaştırılmış			
	β_{EKK}	$SE(\beta_{EKK})$	Özdeğer	VIF_i	β_{EKK}	$SE(\beta_{EKK})$	Özdeğer	VIF_i
$\hat{\beta}_1$	125,8915	5.07	8,060	20198,698	74,4183	3.056	7,5447	7039,660
$\hat{\beta}_2$	6,6914	0.665	0,345	342,195	5,7344	0.574	1,3905	248,663
$\hat{\beta}_3$	7,0113	0.192	0,101	28,591	18,2851	0.508	0,1714	194,669
$\hat{\beta}_4$	16,24	3.133	0,051	7665,789	7,3513	1.445	0,231	1573,6241
$\hat{\beta}_{12}$	-114,9972	6.565	0,00078	33900,847	-12,5959	0.737	0,048	409,402
$\hat{\beta}_{13}$	-99,0116	6.714	0,00072	35423,061	-35,3256	2.459	0,033	4555,863
$\hat{\beta}_{14}$	-56,4377	7.245	0,00045	41234,244	-3,6242	0.485	0,005	177,230
$\hat{\beta}_{23}$	-5,7463	1.211	0,00030	1134,920	-2,7883	0.586	0,001	259,098
$\hat{\beta}_{24}$	-8,6768	3.649	0,000028	10385,443	-0,6888	0.295	0,0004	65,740
$\hat{\beta}_{34}$	-6,7357	3.765	0,0000085	11065,164	-1,8411	1.040	0,00008	815,987
	$\hat{\sigma}^2 = 0.001292234$ $KS = 948235.29$ $MSEOLS = 208.5394$				$\hat{\sigma}^2 = 0.001326762$ $KS = 94308.75$ $MSEOLS = 20.35245$			

X katsayı matrisinin koşul sayısı ve VIF_i değerleri kabul edilebilir değerlerin çok üzerinde olup kötü koşulluluk problemi çok ciddi boyutlardadır. Ayrıca dikkat edilirse değişim aralığı dar olan değişkenlerin VIF değerleri diğerlerinden daha büyük elde edilmektedir. Standartlaştırılmış değişkenler cinsinden elde edilen sonuçların yorumlanabilirliğinin yüksek olması yanında MSE değerinde de dikkate değer oranda küçülmesi ve birçok istatistik paket programının standartlaştırılmış değişkenler cinsinde sonuç üretmesinden dolayı bundan sonra işlemlerimizi standartlaştırılmış değişkenler cinsinden yapacağız.

Şimdi de veride kötü koşulluluğun yanı sıra sapan değer olması durumu ile olmaması durumunu karşılaştırmak amacıyla verideki 12.93 olan 5. gözlem değerini DE6 programının simülasyon editörünü kullanarak 3.417 olarak değiştirelim. Standartlaştırılmış değişkenler kullanılarak elde edilen sonuçlar Tablo 3.3'te verilmiştir.

Tablo 3.10. Bozulmuş gözlemler için elde edilen sonuçlar.

Terim	$\hat{\beta}_{LS}$	$se(\hat{\beta}_{LS})$	Özdeğer	VIF_i
$\hat{\beta}_1$	13.355	131.128	7,5447	7039,660
$\hat{\beta}_2$	19.451	24.645	1,3905	248,663
$\hat{\beta}_3$	0.578	21.806	0,1714	194,669
$\hat{\beta}_4$	-94.826	61.997	0,231	1573,6241
$\hat{\beta}_{12}$	3.657	31.622	0,048	409,402
$\hat{\beta}_{13}$	-6.222	105.488	0,033	4555,863
$\hat{\beta}_{14}$	35.543	20.806	0,005	177,230
$\hat{\beta}_{23}$	0.017	25.157	0,001	259,098
$\hat{\beta}_{24}$	4.267	12.672	0,0004	65,740
$\hat{\beta}_{34}$	82.855	44.644	0,00008	815,987
$\hat{\sigma}^2 = 2.442527$ $KS = 94308.75$ $MSEOLS = 37468.21$				

Tablo 3.9’da orijinal veri ve Tablo 3.10’daki 1 sapan değerli veri için elde edilen sonuçları karşılaştırdığımızda, değişkenleri standartlaştırma işleminin bir miktar iyileştirme sağlamasına rağmen yeterli olmadığını ve veriler için koşul sayısı, VIF_i ve MSE değerlerinin standartlaştırılmamış değişkenlere göre daha küçük olmasına rağmen kabul edilebilecek değerlere göre çok büyük olduğunu ve ciddi bir kötü koşulluluk probleminin olduğunu görüyoruz. Ek olarak veride sapan değer olması durumunda ise MSE değerinin 20,35245’den 37468.21’e yükseldiğini görüyoruz. Ayrıca sapan değer olan veri için β ’ların standart hatalarının da ciddi şekilde arttığını görüyoruz. Yani genel olarak veride kötü koşulluluk problemine ek olarak birde sapan değer problemi varsa en küçük kareler ile elde edilen sonuçların çok tutarlı olmayacağını ve standart hatalar ile MSE değerlerinin daha büyük elde edileceğini söyleyebiliriz.

Şimdi de sapan değer olacak şekilde değiştirilen 5. gözlem değerini ve X katsayı matrisinde karşılık gelen satırı veriden çıkarırsak elde edeceğimiz sonuçları inceleyelim. 17 gözlemden oluşan veriler için elde edilen en küçük kareler analiz sonuçları Tablo 3.11’de verilmiştir. Elde edilen sonuçlara dikkat edilirse MSE değeri ve β katsayı tahminlerinin standart hatalarındaki fark çok ciddi boyutlarda.

Yani 1 tek sapan değerin bile sonuçlara etkisi çok fazla. Ancak yüksek koşul sayısı ve VIF_i değerleri yani kötü koşulluluk problemi, y yönündeki sapan değer ve karşılık gelen X katsayı matrisinin satırı çıkarılmasına rağmen yine ciddi boyutlardadır.

Tablo 3.11. 1 sapan değeri atılmış 17 gözlemlik veriler için en küçük kareler.

Terim	$\hat{\beta}_{LS}$ 18 gözlem	$\hat{\beta}_{LS}$ 17 gözlem	$se(\hat{\beta}_{LS})$	Özdeğer	VIF_i -ler
$\hat{\beta}_1$	13.3553	74.394	3.311	7.362	7230.436
$\hat{\beta}_2$	19.4510	5.740	0.626	1.363	258.290
$\hat{\beta}_3$	0.5775	18.278	0.565	0.616	210.712
$\hat{\beta}_4$	-94.8265	7.311	1.787	0.225	2107.794
$\hat{\beta}_{12}$	3.6771	-12.589	0.800	0.034	422.918
$\hat{\beta}_{13}$	-6.2221	-35.314	2.640	0.0198	4599.2
$\hat{\beta}_{14}$	35.5429	-3.609	0.623	0.0051	255.719
$\hat{\beta}_{23}$	0.0171	-2.787	0.627	0.0019	259.501
$\hat{\beta}_{24}$	4.2673	-0.687	0.319	0.0003	66.997
$\hat{\beta}_{34}$	82.8547	-1.808	1.339	0.00008	1183.009
$\hat{\sigma}^2 = 0.00151587$ $MSEOLS = 25.15523$ $KS = 92025$					

2. bölümde ele aldığımız Robust M tahmin edicileri kullanılarak elde edilen sonuçlar Tablo 3.12’te verilmiştir.

Tablo 3.12. 18 gözlemlik bozulmuş veriler için M tahmin edicileri.

Terim	$\hat{\beta}_{LS}$ (17 gözlem)	Huber $k = 1.345$	Tukey $c = 5$	T $v = 1$	GT $p = 1.2; q = 2$
$\hat{\beta}_1$	74.394	70.368	70.1593	67.315	72.109
$\hat{\beta}_2$	5.740	6.980	6.5454	6.686	6.1733
$\hat{\beta}_3$	18.278	17.493	17.7072	17.770	17.7682
$\hat{\beta}_4$	7.311	2.074	4.3243	6.236	5.3194
$\hat{\beta}_{12}$	-12.589	-11.710	-11.6548	-10.970	-12.0783
$\hat{\beta}_{13}$	-35.314	-32.709	-32.1525	-29.938	-33.6443
$\hat{\beta}_{14}$	-3.609	-1.756	-2.3535	-2.294	-2.7782
$\hat{\beta}_{23}$	-2.787	-3.328	-3.2641	-3.624	-2.9162
$\hat{\beta}_{24}$	-0.687	-0.225	-0.3739	-0.732	-0.5048

$\hat{\beta}_{34}$	-1.808	2.311	0.4685	-1.029	-0.2344
$\hat{\sigma}^2$	2.442572	0.4305	0.0129	0.0000378	0.03315

Robust tahmin edicileri kullanarak elde edilen sonuçlara dikkat edilirse katsayı tahminleri sapan değerden çok az etkilenmiştir ve varyans tahmininde dikkate değer bir oranda azalmıştır. Huber M tahmin edicisinin Tukey M tahmininden farklı olmasının nedeni Şekil 1.9’da verilen ρ fonksiyonları ile açıklanabilir. Huber ρ fonksiyonunun kolları yukarı doğru çıkarken yani sürekli artarken Tukey ρ fonksiyonunun kolları bir süre sonra x eksenine paralel olacak şekilde bir yapıya sahiptir. Bunun anlamı Tukey ρ fonksiyonu kullanıldığında çözüme daha hızlı yaklaşılr. Ancak daha öncede belirttiğimiz gibi robust M tahmin ediciler y yönünde sapan değerlerin olması durumunda etkili olan yöntemlerdir. Kötü koşulluluk probleminin olduğu durumlarda etkili yöntemler değildir. Tablo 3.13’te robust tahmin edicileri için (3.37)’de $\hat{A}^2$ kullanılarak elde edilen standart hatalar verilmiştir. Robust tahmin edicilerin standart hataları $\hat{A}^2 (X'X)^{-1}$ matrisinin köşegen elemanlarının karekökü alınarak elde edilir.

Tablo 3.13. Robust tahmin edicileri için standart hatalar

Terim	se($\hat{\beta}_{LS}$)	se(Huber)	se(Tukey)	se(T)	se(GT)
$\hat{\beta}_1$	131.128	70.74743	9.826825	0.6381237	22.38793
$\hat{\beta}_2$	24.644	13.29661	1.846901	0.119932	4.207695
$\hat{\beta}_3$	21.806	11.76478	1.634129	0.1061153	3.722948
$\hat{\beta}_4$	61.997	33.44917	4.646092	0.3017029	10.58494
$\hat{\beta}_{12}$	31.622	17.06122	2.369806	0.1538879	5.399002
$\hat{\beta}_{13}$	105.488	56.91414	7.905379	0.513351	18.0104
$\hat{\beta}_{14}$	20.806	11.22545	1.559216	0.1012507	3.552277
$\hat{\beta}_{23}$	25.157	13.57274	1.885254	0.1224226	4.295074
$\hat{\beta}_{24}$	12.672	6.836755	0.949626	0.06166579	2.163482
$\hat{\beta}_{34}$	44.644	24.08666	3.345639	0.2172555	7.622191
$\hat{A}^2$		0.711	0.0137175	0.5568	0.0711994

Standart hatalardaki en küçük karelere göre elde edilen önemli iyileşme açıkça görülmektedir.

Robust Ridge tahmin edicisinin etkilerini daha iyi açıklayabilmek için Örnek 3.1’de olduğu gibi işlemleri orijinal veri ve 1 gözlem değeri bozulmuş veri için yapacağız. Daha önce belirttiğimiz nedenlerden dolayı tahminlerimizi standartlaştırılmış değişkenler için yapacağız. Tablo 3.14’te orijinal veriler için en küçük kareler ve Ridge tahmin edicisi kullanılarak elde edilen sonuçlar verilmiştir.

Tablo 3.14. Orijinal veriler için EKK ve Ridge regresyon sonuçları.

Terim	$\hat{\beta}_{LS}$	$se(\hat{\beta}_{LS})$	VIF _i -ler	$\hat{\beta}_R$	$se(\hat{\beta}_R)$	R-VIF _i
$\hat{\beta}_1$	74.4183	2.490	7039.660	21.6098	0.450	152.9509
$\hat{\beta}_2$	5.7344	0.470	248.663	5.0515	0.463	161.8972
$\hat{\beta}_3$	18.2851	0.410	194.669	13.6204	0.391	115.1841
$\hat{\beta}_4$	7.3513	1.180	1573.6241	7.9004	0.648	316.7561
$\hat{\beta}_{12}$	-12.5959	0.601	409.4202	-0.2351	0.218	35.7484
$\hat{\beta}_{13}$	-35.3256	2.007	4555.863	7.1298	0.387	112.9021
$\hat{\beta}_{14}$	-3.6242	0.395	177.230	3.9855	0.217	35.4570
$\hat{\beta}_{23}$	-2.7883	0.478	259.098	-2.8202	0.458	158.4023
$\hat{\beta}_{24}$	-0.6888	0.241	65.740	-0.9530	0.166	20.7418
$\hat{\beta}_{34}$	-1.8411	0.849	815.987	-2.8858	0.478	172.0685
$\hat{\sigma}^2 = 0.0013267662$ MSEOLS = 20.35245				k = 0.0005 MSER = 5.065067		

Elde edilen sonuçlara dikkat edilirse Ridge tahmin edicisi kullanıldığında en küçük kareler tahminine nazaran MSE değeri ile VIF_i değerlerinde dikkate değer bir düzelme olmuştur. Ridge tahmin edicisini bozuk veriler ile 1 sapan değeri çıkarılmış 17 gözlemlik veriler için uyguladığımızda elde edeceğimiz sonuçlar Tablo 3.15’te verilmiştir.

Tablo 3.15. 17 ve 18 gözlemlili veriler için Ridge regresyon sonuçları.

Terim	17 gözlem için sonuçlar			18 gözlem için sonuçlar		
	$\hat{\beta}_R$	$se(\hat{\beta}_R)$	R-VIF _i	$\hat{\beta}_R$	$se(\hat{\beta}_R)$	R-VIF _i
$\hat{\beta}_1$	20.749	0.481	152.987	4.795	3.669	5.510
$\hat{\beta}_2$	5.466	0.498	163.747	9.588	8.294	28.163
$\hat{\beta}_3$	12.979	0.429	121.294	1.178	8.341	28.482
$\hat{\beta}_4$	5.826	0.665	291.754	-0.644	5.281	11.417
$\hat{\beta}_{12}$	0.059	0.234	36.289	6.452	5.860	14.059
$\hat{\beta}_{13}$	7.014	0.416	114.490	2.962	4.953	10.046
$\hat{\beta}_{14}$	4.996	0.254	42.523	18.932	5.899	14.250
$\hat{\beta}_{23}$	-2.510	0.494	160.855	6.647	7.225	21.374
$\hat{\beta}_{24}$	-1.168	0.189	23.500	-11.699	4.331	7.681
$\hat{\beta}_{34}$	-0.843	0.516	175.750	15.398	5.741	13.493
k = 0.0005 MSER = 5.789 MSEOLS = 25.1552				k = 0.005 MSER = 7348.902		

Tablo 3.15'teki sonuçları incelediğimizde 18 gözlem için β_R katsayı tahminlerinin orijinal verilerdekine nazaran çok farklı olduğunu görüyoruz. 1 sapan değer katsayı tahminlerini oldukça etkilemiştir. Ayrıca tahmin edilen katsayıların standart hatalarını da olumsuz yönde etkilemiştir. VIF_i değerleri istediğimiz düzeyde elde edilse de Ridge için MSE değeri hala çok büyüktür. 17 gözlem için elde edilen sonuçlara dikkat edilirse Tablo 3.14'teki 18 gözlemlili verilerden elde edilen sonuçlara oldukça yakındır. Şimdi, Tablo 3.12'deki Tukey M tahmini sonuçları kullanarak, robust Ridge tahmin edicisini sapan değer ve kötü koşulluluk olan veriye uygulayalım. Elde edilen sonuçlar Tablo 3.16'da verilmiştir. Robust Ridge için standart hatalar (3.22)'de verilen $\hat{A}^2$ yardımıyla hesaplanmıştır.

Tablo 3.16. 18 gözlemlili veriler için robust Ridge regresyon sonuçları.

Terim	$\hat{\beta}_{RM}$ k = 0.005	se($\hat{\beta}_{RM}$)	$\hat{\beta}_{RM}$ k = 0.0005	se($\hat{\beta}_{RM}$)	R-VIF _i k = 0.005	R-VIF _i k = 0.0005
$\hat{\beta}_1$	13.389	0.031	20.791	0.164	5.510	152.9509
$\hat{\beta}_2$	2.731	0.071	5.635	0.169	28.163	161.8972
$\hat{\beta}_3$	10.680	0.071	13.270	0.143	28.482	115.1841
$\hat{\beta}_4$	3.283	0.045	6.342	0.237	11.417	316.7561
$\hat{\beta}_{12}$	2.181	0.045	-0.072	0.079	14.059	35.7484
$\hat{\beta}_{13}$	13.775	0.042	7.600	0.141	10.046	112.9021
$\hat{\beta}_{14}$	6.063	0.050	4.490	0.079	14.250	35.4570
$\hat{\beta}_{23}$	0.291	0.061	-3.031	0.167	21.374	158.4023
$\hat{\beta}_{24}$	-0.467	0.037	-0.887	0.061	7.681	20.7418
$\hat{\beta}_{34}$	1.566	0.049	-1.546	0.174	13.493	172.0685
k = 0.005 için MSER = 403.3552 Tukey için $\hat{A}^2 = 0.0137175$ k = 0.0005 için MSER = 3132.521 MSEOLS = 37468.21						

Robust Ridge yardımıyla elde edilen sonuçlara bakılırsa 17 gözlemlili sapan değer çıkarılmış veriler için elde edilen sonuçlara yakındır. Yani robust M tahmin edicileri gibi robust Ridge tahmin edicide sapan değerden etkilenmemiştir. Ayrıca Ridge tahmin edicisinin kötü koşulluluğa karşı kullanılabilecek bir yapıda olmasından dolayı her iki olumsuz durumun etkisi minimuma inmiştir. MSE ve VIF_i değerleri de karşılaştırıldığında en küçük karelere göre kabul edilebilir bir değere sahip oldukları görülür.

Şimdi de robust Liu tahmin edicilerinin etkilerini inceleyelim. Tablo 3.17’de orijinal veriler için en küçük kareler ve Liu tahmin edicisi kullanılarak elde edilen sonuçlar verilmiştir.

Tablo 3.17. Orijinal veriler için EKK ve Liu regresyon sonuçları.

Terim	$\hat{\beta}_{LS}$	$se(\hat{\beta}_{LS})$	VIF _i -ler	$\hat{\beta}_L$	$se(\hat{\beta}_L)$	L-VIF _i
$\hat{\beta}_1$	74.4183	2.490	7039.660	74.2223	1.113	6998.558
$\hat{\beta}_2$	5.7344	0.470	248.663	5.7269	0.209	247.2162
$\hat{\beta}_3$	18.2851	0.410	194.669	18.2514	0.185	193.5375
$\hat{\beta}_4$	7.3513	1.180	1573.6241	7.3431	0.526	1564.44
$\hat{\beta}_{12}$	-12.5959	0.601	409.4202	-12.5475	0.268	407.0165
$\hat{\beta}_{13}$	-35.3256	2.007	4555.863	-35.1992	0.895	4529.265
$\hat{\beta}_{14}$	-3.6242	0.395	177.230	-3.5978	0.176	176.1998
$\hat{\beta}_{23}$	-2.7883	0.478	259.098	-2.769	0.213	257.5902
$\hat{\beta}_{24}$	-0.6888	0.241	65.740	-0.6845	0.107	65.3602
$\hat{\beta}_{34}$	-1.8411	0.849	815.987	-1.8209	0.379	811.2771
$\hat{\sigma}^2 = 0.0013267662$ MSEOLS = 20.35245				d = 0.997076 MSEL = 20.29314		

Elde edilen sonuçlara dikkat edilirse Liu tahmin edicisi kullanıldığında en küçük kareler tahminine nazaran MSE değeri ile VIF_i değerlerinde bir azalma olduğu gibi β katsayı tahminlerinin standart hataları da düşmüştür. Liu tahmin edicisini bozuk veriler ile 1 sapan değeri çıkarılmış 17 gözlemlik veriler için uyguladığımızda elde edeceğimiz sonuçlar Tablo 3.18’de verilmiştir.

Tablo 3.18. 17 ve 18 gözlemli veriler için Liu regresyon sonuçları.

Terim	17 gözlem için sonuçlar			18 gözlem için sonuçlar		
	$\hat{\beta}_L$	$se(\hat{\beta}_L)$	L-VIF _i	$\hat{\beta}_L$	$se(\hat{\beta}_L)$	L-VIF _i
$\hat{\beta}_1$	74.151	1.128	7178.433	74.2223	1.113	6998.558
$\hat{\beta}_2$	5.731	0.213	256.438	5.7269	0.209	247.2162
$\hat{\beta}_3$	18.235	0.192	209.203	18.2514	0.185	193.5375
$\hat{\beta}_4$	7.302	0.609	2092.64	7.3431	0.526	1564.44
$\hat{\beta}_{12}$	-12.530	0.273	419.881	-12.5475	0.268	407.0165
$\hat{\beta}_{13}$	-35.161	0.899	4566.124	-35.1992	0.895	4529.265
$\hat{\beta}_{14}$	-3.575	0.212	253.886	-3.5978	0.176	176.1998
$\hat{\beta}_{23}$	-2.764	0.214	257.640	-2.769	0.213	257.5902
$\hat{\beta}_{24}$	-0.681	0.108	66.520	-0.6845	0.107	65.3602
$\hat{\beta}_{34}$	-1.782	0.456	1174.505	-1.8209	0.379	811.2771
d = 0.996397 MSEL = 25.06497 MSEOLS = 25.1552				d = 0.997076 MSEL = 20.29314		

Şimdi Tablo 3.12'deki Tukey M tahmini sonuçları kullanarak robust Ridge tahmin edicisini sapan değer ve kötü koşulluluk olan veriye uygulayalım. Elde edilen sonuçlar Tablo 3.19'da verilmiştir. Robust Liu için standart hatalar (3.22)'da verilen $\hat{A}^2$ yardımıyla hesaplanmıştır.

