

**ANKARA ÜNİVERSİTESİ  
FEN BİLİMLERİ ENSTİTÜSÜ**

**DOKTORA TEZİ**

**SİMETRİK VE ÇARPIK DAĞILIMLAR İÇİN BİLGİ KARMAŞIKLIK  
(ICOMP) KRİTERİ KULLANILARAK ROBUST MODEL SEÇİMİ**

**Yeşim GÜNEY**

**İSTATİSTİK ANABİLİM DALI**

**ANKARA**

**2018**

**Her hakkı saklıdır**

## TEZ ONAYI

Yeřim GÜNEY tarafından hazırlanan “Simetrik ve Çarpık Dağılımlar için Bilgi Karmaşıklık (ICOMP) Kriteri Kullanılarak Robust Model Seçimi” adlı tez çalışması 05/04/2018 tarihinde ařağıdaki jüri tarafından oy birliğı ile Ankara Üniversitesi Fen Bilimleri Enstitüsü İstatistik Anabilim Dalı’nda **DOKTORA TEZİ** olarak kabul edilmiştir.

**Danışman** : Prof. Dr. Olçay ARSLAN  
Ankara Üniversitesi İstatistik Anabilim Dalı



**Eř Danışman** : Prof. Dr. Hamparsum BOZDOĞAN  
Tennessee Üniversitesi İstatistik Anabilim Dalı

### Jüri Üyeleri :

**Başkan** : Prof. Dr. Özgür YENİAY  
Hacettepe Üniversitesi İstatistik Anabilim Dalı

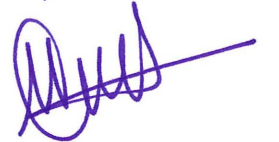


**Üye** : Prof. Dr. Olçay ARSLAN  
Ankara Üniversitesi İstatistik Anabilim Dalı

**Üye** : Prof. Dr. Birdal ŞENOĞLU  
Ankara Üniversitesi İstatistik Anabilim Dalı



**Üye** : Prof. Dr. Meral ÇETİN  
Hacettepe Üniversitesi İstatistik Anabilim Dalı



**Üye** : Prof. Dr. Mehmet YILMAZ  
Ankara Üniversitesi İstatistik Anabilim Dalı



**Yukarıdaki sonucu onaylıyorum.**

**Prof. Dr. Atila YETİŞEMİYEN**  
Enstitü Müdürü

## ETİK

Ankara Üniversitesi Fen Bilimleri Enstitüsü tez yazım kurallarına uygun olarak hazırladığım bu tez içindeki bütün bilgilerin doğru ve tam olduğunu, bilgilerin üretilmesi aşamasında bilimsel etiğe uygun davrandığımı, yararlandığım bütün kaynakları atıf yaparak belirttiğimi beyan ederim.

05.04.2018



Yeşim GÜNEY

# ÖZET

Doktora Tezi

## SİMETRİK VE ÇARPIK DAĞILIMLAR İÇİN BİLGİ KARMAŞIKLIK (ICOMP) KRİTERİ KULLANILARAK ROBUST MODEL SEÇİMİ

Yeşim GÜNEY

Ankara Üniversitesi  
Fen Bilimleri Enstitüsü  
İstatistik Anabilim Dalı

Danışman: Prof. Dr. Olçay ARSLAN

Gelişen teknoloji sayesinde bir istatistiksel problemin çözümü için daha karmaşık modeller ortaya çıkmaktadır. Bu durum alternatif model sayısında artışa neden olmuştur. Bu modellerden hangisinin veriye en uygun, en basit model olduğunu belirlemek için bir çok seçim kriteri tanımlanmıştır. Dağılım varsayımının sağlanmaması, verinin aykırı değer içermesi gibi durumlardan klasik parametre tahmin yöntemleri etkilendiği için, bu tahmin yöntemlerine dayalı model seçim kriterleri de etkilenmektedir. Bu tür problemlerle karşılaşıldığında güvenilir sonuçlar elde etmek için robust model seçim kriterlerine ihtiyaç duyulmaktadır. Yaygın kullanılan kriterlerin robust formları ile ilgili zengin literatür olmasına rağmen, Bilgi Karmaşıklığı (ICOMP) kriterinin robust formlarına ilişkin literatürde az sayıda çalışma vardır. Bu tezde farklı birtakım robust istatistiksel yöntemlere dayalı robust ICOMP kriterleri tanımlanmıştır. Çoklu doğrusal regresyon modelinde hataların kalın kuyruklu ve/veya çarpık olması durumları için Student  $t$ , çarpık normal ve çarpık  $t$  dağılım varsayımları altında model karmaşıklığı hesaplanarak ICOMP bilgi kriterinin formları bu dağılım varsayımları için elde edilmiştir. Bu tezde çoklu doğrusal regresyon modelinde M, S ve MM tahmin edicilerine dayalı robust ICOMP kriterleri önerilmiştir. Çoklu doğrusal regresyonda karşılaşılan bir diğer problem olan çoklu iç ilişkinin model seçimi üzerine etkilerinin azaltılması için ICOMP bilgi kriterinin robust ridge tahmin edicilerine dayalı bir başka robust formu tanımlanmıştır. Karmaşık verilerin modellenmesinde kullanılan ortak konum-ölçek ve konum-ölçek-çarpıklık modellerinde ICOMP kriterinin formları elde edilmiştir. Önerilen kriterlerin, Akaike Bilgi Kriteri (AIC) ve ICOMP kriterleri ile performansları simülasyon çalışmaları ve gerçek veri örnekleri ile karşılaştırılmıştır.

Nisan 2018, 149 sayfa

**Anahtar Kelimeler:** Bilgi karmaşıklığı kriteri, robust istatistiksel analiz, robust model seçimi, robust regresyon, kovaryans karmaşıklığı, ortak konum-ölçek-çarpıklık modeli

## ABSTRACT

Ph.D. Thesis

### ROBUST MODEL SELECTION USING INFORMATION COMPLEXITY (ICOMP) CRITERION FOR SYMMETRIC AND SKEW DISTRIBUTIONS

Yeşim GÜNEY

Ankara University  
Graduate School of Natural and Applied Sciences  
Department of Statistics

Supervisor: Prof. Dr. Olçay ARSLAN

Because of the developing technology, more complicated models can appear for solving a statistical problem. This situation has led to an increase in the number of alternative models. Many criteria have been defined in the literature to determine which of these models is the simplest model that fits the data well. The model selection criteria based on classical parameter estimation methods are also affected, since these methods are influenced by situations such as deviation from distributional assumptions and the fact that the data contain outliers. When such problems are encountered, robust model selection criteria are needed to obtain reliable results. Despite the rich literature on robust forms of commonly used criteria, there are few studies in the literature on robust forms of the Information Complexity (ICOMP) criterion. In this thesis, robust ICOMP criteria based on different robust statistical methods are defined. For multiple regression models, the model complexity is calculated under the Student  $t$ , skew normal and skew  $t$  distribution assumptions for cases where the errors are thick-tailed and / or skewed and the ICOMP information criterion is formulated for these distributional assumptions. In this thesis, a robust ICOMP criterion based on M, S and MM estimators is proposed in a multiple linear regression model. Another robust form based on the robust ridge estimators of the ICOMP information criterion has been defined to reduce the effects of multicollinearity on model selection, another problem encountered in multiple linear regression. The forms of the ICOMP criterion have been obtained in the joint location-scale and location-scale-skewness models used in the modeling of complex data. The performances of the proposed criteria, Akaike Information Criterion (AIC) and ICOMP are compared with the simulation studies and real data examples.

**April 2018, 149 pages**

**Key Words:** Information complexity criterion, robust statistical analysis, robust model selection, robust regression, covariance complexity, joint location-scale-skewness model

## TEŞEKKÜR

Bu tezde problemlerin ortaya konulup çözülmesine kadar olan süreç boyunca sonsuz bilgi ve tecrübesiyle yolumu aydınlatan, desteğiyle moral ve motivasyonumu her zaman en üst seviyede tutmamı sağlayan, kendisiyle çalışmaktan onur duyduğum değerli danışman hocam Sayın Prof. Dr. Olçay ARSLAN (Ankara Üniversitesi, İstatistik A.B.D.)'a en içten saygı ve minnetlerimi sunarım.

Tez çalışmasının fikir aşamasındaki katkılarından dolayı, eş danışmanlığımı kabul edip benden desteklerini esirgemeyen, değerli hocam Sayın Prof. Dr. Hamparsum Bozdoğan (Tennessee Üniversitesi)'a saygı ve teşekkürlerimi sunarım. Değerli TİK üyeleri, hocalarım Sayın Prof. Dr. Meral ÇETİN (Hacettepe Üniversitesi, İstatistik A.B.D.)'e ve Sayın Prof. Dr. Mehmet YILMAZ (Ankara Üniversitesi, İstatistik A.B.D.)'a olumlu eleştirileri ve desteklerinden dolayı teşekkürlerimi bir borç bilirim. Tez çalışmamı titizlikle inceleyerek değerli fikir ve önerilerini benimle paylaşan değerli hocalarım Sayın Prof. Dr. Birdal ŞENOĞLU (Ankara Üniversitesi, İstatistik A.B.D.)'na ve Sayın Prof. Dr. Özgür YENİAY (Hacettepe Üniversitesi, İstatistik A.B.D.)'a teşekkürlerimi sunarım.

Doktora tezimde simülasyon çalışmalarında yardımlarını gördüğüm, destekleriyle hep yanımda olan arkadaşlarım Dr. Öğr. Üyesi Ömer ALTINDAĞ (Bilecik Üniversitesi, İstatistik A.B.D.)'a ve Araş. Gör. Abdullah YALÇINKAYA (Ankara Üniversitesi, İstatistik A.B.D.)'ya, manevi desteğini her zaman hissettiğim, zor anlarımda hep yanımda olan kıymetli arkadaşım Araş. Gör. Şenay ÖZDEMİR (Afyon Kocatepe Üniversitesi, İstatistik A.B.D.)'e ve birlikte çalışmaktan, bu mesleğin zorluklarını da güzelliklerini de beraber yaşamaktan, mutluluk duyduğum değerli dostum Araş. Gör. Yetkin TUAÇ (Ankara Üniversitesi, İstatistik A.B.D.)'a teşekkürlerimi iletirim. Ankara Üniversitesi İstatistik Bölümü öğretim üyeleri ve öğretim elemalarına bu süreçte oldukları destekten ötürü teşekkürlerimi sunarım.

Bu tez "TÜBİTAK-2211 A Genel Yurt İçi Doktora Burs Programı" tarafından desteklenmiştir. Akademik çalışmalarım için sağlanan destekten dolayı TÜBİTAK'a ayrıca teşekkürlerimi sunarım.

Bana her konuda destek olan ve her zaman inanan, gücümü ve azmimi bana duydukları güvenlerinden ve sevgilerinden aldığım, beni bugünlere kadar getiren canım aileme ve eşime en içten teşekkürlerimi sunarım.

Yeşim GÜNEY

Ankara, Nisan 2018

## İÇİNDEKİLER

### TEZ ONAY SAYFASI

|                                                                                                |      |
|------------------------------------------------------------------------------------------------|------|
| ETİK.....                                                                                      | i    |
| ÖZET.....                                                                                      | ii   |
| ABSTRACT .....                                                                                 | iii  |
| TEŞEKKÜR .....                                                                                 | iv   |
| SİMGELER DİZİNİ .....                                                                          | viii |
| ŞEKİLLER DİZİNİ .....                                                                          | ix   |
| ÇİZELGELER DİZİNİ .....                                                                        | x    |
| 1. GİRİŞ .....                                                                                 | 1    |
| 2. REGRESYON MODELİ.....                                                                       | 10   |
| 2.1 Çoklu Doğrusal Regresyon Modeli.....                                                       | 10   |
| 2.2 Çoklu Doğrusal Regresyon Modelinde Parametrelerin En Çok Olabilirlik Tahmin Edicileri..... | 11   |
| 2.2.1 Hataların normal dağıldığı durum .....                                                   | 15   |
| 2.2.2 Hataların Student $t$ dağıldığı durum .....                                              | 18   |
| 2.2.3 Hataların çarpık normal dağıldığı durum.....                                             | 21   |
| 2.2.4 Hataların çarpık $t$ dağıldığı durum .....                                               | 26   |
| 2.3 Çoklu Doğrusal Regresyon Modelinde Robust Parametre Tahmin Yöntemleri.....                 | 31   |
| 2.3.1 M tahmin edicileri .....                                                                 | 31   |
| 2.3.2 S tahmin edicileri .....                                                                 | 37   |
| 2.3.3 MM tahmin edicileri .....                                                                | 41   |
| 3. MODEL SEÇİMİ VE MODEL SEÇİM KRİTERLERİ.....                                                 | 44   |
| 3.1 Bilgi Tabanlı Model Seçim Kriterleri.....                                                  | 44   |
| 3.1.1 Bilgi kuramı .....                                                                       | 44   |
| 3.1.2 Kullback-Leibler bilgisi .....                                                           | 46   |
| 3.1.3 Akaike bilgi kriteri (AIC) .....                                                         | 47   |
| 3.1.4 Düzeltilmiş Akaike bilgi kriteri (AICc).....                                             | 49   |
| 3.1.5 Bayes bilgi kriteri (SBC) .....                                                          | 50   |
| 3.1.6 Mallows'un $C_p$ kriteri .....                                                           | 50   |
| 3.1.7 Bilgi karmaşıklığı kriteri (ICOMP-Information Complexity) .....                          | 51   |
| 3.2 Robust Model Seçim Kriterleri.....                                                         | 56   |
| 3.2.1 M tahmin edicisine dayalı robust AIC (AICR) .....                                        | 57   |
| 3.2.2 S tahmin edicisine dayalı robust AIC (AICS) .....                                        | 58   |

|                                                                                                                      |     |
|----------------------------------------------------------------------------------------------------------------------|-----|
| 3.2.3 MM tahmin edicisine dayalı robust AIC (AICMM) .....                                                            | 58  |
| 3.2.4 Robust Cp .....                                                                                                | 59  |
| 4. FARKLI DAĞILIM VARSAYIMI ALTINDA DOĞRUSAL REGRESYON<br>MODELİ İÇİN <i>ICOMP</i> .....                             | 60  |
| 4.1 Student <i>t</i> Dağılımı için <i>ICOMP</i> ( <i>ICOMP<sub>t</sub></i> ).....                                    | 60  |
| 4.1.1 Çok değişkenli Student <i>t</i> dağılımı için kovaryans karmaşıklığı .....                                     | 61  |
| 4.2 Çarpık Normal Dağılımı için <i>ICOMP</i> ( <i>ICOMP<sub>SN</sub></i> ) .....                                     | 63  |
| 4.2.1 Çok değişkenli çarpık normal dağılımı için kovaryans karmaşıklığı .....                                        | 69  |
| 4.3 Çarpık <i>t</i> Dağılımı için <i>ICOMP</i> ( <i>ICOMP<sub>St</sub></i> ) .....                                   | 69  |
| 4.3.1 Çok değişkenli çarpık <i>t</i> dağılımı için kovaryans karmaşıklığı.....                                       | 69  |
| 4.4 Simülasyon Çalışması .....                                                                                       | 72  |
| 4.4.1 Diyabet verisi .....                                                                                           | 77  |
| 4.4.2 Avustralya Spor Enstitüsü verisi .....                                                                         | 79  |
| 5. ROBUST <i>ICOMP</i> BİLGİ KRİTERİ .....                                                                           | 81  |
| 5.1 M, S ve MM Tahmin Edicilerine Dayalı <i>ICOMP</i> Bilgi Kriterleri .....                                         | 81  |
| 5.1.1 M tahmin edicisine dayalı <i>ICOMP</i> bilgi kriteri.....                                                      | 82  |
| 5.1.2 S tahmin edicisine dayalı <i>ICOMP</i> bilgi kriteri .....                                                     | 83  |
| 5.1.3 MM tahmin edicisine dayalı <i>ICOMP</i> bilgi kriteri .....                                                    | 84  |
| 5.2 Simülasyon Çalışması .....                                                                                       | 84  |
| 5.3 Çoklu Doğrusal Bağlantı ve Ridge Regresyon .....                                                                 | 89  |
| 5.3.1 Çoklu doğrusal bağlantı sorunu .....                                                                           | 89  |
| 5.3.2 Ridge regresyon.....                                                                                           | 92  |
| 5.3.3 Robust ridge tahmin edicileri.....                                                                             | 93  |
| 5.4 Çoklu Doğrusal Bağlantı ve Aykırı Değer Durumunda Çoklu Doğrusal<br>Regresyon Modellerinde Değişken Seçimi ..... | 94  |
| 5.4.1 Simülasyon çalışması .....                                                                                     | 95  |
| 5.4.2 Aylık kazanç verisi .....                                                                                      | 100 |
| 5.4.3 Diyabet verisi .....                                                                                           | 101 |
| 6. ORTAK KONUM-ÖLÇEK VE KONUM-ÖLÇEK-ÇARPIKLIK<br>MODELLERİNDE <i>ICOMP</i> İLE DEĞİŞKEN SEÇİMİ.....                  | 103 |
| 6.1 Ortak Konum-Ölçek Modelinde <i>ICOMP</i> ile Değişken Seçimi.....                                                | 103 |
| 6.1.1 Konum ölçek dağılımlar ailesi .....                                                                            | 103 |
| 6.1.2 Ortak konum-ölçek modeli .....                                                                                 | 105 |
| 6.1.3 Konum ölçek dağılımının ortak konum-ölçek modelinde değişken<br>seçimi .....                                   | 107 |
| 6.1.4 Normal dağılımın ortak konum-ölçek modelinde değişken seçimi .....                                             | 109 |
| 6.1.5 Student-t dağılımının ortak konum-ölçek modelinde değişken seçimi .....                                        | 113 |

|                                                                                                   |     |
|---------------------------------------------------------------------------------------------------|-----|
| 6.2 Ortak Konum-Ölçek-Çarpıklık Modelinde <i>ICOMP</i> ile Değişken Seçimi.....                   | 118 |
| 6.2.1 Çarpık simetrik dağılımlar ailesi.....                                                      | 118 |
| 6.2.2 Ortak konum-ölçek-çarpıklık modeli.....                                                     | 119 |
| 6.2.3 Çarpık-simetrik dağılımların ortak konum-ölçek-çarpıklık modelinde<br>değişken seçimi ..... | 120 |
| 6.2.4 Çarpık normal dağılımın ortak konum-ölçek-çarpıklık modelinde<br>değişken seçimi .....      | 122 |
| 6.2.5 Çarpık- <i>t</i> dağılımın ortak konum-ölçek-çarpıklık modelinde değişken<br>seçimi .....   | 125 |
| 6.2.6 Simülasyon çalışması .....                                                                  | 128 |
| 6.2.7 Arazi kiralari verisi.....                                                                  | 130 |
| 6.2.8 Avustralya Spor Enstitüsü verisi .....                                                      | 132 |
| 7. TARTIŞMA VE SONUÇ.....                                                                         | 135 |
| KAYNAKLAR .....                                                                                   | 138 |
| ÖZGEÇMİŞ.....                                                                                     | 148 |

## SİMGELER DİZİNİ

|                      |                                                                     |
|----------------------|---------------------------------------------------------------------|
| $\mathbb{R}$         | Reel sayılar kümesi                                                 |
| $ \cdot $            | Matris determinantı                                                 |
| $tr(\cdot)$          | Matrisin izi                                                        |
| $\phi(\cdot)$        | $N(0,1)$ dağılımının olasılık yoğunluk fonksiyonu                   |
| $\Phi(\cdot)$        | $N(0,1)$ dağılımının dağılım fonksiyonu                             |
| $t_\nu$              | $\nu$ serbestlik dereceli student-t dağılımının yoğunluk fonksiyonu |
| $T_\nu$              | $\nu$ serbestlik dereceli student-t dağılımının dağılım fonksiyonu  |
| $\mathcal{C}(\cdot)$ | Karmaşıklık ölçüsü                                                  |
| $\Gamma(\cdot)$      | Gama fonksiyonu                                                     |
| $\Psi(\cdot)$        | Digama fonksiyonu                                                   |
| $\Psi'(\cdot)$       | Trigama fonksiyonu                                                  |

### Kısaltmalar

|                       |                                                 |
|-----------------------|-------------------------------------------------|
| AIC                   | Akaike bilgi kriteri                            |
| AICc                  | Düzeltilmiş Akaike bilgi kriteri                |
| AIC.M                 | M tahmin edicilerine dayalı robust ICOMP        |
| AIC.S                 | S tahmin edicilerine dayalı robust ICOMP        |
| AIC.MM                | MM tahmin edicilerine dayalı robust ICOMP       |
| ICOMP                 | Bilgi Karmaşıklığı                              |
| ICOMP <sub>t</sub>    | Student $t$ dağılımına ilişkin ICOMP            |
| ICOMP <sub>SN</sub>   | Çarpık normal dağılımına ilişkin ICOMP          |
| ICOMP <sub>St</sub>   | Çarpık $t$ dağılımına ilişkin ICOMP             |
| ICOMP <sub>R-M</sub>  | Ridge-M tahmin edicilerine dayalı robust ICOMP  |
| ICOMP <sub>R-S</sub>  | Ridge-S tahmin edicilerine dayalı robust ICOMP  |
| ICOMP <sub>R-MM</sub> | Ridge-MM tahmin edicilerine dayalı robust ICOMP |
| IRA                   | İteratif yeniden ağırlıklandırılmış algoritma   |
| LS                    | En küçük kareler                                |
| M                     | En çok olabilirlik tipi                         |
| MM                    | Çok aşamalı M                                   |
| ML                    | En çok olabilirlik                              |
| RICOMP <sub>M</sub>   | M tahmin edicilerine dayalı robust ICOMP        |
| RICOMP <sub>S</sub>   | S tahmin edicilerine dayalı robust ICOMP        |
| RICOMP <sub>MM</sub>  | MM tahmin edicilerine dayalı robust ICOMP       |
| S                     | Ölçek tipi                                      |
| SN                    | Çarpık normal                                   |
| St                    | Çarpık $t$                                      |
| SBC                   | Bayes bilgi kriteri                             |

## ŞEKİLLER DİZİNİ

|                                                                                                              |     |
|--------------------------------------------------------------------------------------------------------------|-----|
| Şekil 2.1 Farklı $\lambda$ değerleri için çarpık normal dağılımın oyf grafikleri .....                       | 23  |
| Şekil 2.2 Farklı $\lambda$ ve $\nu$ değerleri için çarpık t dağılımın oyf grafikleri.....                    | 28  |
| Şekil 4.1 $\tau=1,2,\dots,200$ için $h(\tau)=E[\log(2\Phi_p(W_{11}))]$ değerleri (Contreras-Reyes 2014)..... | 68  |
| Şekil 4.2 Diyabet veri seti $y$ değişkenine ilişkin histogram .....                                          | 78  |
| Şekil 4.3 AIS veri seti $y$ değişkenine ilişkin histogram .....                                              | 80  |
| Şekil 6.1 Arazi kiralari veri seti $y$ değişkenine ilişkin histogram .....                                   | 131 |



## ÇİZELGELER DİZİNİ

|                                                                                                                                            |     |
|--------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Çizelge 2.1 M tahmin edicileri için kullanılan bazı fonksiyonlar .....                                                                     | 32  |
| Çizelge 2.2 Tukey'in amaç fonksiyonu ile farklı c değerleri için BDP ve asimptotik<br>görelî etkinlik değerleri .....                      | 39  |
| Çizelge 4.1 $p=5$ için olası tüm regresyon modelleri .....                                                                                 | 73  |
| Çizelge 4.2 AIC, $ICOMP_N$ ve $ICOMP_{SN}$ model seçim kriterlerinin doğru modeli<br>seçim sıklığı.....                                    | 74  |
| Çizelge 4.3 $ICOMP$ , $ICOMP_t$ ve $ICOMP_{St}$ model seçim kriterlerinin doğru modeli<br>seçim sıklığı.....                               | 76  |
| Çizelge 4.4 Diyabet veri seti değişkenleri .....                                                                                           | 77  |
| Çizelge 4.5 Diyabet veri seti için regresyon modelinde değişken seçimi sonuçları .....                                                     | 78  |
| Çizelge 4.6 AIS veri seti değişkenleri.....                                                                                                | 79  |
| Çizelge 4.7 AIS veri seti için regresyon modelinde değişken seçimi sonuçları .....                                                         | 80  |
| Çizelge 5.1 X yönünde aykırı değer olduğunda <i>RICOMP</i> kriterlerlerinin doğru<br>modeli seçme sıklıkları.....                          | 86  |
| Çizelge 5.2 Y yönünde aykırı değer olduğunda <i>RICOMP</i> kriterlerlerinin doğru<br>modeli seçme sıklıkları.....                          | 87  |
| Çizelge 5.3 X-Y yönünde aykırı değer olduğunda <i>RICOMP</i> kriterlerlerinin doğru<br>modeli seçme sıklıkları.....                        | 88  |
| Çizelge 5.4 Çoklu doğrusal bağlantı durumunda $RICOMP_R$ kriterlerlerinin doğru<br>modeli seçme sıklıkları.....                            | 97  |
| Çizelge 5.5 Çoklu doğrusal bağlantı ve x yönünde aykırı değer durumunda<br>$RICOMP_R$ kriterlerlerinin doğru modeli seçme sıklıkları ..... | 97  |
| Çizelge 5.6 Çoklu doğrusal bağlantı ve y yönünde aykırı değer durumunda<br>$RICOMP_R$ kriterlerlerinin doğru modeli seçme sıklıkları ..... | 98  |
| Çizelge 5.7 Çoklu doğrusal bağlantı ve x-y yönünde aykırı değer durumunda<br>seçim kriterlerinin doğru modeli seçme sıklıkları.....        | 98  |
| Çizelge 5.8 Aylık kazanç veri seti için regresyon modelinde değişken seçimi<br>sonuçları .....                                             | 100 |

|                                                                                                              |     |
|--------------------------------------------------------------------------------------------------------------|-----|
| Çizelge 5.9 Diyabet verilerine ilişkin VIF ve yüksek korelasyon değerleri<br>(Kirtland 2017).....            | 101 |
| Çizelge 5.10 Diyabet veri seti için regresyon modelinde değişken seçimi sonuçları ...                        | 102 |
| Çizelge 6.1 Çarpık-normal dağılımın JLSSM modeli için kriterlerin doğru<br>modeli seçme sıklığı.....         | 129 |
| Çizelge 6.2 Çarpık-t dağılımın JLSSM modeli için kriterlerin doğru modeli<br>seçme sıklığı.....              | 130 |
| Çizelge 6.3 Arazi kiralari veri seti için çarpık normal dağılımın JLSSM'de değişken<br>seçimi sonuçları..... | 131 |
| Çizelge 6.4 AIS veri seti değişkenleri.....                                                                  | 132 |
| Çizelge 6.5 AIS veri seti için çarpık-t dağılımın JLSSM'de değişken seçimi<br>sonuçları .....                | 133 |

## 1. GİRİŞ

Gerçek hayatta gözlenen bir olguyu/olayı karakterize etmek için oluşturulan istatistiksel modeller, rasgele değişkenler arasındaki ilişkinin matematiksel eşitliklerle ifade edilmesi olarak tanımlanabilir. Bu modeller istatistiksel sonuç çıkarımında ve geçmiş gözlemlere dayanarak gelecekle ilgili tahmin yapmada önemli rol oynamaktadır. Bu nedenlerle istatistiksel modelleme süreci uygun model yapısının belirlenmesi, parametre tahmini, alternatif modellerin varlığının tespiti ve bu modeller arasından uygun olan modelin seçimi işlemlerini içerir.

İstatistiksel modellemenin temellerinden biri, model parametrelerinin tahmin edilmesidir. Literatürde en sık kullanılan parametre tahmin yöntemlerinden ikisi, en küçük kareler (Least Squares - LS) yöntemi (Legendre 1805, Gauss 1809) ve en çok olabilirlik (Maximum Likelihood - ML) (Fisher 1922) yöntemidir. LS yönteminin popüler olmasının nedenlerinden bazıları, LS tahmin edicilerinin regresyon modelinde hataların normal dağıldığı varsayımı altında olduğu gibi bazı durumlarda parametrelerin ML tahmin edicileri ile aynı olması ve uzun zamandır çalışıldığı için LS yöntemine ait matematiksel araç ve hesaplama algoritmalarına ilişkin geniş bir literatürün var olmasıdır. Diğer taraftan LS yönteminin dezavantajlarından biri ise, veride az sayıda aykırı değer bile mevcut olduğunda tahmin edilen parametre değerlerinde büyük sapmalar olabilmesidir. LS yönteminden daha güçlü kuramsal özellikler gösteren ML yöntemi ise Fisher (1922) ile popüler olmuştur. Bazı düzgünlük koşulları altında ML tahmin edicisi tutarlıdır ve asimptotik olarak normal dağılmaktadır. Ancak bu cezbedici özelliklere sahip olan ML tahmin edicilerinin dağılım varsayımına dayanması, bu yöntemi dağılım varsayımlarından sapmalara karşı zayıf hale getirmektedir. Bununla birlikte, 20. yüzyıl boyunca istatistikçiler, özellikle normallik varsayımına dayanan istatistiksel yöntemlerin varsayılan modelden ufak sapmalara karşı bile çok hassas olduğunu fark etmiştir (Box 1953, Tukey 1960). Bu durumlar altında, kullanılan klasik istatistiksel yöntemlerle sağlıklı sonuçlar elde edilemeyeceğine dair literatürde birçok örnek bulunmaktadır (bkz Hampel vd. 1986, Maronna vd. 2006, Huber ve Ronchetti 2009).

Model varsayımlarının sağlanmaması ve/veya veri setinin aykırı değer içermesi gibi durumlardan mümkün olduğunca az etkilenen istatistiksel yöntemler araştırmacıların ilgi odağı olmuştur. Bu yöntemler robust (dayanıklı, sağlam) istatistiksel yöntemler olarak literatürdeki yerini almıştır. İstatistik literatüründe "robustluk" (robustness) kavramı teknik olarak ilk defa Box (1953) tarafından ortaya atılmıştır. Huber ise bu kavramı "varsayımlardan küçük sapmalara karşı duyarsızlık" olarak tanımlamaktadır. Dayanıklı istatistiksel yöntemler, varsayılan dağılımdan sapmalar ve/veya veride aykırı değer bulunması durumunda da güvenilir sonuçlar verecek şekilde tasarlanmış istatistiksel yöntemlerin sınıfı olarak tanımlanabilir. Huber'ın (1964) konum parametresinin M tahmin edicisini tanımlaması bu alanda yapılan temel çalışmalardan biridir. Huber'ın diğer çalışmaları (1964, 1967, 1973, 1977) ile literatür zenginleşmiştir. Tukey (1960, 1962) ve Hampel (1971, 1974) tarafından yapılan temel çalışmalarla ilk büyük adımlar atılmıştır. Jureckova (1971), Andrews (1974), Rousseeuw ve Yohai (1984), Rousseeuw ve Leroy (1987), Maronna (2006), Arslan (1996, 1997, 2012, 2016), Arslan vd. (2000) ve , Arslan vd. (2002) tarafından bu alanda önemli çalışmalar yapılmıştır. Dayanıklı istatistiksel yöntemler, yukarıda bahsedilen ve diğer araştırmacılar tarafından yapılan çalışmalar ile zengin bir literatüre sahip olmuştur. Huber (1981), Hampel vd. (1986), Rousseeuw ve Leroy (1987), Staudte ve Sheather (1990) ve Maronna (2006) tarafından yazılan kitaplar da literatürde önemli kaynaklar olarak yerlerini almıştır.

Model parametrelerinin tahmin edilmesinin ardından uygun model seçimi de istatistiksel modellemenin diğer bir temeli olarak ortaya çıkmıştır. Artan teknolojik gelişmeler daha çok değişkenin modele dahil olmasına, dolayısıyla daha karmaşık modellerin oluşturulmasına ve alternatif model sayısında artışa neden olur. Gereğinden fazla değişken ile çalışmak araştırmacı için maliyet ve zaman kaybına neden olur ve birden fazla uygun model tanımlamasının bulunduğu durumlarda, araştırmacı bu modellerden hangisinin daha iyi performansa sahip olduğunu belirlemek ister. Model seçiminde amaç, modelin yapısını (regresyon modellemesi, karma (mixture) modellemesi vs.) belirlemek değil, uygun model değerlendirme kriteri ile alternatif modeller arasından doğru modele en yakın modeli seçmektir. Bir başka deyişle model

seçimi problemi, değişkenlerin birbirine rakip alternatif tüm altkümeleri arasından veriyi en iyi karakterize eden, en basit altkümenin seçimi işlemidir. Böylece model daha iyi yorumlanabilir. Burada daha basit model, geleneksel olarak değişken sayısı az olan model anlamındadır.

Alternatifleri arasından en iyi modelin belirlenebilmesi için literatürde birçok seçim kriteri tanımlanmıştır. Bunlardan yaygın olarak bilinen ve kullanılan model seçim kriteri Akaike (1973, 1974) bilgi kriteri (AIC), bir aday model ile gerçek model arasındaki Kullback-Leibler (K-L) uzaklığının tahmin edilmesi esasına dayanmaktadır. Akaike (1973)'nin makalesi uzaklık ile model seçimi arasında bağlantının kurulduğu ilk makaledir. Bu çalışma yapılana kadar olabilirlik fonksiyonu literatürdeki boşluğu doldurmuştur. Doğrusal modeller çerçevesinde, Mallows (1964, 1973)'un  $C_p$  kriteri de bir başka popüler model seçim kriteridir.  $C_p$ , uygun bir aday model ile gerçek model arasındaki Gauss uyumsuzluğuna dayanmaktadır. AIC'nin küçük örneklem özellikleri doğrusal regresyon modelleri çerçevesinde Sugiura (1978) ve daha sonra Hurvich ve Tsai (1991) tarafından araştırılmış ve küçük örneklem için K-L uzaklığının geliştirilmiş bir tahmincisi ile AIC'nin düzeltilmiş bir şeklini (AICc) önermişlerdir. Daha yaygın olarak Bayesian bilgi kriteri olarak bilinen Schwarz (1978) bilgi kriteri (SBC-BIC-SIC), bir aday modelin önsel olasılığının bir dönüşümüne bir yaklaşım sağlamak üzere tasarlanmıştır. Bilindiği üzere model seçim kriterlerinin kullanılması için öncelikle parametre tahmininin hesaplanması gerekmektedir.

Model seçiminin bu zengin literatürü, bilgi kriterlerinin yanı sıra parametre tahminini ve değişken seçimini eş zamanlı yapan yöntemleri de içermektedir. Bu yöntemler LS regresyonda kullanılan amaç fonksiyonunun, parametrelerin bir fonksiyonu ile cezalandırılması olarak görülebilir. Bu yöntemler arasından LASSO (Tibshirani 1996), Bridge (Frank ve Friedman 1993, Fu 1998, Knight ve Fu 2000), Elastik Net (Zou ve Hastie 2005) bazı örneklerdir. Varsayımdan sapmalara ve aykırı değerlere karşı bahsedilen yöntemlerin robust genellemeleri literatürde yerini almıştır. Owen (2007), Wang vd. (2007) ve Li vd. (2011) tarafından yapılan çalışmalarda olduğu gibi genellikle robust M tahmin edicilerinin, LASSO yöntemindeki ceza terimi ile

cezalandırılmasıyla robust LASSO yöntemi önerilmiştir. Bridge tahmin edicilerini, Li ve Yu (2009) ve Li vd. (2011) M tahmin edicileri ile, Arslan (2016) ise MM tahmin edicileri bir araya getirerek robust Bridge yöntemini önermişlerdir.

AIC ve AIC tipindeki kriterler temelinde uyum iyiliği ve ceza terimi olmak üzere iki kısımdan oluşur. Burada ceza terimi, model karmaşıklığı olarak da adlandırılır ve bu kriterler için örneklem hacmine ve/veya aday modeldeki değişken sayısına bağlıdır. Ancak bu yaklaşımlar ile bağımsız değişkenin tahmin ediciye ne miktarda katkıda bulunduğu ve bağımsız değişkenlerin birbirleriyle ne kadar yakından ilişkili olduğu göz ardı edilmektedir. Karmaşıklık, sadece aday modelin veriye uygunluğunu değil, aynı zamanda bir modelin genelleştirilebilirliğini ve parametre tahmininde değişkenliği de etkiler. Bu nedenle, modelleri değerlendirirken model karmaşıklığı göz önüne alınmalıdır. Basit bir model, yeni veri kümelerini daha iyi karakterize eden ve dolayısıyla öngörü doğruluğunun daha yüksek olmasını sağlayan modeldir. Ayrıca, basit bir modelin davranışını tespit etmek daha kolaydır. Çünkü parametre tahminleri, karmaşık modellerin verilerine göre tekrarlanan veri bağlantılarından sonra daha kararlı olacaktır (Myung 2000). Karmaşıklık çeşitli şekillerde tanımlanabilir. Karmaşıklığı van Emden (1971) "bir bütünün parçalarının bileşiminden farklı olan yolu" olarak tanımlamıştır. Bir kovaryans matrisine uygulandığında karmaşıklık, kovaryans matrisi ile rastgele değişkenlerin etkileşimlerinin bir ölçüsüdür. Van Emden (1971) çok değişkenli normal dağılan bir rasgele vektörü göz önüne alarak kovaryans karmaşıklığının ilk matematiksel tanımını ileri sürmüştür. Bununla birlikte, bu tanımın orijinal rasgele değişkenlerin koordinatlarına bağlı olduğundan çalışmasında bahsetmiştir. Van Emden'in ardından Bozdogan (1990) bağımlı koordinatlı olmayan bir kovaryans karmaşıklığı tanımlamıştır (Alfonso-Reese vd. 2002). Bozdogan'a göre bir modelin karmaşıklığı yalnızca parametre sayısının bir fonksiyonu değil, aynı zamanda değişkenlerin birbirleri ile ne kadar ilgili olduğunun (birbirine bağımlılıklarının) bir ölçütüdür (Bozdogan 2000). Bozdogan (1988), bu fikirden yola çıkarak van Emden'in bilgi karmaşıklığı indeksine dayalı bir ceza fonksiyonu ile Bilgi Karmaşıklığı (Information COMpLexity-*ICOMP*) kriterini önermiştir. *ICOMP* tahmini bir modelin kovaryans karmaşıklığını cezalandırır.

Klasik parametre tahmin yöntemleri, robust tahmin yöntemlerine kıyasla aykırı değerlere, varsayımdan sapmalara karşı daha hassastır. Dolayısıyla klasik tahmin yöntemine dayalı model seçim kriterleri de bu durumlara karşı duyarlıdır. Bu nedenle literatürde klasik seçim kriterlerine alternatif olarak robust seçim kriterleri önerilmiştir. Bir robust model seçim kriterinin elde edilmesi için önerilecek bir yol, mevcut model seçim kriterlerinin robustlaştırılmasıdır. Dayanıklı model seçimi için yapılan çalışmalar incelendiğinde, Ronchetti'nin (1982) çalışması bu alanda önemli bir adım olmuştur. Ronchetti bu çalışmasında robust yaklaşımla klasik model seçim kriterlerine bir alternatif model seçim kriteri önermiştir. AICR olarak bilinen bu model seçim kriteri, AIC'nin ilk robust genelleşmesi olup, ML tahmin edicisinin M tahmin edicisine genelleştirilmesi gibi elde edilir. Hampel (1983), AICR'nin yeniden düzenlenmiş bir biçimini vermiştir. Ronchetti ve Staudte (1994) ise robust Cp kriterini tanımlamışlardır. Shi ve Tsai (1998) doğrusal regresyon için AIC'nin M tahminine dayalı üç farklı genellemesini önermiştir. Machado (1993) Schwarz kriterine dayalı bir robust model seçim kriteri önermiştir. Müller ve Welsh (2005) AIC'nin bir başka robust genelleşmesini vermiştir. Tharmaratnam vd. (2013) ise AIC bilgi kriterinin robust M, S ve MM tahmin edicilerine dayalı genellemelerinin bir karşılaştırmasını yapmıştır.

Bu tezin üç amacı vardır. İlki çoklu doğrusal regresyon modelinde yaygın olarak kullanılan normal dağılım varsayımının sağlanmayıp hataların kalın kuyruklu ve/veya çarpık olması durumu göz önüne alınarak Student  $t$ , çarpık normal ve çarpık  $t$  dağılım varsayımları altında model karmaşıklığının hesaplanarak *ICOMP* bilgi kriterinin bu durumlar için formunun elde edilmesidir. İkincisi, çoklu doğrusal regresyon modelinde hatalar üzerinde dağılım varsayımının olmadığı ve/veya veri setinin aykırı değerler içerdiği durum için *ICOMP* bilgi kriterinin robust formunun elde edilmesidir. Ayrıca yine çoklu doğrusal regresyon analizinde iki önemli problem olan çoklu iç ilişki ve/veya aykırı değer içerme durumlarına karşı *ICOMP* kriterinin bir robust formunun elde edilmesidir. Üçüncüsü ise, karmaşık verilerin modellenmesinde kullanılan ortak konum-ölçek ve ortak konum-ölçek-çarpıklık modellerinde *ICOMP* bilgi kriterinin formlarının elde edilmesidir.