Tablo 3.19. 18 gözlemlili veriler için robust Liu regresyon sonuçları.

Terim	$\hat{\beta}_{LM}$	$se(\hat{\beta}_{LM})$	L-VIF _i
$\hat{\beta}_1$	9.703	0.042	10.010
$\hat{\beta}_2$	3.327	0.009	0.469
$\hat{\beta}_3$	7.173	0.008	0.396
$\hat{\beta}_4$	4.546	0.020	2.325
$\hat{\beta}_{12}$	3.367	0.011	0.725
$\hat{\beta}_{13}$	6.366	0.034	6.532
$\hat{\beta}_{14}$	5.146	0.008	0.385
$\hat{\beta}_{23}$	3.337	0.009	0.499
$\hat{\beta}_{24}$	0.735	0.007	0.286
$\hat{\beta}_{34}$	4.902	0.015	1.275
d=0.0375 MSEOLS = 37468.21 MSEL = 5662.013			

Robust Liu ile elde edilen sonuçlara dikkat edilirse alışılmış Liu tahmininden oldukça farklıdır. VIF_i ve katsayı standart sapmalarında önemli derecede düşüş kaydedilmiştir.

Sonuç olarak, sapan değer ve çoklu iç ilişki problemlerinin karma veri için elde edilen katsayı tahminleri üzerindeki kötü etkilerine karşı mücadele amacıyla önerilen robust M tahmin edicileri ile Ridge ve Liu yanlı tahmin edicilerinin birlikte kullanılması birçok açıdan avantaj sağlamıştır. Özellikle ikinci örnekte, Snee (1975)'in verdiği motor yağı karışımı verisi, MSE açısından çok önemli gelişme sağlanmıştır.

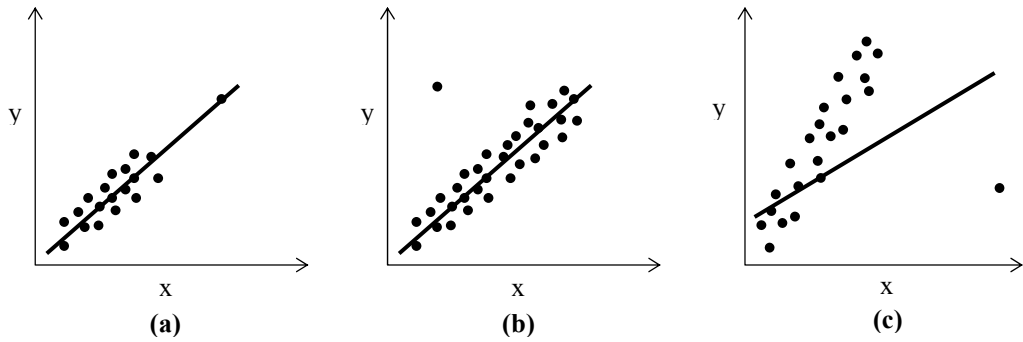
4. BÖLÜM. GENELLEŞTİRİLMİŞ M (GM) TAHMİN EDİCİLERİ

Regresyon analizinde kötü koşulluluk ve sapan değer gibi sıkça karşılaşılan diğer bir problemde x yönünde sapan değer (leverage point) problemidir. Bu çalışmada şimdiye kadar karma veride y uzayında aykırı olan gözlemlerin (sapan değerlerin) varlığının katsayı tahminleri üzerindeki etkileri ve M tahmin edicilerinin bu sapan değerlere karşı performansları incelenmiş buna karşılık katsayı matrisi ile ilgili bir inceleme yapılmamıştır. Birinci bölümde ele alınan, karma denemelerin analizinde sıkça kullanılan standart tasarımlar, kötü koşulluluk problemini düşük seviyede tutan katsayı matrisleri üretirler ancak x yönünde sapan değer problemi (leverage point) araştırmacının karşısında duran diğer bir sorundur. Ayrıca çoğu zaman 1.1 ve 1.2’de verilen doğal kısıtlamalara 1.3’te verilen ek kısıtlamalar eklenmesi tasarım noktalarından uzaklaşılmasına neden olur ve bundan dolayı katsayı matrislerinde ciddi problemler oluşur. Özellikle kötü koşulluluk problemi hiç istenmeyen yüksek seviyelerdedir. Bu bölümde, katsayı matrislerinde kötü koşulluluk probleminin yanı sıra Cook uzaklığı, Mahalanobis uzaklığı gibi tanım yöntemlerinden faydalananak belirleyebileceğimiz x yönünde sapan değerlerin olması durumunda, X uzayındaki sapan değerlere karşı etkili olan, Genelleştirilmiş M (GM) tahmin edicilerinin performansları incelenecektir.

4.1. x Yönünde Sapan Değerleri Tanımlama Yöntemleri

Sapan değerler verinin çoğunluğunun oluşturduğu yapıya uygun olmayan gözlemlerdir. Çok bileşenli sistemlerde bir noktanın x yönünde sapan değer olduğunu belirlemek çok zordur. Özellikle karma denemelerde, en çok kullanılan ikinci dereceden modeller değişkenlerin ikili çarpımlarını içerdiğinden ve değişkenler üzerindeki kısıtlamalar nedeniyle, gözlemin x yönünde sapan değer olduğunu belirlemek daha da güçleşir. Şekil 4.1a’da gösterilen durumu göz önüne alalım. Şekilde bir noktanın X uzayında örneklemin diğer noktalarından daha uzakta ancak neredeyse diğer noktaların içinden geçen regresyon doğrusunun üstünde olduğu görülmektedir. Bu şekildeki gözlemlere **iyi leverage nokta** denir ve iyi x

değerlerlerine sahip değildir. Bu tip gözlemler regresyon katsayı tahminlerini etkilemezler fakat R^2 , katsayıların standart hatası gibi model istatistikleri üzerinde çok önemli etkileri vardır. Şimdi de Şekil 4.1b’de gösterilen durumu ele alalım. Bir nokta X ve Y uzayının içinde ancak örneklemdeki diğer noktalardan farklı davranmaktadır. Bu tip gözlemlere y yönünde sapan değer veya dikey sapan değer denir. Bu gözlemler büyük hata değerlerine sahiptirler. Şekil 4.1c’de ise modele etkisi çok olan büyük hata değeri üreten bir gözlem söz konusudur. Bu gözlem X uzayından uzakta fakat y değeri normal görünmektedir. Bu şekildeki gözlemlere **kötü leverage nokta** denir. Bu gözlemlerin model katsayıları üzerinde önemli etkileri vardır ve regresyon doğrusunun yönünü kendilerine doğru çekerek manivela etkisi yaparlar.



Şekil 4.1. a) iyi leverage nokta b) dikey sapan değer c) kötü leverage nokta

Alışılmış çoklu lineer regresyonda x yönünde sapan değer olan noktaları belirlemek için en çok kullanılan yöntem H şapka matrisinin köşegen elemanları h_{ii} 'leri incelemektir. Ek olarak Mahalanobis uzaklığı, Cook uzaklığı, DFFITS ölçüleri kullanılabilir. Ancak bu istatistikler veride bir tek aykırı değer olması durumunda etkilidirler. Birden çok aykırı değer olması durumunda etkilerin maskelenmesi problemi ortaya çıkar yani x yönünde sapan değerler büyük h_{ii} veya büyük mahalanobis uzaklığına sahip olamayabilirler. Bu durumda bu tip gözlemlere karşı daha dayanıklı olan robust tahmin edicileri kullanmak daha güvenilir sonuçlar üretecektir.

X yönünde sapan değerlerin belirlenmesinde simetrik ve idempotent ($H' = H$, $HH' = H$), izi, rankına eşit olan

$$H = X(X'X)^{-1}X' \quad (4.1)$$

şapka matrisi önemli bir yer tutar. $\hat{y} = Hy$ olduğundan, H' 'ye şapka matrisi denmiştir. H şapka matrisinin h_{ij} elemanları i -inci gözlem y_i tarafından kullanılan x yönünde sapan değer olan noktaların $\hat{y}_i$ uydurulmuş değeri üzerindeki etkilerinin yorumlanmasını sağlar. Ancak daha çok H şapka matrisinin h_{ii} köşegen elemanları ile ilgilenilir. h_{ii} , i -inci gözlemin X uzayının merkezine olan uzaklığının standartlaştırılmış ölçüsüdür. Böylece X uzayından uzakta olan etkili gözlemler büyük h_{ii} değerlerine sahip olacaktır. h_{ii} değerleri 0 ile 1 aralığında olup 1'e yakın olan h_{ii} değerine karşılık gelen i -inci gözlem x yönünde sapan değer olarak değerlendirilir. Yani H matrisinin köşegen elemanları ile açıklayıcı değişkenlerin uzayındaki noktalar arasında önemli bir ilişki vardır.

Genel olarak eğer h_{ii} değeri ortalaması olan $2p/n$ değerinde büyük ise i -inci gözlem x uzayının uzağındadır ve tahminler üzerinde önemli etkileri olabilir. h_{ii} köşegen elemanı sıfır ise bu uydurulan model üzerinde etki yok anlamına gelir. h_{ii} büyük ise yani 1'e yakın ise i -inci rezidünün varyansı hemen hemen sıfırdır. Bunun ana sebebi i -inci gözlemin en küçük karelerin regresyon katsayıları üzerinde önemli etkilerinin olduğudur. Ancak bu durum tek başına o noktanın iyi veya kötü olması anlamına gelmez.

H şapka matrisi sadece X 'e bağlı olup, y yönündeki sapan değerleri ihmal eden, çok dikkatli kullanılması gereken bir istatistiktir. Çünkü Şekil 4.1'de verilen modele etkisi olan noktalar modele etkisi olmayan noktaların etkisi ile görülmeyebilir. Başka bir deyişle, şapka matrisi veri seti sadece tek sapan değere sahip iken kullanışlıdır. Ancak birkaç sapan değer varsa karşılık gelen h_{ii} bununla ilgili bilgi içermeyebilir. Böylece X içinde h_{ii} 'ler arasında görünmeyen çoklu sapan değerler bulunabilir ve h_{ii} yüksek olsa bile rezidüler oldukça küçük elde edilebilir.

Sonuç olarak sadece h_{ii} 'lere bakarak bir gözlemin X uzayında sapan değer olduğunu söyleyemeyiz.

Regresyon analizinde bir gözlemin sapan değer olduğunu iddia edebilmek için kullanılan en faydalı yöntem bu gözlem varken ve yokken çalışmaktır. Bu amaç için en çok başvurulan yöntemlerden biri Cook (1977, 1979)'un verdiği kendi adıyla anılan Cook uzaklığıdır. Cook uzaklığı, i -inci gözlem varken ve yokken regresyon katsayıları arasındaki uzaklığı yani bir gözlemin çıkarılması ile regresyon katsayılarında oluşan değişimi ölçer.

(1.9)'da verilen modeli ele alalım. En küçük kareler için standartlaştırılmış rezidüleri $e_i = y_i - \hat{y}_i$ olmak üzere,

$$r_i = \frac{e_i}{\hat{\sigma}} \quad (4.2)$$

ile verilir. Burada,

$$\hat{\sigma}^2 = \frac{1}{n-p} \sum_{i=1}^n e_i^2 \quad (4.3)$$

dir ve hatalar bağımsız, sıfır ortalamalı, σ standart sapmalı, özdeş dağılıma sahip olduğunda $\hat{\sigma}^2$, σ^2 'nin yansız tahmin edicisidir. Bazen standartlaştırılmış rezidüleri yerine

$$t_i = \frac{e_i}{\hat{\sigma}\sqrt{1-h_{ii}}} \quad (4.4)$$

studentized rezidüleri kullanılır. Bu iki çeşit rezidüye genel olarak standartlaştırılmış rezidüleri denir.

(1.9)'da verilen modelin en küçük kareler çözümü (1.11)'de verilen $\hat{\beta}$ olmak üzere $\hat{y} = X\hat{\beta}$ kullanarak Cook uzaklığı, $\hat{y}(i)$, i-inci gözlem çıkarıldıktan sonra elde edilen uydurulmuş değer ve p tahmin edilecek parametre sayısı olmak üzere

$$CD_i^2 = \frac{(\hat{y} - \hat{y}(i))'(\hat{y} - \hat{y}(i))}{p\hat{\sigma}^2} \quad (4.5)$$

şeklinde verilir. Cook uzaklığı için bir diğer form ise

$$CD_i^2 = \frac{1}{p} r_i^2 \frac{h_{ii}}{1 - h_{ii}} \quad (4.6)$$

şeklinde verilir. (4.7) formu uydurulmuş değer vektöründeki i-inci gözlemin kaldırılmasının etkilerini ölçer diyebiliriz. (4.8) formundan hareketle CD_i^2 'nin veri setinin 3 etkenine bağlı olduğunu söyleyebiliriz. Bunlar, tahmin edilecek katsayı sayısı p, i-inci standartlaştırılmış rezidü ve i-inci rezidü varyansının $(\sigma^2(1 - h_{ii}))$ y_i 'nin varyansına $(\sigma^2 h_{ii})$ oranıdır. Böylece CD_i^2 , y_i 'yi ne kadar iyi uydurduğunu ve x_i 'nin geriye kalan x_j 'lerden ne kadar uzakta olduğunu ölçer. Her iki durumda CD_i^2 'nin büyük çıkmasına katkı sağlar. Eğer CD_i^2 1'den büyük ise i-inci gözlem sapan değerdir ve incelenmesi gerekir.

Besley ve ark. (1980) (4.8) formuna benzer aşağıdaki teşhis formunu vermişlerdir.

$$DFFITS(i) = \frac{e_i}{\hat{\sigma}(i)} \frac{\sqrt{h_{ii}}}{1 - h_{ii}} \quad (4.7)$$

Bu formdaki en önemli değişiklik $\hat{\sigma}$ yerine $\hat{\sigma}(i)$ 'nin yani i-inci gözlem çıkartıldığında elde edilen standart sapmanın kullanılmasıdır. Bu teşhis formunun orijinalliği $\hat{y} - \hat{y}(i)$ 'nin i-inci bileşenin standartlaştırılmasıdır. Böylece bu form bir

gözlemin silinmesinin tahminler üzerindeki etkisinin ölçüsüdür. Eğer DFFIFS(i) değeri $2\sqrt{\frac{p}{n}}$ 'den büyük ise i-inci gözlem sapan değerdir ve incelenmesi gerekir

Bahsettiğimiz bu teşhis yöntemleri en küçük karelerin h_{ii} , $\hat{y}$ gibi bazı fonksiyonları yardımıyla elde edilmiştir. Bu incelemeler y ve x yönünde problemler hakkında bilgi içerirler. Ancak ne yazık ki x yönündeki birden fazla sapan değer çeşidini sağlıklı bir şekilde elde edebilmemizi sağlayamazlar çünkü bu tip sapan değerler birbirlerinin etkilerini maskeleyebilirler. Bu yüzden çok dikkatli kullanılmaları gerekir.

Regresyon analizinde x yönünde sapan değer olan noktaları belirlemek için kullanılan en önemli ölçü Mahalanobis uzaklığıdır. Mahalanobis uzaklığı X katsayı matrisinin i-inci satırları x_i 'nin verinin merkezinden ne kadar uzakta olduğunu verinin oluşturduğu deseni göz önünde bulundurarak ölçer. (1.9)'da verilen modeli ele alalım. $\bar{x}$, X'in

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (4.8)$$

şeklinde verilen aritmetik ortalaması olmak üzere, X katsayı matrisinin varyans-kovaryans matrisi

$$C = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x})' \quad (4.9)$$

dir. Karesel mahalanobis uzaklığı,

$$MD_i^2 = (x_i - \bar{x})C^{-1}(x_i - \bar{x})' \quad (4.10)$$

,

şeklinde verilir. Dikkat edilirse bu ölçü X katsayı matrisinin i-inci satırı ile ortalaması arasındaki uzaklığı C varyans-kovaryans matrisine bağlı olarak verir. MD_i^2 değeri $(p - 1)$ serbestlik dereceli χ^2 dağılımı ile karşılaştırılır. Çünkü (4.12)'de verilen mahalanobis uzaklığı formu χ^2 dağılımından gelmektedir (Graybil, 1976). O halde,

$$MD_i^2 > \chi_{(p-1),0.95}^2 \quad (4.11)$$

ise i-inci gözlem X katsayı matrisinin merkezinden uzaktadır deriz.

Görüldüğü gibi mahalanobis uzaklığı hesabında $\bar{x}$ ve C varyans-kovaryans matrisi önemli bir yer tutmaktadır. Daha öncede belirtildiği gibi çoklu sapan değerlerin olması durumunda, h_{ii} 'lerde de olduğu gibi, bu sapan değerlere karşılık gelen MD_i^2 değerleri etkilerin maskelenmesi nedeniyle büyük olamayabilir. Diğer bir deyişle $\bar{x}$ ve C robust değillerdir ve sapan değerlerin küçük bir kümesi (hatta bir tanesi bile) $\bar{x}$ 'yi kendilerine doğru çekecekler ve C'yi de şişireceklerdir. x_i 'nin h_{ii} ve MD_i^2 değerleri arasında

$$h_{ii} = \frac{MD_i^2}{n-1} + \frac{1}{n} \quad (4.12)$$

şeklinde çok önemli monoton bir ilişki vardır. Yani h_{ii} 'lerin sapan değerlere karşı olan hassassızlıkları göz önüne alındığında MD_i^2 sapan değerlere karşı hassas robust olmayan bir istatistiktir denilebilir (Rousseeuw ve Zomeren, 1990). Özellikle karma veriler, çoğunlukla şiddetli çoklu iç ilişki veya kötü koşulluluk problemine sahip olduklarından MD_i^2 'ler gerçekçi sonuçlar üretemeyecektir. Bu durumda klasik varyans-kovaryans matrisi yerine X matrisindeki aykırı gözlem değerlerine karşı daha dirençli olan Rousseeuw (1985) tarafından verilen ve varyans-kovaryans matrisinin robust tahmini olarak da bilinen X'in minimum hacimli elipsoidini (MVE: Minimum Volume Ellipsoid) kullanarak elde edilen robust Mahalanobis uzaklığı daha gerçekçi sonuçlar üretecektir.

MVE tahmininin kırılma noktasının %50 olduğu bilinmektedir. Yani verinin %50'sinin sapan değer olması durumunda MVE ile elde edilen konum ve kovaryans matrisi tahmini güvenilir olacaktır. Buna göre MVE'ye bağlı robust mahalanobis uzaklığı, $T(X)$ ve $C(X)$, MVE tahmin edicisi ile elde edilen merkez ve varyans-kovaryans tahminleri olmak üzere,

$$RMD_i^2 = (x_i - T(X))C(X)^{-1}(x_i - T(X))' \quad (4.13)$$

dir. Robust MD_i^2 hesabında klasik varyans-kovaryans matrisi yerine kullanılabilecek bir diğer robust tahmin ise Rousseeuw (1984,1985) tarafından verilen minimum kovaryans determinant (MCD: minimum covariance determinant) tahmin edicisidir. Bu tahmin edici ise en küçük determinanta sahip alt kümeler araştırır.

Örnek 4.1. Bölüm 3.4'te incelenen şiddetli çoklu iç ilişki problemine sahip 4 bileşenli motor yağı karışımı verisini ele alalım. (3.40)'ta bileşenler üzerindeki ek kısıtlamalar ve Tablo 3.8'de verinin orijinal ve standartlaştırılmış şekli verilmiştir. Hatırlanacağı gibi veride 5. gözlem y yönünde sapan değer olacak şekilde değiştirilmişti. Bu örnekte bir gözlemi bozulmuş Tablo 3.8'de y_{bozuk} olarak gösterilen y amaç değerleri kullanılacaktır. Tablo 3.10'da en küçük kareler ile elde edilen katsayı tahminleri ile standart hataları ve VIF_i değerleri ile standartlaştırılmış X katsayı matrisinin özdeğerleri verilmiştir.

Tablo 4.1'de yukarıda bahsedilen tanımlama yöntemleri için elde edilen sonuçlar verilmiştir. 5. gözleme karşılık gelen alışılmış, standartlaştırılmış ve studentized rezidü değerlerinin tümü beklenildiği gibi büyüktür çünkü bu gözlem değeri sapan değer olacak şekilde değiştirilmişti. h_{ii} değerlerine baktığımızda $2p/n = 1.111$ olup 1'den büyük olduğu için h_{ii} 'lerin ortalaması olan 0.555'den büyük olan h_{ii} değerlerine dikkat etmek gerekir. 2., 4., 5., 7. ve 9. gözlemlere ait değerler ortalama değer üzerinde elde edilmiştir. Cook uzaklığı ve DFITS değerlerine baktığımızda, 5. ve 7. gözleme ait Cook uzaklığı 1'den büyük ve aynı şekilde özellikle bu iki gözleme ait DFITS değerleri ile 9. ve 13. gözlemlere ait değerler

$2\sqrt{p/n} = 1.491$ değerinden büyüktür. 2. gözlem için C varyans-kovaryans matrisi kullanılarak elde edilen mahalanobis uzaklığı değeri geriye kalan gözlemlerden oldukça büyük olduğu göze çarpmaktadır. 5. ve 7. gözlemlere ait MVE kullanılarak elde edilen robust mahalanobis uzaklığı değerlerine dikkat edilirse $\chi^2_{0.95,p-1} = \chi^2_{0.95,9} = 16.91898$ değerinden oldukça büyüktür. Buna göre 5. gözlemin hem x hemde y yönünde sapan değer olduğunu ve 7. gözlemin x yönünde sapan değer olduğunu söyleyebiliriz.

Tablo 4.1. Motor yağı karışımı verisi için teşhis sonuçları

#	$\hat{e}_i$	r_i	t_i	h_{ii} (0.555)	Cook > 1	DFITS (1.49071)	MD	RMD
1	-0.562	-0.481	-0.457	0.442	0.0183	-0.4064	0.308	8.039
2	-0.072	-0.169	-0.159	0.927	0.0362	-0.5635	1.147	13.298
3	0.941	0.8759	0.862	0.528	0.0859	0.9116	0.528	8.768
4	-0.371	-0.549	-0.523	0.813	0.1309	-1.0909	0.123	12.114
5	-2.054	-2.828	-113.505	0.784	2.9026	-216.263	0.301	208.379
6	-1.108	-1.028	-1.032	0.525	0.1166	-1.0841	0.466	9.3687
7	1.319	2.339	3.892	0.869	3.6495	10.0511	0.340	416.093
8	-0.248	-0.210	-0.197	0.429	0.0033	-0.1712	0.345	7.714
9	0.779	1.197	1.236	0.827	0.6845	2.7012	0.303	12.819
10	1.086	0.939	0.932	0.453	0.0732	0.8487	0.264	9.631
11	-0.824	-0.687	-0.663	0.411	0.0330	-0.5539	0.065	8.971
12	1.003	0.847	0.831	0.427	0.0535	0.7169	0.027	9.046
13	-1.828	-1.617	-1.844	0.477	0.2382	-1.7595	0.148	9.325
14	1.035	0.847	0.830	0.388	0.0455	0.6610	0.072	6.922
15	1.672	1.382	1.482	0.401	0.1277	1.21150	0.321	9.250
16	-0.395	-0.345	-0.326	0.466	0.0104	-0.3043	0.147	8.608
17	-0.043	-0.032	-0.030	0.275	0.0001	-0.0186	0.001	7.529
18	-0.332	-0.319	-0.301	0.558	0.0128	-0.3374	0.382	8.597

Regresyon analizinde bir gözlemin model katsayı tahminlerine etkisini anlayabilmek için en etkili yol ilgili gözlem modelde iken ve modelden çıkarıldığında tahmin yaparak karşılaştırmaktır. Buna göre, motor yağı karışımı verisinden 5. ve 7. gözlemi çıkarttığımızda en küçük kareler ile elde ettiğimiz katsayı tahminleri ile orijinal veri ile elde ettiğimiz katsayı tahminlerini karşılaştırsak bu iki gözlemin gerçekten etkili olup olmadıkları konusunda fikir sahibi olabiliriz. Tablo 4.2’de 5. ve 7. gözlemi çıkartılmış 16 gözlemlili motor yağı karışımı verisinin en küçük kareler sonuçları verilmiştir.

Tablo 4.2. 16 gözlemlili motor yağı karışımı verisi en küçük kareler sonuçları.

Terim	$\hat{\beta}_{LS}$	$se(\hat{\beta}_{LS})$	Özdeğer	VIF _i
$\hat{\beta}_1$	74,96436	3,721593	6,75017	8112,974
$\hat{\beta}_2$	6,028792	0,910098	1,34768	485,1756
$\hat{\beta}_3$	18,0969	0,715631	0,46912	299,9857
$\hat{\beta}_4$	8,888515	3,890649	0,17424	8866,791
$\hat{\beta}_{12}$	-12,9996	1,225619	0,03350	879,9004
$\hat{\beta}_{13}$	-35,3197	2,802102	0,01573	4599,283
$\hat{\beta}_{14}$	-4,21867	1,470038	0,00265	1265,842
$\hat{\beta}_{23}$	-2,90572	0,712832	0,00163	297,6435
$\hat{\beta}_{24}$	-0,81278	0,433496	0,00008	110,0761
$\hat{\beta}_{34}$	-2,87985	2,710856	0,00006	4304,624
$\hat{\sigma}^2 = 0.001707$ $\hat{\sigma} = 0,04131796$ $KS = 110553$ $MSEOLS = 49.88753$				

Dikkat edilirse etkili olarak belirlenen 5. ve 7. gözlemler çıkarıldığında elde edilen sonuçlar ile Tablo 3.10'da verilen, bu iki gözlem modelde iken elde edilen sonuçlar arasında çok önemli farklar vardır. İlk olarak katsayı tahminlerine bakıldığında gerek değer olarak gerekse işaret olarak farklılıklar görüyoruz. Ayrıca katsayıların standart hatalarına ve MSE değerine bakıldığında çok ciddi iyileşmeler hemen göze batmaktadır. Bunlara karşılık koşul sayısı ve VIF_i değerleri hala kötü koşulluluk probleminin olduğunu göstermektedir ancak karma denemelerde 2. dereceden modeller için X katsayı matrisleri değişkenlerin çarpımlarını içerdiklerinden bu beklenen olağan bir durumdur. O halde gerçekte bu iki gözlem katsayı tahminleri üzerinde önemli etkilere sahiptir diyebiliriz.

4.2. Genelleştirilmiş M (GM) Tahmin Edici

Daha öncede belirttiğimiz gibi M tahmin edicilerinin kırılma noktaları $1/n$ 'dir. Yani rezidüler, gözlemlere karşılık gelen tasarım noktası ile ilişkisiz olarak ağırlıklandırılır. Bir gözleme karşılık gelen rezidünün çok küçük ancak karşılık gelen şapka matrisinin köşegen elemanı h_{ii} değeri çok büyük olabilir. Bu durumda M tahmin edicileri ile yüksek etki noktaları 1'ya yakın (hatta 1) ağırlığına sahip

olacaktır. Böylece bu tip aykırı gözlemler tahmin edicinin yararlılığını ve kararlılığını kötü etkileyecektir. Yani M tahmin edicileri y yönündeki sapan değerlere karşı etkili ancak X yönündeki sapan değerlere karşı hassas olan tahmin edicilerdir.

M tahmin edicilerinin bu dezavantajı *sınırlı etkili regresyon tahmin edicilerinin* (bounded influential regression estimators) geliştirilmesini sağlamıştır. Bu tahmin edicilere Genelleştirilmiş M veya GM tahmin edicileri denir. GM tahmin edicilerinin M tahmin edicilerine göre avantajı X yönündeki sapan değerlere karşı hassas olmamalarıdır. GM tahmin edicilerinin dezavantajı ise kırılma noktalarının $1/p$ 'den büyük olmamalarıdır. Sonuç olarak, çok parametrelili bir model için kırılma noktası $1/n$ 'den daha iyi olmayacaktır. Ancak p parametre sayısı küçük ise bu tahmin ediciler çok iyi sonuçlar üretecektir (Montgomery ve ark. (2001)).

Lineer regresyon için en genel halde GM tahmin edicileri, ρ fonksiyonu $[0, \infty)$ aralığında azalmayan çift bir fonksiyon ve u_i 'ler x_i 'lere bağlı ağırlıklar olmak üzere,

$$\sigma^2 \sum_{i=1}^n \rho \left(\frac{y_i - x_i' \beta}{\sigma u_i^\alpha} \right) u_i^{1+\alpha} = \min \quad (4.14)$$

formundadır (Simpson, D.G. ve ark. (1992)). Dikkat edilirse GM tahmin edicileri M tahmin edicilerine x 'e bağlı ağırlıkların eklenmesi ile elde edilmektedir. Bu şekilde tahmin edicinin hem x hem de y uzayında sınırlandırılması sağlanmıştır. Robust istatistikte seçilen ρ fonksiyonunun türevinin yani ψ etki fonksiyonunun sınırlı olması gerekir. GM tahmin edicileri için sınırlı etki fonksiyonu elde edebilmek için $u(x)\psi(e_i/\sigma)x$ 'in sınırlı olması gerekir. Yani $u(x)x$ sınırlı olacak şekilde seçilmelidir (Arslan, O. (1992)).

(4.14) formunda eğer $\rho(x) = x^2$ ve $u_i = 1$ seçilirse,

$$\sigma^2 \sum_{i=1}^n \frac{(y_i - x_i' \beta)^2}{\sigma^2} = \sum_{i=1}^n (y_i - x_i' \beta)^2 = \min \quad (4.15)$$

şeklinde en küçük kareler tahmin edicisi elde edilir. Aynı şekilde eğer $\rho(x) = |x|$ ve $u_i = 1$ seçilirse,

$$\sum_{i=1}^n |y_i - x_i' \beta| = \min \quad (4.16)$$

L_1 tahmin edicisi elde edilir. (4.14)'te verilen formda $u_i = 1$ seçilirse y yönündeki aykırı değerlere karşı hassas olan,

$$\sigma^2 \sum_{i=1}^n \rho\left(\frac{y_i - x_i' \beta}{\sigma}\right) = \min \quad (4.17)$$

şeklinde Huber M tahmin edicisi elde edilir.

(4.14) formunda α 'nın seçimi önemlidir. GM tahmin edicileri için iki tip yaklaşım vardır. İlki Mallows (1975)'in verdiği (4.14) formunda $\alpha = 0$ olarak seçilmesiyle elde edilen

$$\sigma^2 \sum_{i=1}^n \rho\left(\frac{y_i - x_i' \beta}{\sigma}\right) u_i = \min \quad (4.18)$$

şeklindeki Mallows yaklaşımıdır. Mallows tipi M tahmin edicisi olarak da bilinen bu yaklaşım Huber ρ fonksiyonu ile birlikte uygun u_i ağırlıkları kullanarak sapan değerlerle birlikte etkili gözlemlerinde etkilerini azaltmaya çalışır (Şekil 4.2). Uygun Mallows ağırlığı, $\bar{x}_c$, MVE (veya MCD)'nin merkez değerleri, C, MVE (veya MCD) ile elde edilen varyans-kovaryans matrisi ve b'de sabit bir sayı olmak üzere,

$$u_i = \min \left[1, \left\{ \frac{b}{(x_i - \bar{x}_c)' C^{-1} (x_i - \bar{x}_c)} \right\}^{1/2} \right] \quad (4.19)$$

şeklinde verilir (Simpson ve ark., 1992). b sayısı $\alpha = 0.1$ veya 0.05 olacak şekilde $1 - \alpha$ önem düzeyine sahip, $p - 1$ serbestlik dereceli χ^2 dağılımı olarak seçilir.

İkinci yaklaşım ise Schweppe (1975)'in verdiği (4.14) formunda $\alpha = 1$ olarak seçilmesiyle elde edilen

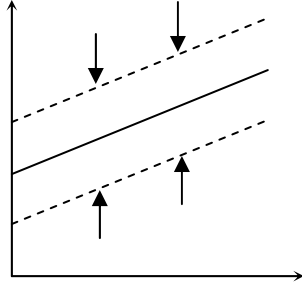
$$\sigma^2 \sum_{i=1}^n \rho \left(\frac{y_i - x_i' \beta}{\sigma u_i} \right) u_i^2 = \min \quad (4.20)$$

şeklindeki Schweppe yaklaşımıdır. Schweppe tipi M tahmin edicisinin Mallows tipi M tahmin edicinin aksine etkili gözleme karşılık gelen rezidü değeri büyükse etkili gözlemin etkilerini azaltmaya çalışır (Şekil 4.2). Uygun Schweppe ağırlığı, $\bar{x}_c$, MVE'nin merkez değerleri, C , MVE (veya MCD) ile elde edilen varyans-kovaryans matrisi olmak üzere,

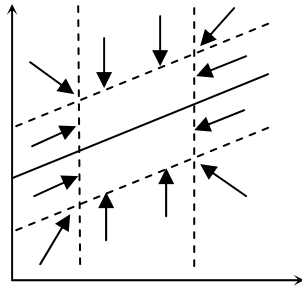
$$u_i = \frac{1}{\sqrt{(x_i - \bar{x}_c)' C^{-1} (x_i - \bar{x}_c)}} \quad (4.21)$$

şeklinde verilir (Hampel ve ark., 1986).

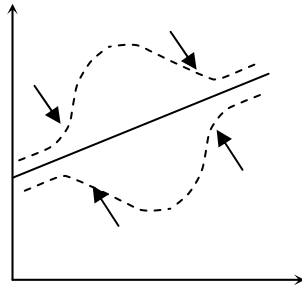
Burada görülüyor ki ağırlıklar x_i noktasının orijinden metrik uzaklığının bir fonksiyonudur. Metrik bilinen sabit konum ile robust kovaryans matrisi olarak verilebilir. Daha öncede söylendiği Mallows formu X uzayındaki konum ile rezidüler için farklı sınırla koyar ve X uzayındaki etkili gözlemler için ağırlıkları amaç değişkenin durumunu dikkate almaksızın belirler. Ancak, amacın veriyi nasıl uydurduğu hakkında bilgi içermeyen düşük ağırlıklı x_i , etkili bir tahmin vermez çünkü aykırı nokta $\hat{\beta}$ 'nin güvenilirliğine katkıda bulunuyor olabilir ve böylece tahmin edicinin etkinliğini arttırabilir. Buna karşılık Schweppe formu ise, x_i 'lerin ağırlıklarını karşılık gelen rezidü değeri büyükse düşürdüğü için bir gözlemin toplam etkisine sınır koyar. Buna göre, Schweppe formu bu tip durumlarda Mallows formuna göre daha iyi sonuçlar üretecektir denilebilir.



Huber M tahmin edicisi X yönündeki sapan değerlerin etkilerini sınırlamaz. Sadece y yönündeki sapan değerlerin etkilerini sınırlar.



Mallows tipi M tahmin edici hem X yönündeki etkili gözlemlerin hem de y yönündeki sapan değerlerin etkilerini sınırlar.



Schweppe tipi M tahmin edici X yönündeki etkili gözlemlere karşılık gelen rezidü büyükse, bu gözlemlerin etkilerini sınırlar.

Şekil 4.2. M ve GM tahmin edicilerinin aykırı gözlemlere karşı davranışları (Hampel, F.R. ve ark. (1986)).

Buna göre Huber M tahmininin elde edildiği gibi, β için GM tahmin edicisi elde edilir. Bu çalışmada σ , verilerden bir defaya mahsus olarak hesaplanıp (örneğin (2.8)'de verilen MAD hesaplanarak) sabit olarak kabul edilecektir. (4.23)'te verilen formun β 'ya göre türevini alıp sıfıra eşitlersek, $\psi = \rho'$ olmak üzere,

$$\sigma^2 \sum_{i=1}^n \rho' \left(\frac{y_i - x_i' \beta}{\sigma u_i^\alpha} \right) \left(\frac{-x_i' \sigma u_i^\alpha - 0}{\sigma^2 u_i^{2\alpha}} \right) u_i^{1+\alpha} = 0 \quad (4.22)$$

ve

$$\sigma \sum_{i=1}^n \psi \left(\frac{y_i - x_i' \beta}{\sigma u_i^\alpha} \right) x_i' u_i = 0 \quad (4.23)$$

elde edilir. Dikkat edilirse $\alpha = 0$ alınırsa Mallows, $\alpha = 1$ alınırsa Schweppe tipi GM tahmini elde edilir. (4.32)'te elde edilen denklem,

$$v(t) = \begin{cases} \frac{\psi(t)}{t}, & t \neq 0 \\ \psi'(0), & t = 0 \end{cases} \quad (4.24)$$

ağırlık fonksiyonu olmak üzere $(y_i - x_i' \beta) / (\sigma u_i^\alpha)$ ile aynı anda çarpılıp bölünürse,

$$\sum_{i=1}^n v \left(\frac{y_i - x_i' \beta}{\sigma u_i^\alpha} \right) x_i' u_i^{1-\alpha} (y_i - x_i' \beta) = 0 \quad (4.25)$$

elde edilir. (4.25) denkleminde $\hat{u}_i = v \left(\frac{y_i - x_i' \beta}{\sigma u_i^\alpha} \right) u_i^{1-\alpha}$ alınırsa,

$$\sum_{i=1}^n (y_i - x_i' \beta) x_i' \hat{u}_i = \sum_{i=1}^n \hat{u}_i x_i (y_i - x_i' \beta) = 0$$

ve

$$\sum_{i=1}^n \hat{u}_i x_i y_i - \sum_{i=1}^n \hat{u}_i x_i x_i' \beta = 0 \quad (4.26)$$

denklemini elde edilir. (4.26) denkleminde eşitliğin her iki tarafının $\left[\sum_{i=1}^n \hat{u}_i x_i x_i' \right]^{-1}$ ile

çarpılmasıyla β 'nin GM tahmini

$$\hat{\beta}_{GM} = \left(\sum_{i=1}^n \hat{u}_i x_i x_i' \right)^{-1} \sum_{i=1}^n \hat{u}_i x_i y_i \quad (4.27)$$

olarak elde edilir. $y = (y_1, \dots, y_n)'$, $X = [x_1, \dots, x_n]'$ ve $\hat{U} = \text{diag}(\hat{u}_1, \dots, \hat{u}_n)$ olarak alınırsa matris notasyonu ile GM tahminin normal denklemleri,

$$(X' \hat{U} X) \hat{\beta} = X' \hat{U} y \quad (4.28)$$

ve β 'nin GM tahmini,

$$\hat{\beta}_{GM} = (X' \hat{U} X)^{-1} X' \hat{U} y \quad (4.29)$$

şeklinde elde edilir.

Örnek 4.2. Örnek 4.1'de ele aldığımız motor yağı karışımı verisini ele alalım. Bölüm 3.4'te bu veri için M tahmin edicileri hesaplanırken σ ölçek tahmini sabit kabul edilmemiş her iterasyonda tekrar hesaplanmıştı. Bu bölümde daha öncede belirtildiği gibi σ bir defaya mahsus olmak üzere hesaplanıp sabit kabul edilecektir. Tablo 4.3'te robust M tahmin edicileri için elde edilen katsayı tahminleri ve $\hat{\sigma}$ değerleri verilmiştir.

Tablo 4.3. Motor yağı karışımı verisi için M tahmin edicileri (σ , (2.8)'de verilen MAD ile hesaplanıp sabit kabul edildi).

Terim	$\hat{\beta}_{LS}$	Huber $k = 1.345$	Tukey $c = 3.67$	T $v = 2$	GT $p = 1.2, q = 2$
$\hat{\beta}_1$	13.355	33.7246	74.3938	67.3288	72.4355
$\hat{\beta}_2$	19.451	14.8755	5.7398	7.3512	6.2796
$\hat{\beta}_3$	0.578	6.4844	18.2781	16.2393	17.7830
$\hat{\beta}_4$	-94.826	-60.7422	7.3113	-4.5520	4.8128
$\hat{\beta}_{12}$	3.657	-1.7645	-12.5894	-10.7139	-12.1489
$\hat{\beta}_{13}$	-6.222	-15.9304	-35.3137	-31.9327	-34.0031
$\hat{\beta}_{14}$	35.543	22.4776	-3.6088	0.9194	-2.7048
$\hat{\beta}_{23}$	0.017	-0.9187	-2.7872	-2.4943	-2.9521
$\hat{\beta}_{24}$	4.267	2.6141	-0.6869	-0.0973	-0.4631
$\hat{\beta}_{34}$	82.855	54.6021	-1.8079	8.0113	0.1867
$\hat{\sigma}$	1.562858	1.017391	1.017391	1.017391	1.017391

Tablo 4.4'te M tahmin edicileri için katsayıların standart hataları ve $\hat{A}^2$ değeri verilmiştir. Tablo 4.5'te ise robust tahmin ediciler için ağırlıklar ve MAD ile standartlaştırılarak elde edilen hatalar verilmiştir. GT için elde edilen sonuçlar iyi olmamasına rağmen p ve q şekil parametrelerinin farklı seçimleri ve σ ölçek parametresinin sabit kabul edilmemesi daha iyi sonuçlar üretecektir. Tablo 3.12 ve Tablo 3.13'te σ 'nın sabit kabul edilememesi ile elde edilen sonuçlar bunu kanıtlamaktadır ve ayrıca örneğin p = 3 ve q = 4 seçilirse $\hat{A}$ değeri 0.678 elde ediliyor dolayısıyla standart hatalarda en büyüğü 69.11 olarak daha uygun bir seviyede elde ediliyor.

Tablo 4.4. M tahmin edicileri için katsayıların standart hataları.

Terim	$\hat{\beta}_{LS}$	Huber k = 1.345	Tukey c = 3.67	T v = 2	GT p = 1.2, q = 2
$\hat{\beta}_1$	131.128	84.62912	89.94505	96.24981	171.8492
$\hat{\beta}_2$	24.645	15.9056	16.9047	18.08965	32.29816
$\hat{\beta}_3$	21.806	14.0732	14.9572	16.00564	28.57725
$\hat{\beta}_4$	61.997	40.01239	42.52574	45.50661	81.24975
$\hat{\beta}_{12}$	31.622	20.40889	21.69086	23.21129	41.44259
$\hat{\beta}_{13}$	105.488	68.08153	72.35803	77.43002	138.2474
$\hat{\beta}_{14}$	20.806	13.42804	14.27152	15.27189	27.26719
$\hat{\beta}_{23}$	25.157	16.23591	17.25576	18.46531	32.96888
$\hat{\beta}_{24}$	12.672	8.178227	8.691938	9.301205	16.60683
$\hat{\beta}_{34}$	44.644	28.81281	30.62268	32.76919	58.50774
$\hat{A}^2$	$\hat{\sigma}^2 = 2.4425$	1.017391	1.149219	1.315976	4.195108

Tablo 4.5 ve Tablo 4.6'te robust M tahmin edicilerinin kullandıkları ağırlıklar ve MAD ile standartlaştırılmış hatalar verilmiştir. 5. gözlemin sapan değer olarak değiştirilmiş olması bu gözleme karşılık gelen ağırlıkların küçük olmasına neden olmuştur. Ayrıca bu gözleme ait robust M hata değerlerinin r_5 'e göre daha yüksek olduğu yani bu gözlem için en küçük kareler ile elde edilen hatanın olması gerekenden daha küçük elde edildiği görülmektedir.

Tablo 4.5. Robust tahmin ediciler için ağırlıklar.