İstatistiksel modeller arasında en yaygın kullanılanı regresyon modelidir. Regresyon modelleri, bir yanıt (açıklanan) değişkeninin bir veya daha fazla açıklayıcı değişkenle ilişkili olup olmadığını istatistiksel olarak belirlememizi sağlar. Regresyon modelinde çok sayıda açıklayıcı değişken olabilir ve bu değişken kümesinin hangi alt kümesinin en önemli sayılacağını bulmak en az model parametre tahmin problemi kadar önemlidir. Bu nedenle bu tezde çoklu doğrusal regresyon modellerinde, model seçim problemine yer verilmiştir.

Doğrusal regresyon modellerinde genellikle hataların normal dağıldığı varsayılmaktadır. Halbuki veri setlerinin kalın kuyruklu olduğu durumlar yaygındır. Bu durumda normal dağılımla her zaman mükemmel şekilde modelleme yapılamaz. Bununla beraber normal dağılıma dayalı istatistiksel çıkarımın, aykırı değerlere karşı savunmasız olduğu bilinmektedir (Lange vd. 1989). Hataların kalın kuyruklu olduğu durumlarda model parametrelerinin robust bir tahmin edicisini elde etmek için ilk adım Edgeworth'den (1887) gelmiştir. En Küçük Mutlak Sapmalar Yöntemi (Least Absolute Deviations - LAD ) olarak bilinen yöntem, hataların mutlak değerce toplamalarının minimize edilmesi esasına dayanmaktadır. Bu yöntem ile elde edilen tahmin ediciler, aynı zamanda hata terimlerinin Laplace dağıldığı varsayımıyla elde edilen ML tahmin edicilerine karşılık gelmektedir. Kalın kuyruklu bir dağılım varsayımı olduğunda kuyruktaki gözlemler, LS tahmin edicilerinin üzerinde aykırı gözlem etkisine sahip olacaktır. LS tahmin edicilerinin aykırı değerlere karşı hassas olduğu bilinmektedir. Bu hassasiyet, LS tahmin edicilerinin karesel bir amaç fonksiyonunun minimizasyonuna dayalı olmasından kaynaklanmaktadır. Veri seti aykırı değer içerdiğinde elde edilen büyük hata, karesel fonksiyonun da etkisiyle tahmin ediciyi büyük ölçüde etkilemektedir. Bu gibi durumlar için istatistikçiler, karesel fonksiyondan daha az hızla artan bir amaç fonksiyonuna ihtiyaç duymuştur. Hataların kalın kuyruklu olduğu varsayımı söz konusu olduğunda aykırı değerlerin tahmin ediciler üzerindeki olumsuz etkisi LAD regresyonda, klasik LS regresyondan olduğundan daha az olacaktır. Hatalara ilişkin kalın kuyruklu dağılım varsayımı için, Lange vd. (1989) tarafından ise normal regresyona alternatif olarak Student  $t$  regresyon önerilmiştir. Bunun yanı sıra finans, ekonomi, sosyoloji, iklim bilimi, çevre, mühendislik ve

biyomedikal bilim gibi birçok alanda, bazen regresyon modellerindeki hata yapıları simetri özelliğini sağlamaz (Lin ve Wang 2009). Bu nedenle son yıllarda normal dağılımı da içeren daha esnek dağılım ailelerinin tanımlanması büyük ilgi görmüştür. Bu ailelerden biri olan çarpık normal dağılım Azzalini (1985, 1986) tarafından önerilmiş ve regresyon analizinde uygulaması yine Azzalini (2005) tarafından incelenmiştir. Veri setinin hem kalın kuyruklu hem de çarpık olduğu durumlar için ise Azzalini ve Capitanio (2003) tarafından çarpık  $t$  dağılımı tanımlanmıştır.

Tüm bu ihtiyaçlar göz önüne alınarak Bölüm 2’de, çoklu doğrusal regresyon modelinde hata terimlerinin normal, Student  $t$ , Azzalini tipi çarpık normal ve çarpık  $t$  dağılımına sahip olduğu varsayımları göz önüne alınmış ve bu varsayımlar altında regresyon parametrelerinin ML tahminlerine yer verilmiştir. ML tahmin yönteminin dağılım varsayımına dayanması, tahmin edicilerin dağılım varsayımlarından ufak sapmalara karşın bile duyarlı olmasına sebep olmaktadır. Bu problem de göz önüne alınarak Bölüm 2’de çoklu doğrusal regresyon modeli için bazı robust tahmin yöntemlerine de yer verilmiştir. Bölüm 3’te ise literatürde yer alan bazı bilgi tabanlı model seçim kriterleri ve bu kriterlerin robust formları tanıtılmıştır.

Bölüm 4’te tezin birinci orijinal kısmında Student  $t$ , Azzalini tipi çarpık normal ve çarpık  $t$  dağılımları için kovaryans karmaşıklıkları elde edilmiştir. Elde edilen karmaşıklık yapıları ile beraber *ICOMP* bilgi kriterinin bu dağılım varsayımları altındaki formları elde edilmiştir.

Alışlagelen değişken seçim yöntemlerinin dağılım varsayımlarından sapma olması, verinin aykırı değerler içermesi gibi durumlardan daha çok etkilendikleri bilinmektedir. Diğer taraftan bu klasik yaklaşımlarda kullanılan model karmaşıklığı tanımlamalarının bahsedilen durumlarda yetersiz kalmaktadır. Tezin ikinci orijinal bölümü olan Bölüm 5’te, bu iki probleme bir çözüm olarak temelinde robust tahmin yöntemine dayanan ve diğer klasik kriterlerden farklı olarak model karmaşıklığını daha iyi temsil eden bir ceza terimine sahip olan robust *ICOMP* bilgi kriteri önerilmiştir.

Aykırı gözlemlerden daha az etkilenen robust yöntemler ve çoklu doğrusal bağlantı problemine çözüm olarak geliştirilen ridge regresyon (Hoerl ve Kennard 1970a) istatistiksel analizde önemli bir yere sahiptir. Ancak bir veri setinde hem aykırı değerler hem de çoklu doğrusal bağlantı problemi söz konusu olduğunda bu yöntemler yalnız başına güvenilir bir analiz için yeterli olmamaktadır. Bu iki problemin aynı anda çözülmesi, her iki problem için önerilen yöntemlerin birleştirilmesiyle mümkündür. Bu gibi durumlar için literatürde robust ridge tahmin edicileri önerilmiştir. Bölüm 5'te, robust ridge tahmin edicilerine dayalı hem aykırı değerlere karşı robust olacak hem de çoklu doğrusal bağlantı probleminden daha az etkilenecek *ICOMP* bilgi kriterinin bir başka robust formu önerilmiştir.

Karmaşık veri yapılarının tanımlanmasında çoğu zaman ortalama ve varyansın bazı değişkenlerden etkilendiği gerçeği ile karşı karşıya kalır. Bu gibi durumlar için önerilen ortak konum-ölçek modelleri (joint location-scale model-JLSM), heteroskedastik verileri modellemek için kullanışlı bir araç olarak bilinir ve çokça dikkat çekmiştir. İlk olarak Park (1966) değişen varyans için log-doğrusal model önermiş, Harvey (1976) konum ve ölçek parametrelerinin ML tahmin ve olabilirlik oran testlerini araştırmış, Aitkin (1987) ortak konum-ölçek modelinin parametrelerinin ML tahmin edicilerini araştırmıştır. Taylor ve Verbyla (2004), Student  $t$  dağılımının, Wu ve Li (2012) ters normal dağılımın konum ve ölçek parametrelerinin ortak modellenmesini önermiştir. Diğer taraftan çoğu var olan değişken seçim yöntemleri, sadece ortalama açıklama değişkenlerini seçmekle sınırlıdır. Ancak, varyansı modellemek, dolayısıyla varyans modelinde önemli değişkenleri tespit etmek, birçok durumda gözlemlerdeki değişkenliğin kaynağını belirlemede önemlidir. Zhang ve Wang, (2011), Wu ve Li (2012), Wu vd. (2013c) ortak konum-ölçek modellerinde değişken seçim problemine ilişkin çalışmalar yapmıştır. Bunun yanı sıra simetrik dağılımlara robust alternatif olarak çarpık dağılımlar ailelerinin tanımlanmasıyla çarpıklığın tahmin edilmesi, modellenmesi problemi ön plana çıkmıştır. Ortak konum-ölçek-çarpıklık modeli bu ihtiyaç sonucunda literatürdeki yerini almıştır. Çarpıklığın tahmin edilmesi, modellenmesi üzerine birçok çalışma olmasına rağmen çarpıklığın modellenmesinde değişken seçimi ile ilgili az sayıda çalışma bulunmaktadır.

Tezin üçüncü amacı ile ilgili elde edilen bulgular Bölüm 6’da verilmiştir. Bu bölümde konum-ölçek dağılımlar ailesinin ortak konum-ölçek modeli için *ICOMP* bilgi kriteri elde edilmiş ve konum-ölçek dağılımlar ailesine örnek olarak normal dağılım ve Student  $t$  dağılımı için konum-ölçek modellerinde *ICOMP* bilgi kriterinin formları elde edilmiştir. Ayrıca, çarpık simetrik dağılımların ortak konum-ölçek-çarpıklık modelinde değişken seçimi için *ICOMP* bilgi kriteri elde edilmiş ve çarpık simetrik dağılımlar ailesine örnek olarak çarpık normal dağılım ve çarpık  $t$  dağılımı için konum-ölçek-çarpıklık modellerinde *ICOMP* bilgi kriterinin formları elde edilmiştir.



## 2. REGRESYON MODELİ

Regresyon analizi, açıklanan değişken ile açıklayan değişken üzerinde etkisi olduğu varsayılan açıklayıcı değişken(ler) arasındaki ilişkinin modellenmesi, incelenmesi ve açıklanan değişkenin değerlerinin tahmin edilmesinde kullanılan bir istatistiksel tekniktir. Regresyon modeli, açıklanan ve açıklayıcı değişken sayısının bir ya da birden çok olması ve modelin doğrusal olup olmamasına bağlı olarak farklı şekillerde kurgulanır. Bu tezde çoklu doğrusal regresyon modeli ele alınmıştır.

Bu bölümde çoklu doğrusal regresyon modeline ilişkin tanımlara, varsayımlara, öncelikle farklı dağılım varsayımları altında model parametrelerinin ML tahmin edicilerine ve daha sonra robust M, S ve MM tahmin edicilerine yer verilmiştir.

### 2.1 Çoklu Doğrusal Regresyon Modeli

Doğrusal regresyon modelinin birden fazla açıklayıcı değişken içerdiği durumda, model çoklu doğrusal regresyon modeli olarak adlandırılır. Bu durumda çoklu doğrusal regresyon modeli kuramsal olarak aşağıdaki gibi ifade edilir:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p + \epsilon$$

Burada  $i = 0, 1, \dots, p$  için  $\beta_i$ 'ler model parametreleridir. Eğer model bir açıklayıcı değişken içeriyorsa basit doğrusal regresyon modeli olarak adlandırılır.

$\mathbf{x}_i^T = [1, x_{i1}, \dots, x_{ip}]$ ,  $\mathbf{X}$  tasarım matrisinin  $i$ . satırını göstermek üzere çoklu doğrusal regresyon modeli  $i$ . gözlem için

$$y_i = \mathbf{x}_i^T \boldsymbol{\beta} + \epsilon_i \quad (2.1)$$

biçiminde yazılır. Örneklem için çoklu doğrusal regresyon modeli, gösterim kolaylığı sağlamak amacıyla matris formunda  $\mathbf{y} = [y_1, y_2, \dots, y_n]^T$  açıklanan değişken

vektörünü,  $\mathbf{X} = [\mathbf{1} \ x_1 x_2 \dots x_p]$  tam sütun ranklı açıklayıcı değişkenler matrisini,  $\boldsymbol{\beta} = [\beta_0, \beta_1, \beta_2, \dots, \beta_p]^T$  parametre vektörünü ve  $\boldsymbol{\epsilon} = [\epsilon_1, \epsilon_2, \dots, \epsilon_n]^T$  hata vektörünü göstermek üzere

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \quad (2.2)$$

biçiminde ifade edilir.

Doğrusal regresyon modeli her istatistiksel yöntemde olduğu gibi bazı varsayımlar altında oluşturulur. Bunlar  $\boldsymbol{\epsilon}$  hata terimlerinin sıfır ortalamalı yani  $E(\boldsymbol{\epsilon}) = \mathbf{0}$ ,  $\mathbf{X}$  tasarım matrisinin stokastik olmayan bir matris yani  $E(\mathbf{X}^T \boldsymbol{\epsilon}) = \mathbf{0}$  ve  $\boldsymbol{\epsilon}$  hata terimlerinin  $\sigma^2$  sabit olmak üzere, kovaryans matrisinin  $Cov(\boldsymbol{\epsilon}) = \sigma^2 \mathbf{I}_{n \times n}$  olmasıdır. Bu varsayımlarla beraber, istatistiksel sonuç çıkarımı için dağılım varsayımına da ihtiyaç duyulmaktadır.

## 2.2 Çoklu Doğrusal Regresyon Modelinde Parametrelerin En Çok Olabilirlik Tahmin Edicileri

Çoklu doğrusal regresyon modelinin parametrelerini tahmin etmek için birçok yöntem kullanılabilir. Bu kısımda, dağılım varsayımı altında en yaygın kullanılan parametre tahmin yöntemi olan ML yöntemine yer verilmiştir.

ML tahmin edicileri asimptotik normallik, belli bir olasılık anlamında gerçek parametreye yakınsama gibi birçok optimal özelliğe sahip olduğundan stokastik sistemlerin modellemesinde, parametre kestirimi ve istatistiksel çıkarım için yaygın kullanılan bir yaklaşımdır. ML tahmin edicisi, olabilirlik fonksiyonu tek bir maksimum noktasına sahip ise fonksiyonu maksimum yapan parametre değeri olarak tanımlanır.

$X_1, X_2, \dots, X_n$  birbirinden bağımsız aynı  $f(x|\theta)$  olasılık (yoğunluk) fonksiyonuna sahip bir dağılımdan rasgele değişkenler ve  $\theta \in \Theta$  olsun.  $\mathbf{X} = (X_1, X_2, \dots, X_n)^T$  olmak

üzere  $L(\theta | \mathbf{X})$  olabilirlik fonksiyonunu gösterebilirsin. Bu durumda  $\theta$  parametresinin ML tahmin edicisi

$$\hat{\theta} = \arg \max_{\theta \in \Theta} L(\theta | \mathbf{X}) = \arg \max_{\theta \in \Theta} \prod_{i=1}^n f(X_i | \theta)$$

maksimizasyon probleminin çözülmesi ile veya buna denk olarak log-olabilirlik fonksiyonu  $\ell(\theta | \mathbf{X}) = \log L(\theta | \mathbf{X})$  olmak üzere

$$\hat{\theta} = \arg \max_{\theta \in \Theta} \ell(\theta | \mathbf{X}) = \arg \max_{\theta \in \Theta} \sum_{i=1}^n \log f(X_i | \theta)$$

maksimizasyon probleminin çözülmesi ile elde edilir. Matematiksel olarak bu maksimizasyon problemi genellikle log-olabilirlik fonksiyonunun bir kritik noktasının elde edilmesidir. Diğer bir deyişle, log-olabilirlik fonksiyonunun birinci türevi olan ve skor fonksiyonu olarak da adlandırılan

$$\ell'(\theta | \mathbf{X}) = \sum_{i=1}^n \frac{\partial}{\partial \theta} \log f(X_i | \theta)$$

fonksiyonunun bir köküdür. Burada  $\ell'(\theta | \mathbf{X})$ ,  $\ell(\theta | \mathbf{X})$  log-olabilirlik fonksiyonunun  $\theta$  parametresine göre türevini göstermektedir (Casella ve Berger 2002).

Birden fazla parametre olduğu durumda  $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_p)^T$  parametre vektörünün ML tahmin edicisi

$$\nabla \ell(\boldsymbol{\theta} | \mathbf{X}) = \begin{bmatrix} \frac{\partial}{\partial \theta_1} \ell(\boldsymbol{\theta} | \mathbf{X}) \\ \frac{\partial}{\partial \theta_2} \ell(\boldsymbol{\theta} | \mathbf{X}) \\ \vdots \\ \frac{\partial}{\partial \theta_p} \ell(\boldsymbol{\theta} | \mathbf{X}) \end{bmatrix} = \mathbf{0}$$

sisteminin çözümüdür.

Parametre tahmin problemlerinde, temel olasılık dağılımından gelen bir verinin bilinmeyen parametre hakkında içerdiği bilgi miktarı, parametre tahmininin kalitesinin bir ölçüsüdür. İstatistikte en yaygın kullanılan bilgi Fisher bilgisidir. Fisher bilgisi ilk defa bilginin matematiksel istatistikte kullanılmasını sağlamış ve özellikle parametre tahmininde kullanılmak üzere tanımlanmıştır (Kullback 1978).

Fisher bilgisi, bir dağılıma ait bir örneklemin parametre değeri hakkında sağladığı bilginin miktarını ölçen istatistiksel bir niceliktir. Fisher bilgisi olabilirlik prensibine dayanmaktadır. Bir kestiricinin varyansının yanı sıra ML tahminlerinin asimptotik davranışını ortaya koyan uygulama alanları vardır. Fisher bilgisi istatistiksel sonuç çıkarımında anahtar bir kavramdır ve aşağıdaki gibi tanımlanmıştır:

**Tanım 2.1** (Beklenen Fisher Bilgisi)  $X_1, X_2, \dots, X_n$  birbirinden bağımsız aynı  $f(x|\theta)$  olasılık yoğunluk fonksiyonuna (oyf) sahip rasgele değişkenler olsun.  $X_1, X_2, \dots, X_n$  örneklemine dayalı Fisher bilgisi

$$\mathcal{F}(\theta) = \text{Var}(\ell'(\theta | \mathbf{X}))$$

şeklinde tanımlıdır.

Fisher bilgisi, Tanım 2.1'e denk olarak

$$\mathcal{F}(\theta) = E \left[ (\ell'(\theta | \mathbf{X}))^2 \right]$$

veya  $l''_n(\theta | \mathbf{X})$ ,  $l_n(\theta | \mathbf{X})$  log-olabilirlik fonksiyonunun  $\theta$  parametresine göre ikinci türevini göstermek üzere

$$\mathcal{F}(\theta) = -E[\ell''(\theta | \mathbf{X})]$$

eşitlikleri ile elde edilebilir. Fisher bilgisi, beklenen Fisher bilgisi olarak da adlandırılır.

**Teorem 2.1**  $X_1, X_2, \dots, X_n$  birbirinden bağımsız olduğundan  $n$  birimlik örneklemdaki Fisher bilgisi bir birimlik örnekleminin  $n$  katıdır.

**Tanım 2.2** (Gözlenen Fisher Bilgisi) Gözlenen Fisher bilgisi negatif log-olabilirlik fonksiyonunun ikinci türevi ile

$$\hat{\mathcal{F}}_n(\theta) = - \sum_{i=1}^n \ell''(\theta | x_i) = - \sum_{i=1}^n \frac{\partial^2}{\partial \theta^2} \ell(\theta | x_i)$$

şeklinde tanımlanır.

**Teorem 2.2** Büyük sayılar yasaından

$$\frac{1}{n} \widehat{\mathcal{F}}_n(\theta) \xrightarrow{P} -E[\ell''(\theta | \mathbf{X})] = \mathcal{F}(\theta)$$

gerçeklenir.

**Teorem 2.3**  $\theta$  parametresinin ML tahmin edicisi  $\widehat{\theta}$ , bazı düzgünlük koşulları altında asimptotik normal dağılır:

$$\sqrt{n}(\widehat{\theta} - \theta) \xrightarrow{d} N(0, \mathcal{F}^{-1}(\theta))$$

Görüldüğü üzere, Fisher bilgisi büyüdükçe tahmin edicinin varyansı küçülmektedir.

Birden fazla parametre olduğu durumda  $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_p)^T$  parametre vektörünün  $\widehat{\boldsymbol{\theta}} = (\widehat{\theta}_1, \widehat{\theta}_2, \dots, \widehat{\theta}_p)^T$  ML tahmin edicisine ilişkin Fisher bilgisi  $p \times p$  boyutlu bir matristir ve bu matris Fisher bilgi matrisi olarak adlandırılır.

**Tanım 2.3**  $\mathbf{X} = (X_1, X_2, \dots, X_n)$  rastgele örneklem olsun ve  $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_p)^T$  parametre vektörü olmak üzere  $f(\mathbf{x}|\boldsymbol{\theta})$  bu örnekleme ilişkin olasılık yoğunluk fonksiyonunu göstereyin.  $\mathcal{F}(\boldsymbol{\theta})$  Fisher bilgi matrisi log-olabilirlik fonksiyonunun birinci kısmi türevleri arasındaki kovaryanslar biçiminde tanımlanan  $p \times p$  tipindeki simetrik matristir ve  $(i, j)$ . terimi

$$\mathcal{F}(\boldsymbol{\theta})_{i,j} = \text{Cov} \left[ \frac{\partial \log f(\mathbf{X}|\boldsymbol{\theta})}{\partial \theta_i}, \frac{\partial \log f(\mathbf{X}|\boldsymbol{\theta})}{\partial \theta_j} \right]$$

şeklindedir.

Bu tanıma denk ve alternatif bir başka tanım ise

$$\mathcal{F}(\boldsymbol{\theta})_{i,j} = -E \left[ \frac{\partial^2 \log f(\mathbf{X}|\boldsymbol{\theta})}{\partial \theta_i \partial \theta_j} \right] \quad (2.3)$$

yani log-olabilirlik fonksiyonunun ikinci kısmi türevlerinin negatif beklenen değerleri olarak verilir (Myung ve Navarro 2004).

**Tanım 2.4** Gözlenen Fisher bilgi matrisi negatif log-olabilirlik fonksiyonunun ikinci türev matrisidir ve  $(i, j)$ . terimi

$$\hat{\mathcal{F}}_n(\boldsymbol{\theta})_{i,j} = - \sum_{k=1}^n \frac{\partial^2 \log f(x_k|\boldsymbol{\theta})}{\partial \theta_i \partial \theta_j}$$

eşitliği ile verilir.

Gözlenen Fisher bilgi matrisi, beklenen Fisher bilgi matrisinin tutarlı tahmin edicisidir.

**Teorem 2.4**  $\boldsymbol{\theta}$  parametre vektörünün  $\hat{\boldsymbol{\theta}}$  ML tahmin edicisi asimptotik yansızdır. Bazı düzgünlük koşulları altında  $\hat{\boldsymbol{\theta}}$ 'nın örneklem dağılımı, kovaryans matrisi örneklem hacmi 1 olan ters Fisher bilgi matrisi ile elde edilmek üzere asimptotik normaldir, yani  $n \rightarrow \infty$  iken  $\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{D} N(\mathbf{0}, \mathcal{F}^{-1}(\boldsymbol{\theta}))$  gerçekleşir.

### 2.2.1 Hataların normal dağıldığı durum

Regresyon analizinde istatistiki sonuç çıkarımı sürecinde, hatalara ilişkin dağılım varsayımlarından en yaygını normallik varsayımıdır.

### Regresyon Parametre Tahmini

Eşitlik (2.2) ile verilen çoklu doğrusal regresyon modelinde  $i = 1, 2, \dots, n$  için  $\epsilon_i$  hata terimlerinin birbirinden bağımsız, sıfır ortalamalı ve  $\sigma^2$  sabit varyanslı normal dağılıma sahip oldukları, başka bir deyişle  $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$  olduğu varsayalım. Bu durumda  $\epsilon_i$  hata terimlerinin olasılık yoğunluk fonksiyonu

$$f(\epsilon_i) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}\epsilon_i^2\right), \quad i = 1, 2, \dots, n$$

biçimindedir.  $i = 1, 2, \dots, n$  için  $\mathbf{x}_i^T = [1 \ x_{i1} \ x_{i2} \ \dots \ x_{ip}]$  olmak üzere  $\epsilon_i = y_i - \mathbf{x}_i^T \boldsymbol{\beta}$  olduğu dikkate alındığında olabilirlik fonksiyonu

$$L(\boldsymbol{\beta}, \sigma^2 | \mathbf{y}, \mathbf{X}) = \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mathbf{x}_i^T \boldsymbol{\beta})^2\right)$$

ve log-olabilirlik fonksiyonu

$$\ell(\boldsymbol{\beta}, \sigma^2 | \mathbf{y}, \mathbf{X}) = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mathbf{x}_i^T \boldsymbol{\beta})^2 \quad (2.4)$$

şeklinde bulunur.  $\arg \max_{\boldsymbol{\beta}, \sigma^2} \ell(\boldsymbol{\beta}, \sigma^2 | \mathbf{y}, \mathbf{X})$  maksimizasyon probleminin çözümüne, log-olabilirlik fonksiyonunun parametrelere göre türevlerinin alınıp sıfıra eşitlenmesi ile ulaşılabilir:

$$\begin{cases} \frac{\partial \ell}{\partial \boldsymbol{\beta}} = \frac{1}{\sigma^2} \sum_{i=1}^n (y_i - \mathbf{x}_i^T \boldsymbol{\beta}) \mathbf{x}_i = \mathbf{0} \\ \frac{\partial \ell}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2} \sum_{i=1}^n (y_i - \mathbf{x}_i^T \boldsymbol{\beta})^2 = 0 \end{cases} \quad (2.5)$$

(2.5) denklem sisteminden parametrelerin çekilmesiyle  $\hat{\boldsymbol{\beta}}$  ve  $\hat{\sigma}$ ,  $\boldsymbol{\beta}$  ve  $\sigma$  parametrelerinin ML tahmin edicilerini göstermek üzere

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= \left( \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i^T \right)^{-1} \sum_{i=1}^n y_i \mathbf{x}_i \\ \hat{\sigma}^2 &= \frac{1}{n} \sum_{i=1}^n (y_i - \mathbf{x}_i^T \hat{\boldsymbol{\beta}})^2 \end{aligned}$$

bulunur. Bu tahmin ediciler matris gösterimi ile

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} \\ \hat{\sigma}^2 &= \frac{(\mathbf{y} - \mathbf{X} \hat{\boldsymbol{\beta}})^T (\mathbf{y} - \mathbf{X} \hat{\boldsymbol{\beta}})}{n} \end{aligned}$$

biçiminde yazılabilir.  $\mathbf{X}$  açıklayıcı değişkenlerin matrisi tam ranklı ise  $(\mathbf{X}^T \mathbf{X})^{-1}$  matrisi her zaman hesaplanabilir (Montgomery vd. 2001).

Doğrusal regresyon modelinin parametrelerinin tahmininde kullanılan çeşitli yöntemler arasında yaygın kullanılan bir diğer tahmin yöntemi ise en küçük kareler (Least Squares - LS) yöntemidir.  $\boldsymbol{\beta}$  parametre vektörünün LS tahmin edicisi

$$\hat{\boldsymbol{\beta}}_{LS} = \arg \min_{\boldsymbol{\beta}} \sum_{i=1}^n \epsilon_i^2 = \arg \min_{\boldsymbol{\beta}} \sum_{i=1}^n (y_i - \mathbf{x}_i^T \boldsymbol{\beta})^2$$

minimizasyon probleminin çözülmesiyle

$$\hat{\boldsymbol{\beta}}_{LS} = (\mathbf{X}^T \mathbf{X})^{-1} (\mathbf{X}^T \mathbf{y})$$

olarak elde edilir.  $\sigma$  parametresinin LS tahmin edicisi ise

$$\hat{\sigma}_{LS}^2 = \frac{(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{LS})^T (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{LS})}{n - (p + 1)}$$

biçimindedir.  $\hat{\boldsymbol{\beta}}_{LS}$  ve  $\hat{\sigma}_{LS}$  tahmin edicileri  $\boldsymbol{\beta}$  ve  $\sigma$  için yansız tahmin edicilerdir.

(2.4) eşitliği ile verilen log-olabilirlik fonksiyonunun parametreye göre ikinci türevleri

$$\begin{aligned} \frac{\partial^2 \ell}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} &= -\frac{1}{\sigma^2} \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i^T \\ \frac{\partial^2 \ell}{\partial (\sigma^2)^2} &= \frac{n}{2(\sigma^2)^2} - \frac{1}{(\sigma^2)^3} \sum_{i=1}^n (y_i - \mathbf{x}_i^T \boldsymbol{\beta})^2 \\ \frac{\partial^2 \ell}{\partial \boldsymbol{\beta} \partial \sigma^2} &= \frac{\partial^2 l}{\partial \sigma^2 \partial \boldsymbol{\beta}} = -\frac{1}{2\sigma^4} \sum_{i=1}^n (y_i - \mathbf{x}_i^T \boldsymbol{\beta}) \mathbf{x}_i \end{aligned}$$

olmak üzere,

$$\mathcal{F}(\boldsymbol{\beta}, \sigma^2) = -E \left( \begin{bmatrix} \frac{\partial^2 l}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} & \frac{\partial^2 l}{\partial \boldsymbol{\beta} \partial \sigma^2} \\ \frac{\partial^2 l}{\partial \sigma^2 \partial \boldsymbol{\beta}} & \frac{\partial^2 l}{\partial (\sigma^2)^2} \end{bmatrix} \right)$$

matrisinde ikinci türevlerin yerine yerine yazılmasıyla hataların normal dağıldığı durumda regresyon parametreleri için elde edilen Fisher bilgi matrisi

$$\mathcal{F}(\boldsymbol{\beta}, \sigma^2) = \begin{bmatrix} \frac{1}{\sigma^2} \mathbf{X}^T \mathbf{X} & 0 \\ 0 & \frac{n}{2\sigma^4} \end{bmatrix}$$

biçimindedir. Burada beklenen değer hata dağılımının varsayımı altında hesaplanmaktadır.

Hataların normal dağıldığı durum için parametrelerin ML tahmin edicilerinin kovaryans matrisinin tahmini *IFIM* (Inverse Fisher Information Matrix), Fisher bilgi matrisinin tersini göstermek üzere

$$IFIM(\hat{\boldsymbol{\beta}}, \hat{\sigma}) = \widehat{Cov}(\hat{\boldsymbol{\beta}}, \hat{\sigma}) = \hat{\mathcal{F}}^{-1}(\hat{\boldsymbol{\beta}}, \hat{\sigma}) = \begin{bmatrix} \hat{\sigma}^2 (\mathbf{X}^T \mathbf{X})^{-1} & 0 \\ 0 & \frac{2\hat{\sigma}^4}{n} \end{bmatrix}$$

şeklinde elde edilir.

### 2.2.2 Hataların Student $t$ dağıldığı durum

Uygulamalar gerçek verilerin modellenmesinde kalın kuyruklu simetrik dağılımların normal dağılıma kıyasla daha elverişli olduğunu göstermektedir. Normal dağılım kalın kuyruklu verilerin modellenmesi için uygun değildir. Bu gibi durumlar için normal regresyona alternatif olarak Lange vd. (1989) doğrusal ve doğrusal olmayan regresyon modelleri için normal dağılım gibi simetrik ve çan şeklinde ancak daha kalın kuyruklu olan Student  $t$  dağılımını kullanarak sağlam istatistiksel modellemeyi önermiştir.

Student  $t$  dağılımı ilk olarak Gosset (1908) tarafından "bir ortalamanın olası hatası" üzerine yaptığı çalışmada geliştirilmiştir. Daha sonra Fisher (1925) ve diğerlerinin katkılarıyla dağılımla ilgili gelişmeler devam etmiştir (Ahsanullah vd. 2014).

**Tanım 2.5** Student  $t$  dağılımının olasılık yoğunluk fonksiyonu  $\mu \in \mathbb{R}$ ,  $\sigma > 0$  sırasıyla konum ve ölçek parametrelerini,  $\nu > 0$  dağılımın şeklini belirleyen serbestlik derecesini göstermek üzere

$$f(x; \mu, \sigma, \nu) = \frac{c_\nu}{\sigma} \left( \nu + \frac{(x - \mu)^2}{\sigma^2} \right)^{-\frac{\nu+1}{2}}, \quad x \in \mathbb{R}$$

şeklinde tanımlıdır. Burada  $\Gamma(\cdot)$  gamma fonksiyonu olmak üzere  $c_\nu = \frac{\Gamma(\frac{\nu+1}{2})\nu^{\nu/2}}{\sqrt{\pi}\Gamma(\frac{\nu}{2})}$  biçimindedir ve  $X \sim t_\nu(\mu, \sigma)$  ile gösterilir.

Student  $t$  dağılımının bazı özellikleri aşağıda verilmiştir:

- (i)  $\nu > 1$  için  $E(X) = \mu$  ve  $\nu = 1$  için  $E(X)$  mevcut değildir.
- (ii)  $\nu > 2$  için  $Var(X) = \frac{\nu\sigma^2}{\nu-2}$  ve  $\nu \leq 2$  için  $Var(X)$  mevcut değildir.
- (iii)  $\nu = 1$  için özel hali Cauchy dağılımıdır.
- (iv)  $\nu \rightarrow \infty$  iken Student  $t$  dağılımı normal dağılıma yakınsar (Ahsanullah vd. 2014).

Hataların Student  $t$  dağıldığı varsayımı altında, (2.2) ile verilen regresyon modeli göz önüne alınsın. Burada  $\nu$  serbestlik derecesinin bilindiği varsayılmaktadır. Aksi halde  $\nu$  parametresi tahmin edilmelidir.

## Regresyon Parametre Tahmini

$i = 1, 2, \dots, n$  için  $\epsilon_i \sim t_\nu(0, \sigma)$  olmak üzere olabilirlik ve log-olabilirlik fonksiyonları sırasıyla

$$L(\boldsymbol{\beta}, \sigma | \mathbf{y}, \mathbf{X}) = \left(\frac{c_\nu}{\sigma}\right)^n \prod_{i=1}^n \left(v + \frac{(y_i - \mathbf{x}_i^T \boldsymbol{\beta})^2}{\sigma^2}\right)^{-\frac{\nu+1}{2}}$$

$$\ell(\boldsymbol{\beta}, \sigma | \mathbf{y}, \mathbf{X}) = n \ln(c_\nu) - n \ln \sigma - \frac{\nu+1}{2} \sum_{i=1}^n \ln \left(v + \frac{(y_i - \mathbf{x}_i^T \boldsymbol{\beta})^2}{\sigma^2}\right) \quad (2.6)$$

biçimindedir.

Log-olabilirlik fonksiyonunun  $\boldsymbol{\beta}$  ve  $\sigma$  parametrelerine göre kısmi türevlerinin alınıp sıfıra eşitlenmesiyle

$$\begin{cases} \frac{\partial \ell}{\partial \boldsymbol{\beta}} = \frac{\nu+1}{\sigma^2} \sum_{i=1}^n \frac{(y_i - \mathbf{x}_i^T \boldsymbol{\beta}) \mathbf{x}_i}{\left[\nu + \frac{(y_i - \mathbf{x}_i^T \boldsymbol{\beta})^2}{\sigma^2}\right]} = \mathbf{0} \\ \frac{\partial \ell}{\partial \sigma} = -\frac{n}{\sigma} + \frac{\nu+1}{\sigma^3} \sum_{i=1}^n \frac{(y_i - \mathbf{x}_i^T \boldsymbol{\beta})^2}{\left[\nu + \frac{(y_i - \mathbf{x}_i^T \boldsymbol{\beta})^2}{\sigma^2}\right]} = 0 \end{cases}$$

denklem sistemi elde edilir. Bu eşitliklerin yeniden düzenlenmesiyle  $\boldsymbol{\beta}$  ve  $\sigma$  parametrelerinin ML tahminleri

$$\hat{\boldsymbol{\beta}} = \left[ \sum_{i=1}^n \omega_i \mathbf{x}_i \mathbf{x}_i^T \right]^{-1} \left[ \sum_{i=1}^n \omega_i \mathbf{x}_i^T y_i \right] \quad (2.7)$$

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n \omega_i (y_i - \mathbf{x}_i^T \hat{\boldsymbol{\beta}})^2 \quad (2.8)$$

eşitlikleri kullanılarak sayısal hesaplamalar ile elde edilir. Burada  $\omega_i$  ağırlıkları

$$\omega_i = \frac{\nu+1}{\left[\nu + \frac{(y_i - \mathbf{x}_i^T \hat{\boldsymbol{\beta}})^2}{\hat{\sigma}^2}\right]}$$

biçiminde tanımlıdır ve  $z_i^2 = \frac{(y_i - \mathbf{x}_i^T \hat{\boldsymbol{\beta}})^2}{\hat{\sigma}^2}$  karesel uzaklığının azalan bir fonksiyonu olarak görülebilir. Yani aykırı değerlere ilişkin uzaklık ( $z_i^2$ ) arttıkça, o gözleme ilişkin  $\omega_i$  ağırlığı azalmaktadır. Bu sayede aykırı değer tahmin edici üzerindeki etkisi azalır.

Dolayısıyla  $\nu < \infty$  için parametrelerin ML tahminleri robusttur. Normal regresyonda ( $\nu = \infty$  durumunda) ise ağırlıklar 1 olacağından, normallik varsayımı altında elde edilen ML tahmin edicileri aykırı değerlere karşı hassas olur. Ayrıca  $\nu$  serbestlik derecesinin artması (azalması), ağırlıkların benzer (farklı) değerler almasına ve dolayısıyla aykırı değerlere karşı daha fazla (daha az) duyarlılığa neden olmaktadır. Bu nedenle  $\nu$  serbestlik derecesi "robustluk parametresi" olarak görülebilir.

### Fisher Bilgi Matrisi

Eşitlik (2.6) ile verilen log-olabilirlik fonksiyonunun ikinci kısmi türevleri

$$\frac{\partial^2 \ell}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} = \sum_{i=1}^n \left( \frac{2}{\sigma^4 (\nu + 1)} \omega_i (y_i - \mathbf{x}_i^T \boldsymbol{\beta})^2 - \frac{1}{\sigma^2} \right) \omega_i \mathbf{x}_i \mathbf{x}_i^T \quad (2.9)$$

$$\frac{\partial^2 \ell}{\partial \sigma^2} = \frac{n}{\sigma^2} - \frac{3}{\sigma^2} \sum_{i=1}^n \omega_i \frac{(y_i - \mathbf{x}_i^T \boldsymbol{\beta})^2}{\sigma^2} + \omega_i^2 \frac{(y_i - \mathbf{x}_i^T \boldsymbol{\beta})^4}{\sigma^4 (\nu + 1)} \quad (2.10)$$

$$\frac{\partial^2 \ell}{\partial \sigma \partial \boldsymbol{\beta}} = \left( -\frac{2}{\sigma^3} \sum_{i=1}^n \omega_i (y_i - \mathbf{x}_i^T \boldsymbol{\beta}) + \frac{2}{\sigma^2 (\nu + 1)} \sum_{i=1}^n \omega_i^2 \frac{(y_i - \mathbf{x}_i^T \boldsymbol{\beta})^3}{\sigma^3} \right) \mathbf{x}_i \quad (2.11)$$

olarak elde edilir. Burada  $\nu$  serbestlik derecesinin bilindiği varsayılmaktadır, bu nedenle Fisher bilgi matrisi  $2 \times 2$  boyutludur. Aksi halde  $\nu$ , parametre olarak ele alınmalı, Fisher bilgi matrisinde log-olabilirlik fonksiyonunun  $\nu$  parametresine göre kısmi türevleri yer almalıdır ve Fisher bilgi matrisinin boyutu  $3 \times 3$  olmalıdır.

Fisher bilgi matrisinin elde edilmesi için (2.9)-(2.11) eşitliklerinin (2.3) eşitliğinde yerine yazılıp ilgili beklenen değerlerin hesaplanması gerekmektedir. Burada önemli bir husus beklenen değerlerin hataların dağılımına göre, yani burada Student  $t$  dağılımına göre alınmasıdır. Buradan Fisher bilgi matrisi

$$\mathcal{F}(\boldsymbol{\beta}, \sigma^2) = \begin{bmatrix} \frac{1}{\sigma^2} \frac{\nu+1}{\nu+3} \mathbf{X}^T \mathbf{X} & 0 \\ 0 & \frac{\nu}{\nu+3} \frac{n}{2\sigma^4} \end{bmatrix} \quad (2.12)$$

olarak elde edilir (Arslan 1993, Fonseca vd. 2008).

Hataların Student  $t$  dağıldığı durum için parametrelerin ML tahmin edicilerinin kovaryans matrisinin tahmini aşağıdaki gibi elde edilir:

$$\hat{\mathcal{F}}^{-1} = \begin{bmatrix} \frac{\nu+3}{\nu+1} \hat{\sigma}^2 (\mathbf{X}^T \mathbf{X})^{-1} & 0 \\ 0 & \frac{\nu+3}{\nu} \frac{2\hat{\sigma}^4}{n} \end{bmatrix}$$

### 2.2.3 Hataların çarpık normal dağıldığı durum

Gerçek hayatta birçok uygulamada veri seti beklendiği gibi simetrik olmayabilir. Bu gibi durumlarda simetrik bir dağılım varsayımı altında istatistiksel çıkarımlar yanıltıcı olabilir. Bu probleme farklı bakış açısı getirmek için simetrik dağılımlara alternatif olarak çarpık dağılımlar kullanılabilir. Dağılımların çarpıklaştırılması ile ilgili literatürdeki önemli bazı öneriler Birnbaum (1950), Nelson (1964), Weinstein (1964), Roberts (1966), O'Hagan ve Leonard (1976) tarafından ortaya atılmakla beraber Azzalini (1985) tarafından önerilen çarpıklaştırma yöntemi oldukça ilgi görmüştür.