	Huber	Tukey	T	GT		Huber	Tukey	T	GT
1	1.000	0.999	1.499	4.448	10	1.000	0.999	1.491	12.779
2	1.000	0.999	1.499	2401	11	1.000	0.999	1.498	3.282
3	1.000	0.999	1.495	69.997	12	1.000	0.999	1.488	3.308
4	1.000	0.999	1.497	9.809	13	1.000	0.999	1.454	5.871
5	0.301	0.000	0.040	0.034	14	1.000	0.999	1.492	9.907
6	1.000	0.999	1.489	69.415	15	1.000	0.999	1.4684	4.005
7	1.000	0.999	1.484	10230.3	16	1.000	0.999	1.4989	6.276
8	1.000	0.999	1.499	1181.44	17	1.000	0.999	1.4999	170.47
9	1.000	1.000	1.494	1692.46	18	1.000	0.999	1.4969	7.295

Tablo 4.6. Robust M tahmin edicileri için MAD ile standartlaştırılmış hatalar.

	r_i (ekk)	Huber	Tukey	T	GT
1	-0.4812	-0.3564	0.0346	-0.0328	0,0280
2	-0.1691	-0.0395	0.0220	0.0105	0,00004
3	0.87598	0.6056	-0.0309	0.0799	-0,00006
4	-0.5488	-0.2495	-0.0196	-0.0583	-0,0095
5	-2.8277	-4.4641	-9.3467	-8.5028	-9,21399
6	-1.0279	-0.7221	0.0098	-0.1165	-0,0083
7	2.3391	0.8632	-0.0038	0.1455	0,000006
8	-0.2102	-0.1631	-0.0019	-0.0294	-0,00080
9	1.1972	0.5097	-0.0004	0.0868	0,00000
10	0.9398	0.7058	-0.0161	0.1078	0,00050
11	-0.6873	-0.5233	0.0492	-0.0505	0,02958
12	0.8474	0.6621	0.0164	0.1273	0,02644
13	-1.6171	-1.2133	-0.0479	-0.2516	-0,10846
14	0.8467	0.6721	-0.0175	0.1007	-0,00382
15	1.3822	1.1020	0.0202	0.2072	0,0465
16	-0.3455	-0.2560	0.0071	-0.0374	0,0006
17	-0.0323	-0.0246	0.0106	0.0037	0,00005
18	-0.3193	-0.2279	-0,0315	-0.0643	-3,4236

Motor yağı karışımı verisi GM tahmin edicileri kullanıldığında elde edilecek sonuçların, Tablo 4.2’de 5. ve 7. gözlemimi çıkartılmış 16 gözlem için en küçük kareler tahmin edicilerine yakın olmasını beklenir. Ayrıca x yönündeki $u(x)$ ağırlıklarda bu iki gözleme karşılık gelen ağırlıkların diğerlerinden daha küçük olması ve y yönündeki ağırlıklarda ise sadece 5. gözleme karşılık gelen ağırlığın diğerlerinden daha küçük olması beklenir. Daha öncede belirtildiği gibi σ ölçek parametresi bir defaya mahsus olarak verilerden hesaplanarak sabit kabul edilecektir. Ayrıca GM tahmin edicileri Huber ρ fonksiyonu kullanılarak elde edilecektir. Buna göre Mallows ve Schweppe tipi GM tahmin edicileri x yönündeki ağırlıklara sırasıyla

(4.19) ve (4.21)'de verilen formlarda sahip olacaklardır. y yönündeki ağırlıklar ise, Mallows için $\alpha = 0$ ve Schweppe için $\alpha = 1$ olmak üzere aşağıdaki gibi olacaktır.

$$w(y) = \frac{k}{\left| \frac{y_i - x_i' \hat{\beta}}{\hat{\sigma}(u(x_i))^\alpha} \right|} \quad (4.30)$$

Tablo 4.7'de Mallows tipi GM tahmin edicisi elde edilen katsayı tahminleri ile x ve y yönündeki ağırlıklar verilmiştir.

Tablo 4.7. Mallows tipi GM tahmini için elde edilen sonuçlar.

Terim	$\hat{\beta}_{GMM}$	#	$w(y)$	$u(x)$	GM için hata
x1	62.8058097	1	1.000	1.000	-0.0768
x2	8.3429841	2	1.000	1.000	0.0045
x3	14.9176127	3	1.000	1.000	0.1505
x4	-12.0801706	4	1.000	1.000	-0.0851
x1x2	-9.5050121	5	0.048175	0.284944	-7.9554
x1x3	-29.7909456	6	1.000	1.000	-0.1988
x1x4	3.8242785	7	1.000	0.201647	0.2432
x2x3	-2.2548128	8	1.000	1.000	-0.0479
x2x4	0.2537239	9	1.000	1.000	0.1449
x3x4	14.2657405	10	1.000	1.000	0.1896
		11	1.000	1.000	-0.1140
		12	1.000	1.000	0.2004
		13	1.000	1.000	-0.3800
		14	1.000	1.000	0.1790
		15	1.000	1.000	0.3284
		16	1.000	1.000	-0.0679
		17	1.000	1.000	0.0006
		18	1.000	1.000	-0.0875

Mallows tipi GM için elde edilen sonuçlara dikkat edilirse beklenildiği gibi 16 gözlemden oluşan en küçük kareler tahminine yakındır ve benzer şekilde 5. ve 7. gözlemlere karşılık gelen ağırlıklarda geriye kalanlardan oldukça küçük elde edilmiştir. 5. gözleme ait hata değerinin de beklenildiği gibi diğerlerinden oldukça büyük olduğu görülmektedir.

Tablo 4.8’de Schweppe tipi GM tahmin edicisi elde edilen katsayı tahminleri ile x ve y yönündeki ağırlıklar verilmiştir. Elde edilen sonuçlara dikkat edilirse Mallows tipi GM tahmininde olduğu gibi 16 gözlemlili sonuçlara yakındır. 16 ve 18 gözlemlili motor yağı karışımı verisi için en küçük kareler ile elde edilen sonuçlar arasındaki fark dikkate alındığında GM tahmin edicilerinin bu iki gözlemin etkilerini en aza indirdiği açıkça görülmektedir.

Tablo 4.8. Schweppe tipi GM tahmini için elde edilen sonuçlar.

Terim	$\hat{\beta}_{GMS}$	#	$w(y)$	$u(x)$	GM için hata
x1	71.5770530	1	1.000	0.35268893	0,0075
x2	6.3726885	2	1.000	0.27422730	0,0177
x3	17.4611650	3	1.000	0.33772022	0,0132
x4	2.5969171	4	1.000	0.28731299	-0,0355
x1x2	-11.8396373	5	0.0103429	0.06927439	-9,0085
x1x3	-33.9714417	6	1.000	0.32671929	-0,0409
x1x4	-1.8018022	7	1.000	0.04902354	0,0563
x2x3	-2.6577933	8	1.000	0.36004646	-0,0131
x2x4	-0.4581876	9	1.000	0.27929644	0,0349
x3x4	2.0997953	10	1.000	0.32223512	0,0339
		11	1.000	0.33386570	0,0095
		12	1.000	0.33247806	0,0611
		13	1.000	1.75325311	-0,1287
		14	1.000	0.38009973	0,0303
		15	1.000	0.32880116	0,0951
		16	1.000	0.34084196	-0,0112
		17	1.000	0.36443882	0,0081
		18	1.000	0.34104708	-0,0451

4.3. GM Tahminine Dayalı Ridge ve Liu Tahmin Edicileri

Regresyon analizinde çoklu iç ilişki probleminin olması durumunda en çok kullanılan yöntemlerin Ridge ve Liu yöntemleridir. Hatırlanacağı gibi bu iki yöntemin en küçük kareler yöntemine dayalı olduğu ancak en küçük kareler yönteminin aykırı değerlere karşı çok hassas olmasından dolayı performanslarının kötü etkilendiği belirtilmişti. 3. Bölümde bu tahmin edicilerde en küçük kareler tahmin edicisinin yerine robust M tahmin edicilerinin kullanılmasının etkileri incelenmişti. Bu bölümde ise en küçük kareler tahmin edicisi yerine GM tahmin

edicilerinin kullanılmasının etkileri incelenecektir. Bunun için Örnek 4.1’de ele alınan motor yağı karışımı verisi kullanılacaktır.

İlk olarak Bölüm 3.2 ve 3.3’te ele alınan Ridge ve robust Ridge tahmin edicisini inceleyelim. Tablo 4.9’da motor yağı karışımı verisinden alışımlı Ridge tahmin edicisi ile elde edilen sonuçlar verilmiştir.

Tablo 4.9. Motor yağı karışımı verisi için en küçük kareler ve alışımlı Ridge tahmin edicisi ile elde edilen sonuçlar.

Terim	$\hat{\beta}_{LS}$	$se(\hat{\beta}_{LS})$	VIF _i -ler	$\hat{\beta}_R$	$se(\hat{\beta}_R)$	R-VIF _i
x1	13.355	131.128	7039,660	4.5613021	4.381725	7.860514
x2	19.451	24.645	248,663	10.1567343	9.831985	39.57702
x3	0.578	21.806	194,669	0.6766834	9.641665	38.05965
x4	-94.826	61.997	1573,6241	-3.1187714	6.718733	18.48143
x1x2	3.657	31.622	409,402	6.4164118	6.321545	16.3609
x1x3	-6.222	105.488	4555,863	3.0814752	5.610025	12.88517
x1x4	35.543	20.806	177,230	19.6165868	6.216967	15.82406
x2x3	0.017	25.157	259,098	6.6122998	8.730988	31.20955
x2x4	4.267	12.672	65,740	-11.5788991	4.535266	8.421051
x3x4	82.855	44.644	815,987	17.3803515	6.661556	18.16821
k = 0.0037 MSEOLS = 37468.21 MSERIDGE = 502.1269						

Dikkat edilirse Ridge tahmini en küçük karelere göre çok daha iyi MSE, VIF_i ve standart hata değerlerine sahiptir. Ayrıca en küçük kareler tahmininde kötü koşulluluk problemlinden kaynaklanan şişme, özellikle x₄ için, hemen göze batmaktadır. Tablo 4.10’da alışımlı Ridge tahmin edicisinde en küçük kareler tahmini yerine robust M tahmin edicileri (bu örnekte Huber M tahmini) kullanıldığında elde edilen sonuçlar verilmiştir. Sonuçlara dikkat edilirse MSE, standart hata ve VIF_i açısından daha iyi sonuçlar elde edilmiştir.

Tablo 4.11’de en küçük kareler yerine GM tahmin edicileri kullanarak elde edilen robust Ridge tahmin sonuçları verilmiştir. Elde edilen sonuçlara dikkat edilirse robust M tahminlerine dayalı Ridge tahminlerine göre her bakımdan daha iyi olan VIF_i değerleri açısından kötü koşulluluk probleminin etkilerinin tamamıyla ortadan kalkmış olduğu görülür.

Tablo 4.10. Huber M tahminine dayalı robust Ridge tahmini sonuçlar.

	$\hat{\beta}_{RM}$	$se(\hat{\beta}_{RM})$	R-VIF_i
$\hat{\beta}_1$	7.6938439	2.738808	7.372848
$\hat{\beta}_2$	7.6085603	6.166462	37.37526
$\hat{\beta}_3$	4.2366447	6.074706	36.27126
$\hat{\beta}_4$	-0.5210319	4.156338	16.97985
$\hat{\beta}_{12}$	4.9360697	4.028257	15.94948
$\hat{\beta}_{13}$	6.5752725	3.542223	12.33286
$\hat{\beta}_{14}$	14.9289221	3.976703	15.54384
$\hat{\beta}_{23}$	4.3930317	5.456299	29.2623
$\hat{\beta}_{24}$	-7.8432065	2.90346	8.285978
$\hat{\beta}_{34}$	11.6257653	4.184912	17.21412
k = 0.0039 MSEOLS = 37468.21 MSERIDGE = 50.83759			

Tablo 4.11. Motor yağı karışımı verisi için GM tahminlerine dayalı robust Ridge tahmin edicisi sonuçları.

	Mallows tipi GM tahmini			Schweppe tipi GM tahmini		
	$\hat{\beta}_{RGMM}$	$se(\hat{\beta}_{RGMM})$	R-VIF_i	$\hat{\beta}_{RGMS}$	$se(\hat{\beta}_{RGMS})$	R-VIF_i
$\hat{\beta}_1$	10.931580	1.197712	1.409993	9.924137	1.361298	1.821457
$\hat{\beta}_2$	2.808700	1.789153	3.14635	2.655433	1.35471	1.803869
$\hat{\beta}_3$	8.390462	1.985499	3.874819	8.212600	1.707809	2.866754
$\hat{\beta}_4$	3.583207	1.19588	1.405683	4.325191	1.14454	1.28758
$\hat{\beta}_{12}$	3.635498	2.238538	4.925393	3.791233	2.098942	4.330249
$\hat{\beta}_{13}$	11.758544	1.739291	2.973422	10.810176	1.567453	2.414911
$\hat{\beta}_{14}$	7.936188	2.563434	6.458868	7.310153	2.209787	4.799688
$\hat{\beta}_{23}$	2.668593	1.719909	2.907522	3.146743	1.554893	2.376366
$\hat{\beta}_{24}$	-2.646799	2.207778	4.790962	-2.217968	2.136578	4.486932
$\hat{\beta}_{34}$	4.425732	1.928084	3.65396	5.224179	1.781817	3.120601
k = 0.02379 MSEOLS = 37468.21 MSERIDGE-GM = 36.414				k = 0.0285 MSEOLS = 37468.21 MSERIDGE-GM = 3.008133		

Ridge tahminleri için yapılan inceleme Liu tahmini içinde yapıldığında yine alışılmış Liu tahmine göre çok daha iyi sonuçlar elde edilir. Tablo 4.12’de motor yağı karışımı verisinden alışılmış Liu tahmin edicisi ile elde edilen sonuçlar verilmiştir.

Tablo 4.12. Motor yağı karışımı verisi için en küçük kareler ve alışılmış Liu tahmin edicisi ile elde edilen sonuçlar.

	$\hat{\beta}_{LS}$	$se(\hat{\beta}_{LS})$	VIF _i -ler	$\hat{\beta}_L$	$se(\hat{\beta}_L)$	L-VIF _i
$\hat{\beta}_1$	13.355	131.128	7039,660	8.488244	41.57892	707.7943
$\hat{\beta}_2$	19.451	24.645	248,663	8.432130	7.870697	25.36221
$\hat{\beta}_3$	0.578	21.806	194,669	4.239941	6.980094	19.94726
$\hat{\beta}_4$	-94.826	61.997	1573,6241	-26.519316	19.67651	158.51
$\hat{\beta}_{12}$	3.657	31.622	409,402	3.651535	10.07022	41.51817
$\hat{\beta}_{13}$	-6.222	105.488	4555,863	2.385151	33.4538	458.1964
$\hat{\beta}_{14}$	35.543	20.806	177,230	15.478449	6.665172	18.18793
$\hat{\beta}_{23}$	0.017	25.157	259,098	2.406394	8.031658	26.41016
$\hat{\beta}_{24}$	4.267	12.672	65,740	2.071275	4.128002	6.976547
$\hat{\beta}_{34}$	82.855	44.644	815,987	30.194690	14.18452	82.37395
d = 0.317 MSEOLS = 37468.21 MSELIU = 11859.06						

Tablo 4.13'te alışılmış Liu tahmin edicisinde en küçük kareler tahmini yerine robust M tahmin edicileri (bu örnekte Huber M tahmini) kullanıldığında elde edilen sonuçlar verilmiştir. Sonuçlara dikkat edilirse MSE, standart hata ve VIF_i açısından daha iyi olan bir sonuç elde edilmiştir.

Tablo 4.13. Huber M tahminine dayalı robust Liu tahmini sonuçları.

	$\hat{\beta}_{LM}$	$se(\hat{\beta}_{LM})$	L-VIF _i
$\hat{\beta}_1$	11.487688	15.24267	228.3675
$\hat{\beta}_2$	5.362894	2.912166	8.335743
$\hat{\beta}_3$	6.266129	2.590139	6.594141
$\hat{\beta}_4$	-6.857456	7.221843	51.26349
$\hat{\beta}_{12}$	2.755164	3.714028	13.55821
$\hat{\beta}_{13}$	2.776288	12.26655	147.8961
$\hat{\beta}_{14}$	8.900187	2.476935	6.030332
$\hat{\beta}_{23}$	2.736956	2.971631	8.679641
$\hat{\beta}_{24}$	1.254363	1.574747	2.437439
$\hat{\beta}_{34}$	14.358056	5.214057	26.72167
d = 0.18 MSEOLS = 37468.21 MSELIU = 6899.073			

Tablo 4.14’te en küçük kareler yerine GM tahmin edicileri kullanarak elde edilen robust Ridge tahmin sonuçları verilmiştir.

Tablo 4.14. Motor yağı karışımı verisi için GM tahminlerine dayalı robust Liu tahmin edicisi sonuçları.

	Mallows tipi GM tahmini			Schweppe tipi GM tahmini		
	$\hat{\beta}_{LGMM}$	$se(\hat{\beta}_{LGMM})$	L-VIF _i	$\hat{\beta}_{LGMS}$	$se(\hat{\beta}_{LGMS})$	L-VIF _i
$\hat{\beta}_1$	13.535994	10.32196	104.7217	12.0051395	16.92759	281.6452
$\hat{\beta}_2$	3.994686	2.13097	4.46341	3.9909349	3.402106	11.37648
$\hat{\beta}_3$	7.421308	1.876067	3.459463	6.0505850	3.125344	9.600808
$\hat{\beta}_4$	2.905508	5.713712	32.08845	2.2574809	9.885847	96.0594
$\hat{\beta}_{12}$	2.144084	2.638207	6.84116	1.9229599	4.434127	19.32539
$\hat{\beta}_{13}$	2.654260	8.269252	67.21164	0.8615781	13.5215	179.7056
$\hat{\beta}_{14}$	5.464152	2.052741	4.141718	4.4166759	3.489517	11.96859
$\hat{\beta}_{23}$	2.945218	2.037166	4.079104	2.7369497	3.397594	11.34632
$\hat{\beta}_{24}$	1.158034	1.144415	1.287297	2.0609957	1.791455	3.15445
$\hat{\beta}_{34}$	6.525920	4.088551	16.43051	5.5964126	6.999646	48.15753
	d = 0.121 MSEOLS = 37468.21 MSELIU-GM = 4549.438			d = 0.121 MSEOLS = 37468.21 MSELIU-GM = 4549.438		

GM tahminine dayalı robust Ridge tahmininde olduğu gibi GM tahminine dayalı robust Liu tahmini de alışılmış Liu ve M tahminine dayalı robust Liu tahmin edicilerinden daha iyi sonuçlar üretmiştir. Ancak GM tahminine dayalı robust Ridge tahmini her açıdan en iyi performansı sergilemiştir.

Sonuç olarak eğer karma veride kötü koşullulukla birlikte x ve y yönünde sapan değerler varsa, Ridge ve Liu yanlı regresyon tahmin edicileri ile birlikte GM sınırlı etkili regresyon tahmin edicisinin kullanılmasıyla elde edilen robust Ridge ve robust Liu tahmin edicilerinin kullanılması daha tutarlı sonuçların elde edilmesini sağlar.

5. BÖLÜM. KARMA DENEMELERDE PSÜDOBİLEŞEN DÖNÜŞÜMÜ VE ROBUST TAHMİN EDİCİLER

Karma denemelerde (1.1) ve (1.2)'de verilen doğal kısıtlamaların yanı sıra genellikle fiziksel, kimyasal veya ekonomik koşullardan kaynaklanan (1.3)'te verilen alt sınırlar, üst sınırlar ya da hem alt hemde üst sınırlar şeklinde ek kısıtlamalar vardır. Ayrıca (1.4)'te verilen bileşenlerin lineer kombinasyonları üzerinde ek kısıtlamalarda bulunabilir. Bu kısıtlamalar deneysel bölgenin alanını ya da hacmini sınırlar ve şeklini değiştirir. Deneysel bölgenin sınırlandırılması şiddetli çoklu iç ilişki ya da kötü koşulluluğa neden olur. Psudobileşenler olarak adlandırılan değişkenler orijinal bileşenlerden yeni değişkenlere bir dönüşümdür ve deneysel bölgenin hacminin sınırlandırılmasından kaynaklanan kötü koşulluğu azaltmak için uygulanır. Deneysel bölge bir simpleks ise psudobileşen dönüşümü aynı zamanda standart tasarımları deneysel bölgeye uyarlamak için kullanılır. Bu bölümde karma sistemi oluşturan bileşenler üzerinde hem alt hemde üst sınırların her ikisinin de olması durumunda psudobileşen dönüşümü uygulanmış karma veriye robust tahmin edicilerinin etkileri incelenecektir.

5.1. Sınırların Tutarlılığı ve Psudobileşen Dönüşümü

Karma denemelerde bileşenler üzerindeki (1.3) ve (1.4)'de verilen ek kısıtlamaların tutarlı olması istenir. (1.1), (1.2), (1.3) ve (1.4) ile verilen kısıtlanmış bölge içindeki tüm noktaların kümesi R olsun. $L_i \leq x_i \leq U_i$ kısıtlamasının tutarlı olabilmesi için her $i = 1, 2, \dots, q$ için $x_i = L_i$ ve $x_i = U_i$ koşulunu sağlayan $x = (x_1, x_2, \dots, x_q)$ noktasının R 'de olması gerekir ve $c_j \leq A_{1j}x_1 + A_{2j}x_2 + \dots + A_{qj}x_q \leq d_j$ kısıtlamanın tutarlı olabilmesi için $y_j = A_{1j}x_1 + A_{2j}x_2 + \dots + A_{qj}x_q$ olmak üzere her j için $y_j = c_j$ ve $y_j = d_j$ koşulunu sağlayan x noktalarının R 'de olması gerekir. (1.3)'teki kısıtlama (1.1)'dekini sağlamalıdır. (1.4)'teki kısıtlama (1.1), (1.2) ve (1.3)'teki kısıtlamaları sağlamalıdır.

Bileşenler üzerinde alt ve üst sınırlar şeklinde kısıtlamalar olduğunda,

$$\sum_{i=1}^q L_i \leq 1 \quad \text{ve} \quad \sum_{i=1}^q U_i \geq 1 \quad (5.1)$$

eşitsizlikleri ortaya çıkar. Bunlar kısıtlanmış bölgenin boş olmadığını gösterir. Yukarıdaki eşitsizlikler sağlanmıyorsa karma deneyin (1.2)'deki doğal koşulu sağlanmaz.