Azzalini (1985) çarpık normal dağılım teorisini genişletmiştir. Çarpık normal dağılımın özellikleri Azzalini (1986) tarafından incelenmiştir. Çok değişkenli çarpık normal dağılım ise benzer düşünce ile Azzalini ve Dalle Valle (1996) tarafından önerilmiş ve Azzalini ve Capitanio (1999) tarafından özellikleri incelenmiştir. Çarpık normal dağılımlar üzerine yazdıkları kitap ile Azzalini ve Capitanio (2014) literatüre zengin bir kaynak sunmuşlardır.

Bu kısımda, çoklu regresyon modelinde hata terimlerinin Azzalini (1985) tipi çarpık normal dağılımına sahip olduğu durum ele alınmıştır.

Azzalini (1985) simetrik dağılımları çarpıklaştırmak için Teorem 2.5'i önermiştir.

**Teorem 2.5**  $f$  sıfır etrafında simetrik olasılık yoğunluk fonksiyonu ve  $G'$  sıfır etrafında simetrik olacak şekilde  $G$  mutlak sürekli dağılım fonksiyonu olmak üzere

$$h(x) = 2f(x)G(\lambda x), x \in \mathbb{R}$$

şeklinde tanımlanan  $h(\cdot)$  fonksiyonu her  $\lambda \in \mathbb{R}$  için olasılık yoğunluk fonksiyonu özelliklerini taşır.

**İspat.**  $X$  ve  $Y$  olasılık yoğunluk fonksiyonları sırasıyla  $f$  ve  $G'$  olacak şekilde birbirinden bağımsız iki rasgele değişken olsun. Bu durumda

$$\int_{-\infty}^{\infty} 2f(x)G(\lambda x)dx = 2E_X(P\{Y < \lambda x | X = x\}) = 1$$

elde edilir (Azzalini 1985). ■

**Tanım 2.6**  $\phi(\cdot)$  ve  $\Phi(\cdot)$  fonksiyonları sırasıyla standart normal dağılımın olasılık yoğunluk ve dağılım fonksiyonlarını göstermek üzere,  $Z$  rasgele değişkeni aşağıdaki olasılık yoğunluk fonksiyonuna sahip ise  $\lambda$  çarpıklık parametrelili çarpık normal dağılıma sahiptir denir ve  $Z \sim SN(\lambda)$  ile gösterilir:

$$f(z; \lambda) = 2\phi(z)\Phi(\lambda z), z \in \mathbb{R}$$

Azzalini tipi çarpık normal dağılımın bazı özellikleri şunlardır:

- (i)  $\lambda > 0$  ise dağılım sağa,  $\lambda < 0$  ise dağılım sola çarpıktır.
- (ii) Özel olarak dağılım,  $\lambda = 0$  ise normal dağılıma ve  $\lambda \rightarrow \infty$  iken yarı normal (half-normal) dağılıma dönüşmektedir.
- (iii)  $Z \sim SN(\lambda)$  ise  $-Z \sim SN(-\lambda)$  dır.
- (iv)  $Z \sim SN(\lambda)$  ise  $Z^2 \sim \chi_1^2$  dir (Azzalini 1985).

Dağılımı uygulamada daha kullanışlı bir hale getirmek için  $Y = \xi + \sigma X$  dönüşümü uygulanarak dağılım konum ölçek ailesine taşınır. Bu dönüşüm altında  $\xi \in \mathbb{R}$  konum parametresi,  $\sigma > 0$  ölçek parametresi ve  $\lambda \in \mathbb{R}$  çarpıklık parametresi olmak üzere çarpık normal dağılımın olasılık yoğunluk fonksiyonu,

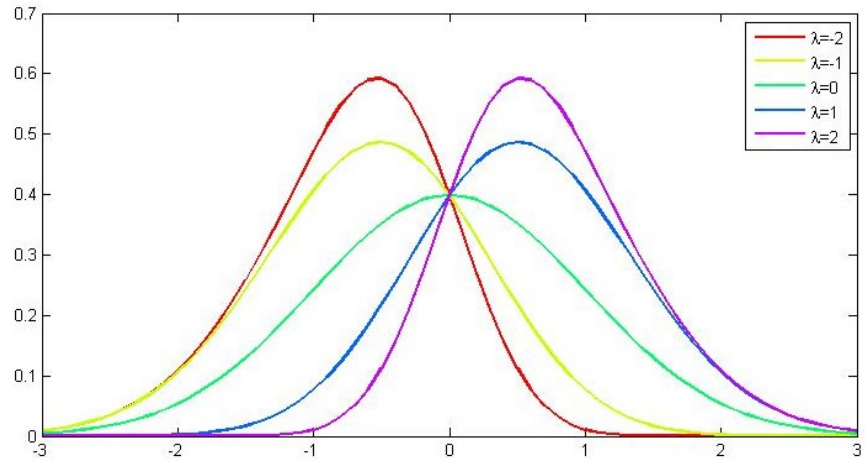
$$f(y; \xi, \sigma, \lambda) = \frac{2}{\sigma} \phi\left(\frac{y - \xi}{\sigma}\right) \Phi\left(\lambda \left(\frac{y - \xi}{\sigma}\right)\right), \quad y \in \mathbb{R}$$

şeklinde ifade edilir ve  $Y \sim SN(\xi, \sigma^2, \lambda)$  biçiminde gösterilir.

$Y \sim SN(\xi, \sigma^2, \lambda)$  ise beklenen değer, varyans, basıklık ve çarpıklık katsayıları sırasıyla aşağıdaki gibidir :

$$\begin{aligned} E(Y) &= \xi + \sqrt{\frac{2}{\pi}} \frac{\lambda \sigma}{\sqrt{\lambda^2 + 1}} \\ Var(Y) &= \left(1 - \frac{2}{\pi} \frac{\lambda^2}{\lambda^2 + 1}\right) \sigma^2 \\ \gamma_1 &= \frac{\sqrt{2}(4 - \pi) \lambda^3}{(\pi + (\pi - 2) \lambda^2)^{3/2}} \\ \gamma_2 &= 3 + \frac{8(\pi - 3) \lambda^4}{(\pi + (\pi - 2) \lambda^2)^2} \end{aligned}$$

Şekil 2.1’de  $\lambda = 0, \pm 1, \pm 2$  değerleri için çarpık normal dağılımın olasılık yoğunluk fonksiyonunun grafikleri verilmiştir.  $\lambda$  çarpıklık parameresinin dağılımın şeklini ne ölçüde değiştirdiği,  $|\lambda|$  büyüdükçe dağılımın çarpıklığının arttığı şekil 2.1’de açıkça görülmektedir.



Şekil 2.1 Farklı  $\lambda$  değerleri için çarpık normal dağılımın oyf grafikleri

Eşitlik (2.2) ile verilen çoklu doğrusal regresyon modelinde  $\epsilon_i$  hata terimlerinin bağımsız, aynı dağılımlı, sıfır ortalamalı,  $\sigma$  bilinmeyen ölçek parametrelili ve  $\lambda$  çarpıklık parametresi ile çarpık normal dağıldığı varsayalım.

## Regresyon Parametre Tahmini

$\epsilon_i$  rasgele hata teriminin olasılık yoğunluk fonksiyonu

$$f(\epsilon_i; \sigma, \lambda) = \frac{2}{\sigma} \phi\left(\frac{\epsilon_i}{\sigma}\right) \Phi\left(\lambda \left(\frac{\epsilon_i}{\sigma}\right)\right)$$

olup buradan olabilirlik fonksiyonu

$$L(\boldsymbol{\beta}, \sigma, \lambda | \boldsymbol{\epsilon}) = \left(\frac{2}{\pi\sigma^2}\right)^{n/2} \exp\left(-\frac{1}{2} \sum_{i=1}^n \left(\frac{\epsilon_i}{\sigma}\right)^2\right) \prod_{i=1}^n \Phi\left(\lambda \left(\frac{\epsilon_i}{\sigma}\right)\right)$$

biçiminde elde edilir.

$\epsilon_i = y_i - \mathbf{x}_i^T \boldsymbol{\beta}$  olduğu dikkate alındığında olabilirlik ve log-olabilirlik fonksiyonları sırasıyla

$$\begin{aligned} L(\boldsymbol{\beta}, \sigma, \lambda | \mathbf{y}, \mathbf{X}) &= \left(\frac{2}{\pi\sigma^2}\right)^{n/2} \exp\left(-\frac{1}{2} \sum_{i=1}^n \left(\frac{y_i - \mathbf{x}_i^T \boldsymbol{\beta}}{\sigma}\right)^2\right) \\ &\quad \times \prod_{i=1}^n \Phi\left(\lambda \left(\frac{y_i - \mathbf{x}_i^T \boldsymbol{\beta}}{\sigma}\right)\right) \\ \ell(\boldsymbol{\beta}, \sigma, \lambda | \mathbf{y}, \mathbf{X}) &= \frac{n}{2} \ln\left(\frac{2}{\pi}\right) - \frac{n}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mathbf{x}_i^T \boldsymbol{\beta})^2 \\ &\quad + \sum_{i=1}^n \ln \Phi\left(\lambda \left(\frac{y_i - \mathbf{x}_i^T \boldsymbol{\beta}}{\sigma}\right)\right) \end{aligned} \quad (2.13)$$

bulunur.  $\arg \max_{\boldsymbol{\beta}, \sigma, \lambda} \ell(\boldsymbol{\beta}, \sigma, \lambda | \mathbf{y}, \mathbf{X})$  maksimizasyon probleminin çözümüne ulaşmak için, log-olabilirlik fonksiyonunun parametrelere göre türevi alınıp sıfıra eşitlenir:

$$\left\{ \begin{array}{l} \frac{\partial \ell}{\partial \boldsymbol{\beta}} = \frac{1}{\sigma^2} \sum_{i=1}^n (y_i - \mathbf{x}_i^T \boldsymbol{\beta}) \mathbf{x}_i - \frac{\lambda}{\sigma} \sum_{i=1}^n \frac{\phi(\lambda z_i)}{\Phi(\lambda z_i)} \mathbf{x}_i = \mathbf{0} \\ \frac{\partial \ell}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^3} \sum_{i=1}^n (y_i - \mathbf{x}_i^T \boldsymbol{\beta})^2 - \frac{\lambda}{\sigma} \sum_{i=1}^n \frac{\phi(\lambda z_i)}{\Phi(\lambda z_i)} z_i = 0 \\ \frac{\partial \ell}{\partial \lambda} = \sum_{i=1}^n \frac{\phi(\lambda z_i)}{\Phi(\lambda z_i)} z_i = 0 \end{array} \right. \quad (2.14)$$

Burada  $z_i = \frac{y_i - \mathbf{x}_i^T \boldsymbol{\beta}}{\sigma}$  biçiminde tanımlıdır.

(2.14) ile verilen denklem sisteminin çözülmesiyle  $\omega_i = \frac{\phi(\lambda z_i)}{\Phi(\lambda z_i)}$  olmak üzere parametrelerin ML tahminleri

$$\begin{aligned}\hat{\boldsymbol{\beta}} &= \left( \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i^T \right)^{-1} \left( \sum_{i=1}^n y_i \mathbf{x}_i - \hat{\lambda} \hat{\sigma} \sum_{i=1}^n \omega_i \mathbf{x}_i \right) \\ \hat{\sigma}^2 &= \frac{\sum_{i=1}^n \left( y_i - \mathbf{x}_i^T \hat{\boldsymbol{\beta}} \right)^2}{n + \hat{\lambda} \sum_{i=1}^n \omega_i z_i}\end{aligned}$$

eşitlikleri kullanılarak sayısal hesaplamalar ile elde edilir.  $\mathbf{W}^T = [\omega_1 \ \omega_2 \ \dots \ \omega_n]$  ağırlık vektörü ve  $\mathbf{Z}^T = [z_1 \ z_2 \ \dots \ z_n]$  olmak üzere ML tahmin edicilerine ilişkin

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \left( \mathbf{X}^T \mathbf{y} - \hat{\lambda} \hat{\sigma} \mathbf{X}^T \widehat{\mathbf{W}} \right) \quad (2.15)$$

$$\hat{\sigma}^2 = \frac{\left( \mathbf{y} - \mathbf{X} \hat{\boldsymbol{\beta}} \right)^T \left( \mathbf{y} - \mathbf{X} \hat{\boldsymbol{\beta}} \right)}{n + \hat{\lambda} \widehat{\mathbf{W}}^T \widehat{\mathbf{Z}}} \quad (2.16)$$

eşitlikleri elde edilir.

$\lambda$  çarpıklık parametresinin tahmini ise analitik olarak elde edilemeyip

$$\widehat{\mathbf{W}}^T \widehat{\mathbf{Z}} = 0 \quad (2.17)$$

eşitliğinden sayısal yöntemlerle hesaplanır.

## Fisher Bilgi Matrisi

Eşitlik (2.13) ile verilen log-olabilirlik fonksiyonunun ikinci kısmi türevleri

$$\frac{\partial^2 \ell}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} = -\frac{1}{\sigma^2} \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i^T - \frac{\lambda^3}{\sigma^2} \sum_{i=1}^n \omega_i z_i \mathbf{x}_i \mathbf{x}_i^T - \frac{\lambda^2}{\sigma^2} \sum_{i=1}^n \omega_i^2 \mathbf{x}_i \mathbf{x}_i^T \quad (2.18)$$

$$\frac{\partial^2 \ell}{\partial \sigma^2} = \frac{n}{\sigma^2} - \frac{3}{\sigma^2} \sum_{i=1}^n z_i^2 - \frac{2\lambda}{\sigma^2} \sum_{i=1}^n \omega_i z_i - \frac{\lambda^2}{\sigma^2} \sum_{i=1}^n \omega_i z_i^2 (\lambda z_i - \omega_i) \quad (2.19)$$

$$\frac{\partial^2 \ell}{\partial \lambda^2} = -\lambda \sum_{i=1}^n \omega_i z_i^3 - \sum_{i=1}^n \omega_i^2 z_i^2 \quad (2.20)$$

$$\frac{\partial^2 \ell}{\partial \sigma \partial \beta} = -\frac{2}{\sigma^2} \sum_{i=1}^n z_i \mathbf{x}_i + \frac{\lambda}{\sigma^2} \sum_{i=1}^n \omega_i \mathbf{x}_i - \frac{\lambda^3}{\sigma^2} \sum_{i=1}^n \omega_i z_i^2 \mathbf{x}_i - \frac{\lambda^2}{\sigma^2} \sum_{i=1}^n \omega_i z_i \mathbf{x}_i \quad (2.21)$$

$$\frac{\partial^2 \ell}{\partial \beta \partial \lambda} = \frac{1}{\sigma} \sum_{i=1}^n z_i \mathbf{x}_i - \frac{1}{\sigma} \sum_{i=1}^n \omega_i \mathbf{x}_i + \frac{\lambda^2}{\sigma} \sum_{i=1}^n \omega_i z_i^2 + \frac{\lambda}{\sigma} \sum_{i=1}^n \omega_i^2 z_i \mathbf{x}_i \quad (2.22)$$

$$\frac{\partial^2 \ell}{\partial \sigma \partial \lambda} = \frac{\lambda^2}{\sigma} \sum_{i=1}^n \omega_i z_i^3 + \frac{\lambda}{\sigma} \sum_{i=1}^n \omega_i^2 z_i^2 \quad (2.23)$$

olarak elde edilir.

$a_{i,k} = E \left( Z^k \frac{\phi(\lambda Z)}{\Phi(\lambda Z)} \right)$  ve  $b_{i,k} = E \left( Z^k \left( \frac{\phi(\lambda Z)}{\Phi(\lambda Z)} \right)^2 \right)$  olmak üzere Fisher bilgi matrisinin elemanları

$$\mathcal{F}_{\beta\beta} = -E \left( \frac{\partial^2 l}{\partial \beta \partial \beta^T} \right) = \frac{1}{\sigma^2} \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i^T + \frac{\lambda^3}{\sigma^2} \sum_{i=1}^n a_{i,1} \mathbf{x}_i \mathbf{x}_i^T + \frac{\lambda^2}{\sigma^2} \sum_{i=1}^n b_{i,0} \mathbf{x}_i \mathbf{x}_i^T$$

$$\mathcal{F}_{\sigma\sigma} = -\frac{n}{\sigma^2} + \frac{3}{\sigma^2} \sum_{i=1}^n z_i^2 + \frac{2\lambda}{\sigma^2} \sum_{i=1}^n a_{i,1} + \frac{\lambda^2}{\sigma^2} \sum_{i=1}^n \omega_i z_i^2 (\lambda z_i - \omega_i)$$

$$\mathcal{F}_{\lambda\lambda} = \lambda \sum_{i=1}^n a_{i,3} + \sum_{i=1}^n b_{i,2}$$

$$\mathcal{F}_{\sigma\beta} = \frac{2}{\sigma^2} \sum_{i=1}^n z_i \mathbf{x}_i - \frac{\lambda}{\sigma^2} \sum_{i=1}^n \omega_i \mathbf{x}_i + \frac{\lambda^3}{\sigma^2} \sum_{i=1}^n a_{i,2} \mathbf{x}_i + \frac{\lambda^2}{\sigma^2} \sum_{i=1}^n a_{i,1} \mathbf{x}_i$$

$$\mathcal{F}_{\beta\lambda} = -\frac{1}{\sigma} \sum_{i=1}^n z_i \mathbf{x}_i + \frac{1}{\sigma} \sum_{i=1}^n \omega_i \mathbf{x}_i - \frac{\lambda^2}{\sigma} \sum_{i=1}^n a_{i,2} - \frac{\lambda}{\sigma} \sum_{i=1}^n b_{i,1} \mathbf{x}_i$$

$$\mathcal{F}_{\sigma\lambda} = -\frac{\lambda^2}{\sigma} \sum_{i=1}^n a_{i,3} - \frac{\lambda}{\sigma} \sum_{i=1}^n b_{i,2}$$

şeklinde elde edilir.

#### 2.2.4 Hataların çarpık $t$ dağıldığı durum

Daha önce bahsedildiği gibi Student  $t$  dağılımı normal dağılıma kalın kuyruklu bir alternatif olarak düşünülebilir. Bu durumda kalın kuyruklu ve çarpık verileri modellemek için çarpık  $t$  dağılımı da normal, Student  $t$  ve çarpık normal dağılıma alternatif olarak düşünülebilir.

Literatürde, simetrik Student  $t$  dağılımının çarpık bir dağılıma genelleştirilmesi üzerine çok sayıda yaklaşım mevcuttur. Hansen (1994), Theodossiou (1998), Fernandez

ve Steel (1998), Branco ve Dey (2001), Azzalini ve Capitanio (2003), Sahu vd. (2003), Jones ve Faddy (2003), Aas ve Haff (2006), Arellano-Valle ve Genton (2010) çalışmaları ile literatürü zenginleştirmiştir. Önerilen bu çarpık dağılımlar arasındaki benzerlikler ve farklılıklar Genton (2004), Azzalini (2005) ve Aas ve Haff (2006)'ın çalışmalarında yer almaktadır.

Bu tezde, Azzalini ve Capitanio (2003) tarafından önerilen çarpık  $t$  dağılımı ele alınmıştır.

**Tanım 2.7** Çarpık  $t$  dağılım  $X$  rasgele değişkeninin olasılık yoğunluk fonksiyonu

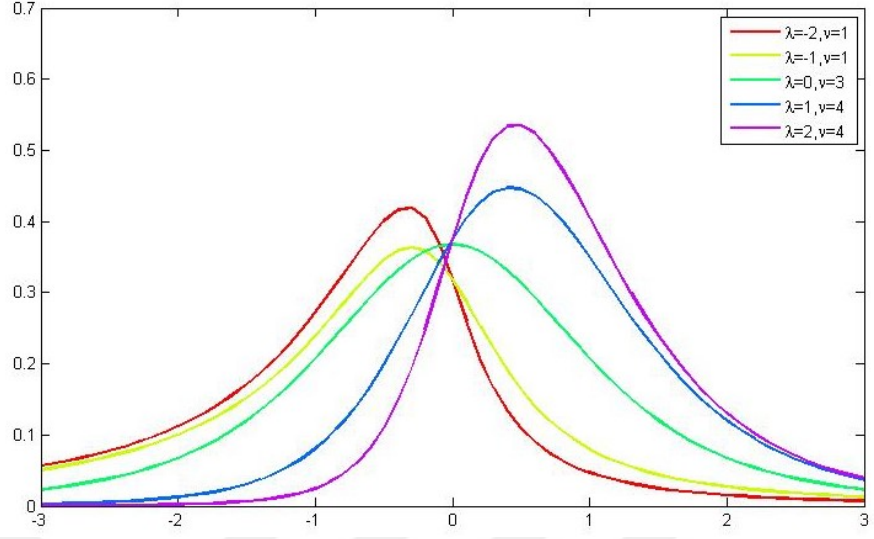
$$f_{St,\nu}(x; \xi, \sigma, \lambda) = \frac{2}{\sigma} t_{\nu}(x_0, \nu) T_{\nu+1} \left( \lambda x_0 \sqrt{\frac{\nu+1}{\nu+x_0^2}} \right), \quad x \in \mathbb{R}$$

biçimindedir ve  $X \sim St(\xi, \sigma^2, \lambda, \nu)$  ile gösterilir. Burada  $x_0 = \sigma^{-1}(x - \xi)$  olmak üzere  $t(\cdot)$  ve  $T(\cdot)$  fonksiyonları sırasıyla Student  $t$  dağılımının olasılık yoğunluk ve dağılım fonksiyonlarını göstermektedir.  $\xi \in \mathbb{R}$  konum parametresi,  $\sigma > 0$  ölçek parametresi,  $\nu$  serbestlik derecesi ve  $\lambda \in \mathbb{R}$  çarpıklık parametresidir.

$X \sim St(\xi, \sigma^2, \lambda, \nu)$  olacak şekildeki  $X$  rasgele değişkeninin beklenen değer, varyans, basıklık ve çarpıklık katsayıları sırasıyla aşağıdaki gibidir :

$$\begin{aligned} E(X) &= \xi + \sigma\kappa, \quad \nu > 1 \\ Var(X) &= \sigma^2 \left( \frac{\nu}{\nu-2} - \kappa^2 \right), \quad \nu > 2 \\ \gamma_1 &= \kappa \left( \frac{\nu(3-\delta^2)}{\nu-3} - \frac{3\nu}{\nu-2} + 2\kappa^2 \right) \left( \frac{\nu}{\nu-2} - \kappa^2 \right)^{-3/2}, \quad \nu > 3 \\ \gamma_2 &= \frac{\frac{3\nu^2}{(\nu-2)(\nu-4)} - \frac{4\kappa^2\nu(3-\delta^2)}{\nu-3} + \frac{6\kappa^2\nu}{\nu-2} - 3\kappa^4}{\left( \frac{\nu}{\nu-2} - \kappa^2 \right)^2} - 3, \quad \nu > 4 \end{aligned}$$

Burada  $\delta = \lambda/\sqrt{1+\lambda^2}$  ve  $\kappa = \delta \left( \frac{\nu}{\pi} \right)^{1/2} \frac{\Gamma(\frac{\nu-1}{2})}{\Gamma(\frac{\nu}{2})}$  şeklinde tanımlıdır. Bazı  $\lambda$  ve  $\nu$  değerleri için  $St(0, 1, \lambda, \nu)$  dağılımına ilişkin olasılık yoğunluk fonksiyonu grafikleri şekil 2.2'de verilmiştir.



Şekil 2.2 Farklı  $\lambda$  ve  $\nu$  değerleri için çarpık  $t$  dağılımın oyf grafikleri

Azzalini tipi çarpık  $t$  dağılımının bazı özellikleri şunlardır:

- (i)  $\lambda > 0$  ise dağılım sağa,  $\lambda < 0$  ise dağılım sola çarpıktır.
- (ii) Özel olarak dağılım,  $\lambda = 0$  ise Student  $t$  dağılımına dönüşmektedir.

Eşitlik (2.2) ile verilen çoklu doğrusal regresyon modelinde  $\epsilon_i$  hata terimlerinin bağımsız, aynı dağılımlı, sıfır ortalamalı,  $\sigma$  bilinmeyen ölçek parametrelili,  $\lambda$  çarpıklık parametrelili ve  $\nu$  serbestlik derecesi ile çarpık  $t$  dağıldığı yani  $\epsilon_i \sim St(0, \sigma^2, \lambda, \nu)$  olduğu ve  $\nu$  serbestlik derecesinin bilindiği varsayalım.

### Regresyon Parametre Tahmini:

Log-olabilirlik fonksiyonu

$$\begin{aligned} \ell(\beta, \sigma, \lambda | \mathbf{y}, \mathbf{X}) = & n \ln(c_\nu) - n \ln \sigma - \frac{\nu + 1}{2} \sum_{i=1}^n \ln(v + z_i^2) \\ & + \sum_{i=1}^n \ln \left( T_{\nu+1} \left( \lambda z_i \sqrt{\frac{\nu + 1}{\nu + z_i^2}} \right) \right) \end{aligned} \quad (2.24)$$

şeklindedir. Burada  $z_i = \frac{y_i - \mathbf{x}_i^T \beta}{\sigma}$  biçiminde tanımlıdır.

$\beta$ ,  $\sigma$  ve  $\lambda$  parametrelerine göre log-olabilirlik fonksiyonunun kısmi türevlerinin sıfıra eşitlenmesiyle parametrelerin ML tahmin edicilerini veren denklem sistemi aşağıdaki gibi elde edilir:

$$\begin{cases} \frac{\partial \ell}{\partial \beta} = \frac{1}{\sigma} \sum_{i=1}^n \omega_i z_i \mathbf{x}_i - \frac{\lambda}{\sigma} \frac{\nu}{\nu+1} \sum_{i=1}^n \frac{t_{\nu+1}(\lambda z_i \sqrt{\omega_i})}{T_{\nu+1}(\lambda z_i \sqrt{\omega_i})} \omega_i^{3/2} \mathbf{x}_i = \mathbf{0} \\ \frac{\partial \ell}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^3} \sum_{i=1}^n \omega_i \left( y_i - \mathbf{x}_i^T \hat{\beta} \right)^2 - \frac{\lambda}{\sigma} \sum_{i=1}^n \gamma_i \omega_i^{3/2} z_i = 0 \\ \frac{\partial \ell}{\partial \lambda} = \sum_{i=1}^n \frac{t_{\nu+1}(\lambda z_i \sqrt{\omega_i})}{T_{\nu+1}(\lambda z_i \sqrt{\omega_i})} \omega_i^{1/2} z_i = 0 \end{cases} \quad (2.25)$$

Burada gösterim kolaylığı olması için  $\omega_i = \frac{\nu+1}{\left[ \nu + \frac{(y_i - \mathbf{x}_i^T \hat{\beta})^2}{\hat{\sigma}^2} \right]}$  ve  $\gamma_i = \frac{\nu}{\nu+1} \frac{t_{\nu+1}(\lambda z_i \sqrt{\omega_i})}{T_{\nu+1}(\lambda z_i \sqrt{\omega_i})}$  biçiminde tanımlanmıştır. Eşitliklerin düzenlenmesiyle  $\beta$  ve  $\sigma$  parametrelerinin ML tahminleri

$$\hat{\beta} = \left( \sum_{i=1}^n \omega_i \mathbf{x}_i \mathbf{x}_i^T \right)^{-1} \left( \sum_{i=1}^n \omega_i y_i \mathbf{x}_i - \hat{\lambda} \hat{\sigma} \sum_{i=1}^n \gamma_i \mathbf{x}_i \omega_i^{3/2} \right) \quad (2.26)$$

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^n \omega_i \left( y_i - \mathbf{x}_i^T \hat{\beta} \right)^2}{n + \hat{\lambda} \sum_{i=1}^n \gamma_i \omega_i^{3/2} z_i} \quad (2.27)$$

eşitlikleri kullanılarak sayısal hesaplamalar ile elde edilir.

$\lambda$  çarpıklık parametresinin ML tahmininin elde edilebilmesi için (2.25) denklem sistemindeki üçüncü eşitliğin çözülmesi gerekmektedir. Lamdanın tahmini için bir diğer yol ise şöyle açıklanabilir: Olabilirlik fonksiyonu iki parçadan oluşmaktadır, bu parçalardan yalnızca ikincisi  $\lambda$  parametresini içermektedir. Böylelikle olabilirlik fonksiyonunun  $\lambda$  parametresine göre maksimize edilmesi ile sadece ikinci parçasının  $\lambda$  parametresine göre maksimize edilmesi teorik olarak denktir. Bu durumda  $\lambda$  parametresinin ML tahminine

$$\arg \max_{\lambda \in \mathbb{R}} \prod_{i=1}^n T_{\nu+1} \left( \lambda z_i \sqrt{\frac{\nu+1}{\nu + z_i^2}} \right) \quad (2.28)$$

maksimizasyon probleminin çözülmesi ile ulaşılabilir.

## Fisher Bilgi Matrisi

Eşitlik (2.24) ile verilen log-olabilirlik fonksiyonunun ikinci kısmi türevleri

$$\ell_z = -z_i \omega_i + \frac{\lambda \omega_i^{1/2} \nu \gamma_i}{\nu + z_i^2}$$

$$\ell_{zz} = \frac{2\omega_i z_i^2}{\nu + z_i^2} - \omega_i - \frac{3\lambda \omega_i^{1/2} \nu z_i \gamma_i}{(\nu + z_i^2)^2} + \frac{\lambda \omega_i^{1/2} \nu \gamma_{zi}}{\nu + z_i^2}$$

$$\ell_{z\lambda} = \frac{\nu \omega_i^{1/2} (\gamma_i + \lambda \gamma_{\lambda i})}{\nu + z_i^2}$$

$$\gamma_{zi} = \frac{\partial \gamma_i}{\partial z_i} = -\frac{\nu(\nu+2)\lambda^2 z_i \gamma_i}{(\nu + z_i^2 + \lambda^2 z_i^2)(\nu + z_i^2)} - \frac{\nu \lambda \omega_i^{1/2} \gamma_i^2}{\nu + z_i^2}$$

$$\gamma_{\lambda i} = \frac{\partial \gamma_i}{\partial \lambda} = -\frac{(\nu+2)\lambda z_i^2 \gamma_i}{(\nu + z_i^2 + \lambda^2 z_i^2)} - z_i \omega_i^{1/2} \gamma_i^2$$

eşitlikleri yardımıyla

$$\ell_{\beta\beta} = \frac{1}{\sigma^2} \sum_{i=1}^n \ell_{zz} \mathbf{x}_i \mathbf{x}_i^T \quad (2.29)$$

$$\ell_{\sigma\sigma} = \frac{1}{\sigma^2} + \frac{z_i^2}{\sigma^2} \ell_{zz} + \frac{2z_i}{\sigma^2} \ell_z \quad (2.30)$$

$$\ell_{\lambda\lambda} = z_i \omega_i^{1/2} \gamma_{\lambda i} \quad (2.31)$$

$$\ell_{\beta\sigma} = \left( \frac{z_i}{\sigma^2} \ell_{zz} + \frac{1}{\sigma^2} \ell_z \right) \mathbf{x}_i \quad (2.32)$$

$$\ell_{\beta\lambda} = -\frac{1}{\sigma} \ell_{z\lambda} \mathbf{x}_i \quad (2.33)$$

$$\ell_{\sigma\lambda} = -\frac{z_i}{\sigma} \ell_{z\lambda} \quad (2.34)$$

biçiminde elde edilir.

(2.29)-(2.34) eşitliklerinin (2.3) eşitliğinde yerine yazılmasıyla Fisher bilgi matrisi elde edilir.

## 2.3 Çoklu Doğrusal Regresyon Modelinde Robust Parametre Tahmin Yöntemleri

Normallik varsayımı altında regresyon parametrelerinin ML tahmin edicileri, LS tahmin edicileri ile aynıdır ve en iyi yansız tahmin edicilerdir. Ancak LS tahmin edicisi verideki aykırı değerlere karşı oldukça hassastır. Bu da parametre tahminlerinin güvenilir olmamasına sebep olmaktadır. Herhangi bir dağılım varsayımı altında elde edilen ML tahmin edicileri ele alınırsa, bu tahmin ediciler de uygulamada sıklıkla karşılaşılan "dağılım varsayımının sağlanmaması" problemi ile karşı karşıya kalmaktadır. Yani ML tahmin edicileri, dağılımdan sapmalara karşı duyarlıdır. Bu durumda güvenilir bir istatistiksel sonuç çıkarımı için aykırı değerlere ve varsayım-dan sapmalara karşı hassasiyeti düşük tahmin edicilere ihtiyaç duyulmaktadır. Bu nedenle bu kısımda literatürde doğrusal regresyon modeli için önerilen bazı robust parametre tahmin yöntemlerine yer verilmiştir.

### 2.3.1 M tahmin edicileri

Literatürde yaygın olarak kullanılan robust tahmin edicilerinden biri, M tahmin edicileridir. Huber (1964) tarafından tanımlanan M tahmin edicileri adını, "en çok olabilirlik tipi" tahmin edicilerden almaktadır.

Huber (1973), doğrusal regresyon modeli için robust M tahmin edicilerini aşağıdaki gibi tanımlamıştır:

Eşitlik (2.1) ile verilen doğrusal regresyon modeli göz önüne alınsın.  $\beta$  parametre vektörünün LS tahmin edicisi Kısım 2.2.1’de bahsedildiği gibi

$$\hat{\beta}_{LS} = \arg \min_{\beta} \sum_{i=1}^n \epsilon_i^2 = \arg \min_{\beta} \sum_{i=1}^n (y_i - \mathbf{x}_i^T \beta)^2$$

minimizasyon probleminin çözülmesi ile elde edilmektedir. Dayanıklı regresyon tahmin edicileri konusunda en önemli adım olan M tahmin edicileri, LS yönteminin

kullandığı artıkların kareleri yerine, daha az hızla artan ve aykırı değerlere karşı daha az duyarlı olacak şekilde bir fonksiyonun kullanılması esasına dayanmaktadır.

Eşitlik (2.1) ile verilen çoklu doğrusal regresyon modeli için parametrelerin M tahmin edicileri

$$\hat{\beta}_M = \arg \min_{\beta} \sum_{i=1}^n \rho(\epsilon_i) = \arg \min_{\beta} \sum_{i=1}^n \rho(y_i - \mathbf{x}_i^T \beta)$$

minimizasyon probleminin çözülmesiyle elde edilir. Amaç fonksiyonu olarak adlandırılan  $\rho(\cdot)$  fonksiyonunun özellikleri aşağıda verilmiştir:

- (i)  $|x_i| > |x_j|$  için  $\rho(x_i) \geq \rho(x_j)$ .
- (ii)  $\rho(0) = 0$ .
- (iii)  $x > 0$  için  $\rho(\cdot)$  fonksiyonu  $\rho(x) < \rho(\infty)$  olacak şekilde artan bir fonksiyondur.
- (iv)  $\rho(\cdot)$  fonksiyonu sınırlı ise  $\rho(\infty) = 1$  olduğu varsayılır (Maronna vd. 2006).

Literatürde M tahmin edicilerinde kullanılabilecek  $\rho(\cdot)$  fonksiyonları için birçok öneri söz konusudur. Bunlardan bazıları çizelge 2.1’de listelenmiştir.

Çizelge 2.1 M tahmin edicileri için kullanılan bazı fonksiyonlar

|         | $\rho(u)$                                                                                                                                                                                                                      | $\psi(u)$                                                                                                                                                                                            |
|---------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Huber   | $\begin{cases} \frac{u^2}{2} & , \quad  u  \leq k \\ k u  - \frac{u^2}{2} & , \quad  u  > k \end{cases}$                                                                                                                       | $\begin{cases} u & , \quad  u  \leq k \\ k \operatorname{sgn}(u) & , \quad  u  > k \end{cases}$                                                                                                      |
| Hampel  | $\begin{cases} \frac{u^2}{2} & , \quad  u  \leq a \\ a u  - \frac{a^2}{2} & , \quad a <  u  \leq b \\ \frac{a(c u  - \frac{u^2}{2})}{c-b} - \frac{7}{6}a^2 & , \quad b <  u  \leq c \\ a(b+c-a) & , \quad  u  > c \end{cases}$ | $\begin{cases} u & , \quad  u  \leq a \\ a \operatorname{sgn}(u) & , \quad a <  u  \leq b \\ \frac{a \operatorname{sgn}(u)(c- u )}{c-b} & , \quad b <  u  \leq c \\ 0 & , \quad  u  > c \end{cases}$ |
| Tukey   | $\begin{cases} \frac{u^2}{2} - \frac{u^4}{2c^2} + \frac{u^6}{6c^4} & , \quad  u  \leq c \\ \frac{c^2}{6} & , \quad  u  > c \end{cases}$                                                                                        | $\begin{cases} u \left(1 - \left(\frac{u}{c}\right)^2\right)^2 & , \quad  u  \leq c \\ 0 & , \quad  u  > c \end{cases}$                                                                              |
| Andrews | $\begin{cases} c[1 - \cos(z/c)] & , \quad  u  \leq c\pi \\ 2c & , \quad  u  > c\pi \end{cases}$                                                                                                                                | $\begin{cases} \sin(z/c) & , \quad  u  \leq c\pi \\ 0 & , \quad  u  > c\pi \end{cases}$                                                                                                              |

Burada  $a, b, c, k$  sabitleri ayar sabiti (tuning constant) olarak ifade edilir. Bu sabitler gözlemlerden hangilerinin aykırı değer olarak değerlendirileceğini belirler ve etkinlik

ile robustluk arasındaki dengeyi sağlar. Ayar sabiti küçük değer aldığında daha fazla gözlem aykırı değer olarak değerlendirilir.

Literatürde bu ayar sabitleri için bir takım öneriler mevcuttur (Holland ve Welsch 1977). Huber için  $k = 1.5$ , Hampel için genellikle  $a = 1.7, b = 3.4$  ve  $c = 8.5$ , Tukey için  $c = 6$  ve Andrews için  $c = 1.5$  veya  $2.1$  alınır. Huber için  $k = 1.345$  kullanılarak standart normal dağılımda % 95 asimptotik etkinlik sağlanabilir.

M tahmin edicileri ölçeğe göre değişmez (invariant) değildir. M tahmin edicilerinin ölçek değişmez (scale invariant) tanımlamasını elde etmek için,  $\hat{\sigma}$ ,  $\sigma$  parametresinin robust tahmini olmak üzere,

$$\hat{\beta}_M = \arg \min_{\beta} \sum_{i=1}^n \rho \left( \frac{\epsilon_i}{\hat{\sigma}} \right)$$

minimizasyon problemi çözüldür. Burada  $\rho(\cdot)$  fonksiyonu konveks ve türevlenebilir ise çözüm tektir (Maronna vd. 2006).

Huber (1964)  $\sigma$  ölçek parametresinin

$$\hat{\sigma} = CMAD(x_i) = C \text{medyan} |x_i - \text{medyan}(x_i)|$$

eşitliği ile veya  $\theta$  konum parametresi ile eşanlı olarak tahmin edilmesini önermiştir.

Ayrıca, normallik varsayımı altında  $C = 1.4826$  alınması önerilmiştir (Rousseeuw ve Leroy 1987).  $r_i = y_i - \mathbf{x}_i^T \hat{\beta}$  olsun. Bu durumda M tahmin edicisi

$$\hat{\beta}_M = \arg \min_{\beta} \sum_{i=1}^n \rho \left( \frac{r_i}{\hat{\sigma}} \right) \quad (2.35)$$

minimizasyon probleminin çözümü ile elde edilir.  $\rho(\cdot)$  fonksiyonunun türevlenebilir olduğu varsayılın ve  $\psi(x; \theta) = \frac{\partial \rho(x; \theta)}{\partial \theta}$  olsun. (2.35) minimizasyon problemi, ifadenin türevinin alınıp sifıra eşitlenmesiyle  $j = 0, 1, \dots, p$  için

$$\sum_{i=1}^n \psi \left( \frac{r_i}{\hat{\sigma}} \right) x_{ij} = 0 \quad (2.36)$$

denklem sistemine dönüştürülür. Burada  $\psi(\cdot)$  skor fonksiyonu olarak adlandırılır ve aşağıdaki özellikleri sağlar:

- (i)  $\psi(-r) = -\psi(r)$
- (ii)  $r > 0$  için  $\psi(r) > 0$  dır.
- (iii)  $\psi(\cdot)$  sürekli ve parçalı türevlenebilir bir fonksiyondur.
- (iv)  $\psi'(0) = 1$ ,  $\psi''(0) = 0$  ve  $\psi'''(0) < 0$  gerçekleşir.

$\psi(\cdot)$  fonksiyonu genellikle doğrusal değildir. Bu nedenle (2.36) eşitliğinden analitik bir çözüme ulaşılamaz ve sayısal yöntemlere başvurulur. Bu yöntemlerden biri ardışık yeniden ağırlıklandırılmış en küçük kareler (iteratively reweighted least squares-IRLS) yöntemidir (Beaton ve Tukey 1974). Bu hesaplama için  $i = 1, 2, \dots, n$  olmak üzere

$$\omega_i \left( \frac{r_i}{\hat{\sigma}} \right) = \begin{cases} \psi \left( \frac{r_i}{\hat{\sigma}} \right) / \frac{r_i}{\hat{\sigma}} & , y_i \neq \mathbf{x}_i' \hat{\boldsymbol{\beta}} \\ 1 & , y_i = \mathbf{x}_i' \hat{\boldsymbol{\beta}} \end{cases}$$

ağırlıkları tanımlansın. Bu ağırlıkların (2.36) eşitliğinde göz önüne alınmasıyla

$$\sum_{i=1}^n \omega_i \left( \frac{r_i}{\hat{\sigma}} \right) x_{ij} = 0, \quad j = 0, 1, \dots, p$$

elde edilir. Görüldüğü üzere  $\psi(\cdot)$  fonksiyonu ağırlıkları belirlediğinden,  $\psi(\cdot)$  fonksiyonu her verinin parametre tahmini üzerindeki marjinal etkisini ölçmektedir. Buradan

$$\sum_{i=1}^n \omega_i \left( y_i - \mathbf{x}_i^T \hat{\boldsymbol{\beta}} \right) x_{ij} = 0, \quad j = 0, 1, \dots, p \quad (2.37)$$

bulunur. (2.37) eşitliğinin düzenlenmesiyle

$$\sum_{i=1}^n \omega_i x_{ij} \mathbf{x}_i^T \hat{\boldsymbol{\beta}} = \sum_{i=1}^n \omega_i y_i x_{ij}, \quad j = 0, 1, \dots, p \quad (2.38)$$

eşitliği elde edilir.  $\mathbf{W} = \text{diag} \{ \omega_i, i = 1, 2, \dots, n \}$  ağırlık matrisinin tanımlanmasıyla (2.38) eşitliği matris formunda

$$\mathbf{X}^T \mathbf{W} \mathbf{X} \boldsymbol{\beta} = \mathbf{X}^T \mathbf{W} \mathbf{y}$$

şeklinde yazılır ve buradan  $\boldsymbol{\beta}$  parametre vektörünün M tahmin edicisine ilişkin

$$\hat{\boldsymbol{\beta}}_M = (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W} \mathbf{y} \quad (2.39)$$

eşitliği elde edilir.

Bu tezde, bu amaç fonksiyonlarından en yaygın kullanılan Tukey'in iki ağırlıklı (bi-weight) fonksiyonu (Beaton ve Tukey 1974) ele alınmıştır. Tukey'in iki ağırlıklı fonksiyonu  $u_i = \frac{r_i}{\hat{\sigma}}$  olmak üzere

$$\rho(u_i) = \begin{cases} \frac{u_i^2}{2} - \frac{u_i^4}{2c^2} + \frac{u_i^6}{6c^4} & , \quad |u_i| \leq c \\ \frac{c^2}{6} & , \quad |u_i| > c \end{cases}$$

ve bu fonksiyona ilişkin  $\psi(\cdot)$  skor fonksiyonları ve  $\omega(\cdot)$  ağırlık fonksiyonları sırasıyla

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

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

biçimindedir. Pratikte normallik varsayımı altında etkinliğin %95 olabilmesi için  $c = 4.685$  alınır (Holland ve Welsch 1977).

M tahmin edicileri, y yönündeki aykırı değerlere karşı robust olmasına rağmen x yönündeki aykırı değerlerin etkisine karşı kırılma noktası (breakdown point - BDP)  $1/n$ 'dir. ("Bir tahmin edicinin kırılma noktası, tahmin edicinin keyfi olarak büyük sapma değerleri almasına yol açabilen aykırı değerlerin oranı olarak tanımlanır" (Matousek vd. 1998).

M tahmin edicileri LS tahmin edicilerine benzemektedir, ancak ağırlık matrisi ile aykırı değerlerin etkisi azaltılmaktadır. LS tahmin edicilerinin aksine, M tahmin edicileri ağırlık matrisine ve ağırlık matrisleri artıklara, dolayısıyla tahmine bağlı olduğundan veriden doğrudan hesaplanamaz. Bu nedenle (2.39) eşitliğinin analitik çözümü mevcut olmayıp sayısal yöntemlerle hesaplanabilir.

IRLS algoritmasının adımları aşağıda verilmiştir:

1.  $\hat{\beta}^{(0)}$  başlangıç tahmini hesaplanır.
2. Her bir iterasyonda ( $t$ ) sırasıyla  $r_i^{(t)} = y_i - \mathbf{x}_i' \hat{\beta}^{(t-1)}$  artıklar hesaplanır.

3.  $r_i^{(t)}$  artıkları yardımıyla  $\hat{\sigma}^{(t)} = 1.4826\text{MAD}\left(r_i^{(t)}\right)$  ölçek tahmini hesaplanır.
4.  $\omega_i^{(t)}$  ağırlıkları aşağıdaki eşitlik kullanılarak hesaplanır.

$$\omega_i^{(t)} = \begin{cases} \psi\left(\frac{r_i^{(t)}}{\hat{\sigma}^{(t)}}\right) / \frac{r_i^{(t)}}{\hat{\sigma}^{(t)}} & , \quad y_i \neq \mathbf{x}_i' \hat{\boldsymbol{\beta}}^{(t-1)} \\ 1 & , \quad y_i = \mathbf{x}_i' \hat{\boldsymbol{\beta}}^{(t-1)} \end{cases}$$

5.  $\mathbf{W}^{(t)} = \text{diag}\left(\left\{\omega_i^{(t)} : i = 1, 2, \dots, n\right\}\right)$  ağırlık matrisi bulunur.
6.  $\hat{\boldsymbol{\beta}}^{(t)} = \left(\mathbf{X}^T \mathbf{W}^{(t)} \mathbf{X}\right)^{-1} \mathbf{X}^T \mathbf{W}^{(t)} \mathbf{y}$  eşitliğinden  $\hat{\boldsymbol{\beta}}^{(t)}$  hesaplanır.
7. Yakınsama sağlanana kadar 2.-6. adımlar tekrar edilir.

Çoklu doğrusal regresyon modelinde değişken seçimi için Bölüm 4'te önerilen M tahmin edicilerine dayalı robust *ICOMP* (*RICOMP<sub>M</sub>*) bilgi kriterinin içerdiği model karmaşıklığının hesaplanması için parametrelerin M tahmin edicilerinin asimptotik kovaryans matrislerine ihtiyaç duyulmaktadır.

M tahmin edicilerinin asimptotik dağılımı Huber (1967) tarafından elde edilmiştir.

**Teorem 2.6**  $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \sigma)^T$  parametre vektörünün M tahmin edicisi  $\hat{\boldsymbol{\theta}}_M$ , asimptotik olarak normal dağılır:

$$\sqrt{n} \left( \hat{\boldsymbol{\theta}}_M - \boldsymbol{\theta} \right) \xrightarrow{D} N_{p+1} \left( \mathbf{0}, \mathbf{B}^{-1} \mathbf{A} \mathbf{B}^{-1} \right) \quad (2.40)$$

(Huber 1981). Burada  $\boldsymbol{\lambda}(\boldsymbol{\theta}) = \{\lambda_j\}_{j=1}^{p+1} = E_F[\psi(x, \boldsymbol{\theta})]$  olmak üzere  $\mathbf{A}$  ve  $\mathbf{B}$  matrisleri aşağıdaki gibi tanımlıdır:

$$\mathbf{A} = E \left[ \psi(x, \boldsymbol{\theta}) \psi(x, \boldsymbol{\theta})^T \right],$$

$$B_{j,k} = \left. \frac{\partial \lambda_j}{\partial \theta_k} \right|_{\boldsymbol{\theta} = \boldsymbol{\theta}_F}$$

### 2.3.2 S tahmin edicileri

M tahmin edicilerinin düşük kırılma noktasına sahip olması nedeniyle bu kısımda daha yüksek kırılma noktasına sahip robust S tahmin edicisine yer verilmiştir.

S tahmin edicileri Rousseeuw ve Yohai (1984) tarafından önerilen, artıkların yayılımını en aza indirgeyen yüksek kırılma noktasına sahip bir yöntemidir. Ölçek (scale) tahminine dayandığı için bu tahmin ediciler S tahmin edicileri adını almıştır.

S tahmin edicilerinin elde edilebilmesi için aşağıdaki özellikleri sağlayan  $\rho(\cdot)$  fonksiyonu göz önüne alınsın:

- (i)  $\rho(\cdot)$  fonksiyonu simetrik, sürekli türevlenebilir ve  $\rho(0) = 0$ .
- (ii)  $\rho(\cdot)$  fonksiyonu  $[0, c]$  aralığında kesin artan ve  $[c, \infty)$  aralığında sabit olacak şekilde bir  $c > 0$  vardır.
- (iii)  $F$  model dağılımı ve  $K = E_F[\rho]$  olmak üzere,  $\frac{K}{\rho(c)} = \frac{1}{2}$  dir.

Amaç fonksiyonu için verilen (ii) özelliği ilgili skor fonksiyonunun re-descending olmasıdır.

(iii) özelliği sağlanmak zorunda olmamakla beraber %50 kırılma noktasını elde etmek için verilmiştir. Rousseeuw ve Yohai (1984) tarafından Tukey'in iki ağırlıklı amaç fonksiyonunu kullanıldığında  $c = 1.547$  seçilmesiyle üçüncü şartın sağlandığı dolayısıyla S tahmin edicisinin % 50 kırılma noktasına sahip olduğu gösterilmiştir.

$r_1, r_2, \dots, r_n$  bir rasgele örneklem olsun.  $s(r_1, r_2, \dots, r_n)$  ölçek tahmini

$$\frac{1}{n} \sum_{i=1}^n \rho\left(\frac{r_i}{s}\right) = K$$

eşitliğinden elde edilir.  $s(r_1, r_2, \dots, r_n)$  ölçek tahmininin aynı zamanda ölçeğin M tahmini olduğu tanımdan açıktır. Buradan  $\theta$  konum parametresinin S tahmin edicisi

$$\hat{\theta}_S = \arg \min_{\theta} s(x_1 - \theta, \dots, x_n - \theta)$$

olarak elde edilir ki  $\hat{\theta}_S$  aynı zamanda

$$\hat{\sigma}_S = s \left( x_1 - \hat{\theta}_S, \dots, x_n - \hat{\theta}_S \right)$$

eşitliğini sağlar.

(2.1) ile verilen çoklu doğrusal regresyon modeli göz önüne alınsın.  $i = 1, 2, \dots, n$  için  $r_i = y_i - \mathbf{x}_i' \boldsymbol{\beta}$  olmak üzere  $s(r_1, r_2, \dots, r_n)$  ölçek tahmini

$$\frac{1}{n} \sum_{i=1}^n \rho \left( \frac{r_i}{s} \right) = K$$

eşitliğinden elde edilir. Burada  $F$  hata terimlerinin dağılımını göstermek üzere  $K = E_F[\rho]$  ile hesaplanır.

$\boldsymbol{\beta}$  parametresinin S tahmin edicisi

$$\hat{\boldsymbol{\beta}}_S = \arg \min_{\boldsymbol{\beta}} \sum_{i=1}^n \rho \left( \frac{y_i - \mathbf{x}_i^T \hat{\boldsymbol{\beta}}}{s} \right)$$

olarak elde edilir ki  $\hat{\boldsymbol{\beta}}_S$  aynı zamanda

$$\hat{\sigma}_S = s \left( y_1 - \mathbf{x}_1^T \hat{\boldsymbol{\beta}}_S, \dots, y_n - \mathbf{x}_n^T \hat{\boldsymbol{\beta}}_S \right)$$

eşitliğini sağlar.

Tukey'in iki ağırlıklı amaç fonksiyonu S tahmin edicilerinin elde edilmesinde kullanılabilir. Çizelge 2.2'de Tukey'in iki ağırlıklı amaç fonksiyonu kullanılarak elde edilen S tahmin edicilerinin BDP'nin farklı değerleri için asimptotik göreceli etkinlikleri verilmiştir.

Çizelge 2.2 Tukey'in amaç fonksiyonu ile farklı  $c$  değerleri için BDP ve asimptotik göreli etkinlik değerleri

| BDP | Etkinlik | $c$   | $K$    |
|-----|----------|-------|--------|
| %50 | %28.7    | 1.547 | 0.1995 |
| %45 | %37.0    | 1.756 | 0.2312 |
| %40 | %46.2    | 1.988 | 0.2634 |
| %35 | %56.0    | 2.251 | 0.2957 |
| %30 | %66.1    | 2.560 | 0.3278 |
| %25 | %75.9    | 2.973 | 0.3593 |
| %20 | %84.7    | 3.420 | 0.3899 |
| %15 | %91.7    | 4.096 | 0.4194 |
| %10 | %96.6    | 5.182 | 0.4475 |

Çoklu doğrusal regresyon modelinde değişken seçimi için Bölüm 4'te önerilen  $S$  tahmin edicilerine dayalı robust  $ICOMP$  ( $RICOMP_S$ ) bilgi kriterinin içerdiği model karmaşıklığının hesaplanması için parametrelerin  $S$  tahmin edicilerinin asimptotik kovaryans matrislerine ihtiyaç duyulmaktadır.

$S$  tahmin edicilerinin asimptotik dağılımının elde edilmesi için aşağıdaki varsayımlar sağlansın.

- (i)  $\rho(\cdot)$  fonksiyonu simetrik, sürekli türevlenebilir ve  $\rho(0) = 0$ .
- (ii)  $\rho(\cdot)$  fonksiyonu  $[0, c]$  aralığında kesin artan ve  $[c, \infty)$  aralığında sabit olacak şekilde bir  $c > 0$  var olsun.
- (iii)  $F$  model dağılımı ve  $K = E_F[\rho]$  olmak üzere  $\frac{K}{\rho(c)} = \frac{1}{2}$  dir.
- (iv)  $u > 0$  için  $\psi(u)/u$  artmayan bir fonksiyon olsun.
- (v)  $E\|X\| < \infty$  olsun.
- (vi)  $\psi(\cdot)$  fonksiyonu türevlenebilir,  $|\psi'(u)|$  sonlu sayıda  $u$  noktası hariç sınırlı ve  $E_\phi(\psi'(u)) < \infty$  olsun.
- (vii)  $E_F(XX^T)$  tekil olmayan ve  $E(\|X\|^3) < \infty$  olsun.

**Teorem 2.7**  $\beta$  ve  $\sigma$  parametrelerinin  $\hat{\beta}_S$  ve  $\hat{\sigma}_S$  S tahmin edicileri asimptotik olarak normal dağılır:

$$\sqrt{n} \left( \hat{\beta}_S - \beta \right) \xrightarrow{D} N_{p+1} \left( \mathbf{0}, \frac{E [\psi^2 (x, \beta)]}{\{E [\psi (x, \beta)]\}^2} (E_F (\mathbf{X}^T \mathbf{X}))^{-1} \right) \quad (2.41)$$

$$\sqrt{n} (\hat{\sigma}_S - \sigma) \xrightarrow{D} N \left( 0, \frac{E [(\rho (x) - K)^2]}{\{E [x\psi (x)]\}^2} \right) \quad (2.42)$$

(Rousseeuw ve Yohai 1984).

Yukarıdaki (i)-(iii) şartlarını sağlayan bir  $\rho(\cdot)$  fonksiyonu kullanıldığında S tahmin edicilerinin kırılma noktaları

$$\varepsilon^* = \frac{(n/2) - p + 2}{n}$$

şeklinde tanımlanır.  $n \rightarrow \infty$  iken  $\varepsilon^* \rightarrow (1/2)$  gerçekleşir ki bu mümkün en iyi değerdir.  $0 < \alpha \leq \frac{1}{2}$  olmak üzere (iii) şartı  $\frac{K}{\rho(c)} = \alpha$  olarak değiştirilirse bu durumda  $n \rightarrow \infty$  iken  $\varepsilon^* \rightarrow \alpha$  gerçekleşir.

S tahmin edicilerinin IRLS algoritması ile hesaplama adımları aşağıda verilmiştir.

1.  $\hat{\beta}^{(0)}$  başlangıç tahmini hesaplanır.
2. Her bir  $t$ . iterasyonda  $i = 1, 2, \dots, n$  için  $r_i^{(t)} = y_i - \mathbf{x}_i^T \hat{\beta}^{(t-1)}$  artıklar hesaplanır.
3.  $r_i^{(t)}$  artıkları yardımıyla  $\hat{\sigma}^{(t)}$  ölçek tahmini hesaplanır:

$$\hat{\sigma}^{(t)} = \begin{cases} 1.4826 \text{MAD}(\underline{x}) & , \quad t = 1 \\ \sqrt{\frac{1}{nK} \sum_{i=1}^n \omega_i e_i^2} & , \quad t > 1 \end{cases}$$

4.  $i = 1, 2, \dots, n$  için  $u_i^{(t)} = \frac{r_i^{(t)}}{\hat{\sigma}^{(t)}}$  değerleri hesaplanır.
5.  $i = 1, 2, \dots, n$  için  $\omega_i^{(t)}$  hesaplanır ve  $\mathbf{W}^{(t)}$  ağırlık matrisi oluşturulur.
6.  $\hat{\beta}^{(t)} = \left( \mathbf{X}' \mathbf{W}^{(t)} \mathbf{X} \right)^{-1} \mathbf{X}' \mathbf{W}^{(t)} \mathbf{y}$  eşitliğinden  $\hat{\beta}^{(t)}$  hesaplanır.
7. Yakınsama sağlanıncaya kadar 2.-6. adımlar tekrar edilir.

### 2.3.3 MM tahmin edicileri

M tahmin edicilerinin yüksek asimptotik göreliliklerinin yanında düşük kırılma noktasına sahip olması, S tahmin edicilerinin ise yüksek kırılma noktasına sahip olmalarının yanısıra asimptotik göreliliklerinin düşük olması göz önüne alınarak bu kısımda bu iki tahmin edicinin robustluk özelliklerini bir araya getiren MM tahmin edicisi ele alınmıştır.

MM tahmin edicileri, M tahmin edicilerinin yüksek asimptotik göreliliği ile S tahmin edicilerinin yüksek kırılma noktası özelliklerini bir araya getiren tahmin edicilerdir. Yohai (1987) tarafından tanımlanan MM tahmin edicileri bu iki özelliğe aynı anda sahip olan ilk robust tahmin edicilerdir. "MM", hesaplanmasında birden fazla M-kestirim yönteminin (multiple M-estimation procedures) uygulanmasını ifade eder. MM tahmin edicileri "belki de şu anda en yaygın olarak kullanılan güçlü regresyon tekniği" (Andersen 2008) olarak düşünülebilir. Bu kısımda MM tahmin edicileri ve bu tahmin edicilerin robustluk özelliklerine yer verilmiştir.

MM tahminleri üç aşamalı bir süreçle aşağıdaki gibi tanımlanır.

**1. Aşama :** Yüksek kırılma noktasına sahip (mümkünse 0.5) bir başlangıç tahmini  $\hat{\beta}_0$  hesaplanır.

**2. Aşama :**

$$r_i(\hat{\beta}_0) = y_i - x_i' \hat{\beta}_0$$

artıkları hesaplanır.  $s_n = s(r_i(\hat{\beta}_0))$  M ölçek tahmini, Yohai (1987)'de verilen varsayımları sağlayan  $\rho_0(\cdot)$  fonksiyonu kullanılarak,  $b/a = 0.5$  eşitliğini sağlayan bir  $b$  sabiti için,

$$\frac{1}{n} \sum_{i=1}^n \rho_0 \left( \frac{r_i(\hat{\beta}_0)}{s_n} \right) = b$$

çözümünden hesaplanır. Burada,  $a = \max \rho(u)$  dur. Huber (1981), bu ilk ölçek tahmininin 0.5 bozulma noktasına sahip olması için,  $b/a = 0.5$  olması gerektiğini ispatlamıştır.

**3. Aşama :**  $\rho_1, \rho_0$  için verilen koşulları sağlayan diğer bir fonksiyon olmak üzere

$$\rho_1(u) \leq \rho_0(u)$$

$$\sup \rho_1(u) = \sup \rho_0(u) = a$$

olmak üzere MM tahmini  $\theta_1$

$$\sum \psi_1 \left( \frac{r_i(\beta)}{s_n} \right) x_i = 0$$

eşitliğinin bir çözümü olarak tanımlanır. Bu  $\hat{\beta}_1$  tahmini ile  $S(\hat{\beta}_1) \leq S(\hat{\beta}_0)$  olmaktadır. Burada

$$S(\beta) = \sum_{i=1}^n \rho_1 \left( \frac{r_i(\beta)}{s_n} \right)$$

dir.  $\rho_1 \left( \frac{0}{0} \right) = 0$  olarak tanımlanır (Çetin ve Orsoy 2001).

MM tahmini sürecinin ilk iki aşaması, yüksek kırılma noktası olan tahmin edicinin elde edilmesini, üçüncü aşama yüksek asimptotik göreceli etkinliği sağlamayı amaçlamaktadır. Bu nedenle  $\rho_0(\cdot)$  ve  $\rho_1(\cdot)$  amaç fonksiyonlarının aynı olması gerekli değildir.

Çoklu doğrusal regresyon modelinde değişken seçimi için Bölüm 4'te önerilen MM tahmin edicilerine dayalı robust *ICOMP* (*RICOMP<sub>MM</sub>*) bilgi kriterinin içerdiği model karmaşıklığının hesaplanması için parametrelerin MM tahmin edicilerinin asimptotik kovaryans matrislerine ihtiyaç duyulmaktadır.

**Teorem 2.8**  $\theta = (\beta^T, \sigma)^T$  parametre vektörünün MM tahmin edicisi asimptotik olarak normal dağılır:

$$\sqrt{n} \left( \hat{\theta}_{MM} - \theta \right) \xrightarrow{D} N_{p+1} \left( \mathbf{0}, \mathbf{B}^{-1} \mathbf{A} \mathbf{B}^{-1} \right) \quad (2.43)$$

Burada  $\lambda(\theta) = E_F [\psi(x, \theta)]$  olmak üzere  $\mathbf{A}$  ve  $\mathbf{B}$  matrisleri aşağıdaki gibi tanımlıdır:

$$\mathbf{A} = E \left[ \psi(x, \theta) \psi(x, \theta)^T \right],$$

$$B_{j,k} = \left. \frac{\partial \lambda_j}{\partial \theta_k} \right|_{\theta=\theta_F}.$$

Yohai (1987), MM tahmin edicilerini tanımladığında, % 50 kırılma noktasına sahip bir tahmin edicinin kullanılması durumunda, MM tahmin edicisinin de % 50 kırılma noktasına sahip olacağını göstermiştir. MM tahmin edicisinin kırılma noktası, ilk iki aşamadaki ayar sabitlerinin seçimine bağlı olmakla birlikte, asimptotik görelî etkinliđi üçüncü aşamadaki ayar sabitlerinin seçimi ile belirlenir. Bu nedenle, M tahmin edicilerinin aksine, MM tahmincilerinin kırılma noktası ve görelî etkinliđi birbirinden bağımsızdır. Bu nedenle, kırılma noktası % 50’de kalırken, asimptotik görelî etkinliđi 1’e yakın olarak ayarlanabilir (Yohai 1987).



### 3. MODEL SEÇİMİ VE MODEL SEÇİM KRİTERLERİ

Bu bölümde, doğrusal regresyon modelinde değişken seçimi için literatürde yer alan bazı bilgi tabanlı model seçim kriterleri ve bu model seçim kriterlerinin robust formlarına yer verilmiştir.

#### 3.1 Bilgi Tabanlı Model Seçim Kriterleri

Bu kısımda bilgi tabanlı model seçim kriterleri ele alınmıştır. Bu seçim kriterlerini ele almadan önce "bilgi"nin istatistiksel tanımına yer verilmiştir.

##### 3.1.1 Bilgi kuramı

Bir rasgele değişkenin entropisi, dağılımın düzgünlüğünü ölçer ve dağılıma ilişkin belirsizliğin bir ölçüsü olarak tanımlanır (Thomas 2001). Belirsizlik ise bilgi eksikliğidir ve bilginin negatif miktarıdır (Horton 1989).

Belirsizliğin tanımlanabilmesi için öncelikle aşağıdaki 4 aksiyom göz önüne alınsın.

- (i)  $B(1) = 0$
- (ii)  $B(p)$ ,  $p$ 'nin kesin azalan bir fonksiyonudur.
- (iii)  $B(p)$ ,  $p$ 'nin sürekli bir fonksiyonudur. Bu aksiyom,  $p$ 'deki küçük bir değişimin  $B(p)$ 'de küçük bir değişmeye karşılık olacağını, sezgisel olarak beklenmesine dayanır.
- (iv)  $B(pq) = B(p) + B(q)$  ,  $0 \leq p, q \leq 1$

Bu aksiyomlara dayanarak, belirsizliğin formülleştirildiği aşağıdaki teorem verilebilir.

**Teorem 3.1**  $B(\cdot)$  yukarıdaki aksiyomları sağlayan bir fonksiyon ve  $c$  herhangi bir sabit olmak üzere,  $B(p) = -c \log(p)$  yazılabilir. Burada ikinci aksiyom göz önüne alındığında  $c > 0$  olduğu açıktır.

Bu durumda, belirsizlik rasgele deęişkenler için aşığıdaki gibi tanımlanabilir:

$X$  rasgele deęişkeninin olasılık fonksiyonu  $f_X(\cdot)$  olmak üzere,  $X$  rasgele deęişkeninin entropisi  $H(X) = E[-\log f_X(X)]$  ile verilir. Bu deęerin  $X$  rasgele deęişkeninin aldığı deęerlere deęil de yalnızca olasılıklarına baęlı olması önemli bir özelliktir. Olasılık daęılımları aynı, fakat deęerleri farklı iki rasgele deęişken için belirsizlik birbirine eşıt olacaktır.

$X$  rasgele deęişkeni ile beraber bir  $Y$  rasgele deęişkeni göz önüne alınsın ve genellik amacıyla bu iki deęişkenin baęımlı olduęu varsayılın.  $(X, Y)$  rasgele vektörüne ilişkin entropi, tanım gereęi  $H(X, Y) = E[-\log f_{X,Y}(X, Y)]$  olacaktır. Burada  $Y$  rasgele deęişkeni gözleendięi zaman  $X$  rasgele deęişkeni için kalan belirsizlik  $H_{Y=y}(X) = E_{X|Y}[-\log f_{X|Y}(X | Y)]$  eşıtlięi ile verilecektir. Buradan da,  $Y$  gözleendięi zaman  $X$  için ortalama belirsizlik miktarı veya  $Y$  'nin geręekleşmesine baęlı  $X$  rasgele deęişkeninin entropisi,  $H_Y(X) = E_Y[H_{Y=y}(X)]$  biçiminde elde edilir.

**Önerme 3.1**  $X$  ve  $Y$  için belirsizlik,  $Y$ 'nin belirsizlięi ile  $Y$  gözleendięi zamandaki  $X$ 'deki ortalama belirsizlięin toplamına eşıttir. Yani  $H(X, Y) = H(Y) + H_Y(X)$  eşıtlięi saęlanır.

**Teorem 3.2**  $X$  ve  $Y$  rasgele deęişkenleri için  $H_Y(X) \leq H(X)$  geręeklenir.  $X$  ile  $Y$  rasgele deęişkenleri baęımsız ise eşıtlik saęlanır.

Teorem 3.2 " $Y$ 'nin geręekleşmesinin  $X$ 'in belirsizlięini ancak azaltacaęını veya bu belirsizlięi deęiştirmeyeceęini (baęımsızlık halinde), fakat hiç bir zaman bunu arttıramayacaęını" ifade etmektedir. Bu durumda  $Y$ 'nin geręekleşmesiyle  $X$ 'in belirsizlięinin ne kadar azaldıęı

$$I(X, Y) = H(X) - H_Y(X) \text{ veya } I(X, Y) = H(Y) - H_X(Y)$$

farkı ile hesaplanabilir. Bu farka  $X$  rasgele deęişkeni hakkında  $Y$  rasgele deęişkeninin verdięi bilgi miktarı veya  $X$  hakkında  $Y$ 'nin içerdigi bilgi denir.

Shannon (1948)'un bilgiyi yukarıda özetlenen şekilde formülleştirmesi bilginin nicel olarak ölçülebilmesini de mümkün kılmıştır (Çetinkaya 2011).

### 3.1.2 Kullback-Leibler bilgisi

Bilgi teorisi ve Kullback-Leibler (K-L) uzaklığı model seçiminin teorik temelini oluşturmaktadır. Kullback-Leibler uzaklığı iki model arasındaki uzaklık olarak ifade edilebilir (Kullback 1959).

**Tanım 3.1** Dağılım fonksiyonu  $G(x)$  olan dağılımdan birbirinden bağımsız, rasgele üretilen  $n$  gözlem  $(x_1, x_2, \dots, x_n)$  ve  $F(x)$  ise bir başka dağılımın dağılım fonksiyonu olsun.  $F(x)$  ve  $G(x)$  dağılımlarına ilişkin olasılık (yoğunluk) fonksiyonları sırasıyla  $f(x)$  ve  $g(x)$  ile gösterilsin. Bu durumda  $f(x)$  modelinin uyum iyiliği gerçek olan  $g(x)$  olasılık fonksiyonuna yakınlığı yani K-L uzaklığı

$$I(g, f) = E_G \left[ \log \left( \frac{g(X)}{f(X)} \right) \right] = \begin{cases} \int_{-\infty}^{\infty} \log \left( \frac{g(x)}{f(x)} \right) g(x) dx & , \text{ sürekli durum} \\ \sum_{i=1}^n \log \left( \frac{g(x_i)}{f(x_i)} \right) g(x_i) & , \text{ kesikli durum} \end{cases}$$

ile tespit edilir. Burada  $E_G$ ,  $G(x)$  olasılık dağılımına göre beklenen değer operatörünü göstermektedir.

Boltzman'ın K-L uzaklığı ile,  $g(x)$ 'e göre  $f(x)$ 'in “yakınlığı (closeness)” ya da “uyum iyiliği (goodness of fit)” ölçülür. K-L uzaklığı aşağıdaki özellikleri sağlar (Konishi ve Kitagawa 2008) :

- $I(g, f) \geq 0$
- $I(g, f) = 0 \Leftrightarrow g$  ile  $f$  aynı dağılımlı
- $I(g, f) \neq I(f, g)$  olduğundan K-L uzaklığı metrik değildir.
- $X_1, X_2, \dots, X_n$  birbirinden bağımsız ve aynı dağılımlı rasgele değişkenler olmak üzere örneklemin K-L bilgisi  $I_n(g, f) = nI(g, f)$ 'dir.

Bu özellikler kullanılarak K-L uzaklığı ne kadar küçükse  $f(x)$  modeli  $g(x)$ 'e o kadar yakındır denilebilir.

Tanımdan yola çıkarak Kullback-Leibler uyumsuzluğu (Kullback ve Leibler 1951) aşağıdaki gibi tanımlanır:

$$I(g, f) = E_G \left[ \log \left( \frac{g(X)}{f(X)} \right) \right] = E_G [\log g(X)] - E_G [\log f(X)]$$

Eşitliğin sağ tarafındaki ilk terim sadece  $g(x)$ 'e bağlıdır. Bu durumda iki model arasındaki uzaklık  $\Delta_{K-L}(f, g) = -E_G [\log f(X)]$  değerine bağlıdır. Bu terim ne kadar küçük olursa K-L uzaklığı o kadar küçük olur, yani bu durumu sağlayan model en iyidir.

$$\Delta_{K-L}(f, g) = -E_G [\log f(X)] = \begin{cases} -\int_{-\infty}^{\infty} g(x) \log f(x) dx & , \text{ sürekli durum} \\ -\sum_{i=1}^n g(x_i) \log f(x_i) & , \text{ kesikli durum} \end{cases}$$

ile tanımlanan  $\Delta_{K-L}(f, g)$  ifadesine Kullback-Leibler uyumsuzluğu adı verilir.

Literatürde Pearson ki-kare, Gauss vb. birçok uyumsuzluk tanımlanmıştır. Uyumu değerlendirmek için hangi uyumsuzluk fonksiyonunun kullanılacağına karar verdikten sonra, modelleri sıralamamızı sağlayacak olan uyumsuzluğun beklentisini tahmini ile seçim kriteri elde edilebilir. Akaike bilgi kriteri de bu yöntem ile elde edilebilir.

### 3.1.3 Akaike bilgi kriteri (AIC)

Olabilirlik fonksiyonunun model parametrelerinin gerçek değerlerinden sapmasını gösteren duyarlı bir ölçüt olduğu bilinir. ML yöntemi Akaike tarafından model tanımlaması için pratik bir bakış açısı olarak değerlendirilmiştir (Akaike 1974). Çünkü ML tahmin edicileri, uygun koşullar altında asimptotik olarak etkindir ve

gerçek değer çevresinde parametrelerin küçük değişimlerine karşı duyarlıdır. Ayrıca en çok olabilirlik fonksiyonu nicelik olarak da ifade edilmeye uygundur (Akaike 1974).

Model derecesinin belirlenmesinde ML yönteminin kullanılması her zaman model için en büyük dereceyi seçmek anlamını taşır. Modelin karmaşıklığı arttıkça model veriye daha duyarlı olur. Böylelikle ampirik olabilirliği maksimum yapan modelin seçimi aday sınıfından en karmaşık modelin seçilmesini sağlar. Bu durumda ML yöntemini model derecesini belirlemek için tek başına kullanmak uygun değildir (Schwarz 1978).

Boyut sayısı farklı modeller arasından en iyi model seçimi için ML yönteminin genişletilmesi düşüncesi Akaike (1974) tarafından ortaya atılmıştır. Olabilirlik prensibi ve Kullback-Leibler uzaklığına dayanan AIC olarak bilinen bu kriter ilk model seçim kriteri olarak kabul edilir ve

$$AIC(p) = -2 \log L(\hat{\theta}_k) + 2p \quad (3.1)$$

biçiminde tanımlanmıştır. Burada  $L(\cdot)$  olabilirlik fonksiyonu,  $\hat{\theta}_k$ ,  $M_k$  modeli altında  $\theta_k$  parametre vektörünün ML tahmini ve  $p$ ,  $M_k$  modelinin bağımsız parametre sayısıdır. En küçük  $AIC$  değeri en iyi modeli ifade etmektedir.

Hataların sıfır ortalamalı,  $\sigma^2$  varyanslı normal dağıldığı varsayımı altınca AIC bilgi kriteri

$$AIC(p) = 2n \log \hat{\sigma} + 2p + (n + n \log(2\pi)) \quad (3.2)$$

şeklinde elde edilir. Burada  $(n + n \log(2\pi))$  sabit olduğundan ihmal edilebilir.

Model içinde tahmin edilen serbest parametrelerin sayısı (ceza bileşeni) karmaşıklığın bir ölçümüdür ve aynı zamanda ML tahmin edicisi kullanıldığında uyumun eksikliğindeki yanlılıktır.  $AIC$  asimptotik kuramın bir sonucu olarak, beklenen K-L bilgisinin örneklem kestirimidir. Shibata (1980)  $AIC$ 'nin asimptotik olarak etkili olduğunu göstermiştir.

Akaike bilgi kriteri aşağıda verilen koşullar altında uygulanır.

- (i) ML kestirimi yapılır.
- (ii) Gerçek modeli içeren parametrik dağılımlar ailesinde uygulanır.

Akaike'nin orijinal AIC'sine dayanarak, cezalandırılmış olabilirliğe dayalı birçok model seçim kriteri önerilmiştir.

### 3.1.4 Düzeltilmiş Akaike bilgi kriteri (AICc)

Düzeltilmiş Akaike bilgi kriteri Sugiura (1978) tarafından doğrusal regresyon modelleri için

$$AICc = n \log \hat{\sigma}^2 + n \frac{1 + p/n}{1 - (p + 2)/n} \quad (3.3)$$

biçiminde tanımlanmıştır. Bununla birlikte

$$AICc = \underbrace{n (\log \hat{\sigma}^2 + 1) + 2(p + 1)}_{AIC} + 2 \frac{(p + 1)(p + 2)}{n - p - 2} \quad (3.4)$$

olarak da yazılabilir.

$AICc$ , regresyon ve zaman serilerinde asimptotik olarak yeterlidir.  $AICc$ , doğrusal regresyon modelleri için yansız olup doğrusal olmayan ve zaman serisi modelleri için yanlıdır. Ancak bu yan,  $AIC$ 'nin yanından daha küçüktür.

Bhansali ve Downhan (1977) tarafından  $AIC$  nin bir genelleştirilmesi verilen bir  $\alpha$  değeri için

$$AIC(p; \alpha) = -2 \log L(\hat{\theta}_k) + \alpha p \quad (3.5)$$

biçiminde tanımlanmıştır.

### 3.1.5 Bayes bilgi kriteri (SBC)

$AIC$  doğrudan örneklem hacmini içermez ve tutarlılık özelliklerinin eksik olması ile eleştirilir (Bozdoğan 1987a). Örneklem hacmini içeren  $AIC$ 'ye popüler bir alternatif olarak  $SBC$  önerilmiştir.  $AIC$  ve  $SBC$  farklı bakış açıları ile elde edilir.  $AIC$ 'nin amacı, gerçek model ile K-L uzaklığının minimum olmasını amaçlarken,  $SBC$  sonsal (posterior) olasılığı en yüksek olan modeli seçer. Schwarz bilgi kriteri ( $BIC$  veya  $SBIC$ ) olarak da bilinen bu kriter, Schwarz (1978) tarafından geliştirilmiş olup farklı parametre sayılarına sahip parametrik modellerin karşılaştırılmasında kullanılan bir kriterdir ve

$$SBC = -2 \log L(\hat{\theta}_k) + k \log(n) \quad (3.6)$$

şeklinde tanımlanır.  $SBC$  eşitliğin sağ tarafındaki örnek büyüklüğüne bağlı olan ikinci kısım itibarıyla  $AIC$ 'den farklılık gösterir.  $SBC$  gözlem noktalarının bağımsız ve benzer dağılımlı olduğunu varsayar.  $AIC$ 'de olduğu gibi  $SBC$  kriterinde de en küçük  $SBC$  değeri en iyi modeli ifade etmektedir (Schwarz 1978).

### 3.1.6 Mallows'un $C_p$ kriteri

Sabit terimide içeren  $k$  açıklayıcı değişkenin bulunduğu model varsayımı altında  $p$  ( $p \leq k$ ) kadar açıklayıcı değişkenin seçildiği ve buradan artık kareler toplamı hesaplanırsa Mallows'un model seçim kriteri (Mallows 1973, 1995) aşağıdaki gibi tanımlanmıştır:

$$C_p = \frac{AKT}{\hat{\sigma}^2} - (n - 2p) \quad (3.7)$$

Cimrilik (parsimony) prensibi dahilinde  $p < k$  yaklaşımı altında modeller karşılaştırılırken en düşük  $C_p$  değerini veren model seçilir (Çetin 2000). Doğrusal regresyon modeli için  $AIC(p)$  ile Mallows'un  $C_p$  kriteri arasındaki ilişki (birinci dereceden eşitlik) ilk olarak Bozdoğan ve Ramirez (1987b) tarafından gösterilmiştir.  $AIC$  ve  $C_p$ 'nin asimptotik olarak eşdeğer olduğu ise Nishii (1984) tarafından ispatlanmıştır.

### 3.1.7 Bilgi karmaşıklığı kriteri (*ICOMP*-Information Complexity)

Bozdogan, çok değişkenli doğrusal veya doğrusal olmayan yapısal modellerde model seçimi için bilgi tabanlı seçim kriterlerine alternatif olarak *ICOMP* olarak adlandırılan entropik istatistiksel karmaşıklık kriterini geliştirmiştir.

*ICOMP*'ın oluşturulması ve geliştirilmesi orijinal olarak van Emden (1971) tarafından tanımlanan bilgi-tabanlı kovaryans karmaşıklık indeksinin bir genelleştirilmesi yoluyla bir elemanın veya rasgele vektörlerin yapısal karmaşıklığına dayalıdır.

Bilgi teorik yorumunu kullanarak *ICOMP*, karmaşıklığı parametrelerin tahminlerinin marjinal dağılımlarına karşı ortak dağılımın ürettiğinin ayrışma bilgisi olarak görür. Ayrışım bilgisi dağılımlar aynı ise sıfır, aksi halde pozitifdir. Genel olarak çok değişkenli doğrusal ya da doğrusal olmayan model

$$\text{istatistiksel model} = \text{sinyal} + \text{gürültü}$$

biçiminde tanımlanır (Bozdogan ve Bearse 2003).

*ICOMP*

$$\text{kayıp} = \underbrace{\text{uyum eksikliği} + \text{cimrilik eksikliği}}_{AIC} + \underbrace{\text{karmaşıklık fazlalığı}}_{ICOMP}$$

şeklindeki bir kayıp fonksiyonu tahmin etmek için bilgi teorisinin ekleme özellikleri kullanılarak değişik yollarla tasarlanır (Bozdogan 2010).

*ICOMP*, *AIC*'den farklı olarak kayıp fonksiyonda karmaşıklık fazlalığı (profusion of complexity) denilen üçüncü bir terime sahiptir. Bu terim modelin parametre tahminleri ve rasgele hata terimleri arasındaki bağımlılık ve kovaryans yapısını anlatmaktadır. Yani yalnızca serbest parametrelerin sayısını cezalandırma yerine *ICOMP*, modelin kovaryans karmaşıklığını cezalandırır (Bozdogan 2010).