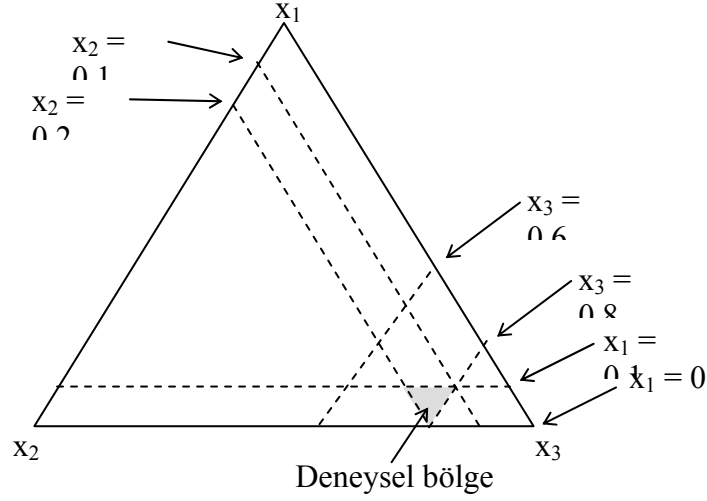
Karma denemelerde alt sınırlardan ve üst sınırlardan kaynaklanan iki tip tutarsızlıkla karşılaşılabilir. Örneğin, McLean ve Anderson (1966)'un verdikleri çok iyi bilinen dört bileşenli parlayarak yanma verisi için bileşenler üzerindeki ek kısıtlamalar,

$$\begin{aligned} 0.40 &\leq x_1 \leq 0.60 \\ 0.10 &\leq x_2 \leq 0.50 \\ 0.10 &\leq x_3 \leq 0.50 \\ 0.03 &\leq x_4 \leq 0.08 \end{aligned} \quad (5.2)$$

şeklindedir. $L_i \leq x_i \leq U_i$ kısıtlamasının tutarlı olabilmesi için her $i = 1, 2, \dots, q$ için $x_i = L_i$ ve $x_i = U_i$ koşulunu sağlayan $x = (x_1, x_2, \dots, x_q)$ noktası olması gerektiği belirtilmişti. Ancak $x_2 = 0,50$ ve $x_3 = 0,50$ iken hiçbir $x = (x_1, x_2, x_3, x_4)$ noktası R bölgesinde olmayacaktır. O halde bu kısıtlamalarda üst sınırlardan kaynaklanan bir tutarsızlık vardır. İkinci tip tutarsızlık için üç bileşenli, bileşenler üzerinde

$$\begin{aligned} 0 &\leq x_1 \leq 0.1 \\ 0.1 &\leq x_2 \leq 0.2 \\ 0.6 &\leq x_3 \leq 0.8 \end{aligned} \quad (5.3)$$

ek kısıtlamaları bulunan karma sistem ele alalım (Cornell, 1990). Şekil 5.1'de karma sistem için deneysel bölge verilmiştir. Dikkat edilirse eğer $x_3 = 0.6$ ise, hiçbir $x = (x_1, x_2, x_3)$ noktası R bölgesinde olmayacaktır ve bu kısıtlamalarda alt sınırlardan kaynaklanan bir tutarsızlık vardır.



Şekil 5.1. Üç bileşenli karma sistem için deneysel bölge

Tutarsızlıkları ve tiplerini belirlemenin en kolay yolu deneysel değişim aralıklarının karşılaştırılmasıdır. Her bileşen için deneysel değişim aralığı,

$$R_i = U_i - L_i \quad (5.4)$$

bileşenler üzerindeki alt sınırların deneysel değişim aralığı,

$$R_L = 1 - \sum_{i=1}^q L_i \quad (5.5)$$

ve bileşenler üzerindeki üst sınırların deneysel değişim aralığı

$$R_U = \sum_{i=1}^q U_i - 1 \quad (5.6)$$

olsun. R_L 'den büyük herhangi bir R_i tutarsız bir üst sınır belirtir. R_U 'dan büyük herhangi bir R_i tutarsız bir alt sınır belirtir (Crosier, 1984). Buna göre (5.2)'de verilen 4 bileşenli karma sistem için, $R_L = 0.37$, $R_U = 0.68$ 'dir. Bileşenlerin deneysel

değişim aralıkları $R_1 = 0.20$, $R_2 = 0.40$, $R_3 = 0.40$ ve $R_4 = 0.05$ 'dir. Buna göre x_2 bileşeni için $R_2 > R_L$ ve x_3 bileşeni için $R_3 > R_L$ tutarsız üst sınırları belirlenir. Benzer şekilde (5.3)'de verilen 3 bileşenli karma sistem için $R_L = 0.3$, $R_U = 0.1$ olup bileşenlerin deneysel değişim aralıkları $R_1 = 0.1$, $R_2 = 0.1$ ve $R_3 = 0.2$ olarak elde edilir. Sadece x_3 bileşeni için $R_3 > R_U$ olup tutarsız bir alt sınır vardır. Şekil 5.1'e dikkat edilirse x_3 bileşeni için alt sınır 0.6 yerine 0.7 olsaydı problem söz konusu olmayacaktı.

İncelenen birinci tip tutarsızlık,

$$U_i + \sum_{j \neq i}^q L_j > 1, \text{ her } i = 1, 2, \dots, q \quad (5.7)$$

durumunda ortaya çıkar ve ikinci tip tutarsızlık,

$$L_i + \sum_{j \neq i}^q U_j < 1, \text{ her } i = 1, 2, \dots, q \quad (5.8)$$

durumunda ortaya çıkar. Kısıtlamaların tutarsızlığı,

$$U_i + \sum_{j \neq i}^q L_j = 1 \quad (5.9)$$

veya

$$L_i + \sum_{i \neq j}^q U_j = 1 \quad (5.10)$$

olacak şekilde alt ve üst sınırların arttırılması veya azaltılmasıyla düzeltilebilir. (5.2)'de verilen dört bileşenli örnekte sistemi oluşturan ikinci ve üçüncü bileşenler için eğer U_2 ve U_3 üst sınırlarını 0,5 yerine daha düşük bir değer almamıza engel olacak çok önemli bir sebep yoksa 0,47 olarak alabiliriz. Ancak örneğin $U_2 = 0,5$

sınırı çok önemli bir sınırlama ise L_1 , L_3 ve L_4 'den herhangi biri veya tümü azaltılır ve $U_2 + L_1 + L_3 + L_4 = 1$ sağlatılmaya çalışılır. Bu değişiklikler çalışmanın amacına uygun olarak araştırmacı tarafından yapılır.

Karma sistemin varsa tutarsızlıkları ortadan kaldırıldıktan sonra orijinal bileşenleri değişkenlerin yeni bir kümesine dönüştüren uygun psudobileşen dönüşümü, q bileşenli sistemde, L_i , x_i değişkenine ait alt sınır göstermek üzere,

$$z_i = \frac{x_i - L_i}{1 - \sum_{j=1}^q L_j} \quad (5.11)$$

şeklinde verilir. (5.11)'deki psudobileşen dönüşümü standart tasarımı, (1.3)'teki kısıtlamalar tarafından indirgenen bir simplekse uyarlamada kullanılır. Bu dönüşüm, Scheffe polinom modeli oluşturmada kötü koşulluluğu azaltmak için ve deneysel bölgenin bir simpleks olmadığı durumlarda kullanılır (Gorman, 1970, Snee ve Rayner, 1982). Psudobileşen dönüşümü bazen kötü koşulluluğu azaltmada etkin değildir ve her zaman bir simpleks tasarımı deneysel bölgeye uyarladığı da doğru değildir. Dönüşümünün ara sıra olan bu başarısızlığı, bileşenlerin değişim aralığı üzerindeki üst sınırların etkisinin ve deneysel bölgenin şeklinin hesaba katılmamasından kaynaklanır (Piepel, 1983).

Karma sistemi oluşturan bileşenler üzerinde sadece alt (tüm üst sınırlar 1'e eşit), sadece üst (tüm alt sınırlar 0'a eşit) veya hem alt hem de üst sınırlar olabilir. Bu çalışmada hem alt hem de üst sınırların birlikte olduğu karma sistemler ele alınacaktır.

5.2. Robust Tahmin Edicilerin Psudobileşen Dönüşümü Yapılmış Karma Veriye Etkilerinin İncelemesi

St. John (1984) karma sistemlerde psudobileşen dönüşümünün koşul sayısı ve VIF_i değerleri üzerinde iyileştirici bir etkisi olduğunu göstermiştir. Tablo 5.1'de McLean ve Anderson (1966)'un verdikleri çok iyi bilinen parlayarak yanma verisi

orijinal ve psüdobileşen dönüşümü yapılaş şekli verilmiştir. (5.2)'de bileşen değişim aralıkları verilmiş ve tutarsızlıklar belirlenerek 2. ve 3. değişkenlerin 0.5 olan üst sınırlarının 0.47 alınması durumunda tutarsızlığın ortadan kalkacağı belirtilmişti. Buna göre dört bileşenli sistem için yeni bileşen değişim aralıkları,

$$\begin{aligned}
 0.40 &\leq x_1 \leq 0.60 \\
 0.10 &\leq x_2 \leq 0.47 \\
 0.10 &\leq x_3 \leq 0.47 \\
 0.03 &\leq x_4 \leq 0.08
 \end{aligned}
 \tag{5.12}$$

şeklinde oluşur.

Tablo 5.1. Parlayarak yanma verisi

Sıra No	Orijinal Bileşenler				Prsüdobileşenler				y
	x_1	x_2	x_3	x_4	z_1	z_2	z_3	z_4	
1	0.4	0.1	0.47	0.03	0.0	0.0	0.5689	0.0	75.00
2	0.4	0.1	0.42	0.08	0.0	0.0	0.4920	0.4	180.00
3	0.6	0.1	0.27	0.03	0.4	0.0	0.2614	0.0	195.00
4	0.6	0.1	0.22	0.08	0.4	0.0	0.1845	0.4	300.00
5	0.4	0.4	0.10	0.03	0.0	0.5689	0.0	0.0	145.00
6	0.4	0.42	0.10	0.08	0.0	0.4920	0.0	0.4	230.00
7	0.6	0.27	0.10	0.03	0.4	0.2614	0.0	0.0	220.00
8	0.6	0.22	0.10	0.08	0.4	0.1845	0.0	0.4	350.00
9	0.5	0.10	0.345	0.055	0.2	0.0	0.3767	0.2	220.00
10	0.5	0.345	0.10	0.055	0.2	0.3767	0.0	0.2	260.00
11	0.4	0.2725	0.2725	0.055	0.0	0.2652	0.2652	0.2	190.00
12	0.6	0.1725	0.1725	0.055	0.4	0.1115	0.1115	0.2	310.00
13	0.5	0.235	0.235	0.03	0.2	0.2076	0.2076	0.0	260.00
14	0.5	0.21	0.21	0.08	0.2	0.1691	0.1691	0.4	410.00
15	0.5	0.2225	0.2225	0.055	0.2	0.1884	0.1884	0.2	425.00

Bu bölümde daha öncede belirtilen değişkenleri standartlaştırma işleminin avantajlarından dolayı orijinal ve psüdobileşen dönüşüm yapılmış verilerin standartlaştırılmış şekilleri kullanılacaktır. Tablo 5.1'de standartlaştırma yapılmamış veriler verilmiştir. Tablo 5.2'de orijinal parlayarak yanma verisi için en küçük kareler sonuçları verilmiştir. Yüksek VIF_i , koşul sayısı değerleri kötü koşulluluğun çok ciddi boyutlarda olduğunu göstermektedir.

Tablo 5.2. Standartlaştırılmış orijinal parlayarak yanma verisi için en küçük kareler sonuçları

Terim	$\hat{\beta}_{LS}$	$se(\hat{\beta}_{LS})$	Özdeğer	VIF _i
$\hat{\beta}_1$	-3055.9987	1752.3977	8.42211044154	855.5329
$\hat{\beta}_2$	-2280.2189	963.7927	1.00312516350	258.7847
$\hat{\beta}_3$	-2353.0399	963.7927	0.32870438064	258.7847
$\hat{\beta}_4$	3262.2054	12137.5161	0.16753685959	41042.23
$\hat{\beta}_{12}$	3800.6082	1731.3680	0.04077821295	835.1224
$\hat{\beta}_{13}$	3698.1418	1731.3680	0.02756489018	835.1224
$\hat{\beta}_{14}$	-760.9615	6852.0166	0.00566704262	13080.01
$\hat{\beta}_{23}$	524.0979	320.3815	0.00410255422	28.59606
$\hat{\beta}_{24}$	-935.6654	3309.3159	0.00039386170	3051.039
$\hat{\beta}_{34}$	-942.7778	3309.3159	0.00001659306	3051.039
$\hat{\sigma}^2 = 3589.456$ $\hat{\sigma} = 59.91207$ $KS = 507568.3$ $MSEOLS = 227199177$				

Tablo 5.3'te psudobileşen dönüşümü ve standartlaştırma işlemi uygulanmış parlayarak yanma verisi için en küçük kareler ile elde edilen sonuçlar verilmiştir. Dikkat edilirse, psudobileşen dönüşümü yapılmadan elde edilen sonuçlara göre çok daha iyi VIF_i, standart hata, koşul sayısı ve MSE değerleri elde edilmiştir.

Tablo 5.3. Psudobileşen dönüşümü ve standartlaştırma işlemi uygulanmış parlayarak yanma verisi için en küçük kareler sonuçları

Terim	$\hat{\beta}_{LS}$	$se(\hat{\beta}_{LS})$	Özdeğer	VIF _i
$\hat{\beta}_1$	-281.3250	338.92530	4.7817341852	32.00216
$\hat{\beta}_2$	218.5372	101.86761	1.6625011883	2.89097
$\hat{\beta}_3$	108.9720	101.86761	1.2508628575	2.89097
$\hat{\beta}_4$	982.5472	2405.71540	0.8753479762	1612.352
$\hat{\beta}_{12}$	454.2743	206.94475	0.5449124629	11.93109
$\hat{\beta}_{13}$	442.0268	206.94475	0.4623332391	11.93109
$\hat{\beta}_{14}$	-108.3392	975.53162	0.2238297272	265.127
$\hat{\beta}_{23}$	129.2178	78.99095	0.1781531714	1.738305
$\hat{\beta}_{24}$	-335.8066	1187.70021	0.0199492368	392.9932
$\hat{\beta}_{34}$	-338.3592	1187.70021	0.0003759552	392.9932
$\hat{\sigma}^2 = 3589.456$ $\hat{\sigma} = 59.91$ $KS = 127.89$ $MSEOLS = 9787908$				

Montgomery ve Voth (1994) karma denemelerin standart tasarımlarında x yönünde etkili noktaların modele etkilerini incelemişlerdir. x yönündeki etkili gözlemlerin teşhisinde Cook uzaklığı ölçüsünü kullanmışlardır. Ancak (4.7) ve (4.8)'de verilen Cook uzaklığı şapka matrisi H 'nin köşegen elemanları h_{ii} 'lere bağlıdır ve h_{ii} değerleri etkilerin maskelenmesinden etkilenen çok dikkatli kullanılması gereken istatistiklerdir. Ayrıca etkili gözlem noktalarında tekrar sayısı artırılarak h_{ii} değerleri daha düşük elde edilmektedir. Ancak bu durum deneyin maliyet, uygulanabilirlik gibi etkileri göz önüne alındığında her zaman geçerli olmayan ve karma deneme oluşturulmadan önce düşünülmesi gereken bir çözümdür. Bu nedenle sapan değerlere daha düşük ağırlıklar vererek etkilerini azaltmaya çalışan robust yöntemlerin uygulanması önerilmiştir. Bu bölümde amaç, daha öncede belirtildiği gibi tezde ele alınan robust tahminlerin ve bu tahminlere dayalı kötü koşulluluğa karşı direçli robust Ridge ve robust Liu tahmin edicilerinin psüdobileşen dönüşümü yapılmış karma verilere etkilerinin incelenmesidir.

Tablo 5.4. Parlayarak yanma verisi için teşhis sonuçları

Sıra	$\hat{e}_i$	h_{ii}	r_i	t_i	Cook	DFFITS	MD	RMD
1	13.0042	0.9357	0.8560	0.8288	1.0664	3.162	12.189	58.584
2	7.1385	0.8970	0.3718	0.3373	0.1208	0.997	11.769	37.912
3	4.4646	0.7242	0.1417	0.1270	0.0053	0.205	9.398	14.830
4	-25.6653	0.6039	-0.6807	-0.6392	0.0707	-0.789	7.695	9.994
5	20.6709	0.9357	1.3607	1.5337	2.6944	5.851	12.1894	58.584
6	1.1632	0.8970	0.0606	0.0542	0.0032	0.160	11.769	37.912
7	-6.7829	0.7242	-0.2153	-0.1934	0.0121	-0.313	9.398	14.829
8	-2.0264	0.6039	-0.0537	-0.0481	0.0004	-0.059	7.695	9.994
9	-37.8292	0.5440	-0.9344	-0.9199	0.1039	-1.004	8.273	7.862
10	-45.3768	0.5439	-1.1209	-1.1586	0.1495	-1.264	8.273	7.862
11	-41.9767	0.8277	-1.6917	-2.3137	1.3821	-5.085	11.697	18.809
12	30.0099	0.4315	0.6627	0.6206	0.0329	0.538	8.025	6.423
13	-31.3567	0.5762	-0.8093	-0.7765	0.0911	-0.916	10.009	8.489
14	19.3899	0.4847	0.4496	0.4105	0.0188	0.396	8.216	7.627
15	95.1728	0.2699	1.8614	3.0047	0.1293	1.835	3.407	4.648

Tablo 5.4'te psudobileşen dönüşümü uygulanmış parlayarak yanma verisi için teşhisler verilmiştir. $\hat{e}_i = y_i - \hat{y}_i$ alışılmış rezidü değerlerinde 15. gözleme ait rezidü değeri diğerlerine göre oldukça büyüktür ve y yönünde sapan değer olabilir. (4.8)'de verilen Cook uzaklığı formu h_{ii} ve r_i standartlaştırılmış rezidü değerlerine

bağlıdır ve dikkat edilirse deneysel noktalara karşılık gelen Cook uzaklığı değerleri arasında 1., 5. ve 11. gözlemlere ait değerler 1'den büyüktür. 15. gözlem büyük r_i değerine sahip olsa bile h_{ii} değeri küçük olduğundan Cook uzaklığı 1'den küçük elde edilmiştir. Benzer şekilde büyük $\hat{e}_i$ ve h_{ii} değerlerine sahip olan 1., 5., 11. ve 15. gözlemlere ait DFFITS değerleri mutlak değerce $2\sqrt{p/n} = 1.6329$ değerinden oldukça büyüktür ancak 15. gözleme ait DFFITS değeri 1.6329 değerini çok fazla aşmamıştır. (4.12)'de verilen, (4.11)'deki C klasik varyans-kovaryans matrisine dayalı alışılmış Mahalanobis uzaklığı ile elde edilen değerlerde ise $\chi^2_{0.95, p-1} = \chi^2_{0.95, 9} = 16.91898$ değerinden büyük bir uzaklık elde edilmemiştir. Buna rağmen robust varyans-kovaryans matrisinin robust tahmini MVE'ye dayalı robust Mahalanobis uzaklığı değerlerinde ise 1., 2., 5., 6. ve 11. gözlemlere ait değerler (4.11)'deki kriteri aşan değerlerdir ve x yönünde sapan değer olabilirler. Robust Mahalanobis uzaklıkları

$$w_i = \frac{1 - h_{ii}}{\sqrt{h_{ii}}} \quad (5.13)$$

ağırlıkları ve

$$\bar{x}_c = \frac{\sum w_i x_i}{\sum w_i} \quad (5.14)$$

merkez değerleri kullanılarak elde edilen robust varyans-kovaryans matrisine dayalı olarak hesaplanmıştır. Bunun nedeni algoritma işleyişi sırasında 3. Bölümde belirtilen kötü koşulluluk ve yuvarlama hatalarından kaynaklanan singüler matris probleminin olmasıdır.

Bir gözlemin kurulan modele etkisini göstermenin en basit yolu ilgili gözlem çıkartılarak tekrar model kurulmasıdır. Tablo (5.4)'teki teşhis sonuçlarına göre 15. gözlem y yönünde sapan değerdir. Buna göre 15. gözlemi modelden çıkartılarak yeniden en küçük kareler tahmin edicisi uygulandığında katsayı tahminleri ve bu

tahminlerin standart hataları gibi istatistiklerde meydana gelebilecek değişiklikler incelenmelidir. Tablo 5.5'te 15. gözlemi çıkarılmış 14 gözlemlili parlayarak yanma verisi için elde edilen sonuçlar verilmiştir.

Tablo 5.5. 14 gözlemlili parlayarak yanma verisi sonuçları

Terim	$\hat{\beta}_{LS}$	$se(\hat{\beta}_{LS})$	Özdeğer	VIF _i
$\hat{\beta}_1$	-131.9422	211.20683	4.5765789352	32.38238
$\hat{\beta}_2$	225.9213	62.09102	1.7225206064	2.798662
$\hat{\beta}_3$	118.3173	62.09102	1.3111154257	2.798662
$\hat{\beta}_4$	1950.7786	1496.79241	0.8970950649	1626.358
$\hat{\beta}_{12}$	352.2804	128.52610	0.5688734798	11.99156
$\hat{\beta}_{13}$	340.3436	128.52610	0.4904577754	11.99156
$\hat{\beta}_{14}$	-542.7674	614.62006	0.2314066936	274.2246
$\hat{\beta}_{23}$	100.5923	45.85081	0.1825222881	1.526114
$\hat{\beta}_{24}$	-823.9751	745.12946	0.0190595560	403.0476
$\hat{\beta}_{34}$	-826.4969	745.12946	0.0003701749	403.0476
$\hat{\sigma}^2 = 1377.549$ $\hat{\sigma} = 37.11535$ KS = 12363.29 MSEOLS = 3816040				

Dikkat edilirse kötü koşulluluk problemi hala çok ciddi olarak devam etmektedir ancak katsayı tahminlerinde mutlak değerce önemli değişiklikler oluşmuştur. Ayrıca standart sapma ve katsayıların standart hatalarında yaklaşık olarak yarı yarıya iyileşme oluşmuştur.

Tablo 5.6'te robust M tahmin edicileri kullanılarak elde edilen katsayı tahminleri ve standart sapma verilmiştir. Standart sapma (1.23)'te verilen MAD ile bir defaya mahsus olarak hesaplanıp sabit kabul edilmiştir. Daha öncede belirtildiği gibi robust M tahmin edicileri eğer veride y yönünde sapan değer varsa o gözlemin etkisini en aza indirecek sonuçlar üretirler. Yani robust M tahmin edicilerinin parlayarak yanma verisi için 14 gözlemlili en küçük kareler ile elde edilen katsayı tahminlerine yakın sonuçlar üretmesi beklenir.