$ICOMP$ 'ın en genel formu,  $\hat{\theta}$ ,  $\theta$  parametre vektörünün ML tahmini,  $C$  reel değerli karmaşıklık ölçüsü ve  $\hat{\Sigma}_{\text{model}} = \widehat{Cov}(\hat{\theta})$  modelin parametre vektörünün tahmini kovaryans matrisini göstermek üzere

$$ICOMP = -2 \log L(\hat{\theta}) + 2C(\hat{\Sigma}_{\text{model}}) \quad (3.8)$$

biçimindedir.

### Karmaşıklık Ölçüsü

İstatistiksel modelleme ve model değerlendirme problemlerinde model karmaşıklığı kavramı önemli rol oynamaktadır. Felsefi düzeyde karmaşıklık, model bileşenlerinin bağlantı şekilleri ve etkileşimleri gibi kavramları içerir. Model karmaşıklığının genel bir ölçüsü olmadan model davranışının tahmini ve model kalitesinin değerlendirilmesi zordur. Aşağıda istatistiksel olarak tanımlanmış bir ölçümden hareketle karmaşıklığın basit, sistemsel teorik tanımı verilmiştir (Bozdoğan ve Bearse 2003).

**Tanım 3.2** Herhangi bir tip sistemin karmaşıklığı tüm sistem ve bu sistemin bir basit sayılabilir bileşeni veya kısmı arasında karşılıklı bağımsızlığın derecesinin bir ölçümüdür.

$C$ , bir sistemin karmaşıklığının gerçel-değerli ölçümünü gösterebilir. Bu durumda  $C(S)$ , tüm sistem ile ayrıştırılan bileşenlerin arasındaki farkın miktarını ölçer. Bilginin teorik yorum kullanılarak bu miktar eldeki olasılık modelinin ortak dağılımı ile modelin marjinal dağılımlarının çarpımı arasındaki ayrıştırma bilgisi olarak tanımlanır. Ayrıştırma bilgisi, dağılımlar aynı ise sıfır, diğer durumda pozitifdir (van Emden 1971). Karmaşıklığın bu tanımının, bir modeldeki parametrelerin tahmininin sayısı anlamına gelen ve literatürde sıkça kullanılan tanımdan farklı olduğuna dikkat edilmelidir (Bozdoğan 2003).

## Çok Değişkenli Dağılımların Karmaşıklık Ölçüsü Bilgi Teorisi

$p$  değişkenli sürekli bir dağılımın olasılık yoğunluk fonksiyonu  $f(\mathbf{x}) = f(x_1, x_2, \dots, x_p)$

ve marjinal olasılık yoğunluk fonksiyonları  $j = 1, 2, \dots, p$  için  $f_j(x_j)$  olsun.

$X_1, X_2, \dots, X_p$  rasgele değişkenleri arasındaki bağımlılık bilgi ölçüsü (informational measure of dependence)

$$I(\mathbf{x}) = I(x_1, x_2, \dots, x_p) = E_f \left[ \log \frac{f(x_1, x_2, \dots, x_p)}{f_1(x_1) f_2(x_2) \dots f_p(x_p)} \right]$$

biçiminde tanımlanır. Bu ifade "değişkenler arasındaki beklenen bağımlılığın ölçüsü" olarak da adlandırılır ve aşağıdaki özellikleri sağlar:

- (i)  $I(\mathbf{x}) = I(x_1, x_2, \dots, x_p) \geq 0$
- (ii)  $f(x_1, x_2, \dots, x_p) = f_1(x_1) f_2(x_2) \dots f_p(x_p)$  olması için gerek ve yeter koşul  $X_1, X_2, \dots, X_n$  rasgele değişkenlerinin istatistiksel olarak bağımsız olmasıdır. Böyle bir durumda  $I(\mathbf{x}) = 0$  gerçekleşir. Diğer taraftan  $I(\mathbf{x}) \neq 0$  olması bağımlılığı gerektirir (Bozdoğan 2003).

## Çok Değişkenli Normal Dağılım İçin Kovaryans Karmaşıklığı

$\mathbf{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$  çok değişkenli normal dağılımın olasılık yoğunluk fonksiyonu  $j = 1, 2, \dots, p$  için  $\boldsymbol{\mu}^T = (\mu_1, \mu_2, \dots, \mu_p)$ ,  $-\infty < \mu_j < \infty$ , ve  $\boldsymbol{\Sigma}$  simetrik pozitif tanımlı olmak üzere

$$f(\mathbf{x}) = f(x_1, x_2, \dots, x_p) = (2\pi)^{-p/2} |\boldsymbol{\Sigma}|^{-1/2} \exp \left\{ -\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right\}$$

biçimindedir. Entropinin tanımından

$$\begin{aligned} H(\mathbf{x}) &= H(x_1, x_2, \dots, x_p) \\ &= -E[\log f(\mathbf{X})] \\ &= -\int f(\mathbf{x}) \log f(\mathbf{x}) d(\mathbf{x}) \end{aligned}$$

yazılır.

Buradan ortak entropi bazı hesaplamalar ile

$$\begin{aligned}
H(\mathbf{x}) &= \int f(\mathbf{x}) \left( \frac{1}{2} \log((2\pi)^p |\Sigma|) \right) d(\mathbf{x}) \\
&\quad + \int \frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}) f(\mathbf{x}) d(\mathbf{x}) \\
&= \frac{1}{2} \log((2\pi)^p |\Sigma|) + \frac{1}{2} \int \text{tr} \left( (\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right) f(\mathbf{x}) d(\mathbf{x}) \\
&= \frac{1}{2} \log((2\pi)^p |\Sigma|) + \frac{1}{2} \int \text{tr} \left( \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}) (\mathbf{x} - \boldsymbol{\mu})^T \right) f(\mathbf{x}) d(\mathbf{x}) \\
&= \frac{1}{2} \log((2\pi)^p |\Sigma|) + \frac{1}{2} \text{tr} \left[ \int \left( \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}) (\mathbf{x} - \boldsymbol{\mu})^T \right) f(\mathbf{x}) d(\mathbf{x}) \right] \\
&= \frac{1}{2} \log((2\pi)^p |\Sigma|) + \frac{1}{2} \text{tr} \left[ \Sigma^{-1} E \left( (\mathbf{X} - \boldsymbol{\mu}) (\mathbf{X} - \boldsymbol{\mu})^T \right) \right]
\end{aligned}$$

biçiminde elde edilir (Bozdoğan 2004). Burada  $E \left( (\mathbf{X} - \boldsymbol{\mu}) (\mathbf{X} - \boldsymbol{\mu})^T \right) = \Sigma$  olduğu dikkate alınır

$$\begin{aligned}
H(\mathbf{x}) &= \frac{1}{2} \log((2\pi)^p |\Sigma|) + \frac{1}{2} \text{tr} [\Sigma^{-1} \Sigma] \\
&= \frac{1}{2} \log((2\pi)^p |\Sigma|) + \frac{1}{2} \text{tr} [\mathbf{I}_{p \times p}] \\
&= \frac{p}{2} (\log(2\pi) + 1) + \frac{1}{2} \log |\Sigma|
\end{aligned}$$

bulunur (Bozdoğan 2004).

Benzer şekilde  $j = 1, 2, \dots, p$  için  $X_j \sim (\mu_j, \sigma_j^2)$  rasgele değişkenleri için marjinal entropi ise,

$$\begin{aligned}
H(x_j) &= -E[\log f(X_j)] \\
&= - \int_{-\infty}^{\infty} f(x_j) \log f(x_j) dx_j \\
&= \int_{-\infty}^{\infty} \left( \frac{1}{2} \log(2\pi\sigma_j^2) + \frac{1}{2\sigma_j^2} (x_j - \mu_j)^2 \right) f(x_j) dx_j \\
&= \frac{1}{2} \log(2\pi\sigma_j^2) + \frac{1}{2\sigma_j^2} \overbrace{\text{Var}(X_j)}^{=\sigma_j^2} \\
&= \frac{1}{2} (\log(2\pi\sigma_j^2) + 1)
\end{aligned}$$

biçiminde elde edilir.

Ortak ve marjinal entropiler arasındaki farktan çok değişkenli normal dağılımın  $\Sigma$  kovaryans matrisinin karmaşıklık bilgisi hesaplanır. Burada  $\sigma_{jj} = \sigma_j^2$ ,  $\Sigma$  matrisinin  $j$ . köşegen elemanı ve  $p$  ise boyutudur.  $\Sigma$  kovaryans matrisinin karmaşıklık bilgisi  $C_0(\Sigma)$  ile gösterilirse gerekli sadeleştirmelerden sonra

$$C_{0N}(\Sigma) = \frac{1}{2} \sum_{j=1}^p \log(\sigma_{jj}) - \frac{1}{2} \log |\Sigma| \quad (3.9)$$

bulunur (Bozdogan 2004).  $\Sigma$  köşegen matris ise (yani değişkenler doğrusal olarak bağımsız ise)  $C_{0N}(\Sigma) = 0$  bulunur. Değişkenlerden biri diğerlerinin bir doğrusal kombinasyonu ise (yani  $|\Sigma| = 0$  ise)  $C_{0N}(\Sigma)$  sonlu olmaz.

$C_{0N}(\Sigma)$ , marjinal ve ortak dağılımlara bağlı olduğundan ve ilk terimi ortonormal dönüşümler altında değişeceğinden  $\Sigma$  kovaryans matrisinin karmaşıklığının ölçülmesinde etkili değildir (van Emden 1971). Karmaşıklığın istatistiksel bir modelin genel bir özelliği olduğu hatırlanacak olursa,  $C_{0N}(\Sigma)$  ifadesinin  $j = 1, 2, \dots, n$  için  $\sigma_j^2$ 'ler ile ilişkili rasgele değişkenlerin koordinatlarından bağımsız olması beklenir. Bu durumu iyileştirmek için, Bozdogan (1990) tarafından çok değişkenli normal dağılımın kovaryans matrisine ilişkin  $C_{1N}(\Sigma) = \max C_0(\Sigma)$  karmaşıklık ölçütününün maksimal bilgisi tanımlanmıştır.  $C_{1N}(\Sigma)$  ortonormal dönüşümler altında değişmeyen,  $p$ 'nin monoton artan bir fonksiyonudur (Magnus ve Neudecker 1999).

**Önerme 3.2** Çok değişkenli normal dağılımın  $\Sigma$  kovaryans matrisinin karmaşıklık ölçütününün maksimal bilgisi

$$C_{1N}(\Sigma) = \max_T C_{0N}(\Sigma) = \frac{p}{2} \log \left[ \frac{\text{tr}(\Sigma)}{p} \right] - \frac{1}{2} \log |\Sigma| \quad (3.10)$$

biçimindedir. Burada  $T$ , tüm  $x_1, x_2, \dots, x_p$  koordinat sisteminin ortogonal dönüşümlerini temsil etmektedir (Bozdogan 1990).

Genel olarak karmaşıklığın büyük değerler alması değişkenler arasında etkileşimin fazla olduğunu, küçük değerler alması ise değişkenler arasında az etkileşim olduğunu

gösterir.  $C_{1N}(\boldsymbol{\Sigma})$ 'nın minimum değerine karşılık gelen yapı, "en az karmaşık yapı"dır. Başka bir deyişle  $\boldsymbol{\Sigma} \rightarrow \mathbf{I}$  iken  $C_{1N}(\boldsymbol{\Sigma}) \rightarrow 0$  dır. Aksi halde  $C_{1N}(\boldsymbol{\Sigma})$  pozitif değerlidir ve  $\boldsymbol{\Sigma}$  nın boyutu  $p$ 'ye göre artan bir fonksiyondur. Böyle bir ölçü, model seçiminde benzerlik, farklılık ve yüksek dereceden korelasyon gücünü belirlemek için çok önemlidir (Bozdoğan 2010).

*ICOMP*'ın en genel formu *ICOMP(IFIM)* dir. *ICOMP(IFIM)*, ML tahminlerin iyi bilinen asimptotik optimallik özelliğini açıklar ve bir modelin Fisher bilgi matrisinin tersinin (*IFIM*) bilgi tabanlı karmaşıklığını kullanır. Bu, Cramér-Rao alt sınır matrisi olarak bilinir (Cramér 1946, Rao 1945, 1947, 1948). *ICOMP* aynı zamanda iki K-L uzaklığının toplamı olacak şekilde de ifade edilebilir (Bozdoğan 1988).

Çok değişkenli normal linear veya doğrusal olmayan modeller için *ICOMP(IFIM)*

$$ICOMP(IFIM) = -2 \log L(\hat{\boldsymbol{\theta}}) + 2C_{1N}(\hat{\mathcal{F}}^{-1}) \quad (3.11)$$

biçiminde tanımlıdır (Bozdoğan ve Haughton 1998). Burada  $C_{1N}$ ,  $\hat{\mathcal{F}}^{-1}$  için maksimal karmaşıklık bilgisi olup

$$C_{1N}(\hat{\mathcal{F}}^{-1}) = \frac{p}{2} \log \left[ \frac{tr \hat{\mathcal{F}}^{-1}}{p} \right] - \frac{1}{2} \log |\hat{\mathcal{F}}^{-1}| \quad (3.12)$$

şeklinde tanımlıdır. Burada  $p = \dim \hat{\mathcal{F}}^{-1}(\hat{\boldsymbol{\theta}}) = rank \hat{\mathcal{F}}^{-1}(\hat{\boldsymbol{\theta}})$ 'dır.

### 3.2 Robust Model Seçim Kriterleri

Belli varsayımlar altında ve hata dağılımının bilinmesi gibi durumlarda model seçimi iyi sonuçların elde edildiği oldukça zengin bir literatüre sahiptir. Ancak belli varsayımların sağlanmaması, hataların dağılımının bilinmemesi ve olası aykırı değerler göz önüne alındığında bu problem daha zordur. Klasik model seçim yöntemlerinin çoğu LS gibi tahmin edicilere dayanmaktadır. LS gibi birçok tahmin yönteminin tüm gözlemlere eşit derecede önem verdiği bilinmektedir. Bu, yalnızca bir nokta ile bile model seçiminin çok etkilenmesine yol açmaktadır.

Bu kısımda AIC ve Mallows'un  $C_p$  kriterinin literatürdeki bazı robust formları ele alınmıştır.

### 3.2.1 M tahmin edicisine dayalı robust AIC (AICR)

Ronchetti (1982) tarafından ML tahmin edicisinin M tahmin edicileri ile değiştirilmesiyle robust AIC kriteri tanımlanmıştır. Hataların olasılık yoğunluk fonksiyonunun  $g$  olduğu varsayımı altında

$$AIC(p, \alpha) = n(\log \hat{\sigma}^2) + \alpha p \quad (3.13)$$

eşitliği ile verilen genelleştirilmiş AIC (Bhansali ve Downhan 1977) regresyon modeline uygulanırsa eşitliğin sağ kısmı

$$K(n, \hat{\sigma}) = -2 \sum_{i=1}^n \log g(r_{i,p}) + \alpha p \quad (3.14)$$

biçiminde yazılır. Burada  $r_{i,p} = \frac{(y_i - \mathbf{x}_i^T \hat{\beta})}{\hat{\sigma}}$ ,  $\hat{\sigma}$   $\sigma$  parametresinin bir tahmini ve  $\hat{\beta}$  hataların dağılımının  $g$  olduğu varsayımı altında  $\beta$ 'nin ML tahminidir. Bu şekilde  $K(n, \hat{\sigma})$   $x_i$ 'lere dayalı sabit bir değer,  $R_p = \sum_{i=1}^n (y_i - \mathbf{x}_i^T \tilde{\beta})^2$  de LS tahmini olan  $\tilde{\beta}$ 'ye göre artık kareler toplamı olmak üzere AIC'nin bu genellemesi

$$AIC(p, \alpha) = K(n, \hat{\sigma}) + \frac{R_p}{\hat{\sigma}^2} + \alpha p \quad (3.15)$$

biçiminde yazılabilir.  $-\log g(\cdot)$  fonksiyonunun genel bir  $\rho(\cdot)$  fonksiyonu ile değiştirilmesiyle Ronchetti (1982) tarafından AIC' nin bir robust genellemesi verilen bir  $\alpha$  sabiti ile

$$AICR(p, \alpha, \rho) = 2 \sum_{i=1}^n \rho(r_{i,p}) + \alpha p \quad (3.16)$$

biçiminde tanımlanmıştır. Burada  $\hat{\beta}_M$  M tahmini ve  $\hat{\sigma}$   $\sigma$  parametresinin herhangi bir robust tahmin edicisi olmak üzere  $r_{i,p} = \frac{(y_i - \mathbf{x}_i^T \hat{\beta}_M)}{\hat{\sigma}}$  'dir. Ronchetti (1982, 1985) çalışmalarında  $\alpha = 2 \frac{E[\psi^2]}{E[\psi']} < 2$  alınmasını önermiştir.

Hampel (1983) çalışmasında farklı bir ceza terimi önermiştir. Hampel tarafından önerilen robust AIC kriteri

$$HAIC(p, \alpha, \rho) = 2 \sum_{i=1}^n \rho(r_{i,p}) + \alpha_{HP} \quad (3.17)$$

biçiminde olup burada  $\alpha_H = \frac{E[\psi^2]}{E[\psi']^2} + \frac{E[\psi^2]}{(E[\psi'])^2}$  olarak tanımlıdır.

### 3.2.2 S tahmin edicisine dayalı robust AIC (AICS)

Tharmaratnam vd. (2013) tarafından S tahmin edicilerine dayalı robust AIC kriteri

$$AICS(\rho) = 2n \log(\hat{\sigma}_S) + 2tr(J_{s,n}^{-1}K_{s,n}) \quad (3.18)$$

formunda tanımlanmıştır. Burada  $\hat{\beta}_S$  ve  $\hat{\sigma}_S$  parametrelerin robust S tahminlerini göstermektedir ve ceza terimindeki  $J_{s,n}$  ve  $K_{s,n}$

$$J_{s,n} = \frac{1}{n} \sum_{i=1}^n \rho'' \left( \frac{y_i - \mathbf{x}_i^T \hat{\beta}_S}{\hat{\sigma}_S} \right) \frac{\mathbf{x}_i \mathbf{x}_i^T}{\hat{\sigma}_S^2}$$

$$K_{s,n} = \frac{1}{n} \sum_{i=1}^n \left\{ \rho' \left( \frac{y_i - \mathbf{x}_i^T \hat{\beta}_S}{\hat{\sigma}_S} \right) \right\}^2 \frac{\mathbf{x}_i \mathbf{x}_i^T}{\hat{\sigma}_S^2}$$

şeklinde tanımlıdır. Tharmaratnam vd. (2013) çalışmalarında M ve MM tahmin edicilerine dayalı robust AIC kriterleri benzer şekilde önerilmiştir.

### 3.2.3 MM tahmin edicisine dayalı robust AIC (AICMM)

Tharmaratnam vd. (2013) tarafından MM tahmin edicilerine dayalı robust AIC kriteri

$$AICMM(\rho) = 2 \sum_{i=1}^n \rho \left( \frac{y_i - \mathbf{x}_i^T \hat{\beta}_{MM}}{\hat{\sigma}_{MM}} \right) + 2tr(J_{mm,n}^{-1}K_{mm,n}) \quad (3.19)$$

formunda tanımlanmıştır.  $\hat{\theta}_{MM} = (\hat{\beta}_{MM}^T, \hat{\sigma}_{MM})^T$  parametrelerin MM tahmin edicilerini göstermek üzere

$$J_{mm,n} = - \sum_{i=1}^n \frac{\partial \psi(y_i | x_i, \hat{\theta}_{MM})}{\partial \hat{\theta}_{MM}}$$

$$K_{mm,n} = \sum_{i=1}^n \psi(y_i | x_i, \hat{\theta}_{MM}) \psi'(y_i | x_i, \hat{\theta}_{MM})$$

eşitlikleriyle tanımlıdır. Burada  $\psi(\cdot), \rho(\cdot)$  fonksiyonunun  $\hat{\theta}_{MM}$  'e göre türevidir.

### 3.2.4 Robust $C_p$

Ronchetti ve Staudte (1994)  $C_p$ 'nin sapan gözlemlere karşı duyarlılığını göz önüne alarak  $C_p$ 'nin bir robust genellemesi olan  $RC_p$  kriterini önermiştir.

$$W_p = \sum_{i=1}^n \hat{\omega}_i^2 r_i^2$$

ağırlıklandırılmış artık karelerin toplamı,  $\hat{\sigma}$  modelin varyansının robust ve tutarlı tahmin edicisi,  $U_p$  ile  $V_p$  ağırlıklara ve parametre sayısına bağlı sabitler olmak üzere  $C_p$ 'nin Ronchetti ve Staudte (1994) tarafından önerilen robust genellemesi

$$RC_p = \frac{W_p}{\hat{\sigma}^2} - (U_p - V_p) \quad (3.20)$$

şeklindedir. Burada

$$U_p = \sum_{i=1}^n \text{var}(\hat{\omega}_i r_i)$$

$$V_p = \sum_{i=1}^n \text{var}(\hat{\omega}_i \mathbf{x}_i^T (\boldsymbol{\beta}_p - \boldsymbol{\beta}))$$

eşitlikleri ile hesaplanır.  $V_p$  değerine en yakın ya da küçük olan  $RC_p$  değerini veren model seçilir.

## 4. FARKLI DAĞILIM VARSAYIMI ALTINDA DOĞRUSAL REGRESYON MODELİ İÇİN *ICOMP*

Doğrusal regresyon modellerinde genellikle hataların normal dağıldığı varsayılmaktadır. Halbuki veri setlerinin çarpık/kalın kuyruklu olduğu durumlarla da sıklıkla karşılaşilmektedir. Bu durumlarda normal dağılımla her zaman mükemmel şekilde modelleme yapılamaz. Bununla beraber normal dağılıma dayalı istatistiksel çıkarımın, aykırı değerlere karşı savunmasız olduğu bilinmektedir (Lange vd. 1989). Hataların kalın kuyruklu olduğu durumlarda, Student  $t$  dağılım varsayımı normal dağılıma alternatif olarak kullanılmaktadır. Bunun yanı sıra hataların çarpık bir yapıda olduğu durumlarda ise çarpık dağılımlar, dağılım varsayımlarında ele alınabilir. Bu durumda Student  $t$ , çarpık normal ve çarpık  $t$  dağılımlarının hata dağılımı olarak alınabileceğine Bölüm 2’de yer verilmişti.

Bu bölümde hataların Student  $t$ , çarpık normal ve çarpık  $t$  dağıldığı varsayımı altında *ICOMP* bilgi kriterinin formları elde edilmiştir. Bu varsayımlar altında değişken seçimi problemi ele alınmıştır. *ICOMP* bilgi kriterinin önerilen formları ve orijinal *ICOMP* bilgi kriteri simülasyon çalışması ve gerçek veri örnekleri ile karşılaştırılmıştır.

### 4.1 Student $t$ Dağılımı için *ICOMP* (*ICOMP<sub>t</sub>*)

Çok değişkenli Student  $t$  dağılımı Dunnett ve Sobel (1955) tarafından tanımlanmıştır.  $\mathbf{X}$  rasgele vektörü

$$f(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Omega}, \nu) = \frac{\Gamma\left(\frac{\nu+p}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right) |\boldsymbol{\Omega}|^{1/2} (\nu\pi)^{p/2}} \left(1 + \frac{\mathbf{x}_0^T \mathbf{x}_0}{\nu}\right)^{-\frac{\nu+p}{2}}$$

olasılık yoğunluk fonksiyonuna sahip ise  $\boldsymbol{\mu} \in \mathbb{R}^p$  konum vektörü,  $\boldsymbol{\Omega} \in \mathbb{R}^p \times \mathbb{R}^p$  simetrik ve pozitif tanımlı kovaryans matrisi ve şekil parametresi  $\nu$  serbestlik dereceli çok değişkenli Student  $t$  dağılımına sahiptir denir ve  $\mathbf{X} \sim T_p(\boldsymbol{\mu}, \boldsymbol{\Omega}, \nu)$  ile gösterilir.  $\mathbf{x}_0 = |\boldsymbol{\Omega}|^{-1/2}(\mathbf{x} - \boldsymbol{\mu})$  biçiminde tanımlıdır. Burada  $\nu$  serbestlik derecesinin

bilindiği kabul edilmiştir. Aksi halde  $\nu$  serbestlik derecesinin de tahmin edilmesi gerekmektedir.

#### 4.1.1 Çok değişkenli Student $t$ dağılımı için kovaryans karmaşıklığı

**Lemma 4.1**  $\mathbf{X} \sim \mathbf{T}_p(\mathbf{0}, \mathbf{\Omega}, \nu)$  için ortak entropi

$$H(\mathbf{x}) = \frac{1}{2} \ln |\mathbf{\Omega}| + \ln \left( \frac{(\nu\pi)^{p/2}}{\Gamma(\frac{p}{2})} B\left(\frac{p}{2}, \frac{\nu}{2}\right) \right) + \frac{\nu+p}{2} \left[ \Psi\left(\frac{\nu+p}{2}\right) - \Psi\left(\frac{\nu}{2}\right) \right] \quad (4.1)$$

biçiminde elde edilir.  $B(\cdot, \cdot)$  beta fonksiyonunu ve  $\Psi(\cdot)$  digama fonksiyonunu göstermektedir (Contreras-Reyes 2014).

**Lemma 4.2**  $j = 1, 2, \dots, p$  olmak üzere  $X_j$  için marjinal entropi

$$H(x_j) = \frac{1}{2} \ln \sigma_j^2 + \ln \left( \frac{\sqrt{\nu\pi}}{\Gamma(\frac{1}{2})} B\left(\frac{1}{2}, \frac{\nu}{2}\right) \right) + \frac{\nu+1}{2} \left[ \Psi\left(\frac{\nu+1}{2}\right) - \Psi\left(\frac{\nu}{2}\right) \right] \quad (4.2)$$

elde edilir (Contreras-Reyes 2014).

Böylelikle Lemma 4.1 ve Lemma 4.2 kullanılarak  $\sigma_{jj} = \sigma_j^2$ ,  $\mathbf{\Omega}$  kovaryans matrisinin köşegen elemanlarını göstermek üzere bağımlılık ölçüsü

$$\begin{aligned} I(\mathbf{x}) &= \sum_{j=1}^p \left( \frac{1}{2} \ln \sigma_j^2 + \ln \left( \frac{\sqrt{\nu\pi}}{\Gamma(\frac{1}{2})} B\left(\frac{1}{2}, \frac{\nu}{2}\right) \right) + \frac{\nu+1}{2} \left[ \Psi\left(\frac{\nu+1}{2}\right) - \Psi\left(\frac{\nu}{2}\right) \right] \right) \\ &\quad - \frac{1}{2} \log |\mathbf{\Omega}| - \ln \left( \frac{(\nu\pi)^{p/2}}{\Gamma(\frac{p}{2})} B\left(\frac{p}{2}, \frac{\nu}{2}\right) \right) \\ &\quad - \frac{\nu+p}{2} \left[ \Psi\left(\frac{\nu+p}{2}\right) - \Psi\left(\frac{\nu}{2}\right) \right] \end{aligned}$$

formundadır. Eşitlik (3.9) gözönüne alındığında Student  $t$  dağılımı için kovaryans karmaşıklığı aşağıdaki gibi elde edilir:

$$\begin{aligned} C_{0t}(\mathbf{\Omega}) &= C_{0N}(\mathbf{\Omega}) + \ln \left( \frac{\Gamma(\frac{p}{2}) [B(\frac{1}{2}, \frac{\nu}{2})]^p}{\pi^{p/2} B(\frac{p}{2}, \frac{\nu}{2})} \right) \\ &\quad + p \frac{\nu+1}{2} \left[ \Psi\left(\frac{\nu+1}{2}\right) - \Psi\left(\frac{\nu}{2}\right) \right] \\ &\quad - \frac{\nu+p}{2} \left[ \Psi\left(\frac{\nu+p}{2}\right) - \Psi\left(\frac{\nu}{2}\right) \right]. \end{aligned} \quad (4.3)$$

**Lemma 4.3**  $a, b \in \mathbb{R}^+$  olmak üzere  $\Psi(a) + \ln b \leq \Psi(a + b)$  gereklenir (Gao 2005).

**Lemma 4.4** Her  $a \in \mathbb{R}^+$  iin  $\ln a - \frac{1}{a} \leq \Psi(a) \leq \ln(a + 1) - \frac{1}{a}$  gereklenir (Muqattash ve Yahdi 2006).

Lemma (4.3)'ten

$$\Psi\left(\frac{\nu + p}{2}\right) - \Psi\left(\frac{\nu}{2}\right) \geq \ln\left(\frac{p}{2}\right) \quad (4.4)$$

eşitsizlięi yazılır.

Lemma (4.4)'ten

$$\begin{aligned} \Psi\left(\frac{\nu + 1}{2}\right) - \Psi\left(\frac{\nu}{2}\right) &\leq \ln\left(\frac{\nu + 1}{2} + 1\right) - \frac{2}{\nu + 1} - \Psi\left(\frac{\nu}{2}\right) \\ &\leq \ln\left(\frac{\nu + 3}{2}\right) - \ln\left(\frac{\nu}{2}\right) + \frac{2}{\nu} - \frac{2}{\nu + 1} \\ &= \ln\left(\frac{\nu + 3}{\nu}\right) + \frac{2}{\nu(\nu + 1)} \end{aligned} \quad (4.5)$$

elde edilir.

Eşitlik (4.4) ve (4.5)'den  $C_{0t}(\mathbf{\Omega})$  iin bir üst sınır

$$C_{0t}(\mathbf{\Omega}) \leq C_{0N}(\mathbf{\Omega}) + p \frac{\nu + 1}{2} \left[ \ln\left(\frac{\nu + 3}{\nu}\right) + \frac{2}{\nu(\nu + 1)} \right] - \frac{\nu + p}{2} \ln\left(\frac{p}{2}\right) + \kappa_\nu \quad (4.6)$$

biiminde elde edilir. Burada  $\kappa_\nu = \ln\left(\frac{\Gamma(\frac{p}{2})[B(\frac{1}{2}, \frac{\nu}{2})]^p}{\pi^{p/2} B(\frac{p}{2}, \frac{\nu}{2})}\right)$  sabittir.

Ayrıca,

$$\max C_{0N}(\mathbf{\Omega}) = \frac{p}{2} \log \left[ \frac{\text{tr} \mathbf{\Omega}}{p} \right] - \frac{1}{2} \log |\mathbf{\Omega}|$$

olduęundan Student  $t$  daęılımının maksimum kovaryans karmaşıklığı

$$\begin{aligned} C_{1t}(\mathbf{\Omega}) &= \max C_{0t}(\mathbf{\Omega}) \\ &= \frac{p}{2} \log \left[ \frac{\text{tr}(\mathbf{\Omega})}{p} \right] - \frac{1}{2} \log |\mathbf{\Omega}| + \frac{p(\nu + 1)}{2} \ln\left(\frac{\nu + 3}{\nu}\right) \\ &\quad - \frac{\nu + p}{2} \ln\left(\frac{p}{2}\right) + \frac{p}{\nu} + \kappa_\nu \end{aligned} \quad (4.7)$$

formunda elde edilir. Burada  $p = \text{rank}(\mathbf{\Omega})$ 'dır ve  $\nu$  serbestlik derecesi biliniyor varsayılmaktadır.

Student  $t$  dağılımı varsayımı altında olabilirlik fonksiyonu  $L_{St}(\cdot)$ ,  $\boldsymbol{\theta}$  parametre vektörünün ML tahmin edicisi  $\hat{\boldsymbol{\theta}}$  ve  $\hat{\boldsymbol{\theta}}$ 'nın tahmini kovaryans matrisi  $\hat{\mathcal{F}}^{-1}$  olmak üzere Student  $t$  dağılımı için *ICOMP* kriteri

$$ICOMP_t(IFIM) = -2 \log L_t(\hat{\boldsymbol{\theta}}) + 2C_{1t}(\hat{\mathcal{F}}^{-1}) \quad (4.8)$$

biçiminde tanımlanır. Burada

$$\begin{aligned} C_{1t}(\hat{\mathcal{F}}^{-1}) = & \frac{p}{2} \log \left[ \frac{\text{tr}(\hat{\mathcal{F}}^{-1})}{p} \right] - \frac{1}{2} \log |\hat{\mathcal{F}}^{-1}| + \frac{p(\nu+1)}{2} \ln \left( \frac{\nu+3}{\nu} \right) \\ & - \frac{\nu+p}{2} \ln \left( \frac{p}{2} \right) + \frac{p}{\nu} + \kappa_\nu \end{aligned}$$

şeklinde tanımlanır. Doğrusal ve doğrusal olmayan regresyon modellerinde hataların Student  $t$  dağıldığı durumlarda  $ICOMP_t$  seçim kriteri orijinal *ICOMP* kriterine bir alternatif olabilir.

## 4.2 Çarpık Normal Dağılım için *ICOMP* ( $ICOMP_{SN}$ )

Çok değişkenli çarpık normal dağılım Azzalini ve Dalla-Valle (1996) tarafından tanımlanmıştır. Bu dağılım ve türevleri ciddi sayıda araştırmanın ilgi odağı olmuştur. Bu çalışmada anlatım kolaylığı açısından dağılımın orijinal tanımının farklı şekli (Arellano-Valle ve Genton 2005) kullanılmıştır.

$\boldsymbol{\xi} \in \mathbb{R}^p$  konum vektörü,  $\boldsymbol{\Omega} \in \mathbb{R}^p \times \mathbb{R}^p$  simetrik, pozitif tanımlı kovaryans matrisi,  $\boldsymbol{\eta} \in \mathbb{R}^p$  şekil/çarpıklık parametre vektörü olmak üzere  $\mathbf{Z} \sim SN_p(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\eta})$  rasgele vektörünün olasılık yoğunluk fonksiyonu

$$f(\mathbf{z}; \boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\eta}) = 2\phi_p(\mathbf{z}; \boldsymbol{\xi}, \boldsymbol{\Omega}) \Phi \left( \boldsymbol{\eta}^T \boldsymbol{\Omega}^{-1/2} (\mathbf{z} - \boldsymbol{\xi}) \right), \quad \mathbf{z} \in \mathbb{R}^p$$

biçiminde tanımlıdır. Burada  $\mathbf{z}_0 = \boldsymbol{\Omega}^{-1/2} (\mathbf{z} - \boldsymbol{\xi})$  olmak üzere  $\phi_p(\mathbf{z}; \boldsymbol{\xi}, \boldsymbol{\Omega})$ ,  $p$  değişkenli  $N_p(\boldsymbol{\xi}, \boldsymbol{\Omega})$  dağılımının olasılık yoğunluk fonksiyonunu,  $\phi_p(\mathbf{z}_0) N_p(\mathbf{0}, \mathbf{I}_p)$  dağılımının olasılık yoğunluk fonksiyonunu ve  $\Phi$  tek değişkenli  $N_1(0, 1)$  dağılımının dağılım fonksiyonunu göstermektedir. Ayrıca,  $\phi_p(\mathbf{z}; \boldsymbol{\xi}, \boldsymbol{\Omega}) = |\boldsymbol{\Omega}|^{-1/2} \phi_p(\mathbf{z}_0)$  eşitliği sağlanmaktadır.  $\boldsymbol{\Omega}^{-1/2}$  ile ise  $\boldsymbol{\Omega}^{1/2}$  matrisinin tersi gösterilmektedir.

#### 4.2.1 Çok deęişkenli arpık normal daęılım için kovaryans karmaşıklığı

Kullback-Leibler bilgisinin kolay hesaplanabilmesi için aşığıdaki lemma verilmiştir.

**Lemma 4.5**  $\mathbf{Z} \sim SN_p(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\eta})$  ve  $\boldsymbol{\delta} = \frac{\boldsymbol{\Omega}\boldsymbol{\eta}}{\sqrt{1+\boldsymbol{\eta}^T\boldsymbol{\Omega}\boldsymbol{\eta}}}$  olmak üzere

(i)  $U_0 \sim N_1(0, 1)$  ve  $\mathbf{U} \sim \mathbf{N}_p(\mathbf{0}, \boldsymbol{\Omega} - \boldsymbol{\delta}\boldsymbol{\delta}^T)$  bağımsız olmak üzere

$$\mathbf{Z} \stackrel{d}{=} \boldsymbol{\xi} + \boldsymbol{\delta}|U_0| + \mathbf{U}$$

gereklenir.

(ii)  $E(\mathbf{Z}) = \boldsymbol{\xi} + \sqrt{\frac{2}{\pi}}\boldsymbol{\delta}$  ve  $Cov(\mathbf{Z}) = \boldsymbol{\Omega} - \frac{2}{\pi}\boldsymbol{\delta}\boldsymbol{\delta}^T$  dir.

(iii) Her  $\mathbf{a} \in \mathbb{R}^p$  vektörü ve  $\mathbf{B} \in \mathbb{R}^p \times \mathbb{R}^p$  simetrik matrisi için

$$E\left[(\mathbf{Z} - \mathbf{a})^T \mathbf{B} (\mathbf{Z} - \mathbf{a})\right] = tr(\mathbf{B}\boldsymbol{\Omega}) + (\boldsymbol{\xi} - \mathbf{a})^T \mathbf{B} (\boldsymbol{\xi} - \mathbf{a}) + 2\sqrt{\frac{2}{\pi}}(\boldsymbol{\xi} - \mathbf{a})^T \boldsymbol{\Omega}^{-1}\boldsymbol{\delta} \quad (4.9)$$

eşitlięi gereklenir.

(iv) Her  $\tilde{\boldsymbol{\eta}}, \tilde{\boldsymbol{\xi}} \in \mathbb{R}^p$  vektörleri için  $\tilde{\boldsymbol{\eta}}^T(\mathbf{Z} - \tilde{\boldsymbol{\xi}})$  rasgele deęişkeni aşığıdaki daęılıma sahiptir:

$$\tilde{\boldsymbol{\eta}}^T(\mathbf{Z} - \tilde{\boldsymbol{\xi}}) \sim SN_1\left(\tilde{\boldsymbol{\eta}}^T(\boldsymbol{\xi} - \tilde{\boldsymbol{\xi}}), \tilde{\boldsymbol{\eta}}^T\boldsymbol{\Omega}\tilde{\boldsymbol{\eta}}, \frac{\tilde{\boldsymbol{\eta}}^T\boldsymbol{\delta}}{\sqrt{\tilde{\boldsymbol{\eta}}^T\boldsymbol{\Omega}\tilde{\boldsymbol{\eta}} - (\tilde{\boldsymbol{\eta}}^T\boldsymbol{\delta})^2}}\right) \quad (4.10)$$

(Contreras-Reyes ve Arellano-Valle 2012).

$\mathbf{Z} \sim SN_p(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\eta})$  çok deęişkenli arpık normal daęılım için kovaryans karmaşıklığı aşığıda elde edilmiştir:

Ortak entropi normal daęılım için hesaplandığı gibi

$$\begin{aligned} H(\mathbf{z}) &= H(z_1, z_2, \dots, z_p) \\ &= -E[\log f(\mathbf{Z})] \\ &= -\int f(\mathbf{z}) \log f(\mathbf{z}) d(\mathbf{z}) \end{aligned}$$

formülü ile hesaplanır.

Buradan

$$\begin{aligned} H(\mathbf{z}) &= - \int \log(2\phi_p(\mathbf{z}; \boldsymbol{\xi}, \boldsymbol{\Omega}) \Phi_p(\mathbf{z}; \boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\eta})) f(\mathbf{z}) d(\mathbf{z}) \\ &= - \int \log(\phi_p(\mathbf{z}; \boldsymbol{\xi}, \boldsymbol{\Omega})) f(\mathbf{z}) d(\mathbf{z}) - \int \log(2\Phi_p(\mathbf{z}; \boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\eta})) f(\mathbf{z}) d(\mathbf{z}) \end{aligned}$$

yazılır ve ortak entropi

$$\begin{aligned} H(\mathbf{z}) &= - \int - \frac{1}{2} \left\{ p \log(2\pi) + \log|\boldsymbol{\Omega}| + (\mathbf{z} - \boldsymbol{\xi})^T \boldsymbol{\Omega}^{-1} (\mathbf{z} - \boldsymbol{\xi}) \right\} f(\mathbf{z}) d(\mathbf{z}) \\ &\quad - E[\log(2\Phi_p(\mathbf{z}; \boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\eta}))] \\ &= \frac{1}{2} \left( p \log(2\pi) + \log|\boldsymbol{\Omega}| + E[(\mathbf{Z} - \boldsymbol{\xi})^T \boldsymbol{\Omega}^{-1} (\mathbf{Z} - \boldsymbol{\xi})] \right) \\ &\quad - E[\log(2\Phi_p(\mathbf{z}; \boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\eta}))] \end{aligned}$$

biçiminde elde edilir.

Lemma 4.5 (iii) şıkında  $\mathbf{a} = \boldsymbol{\xi}$  ve  $\mathbf{B} = \boldsymbol{\Omega}^{-1}$  alınırsa

$$\begin{aligned} E[(\mathbf{Z} - \boldsymbol{\xi})^T \boldsymbol{\Omega}^{-1} (\mathbf{Z} - \boldsymbol{\xi})] &= \text{tr}(\boldsymbol{\Omega}^{-1} \boldsymbol{\Omega}) + \underbrace{(\boldsymbol{\xi} - \boldsymbol{\xi})^T \boldsymbol{\Omega}^{-1} (\boldsymbol{\xi} - \boldsymbol{\xi})}_{=0} \\ &\quad + 2 \underbrace{\sqrt{\frac{2}{\pi}} (\boldsymbol{\xi} - \boldsymbol{\xi})^T \boldsymbol{\Omega}^{-1} \frac{\boldsymbol{\Omega} \boldsymbol{\eta}}{\sqrt{1 + \boldsymbol{\eta}^T \boldsymbol{\Omega} \boldsymbol{\eta}}}}_{=0} \\ &= p \end{aligned}$$

elde edilir. Buradan ortak entropi

$$\begin{aligned} H(\mathbf{z}) &= \frac{1}{2} (p \log(2\pi) + \log|\boldsymbol{\Omega}| + p) - E[\log(2\Phi_p(\mathbf{z}; \boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\eta}))] \\ &= \frac{1}{2} \log((2\pi e)^p |\boldsymbol{\Omega}|) - E[\log(2\Phi_p(\mathbf{z}; \boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\eta}))] \end{aligned} \quad (4.11)$$

bulunur.

$\tau = \|\boldsymbol{\eta}\| = (\boldsymbol{\eta}^T \boldsymbol{\Omega} \boldsymbol{\eta})^{1/2}$  ve  $W \sim SN(0, 1, \tau)$  olmak üzere eşitlik da, (4.11) daha sade haliyle

$$\begin{aligned} H(\mathbf{z}) &= \frac{1}{2} (p \log(2\pi) + \log|\boldsymbol{\Omega}| + p) - E[\log(2\Phi_p(\mathbf{z}; \boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\eta}))] \\ &= \frac{1}{2} \log((2\pi e)^p |\boldsymbol{\Omega}|) - E[\log(2\Phi_p(\tau W))] \end{aligned} \quad (4.12)$$

şeklinde yazılabilir.

$W_{11} = \tau W$  denilirse  $W_{11} \sim SN(0, \tau^2, \tau)$  olmak üzere (4.12)'den

$$H(\mathbf{z}) = \frac{1}{2} \log((2\pi e)^p |\mathbf{\Omega}|) - E[\log(2\Phi(W_{11}))] \quad (4.13)$$

elde edilir.

Benzer şekilde  $j = 1, 2, \dots, p$  için  $Z_j$  rasgele değişkenlerinin marjinal entropisi ise,

$$\begin{aligned} H(z_j) &= -E[\log f(Z_j)] \\ &= -\int_{-\infty}^{\infty} f(z_j) \log f(z_j) dz_j \\ &= -\int_{-\infty}^{\infty} \left( \log \left( \frac{2}{\sigma_j} \phi \left( \frac{z_j}{\sigma_j} \right) \right) \Phi \left( \frac{z_j}{\sigma_j} \right) \right) f(z_j) dz_j \\ &= -\int_{-\infty}^{\infty} \left( \log(\sigma_j) + \frac{1}{2} \log(2\pi) + \frac{1}{2\sigma_j^2} z_j^2 - \log \left( 2\Phi \left( \frac{z_j}{\sigma_j} \right) \right) \right) f(z_j) dz_j \\ &= \frac{1}{2} \log(\sigma_j^2) + \frac{1}{2} \log(2\pi) + \frac{1}{2\sigma_j^2} \underbrace{E(Z_j^2)}_{=\sigma_j^2} - E \left( \log 2\Phi \left( \frac{Z_j}{\sigma_j} \right) \right) \\ &= \frac{1}{2} [\log(\sigma_j^2) + \log(2\pi) + 1] - E \left( \log 2\Phi \left( \frac{Z_j}{\sigma_j} \right) \right) \end{aligned} \quad (4.14)$$

bulunur.

Eşitlik (4.13) ve (4.14) gözönüne alındığında,  $\bar{\eta} = \mathbf{\Omega}^{1/2} \eta$  ve  $W_j \sim SN \left( \left\| \bar{\eta} \right\| \right)$  olmak üzere  $\left\| \bar{\eta} \right\| W_j = \frac{Z_j}{\sigma_j}$  şeklinde ifade edilirse çok değişkenli çarpık normal dağılımın kovaryans karmaşıklık bilgisi

$$\begin{aligned} I(\mathbf{z}) &= \sum_{j=1}^p H(z_j) - H(\mathbf{z}) \\ &= \sum_{j=1}^p \frac{1}{2} [\log(\sigma_j^2) + \log(2\pi) + 1] - E \left( \log 2\Phi \left( \frac{Z_j}{\sigma_j} \right) \right) \\ &\quad - \left( \frac{1}{2} \log((2\pi e)^p |\mathbf{\Omega}|) - E[\log(2\Phi(W_{11}))] \right) \end{aligned}$$

biçiminde elde edilir. Burada  $\sigma_{jj} = \sigma_j^2$   $\mathbf{\Omega}$ 'nın  $j$ . köşegen elemanı ve  $p$  ise  $\mathbf{\Omega}$ 'nın boyutudur.

$\mathbf{\Omega}$  kovaryans matrisinin karmaşıklık bilgisi  $C_0(\mathbf{\Omega})$  ile gösterilirse bazı terimlerin sadeleşmesiyle

$$\begin{aligned} C_{0SN}(\mathbf{\Omega}) &= \frac{1}{2} \sum_{j=1}^p \log(\sigma_{jj}) - \frac{1}{2} \log |\mathbf{\Omega}| + E[\log(2\Phi_p(W_{11}))] \\ &\quad - \sum_{j=1}^p E\left(\log 2\Phi\left(\left\|\bar{\eta}\right\| W_j\right)\right) \end{aligned}$$

elde edilir. Buradan,

$$C_{0SN}(\mathbf{\Omega}) = C_{0N}(\mathbf{\Sigma}) + E[\log(2\Phi_p(W_{11}))] - \sum_{j=1}^p E\left(\log 2\Phi\left(\left\|\bar{\eta}\right\| W_j\right)\right)$$

bulunur.

$E\left(\log 2\Phi\left(\left\|\bar{\eta}\right\| W_j\right)\right)$  beklenen değeri pozitif olduğundan  $C_{0SN}(\mathbf{\Omega})$  için bir üst sınır

$$C_{0SN}(\mathbf{\Omega}) \leq C_{0N}(\mathbf{\Omega}) + E[\log(2\Phi_p(W_{11}))] \quad (4.15)$$

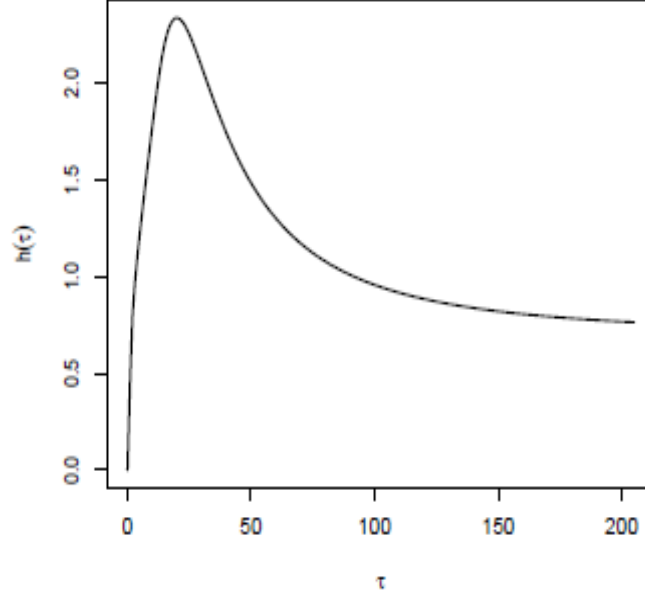
biçiminde elde edilir.

Ayrıca  $\max C_{0N}(\mathbf{\Omega}) = \frac{p}{2} \log \left[ \frac{tr\mathbf{\Omega}}{p} \right] - \frac{1}{2} \log |\mathbf{\Omega}|$  olduğundan

$$\begin{aligned} C_{1SN}(\mathbf{\Omega}) &= \max C_{0SN}(\mathbf{\Omega}) \\ &= \frac{p}{2} \log \left[ \frac{tr\mathbf{\Omega}}{p} \right] - \frac{1}{2} \log |\mathbf{\Omega}| + E[\log(2\Phi_p(W_{11}))] \end{aligned} \quad (4.16)$$

şeklinde elde edilir.

Şekil 4.1'de  $\tau = 1, 2, \dots, 200$  için  $E[\log(2\Phi_p(W_{11}))] = E[\log(2\Phi_p(\tau W))]$  beklenen değerinin aldığı değerler gösterilmektedir. Bu beklenen değer için üst ve alt sınır  $0 \leq E[\log(2\Phi_p(W_{11}))] \leq \log(4e) \approx 2.339$  olarak hesaplanmıştır (Contreras-Reyes 2014). Contreras-Reyes ve Arellano-Valle (2012) bu beklenen değeri "negentropi" adı ile normallikten ayrılma olarak tanımlamışlardır. Çarpık normal dağılım için hesaplanan entropide dolayısıyla karmaşıklık bilgisinde normal dağılımdan farklı olarak çarpıklık parametresine bağlı bir terim gelmiştir.



Şekil 4.1  $\tau = 1, 2, \dots, 200$  için  $h(\tau) = E [\log (2\Phi_p (W_{11}))]$  değerleri  
(Contreras-Reyes 2014)

Çarpık normal dağılım varsayımı altında olabilirlik fonksiyonu  $L_{St}(\cdot)$ ,  $\theta$  parametre vektörünün ML tahmin edicisi  $\hat{\theta}$  ve  $\hat{\theta}$ 'nin tahmini kovaryans matrisi  $\hat{\mathcal{F}}^{-1}$  olmak üzere orijinal *ICOMP* kriterine alternatif olarak önerilen çok değişkenli çarpık normal doğrusal veya doğrusal olmayan modeller için  $ICOMP_{SN}$

$$ICOMP_{SN}(IFIM) = -2 \log L_{SN}(\hat{\theta}) + 2C_{1SN}(\hat{\mathcal{F}}^{-1}) \quad (4.17)$$

biçiminde tanımlıdır.

Burada  $C_{1SN}$ ,  $\hat{\mathcal{F}}^{-1}$  için maksimal karmaşıklık bilgisi olup

$$C_{1SN}(\hat{\mathcal{F}}^{-1}) = \frac{p}{2} \log \left[ \frac{tr \hat{\mathcal{F}}^{-1}}{p} \right] - \frac{1}{2} \log |\hat{\mathcal{F}}^{-1}| + E [\log (2\Phi_p (W_{11}))] \quad (4.18)$$

biçiminde tanımlanır. Burada  $p = \dim \hat{\mathcal{F}}^{-1}(\hat{\theta}) = rank \hat{\mathcal{F}}^{-1}(\hat{\theta})$ 'dır.

### 4.3 Çarpık $t$ Dağılımı için $ICOMP$ ( $ICOMP_{St}$ )

Çok değişkenli çarpık  $t$  dağılımı Branco ve Dey (2001), Azzalini ve Capitanio (2003), Gupta (2003) tarafından tanımlanmıştır.  $\boldsymbol{\mu} \in \mathbb{R}^p$  konum parametresini,  $\boldsymbol{\Omega} \in \mathbb{R}^p \times \mathbb{R}^p$  simetrik ve pozitif tanımlı kovaryans matrisini ve  $\boldsymbol{\lambda} \in \mathbb{R}^p$  çarpıklık parametresini göstermek üzere, çok değişkenli çarpık  $t$  dağılımına sahip bir  $\mathbf{X}$  rasgele vektörünün olasılık yoğunluk fonksiyonu

$$f(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Omega}, \boldsymbol{\lambda}) = 2 |\boldsymbol{\Omega}|^{-1/2} t_{p,v}(\mathbf{x}_0; \nu) T_{v+p} \left( \sqrt{\frac{\nu+p}{\nu + \|\mathbf{x}_0\|^2}} \bar{\boldsymbol{\lambda}}^T \mathbf{x}_0; \nu+p \right), \quad \mathbf{x} \in \mathbb{R}^p$$

biçimindedir ve  $\mathbf{X} \sim \mathbf{ST}_p(\boldsymbol{\mu}, \boldsymbol{\Omega}, \boldsymbol{\lambda})$  ile gösterilir.  $\bar{\boldsymbol{\lambda}} = \boldsymbol{\Omega}^{-1/2} \boldsymbol{\lambda}$ ,  $\mathbf{x}_0 = \boldsymbol{\Omega}^{-1/2}(\mathbf{x} - \boldsymbol{\mu})$  olmak üzere  $t_{p,\nu}(\cdot)$ ,  $\nu$  serbestlik dereceli  $p$ -boyutlu Student  $t$  dağılımının olasılık yoğunluk fonksiyonunu ve  $T_{v+p}(\cdot)$   $\nu + p$  serbestlik dereceli tek değişkenli Student  $t$  dağılımının dağılım fonksiyonunu göstermektedir. Burada  $\nu$  serbestlik derecesi biliniyor kabul edilmiştir. Aksi halde  $\nu$  serbestlik derecesinin de tahmin edilmesi gerekmektedir.

#### 4.3.1 Çok değişkenli çarpık $t$ dağılımı için kovaryans karmaşıklığı

Bu kısımda  $\mathbf{X} \sim \mathbf{ST}_p(\mathbf{0}, \boldsymbol{\Omega}, \boldsymbol{\eta}, \nu)$  çok değişkenli çarpık  $t$  dağılımı için kovaryans karmaşıklığı hesaplanmıştır.

Ortak entropi

$$\begin{aligned} H_{St}(\mathbf{x}) &= H_{St}(x_1, x_2, \dots, x_p) \\ &= -E[\log f(\mathbf{X})] \\ &= -\int \log f(\mathbf{x}) f(\mathbf{x}) d(\mathbf{x}) \\ &= \frac{1}{2} \log |\boldsymbol{\Omega}| - \int \log(t_{p,v}(\mathbf{x}_0; \nu)) f(\mathbf{x}) d(\mathbf{x}) \\ &\quad - \int \log \left( 2T_{v+p} \left( \sqrt{\frac{\nu+p}{\nu + \|\mathbf{x}_0\|^2}} \bar{\boldsymbol{\lambda}}^T \mathbf{x}_0; \nu+p \right) \right) f(\mathbf{x}) d(\mathbf{x}) \end{aligned}$$

biçiminde elde edilir. (4.1) eşitliği ile verilen  $\mathbf{T}_p(\mathbf{0}, \boldsymbol{\Omega}, \nu)$  için elde edilen ortak entropi göz önüne alınırsa

$$H_{St}(\mathbf{x}) = H_t(x) - E \left[ \log \left( 2T_{\nu+p} \left( \sqrt{\frac{\nu+p}{\nu + \|\mathbf{X}_0\|^2}} \bar{\boldsymbol{\lambda}}^T \mathbf{X}_0; \nu+p \right) \right) \right] \quad (4.19)$$

elde edilir.

**Lemma 4.6**  $\mathbf{X}_0 \sim ST_p(\mathbf{0}, \mathbf{I}, \bar{\boldsymbol{\lambda}}, \nu)$  ise  $W_{ST} \sim ST_p(0, 1, \|\bar{\boldsymbol{\lambda}}\|, \nu+p-1)$  ve  $\bar{\boldsymbol{\lambda}} = \boldsymbol{\Omega}^{-1/2} \boldsymbol{\lambda}$  olmak üzere

$$\sqrt{\frac{\nu+p}{\nu + \|\mathbf{X}_0\|^2}} \bar{\boldsymbol{\lambda}}^T \mathbf{X}_0 \stackrel{d}{=} \frac{\sqrt{\nu+p} \|\bar{\boldsymbol{\lambda}}\| W_{ST}}{\sqrt{\nu+p-1 + W_{ST}^2}} \quad (4.20)$$

gerçeklenir (Arellano-Valle vd. 2010).

Eşitlik (4.19) ve (4.20)'dan ortak entropi

$$H_{ST_p}(\mathbf{x}) = H_t(\mathbf{x}) - E \left[ \log \left( 2T_{\nu} \left( \frac{\sqrt{\nu+p} \|\bar{\boldsymbol{\lambda}}\| W_{ST}}{\sqrt{\nu+p-1 + W_{ST}^2}} \right) \right) \right] \quad (4.21)$$

elde edilir. Benzer şekilde  $j = 1, 2, \dots, p$  için  $X_j \sim St(0, \sigma_j^2, \lambda, \nu)$  rasgele değişkenlerinin marjinal entropisi ise  $x_{0j} = \sigma^{-1} x_j$  olmak üzere

$$\begin{aligned} H_{St}(x_j) &= -E[\log f(X_j)] \\ &= -\int_{-\infty}^{\infty} f(x_j) \log f(x_j) dz_j \\ &= -\int_{-\infty}^{\infty} \left( \log \left( 2t_{\nu}(x_0, \nu) T_{\nu+1} \left( \lambda x_{0j} \sqrt{\frac{\nu+1}{\nu + x_{0j}^2}} \right) \right) \right) f(x_j) dx_j \\ &= H_t(x_j) - E \left[ \log \left( 2T_{\nu+1} \left( \lambda x_{0j} \sqrt{\frac{\nu+1}{\nu + x_{0j}^2}} \right) \right) \right] \end{aligned} \quad (4.22)$$

bulunur.

Eşitlik (4.21) ve (4.22)'den çok değişkenli çarpık  $t$  dağılımın kovaryans karmaşıklık bilgisi

$$\begin{aligned}
C_{0St}(\mathbf{x}) &= \sum_{j=1}^p \left( H_t(x_j) - E \left[ \log \left( 2T_{v+1} \left( \lambda x_{0j} \sqrt{\frac{\nu+1}{\nu+x_{0j}^2}} \right) \right) \right] \right) \\
&\quad - \left( H_t(x) - E \left[ \log \left( 2T_v \left( \frac{\sqrt{\nu+p} \|\bar{\mathbf{X}}\| W_{ST}}{\sqrt{\nu+p-1+W_{ST}^2}} \right) \right) \right] \right) \\
&= C_{0t}(\mathbf{x}) + E \left[ \log \left( 2T_v \left( \frac{\sqrt{\nu+p} \|\bar{\mathbf{X}}\| W_{ST}}{\sqrt{\nu+p-1+W_{ST}^2}} \right) \right) \right] \\
&\quad - \sum_{j=1}^p E \left[ \log \left( 2T_{v+1} \left( \lambda x_{0j} \sqrt{\frac{\nu+1}{\nu+x_{0j}^2}} \right) \right) \right] \tag{4.23}
\end{aligned}$$

biçiminde elde edilir.

Buradan

$$C_{1St}(\mathbf{\Omega}) = \max C_{0St}(\mathbf{\Omega}) \tag{4.24}$$

ile elde edilir.

Çarpık  $t$  dağılımı için hesaplanan karmaşıklık bilgisinde de çarpık normal dağılımda olduğu gibi çarpıklık parametresine bağlı fazladan terim gelmiştir. Burada  $\nu$  serbestlik derecesi biliniyor varsayılmaktadır. Çarpık  $t$  dağılımı varsayımı altında olabilirlik fonksiyonu  $L_{St}(\cdot)$ ,  $\boldsymbol{\theta}$  parametre vektörünün ML tahmin edicisi  $\hat{\boldsymbol{\theta}}$  ve  $\hat{\boldsymbol{\theta}}$ 'nın tahmini kovaryans matrisi  $\hat{\mathcal{F}}^{-1}$  olmak üzere Student  $t$  dağılımı için *ICOMP* kriteri, *ICOMP* bilgi kriterine benzer şekilde

$$ICOMP_{St}(IFIM) = -2 \log L_{St}(\hat{\boldsymbol{\theta}}) + 2C_{1St}(\hat{\mathcal{F}}^{-1}) \tag{4.25}$$

eşitliği ile hesaplanır.

#### 4.4 Simülasyon Çalışması

Bu kısımda çoklu doğrusal regresyon modeli için hata terimlerinin Bölüm 4'te ele alınan farklı dağılım varsayımları altında,  $ICOMP$  ve  $ICOMP_t$ ,  $ICOMP_{SN}$  ve  $ICOMP_{St}$ , kriterleri ile değişken seçimine yer verilmiş ve simülasyon çalışması ile kriterlerin performansları karşılaştırılmıştır. Simülasyon çalışması Matlab2013a programlama dili aracılığıyla gerçekleştirilmiştir.

**Simülasyon 1.** Simülasyon çalışması için örneklem hacmi  $n = 40, 100$  ve  $200$  ve tekrar sayısı  $100$  alınmıştır. Hata teriminin dağılımı için aşağıdaki durumlar ele alınmıştır:

- (i)  $\epsilon \sim N(0, 1)$
- (ii)  $\epsilon \sim SN(0, 1, \lambda)$  ,  $\lambda = 0.1, 0.3, 0.5, 0.7, 0.9$ .

Çoklu doğrusal regresyon modeli  $p = 5$  açıklayıcı değişken ile oluşturulmuş ve parametre vektörü  $\beta = (1, 1, 1, 0, 0)^T$  alınmıştır. Burada son iki değişken, değişken seçim kriterlerinin performanslarını takip edebilmek için sıfır seçilmiştir. Her bir açıklayıcı değişkene ilişkin gözlemler birbirinden bağımsız olarak standart normal dağılımdan üretilmiştir.  $\{(x_i, y_i), i = 1, 2, \dots, n\}$  verisi

$$y = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \epsilon$$

çoklu doğrusal regresyon modeli ile elde edilmiştir.

Bölüm 2' de verilen klasik regresyon modeli üzerinde hataların ilişkisizlik varsayımının korunması durumunda çarpık normal dağılıma dayalı verilen parametre tahminleri de sayısal olarak bulunmaktadır. Söz konusu tahminleri bulmak için (2.15)-(2.17) denklemleri kullanılarak iteratif yeniden ağırlıklandırılmış algoritma (IRA) adımları yazılıp yaklaşık çözümler elde edilmiştir.

Değişken sayısı  $p = 5$  olduğundan  $2^5 - 1 = 31$  olası model sözkonusudur. Gerçek model birinci, ikinci ve üçüncü açıklayıcı değişkenleri içermektedir. Bu örnek için olası regresyon modelleri açıklayıcı değişkenlerin çizelge 4.1'de verilen altkümeleri ile elde edilen regresyon modelleridir.

Çizelge 4.1’de verilen olası modellerin her biri için parametre tahminleri ve model seçim kriteri değeri hesaplanır. Yani  $2^p - 1$  modelin herbiri için model seçim kriterinin değeri elde edilir. Kriterin en küçük değeri aldığı model, o seçim kriterinin seçtiği modeldir.

Simülasyon çalışmasında parametre vektörü  $\beta = (1, 1, 1, 0, 0)^T$  alındığından burada doğru model, 16 numaralı modeldir. Çizelge 4.1’de verilen 31 olası model için  $AIC$ ,  $ICOMP$ ,  $ICOMP_{SN}$  kriter değerleri hesaplanmış ve 100 denemede bu modeller arasından doğru modelin, yani 16. modelin, seçim sıklıkları her bir kriter için sayılmış ve frekansları hesaplanmıştır. Elde edilen sonuçlar çizelge 4.2’de verilmiştir.

Çizelge 4.1 p=5 için olası tüm regresyon modelleri

| Değişken Sayısı | Model Numarası | Model Değişkenleri            |
|-----------------|----------------|-------------------------------|
| 1               | 1              | $\{x_1\}$                     |
|                 | 2              | $\{x_2\}$                     |
|                 | $\vdots$       | $\vdots$                      |
|                 | 5              | $\{x_5\}$                     |
| 2               | 6              | $\{x_1, x_2\}$                |
|                 | 7              | $\{x_1, x_3\}$                |
|                 | $\vdots$       | $\vdots$                      |
|                 | 15             | $\{x_4, x_5\}$                |
| 3               | 16             | $\{x_1, x_2, x_3\}$           |
|                 | 17             | $\{x_1, x_2, x_4\}$           |
|                 | $\vdots$       | $\vdots$                      |
|                 | 25             | $\{x_3, x_4, x_5\}$           |
| 4               | 26             | $\{x_1, x_2, x_3, x_4\}$      |
|                 | 27             | $\{x_1, x_2, x_3, x_5\}$      |
|                 | $\vdots$       | $\vdots$                      |
|                 | 30             | $\{x_2, x_3, x_4, x_5\}$      |
| 5               | 31             | $\{x_1, x_2, x_3, x_4, x_5\}$ |

Çizelge 4.2  $AIC$ ,  $ICOMP_N$  ve  $ICOMP_{SN}$  model seçim kriterlerinin doğru modeli seçim sıklığı

| $n = 40$                      | $AIC$ | $ICOMP$ | $ICOMP_{SN}$ |
|-------------------------------|-------|---------|--------------|
| $\epsilon \sim N(0, 1)$       | 0.95  | 0.97    | 0.97         |
| $\epsilon \sim SN(0, 1, 0.1)$ | 0.92  | 0.93    | 0.97         |
| $\epsilon \sim SN(0, 1, 0.3)$ | 0.91  | 0.92    | 0.96         |
| $\epsilon \sim SN(0, 1, 0.5)$ | 0.90  | 0.91    | 0.95         |
| $\epsilon \sim SN(0, 1, 0.7)$ | 0.88  | 0.90    | 0.94         |
| $\epsilon \sim SN(0, 1, 0.9)$ | 0.87  | 0.89    | 0.93         |
| $n = 100$                     |       |         |              |
| $\epsilon \sim N(0, 1)$       | 0.97  | 0.99    | 0.99         |
| $\epsilon \sim SN(0, 1, 0.1)$ | 0.94  | 0.97    | 0.99         |
| $\epsilon \sim SN(0, 1, 0.3)$ | 0.93  | 0.95    | 0.98         |
| $\epsilon \sim SN(0, 1, 0.5)$ | 0.91  | 0.94    | 0.97         |
| $\epsilon \sim SN(0, 1, 0.7)$ | 0.90  | 0.93    | 0.96         |
| $\epsilon \sim SN(0, 1, 0.9)$ | 0.89  | 0.91    | 0.95         |
| $n = 200$                     |       |         |              |
| $\epsilon \sim N(0, 1)$       | 0.98  | 1.00    | 1.00         |
| $\epsilon \sim SN(0, 1, 0.1)$ | 0.97  | 0.98    | 1.00         |
| $\epsilon \sim SN(0, 1, 0.3)$ | 0.95  | 0.97    | 0.99         |
| $\epsilon \sim SN(0, 1, 0.5)$ | 0.94  | 0.96    | 0.99         |
| $\epsilon \sim SN(0, 1, 0.7)$ | 0.93  | 0.95    | 0.98         |
| $\epsilon \sim SN(0, 1, 0.9)$ | 0.90  | 0.94    | 0.97         |

Çizelge 4.2’de verilen sonuçlara göre hataların normal dağıldığı durumda orijinal  $ICOMP$  model seçim kriterinin  $AIC$  ve  $ICOMP_{SN}$  model seçim kriterlerine göre daha çok doğru model seçtiği, görülmektedir. Bununla beraber  $ICOMP_{SN}$  kriteri,  $ICOMP$  kriterine yakın sonuçlar vermiştir.

Hata terimi çarpık normal dağıldığında ise  $ICOMP_{SN}$ ,  $AIC$  ve  $ICOMP$  seçim kriterlerine göre doğru modeli seçmede daha iyi sonuçlara sahiptir.  $\lambda$  çarpıklık parametresinin artması ile  $ICOMP_{SN}$  seçim kriterinin çarpık normal hata terimli regresyon modelinde, diğer model seçim kriterine kıyasla daha iyi sonuçlar elde edilebileceği daha da belirginleşmiştir.

Simülasyon sonuçlarına göre örneklem hacminin artması ile model seçim kriterlerinin performansının artması kriterlerin tutarlılık özelliğini desteklemektedir. Hata terimlerinin üretildiği dağılımlarla model seçim kriterinde kullanılan karmaşıklık ölçüsü aynı olduğunda, beklendiği üzere diğer durumlara göre daha iyi sonuçlar ortaya çıkmaktadır.

**Simülasyon 2.** Simülasyon çalışması için örneklem hacmi  $n = 40, 100$  ve  $200$  ve tekrar sayısı  $100$  alınmıştır. Hata terimlerinin dağılımı için aşağıdaki durumlar ele alınmıştır:

- (i)  $\epsilon \sim N(0, 1)$
- (ii)  $\epsilon \sim t(0, 1, 3)$
- (ii)  $\epsilon \sim St(0, 1, \lambda, \nu)$  ,  $\lambda = 0.1, 0.3, 0.5, 0.7, 0.9$ .

Çoklu doğrusal regresyon modeli  $p = 5$  açıklayıcı değişken ile oluşturulmuş ve parametre vektörü  $\beta = (1, 1, 1, 0, 0)^T$  alınmıştır. Burada son iki değişken değişken seçim kriterlerinin performanslarını takip edebilmek için sıfır seçilmiştir. Her bir açıklayıcı değişkene ilişkin gözlemler birbirinden bağımsız olarak standart normal dağılımdan üretilmiştir.  $\{(x_i, y_i), i = 1, 2, \dots, n\}$  verisi

$$y = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \epsilon$$

çoklu doğrusal regresyon modeli ile elde edilmiştir. Parametre tahminleri Student  $t$  dağılımı için (2.7)-(2.8), çarpık  $t$  dağılımı için (2.26)-(2.27) denklemleri kullanılarak IRA ile elde edilmiştir. Serbestlik derecesi  $\nu = 3$  sabit alınmıştır. Simülasyon sonuçları çizelge 4.3'te verilmiştir.

Çizelge 4.3  $ICOMP$ ,  $ICOMP_t$  ve  $ICOMP_{St}$  model seçim kriterlerinin doğru modeli seçim sıklığı

| $n = 40$  | $\epsilon$         | $ICOMP$ | $ICOMP_t$ | $ICOMP_{St}$ |
|-----------|--------------------|---------|-----------|--------------|
|           | $N(0, 1)$          | 0.97    | 0.93      | 0.92         |
|           | $t(0, 1, 3)$       | 0.91    | 0.95      | 0.94         |
|           | $St(0, 1, 0.1, 3)$ | 0.88    | 0.93      | 0.95         |
|           | $St(0, 1, 0.3, 3)$ | 0.84    | 0.90      | 0.95         |
|           | $St(0, 1, 0.5, 3)$ | 0.83    | 0.88      | 0.94         |
|           | $St(0, 1, 0.7, 3)$ | 0.82    | 0.87      | 0.93         |
|           | $St(0, 1, 0.9, 3)$ | 0.81    | 0.84      | 0.92         |
| <hr/>     |                    |         |           |              |
| $n = 100$ |                    |         |           |              |
|           | $N(0, 1)$          | 0.98    | 0.93      | 0.92         |
|           | $t(0, 1, 3)$       | 0.92    | 0.96      | 0.95         |
|           | $St(0, 1, 0.1, 3)$ | 0.90    | 0.94      | 0.97         |
|           | $St(0, 1, 0.3, 3)$ | 0.89    | 0.92      | 0.96         |
|           | $St(0, 1, 0.5, 3)$ | 0.86    | 0.91      | 0.96         |
|           | $St(0, 1, 0.7, 3)$ | 0.84    | 0.89      | 0.95         |
|           | $St(0, 1, 0.9, 3)$ | 0.82    | 0.86      | 0.94         |
| <hr/>     |                    |         |           |              |
| $n = 200$ |                    |         |           |              |
|           | $N(0, 1)$          | 1.00    | 0.94      | 0.94         |
|           | $t(0, 1, 3)$       | 0.93    | 0.98      | 0.96         |
|           | $St(0, 1, 0.1, 3)$ | 0.91    | 0.96      | 0.98         |
|           | $St(0, 1, 0.3, 3)$ | 0.90    | 0.95      | 0.98         |
|           | $St(0, 1, 0.5, 3)$ | 0.87    | 0.94      | 0.97         |
|           | $St(0, 1, 0.7, 3)$ | 0.85    | 0.91      | 0.96         |
|           | $St(0, 1, 0.9, 3)$ | 0.83    | 0.88      | 0.95         |

Çizelge 4.3'te verilen simülasyon sonuçlarına göre hataların normal dağıldığı varsayımı altında  $ICOMP$  model seçim kriteri yine  $ICOMP_t$  ve  $ICOMP_{St}$  model seçim kriterlerine göre daha fazla doğru modeli seçmiştir. Hata terimi Student  $t$  dağıldığında,

$ICOMP_t$  model seçim kriterinin,  $ICOMP$  ve  $ICOMP_{St}$  model seçim kriterlerine göre daha çok doğru model seçtiği görülmektedir. Hata terimi çarpık  $t$  dağıldığında ise,  $ICOMP_{St}$ ,  $ICOMP$  ve  $ICOMP_t$  seçim kriterlerine göre doğru modeli seçmede daha iyi sonuçlara sahiptir. Çarpıklık arttıkça  $ICOMP_{St}$  seçim kriterinin çarpık  $t$  dağılımlı hata terimleri ile regresyonda, diğer model seçim kriterine kıyasla daha iyi sonuçlar elde edildiği daha da belirgindir. Örneklem hacminin artması ile model seçim kriterlerinin performansının artması kriterlerin tutarlılık özelliğini burada da desteklemektedir.

#### 4.4.1 Diyabet verisi

Bu kısımda çoklu doğrusal regresyon modelleri için değişken seçimi problemlerinde sıklıkla kullanılan diyabet veri seti (Efron vd. 2004) üzerinde  $ICOMP$  ve  $ICOMP_{SN}$  kriterleri kullanılarak değişken seçimi yapılmıştır.

Veri seti  $p = 10$  açıklayıcı değişken (yaş, cinsiyet, vücut kütle indeksi, ortalama kan basıncı ve altı kan serumu ölçümleri) ile  $n = 442$  diyabetli hastayı ve başlangıçtan bir yıl sonra hastalığın ilerleyişinin kantitatif bir ölçütü olan bir açıklanan değişkeni ( $y$ ) kapsar. Çizelge 4.4'te, diyabet veri seti için kurulan regresyon modelindeki açıklayıcı değişkenler verilmiştir.

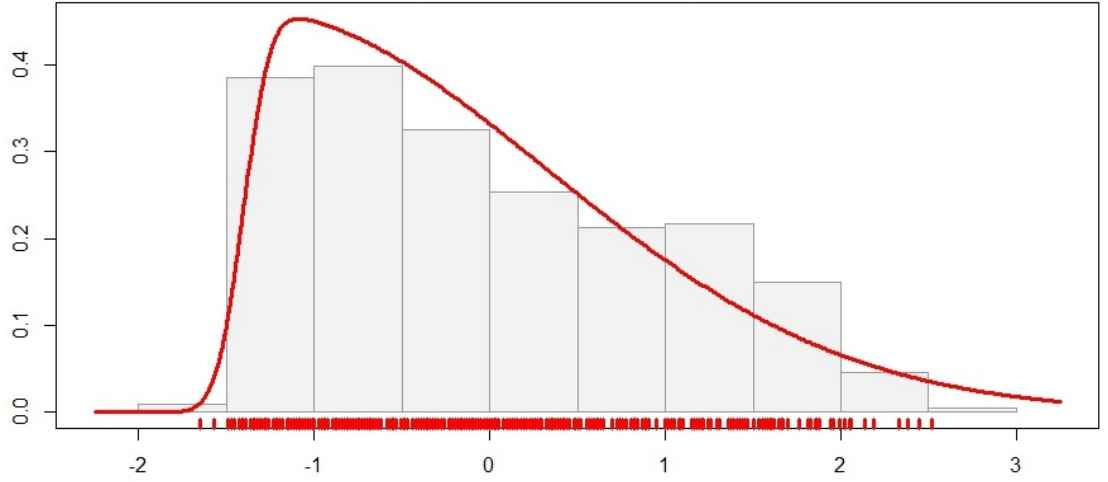
Çizelge 4.4 Diyabet veri seti değişkenleri

---

|          |                              |
|----------|------------------------------|
| $x_1$    | : yaş                        |
| $x_2$    | : cinsiyet                   |
| $x_3$    | : bmi (vücut kütle indeksi ) |
| $x_4$    | : bp (ortalama kan basıncı)  |
| $x_5$    | : s1 (tc)                    |
| $x_6$    | : s2 (ldl)                   |
| $x_7$    | : s3 (hdl)                   |
| $x_8$    | : s4 (tch)                   |
| $x_9$    | : s5 (ltg)                   |
| $x_{10}$ | : s6 (glu)                   |

---

Açıklanan değişkenin çarpık normal dağılıma uygunluğu şekil 4.2'den görülebilir.



Şekil 4.2 Diyabet veri seti y değişkenine ilişkin histogram

Bu veri seti için literatürde mevcut olan sonuçlar ve önerilen *ICOMP* kriterleri ile elde edilen sonuçlar çizelge 4.5'te özetlenmiştir.

Çizelge 4.5 Diyabet veri seti için regresyon modelinde değişken seçimi sonuçları

| Model Seçim Kriterleri    | Seçilen Altkümeler                             |
|---------------------------|------------------------------------------------|
| <i>AIC</i>                | : $\{x_2, x_3, x_4, x_5, x_6, x_9\}$           |
| <i>SBC</i> (Dey, 2013)    | : $\{x_2, x_3, x_4, x_7, x_9\}$                |
| <i>ICOMP</i>              | : $\{x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9\}$ |
| <i>ICOMP<sub>SN</sub></i> | : $\{x_2, x_3, x_4, x_5, x_6, x_7, x_9\}$      |

Çizelge 4.5'te görüldüğü üzere, model seçimi için *AIC* kullanıldığında  $x_2, x_3, x_4, x_5, x_6, x_9$  değişkenleri seçilmiştir. Son olarak, *ICOMP* ve *ICOMP<sub>SN</sub>* model seçim kriterleri kullanıldığında *ICOMP*  $x_1$  ve  $x_{10}$  dışındaki değişkenlerden oluşan modeli ve *ICOMP<sub>SN</sub>* ise  $x_1, x_8$  ve  $x_{10}$  dışındaki değişkenlerden oluşan modeli seçmiştir. Bu sonuçlar, *ICOMP* ve *ICOMP<sub>SN</sub>* arasındaki yakınlığı göstermek ile beraber, *ICOMP<sub>SN</sub>* kriterinin daha basit olan modeli seçtiğini göstermektedir. Ayrıca literatürde yer alan çalışmalara göre LASSO  $x_3, x_4, x_7, x_9$ , Dey (2013)'e göre *SBC*  $x_2, x_3, x_4, x_7, x_9$  ve Hahn ve Carvalho (2014)'ya göre DSS yöntemi,  $x_2, x_3, x_4, x_5, x_6, x_9$  değişkenlerinden oluşan modeli seçmiştir.

#### 4.4.2 Avustralya Spor Enstitüsü verisi

Bu kısımda çarpık dağılımlar ile ilgili çalışmalarda literatürde sıkça kullanılan Avustralya Spor Enstitüsü veri seti (AIS-Australian Institute of Sport Data) (Cook ve Weisberg 1994) ele alınmıştır.

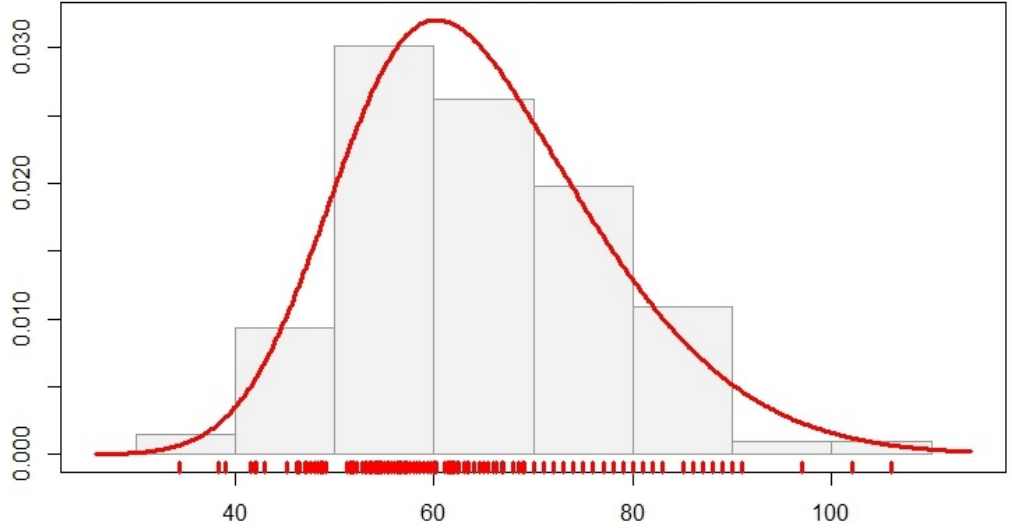
AIS veri setinde Avustralyalı sporcularının  $p = 13$  biyolojik karakterini içeren  $n = 202$  gözlem (100 kadın ve 102 erkek) bulunmaktadır. Çizelge 4.6’da, Avustralya Spor Enstitüsü veri seti için kurulan regresyon modelindeki değişkenler verilmiştir.