Tablo 5.6. Parlayarak yanma verisi için robust M tahminleri

Terim	$\hat{\beta}_{LS} (15)$	$\hat{\beta}_{LS} (14)$	Huber	Tukey	T	GT
$\hat{\beta}_1$	-281.3250	-131.9422	-197.9521	-151.8642	-182.4999	-183.7196
$\hat{\beta}_2$	218.5372	225.9213	225.0756	234.3074	236.0598	234.3321
$\hat{\beta}_3$	108.9720	118.3173	115.5104	123.9349	124.7618	123.2807
$\hat{\beta}_4$	982.5472	1950.7786	1555.8251	2245.2633	2254.7979	2141.0255
$\hat{\beta}_{12}$	454.2743	352.2804	401.5061	366.7626	382.1784	384.5568
$\hat{\beta}_{13}$	442.0268	340.3436	389.2586	355.9734	372.8894	374.7762
$\hat{\beta}_{14}$	-108.3392	-542.7674	-358.9501	-651.1265	-649.1168	-601.8340
$\hat{\beta}_{23}$	129.2178	100.5923	117.3983	105.5029	106.3850	108.1965
$\hat{\beta}_{24}$	-335.8066	-823.9751	-619.0362	-962.2139	-969.1997	-912.0930
$\hat{\beta}_{34}$	-338.3592	-826.4969	-621.5888	-963.7804	-969.7316	-912.9272
			k = 1.345	c = 3.67	v = 2	p= 2 q=2.5
$\hat{\sigma} = 59.912$ MAD = 30.5353						

Sonuçlara dikkat edilirse gerçekte robust M tahminleri 15. gözlemin katsayı tahminlerine etkilerini düşürmüştür. Tablo 5.7’de robust M tahmin edicileri ile elde edilen katsayıların standart hataları $\hat{A}$ değerleri ve Tablo 5.8’de robust M tahmin edicileri ile elde edilen standartlaştırılmış hata ve ağırlıklar verilmiştir. Robust M tahminlerinin hataları MAD ile standartlaştırılmıştır ve katsayıların standart hataları $\hat{A}$ ile elde edilmiştir. Tablo 5.7 ve 5.8’de elde edilen sonuçlar en küçük kareler ile elde edilen sonuçlardan daha iyidir ve gerçekte 15. gözleme ait ağırlık diğerlerinden daha küçüktür.

Tablo 5.7. Parlayarak yanma verisi katsayıların standart hataları ve $\hat{A}$ değerleri

Terim	$se(\hat{\beta}_{LS}) (15)$	$se(\hat{\beta}_{LS}) (14)$	$se(Huber)$	$se(Tukey)$	$se(T)$	$se(GT)$
$\hat{\beta}_1$	338.92530	211.20683	171.346	162.6146	157.2658	172.7395
$\hat{\beta}_2$	101.86761	62.09102	51.49987	48.87556	47.26791	51.9187
$\hat{\beta}_3$	101.86761	62.09102	51.49987	48.87556	47.26791	51.9187
$\hat{\beta}_4$	2405.71540	1496.79241	1216.226	1154.25	1116.284	1226.117
$\hat{\beta}_{12}$	206.94475	128.52610	104.6223	99.29104	96.02508	105.4732
$\hat{\beta}_{13}$	206.94475	128.52610	104.6223	99.29104	96.02508	105.4732
$\hat{\beta}_{14}$	975.53162	614.62006	493.1867	468.0551	452.6595	497.1976
$\hat{\beta}_{23}$	78.99095	45.85081	39.93442	37.89946	36.65284	40.25919
$\hat{\beta}_{24}$	1187.70021	745.12946	600.45	569.8525	551.1085	605.3332
$\hat{\beta}_{34}$	1187.70021	745.12946	600.45	569.8525	551.1085	605.3332
$\hat{A}^2$	$\hat{\sigma}^2 = 3589,4$	1377.549	917.421	826.3041	772.8393	932.4038
$\hat{A}^2$	$\hat{\sigma} = 59.912$	37.11535	30.2829	28.7455	27.7999	30.5353

Tablo 5.8. Robust M tahmin edicileri için ağırlıklar ve hatalar

	Ağırlıklar				Standartlaştırılmış Hatalar				
	Huber	Tukey	T	GT	EKK	Huber	Tukey	T	GT
1	1.000	0.9968	1.487	2.3759	0.85600	0.3041	0.1471	0.1317	0.1593
2	1.000	0.9999	1.498	2.3974	0.37184	0.1238	0.0241	0.0435	0.0526
3	1.000	0.9992	1.488	2.3878	0.14169	0.0716	0.0728	0.1230	0.1132
4	1.000	0.8897	1.117	1.8642	-0.6806	-0.9034	-0.8744	-0.8269	-0.8477
5	1.000	0.9783	1.413	2.2672	1.36067	0.5551	0.3831	0.3505	0.3827
6	1.000	0.9963	1.488	2.3859	0.06059	-0.0719	-0.1585	-0.1265	-0.1215
7	1.000	0.9890	1.470	2.3554	-0.2152	-0.2968	-0.2728	-0.2000	-0.2177
8	1.000	0.9992	1.499	2.3991	-0.0537	-0.1292	-0.0744	-0.0007	-0.0300
9	1.000	0.9925	1.437	2.2794	-0.934	-0.7223	-0.2256	-0.2952	-0.3638
10	1.000	0.9700	1.333	2.1191	-1.120	-0.9695	-0.4513	-0.4999	-0.5757
11	1.000	0.9768	1.383	2.1857	-1.691	-0.9110	-0.3968	-0.4107	-0.4951
12	1.000	0.6206	0.590	1.1355	0.66269	1.2578	1.6907	1.7547	1.6686
13	1.000	0.9839	1.377	2.2183	-0.809	-0.6340	-0.3298	-0.4209	-0.4525
14	1.000	0.7729	0.863	1.5471	0.4495	0.9808	1.2759	1.2137	1.1740
15	0.356	0.0000	0.142	0.2898	1.861	3.7778	4.4307	4.3608	4.2665

Tablo 5.2’de elde edilen teşhis sonuçlarına göre 1., 2., 5., 6. ve 11. gözlemler x yönünde sapan değer olabilecek gözlemlerdir. 4. Bölümde ele aldığımız hem x hemde y yönündeki sapan değerlere karşı etkili olan GM tahmin edicileri için elde edilen sonuçlar Tablo 5.9’da verilmiştir.

Tablo 5.9. Parlayarak yanma verisi için GM tahmin edicileri

Sıra	Mallows tipi GM tahmini				Schweppe tipi GM tahmini			
	w.hb	wtx	Hata	$\hat{\beta}_{GMM}$	w.hb	wtx	Hata	$\hat{\beta}_{GMS}$
1	1.000	0.5374	0.3041	-197.9521	1.0000	0.1307	0.1966	33.0533
2	1.000	0.6680	0.1238	225.0756	1.0000	0.1624	-0.0692	230.8425
3	1.000	1.0000	0.0716	115.5104	1.0000	0.2597	0.2995	121.2773
4	1.000	1.0000	-0.9034	1555.8251	1.0000	0.3163	-0.7610	471.7584
5	1.000	0.5374	0.5551	401.5061	1.0000	0.1307	0.4477	239.1248
6	1.000	0.6680	-0.0719	389.2586	1.0000	0.1624	-0.2649	226.8773
7	1.000	1.0000	-0.2968	-358.9501	1.0000	0.2597	-0.0689	39.8390
8	1.000	1.0000	-0.1292	117.3983	1.0000	0.3163	0.0131	203.1427
9	1.000	1.0000	-0.7223	-619.0362	1.0000	0.3566	-0.2919	-79.3402
10	1.000	1.0000	-0.9695	-621.5888	1.0000	0.3566	-0.5391	-81.8929
11	1.000	0.9484	-0.9110		0.0839	0.2306	-3.6979	
12	1.000	1.0000	1.2578		1.0000	0.3946	0.5172	
13	1.000	1.0000	-0.6340		1.0000	0.3432	-0.8749	
14	1.000	1.0000	0.9808		1.0000	0.3621	1.0820	
15	0.356	1.0000	3.7778		0.1974	0.4638	3.1603	

Mallows GM tahmini için (4.19)'da verilen ağırlık formu ve Schweppe GM tahmini için (4.21)'de verilen ağırlık formu kullanılmıştır. Özellikle Schweppe tahmini diğer tahminlerden oldukça farklı elde edilmiştir.

Psüdobileşen dönüşümü yapılmış parlayarak yanma verisine alışılmış ve robust Ridge tahmin edicileri uygulandığında elde edilen sonuçlar Tablo 5.10'da verilmiştir. Ayrıca orijinal veri sadece standartlaştırıldıktan sonra elde edilen en küçük kareler ve Huber M tahminine dayalı robust Ridge sonuçları da Tablo 5.10'a eklenmiştir.

Tablo 5.10. Parlayarak yanma verisi için Ridge tahmin edicisi sonuçları

Terim	Standartlaştırılmış orijinal veri			Psüdobileşen dönüşümü uygulanmış veri		
	$\hat{\beta}_{LS}$	$\hat{\beta}_R$ k=0.0190	$\hat{\beta}_{RM}$ k=0.0091	$\hat{\beta}_{LS}$	$\hat{\beta}_R$ k=0.0075	$\hat{\beta}_{RM}$ k=0.0035
$\hat{\beta}_1$	-3044.0306	358.73703	83.02694	-281.3250	-129.22198	-101.22440
$\hat{\beta}_2$	-2276.0401	-238.46644	-664.65576	218.5372	216.01184	222.03674
$\hat{\beta}_3$	-2349.0884	-314.04544	-740.73694	108.9720	108.62135	113.50268
$\hat{\beta}_4$	3343.9516	197.47643	221.02499	982.5472	178.35651	274.70734
$\hat{\beta}_{12}$	3791.4314	196.96319	650.17613	454.2743	363.01133	338.43799
$\hat{\beta}_{13}$	3689.1398	116.22519	553.47727	442.0268	350.17265	325.90873
$\hat{\beta}_{14}$	-808.6970	430.90933	421.35753	-108.3392	186.98786	144.86187
$\hat{\beta}_{23}$	527.0704	350.79419	505.41051	129.2178	129.71959	108.50619
$\hat{\beta}_{24}$	-958.0988	-43.50899	-26.38045	-335.8066	64.78187	13.98360
$\hat{\beta}_{34}$	-965.1469	-65.73362	-35.42562	-338.3592	60.98068	10.83649
MSE	227199177	566627.4	593501.5	9787908	204381.4	24363.14
Max VIF_i	42658.93	8.708139	54.88231	392.9932	15.99934	21.8152
Max se(β)	12369.2784	176.7246	219.8026	2405.715	150.5086	141.4697

Ridge tahmin edicisi için elde edilen sonuçlara dikkat edilirse en küçük MSE değeri psüdobileşen dönüşümü uygulanmış veri için Huber M tahminine dayalı robust Ridge tahmini için elde edilmiştir. VIF_i değerleri gerek alışılmış Ridge gerekse robust Ridge tahminleri için 100'ün altında elde edilmiştir ve buda kötü koşulluluk probleminin önemli derecede düzeltildiği anlamına gelmektedir. Benzer şekilde katsayıların standart hataları da alışılmış ve robust Ridge tahmin edicileri tarafından en küçük karelere göre oldukça düzeltilmişlerdir. Özellikle psüdobileşen dönüşümü yapılmış veri için robust Ridge tahmini en küçük standart hata değerini üretmiştir.

Tablo 5.11'de parlayarak yanma verisi için Liu tahmin edicileri ile elde edilen sonuçlar verilmiştir. MSE değerleri açısından Liu tahminleri çok iyi bir performans göstermeseler de psüdobileşen dönüşümü yapıldıktan sonra değerlerde dikkate değer bir düşüş olmuştur. Benzer düzelme katsayıların standart hatalarında ve VIF_i'lerde de görülmektedir.

Tablo 5.11. Parlayarak yanma verisi için robust Liu tahmin edicisi sonuçları

Terim	Standartlaştırılmış orijinal veri			Psüdobileşen dönüşümü uygulanmış veri		
	$\hat{\beta}_{LS}$	$\hat{\beta}_L$ d=0.207	$\hat{\beta}_{LM}$ d=0.3373	$\hat{\beta}_{LS}$	$\hat{\beta}_L$ d=0.134	$\hat{\beta}_{LM}$ d=0.265
$\hat{\beta}_1$	-3044.0306	-524.21954	-767.8350	-281.3250	92.79071	57.705433
$\hat{\beta}_2$	-2276.0401	-411.59631	-621.6783	218.5372	152.06994	162.603652
$\hat{\beta}_3$	-2349.0884	-452.29637	-667.6566	108.9720	101.21513	102.867707
$\hat{\beta}_4$	3343.9516	771.70900	2177.2339	982.5472	248.01620	510.324870
$\hat{\beta}_{12}$	3791.4314	873.41053	1204.1364	454.2743	186.40108	209.589260
$\hat{\beta}_{13}$	3689.1398	827.69482	1149.0951	442.0268	168.99861	192.966569
$\hat{\beta}_{14}$	-808.6970	-38.00207	-751.5287	-108.3392	89.81140	-7.012378
$\hat{\beta}_{23}$	527.0704	168.70293	207.3475	129.2178	116.07550	106.841125
$\hat{\beta}_{24}$	-958.0988	-122.17302	-522.8118	-335.8066	42.27675	-91.036696
$\hat{\beta}_{34}$	-965.1469	-145.85485	-543.7709	-338.3592	23.49783	-107.36105
MSE	227199177	48217329	64304950	9787908	1329118	2604632
Max VIF_i	42658.93	560.7821	4670.106	392.9932	29.1864	113.5654
Max se(β)	12369.2784	2512.687	2069.889	2405.715	323.6717	322.7801

Huber M tahminine dayalı robust Ridge ve robust Liu tahminleri için yapılan benzer inceleme GM tahminine dayalı olarak yapıldığında x yönünde sapan değer olabilecek gözlemlerin katsayı tahminleri üzerindeki etkileri ortaya çıkacaktır. Tablo 5.12’de Schweppe tipi GM tahminine dayalı robust Ridge ve robust Liu tahmin edicileri için elde edilen sonuçlar verilmiştir. Katsayıların standart hataları Huber ρ için elde edilen $\hat{A}^2$ değeri kullanılarak elde edilmiştir. Elde edilen sonuçlara dikkat edilirse özellikle VIF_i ’lerde dikkate değer bir düzelme olmuştur. Ayrıca katsayıların standart hataları da oldukça azalmıştır. GM tahminine dayalı Liu tahmini psüdobileşen dönüşümü ile daha iyi bir performans gösterirken Ridge tahmini psüdobileşen dönüşüm ile daha yüksek MSE değeri üretirken katsayıların standart hataları ile VIF_i ’lerde çok dikkate değer bir düzelme olmamıştır.

Sonuç olarak x ve y yönünde sapan değerlerin olduğu verilerde y yönündeki sapan değerlere karşı etkili olan M tahminleri yerine hem x hemde y yönündeki sapan değerlere karşı etkili olan GM tahminini kullanmak kötü koşulluluk

probleminin hemen hemen ortadan kaldırılmasını ve diğer istatistiklerin daha iyi elde edilmesini sağlayacaktır.

Tablo 5.12. Parlayarak yanma verisi için GM tahminine dayalı Ridge ve Liu

Terim	Standartlaştırılmış orijinal veri			Psüdobileşen dönüşümü uygulanmış veri		
	$\hat{\beta}_{LS}$	$\hat{\beta}_{RGMS}$ k=0.0134	$\hat{\beta}_{LGMS}$ d=0.0634	$\hat{\beta}_{LS}$	$\hat{\beta}_{RGMS}$ k=0.01522	$\hat{\beta}_{LGMS}$ d=0.02
$\hat{\beta}_1$	-3044.0306	257.03034	8.813695	-281.3250	136.98860	109.63857
$\hat{\beta}_2$	-2276.0401	-54.08314	-26.005257	218.5372	208.13053	90.97003
$\hat{\beta}_3$	-2349.0884	-118.70753	-40.772219	108.9720	120.92783	74.67274
$\hat{\beta}_4$	3343.9516	180.92895	139.663902	982.5472	145.35356	110.85725
$\hat{\beta}_{12}$	3791.4314	131.34895	206.858952	454.2743	183.87778	102.79597
$\hat{\beta}_{13}$	3689.1398	62.09933	190.163136	442.0268	165.10662	91.18759
$\hat{\beta}_{14}$	-808.6970	341.06983	112.696975	-108.3392	142.47740	94.28381
$\hat{\beta}_{23}$	527.0704	294.29578	128.481436	129.2178	203.37176	95.17784
$\hat{\beta}_{24}$	-958.0988	12.97519	60.809004	-335.8066	98.94214	70.69020
$\hat{\beta}_{34}$	-965.1469	-25.05358	50.862421	-338.3592	84.83324	60.34006
MSE	227199177	2333.674	14434408	9787908	21560.39	228549
Max VIF_i	42658.93	10.49642	482.8303	392.9932	9.75909	1.98254
Max se(β)	12369.2784	108.5138	665.5505	2405.715	94.6212	42.6477

5.3. Motor Yağı Karışımı Verisi

Bu bölümde 3. ve 4. bölümlerde M ve GM tahminlerinin etkilerinin incelenmesi amacıyla kullanılan motor yağı karışımı verisini ele alarak psüdobileşen dönüşümü etkilerini incelenecektir. 4 bileşenli veride bileşenler üzerindeki ek kısıtlamalar (3.40)'da verilmiştir. Veriler Tablo 3.8'de verilmiştir ve bu bölümde, tezin amacı doğrultusunda, 5. gözlemi değiştirilerek sapan değer olarak ayarlanmış y değerleri kullanılacaktır. Ek kısıtlamalar için bileşen değişim aralıkları $R_1 = 0.11$, $R_2 = 0.30$, $R_3 = 0.33$ ve $R_4 = 0.15$ 'tir ve $R_L = 0.56$, $R_U = 0.33$ 'tür. R_L 'den büyük herhangi bir R_i ve R_U 'dan büyük herhangi bir R_i olmadığından tutarsızlık söz konusu değildir. Tablo 5. 13'te psüdobileşen dönüşümü uygulanmış motor yağı karışımı verisi verilmiştir.

Tablo 5.13. Psüdobileşen dönüşüm uygulanmış motor yağı karışımı verisi

Sıra	z_1	z_2	z_3	z_4
1	0.2430735	0.00000000	0.3039247	0.3314312
2	0.3342261	0.35171185	0.00000000	0.3314312
3	0.00000000	0.26964575	0.3039247	0.00000000
4	0.00000000	0.09378983	0.3039247	0.3314312
5	0.3342261	0.14068474	0.3039247	0.00000000
6	0.00000000	0.35171185	0.2394558	0.00000000
7	0.00000000	0.35171185	0.1013082	0.3314312
8	0.3342261	0.00000000	0.2762952	0.3314312
9	0.3342261	0.35171185	0.1381476	0.00000000
10	0.3342261	0.00000000	0.3039247	0.2651450
11	0.00000000	0.26671482	0.2371533	0.1657156
12	0.3342261	0.16882169	0.2044584	0.1856015
13	0.1671131	0.35171185	0.1197279	0.1657156
14	0.1823051	0.10082406	0.3039247	0.1856015
15	0.1671131	0.27843855	0.2463632	0.00000000
16	0.1823051	0.15944270	0.1970906	0.3314312
17	0.1914204	0.19109677	0.2274830	0.1922301
18	0.3342261	0.17585592	0.1381476	0.3314312

Psüdobileşen dönüşümü uygulanmış motor yağı karışımı verisi için en küçük kareler ile elde edilen sonuçlar Tablo 5.14’te verilmiştir. Dikkat edilirse psüdobileşen dönüşümünden sonra hala koşul sayısı yüksek seviyelerde ve kötü koşulluluk problemi ciddi boyutlardadır.

Tablo 5.14. Psüdobileşen dönüşümü ve standartlaştırma işlemi uygulanmış motor yağı karışımı verisi için en küçük kareler sonuçları

Terim	$\hat{\beta}_{LS}$	$se(\hat{\beta}_{LS})$	Özdeğer	VIF _i
$\hat{\beta}_1$	6,15183	43,09344	6,34550	761,53360
$\hat{\beta}_2$	23,29374	9,29810	1,49396	35,45315
$\hat{\beta}_3$	1,25333	10,06684	1,02750	41,55782
$\hat{\beta}_4$	-25,70834	30,03334	0,85254	369,89120
$\hat{\beta}_{12}$	1,87997	17,05117	0,13235	119,22720
$\hat{\beta}_{13}$	-1,48698	23,92916	0,11006	234,81260
$\hat{\beta}_{14}$	19,91582	11,75335	0,02243	56,64868
$\hat{\beta}_{23}$	-0,00985	9,26471	0,01285	35,19902
$\hat{\beta}_{24}$	4,27542	12,64476	0,00198	65,56734
$\hat{\beta}_{34}$	33,96643	18,25380	0,00083	136,63850
$\hat{\sigma} = 1,562 \quad \hat{\sigma}^2 = 2,439 \quad KS = 7671,4 \quad MSEOLS = 4527,256$				

Tablo 5.15'te motor yağı karışımı verisi için elde edilen teşhis sonuçları verilmiştir.