Çizelge 4.6 AIS veri seti değişkenleri

|          |                                       |
|----------|---------------------------------------|
| $y$      | : yağsız vücut kitlesi                |
| $x_1$    | : rcc (kırmızı hücre sayısı)          |
| $x_2$    | : wcc (beyaz hücre sayısı)            |
| $x_3$    | : Hc (Hematokrit)                     |
| $x_4$    | : Hg (Hemoglobin)                     |
| $x_5$    | : Fe (plazma ferritin konsantrasyonu) |
| $x_6$    | : bmi (vücut kitle indeksi)           |
| $x_7$    | : sff (deri kıvrımlarının toplamı)    |
| $x_8$    | : bfat (vücut yağ yüzdesi)            |
| $x_9$    | : Ht (boy)                            |
| $x_{10}$ | : Wt (ağırlık)                        |

Çizelge 4.6’da verilen değişkenler ile oluşturulan regresyon modelinde  $AIC$ ,  $ICOMP$  ve önerilen  $ICOMP_t$  ve  $ICOMP_{St}$  kriterleri kullanılarak değişken seçimi yapılmıştır.

Regresyon modelinde açıklanan değişken olan yağsız vücut kitlesi değişkeninin çarpık  $t$  dağılıma uygunluğu şekil 4.3’te görülmektedir.



Şekil 4.3 AIS veri seti y değişkenine ilişkin histogram

$AIC$ ,  $ICOMP$  ve önerilen  $ICOMP_t$  ve  $ICOMP_{St}$  kriterleri kullanılarak değişken seçimi yapılmıştır ve sonuçlar çizelge 4.7’de verilmiştir.

Çizelge 4.7 AIS veri seti için regresyon modelinde değişken seçimi sonuçları

| Model Seçim Kriterleri | Seçilen Altkümeler                           |
|------------------------|----------------------------------------------|
| $AIC$                  | : $\{x_1, x_4, x_6, x_7, x_8, x_9, x_{10}\}$ |
| $ICOMP$                | : $\{x_4, x_6, x_7, x_8, x_9, x_{10}\}$      |
| $ICOMP_t$              | : $\{x_1, x_4, x_7, x_8, x_9, x_{10}\}$      |
| $ICOMP_{St}$           | : $\{x_7, x_8, x_9, x_{10}\}$                |

Çizelge 4.7’de görüldüğü üzere  $AIC$   $x_1, x_4, x_6, x_7, x_8, x_9, x_{10}$  ve  $ICOMP$  bilgi kriteri  $x_4, x_6, x_7, x_8, x_9, x_{10}$  değişkenlerini seçmiştir. Student  $t$  dağılımına dayalı önerilen  $ICOMP_t$  ise ve çarpık  $t$  dağılıma dayalı  $ICOMP_{St}$  kriterinin seçtiği değişkenler ile beraber  $x_1$  ve  $x_4$  değişkenlerini içeren modeli seçmiştir. Burada serbestlik derecesi  $\nu = 3$  alınmıştır. Tüm kriterler ortak olarak  $x_7, x_8, x_9, x_{10}$  değişkenlerini içeren modelleri seçmiştir.  $ICOMP$  ve  $ICOMP_t$  kriterler birbirine yakın modelleri seçmiştir.  $ICOMP_{St}$  ise tüm kriterlerin seçtiği modellerin ortak değişkenlerinden oluşan en basit modeli seçmiştir.

## 5. ROBUST *ICOMP* BİLGİ KRİTERİ

### 5.1 M, S ve MM Tahmin Edicilerine Dayalı *ICOMP* Bilgi Kriterleri

İstatistiksel çıkarımlar yaparken gözlemler kadar modelleme varsayımları da aynı derecede önemlidir. Bu varsayımlar dağılımsal ve model formu ile ilgili olmak üzere iki başlık altında toplanabilir. Model formuna ilişkin varsayımlara modelin doğrusal olması örnek verilebilir. Doğrusal modellerde bağımsızlık, normallik, sabit varyanslılık gibi varsayımlar da dağılımsal varsayımlara örnek verilebilir. Ancak bu varsayımların tamamı uygulamada çoğu zaman sağlanmaz. Diğer taraftan veride aykırı değerlerle de karşılaşılabilir. Alışlagelen değişken seçim yöntemlerinin bu gibi durumlardan etkilendikleri bilinmektedir. Bu nedenle robust model seçim kriterlerine ihtiyaç duyulmaktadır. *ICOMP* kriteri de diğer klasik yaklaşımlarda olduğu gibi ML tahmin edicilerine ve dağılım varsayımına dayanmaktadır. Bu nedenle bu kısımda *ICOMP* model seçim kriterinin dağılım varsayımlarından sapmalara ve aykırı değerlere karşı daha robust olması için robust M, S ve MM tahmin edicilerine dayalı genellemeleri önerilmiş, simülasyon çalışması ve gerçek veri örneği ile önerilen kriterlerin ve orijinal *ICOMP* kriterinin performansları karşılaştırılmıştır.

Çoklu doğrusal regresyon modelinde sıklıkla karşılaşılan bir diğer problem ise açıklayıcı değişkenler arasında çoklu doğrusal bağlantının olmasıdır. Bu durumda, klasik analizler sonucunda önemli değişkenlerin anlamlı olmadığı görülebilir ve standart hatalar büyük olabilir. Bununla beraber veri seti aykırı değerler de içerebilir. Bu durumda klasik değişken seçim yöntemleri sağlıklı sonuçlar vermeyebilir. Bu nedenle bu kısımda regresyon modelinde çoklu doğrusal bağlantı ve aykırı değer olduğu durumlar için robust ridge tahmin edicilerine dayalı *ICOMP* bilgi kriterinin bir robust formu önerilmiştir. Simülasyon çalışması ve gerçek veri örneği ile önerilen kriterlerin ve orijinal *ICOMP* kriterinin performansları karşılaştırılmıştır.

### 5.1.1 M tahmin edicisine dayalı *ICOMP* bilgi kriteri

ML tahmin edicilerinin bir genellemesi olan M tahmin edicileri, ML tahmin edicilerine kıyasla varsayımdan sapmalara karşı daha robust tahmin edicilerdir. Bu nedenle bu kısımda robust M tahmin edicilerine dayalı robust *ICOMP* bilgi kriteri tanımlanmıştır.

*ICOMP* model seçim kriterinde  $-\log L(\hat{\theta}) = \sum_{i=1}^n -\log f(x_i, \hat{\theta})$  olduğu göz önüne alınırsa M tahmin edicilerinin tanımlanmasına benzer düşünce ile  $-\log f(\cdot)$  fonksiyonu yerine  $\rho(\cdot)$  amaç fonksiyonunun yazılmasıyla *ICOMP*'ın M tahmin edicilerine dayalı genellemesi

$$RICOMP_M = 2 \sum_{i=1}^n \rho \left( \frac{y_i - \mathbf{x}_i^T \hat{\beta}_M}{\hat{\sigma}_M} \right) + 2C(\hat{\Sigma}_M) \quad (5.1)$$

biçiminde tanımlanır. Burada  $\hat{\Sigma}_M, \hat{\theta}_M = (\hat{\beta}_M, \hat{\sigma}_M)$  tahmin edicisinin (2.40) ifadesinde verilen kovaryans matrisinin tahmin edicisini göstermektedir.

Orijinal *ICOMP* bilgi kriterinde olduğu gibi, (5.1) eşitliğinin sağ tarafındaki ilk terim uyum iyiliğini, ikinci terim ise ceza terimini göstermektedir.  $RICOMP_M$  kriterine göre en iyi model, en küçük  $RICOMP_M$  değerini veren modeldir.

Eşitlik (5.1), robust M tahmin edicileri ile Bozdoğan'ın (1988) bilgi karmaşıklığı kriterini birleştiren bir bilgi kriteridir. Bunun sonucu olarak *ICOMP* kriterinin cezbedici özelliklerinin yanı sıra, modelin varsayımlardan sapma ve aykırı değerlere karşı robust olma özelliğine de sahiptir.

M tahmin edicileri y yönündeki aykırı değerlere karşı robust olduğundan *ICOMP* bilgi kriterinin M tahmin edicisine dayalı robust genellemesi olan  $RICOMP_M$  kriteri veri setinde y yönünde aykırı değerlere karşı robusttır. Ayrıca, robust Akaike-tipi kriterler ile karşılaştırıldığında ceza teriminde parametrelerin korelasyonunu içermesiyle de avantaj sağlamaktadır. Bu sayede değişkenler arasındaki ilişkiden daha az etkilenir. Bu özelliği ile robust Akaike-tipi kriterlerden ayrılır.

### 5.1.2 S tahmin edicisine dayalı *ICOMP* bilgi kriteri

M tahmin edicileri, y yönündeki aykırı değerlere karşı robusttur. Ancak bu tahmin edicilerin x yönündeki aykırı değerlerin etkisine karşı kırılma noktası  $1/n$ 'dir. Bu durumda kırılma noktası daha yüksek olan robust bir tahmin yöntemine dayalı bir model seçim kriteri ile aykırı değerlere karşı daha iyi sonuçlar elde edilebilir. Bu düşünceden yola çıkarak, bu kısımda S tahmin edicilerine dayalı robust *ICOMP* ( $RICOMP_S$ ) bilgi kriteri önerilmiştir.

S tahmin edicilerinin elde edilmesi için kullanılan

$$\frac{1}{n} \sum_{i=1}^n \rho \left( \frac{r_i}{\hat{\sigma}_S} \right) = K$$

kısıtı göz önüne alınsın. Buradan

$$\sum_{i=1}^n \rho \left( \frac{r_i}{\hat{\sigma}_S} \right) = nK$$

elde edilir.  $RICOMP_M$  kriterinin tanımlanmasına benzer düşünce ile hareket edilirse, (5.1) eşitliğinin sağ tarafındaki ilk terim yani  $2 \sum_{i=1}^n \rho \left( \frac{r_i}{\hat{\sigma}_S} \right)$  tüm modeller için sabit olacaktır. Dolayısıyla değişken seçimi sadece model karmaşıklığına bağlı olacaktır. Bu nedenle (3.2) eşitliğinde verilen robust AIC bilgi kriterinin elde edilmesine benzer düşünce ile *ICOMP*'ın S tahmin edicilerine dayalı genellemesi

$$RICOMP_S = 2n \log \hat{\sigma}_S + 2C \left( \hat{\Sigma}_S \right) \quad (5.2)$$

biçiminde tanımlanır. Burada parametrelerin S tahmin edicileri  $\hat{\theta}_S = \left( \hat{\beta}_S, \hat{\sigma}_S \right)$  ve S tahmin edicilerinin (2.41), (2.42) eşitlikleri ile elde edilen kovaryans matrisinin tahmin edicisi  $\hat{\Sigma}_S$  kullanılmıştır.  $RICOMP_S$  kriterine göre en iyi model, en küçük  $RICOMP_S$  değerini veren modeldir.

S tahmin edicilerinin x yönündeki aykırı değerlere karşı kırılma noktası yüksek olduğundan *ICOMP* bilgi kriterinin S tahmin edicisine dayalı robust genellemesi olan  $RICOMP_S$  kriterinin de x yönünde aykırı değerlere karşı kırılma noktası yüksek olacaktır. Kısım 2.3.2'de bahsedildiği gibi uygun ayar sabiti değerleri ile yüksek kırılma noktası elde edilebilir.

### 5.1.3 MM tahmin edicisine dayalı *ICOMP* bilgi kriteri

M tahmin edicilerinin yüksek asimptotik görelî etkinliklerinin yanında düşük kırılma noktasına sahip olması, S tahmin edicilerinin ise yüksek kırılma noktasına sahip olmalarının yanısıra asimptotik görelî etkinliklerinin düşük olması göz önüne alınarak bu kısımda bu iki tahmin edicinin robustluk özelliklerini bir araya getiren MM tahmin edicisi ele alınmıştır. MM tahmin edicilerine dayalı robust *ICOMP* ( $RICOMP_{MM}$ ) bilgi kriteri önerilmiştir.

*ICOMP*'ın MM tahmin edicilerine dayalı genellemesi

$$RICOMP_{MM} = 2 \sum_{i=1}^n \tilde{\rho} \left( \frac{y_i - \mathbf{x}_i^T \hat{\boldsymbol{\beta}}_{MM}}{\hat{\sigma}_{MM}} \right) + 2C(\hat{\boldsymbol{\Sigma}}_{MM}) \quad (5.3)$$

biçiminde elde edilir. Parametrelerin MM tahmin edicileri  $\hat{\boldsymbol{\theta}}_{MM} = (\hat{\boldsymbol{\beta}}_{MM}, \hat{\sigma}_{MM})$  olmak üzere  $\hat{\boldsymbol{\Sigma}}_{MM}$ , MM tahmin edicilerinin (2.43) eşitliği ile elde edilen kovaryans matrisinin tahmin edicisidir.  $RICOMP_{MM}$  kriterine göre en iyi model, en küçük  $RICOMP_{MM}$  değerini veren modeldir.

MM tahmin edicileri hem x hem y yönündeki aykırı değerlere karşı robust olduğundan,  $RICOMP_{MM}$  kriteri de veri seti hem x hem y yönünde aykırı değer içerdiğinde başarılı performansa sahip olur.

## 5.2 Simülasyon Çalışması

Bu kısımda ML tahmin edicisine dayalı orijinal *ICOMP* model seçim kriteri ile önerilen robust *ICOMP* seçim kriterlerinin performansları doğru modelin seçim sıklığına göre karşılaştırılmıştır. Simülasyon çalışması Matlab2013a programlama dili aracılığıyla gerçekleştirilmiştir.

Simülasyon çalışması için örneklem hacmi  $n = 20, 50$  ve  $100$  ve tekrar sayısı  $100$  alınmıştır. Hata terimlerinin dağılımı  $\varepsilon_i \sim N(0, 1)$  standart normal dağılımdır.

Değişken sayısı  $p = 5$  olmak üzere parametre vektörü  $\beta = (1, 1, 1, 0, 0)^T$  alınmıştır. Her bir açıklayıcı değişkene ilişkin gözlemler birbirinden bağımsız olarak standart normal dağılımdan üretilmiştir.

$\{(x_i, y_i), i = 1, 2, \dots, n\}$  verisi (2.2)'de verilen çoklu doğrusal regresyon modeli ele alınarak hesaplanmıştır.

Simülasyon çalışmasında aşağıdaki 4 durum göz önüne alınmıştır.

- (1) Veri setinde aykırı değer yoktur.
- (2) Veri seti %10-%20-%30 oranında x yönünde aykırı değer içermektedir.
- (3) Veri seti %10-%20-%30 oranında y yönünde aykırı değer içermektedir.
- (4) Veri seti %10-%20-%30 oranında x ve y yönünde aykırı değer içermektedir.

(2) durumunda  $x_{1,i}, x_{2,i}, x_{3,i}, x_{4,i}, x_{5,i}$  gözlemlerinin rasgele %10-%20-%30'u,  $N(10, 1)$  dağılımından üretilen değerler ile değiştirilerek veri seti oluşturulmuştur. Burada  $y_i$ , gözlemleri değiştirilmemiştir.

(3) durumunda  $y_i$  gözlemlerinin rasgele %10-%20-%30'u,  $N(10, 1)$  dağılımından üretilen değerler ile değiştirilerek veri seti oluşturulmuştur. Burada  $x_{1,i}, x_{2,i}, x_{3,i}, x_{4,i}, x_{5,i}$  gözlemleri değiştirilmemiştir.

(4) durumunda  $x_{1,i}, x_{2,i}, x_{3,i}, x_{4,i}, x_{5,i}$  gözlemlerinin rasgele %10-%20-%30'u,  $N(10, 1)$  dağılımından üretilen değerler ile değiştirilir ve bu değerler kullanılarak  $y_i$  gözlemleri regresyon modelinden hesaplanır.

M tahminleri  $c = 4.685$  için Tukey'in iki ağırlıklı fonksiyonu kullanılarak elde edilmiştir.

S tahminleri  $c = 1.55$  için (böylece kırılma noktasının %50 olması sağlanmıştır) Tukey'in iki ağırlıklı fonksiyonu kullanılarak elde edilmiştir.

MM tahminleri için öncelikle  $c = 1.55$  (kırılma noktasının %50 olması için) alınarak Tukey'in iki ağırlıklı fonksiyonu kullanılarak artıkların ölçek tahmini elde edilmiştir. Daha sonra %95 etkinlik için  $c = 4.685$  alınarak Tukey'in iki ağırlıklı fonksiyonun kullanılmasıyla regresyon katsayılarının MM tahminleri elde edilmiştir.

$2^5 - 1 = 31$  olası model için kriter değerleri hesaplanmış ve 100 denemede bu modeller arasından doğru modelin seçim sıklıkları her bir kriter için hesaplanmıştır. Çizelge 5.1-5.3'te veri setinin aykırı değer içermediği (% 0) ve içerdiği (% 10, %20 %30) durumlar için orijinal *ICOMP* kriterinin ve Kısım 5.1' de önerilen M, S ve MM tahmin yöntemlerine dayalı robust *ICOMP* (*RICOMP*) kriterlerinin doğru modeli seçim sıklıkları listelenmiştir.

Çizelge 5.1 X yönünde aykırı değer olduğunda *RICOMP* kriterlerinin doğru modeli seçme sıklıkları

| $n$ | %  | <i>ICOMP</i> | <i>RICOMP</i> <sub>M</sub> | <i>RICOMP</i> <sub>S</sub> | <i>RICOMP</i> <sub>MM</sub> |
|-----|----|--------------|----------------------------|----------------------------|-----------------------------|
| 20  | 0  | 0.88         | 0.52                       | 0.58                       | 0.64                        |
|     | 10 | 0.35         | 0.48                       | 0.86                       | 0.90                        |
|     | 20 | 0.30         | 0.44                       | 0.84                       | 0.86                        |
|     | 30 | 0.27         | 0.42                       | 0.83                       | 0.85                        |
| 50  | 0  | 0.90         | 0.53                       | 0.61                       | 0.66                        |
|     | 10 | 0.38         | 0.50                       | 0.90                       | 0.93                        |
|     | 20 | 0.32         | 0.45                       | 0.86                       | 0.89                        |
|     | 30 | 0.28         | 0.43                       | 0.85                       | 0.86                        |
| 100 | 0  | 0.93         | 0.55                       | 0.63                       | 0.69                        |
|     | 10 | 0.40         | 0.51                       | 0.91                       | 0.94                        |
|     | 20 | 0.33         | 0.46                       | 0.87                       | 0.90                        |
|     | 30 | 0.30         | 0.44                       | 0.86                       | 0.87                        |

Çizelge 5.1- 5.3'te verilen sonuçlara göre, beklendiği gibi, orijinal *ICOMP* kriteri aykırı değeri olmayan veriler için *RICOMP*<sub>M</sub>, *RICOMP*<sub>S</sub>, *RICOMP*<sub>MM</sub> kriterlerinden daha iyi performansa sahiptir. Bunun yanı sıra veri setinde aykırı değer olması halinde orijinal *ICOMP*, LS tahmin edicisine dayandığından beklendiği gibi en düşük performansa sahiptir ve aykırı değer oranı arttıkça, doğru modeli seçme sıklığı azalmaktadır.

Çizelge 5.1'e göre veri setinin en az miktarda (%10) x yönünde aykırı değer içermesi durumunda  $RICOMP_M$  ile orijinal  $ICOMP$  kriterinin yakın performans sergilediği gözlenirse de  $RICOMP_M$  orijinal  $ICOMP$  kriterine kıyasla daha fazla doğru modeli seçmektedir. Veri setinin x yönünde aykırı değer içerdiği tüm durumlar için robust kriterler arasında  $RICOMP_M$  en düşük performansa sahiptir.  $RICOMP_M$  M tahminine dayandığından, M tahmin edicilerinin x yönündeki sapan gözlemlere karşı hassas olması bu sonucu desteklemektedir. Diğer taraftan  $RICOMP_S$  ve  $RICOMP_{MM}$  kriterlerinin, aykırı değerlere karşı oldukça dayanıklı oldukları görülmektedir. Burada  $RICOMP_{MM}$  en iyi performansa sahiptir.

Çizelge 5.2 Y yönünde aykırı değer olduğunda  $RICOMP$  kriterlerinin doğru modeli seçme sıklıkları

| $n$ | %  | $ICOMP$ | $RICOMP_M$ | $RICOMP_S$ | $RICOMP_{MM}$ |
|-----|----|---------|------------|------------|---------------|
| 20  | 0  | 0.91    | 0.62       | 0.67       | 0.70          |
|     | 10 | 0.28    | 0.82       | 0.73       | 0.90          |
|     | 20 | 0.25    | 0.80       | 0.70       | 0.84          |
|     | 30 | 0.16    | 0.78       | 0.69       | 0.80          |
| 50  | 0  | 0.92    | 0.66       | 0.68       | 0.72          |
|     | 10 | 0.29    | 0.86       | 0.74       | 0.92          |
|     | 20 | 0.26    | 0.81       | 0.72       | 0.87          |
|     | 30 | 0.18    | 0.80       | 0.71       | 0.83          |
| 100 | 0  | 0.95    | 0.69       | 0.69       | 0.74          |
|     | 10 | 0.31    | 0.87       | 0.78       | 0.93          |
|     | 20 | 0.28    | 0.83       | 0.74       | 0.88          |
|     | 30 | 0.19    | 0.82       | 0.73       | 0.86          |

Çizelge 5.2'de verilen sonuçlara göre, veri setinin y yönünde aykırı değer içermesi durumunda  $RICOMP_M$  ile  $RICOMP_{MM}$  kriterleri yakın performans sergilemekle beraber,  $RICOMP_{MM}$  en iyi performansa sahiptir.

Çizelge 5.3 X-Y yönünde aykırı değer olduğunda RCOMP kriterlerinin doğru modeli seçme sıklıkları

| $n$ | %  | $ICOMP$ | $RICOMP_M$ | $RICOMP_S$ | $RICOMP_{MM}$ |
|-----|----|---------|------------|------------|---------------|
| 20  | 0  | 0.91    | 0.52       | 0.56       | 0.60          |
|     | 10 | 0.32    | 0.74       | 0.81       | 0.90          |
|     | 20 | 0.26    | 0.68       | 0.77       | 0.87          |
|     | 30 | 0.19    | 0.66       | 0.75       | 0.86          |
| 50  | 0  | 0.94    | 0.55       | 0.58       | 0.64          |
|     | 10 | 0.33    | 0.76       | 0.82       | 0.92          |
|     | 20 | 0.28    | 0.73       | 0.80       | 0.89          |
|     | 30 | 0.21    | 0.70       | 0.77       | 0.87          |
| 100 | 0  | 0.96    | 0.57       | 0.60       | 0.72          |
|     | 10 | 0.36    | 0.77       | 0.83       | 0.93          |
|     | 20 | 0.32    | 0.75       | 0.81       | 0.90          |
|     | 30 | 0.25    | 0.72       | 0.78       | 0.88          |

Çizelge 5.3'te verilen sonuçlara göre veri setinin hem x hem y yönünde aykırı değer içermesi durumunda,  $RICOMP_{MM}$  kriteri aykırı değerlere karşı başarılı bir şekilde değişken seçimi yaparak ele alınan diğer kriterler ile kıyaslandığında en iyi performansa sahiptir. MM tahminine dayandığından, MM tahmin edicilerinin hem x hem y yönündeki aykırı değerlere karşı dayanıklı olması bu sonucu desteklemektedir.

Aykırı gözlemlerin olduğu tüm durumlarda en yüksek performans  $RICOMP_{MM}$  kriterinin olsa da  $RICOMP_S$  ile yakın sonuçlar elde edilmektedir.  $RICOMP_S$  ve  $RICOMP_{MM}$  kriterleri verinin yüksek oranda (%30) aykırı değer içermesi durumunda daha iyi sonuçlar vermektedir. Bu yöntemlerde, tahmin yönteminin kırılma noktası, verilerdeki aykırı değerlerin yüzdesine uygun olarak ayarlanmalıdır. Önerilen robust yöntemler karşılaştırıldığında  $RICOMP_S$  ve  $RICOMP_{MM}$  kriterleri M tahmin edicisine dayalı olan  $RICOMP_M$  kriterine göre tüm durumlar için daha iyi performansa sahiptir. Bununla birlikte, çizelge 5.1-5.3'te görüldüğü üzere örneklem hacmi arttıkça ele alınan kriterlerin performansları da artmaktadır.

### 5.3 Çoklu Doğrusal Bağlantı ve Ridge Regresyon

Regresyon analizi genellikle açıklayıcı değişkenlerin bağımsız olduğu varsayımı altında yapılmaktadır. Ancak bu varsayım gerçek veri uygulamalarında her zaman sağlanmaz. Bu gibi durumlarda, açıklayıcı değişkenler arasındaki ilişkiler zayıf olduğunda bu ilişkiler gözardı edilebilir. Ancak bu ilişkiler kuvvetli olduğunda gözardı edilmesi istatistiksel sonuç çıkarımında sorunlara sebep olacaktır.

Açıklayıcı değişkenlerin tamamının veya bir kısmının arasında yüksek doğrusal bağlantılar olması durumunda çoklu doğrusal bağlantı (multicollinearity) sorunun var olduğu söylenir (Montgomery vd. 2001). Çoklu doğrusal bağlantı mevcut olduğunda, klasik analizler sonucunda önemli değişkenlerin anlamlı olmadığı görülebilir ve standart hatalar büyük olabilir. Bu nedenle, bu kısımda regresyon modelinde çoklu doğrusal bağlantı olduğu durumda değişken seçimi problemi ele alınmıştır.

#### 5.3.1 Çoklu doğrusal bağlantı sorunu

##### Çoklu doğrusal bağlantı

$\mathbf{y} = [y_1, y_2, \dots, y_n]^T$  açıklanan değişken vektörünü,  $\mathbf{X} = [X_1, X_2, \dots, X_p]$  tam sütun ranklı açıklayıcı değişkenler matrisini,  $\boldsymbol{\beta} = [\beta_1, \beta_2, \dots, \beta_p]^T$  parametre vektörünü ve  $\boldsymbol{\epsilon} = [\epsilon_1, \epsilon_2, \dots, \epsilon_n]^T$  hata vektörünü göstermek üzere

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

örneklem için çoklu doğrusal regresyon modeli göz önüne alınsın.  $\boldsymbol{\epsilon}$  hata terimlerinin beklenen değerinin sıfır  $E(\boldsymbol{\epsilon}) = \mathbf{0}$  ve  $\boldsymbol{\epsilon}$  hata terimlerinin varyans-kovaryans matrisinin  $Cov(\boldsymbol{\epsilon}) = \sigma^2 \mathbf{I}_{n \times n}$  olduğu varsayımları sağlansın.

Çoklu doğrusal bağlantı  $\mathbf{X} = [X_1, X_2, \dots, X_p]$  tasarım matrisindeki sütunların doğrusal ilişkisi olarak tanımlanabilir (Montgomery vd. 2001).

Yani  $a_1, a_2, \dots, a_p$  katsayılarının hepsi sıfır olmamak üzere

$$\sum_{j=1}^p a_j X_j = 0$$

eşitliği gerçekleşiyorsa  $X_1, X_2, \dots, X_p$  bağımsız değişkenleri doğrusal bağımlıdır ve tam çoklu doğrusal bağlantı problemi vardır denir. Bu durum  $|\mathbf{X}^T \mathbf{X}| = 0$  yani özdeğerlerden en az birinin sıfır olması anlamına gelir (Rao vd. 2008).  $\mathbf{X}^T \mathbf{X}$  matrisinin rankı  $p$  değerinden küçük olup  $(\mathbf{X}^T \mathbf{X})^{-1}$  matrisi mevcut değildir. Dolayısıyla tam çoklu doğrusal bağlantı durumunda katsayılar belirlenemez ve standart hata tanımlanamaz.

$$\sum_{j=1}^p a_j X_j \simeq 0$$

olması durumunda  $\mathbf{X}$  matrisindeki ilişki, yaklaşık çoklu doğrusal bağlantı olarak adlandırılır. Bu durum  $|\mathbf{X}^T \mathbf{X}| \simeq 0$  yani özdeğerlerden en az birinin sıfıra çok yakın olması anlamına gelir (Rao vd. 2008). Yaklaşık çoklu doğrusal bağlantı durumu uygulamada daha yaygın bir durumdur ve bu durum parametrelerin varyans ve kovaryanslarının büyük olmasına dolayısıyla büyük güven aralıklarına, yüksek  $R^2$  değerine karşın açıklayıcı değişkenlerin çok azının (düşük t değerine göre) anlamlı çıkmasına neden olmaktadır. Özetle modelden yanlış değişkenlerin çıkarılmasına ve hatalı modellemeye yol açmaktadır (Gujarati 2003).

### **Çoklu doğrusal bağlantının nedenleri**

Çoklu doğrusal bağlantının nedenleri dört başlıkta toplanabilir:

1. Veri toplama yöntemi,
2. Model ya da kitle üzerindeki kısıtlamalar,
3. Modelin özellikleri,
4. Geniş tanımlı model ( $p > n$ ) (Montgomery vd. 2004).

## Çoklu doğrusal bağlantının etkileri

Çoklu doğrusal bağlantının etkileri aşağıdaki gibi özetlenebilir (Montgomery vd. 2004):

1. Regresyon katsayılarının LS tahminleri mutlak değerce çok büyük olma eğilimindedir. Ayrıca tahminler gerçek parametre değerinden uzaklaşır.
2. Regresyon katsayılarının LS tahmin edicileri doğrusal yansız olmalarına rağmen büyük varyans ve kovaryanslara sahiptir.
3. Varyans ve kovaryansın büyük olması, güven aralıklarının daha geniş olmasına sebep olacaktır.
4. En az bir regresyon katsayısı için düşük  $t$  değerinin çıkmasına karşın  $R^2$  değeri yüksek olacağı için model hatalı olsa da modelin uygun bir model olduğu yanılgısına neden olacaktır.

## Çoklu doğrusal bağlantının saptanması

Literatürde çoklu bağlantının tespit edilmesi için birçok yöntem önerilmiştir. Bunlardan en güvenilir ve yaygın olanları aşağıda verilmiştir.

**$R^2$  Belirlilik Katsayısı :**  $R^2$  değerinin yüksek çıkmasına rağmen ya hiç ya da birkaç tane katsayının  $t$  istatistiklerinin anlamlılık değerlerinin sıfırdan farklı olması çoklu bağlantının bir göstergesidir (Gujarati 2003).

**Varyans Büyütme Faktörü (Variance Inflation Factors-VIF):**  $C = (\mathbf{X}^T \mathbf{X})^{-1}$  matrisinin köşegen elemanları

$$VIF_j = C_{jj} = \frac{1}{1 - R_j^2}, j = 1, 2, \dots, p$$

şeklinde gösterilebilir. Burada  $R_j^2$ ,  $X_j$  açıklayıcı değişkeninin geri kalan  $p-1$  değişken üzerindeki regresyonundan elde edilen çoklu belirtme katsayısıdır. Eğer  $X_j$  açıklayıcı değişkeninin geri kalan  $p-1$  değişkenin bir altkümesi arasında yaklaşık çoklu doğrusal bağlantı var ise  $R_j^2$  değeri 1'e yakın olacaktır. VIF çoklu bağlantının, regresyon katsayılarının tahmin edicilerinin varyanslarını ne kadar arttırdığını göstermektedir. Dolayısıyla bir ya da daha fazla yüksek VIF değeri, çoklu bağlantıyı

işaret etmektedir. VIF değerlerinden herhangi biri 10 un üzerinde ise ciddi bir çoklu doğrusal bağlantı olduğu söylenebilir.

### 5.3.2 Ridge regresyon

Çoklu doğrusal bağlantı durumunda  $\beta$  parametre vektörünün LS tahmin edicisi  $\hat{\beta} = (\mathbf{X}^T \mathbf{X})^{-1}(\mathbf{X}^T \mathbf{y})$  güvenilir değildir. Çoklu doğrusal bağlantı problemi için önerilen yöntemlerden en popüler olanı Ridge tahmin edicisidir. Hoerl ve Kennard (1970a, 1970b) tarafından önerilen bu yöntem  $\mathbf{X}^T \mathbf{X}$  matrisinin köşegen elemanlarına bir pozitif  $k$  sayısının eklenmesi esasına dayanmaktadır.  $k$  parametresi yanalılık parametresi olarak adlandırılır. Ridge tahmin edicisi

$$(\mathbf{X}^T \mathbf{X} + k\mathbf{I}_n)\hat{\beta}_{Ridge} = \mathbf{X}^T \mathbf{y}$$

denkleminde

$$\hat{\beta}_{Ridge} = (\mathbf{X}^T \mathbf{X} + k\mathbf{I}_n)^{-1}(\mathbf{X}^T \mathbf{y}) \quad (5.4)$$

olarak elde edilir. Ridge tahmin edicisi

$$\begin{aligned} \hat{\beta}_{Ridge} &= (\mathbf{X}^T \mathbf{X} + k\mathbf{I}_n)^{-1}(\mathbf{X}^T \mathbf{y}) \\ &= (\mathbf{X}^T \mathbf{X} + k\mathbf{I}_n)^{-1}(\mathbf{X}^T \mathbf{X})\hat{\beta}_{LS} \\ &= \mathbf{L}_k \hat{\beta}_{LS} \end{aligned}$$

şeklinde yazılabildiğinden LS tahmin edicisinin bir doğrusal dönüşümüdür.

$E(\hat{\beta}_{Ridge}) = E(\mathbf{L}_k \hat{\beta}_{LS}) = \mathbf{L}_k \beta$  olacağından ridge tahmin edicisinin yanlı bir tahmin edici olacağı açıktır. Buradaki temel fikir tahmin edicinin varyansında önemli derecede bir azaltma için küçük bir yan miktarını göze almaktır. Başka bir deyişle ridge regresyon ile tahminlerin varyansı azaltılırken,  $k$  ile orantılı olarak yanlı tahminler elde edilmektedir. Bunun sonucunda iki durum ortaya çıkmaktadır: yansız büyük varyanslı veya yanlı düşük varyanslı tahmin ediciler elde edilebilir. LS tahmin edicisi, ridge tahmin edicisinin  $k = 0$  için özel halidir. Ridge tahmin edicisinin hata kareler ortalaması (HKO), LS tahmin edicisinin HKO'dan daha küçük olacak

şekilde bir  $k$  değeri vardır (Hoerl ve Kennard 1970a). Parametrelerin ridge tahmin edicilerinin kovaryans matrisi ise

$$Cov(\hat{\beta}_{Ridge}) = \sigma^2(\mathbf{X}^T \mathbf{X} + k\mathbf{I}_n)^{-1} \mathbf{X}^T \mathbf{X} (\mathbf{X}^T \mathbf{X} + k\mathbf{I}_n)^{-1}$$

şeklinde elde edilir.

Literatürde  $k$  parametresinin seçimi için birçok yöntem önerilmiştir. Hoerl vd. (1975), tarafından önerilen

$$k = \frac{p\hat{\sigma}^2}{\hat{\beta}'\hat{\beta}}$$

eşitliği literatürde yaygın olarak kullanılmaktadır. Burada  $\hat{\sigma}^2 = \frac{(\mathbf{y} - \mathbf{X}\hat{\beta}_{LS})'(\mathbf{y} - \mathbf{X}\hat{\beta}_{LS})}{n-p}$  eşitliği ile elde edilmektedir.

### 5.3.3 Robust ridge tahmin edicileri

Aykırı gözlemlerden daha az etkilenen robust yöntemler ve çoklu doğrusal bağlantı problemine çözüm olarak geliştirilen ridge regresyon istatistiksel analizde önemli bir yere sahiptir. Ancak bir veri setinde hem aykırı değerler hem de çoklu doğrusal bağlantı problemi söz konusu olduğunda bu yöntemler yalnız başına güvenilir bir analiz için yeterli olmamaktadır. Bu iki problemin aynı anda çözülmesi, her iki problem için önerilen yöntemlerin birleştirilmesiyle mümkündür. Bu gibi durumlar için literatürde robust ridge tahmin edicileri önerilmiştir. Böylece bu tahmin ediciler hem aykırı değerlere karşı robust olacak hem de çoklu doğrusal bağlantı probleminden daha az etkilenecektir.

Literatürde önerilen robust ridge tahmin edicileri şunlardır: Holland (1973) ağırlıklandırılmış ridge regresyon tahmin edicilerini önermiştir. Askin ve Montgomery (1980) M tahmin edicilerine dayalı ridge tahmin edicilerini önermişler ve bu tahmin edicileri ağırlıklandırılmış LS yöntemi ile hesaplamışlardır. Midi ve Zahari (2007), MM tahmin edicilerini ridge tahmin edicileri ile birleştirerek ridge MM (RMM) tahmin edicilerini önermiştir. Maronna (2011), yüksek boyutlu veriler için ridge regresyonda MM tahmin edicilerini önermiştir. Bu çalışmaların dışında Arslan ve

Billor (1996), hem çoklu doğrusal bağlantı hem aykırı değer problemleri ile eş zamanlı başa çıkabilecek iki alternatif ridge tipi GM tahmin edicilerini önermiştir.

Ağırlıklandırılmış robust ridge tahmin edicisi

$$\widehat{\beta}_{R-Ridge} = (\mathbf{X}^T \mathbf{W} \mathbf{X} + k_R \mathbf{I}_n)^{-1} (\mathbf{X}^T \mathbf{W} \mathbf{y}) \quad (5.5)$$

eşitliği ile tanımlanmıştır. Burada  $\mathbf{W}$  robust tahmin edicilerden elde edilen köşegen ağırlık matrisini göstermektedir.  $k_R$  robust yanalılık parametresi

$$k_R = \frac{p \widehat{\sigma}_R^2}{\widehat{\beta}_R' \widehat{\beta}_R}$$

ve  $\widehat{\sigma}_R^2 = \frac{(\mathbf{y} - \mathbf{X} \widehat{\beta}_R)' (\mathbf{y} - \mathbf{X} \widehat{\beta}_R)}{n-p}$  eşitlikleri ile elde edilmektedir.

Parametrelerin ridge tahmin edicilerinin kovaryans matrisi ise

$$Cov \left( \widehat{\beta}_{R-Ridge} \right) = \sigma^2 (\mathbf{X}^T \mathbf{W} \mathbf{X} + k_R \mathbf{I}_n)^{-1} \mathbf{X}^T \mathbf{X} (\mathbf{X}^T \mathbf{W} \mathbf{X} + k_R \mathbf{I}_n)^{-1}$$

şeklinde elde edilir.

Bu çalışmada ağırlık matrisi robust S, M ve MM tahmin edicilerine bağlı olarak elde edilmiştir.

#### 5.4 Çoklu Doğrusal Bağlantı ve Aykırı Gözlem Durumunda Çoklu Doğrusal Regresyon Modellerinde Değişken Seçimi

Bu kısımda çoklu doğrusal regresyon modellerinde çoklu doğrusal bağlantı ve aykırı değer olması durumunda *ICOMP* ile değişken seçimi problemi ele alınmıştır.

Çoklu doğrusal regresyon modellerinde çoklu doğrusal bağlantı problemi ve aykırı gözlem var olması durumunda robust ridge tahmin edicilerine dayalı robust *ICOMP*,

$$RICOMP_R = \sum_{i=1}^n \left( \frac{y_i - \mathbf{x}_i' \widehat{\beta}_{R-Ridge}}{\widehat{\sigma}_{R-Ridge}} \right)^2 + 2C \left( \widehat{Cov} \left( \widehat{\theta} \right) \right) \quad (5.6)$$

eşitliği ile tanımlanmıştır. Burada  $C, \widehat{Cov}(\hat{\theta})$  için maksimal karmaşıklık bilgisi olup

$$C(\widehat{Cov}(\hat{\theta})) = \frac{p}{2} \log \left[ \frac{tr(\widehat{Cov}(\hat{\theta}))}{p} \right] - \frac{1}{2} \log |\widehat{Cov}(\hat{\theta})| \quad (5.7)$$

şeklinde tanımlıdır. Burada  $\widehat{Cov}(\hat{\theta})$  ise

$$\widehat{Cov}(\hat{\theta}) = \begin{bmatrix} \hat{\sigma}_R^2(\mathbf{X}^T \mathbf{W} \mathbf{X} + k \mathbf{I}_n)^{-1} \mathbf{X}^T \mathbf{X} (\mathbf{X}^T \mathbf{W} \mathbf{X} + k_R \mathbf{I}_n)^{-1} & \mathbf{0} \\ \mathbf{0}^T & Var(\hat{\sigma}_R) \end{bmatrix}$$

şeklinde dir.  $RICOMP_R$  kriterine göre en iyi model, en küçük  $RICOMP_R$  değerini veren modeldir.

#### 5.4.1 Simülasyon çalışması

Bu kısımda, çoklu doğrusal regresyon modellerinde çoklu doğrusal bağlantı problemi ve aykırı gözlem var olması durumunda, ML tahmin edicisine dayalı orijinal  $ICOMP$  model seçim kriteri ile önerilen robust ridge tahmin edicilerine dayalı  $RICOMP_R$  seçim kriterlerinin performansları doğru modelin seçim sıklığına göre karşılaştırılmıştır. Simülasyon çalışması Matlab2013a programlama dili aracılığıyla gerçekleştirilmiştir.