Tablo 5.15. Motor yağı karışımı verisi için teşhis sonuçları

Sıra	$\hat{e}_i$	h_{ii}	r_i	t_i	Cook	DFFITS	MD	RMD
1	-0,56168	0,44192	-0,48148	-0,45705	0,01836	-0,40672	6,598656	6,116465
2	-0,07149	0,92680	-0,16920	-0,15856	0,03625	-0,56420	14,81278	90,14792
3	0,94227	0,52785	0,87815	0,86414	0,08621	0,91369	8,038952	10,0929
4	-0,36960	0,81354	-0,54813	-0,52263	0,13109	-1,09168	12,90192	34,27278
5	-2,05036	0,78439	-2,82768	-114,9162	2,90889	-219,1871	12,4593	30,01805
6	-1,10844	0,52455	-1,02942	-1,03384	0,11691	-1,08591	7,987227	8,780057
7	1,31794	0,86955	2,33668	3,87916	3,63941	10,01507	13,91759	54,99021
8	-0,24972	0,42992	-0,21180	-0,19868	0,00338	-0,17254	6,386608	5,39551
9	0,77581	0,82650	1,19272	1,23044	0,67767	2,68553	13,09467	39,21376
10	1,08554	0,45285	0,93978	0,93203	0,07310	0,84791	6,746075	6,025192
11	-0,82767	0,41117	-0,69071	-0,66627	0,03331	-0,55676	6,07538	6,743063
12	0,99662	0,42878	0,84443	0,82763	0,05352	0,71705	6,362439	6,425888
13	-1,82815	0,47613	-1,61746	-1,84432	0,23778	-1,75827	7,281065	9,388205
14	1,03038	0,38954	0,84450	0,82771	0,04551	0,66119	5,777575	5,353137
15	1,67547	0,39835	1,38324	1,48340	0,12668	1,20704	5,799508	6,159707
16	-0,38987	0,46578	-0,34158	-0,32187	0,01017	-0,30055	6,980206	7,326635
17	-0,03770	0,27460	-0,02835	-0,02652	0,00003	-0,01631	3,797246	2,725358
18	-0,32935	0,55778	-0,31715	-0,29855	0,01269	-0,33530	8,584038	10,56803

Teşhis sonuçlarına dikkat edilirse beklendiği gibi 5. gözlem rezidü sonuçlarına y yönünde sapan değerdir ve 5. ve 7. gözlemlere ait Cook uzaklıkları 1'den büyüktür. DFITS değerlerinde ise 5., 7., 9. ve 13. gözlemler 1.490 olan kontrol değerinden büyüktür. MD değerlerinde ise beklenenin aksine büyük bir değer bulunmamakta ancak RMD değerlerinde 2., 4., 5., 7. ve 9. gözlemler kontrol değeri olan 16.91898'den büyük elde edilmiştir. Parlayarak yanma verisinde olduğu gibi burada da RMD (5.13) ve (5.14)'te verilen merkez ve ağırlıklar kullanılarak hesaplanmıştır. Sonuç olarak hem y hemde x yönünde problemli olan gözlemler bulunmaktadır.

Psüdobileşen dönüşümü yapılmış motor yağı karışımı verisi için robust M tahmin edicileri ile elde edilen katsayı tahminleri Tablo 5.16'da, katsayıların standart hataları ile $\hat{A}$ değerleri Tablo 5.17'de ve hatalar ile ağırlıklar Tablo 5.18'de verilmiştir.

Tablo 5.16. Motor yağı karışımı verisi için robust M tahmin edicileri

Terim	$\hat{\beta}_{LS}$	Huber	Tukey	T	GT
$\hat{\beta}_1$	6,15183	15.54779	34.56264	32.22117	32.34068
$\hat{\beta}_2$	23,29374	19.06279	10.50043	10.95868	10.94419
$\hat{\beta}_3$	1,25333	6.53577	17.22606	17.15188	17.14108
$\hat{\beta}_4$	-25,70834	-13.38082	11.56697	11.34179	11.37159
$\hat{\beta}_{12}$	1,87997	-0.99616	-6.81663	-5.97382	-6.01882
$\hat{\beta}_{13}$	-1,48698	-3.66043	-8.05885	-6.86564	-6.92427
$\hat{\beta}_{14}$	19,91582	12.64011	-2.08404	-1.39908	-1.43571
$\hat{\beta}_{23}$	-0,00985	-0.34116	-1.01164	-1.31238	-1.29410
$\hat{\beta}_{24}$	4,27542	2.63404	-0.68769	-0.78839	-0.79631
$\hat{\beta}_{34}$	33,96643	22.48972	-0.73624	-0.52819	-0.54342
		k = 1,345	c = 3,67	v = 2	p = 1.2,q = 2
$\hat{\sigma} = 1.56157$ MAD = 0.68798					

Tablo 5.17. Katsayıların standart hataları ve $\hat{A}$ değerleri

Terim	β_{EKK}	Huber	Tukey	T	GT
$\hat{\beta}_1$	43,09344	23.27547	4.539282	2.36828	1.375045
$\hat{\beta}_2$	9,29810	5.022053	0.9794225	0.5109941	0.2966879
$\hat{\beta}_3$	10,06684	5.437262	1.060398	0.5532416	0.3212172
$\hat{\beta}_4$	30,03334	16.22149	3.163586	1.650538	0.9583174
$\hat{\beta}_{12}$	17,05117	9.209616	1.796099	0.9370788	0.5440766
$\hat{\beta}_{13}$	23,92916	12.92453	2.520597	1.315071	0.7635424
$\hat{\beta}_{14}$	11,75335	6.348174	1.238048	0.6459269	0.3750311
$\hat{\beta}_{23}$	9,26471	5.004021	0.9759059	0.5091594	0.2956226
$\hat{\beta}_{24}$	12,64476	6.829638	1.331946	0.6949159	0.4034746
$\hat{\beta}_{34}$	18,25380	9.859171	1.922778	1.003171	0.5824504
$\hat{A}^2$	$\hat{\sigma}^2=2.439$	0,5061	0,000734	0,0000543	0,0006165
$\hat{A}$	$\hat{\sigma}=1.562$	0,71139	0,0271	0,00737	0,02483

Elde edilen sonuçlara dikkat edilirse sapan değer olarak değiştirilen 5. gözlemin katsayı tahminleri üzerindeki etkileri açıkça görülmektedir. Robust M tahmin edicileri ile elde edilen katsayıların standart hatalarındaki iyileşmede önemli derecededir. Tablo 5.18’de robust M tahmin edicileri ile elde edilen hatalar ve katsayı hesaplamalarında kullanılan ağırlıklar verilmiştir. 5. gözleme ait hata

değerlerinin büyüklüğü ve verilen ağırlıkların çok küçük olması beklenen bir durumdur ve 5. gözlemin etkilerinin çok fazla olduğunu işaretidir.

Tablo 5.18. Robust M tahmin edicileri için hatalar ve ağırlıklar

#	Hatalar					Ağırlıklar			
	Klasik	Huber	Tukey	T	GT	Huber	Tukey	T	GT
1	-0,48148	-0.3569	0.03474	0.07669	0.20017	1.00000	0.99982	1.49560	2.74178
2	-0,16920	-0.0398	0.02141	0.17764	0.17335	1.00000	0.99993	1.47670	2.76411
3	0,87815	0.6081	-0.03021	-0.39143	-0.23864	1.00000	0.99986	1.39326	2.70496
4	-0,54813	-0.2490	-0.02004	0.23950	0.11206	1.00000	0.99994	1.45818	2.80399
5	-2,82768	-4.4274	-9.32065	-1653.58	-770.2814	0.303789	0.00000	0.00000	0.00001
6	-1,02942	-0.7238	0.00985	-0.02703	-0.04651	1.00000	0.99999	1.49945	2.82823
7	2,33668	0.8634	-0.00350	-0.01814	-0.03823	1.00000	1.00000	1.49975	2.82989
8	-0,21180	-0.1646	-0.00254	0.40665	0.14055	1.00000	1.00000	1.38545	2.78745
9	1,19272	0.5090	0.00025	-0.35873	-0.25729	1.00000	1.00000	1.40932	2.68520
10	0,93978	0.7067	-0.01617	-0.30368	-0.13720	1.00000	0.99996	1.43388	2.78957
11	-0,69071	-0.5269	0.04849	11.87639	5.51079	1.00000	0.99965	0.02097	0.10770
12	0,84443	0.6589	0.01565	10.42501	4.72090	1.00000	0.99996	0.02710	0.14476
13	-1,61746	-1.2150	-0.04770	-19.48262	-8.84780	1.00000	0.99966	0.00786	0.04278
14	0,84450	0.6706	-0.01605	-3.91567	-1.64743	1.00000	0.99996	0.17309	0.86867
15	1,38324	1.1053	0.01883	1.47784	0.74018	1.00000	0.99995	0.71702	1.94521
16	-0,34158	-0.2533	0.00727	-0.79763	-0.32893	1.00000	0.99999	1.13800	2.59901
17	-0,02835	-0.0213	0.01022	0.42728	0.29835	1.00000	0.99998	1.37453	2.63767
18	-0,31715	-0.2261	-0.03050	-0.11659	-0.22431	1.00000	0.99986	1.48987	2.71931

Motor yağı karışımı verisi için 5. gözlemin y yönünde sapan değer olduğunu biliyoruz ve teşhis sonuçlarında x yönünde de sapan olabilecek gözlemlerin olduğunu biliyoruz. Şimdi hem x hemde y yönünde sapan değerlere karşı etkili olan GM tahmin edicileri kullanıldığında elde edilen sonuçlar Tablo 5.19’da verilmiştir. Dikkat edilirse gerek Mallows gerekse Schweppe tipi GM tahminlerinde 5. gözleme ait Huber ağırlıkları çok küçüktür ve x yönündeki wtx ağırlıklarında 2., 4., 5., 7. ve 9. gözlemlere ait ağırlıklar diğerlerinde daha küçüktür. Bu sonuç Tablo 5.15’teki teşhis sonuçlarını destekleyecek niteliktedir ve bu beş gözlem dikkat edilmesi gereken gözlemlerdir.

Tablo 5.19. Motor yağı karışımı verisi için GM tahmin edicileri

Sıra	Mallovs tipi GM tahmini				Schweppe tipi GM tahmini			
	w(y)	u(x)	Hata	$\hat{\beta}_{GMM}$	w(y)	u(x)	Hata	$\hat{\beta}_{GMS}$
1	1.000	1.000	-0.28045	20.28727	1.000	0.380	-0.04188	31.09222
2	1.000	0.463	-0.02785	16.92862	1.000	0.113	0.00944	12.06321
3	1.000	1.000	0.48343	9.20032	1.000	0.313	0.09466	15.27489
4	1.000	0.751	-0.20435	-7.16261	1.000	0.183	-0.06485	7.01351
5	0.201	0.805	-5.38327	-2.44692	0.031	0.196	-8.36341	-5.75434
6	1.000	1.000	-0.58054	-4.75676	1.000	0.310	-0.13368	-7.25614
7	1.000	0.599	0.69406	8.97011	1.000	0.146	0.16609	0.60336
8	1.000	1.000	-0.13299	-0.50827	1.000	0.379	-0.03426	-0.88926
9	1.000	0.698	0.40963	1.80611	1.000	0.169	0.09978	-0.0814
10	1.000	1.000	0.56554	16.70067	1.000	0.378	0.12526	3.50295
11	1.000	1.000	-0.41451		1.000	0.366	-0.06408	
12	1.000	1.000	0.53329		1.000	0.368	0.14149	
13	1.000	1.000	-0.98700		1.000	0.327	-0.27605	
14	1.000	1.000	0.53648		1.000	0.387	0.11827	
15	1.000	1.000	0.89309		1.000	0.372	0.23138	
16	1.000	1.000	-0.20244		1.000	0.348	-0.04371	
17	1.000	1.000	-0.01518		1.000	0.437	0.00404	
18	1.000	1.000	-0.18792		1.000	0.313	-0.06878	

Daha öncede belirtildiği gibi karma veriler bileşenler üzerindeki ek kısıtlamalar ve katsayı matrislerinde bileşenlerin çarpımlarını içerdiğinden kötü koşulluluk problemi çoğunlukla çok ciddi boyutlardadır ve motor yağı karışımı verisinde de bu problemin olduğu gösterilmiştir. Bilindiği gibi kötü koşulluluk problemine karşı en çok kullanılan yöntemler Ridge ve Liu yanlı tahmin edicileridir. Tablo 5.20’de alışılmış Ridge ve alışılmış Liu tahminleri için elde edilen sonuçlar verilmiştir.

Tablo 5.20. Motor yağı karışımı verisi için alışılmış Ridge ve Liu tahminleri

Terim	$\hat{\beta}_R$	$se(\hat{\beta}_R)$	R-VIF _i	$\hat{\beta}_L$	$se(\hat{\beta}_L)$	L-VIF _i
$\hat{\beta}_1$	4.0918496	3.64783	5.456773	5.673555	14.68378	88.41832
$\hat{\beta}_2$	16.8099032	5.910895	14.32759	11.327657	3.302501	4.472523
$\hat{\beta}_3$	2.2123290	6.106414	15.29111	4.391176	3.556041	5.185614
$\hat{\beta}_4$	2.7232660	4.694164	9.036146	-4.747997	10.257	43.14267
$\hat{\beta}_{12}$	3.9843880	3.369259	4.65517	3.233529	5.874397	14.1512
$\hat{\beta}_{13}$	0.3791923	3.771248	5.832261	2.492044	8.192296	27.52187
$\hat{\beta}_{14}$	14.1546001	3.54529	5.154306	10.861604	4.110757	6.929634
$\hat{\beta}_{23}$	3.9597927	5.219173	11.17043	3.121841	3.288051	4.433469
$\hat{\beta}_{24}$	-4.5693192	3.722555	5.682624	3.611355	4.40171	7.945288
$\hat{\beta}_{34}$	15.9092673	4.291204	7.551359	15.646887	6.279592	16.17073
k = 0,00996 MSEOLS = 4527,256 MSERIDGE = 213,1124			d = 0,34 MSEOLS = 4527,256 MSELIU = 1570,607			

Tablo 5.20'deki sonuçlara dikkat edilirse en küçük karelere göre çok daha iyi standart hata, VIF_i ve MSE değerleri vermişlerdir. Tablo 5.21'de robust M tahminine dayalı Ridge ve Liu ile elde edilen sonuçlar verilmiştir.

Tablo 5.21. Motor yağı karışımı verisi için robust Ridge ve Liu tahminleri

Terim	$\hat{\beta}_{RM}$	$se(\hat{\beta}_{RM})$	R-VIF _i	$\hat{\beta}_{LM}$	$se(\hat{\beta}_{LM})$	L-VIF _i
$\hat{\beta}_1$	8.153707	0.01934952	0.510087	13.260306	0.1808486	44.55889
$\hat{\beta}_2$	6.768967	0.04483095	2.738167	6.373177	0.04182628	2.38343
$\hat{\beta}_3$	10.415851	0.04402734	2.640881	9.416847	0.04487007	2.742947
$\hat{\beta}_4$	5.775372	0.02346278	0.750003	6.872914	0.1265284	21.81122
$\hat{\beta}_{12}$	4.138546	0.03889175	2.06072	1.680146	0.07299114	7.258455
$\hat{\beta}_{13}$	7.703125	0.03943609	2.118808	2.986618	0.1012665	13.97125
$\hat{\beta}_{14}$	5.881359	0.04457493	2.706981	3.486287	0.0516071	3.628464
$\hat{\beta}_{23}$	4.800140	0.03857518	2.027309	3.648446	0.04170514	2.369644
$\hat{\beta}_{24}$	2.349106	0.03837121	2.005926	2.211256	0.05512359	4.139795
$\hat{\beta}_{34}$	4.930155	0.03843115	2.012198	3.900648	0.07789729	8.267013
k = 0,05 MSEOLS = 4527,256 MSERIDGE = 72.54679			d = 0,241 MSEOLS = 4527,256 MSELIU = 1101.752			

Robust M tahmin edicilerinin en küçük karelere göre katsayıların standart hatalarının düşürülmesinde gösterdikleri başarı robust Ridge ve robust Liu

tahminlerine de yansımıştır. Ayrıca MSE ve VIF_i değerlerindeki azalma dikkat çekicidir. Tablo 5.22 ve Tablo 5.23'te sırasıyla Mallows ve Schweppe tipi GM tahminlerine dayalı robust Ridge ve Robust Liu tahmin edicileri ile elde edilen sonuçlar verilmiştir.

Tablo 5.22. Mallows tipi GM tahminine dayalı robust Ridge ve Liu tahminleri

Terim	$\hat{\beta}_{RGMM}$	$se(\hat{\beta}_{RGMM})$	R- VIF_i	$\hat{\beta}_{LGMM}$	$se(\hat{\beta}_{LGMM})$	L- VIF_i
$\hat{\beta}_1$	6.673146	1.376519	3.743936	10.910472	3.844892	29.21003
$\hat{\beta}_2$	12.373626	2.537231	12.7199	6.283746	0.9302253	1.709779
$\hat{\beta}_3$	7.415275	2.557505	12.92399	8.415504	1.076856	2.291285
$\hat{\beta}_4$	4.973394	1.784043	6.288891	5.638886	2.703833	14.44519
$\hat{\beta}_{12}$	3.507502	1.596628	5.036989	2.098555	1.559467	4.805251
$\hat{\beta}_{13}$	3.399484	1.722467	5.862268	3.701876	2.174504	9.342954
$\hat{\beta}_{14}$	9.853543	1.678098	5.564142	4.496350	1.174135	2.723953
$\hat{\beta}_{23}$	3.498628	2.186649	9.447604	3.981786	0.9605398	1.823033
$\hat{\beta}_{24}$	-1.500042	1.697303	5.692228	2.356240	1.18719	2.784865
$\hat{\beta}_{34}$	9.605703	1.858381	6.823909	5.223936	1.686612	5.620749
	k = 0,01224 MSEOLS = 4527,256 MSERIDGE = 38.48184			d = 0,191 MSEOLS = 4527,256 MSELIU = 873.0547		

Tablo 5.23. Schweppe tipi GM tahminine dayalı robust Ridge ve Liu tahminleri

Terim	$\hat{\beta}_{RGMS}$	$se(\hat{\beta}_{RGMS})$	R- VIF_i	$\hat{\beta}_{LGMS}$	$se(\hat{\beta}_{LGMS})$	L- VIF_i
$\hat{\beta}_1$	7.331187	0.6948653	0.9540364	10.910472	3.844892	29.21003
$\hat{\beta}_2$	6.226782	1.118088	2.470106	6.283746	0.9302253	1.709779
$\hat{\beta}_3$	8.664814	1.039661	2.135732	8.415504	1.076856	2.291285
$\hat{\beta}_4$	5.715516	0.7503253	1.112405	5.638886	2.703833	14.44519
$\hat{\beta}_{12}$	3.868494	1.390496	3.820351	2.098555	1.559467	4.805251
$\hat{\beta}_{13}$	7.379170	1.251622	3.095351	3.701876	2.174504	9.342954
$\hat{\beta}_{14}$	6.537453	1.511307	4.51304	4.496350	1.174135	2.723953
$\hat{\beta}_{23}$	5.900824	1.071782	2.269741	3.981786	0.9605398	1.823033
$\hat{\beta}_{24}$	2.008130	1.371336	3.715794	2.356240	1.18719	2.784865
$\hat{\beta}_{34}$	6.129555	1.205846	2.873076	5.223936	1.686612	5.620749
	k = 0,0387 MSEOLS = 4527,256 MSERIDGE = 3.464503			d = 0,192 MSEOLS = 4527,256 MSELIU = 873.0753		

Elde edilen sonuçlara dikkat edilirse M tahminine dayalı robust Ridge ve Liu tahminlerine göre her açıdan daha iyi sonuçlar elde edilmiştir. Özellikle Schweppe tahminine dayalı robust Ridge tahmin edicisi MSE, standart hata ve VIF_i değerleri açısından çok iyi sonuçlar üretmiştir. Ayrıca psüdobileşen dönüşümü yapıldıktan sonra Liu tahmin edicisinin performansı gözle görülür derecede artmıştır. Dönüşüm yapılmadan önce oldukça yüksek elde edilen MSE değerindeki düşüş dikkatten kaçmamaktadır.

Sonuç olarak psudobileşen dönüşümü ile sapan değerlerin etkileri azaltılamamıştır ancak koşul sayısındaki düşüş veya kötü koşulluluk probleminin azaltılması ile 3. ve 4. Bölümlerde ele aldığımız Robust M ve GM tahmin edicileri daha etkili sonuçlar üretmişlerdir.

6. BÖLÜM. SONUÇLAR VE ÖNERİLER

Karma denemelerde sistemi oluşturan bileşenler üzerinde (1.1) ve (1.2)'de verilen karma denemelerin doğal kısıtlamaları yanı sıra genellikle (1.3) ve (1.4)'te verilen ek kısıtlamalar bulunur. Bu kısıtlamalar regresyon analizinde araştırmacıların sıkça karşılaştıkları kötü koşulluluk ve iç ilişki probleminin doğmasına sebep olur. Bu tezde bu kısıtlamalara ek olarak y ve/veya x yönünde sapan değerlerin karma modele etkileri incelenmiştir. Ayrıca kötü koşulluk problemini ortadan kaldırmak için kullanılan Ridge ve Liu regresyon yöntemlerinin robust şekillerinin performansları incelenmiştir.

İlk bölümde $\{3, 2\}$ simpleks latis tasarımı, tasarım noktalarında çok tekrarlı gözlemlerin olması durumu ele alınmıştır. (1.9)'daki modelin (1.35)'te verilen en küçük kareler tahminleri simpleksin köşe ve orta noktalarında yapılan tekrarlı gözlemlerin aritmetik ortalamasına dayalıdır. Ancak bilindiği gibi aritmetik ortalama bir tek sapan değerden bile çok fazla etkilenen bir istatistiktir. Yani alınan ölçümlerde dış koşullardan kaynaklanan hatalar varsa bunlar tahmin edilen modeli çok fazla etkilemektedir. Bu tip hatalı gözlemlerin etkilerini azaltmak veya ortadan kaldırmak için robust konum ve ölçek tahmin edicileri kullanılmıştır.

İkinci bölümde, karma verilerin gözlem değerlerinin standart tasarımlardan uzaklaştıkları veya XVERT, XVERT1 gibi kısıtlanmış bölgenin tüm köşe, yüzey orta ve bölge merkezi noktalarının koordinatları arasından seçilmiş tüm aday noktaların bir alt kümesi seçilerek oluşturulmuş veriler ele alınmıştır. Bu bölümde y yönünde sapan değerlerin karma verilere etkileri incelenerek robust M tahmin edicilerinin performansları araştırılmıştır. İlk olarak hataların dağılımının formunun bilinmediği varsayılarak Huber ve Tukey ρ fonksiyonları ile katsayı tahminleri elde edilmiştir daha sonra hataların t ve GT dağılımından geldiği varsayılarak katsayı tahminleri elde edilmiştir. Son olarak hataların çarpık bir dağılım olan SGT dağılımından gelmesi durumunda robust M tahmin edicisi ele alınmıştır. Tüm bu incelemelerde en küçük kareler tahmin edicilerinin hassas oldukları durumlardan robust tahmin ediciler beklenildiği gibi çok fazla etkilenmemişlerdir.

Üçüncü bölümde, Ridge ve Liu tahmin edicileri ile bunların robust şekilleri incelenmiştir. Bilindiği gibi her iki yanlı tahmin edici sapan değerlere karşı çok hassas olan en küçük kareler tahminine dayalıdır. Bu bölümde en küçük kareler tahminleri yerine robust M tahminlerinin kullanılmasının katsayı tahminleri, VIF_i , MSE gibi istatistiklere etkileri gösterilmiştir. Gerçektende özellikle Ridge tahmin edicisinin robust şekli çok iyi bir performans göstermiştir ve kötü koşulluluk problemi ortadan kaldırıldığı gibi y yönündeki olası sapan değerlerin etkileri en küçük kareler yerine robust M tahminleri kullanılarak ortadan kaldırılmıştır.

Dördüncü bölümde, hem x hemde y yönünde sapan değerlerin olması durumunda her iki sapan değere karşı etkili olan GM tahmin edicileri incelenmiştir. Montgomery ve Voth (1994) karma denemelerin standart tasarımlarında etkili gözlem problemi olduğunu ve yüksek kırılma noktasına sahip tahmin edicilerin araştırılmasını önermiştir. Gerçektende Arslan ve Billor (1996, 2000) tarafından verilen GM tahminine dayalı robust Ridge ve Liu tahmin edicileri bu sayede kötü koşulluluk ve sapan değerlere karşı çok iyi performans sergilemişlerdir. Ayrıca bu bölümde x yönündeki sapan değerlerin belirlenmesinde çok sık kullanılan ancak etkilerin maskelenmesinde çok etkilenen H şapka matrisi köşegen elemanları h_{ii} 'lere alternatif olarak x'in varyans-kovaryans matrisinin robust tahmini olan MVE'ye dayalı robust Mahalanobis uzaklıkları ele alınmıştır.

Son olarak çok iyi bilinen psüdobileşen dönüşümünün etkileri incelenmiştir. St. John (1984) psüdobileşen dönüşümü ve değişkenleri standartlaştırma işlemlerinin kötü koşulluluğun belirlenmesinde kullanılan VIF_i ve koşul sayısı değerlerinde iyileşmeler yaptığını ve Ridge tahmin edicisinin seçilmiş k değerleri için en küçük kareler tarafından üretilen MSE değerinden daha iyi MSE değeri verdiğini göstermiştir. Bu bölümde benzer inceleme robust Ridge ve Liu tahmin edicileri için yapılmıştır. Psüdobileşen dönüşümü özellikle Liu tahminlerinin etkinliğini arttırmıştır.

Bu çalışmada, üretim formüllerinin optimizasyonunda doğrusal modellere göre çok fazla tercih edilen karesel modeller kullanılmıştır. Ancak bazı durumlarda (1.28) ve (1.29)'da verdiğimiz doğrusal ve karesel modeller dışında (6.1) ve (6.2)'de verilen sırasıyla kübik ve özel kübik modeller daha etkin olabilirler.

$$\eta = \sum_{i=1}^q \beta_i x_i + \sum_{i<j}^q \beta_{ij} x_i x_j + \sum_{i<j}^q \delta_{ij} x_i x_j (x_i - x_j) + \sum_{i<j<k}^q \beta_{ijk} x_i x_j x_k \quad (6.1)$$

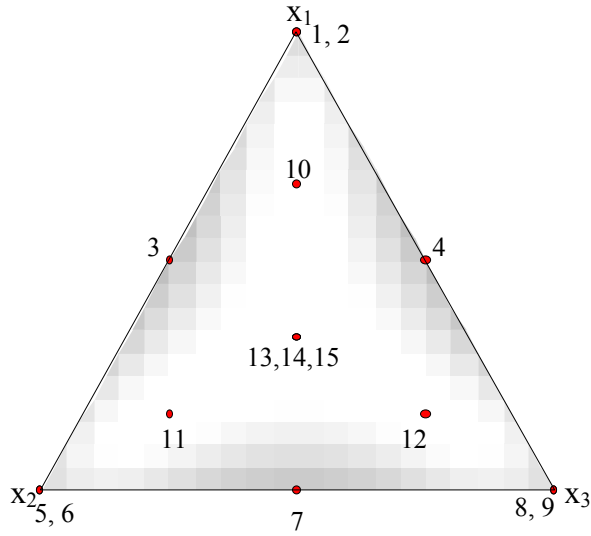
$$\eta = \sum_{i=1}^q \beta_i x_i + \sum_{i<j}^q \beta_{ij} x_i x_j + \sum_{i<j<k}^q \beta_{ijk} x_i x_j x_k \quad (6.2)$$

Örneğin Montgomery ve Voth (1994)'ün verdikleri yakıt karışımı verisini ele alalım. Örnek, roket motorlarının gücünü arttırmak için benzin (x_1), oksidan (x_2) ve bağlayıcı (x_3) maddelerin karışım formlarının incelenmesi ile ilgilidir. Birçok karma denemede olduğu gibi yakıtın fiziksel ve mekanik gibi özellikleri incelenebilir ancak burada sadece motorun performansını arttırıcı etkisi olan yanma hızı (y) ele alınacaktır. Psüdobileşen dönüşümü yapılmış veriler Tablo 6.1'de verilmiştir.

Tablo 6.1. Psüdobileşen dönüşümü yapılmış yakıt karışımı verisi.

Sıra	Benzin (x_1)	Oksidan (x_2)	Bağlayıcı (x_3)	Yanma Hızı (y)
1	1	0	0	32.5
2	1	0	0	37.9
3	0.5	0.5	0	44.0
4	0.5	0	0.5	63.2
5	0	1	0	54.5
6	0	1	0	32.5
7	0	0.5	0.5	94.0
8	0	0	1	64.0
9	0	0	1	78.5
10	0.666	0.166	0.166	67.1
11	0.166	0.666	0.166	73.0
12	0.166	0.166	0.666	87.5
13	0.333	0.333	0.333	112.5
14	0.333	0.333	0.333	98.5
15	0.333	0.333	0.333	103.6

Şekil 6.1'de üçgensel koordinatlarda tasarım noktalarının yerleşimi verilmiştir. Design Expert 6 paket programı ile elde edilen ANOVA ve model uygunluğu testi sonuçları Tablo 6.2'de verilmiştir.



Şekil 6.1. Tasarım noktalarının üçgensel koordinatlardaki yerleşimleri

Tablo 6.2. Yakıt karışımı verisi için ANOVA ve Model uygunluğu testi

ANOVA				
Model	Kareler Toplamı	SD	Ortalama Kareler	F
Doğrusal	2395.93	2	1197.97	2.01
Karesel	5486.83	3	1828.94	9.83
Özel Kübik	<u>1063.60</u>	<u>1</u>	<u>1063.60</u>	<u>13.92</u>
Rezidü	462.11	5	92.42	
Model Uygunluğu Testi				
Model	Kareler Toplamı	SD	Ortalama Kareler	F
Doğrusal	6699.74	7	957.11	10.36
Karesel	1212.91	4	303.23	3.28
Özel Kübik	<u>149.30</u>	<u>3</u>	<u>49.77</u>	<u>0.54</u>
Hata	462.11	5	92.42	

Tablo 6.2’de görüldüğü gibi özel kübik model diğer modellerden daha iyi sonuçlar vermiştir. Tablo 6.3’te özel kübik ve karesel model için elde edilen sonuçlar verilmiştir. Dikkat edilirse katsayı matrisinin koşul sayısı ve VIF_i ’lerdeki fark önemli derecededir. Ayrıca MSE değerleri açısından da karesel model daha iyi sonuç vermiştir. Yani karesel model için düşük seviyelerde olan kötü koşulluluk problemi üçlü etkileşim teriminin eklenmesiyle elde edilen özel kübik modelde çok daha fazlalaşmıştır. Benzer şekilde h_{ii} değerleri de karesel model için maksimum 0.693 iken özel kübik için 0.909’dur. Bu değerler beraberinde özel kübik modelde x yönünde sapan değer probleminin de ortaya çıkmasına sebep olur.

Tablo 6.3. Özel kübik ve karesel modeller için en küçük kareler sonuçları

Terim	Özel Kübik Model				Karesel Model			
	$\hat{\beta}_{LS}$	$se(\hat{\beta}_{LS})$	Özdeğer	VIFi	$\hat{\beta}_{LS}$	$se(\hat{\beta}_{LS})$	Özdeğer	VIFi
$\hat{\beta}_1$	35.484	6.114	5.14347	0.4825	33.678	9.488	5.1415	0.4794
$\hat{\beta}_2$	42.764	6.114	2.51138	0.4825	40.958	9.488	2.5113	0.4794
$\hat{\beta}_3$	70.349	6.114	2.51138	0.4825	68.544	9.488	2.5113	0.4794
$\hat{\beta}_4$	15.862	38.553	0.09623	19.18412	83.912	52.817	0.0944	14.85
$\hat{\beta}_{12}$	36.171	38.553	0.05808	19.18412	104.220	52.817	0.0580	14.85
$\hat{\beta}_{13}$	136.625	38.553	0.05808	19.18412	204.674	52.817	0.0580	14.85
$\hat{\beta}_{14}$	858.89	231.109	0.00142	689.3697	$\hat{\sigma}^2 = 187.772$ $\hat{\sigma} = 13.70298$ KS = 88.517 MSEOLS = 8639.138			
	$\hat{\sigma}^2 = 77.4789$ $\hat{\sigma} = 8.80221$ KS = 3613.576 MSEOLS = 57982.86							

Tablo 6.3'teki sonuçlardan hareketle, karma denemelerde bileşen sayısının 3'ten büyük olduğu durumlarda karesel modelleri kullanmak tahmin edilecek parametre sayısı da göz önüne alındığında daha uygundur ve çoğu zaman kübik veya özel kübik modelden daha iyi koşul sayısı ve VIF_i değerlerine sahip olur.

Sonuç olarak çeşitli karma veriler x ve/veya y yönündeki sapan değerler ve çoklu iç ilişki veya kötü koşulluluk problemi açısından incelenmiştir. Genellikle kısıtlanmış bölge simpleks ise psüdobileşen dönüşümü kötü koşulluluk problemini gidermede yeterli olacaktır. Ancak kısıtlanmış bölge simplek değilse psüdobileşen dönüşümü kötü koşulluluk problemini düzeltse bile yeterli olmayacaktır. Bunlara birde sapan değer problemi eklendiğinde kötü koşulluluk problemine karşı etkili olan Ridge ve Liu tahminleri de yetersiz kalacaklardır. Bu durumda robust tahmin edicilere dayalı Ridge ve Liu tahminleri gerek iç ilişki gerekse sapan değer problemine karşı etkili tahminler üretecektir.

İleriki çalışmalarda bu tezde ele alınan M ve GM tahminleri yanı sıra LMS, S, CM, MM, τ gibi bozulma noktaları yüksek olan tahmin ediciler ile Ridge ve Liu regresyon yöntemleri birlikte kullanılmasının etkileri incelenecektir.

KAYNAKLAR

- ARSLAN, O., BİLLOR, N. (1996). Robust Ridge Regression Estimation Based on the GM-Estimators. *Jour. Of Math. & Comp. Sci. (Math. Ser.)*, Vol.9-1, 1-9.
- ARSLAN, O., BİLLOR, N. (2000). Robust Liu Estimator for Regression Based on an M-Estimators. *Journal of Applied Statistics*, Vol.27-1, 39-47.
- ARSLAN, O., EDLUND, H., EKBLOM, H. (2001). Algorithms to Compute CM- and S-Estimates for Regression. *Lulea University of Technology Research Reports*, Lulea, Sweeden.
- ARSLAN, O., GENÇ, A.İ. (2003). Robust Location and Scale Estimation Based on the Univariate Generalized t (GT) Distribution. *Communication in Statistics-Theory and Methods*, Vol.32-8, 1505-1525.
- ARSLAN, O. (2004). Convergence Behavior of an Iterative Reweighting Algorithm to Compute Multivariate M-Estimates for Location and Scatter. *Journal of Statistical Planning and Inference*, Vol. 118, 115-128.
- ARSLAN, O. (2004). Family of Multivariate Generalized t Distribution. *Journal of Multivariate Analysis*, Vol. 89, 329-337.
- BELSLEY, D.A., KUH, E., WELSCH, R.E. (1980). *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*, Wiley, New York.
- BIRKES, D., DODGE, Y. (1993). *Alternative Medhod of Regression*.Wiley, New York.
- BLANQUERO, R., CARRIZOSA, E., CONDE, E. (2001). Finding GM Estimators with Global Optimization Techniques. *Journal of Global Optimization*, Vol. 21, 223-237.
- BUTLER, R.J., McDONALD, J.B., NELSON, R.D., WHITE,S.B. (1990). Robust and Partially Adaptive Estimation of Regression Models. *The Review of Economics and Statistics*, Vol. 72, 321-327.
- CAMARINOPOULOS, L., ZIOUTAS, G. (2002). Formulating Robust Regression Estimation as an Optimum Allocation Problem. *J. Statist Comput. Simul.*, Vol. 72-9, 687-705.

- CLARINGBOLD, P.J. (1955). Use of the simpleks design in study of the joint action of related hormones. *Biometrics*, Vol. 11, pp.. 174-185.
- CORNELL, J. A. (1971). Experiments with mixtures: A Review. *Technometrics*, Vol. 15, pp. 437-455.
- CORNELL, J. A. (1979). Experiments with mixtures: An Update and Bibliography. *Technometrics*, Vol. 21, pp. 95-106.
- CORNELL, J. A. (1990). *Experiments with Mixtures - Designs, Models, and The Analysis of Mixture Data*. 2nd edition, John Wiley& Sons, Inc., New York, USA.
- CORNELL, J.A. (2000). Developing mixture models, are we done?. *Journal of Statist. And Comput. Simul.*, Vol. 66, pp. 127-144.
- CROSIER, B.R. (1984). Mixture experiments: geometry and pseudocomponents. *Technometrics*, Vol. 26, pp. 209-216.
- CROSIER, B.R. (1986). The geometry of constrained mixture experiments. *Technometrics*, Vol. 28, pp. 95-102.
- DRAPER, N.R., SMITH, H. (1998). *Applied Regression Analysis*. 3th Edition, John Wiley and Sons, New York.
- GORMAN, J. W. (1970). Fitting equations to mixture data with restraints on compositions. *Journal of Quality Technology*, Vol. 2, pp. 186-194.
- GORMAN, J.W. and CORNELL, J.A. (1982). A note on model reduction for experiments with both mixture components and process variables. *Technometrics*, Vol. 24, pp. 243-247.
- GRAYBIL, F.A. (1976). *Theory and Application of the Linear models*, Duxbury Press, Belmont, California.
- HADI, A.S., WELLS, M.T. (1990). Assessing the Effects of Multiple Rows on the Condition Number of a Matrix. *Journal of American Statistical Association*, Vol. 85-411, 786-792.
- HAMPEL, F.R., RONCHETTI, E.M., ROUSSEEUW, P.J., STAHEL, W.A. (1986). *Robust Statistics: The Approach Based on Influential Functions*, Wiley, New York.

- HARE, L.B. (1985). Graphical display of the result of mixture experiments. *Experiments in Industry: Design, Analysis and Interpretation of Results*, Edited by R.D. Snee, L.B. Hare and R. Trout, sponsored by ASQC Chemical and Process Industries Division, ASQC, pp. 99-109.
- HOERL, A.E., KENNARD, R.W. (1970a). Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics*, Vol. 12-1, 55-67.
- HOERL, A.E., KENNARD, R.W. (1970b). Ridge Regression: Applications to Nonorthogonal Problems. *Technometrics*, Vol. 12-1, 69-82.
- HUBER, P.J. (1964). Robust Estimation of a Location Parameters. *The Annals of Mathematical Statistics*, Vol 35, 73-101.
- HUBER, P.J. (1981). *Robust Statistics*, Wiley, New York.
- HUBERT, M., ROUSSEEUW, P.J., VAN AELST, S. (2004). Robustness, *Encyclopedia of Actuarial Sciences*, edited by Sundt, B. and Teugels, J., Wiley, New York, pp.1515-1529.
- KEJIAN, L. (1993). A New Class of Biased Estimates in Linear Regression. *Commun. Statist.-Theory and Methods*, Vol. 22-2, 393-402.
- KRASKER, W.S., WELSCH, R.E. (1982). Efficient bounded-influence regression estimation. *Journal of the American Statistical Association*, Vol. 77-379, 595-604.
- LAWRENCE, K.D., ARTHUR, J.L. (1990). Robust Regression Analysis and Applications. Marcel Dekker Inc., New York
- LIU, K. (1993). A new class of biased estimate in linear regression. *Communications in Statistics A*, 22, 393-402.
- MARQUARDT, D.W. (1970). Generalized inverses, Ridge regression, biased linear estimation, and nonlinear estimation. *Technometrics*, 12, 591-612.
- MARQUARDT, D.W., SNEE, R.D. (1974). Test Statistics for Mixture Models. *Technometrics*, Vol. 16, pp. 533-537.
- MCDONALD, J.B., NEWAY, W.K. (1988). Partially Adaptive Estimation of Regression Models Via the Generalized T Distribution. *Econometric Theory*, Vol. 4, 428-457.

- McLEAN, R. D., ANDERSON, V. L. (1966). Extreme vertices design of mixture experiments. *Technometrics*, Vol. 8, pp. 447-454.
- MENDES, B., TYLER, D.E. (1996). Constrained M-Estimates for Regression. *Robust Statistics, Data Analysis and Computer Intensive Methods, Lecture Notes in Statistics*, No. 109, Springer-Verlag, New York, pp. 299-320.
- MONTGOMERY, D.C., VOTH, S.R. (1994). Multicollinearity and Leverage in Mixture experiments. *Journal of Quality Technology*, Vol. 26, pp. 96-108.
- MONTGOMERY, D.C. (2001). *Design and Analysis of Experiments*, 5th Ed., Wiley, New York.
- MYERS, R.H., MONTGOMERY, D.C. (2002). *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*, 2nd. Ed., Wiley, New York.
- PIEPEL, G., F. (1983a). Defining consistent constraint regions in mixture experiments. *Technometrics*, Vol. 25, pp. 97-101.
- PIEPEL, G., F. (1983b). Calculating centroids in constrained mixture experiments. *Technometrics*, Vol. 25, pp. 279-283.
- QUENOUILLE, M.H. (1953). *The design and Analysis of Experiments*, Charles Griffin and Company. London, England.
- ROUSSEEuw, P.J. (1984). Least Median of Squares Regression. *Journal of American Statistical Association*, Vol. 79-388, 871-880.
- ROUSSEEuw, P.J., YOHAI, V.J. (1984). Robust Regression by Means of S-Estimators. *Robust and Nonlinear Time series, Lecture Notes in Statistics*, No. 26, Springer-Verlag, New York, pp.256-272.
- ROUSSEEuw, P.J., LEROY, A.M. (1987). *Robust Regression and Outlier Detection*. Wiley, New York.
- ROUSSEEuw, P.J., ZOMEREN, B.C. (1990). Unmasking Multivariate Outliers and Leverage Points. *Journal of American Statistical Association*, Vol. 85-411, 633-651.
- SCHEFFÉ, H. (1958). Experiments with mixtures. *Journal of the Royal Statistical Society - B*, Vol. 20, pp. 344-360.

- SCHEFFÉ, H. (1963). The simplex-centroid design for experiments with mixtures. *Journal of the Royal Statistical Society - B*, Vol. 25, pp. 235-263.
- SILVAPULLE, M.J. (1991). Robust Ridge Regression Based on M-Estimator. *Australian and New Zealand Journal of Statistics*, Vol.33, 319-333.
- SIMPSON, D.G., RUPPERT, D., CARROLL, R.J. (1992). On One-Step GM Estimates and Stability of Inferences in Linear Regression. *Journal of American Statistical Association*, Vol. 87-418, 439-450.
- SIMPSON, D.G., YOHAİ, V.J. (1998). Functional Stability of One-Step GM Estimators in Approximately Linear Regression. *The Annals of Statistics*, Vol. 26-3, 1147-1169.
- SNEE, R. D. (1973). Techniques for the Analysis of Mixture Data.. *Technometrics*, Vol. 15, pp. 517-528.
- SNEE, R. D. (1975). Experimental designs for quadratic models in constrained mixture spaces. *Technometrics*, Vol. 17, pp. 149-159.
- SNEE, R. D., MARQUARDT, D. W. (1974). Extreme vertices designs for linear mixture models. *Technometrics*, Vol. 16, pp. 399-408.
- SNEE, R. D., RAYNER, A. A. (1982). Assessing the accuracy of mixture model regression calculations. *Journal of Quality Technology*, Vol. 14, pp. 67-79.
- ST. JOHN, R. C. (1984). Experiments with mixtures, ill conditionning, and Ridge regression. *Journal of Quality Technology*, Vol. 16, pp. 81-96.
- STAT-EASE (2004). DESIGN-EXPERT Software for Response Surface Methodology and Mixture Experiments. Version 6, 45 days Trial. Stat-Ease, Inc. Minneapolis, MN.
- THEODOSSIOU, P. (1998). Financial Data and Skewed Generalize T Distribution. *Management Science*, Vol.44-12, 1650-1661.
- WELSH, A.H., MORRISON, H.L. (1990). Robust L Estimation of Scale with an Application in Astronomy. *Journal of American Statistical Association*, Vol. 85-411, 729-742.
- YOHAİ, V.J. (1987). High Breakdown Point and High Efficiency Robust Estimates for Regression. *The Annals of Statistics*, Vol. 15-20, 642-656.

ÖZGEÇMİŞ

1974 yılında Adana'da doğdum. İlk, orta ve lise eğitimimi Adana'da tamamladıktan sonra 1992'de Çukurova Üniversitesi Fen-Edebiyat Fakültesi Matematik Bölümüne girdim. 1997 yılında mezun olup 1998 yılında Çukurova Üniversitesi Fen-Edebiyat Fakültesi Matematik Bölümünde yüksek lisans 2001 yılında doktora eğitime başladım. 1999 yılında aynı bölümde araştırma görevlisi olarak göreve başladım.