Simülasyon çalışması için örneklem hacmi  $n = 20, 50, 100$ , tekrar sayısı 100 ve açıklayıcı değişken sayısı  $p = 5$  olmak üzere parametre vektörü  $\beta = (1, 1, 1, 0, 0)^T$  alınmıştır.

Açıklayıcı değişkenler McDonald ve Galarneau (1975) tarafından önerilen tasarım temel alınarak oluşturulmuştur. Burada öncelikle  $i = 1, 2, \dots, 6$  olmak üzere  $z_i \sim N(0, 1)$  üretilmiştir. Üretilen bu rasgele sayılar

$$x_i = \sqrt{1 - \alpha^2} z_i + \alpha z_6, \quad i = 1, 2, \dots, 5$$

eşitlikleri yardımıyla elde edilmiştir. Burada  $\alpha$  katsayısı çoklu doğrusal bağlantının derecesini belirlemektedir. Simülasyon çalışmasında  $\alpha = 0.5$  alınmıştır.  $\epsilon$  hata terimi standart normal dağılımdan üretilmiştir.

$\{(x_i, y_i) : i = 1, 2, \dots, n\}$  veri seti

$$y = \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_5 x_5 + \epsilon$$

çoklu doğrusal regresyon modeli ele alınarak oluşturulmuştur.

Ağırlıkların hesaplanması için başlangıçta parametrelerin robust M, S ve MM tahminleri hesaplanmış, bu tahminler doğrultusunda elde edilen ağırlıklara dayalı robust ridge tahminleri elde edilmiştir. M tahminleri  $c = 4.685$  için Tukey'in iki ağırlıklı fonksiyonu kullanılarak elde edilmiştir. S tahminleri  $c = 1.55$  için (böylece kırılma noktasının %50 olması sağlanmıştır) Tukey'in iki ağırlıklı fonksiyonu kullanılarak elde edilmiştir. MM tahminleri için öncelikle  $c = 1.55$  (kırılma noktasının %50 olması için) alınarak Tukey'in iki ağırlıklı fonksiyonu kullanılarak artıkların ölçek tahmini elde edilmiştir. Daha sonra %95 etkinlik için  $c = 4.685$  alınarak Tukey'in iki ağırlıklı fonksiyonun kullanılmasıyla regresyon katsayılarının MM tahminleri elde edilmiştir. Robust M, S ve MM tahminlerine ilişkin ağırlıklar kullanılarak robust ridge tahminleri ve tahmini kovaryans matrisi hesaplanmıştır.

Simülasyon çalışmasında aşağıdaki 4 durum göz önüne alınmıştır.

- (1) Veri setinde aykırı değer yoktur.
- (2) Veri seti %10-%20-%30 oranında x yönünde aykırı değer içermektedir.
- (3) Veri seti %10-%20-%30 oranında y yönünde aykırı değer içermektedir.
- (4) Veri seti %10-%20-%30 oranında x ve y yönünde aykırı değer içermektedir.

(2) durumunda  $x_{1,i}, x_{2,i}, x_{3,i}, x_{4,i}, x_{5,i}$  gözlemlerinin rasgele %10-%20-%30'u,  $N(10, 1)$  dağılımından üretilen değerler ile değiştirilerek veri seti oluşturulmuştur. Burada  $y_i$ , gözlemleri değiştirilmemiştir.

(3) durumunda  $y_i$  gözlemlerinin rasgele %10-%20-%30'u,  $N(10, 1)$  dağılımından üretilen değerler ile değiştirilerek veri seti oluşturulmuştur. Burada  $x_{1,i}, x_{2,i}, x_{3,i}, x_{4,i}, x_{5,i}$  gözlemleri değiştirilmemiştir.

(4) durumunda  $x_{1,i}, x_{2,i}, x_{3,i}, x_{4,i}, x_{5,i}$  gözlemlerinin rasgele %10-%20-%30'u,  $N(10, 1)$  dağılımından üretilen değerler ile değiştirilir ve bu değerler kullanılarak  $y_i$  gözlemleri regresyon modelinden hesaplanır.

Çizelge 5.4-5.7’de veri setinin aykırı değer içermediği (% 0), farklı yönlerde (x, y ve x-y yönünde) ve farklı oranlarda (% 10, %20 %30) aykırı değerler içerdiği durumlar için Ridge tahmin edicisine dayalı ( $ICOMP_{Ridge}$ ) ve Kısım 5.4’te önerilen robust ridge tahmin edicilerine dayalı robust  $ICOMP$  kriterlerinin doğru modeli seçim sıklıkları listelenmiştir.  $RICOMP_{R-M}$ ,  $RICOMP_{R-S}$  ve  $RICOMP_{R-MM}$  sırasıyla M, S ve MM tahmin edicilerine ilişkin ağırlıklarla elde edilen robust ridge tahmin edicilerine dayalı  $RICOMP_R$  kriterini göstermektedir.

Çizelge 5.4 Çoklu doğrusal bağlantı durumunda  $RICOMP_R$  kriterlerinin doğru modeli seçme sıklıkları

| $n$ | $ICOMP_{Ridge}$ | $RICOMP_{R-M}$ | $RICOMP_{R-S}$ | $RICOMP_{R-MM}$ |
|-----|-----------------|----------------|----------------|-----------------|
| 20  | 0.57            | 0.78           | 0.80           | 0.87            |
| 50  | 0.59            | 0.81           | 0.84           | 0.89            |
| 100 | 0.61            | 0.83           | 0.88           | 0.91            |

Çizelge 5.5 Çoklu doğrusal bağlantı ve x yönünde aykırı değer durumunda  $RICOMP_R$  kriterlerinin doğru modeli seçme sıklıkları

| $n$ | %  | $ICOMP_{Ridge}$ | $RICOMP_{R-M}$ | $RICOMP_{R-S}$ | $RICOMP_{R-MM}$ |
|-----|----|-----------------|----------------|----------------|-----------------|
| 20  | 10 | 0.10            | 0.17           | 0.85           | 0.89            |
|     | 20 | 0.05            | 0.09           | 0.79           | 0.86            |
|     | 30 | 0.01            | 0.05           | 0.77           | 0.84            |
| 50  | 10 | 0.11            | 0.20           | 0.88           | 0.91            |
|     | 20 | 0.06            | 0.10           | 0.82           | 0.89            |
|     | 30 | 0.02            | 0.06           | 0.80           | 0.87            |
| 100 | 10 | 0.13            | 0.21           | 0.89           | 0.92            |
|     | 20 | 0.08            | 0.14           | 0.84           | 0.90            |
|     | 30 | 0.04            | 0.08           | 0.81           | 0.88            |

Çizelge 5.6 Çoklu doğrusal bağlantı ve y yönünde aykırı değer durumunda  
 $RICOMP_R$  kriterlerinin doğru modeli seçme sıklıkları

| $n$ | %  | $ICOMP_{Ridge}$ | $RICOMP_{R-M}$ | $RICOMP_{R-S}$ | $RICOMP_{R-MM}$ |
|-----|----|-----------------|----------------|----------------|-----------------|
| 20  | 10 | 0.09            | 0.81           | 0.59           | 0.89            |
|     | 20 | 0.04            | 0.75           | 0.51           | 0.87            |
|     | 30 | 0.01            | 0.69           | 0.46           | 0.81            |
| 50  | 10 | 0.10            | 0.83           | 0.62           | 0.91            |
|     | 20 | 0.06            | 0.78           | 0.54           | 0.88            |
|     | 30 | 0.02            | 0.70           | 0.48           | 0.82            |
| 100 | 10 | 0.11            | 0.85           | 0.65           | 0.92            |
|     | 20 | 0.07            | 0.81           | 0.57           | 0.90            |
|     | 30 | 0.04            | 0.73           | 0.51           | 0.85            |

Çizelge 5.7 Çoklu doğrusal bağlantı ve x-y yönünde aykırı değer durumunda  
 $RICOMP_R$  kriterlerinin doğru modeli seçme sıklıkları

| $n$ | %  | $ICOMP_{Ridge}$ | $RICOMP_{R-M}$ | $RICOMP_{R-S}$ | $RICOMP_{R-MM}$ |
|-----|----|-----------------|----------------|----------------|-----------------|
| 20  | 10 | 0.28            | 0.71           | 0.79           | 0.91            |
|     | 20 | 0.25            | 0.68           | 0.76           | 0.84            |
|     | 30 | 0.21            | 0.64           | 0.71           | 0.81            |
| 50  | 10 | 0.32            | 0.73           | 0.80           | 0.92            |
|     | 20 | 0.27            | 0.69           | 0.77           | 0.88            |
|     | 30 | 0.23            | 0.66           | 0.72           | 0.83            |
| 100 | 10 | 0.33            | 0.74           | 0.82           | 0.93            |
|     | 20 | 0.28            | 0.70           | 0.79           | 0.89            |
|     | 30 | 0.24            | 0.67           | 0.75           | 0.86            |

Çizelge 5.4, veri setinde çoklu bağlantı olduğu durumda robust ridge tahmin edicilere dayalı  $RICOMP_R$  kriterlerinin, klasik ridge tahmin edicisine dayalı  $ICOMP$  ( $ICOMP_{Ridge}$ ) kriterine kıyasla doğru modeli daha sık seçme eğiliminde olduklarını göstermektedir.

Çizelge 5.5, veride çoklu doğrusal bağlantı ve x yönünde aykırı değer olması durumunda  $RICOMP_{R-S}$  ve  $RICOMP_{R-MM}$  kriterlerinin doğru modeli seçmede oldukça başarılı olduklarını göstermektedir. Bu sonuçlara göre  $RICOMP_{R-M}$  ve  $ICOMP_{Ridge}$  kriterlerinin x yönündeki aykırı değerlerden oldukça etkilendikleri de görülmektedir.

Veride çoklu doğrusal bağlantı ve y yönünde aykırı değer olması durumunda elde edilen sonuçlar çizelge 5.6'da verilmiştir. Bu sonuçlara göre  $RICOMP_{R-S}$  kriterinin  $ICOMP_{Ridge}$  kriterinden daha iyi performansa sahip olduğu, özellikle  $RICOMP_{R-M}$  ve  $RICOMP_{R-MM}$  kriterlerinin doğru modeli seçmede oldukça başarılı oldukları görülmektedir. M tahmin edicilerinin y yönündeki aykırı değerlere karşı robust olması bu sonucu doğrulamaktadır.

Çizelge 5.7, veride çoklu doğrusal bağlantı ve x-y yönünde aykırı değer olması durumunda  $RICOMP_{R-M}$  kriterinin  $ICOMP_{Ridge}$  kriterinden daha sık doğru modeli seçtiğini, özellikle  $RICOMP_{R-S}$  ve  $RICOMP_{R-MM}$  kriterlerinin ise daha iyi sonuçlar verdiğini göstermektedir.

Çizelge 5.4-5.7'de verilen sonuçlara göre, incelenen tüm durumlar için, örneklem hacminin artması ile model seçim kriterlerinin performansının arttığı görülmektedir.

Sonuç olarak veride çoklu doğrusal bağlantı ve aykırı değer olması durumunda  $RICOMP_{R-MM}$  kriteri, ele alınan diğer kriterlere kıyasla en iyi performansa sahiptir. Dayanıklı tahmin edicilerin ridge tahmin edicileri ile ve  $ICOMP$  bilgi kriteri ile biraraya getirilmesi ile çoklu doğrusal bağlantı ve aykırı değer olduğu durumlarda çoklu doğrusal regresyon modelinde değişken seçimi problemine başarılı bir bakış açısı getirilmiştir.

### 5.4.2 Aylık kazanç verisi

"1996-1998 sonrası komünizm sonrası Doğu-Orta Avrupa'da genç işsizlere yardımcı olacak programların değerlendirilmesi" adlı veri seti üzerinde çalışılmıştır. Veri setine ZA3121 kodu ile <http://zacat.gegis.org/webview/index.jsp> web sitesinden ulaşılabilir. Bu veri kümesinin  $n = 114$  ve  $p = 17$  olan bir altkütmesi alınmıştır. Açıklanan değişken 1996-1998 döneminde geçerli aylık kazançlar (ABD \$) olmak üzere, 16 açıklayıcı değişkenden oluşmaktadır. Açıklayıcı değişkenler şunlardır: yaş ( $x_1$ ), cinsiyet ( $x_2$ ), medeni durum ( $x_3$ ), en yüksek eğitim seviyesi ( $x_4$ ), eğitimi tamamlama yaşı ( $x_5$ ), uzmanlık alanı ( $x_6$ ), çalıştığı kurum sayısı ( $x_7$ ), son 12 ayda evinden ayrı olduğu tatil gün sayısı ( $x_8$ ), ailesine ayırdığı zaman ( $x_9$ ), arkadaşlarına ayırdığı zaman ( $x_{10}$ ), kendine ayırdığı zaman ( $x_{11}$ ), eğitim ve iş eşleşmesi ( $x_{12}$ ), motorlu araç kullanma ( $x_{13}$ ), uydu ve kablolu TV kullanma ( $x_{14}$ ), kişisel bilgisayar kullanımı ( $x_{15}$ ), cep telefonu kullanımı ( $x_{16}$ ).

Açıklanan değişkene ilişkin 8 gözlem (% 7) (-2; 2) aralığının dışında olduğundan bu değerlerin y-yönünde aykırı değer oldukları düşünülebilir.  $x_1$  ve  $x_8$  değişkenlerine ilişkin 6 gözlemin ise x-yönünde aykırı değer olduğu düşünülebilir (bkz Tharmaratnam 2013). Bu nedenle, S ve MM tahmin yöntemlerinde kullanılan ayar sabitleri % 30'luk BDP olacak şekilde seçilmiştir.

Çizelge 5.8'de *ICOMP* ve önerilen robust *ICOMP* kriterleri ile elde edilen değişken seçim sonuçları verilmiştir.

Çizelge 5.8 Aylık kazanç veri için seti regresyon modelinde değişken seçimi sonuçları

| Model Seçim Kriterleri     | Seçilen Altkümeler                                 |
|----------------------------|----------------------------------------------------|
| <i>ICOMP</i>               | : $\{x_2, x_3, x_8, x_{14}, x_{15}, x_{16}\}$      |
| <i>RICOMP<sub>M</sub></i>  | : $\{x_2, x_3, x_4, x_8, x_{10}, x_{15}, x_{16}\}$ |
| <i>RICOMP<sub>S</sub></i>  | : $\{x_3, x_4, x_{10}, x_{11}, x_{12}, x_{15}\}$   |
| <i>RICOMP<sub>MM</sub></i> | : $\{x_3, x_4, x_{10}, x_{11}, x_{12}, x_{15}\}$   |

$RICOMP_S$  ve  $RICOMP_{MM}$  kriterleri  $x_3, x_4, x_{10}, x_{11}, x_{12}, x_{15}$  değişkenlerini seçmiştir.  $RICOMP_M$  kriterinin verideki x yönündeki aykırı değerlerden etkilendiği düşünülebilir. Nitekim aykırı değerlere karşı hassas olan LS tahmin edicilerine dayalı  $ICOMP$  ve x yönündeki aykırı değerlere karşı robust olmayan M tahmin edicilerine dayalı  $RICOMP_M$  kriterleri birbirine yakın modelleri seçmiştir.

### 5.4.3 Diyabet verisi

Bu kısımda dördüncü bölümde de ele alınan diyabet veri seti (Efron vd. 2004) ele alınmıştır. Veri seti  $p = 10$  açıklayıcı değişken (yaş, cinsiyet, vücut kütle indeksi, ortalama kan basıncı ve altı kan serumu ölçümleri) ile  $n = 442$  diyabetli hastayı ve başlangıçtan bir yıl sonra hastalığın ilerleyişinin kantitatif bir ölçütü olan bir açıklanan değişkeni ( $y$ ) kapsar. Diyabet veri seti için kurulan regresyon modelindeki açıklayıcı değişkenler çizelge 4.4'te verilmişti. Bu kısımda veri seti için  $ICOMP_{Ridge}$ ,  $ICOMP_{R-M}$ ,  $ICOMP_{R-S}$ ,  $ICOMP_{R-MM}$  kriterleri kullanılarak değişken seçimi yapılmıştır.

Diyabet veri seti için çizelge 4.4'te verilen açıklayıcı değişkenlere ilişkin VIF değerleri ve en yüksek korelasyon değerleri çizelge 5.9'da verilmiştir (Kirtland 2017).

Çizelge 5.9 Diyabet verilerine ilişkin VIF değerleri ve yüksek korelasyon değerleri (Kirtland 2017)

| Değişken    | VIF   | Yüksek Korelasyonlar                        |
|-------------|-------|---------------------------------------------|
| $x_1$ (yaş) | 1.22  | $x_3$ (0.34)                                |
| $x_2$ (bmi) | 1.51  | $x_8$ (0.45)                                |
| $x_3$ (bp)  | 1.46  | $x_2$ (0.40)                                |
| $x_4$ (s1)  | 59.20 | $x_5$ (0.90)                                |
| $x_5$ (s2)  | 39.19 | $x_4$ (-0.74) , $x_7$ (0.66)                |
| $x_6$ (s3)  | 15.40 | $x_7$ (-0.74)                               |
| $x_7$ (s4)  | 8.89  | $x_6$ (-0.74) , $x_4$ (0.66) , $x_2$ (0.62) |
| $x_8$ (s5)  | 10.08 | $x_7$ (0.62)                                |
| $x_9$ (s6)  | 1.48  | $x_8$ (0.46)                                |

Çizelge 5.9’da verilen VIF değerleri açıklayıcı değişkenlerin neredeyse yarısında çoklu doğrusal bağlantı olduğunu işaret etmektedir. Özellikle  $x_4$ (s1) ve  $x_5$  (s2) değişkenleri arasında güçlü ve eşgüdümlü bir bağlantı vardır. Ayrıca  $x_7$  değişkeninin  $x_5$ ,  $x_6$  ve  $x_8$  değişkenleri ile arasındaki ilişki yüksektir (Kirtland 2017).

Çizelge 5.10’da *ICOMP* ve önerilen robust *ICOMP* kriterleri ile elde edilen değişken seçim sonuçları verilmiştir.

Çizelge 5.10 Diyabet veri seti için regresyon modelinde değişken seçimi sonuçları

| Model Seçim Kriterleri | Seçilen Altkümeler                        |
|------------------------|-------------------------------------------|
| $ICOMP_{Ridge}$        | : $\{x_2, x_3, x_4, x_6, x_7, x_8, x_9\}$ |
| $RICOMP_{R-M}$         | : $\{x_2, x_3, x_4, x_6, x_7, x_9\}$      |
| $RICOMP_{R-S}$         | : $\{x_2, x_3, x_4, x_6, x_9\}$           |
| $RICOMP_{R-MM}$        | : $\{x_2, x_3, x_4, x_7, x_9\}$           |

Çizelge 5.10’da verilen sonuçlara göre  $ICOMP_{Ridge}$  kriteri diğer kriterlere kıyasla daha fazla değişkenden oluşan modeli seçmiştir. Aynı zamanda  $ICOMP_{Ridge}$  kriteri aralarında çoklu doğrusal bağlantı olduğu düşünülen değişkenlerin çoğunu birlikte içeren modeli seçmiştir.  $RICOMP_{R-M}$ ,  $RICOMP_{R-S}$  ve  $RICOMP_{R-MM}$  kriterleri tarafından seçilen modeller, klasik ridge tahminine dayalı  $ICOMP_{Ridge}$  kriterinin seçtiği modelden daha basit olmakla beraber, genel olarak aralarında yüksek ilişkinin olduğu değişkenlerin tamamını içermeyen modellerdir.  $RICOMP_{R-MM}$  kriteri  $x_2, x_3, x_4, x_7, x_9$  değişkenlerinden oluşan modeli seçmiştir. Bu seçim  $x_4$  ve  $x_5$  değişkenleri arasındaki doğrusal bağlantının olduğu ve  $x_7$  ile  $x_5$ ,  $x_6$ ,  $x_8$  değişkenleri arasındaki ilişkinin yüksek olduğu da göz önüne alındığında makul bir seçim olarak düşünülebilir.

## 6. ORTAK KONUM-ÖLÇEK VE KONUM-ÖLÇEK-ÇARPIKLIK MODELLERİNDE *ICOMP* İLE DEĞİŞKEN SEÇİMİ

Simetrik heteroskedastik (değişen varyanslı) verilerin modellenmesinde ortak konum-ölçek modelinin iyi bir araç olduğu bilinmektedir. Heteroskedastik verilerin çarpık olması durumunda çarpıklığın da modellenmesi problemi ortaya çıkmış ve ortak konum-ölçek-çarpıklık modeli literatürdeki yerini almıştır. Literatürde ortak konum-ölçek ve konum-ölçek-çarpıklık modellerinin parametrelerinin tahmini üzerine birçok çalışma yapılmıştır. Ancak bu modellerde değişken seçimi üzerine yapılan çalışmalar olabilirlik fonksiyonunun cezalandırılmasıyla elde edilen yöntemlerle sınırlı kalmış, bilgi tabanlı bir seçim kriteri üzerine henüz bir çalışma literatürde mevcut değildir. Bu nedenle, bu bölümde öncelikle *ICOMP* ile konum ölçek dağılımların ortak konum-ölçek modelinde, daha sonra çarpık simetrik dağılımların ortak konum-ölçek-çarpıklık modelinde değişken seçimi problemi ele alınmıştır. Konum ölçek ailesine örnek olarak normal dağılım ve Student  $t$  dağılımı, çarpık simetrik dağılımlara örnek olarak çarpık normal dağılım ve çarpık  $t$  dağılımı ele alınarak *ICOMP* kriterinin formları elde edilmiştir. Yapılan simülasyon çalışmaları ve gerçek veri örnekleri ile karşılaştırmalar yapılmıştır.

### 6.1 Ortak Konum-Ölçek Modelinde *ICOMP* ile Değişken Seçimi

Bu kesimde ortak konum-ölçek modelinin tanımlanması için öncelikle konum ölçek dağılımlar ailesine yer verilmiştir.

#### 6.1.1 Konum ölçek dağılımlar ailesi

İstatistiksel dağılımlar aynı özellikleri taşıma, aynı ilkelere göre inşa edilme, aynı yapıya sahip olma gibi özellikler altında aile olarak gruplandırılabilirler. Olasılık teorisinde bu ailelerden en yaygın olanı konum ve ölçek aileleridir. Konum ve ölçek aileleri, konum aileleri, ölçek aileleri ve konum, ölçek aileleri olmak üzere üç farklı aile

yapısını içermektedir. Bu aileler, bir  $f$  olasılık yoğunluk fonksiyonunun dönüşümü ile üretilen olasılık yoğunluk fonksiyonlarından oluşur.

**Tanım 6.1**  $g$  olasılık yoğunluk fonksiyonu ve  $\mu \in \mathbb{R}$  olmak üzere

$$f(x; \mu) = g(x - \mu)$$

biçiminde yazılabilen  $f$  olasılık yoğunluk fonksiyonlarının oluşturduğu topluluğa konum ailesi ve  $\mu$  parametresine konum parametresi denir (Shao 2003).

**Tanım 6.2**  $g$  olasılık yoğunluk fonksiyonu ve  $\sigma > 0$  olmak üzere

$$f(x; \sigma) = \frac{1}{\sigma} g\left(\frac{x}{\sigma}\right)$$

biçiminde yazılabilen  $f$  olasılık yoğunluk fonksiyonlarının oluşturduğu topluluğa ölçek ailesi ve  $\sigma$  parametresine ölçek parametresi denir (Shao 2003).

**Teorem 6.1**  $g$  olasılık yoğunluk fonksiyonu  $\mu \in \mathbb{R}$  ve  $\sigma > 0$  olmak üzere

$$f(x; \mu, \sigma) = \frac{1}{\sigma} g\left(\frac{x - \mu}{\sigma}\right)$$

biçiminde yazılabilen  $f$  fonksiyonu olasılık yoğunluk fonksiyonudur (Shao 2003).

**İspat.**  $g$  olasılık yoğunluk fonksiyonu olduğundan her  $x \in \mathbb{R}$  için  $g(x) \geq 0$ 'dır.

Buradan her  $x \in \mathbb{R}$ ,  $\mu \in \mathbb{R}$  ve  $\sigma > 0$  için  $\frac{1}{\sigma} g\left(\frac{x - \mu}{\sigma}\right) \geq 0$  olacağı açıktır. Diğer

tarafından  $y = \frac{x - \mu}{\sigma}$  dönüşümü yapılrısa  $\int_{-\infty}^{\infty} \frac{1}{\sigma} g\left(\frac{x - \mu}{\sigma}\right) dx = \int_{-\infty}^{\infty} g(y) dy = 1$  elde edilir.

■

**Tanım 6.3**  $g$  olasılık yoğunluk fonksiyonu,  $\mu \in \mathbb{R}$  ve  $\sigma > 0$  olmak üzere

$$f(x; \mu, \sigma) = \frac{1}{\sigma} g\left(\frac{x - \mu}{\sigma}\right)$$

biçiminde yazılabilen  $f$  olasılık yoğunluk fonksiyonlarının oluşturduğu topluluğa konum ölçek dağılımlar ailesi,  $\mu$  ve  $\sigma$  parametrelerine sırasıyla konum ve ölçek parametreleri denir (Shao 2003).

Normal dağılım, Student  $t$  dağılımı, Cauchy dağılımı, Laplace dağılımı konum-ölçek ailelerine örnek verilebilir. Ayrıca Weibull ve log-normal gibi bazı dağılımlar uygun dönüşümler altında bu formda yazılabilir.

Konum ölçek dağılımlar ailesinin bazı özellikleri aşağıda verilmiştir:

- (i)  $f(x; 0, 1) = g(x)$
- (ii)  $X \sim f(x; \mu, \sigma)$  ise  $\frac{X-\mu}{\sigma} \sim f(x; 0, 1)$  gerçekleşir.
- (iii)  $X \sim f(x; 0, 1)$  ise  $\sigma X + \mu \sim f(x; \mu, \sigma)$  gerçekleşir (Shao 2003).

**Teorem 6.2**  $Z$  olasılık yoğunluk fonksiyonu  $g$  olan bir rasgele değişken olsun. Eğer  $X \sim \frac{1}{\sigma}g\left(\frac{x-\mu}{\sigma}\right)$  ise  $E[X] = \mu + \sigma E[Z]$  ve  $Var(X) = \sigma^2 Var(Z)$ 'dir (Shao 2003).

**Teorem 6.3**  $X_1, X_2, \dots, X_n$  olasılık yoğunluk fonksiyonu  $\frac{1}{\sigma}f\left(\frac{x-\mu}{\sigma}\right)$  olan dağılımdan bir rasgele örneklem ve  $x \in \mathbb{R}$ ,  $\mu \in \mathbb{R}$  ve  $\sigma > 0$  için  $f(x) > 0$  ve  $f'(x)$  mevcut olsun. Bu durumda  $X_1, X_2, \dots, X_n$  örnekleminin içerdiği  $(\mu, \sigma)$  parametre vektörüne ilişkin Fisher bilgi matrisi

$$\mathcal{F}(\mu, \sigma) = \frac{n}{\sigma^2} \begin{bmatrix} \int \frac{(f'(x))^2}{f(x)} dx & \int \frac{f'(x)(xf'(x)+f(x))}{f(x)} dx \\ \int \frac{f'(x)(xf'(x)+f(x))}{f(x)} dx & \left( \int \frac{(xf'(x)+f(x))^2}{f(x)} dx \right) \end{bmatrix}$$

biçimindedir (Shao 2003).

### 6.1.2 Ortak konum-ölçek modeli

İstatistiksel modellerin en yaygın varsayımı hataların sabit varyanslı (homoskedastik) olmasıdır. Bu varsayımın bozulmasına değişen varyanslılık (heteroskedastik) denir. Regresyonda değişen varyans problemi karşısında literatürde önerilen yöntemlerden biri bağımlı değişkenin Box-Cox dönüşümü ile düzeltilmesidir. Bu da bazen normale yakın yararlı sonuçlar elde edilmesini sağlamakla beraber, her zaman değişen varyansa ilişkin uygun modeli bulmak ve uydurmak zordur. Bu durum karşısında

değişen varyansın (6.3) eşitliğinde olduğu gibi modellenmesi ile değişen varyans probleminde daha güçlü bir regresyon yaklaşımı elde edilmiştir (Aitkin 1987). Park (1966), değişen varyans için log-doğrusal model önermiştir. Harvey (1976), konum ve ölçek parametrelerinin ML tahmin ve olabilirlik oran testlerini araştırmıştır. Aitkin (1987), ortak konum-ölçek modelinin parametrelerinin ML tahmin edicilerini araştırmıştır. Xu ve Zhang (2011), heteroskedastik regresyon modellerinde aynı anda değişken seçim yöntemini (regularized REML) önermiştir. Zhang ve Wang (2011), heteroskedastik doğrusal model yapısına dayalı değişken seçim yöntemi önermiştir. Wu ve Li (2012), ters Gauss dağılımının ortak konum-ölçek modeli için değişken seçimi yapmıştır. Li ve Wu (2014), çarpık normal dağılımın ortak konum-ölçek modelini tanımlamış ve Wu vd. (2013c), çarpık normal dağılımın ortak konum-ölçek modeli için değişken seçimi yöntemi önermiştir. Wu vd. (2013a), çarpık  $t$ -normal dağılımın ortak konum-ölçek modelinin ML tahmin edicilerini elde etmiş ve Wu (2014), çarpık  $t$ -normal dağılımın ortak konum-ölçek modelinde parametre tahminini ve değişken seçimi için bir yöntem önermiştir.

Bu kısımda konum ölçek dağılım varsayımı altında ortak konum-ölçek modelinde *ICOMP* ile değişken seçimi yapılmıştır. Konum ölçek dağılım ailelerine örnek olarak normal dağılım ve Student  $t$  dağılımı alınarak, bu dağılımların ortak konum-ölçek modelleri için *ICOMP* kriterinin formları elde edilmiştir.

Ortak konum-ölçek modeli

$$y \sim f(\mu, \sigma) \quad (6.1)$$

$$\mu = \mathbf{x}^T \boldsymbol{\beta} \quad (6.2)$$

$$\log \sigma^2 = \mathbf{z}^T \boldsymbol{\gamma} \quad (6.3)$$

şeklinde tanımlanmaktadır. Burada (6.2) eşitliği konum modelini, (6.3) eşitliği ölçek modelini,  $\mathbf{x} = [x_1, x_2, \dots, x_p]^T$  ve  $\mathbf{z} = [z_1, z_2, \dots, z_q]^T$  değişken vektörlerini,  $\boldsymbol{\beta} = [\beta_1, \beta_2, \dots, \beta_p]^T$  konum modelinin bilinmeyen parametre vektörünü ve  $\boldsymbol{\gamma} = [\gamma_1, \gamma_2, \dots, \gamma_q]^T$  ölçek modelinin bilinmeyen parametre vektörünü göstermektedir. (6.3) eşitliğinde verilen log-doğrusal form varyansın pozitifliğini sağlamaktadır.

### 6.1.3 Konum ölçek dağılımının ortak konum-ölçek modelinde değişken seçimi

Konum ölçek dağılımı için ortak konum-ölçek modeli ele alındığında, model

$$y \sim \frac{1}{\sigma} f\left(\frac{y - \mu}{\sigma}\right) \quad (6.4)$$

$$\mu = \mathbf{x}^T \boldsymbol{\beta} \quad (6.5)$$

$$\log \sigma^2 = \mathbf{z}^T \boldsymbol{\gamma} \quad (6.6)$$

biçiminde yazılır.

$(y_i, \mathbf{x}_i, \mathbf{z}_i)$  (6.4)-(6.6) ile verilen modelden  $n$  birimlik rasgele örneklem olsun.

Olabilirlik ve log-olabilirlik fonksiyonları sırasıyla

$$\begin{aligned} L(\boldsymbol{\beta}, \boldsymbol{\gamma} | \mathbf{y}, \mathbf{X}, \mathbf{Z}) &= \prod_{i=1}^n \frac{1}{\sigma_i} f\left(\frac{y_i - \mu_i}{\sigma_i}\right) \\ &= \left(\prod_{i=1}^n e^{\mathbf{z}_i^T \boldsymbol{\gamma}}\right)^{-1/2} \prod_{i=1}^n f\left(\frac{y_i - \mathbf{x}_i^T \boldsymbol{\beta}}{e^{\frac{1}{2} \mathbf{z}_i^T \boldsymbol{\gamma}}}\right) \\ \ell(\boldsymbol{\beta}, \boldsymbol{\gamma} | \mathbf{y}, \mathbf{X}, \mathbf{Z}) &= -\frac{1}{2} \sum_{i=1}^n \mathbf{z}_i^T \boldsymbol{\gamma} + \sum_{i=1}^n \log f\left(\frac{y_i - \mathbf{x}_i^T \boldsymbol{\beta}}{e^{\frac{1}{2} \mathbf{z}_i^T \boldsymbol{\gamma}}}\right) \end{aligned} \quad (6.7)$$

olarak elde edilir.

$\boldsymbol{\beta}$  ve  $\boldsymbol{\gamma}$  parametrelerinin ML tahminleri

$$\frac{\partial \ell}{\partial \boldsymbol{\beta}} = - \sum_{i=1}^n \frac{1}{e^{\frac{1}{2} \mathbf{z}_i^T \boldsymbol{\gamma}}} \frac{f'\left(\frac{y_i - \mathbf{x}_i^T \boldsymbol{\beta}}{e^{\frac{1}{2} \mathbf{z}_i^T \boldsymbol{\gamma}}}\right)}{f\left(\frac{y_i - \mathbf{x}_i^T \boldsymbol{\beta}}{e^{\frac{1}{2} \mathbf{z}_i^T \boldsymbol{\gamma}}}\right)} \mathbf{x}_i = 0 \quad (6.8)$$

$$\frac{\partial \ell}{\partial \boldsymbol{\gamma}} = -\frac{1}{2} \sum_{i=1}^n \left[ \frac{f'\left(\frac{y_i - \mathbf{x}_i^T \boldsymbol{\beta}}{e^{\frac{1}{2} \mathbf{z}_i^T \boldsymbol{\gamma}}}\right)}{f\left(\frac{y_i - \mathbf{x}_i^T \boldsymbol{\beta}}{e^{\frac{1}{2} \mathbf{z}_i^T \boldsymbol{\gamma}}}\right)} \left(\frac{y_i - \mathbf{x}_i^T \boldsymbol{\beta}}{e^{\frac{1}{2} \mathbf{z}_i^T \boldsymbol{\gamma}}}\right) + 1 \right] \mathbf{z}_i = 0 \quad (6.9)$$

denklem sisteminin çözülmesiyle elde edilir.

Diğer taraftan Fisher bilgi matrisi eşitlik (6.7)'nin kısmi türevleri kullanılarak

$$\begin{aligned}\mathcal{F}(\boldsymbol{\beta}, \boldsymbol{\gamma}) &= -E \begin{bmatrix} \frac{\partial^2 \ell}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} & \frac{\partial^2 \ell}{\partial \boldsymbol{\beta} \partial \boldsymbol{\gamma}^T} \\ \frac{\partial^2 \ell}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\beta}^T} & \frac{\partial^2 \ell}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\gamma}^T} \end{bmatrix} \\ &= \begin{bmatrix} \left( \frac{1}{\mathbf{e}^{\mathbf{z}^T \boldsymbol{\gamma}/2}} \int \frac{(f')^2}{f} dy \right) xx^T & \left( \frac{1}{2\mathbf{e}^{\mathbf{z}^T \boldsymbol{\gamma}/2}} \int \frac{f'(f'+f)}{f} dy \right) xz^T \\ \left( \frac{1}{2\mathbf{e}^{\mathbf{z}^T \boldsymbol{\gamma}/2}} \int \frac{f'(f'+f)}{f} dy \right) zx^T & \left( \frac{1}{4} \int \frac{(f'+f)^2}{f} dy \right) zz^T \end{bmatrix}\end{aligned}$$

eşitliği ile elde edilir. Buradan parametrelerin ML tahmin edicilerinin asimptotik kovaryans matrisi

$$\mathcal{F}^{-1} = \frac{1}{|\mathcal{F}|} \begin{bmatrix} \left( \frac{1}{4} \int \frac{(f'+f)^2}{f} dy \right) zz^T & - \left( \frac{1}{\mathbf{e}^{\mathbf{z}^T \boldsymbol{\gamma}/2}} \int \frac{f'(f'+f)}{f} dy \right) xz^T \\ - \left( \frac{1}{2\mathbf{e}^{\mathbf{z}^T \boldsymbol{\gamma}/2}} \int \frac{f'(f'+f)}{f} dy \right) zx^T & \left( \frac{1}{\mathbf{e}^{\mathbf{z}^T \boldsymbol{\gamma}/2}} \int \frac{(f')^2}{f} dy \right) xx^T \end{bmatrix} \quad (6.10)$$

biçiminde bulunur.

$\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}})$ ,  $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\gamma})$  parametre vektörünün ML tahmin edicisini göstermek üzere konum ölçek dağılımının ortak konum-ölçek modeli için *ICOMP* bilgi kriteri orijinal *ICOMP(IFIM)* formu ile yani

$$ICOMP = -2 \log L(\hat{\boldsymbol{\theta}}) + 2C_{1N}(\hat{\mathcal{F}}^{-1})$$

eşitliği ile tanımlanır.  $\mathcal{F}$  fisher bilgi matrisi (6.10) eşitliğinde verilmiştir. Kritere göre, en küçük *ICOMP* değerini veren model en iyi modeldir.

Konum-ölçek dağılımlara örnek olarak normal dağılımın ve Student *t* dağılımının ele alındığı, ortak konum-ölçek modelinde *ICOMP* bilgi kriteri ile değişken seçimine Kısım 6.1.4 ve 6.1.5'de yer verilmiştir. Bu kısımlarda normal dağılım ve Student *t* dağılımının ortak konum-ölçek modelleri için *ICOMP* kriterinin formları elde edilmiştir.

#### 6.1.4 Normal dağılımın ortak konum-ölçek modelinde değişken seçimi

Bu kısımda konum-ölçek ailesine örnek olmak üzere normal dağılım ele alınmıştır ve normal dağılımın ortak konum-ölçek modeli için *ICOMP* bilgi kriterinin formu elde edilmiştir.

Bir konum ölçek dağılımı olan normal dağılıma ilişkin konum-ölçek modeli

$$y_i \sim N(\mu_i, \sigma_i^2) \quad (6.11)$$

$$\mu_i = \mathbf{x}_i^T \boldsymbol{\beta} \quad (6.12)$$

$$\log \sigma_i^2 = \mathbf{z}_i^T \boldsymbol{\gamma} \quad (6.13)$$

$$i = 1, 2, \dots, n$$

biçiminde yazılır. Burada daha önce tanımlandığı gibi  $\mathbf{x}_i = [x_{i1}, x_{i2}, \dots, x_{ip}]^T$  ve  $\mathbf{z}_i = [z_{i1}, z_{i2}, \dots, z_{iq}]^T$  gözlem vektörlerini,  $\boldsymbol{\beta} = [\beta_1, \beta_2, \dots, \beta_p]^T$  ve  $\boldsymbol{\gamma} = [\gamma_1, \gamma_2, \dots, \gamma_q]^T$  sırasıyla konum ve ölçek modelinin bilinmeyen parametre vektörlerini göstermektedir.

Normal dağılıma ilişkin konum-ölçek modelinde de diğer modellerde olduğu gibi *ICOMP* bilgi kriteri ile değişken seçimi yapabilmek için öncelikle parametrelerin ML tahmin edicileri elde edilmelidir:

$(y_i, \mathbf{x}_i, \mathbf{z}_i)$  normal dağılımın ortak konum-ölçek modelinden bir rasgele örneklem olmak üzere olabirlik ve log-olabirlik fonksiyonları sırasıyla

$$\begin{aligned} L(\boldsymbol{\beta}, \boldsymbol{\gamma} | \mathbf{y}, \mathbf{X}, \mathbf{Z}) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{1}{2} \frac{(y_i - \mu_i)^2}{\sigma_i^2}\right) \\ &= \left(\frac{1}{\sqrt{2\pi}}\right)^n \left(\prod_{i=1}^n \frac{1}{\sigma_i^2}\right)^{1/2} \exp\left(-\frac{1}{2} \sum_{i=1}^n \frac{(y_i - \mathbf{x}_i^T \boldsymbol{\beta})^2}{\mathbf{e}^{\mathbf{z}_i^T \boldsymbol{\gamma}}}\right) \\ \ell(\boldsymbol{\beta}, \boldsymbol{\gamma} | \mathbf{y}, \mathbf{X}, \mathbf{Z}) &= -n \log \sqrt{2\pi} - \frac{1}{2} \sum_{i=1}^n \mathbf{z}_i^T \boldsymbol{\gamma} - \frac{1}{2} \sum_{i=1}^n \frac{(y_i - \mathbf{x}_i^T \boldsymbol{\beta})^2}{\mathbf{e}^{\mathbf{z}_i^T \boldsymbol{\gamma}}} \end{aligned} \quad (6.14)$$

şeklinde yazılabilir. Parametrelerin ML tahmin edicileri  $\arg \max_{\boldsymbol{\beta}, \boldsymbol{\gamma}} \ell(\boldsymbol{\beta}, \boldsymbol{\gamma} | \mathbf{y}, \mathbf{X}, \mathbf{Z})$  maksimizasyon probleminin çözümü ile elde edilir. Bu maksimizasyon probleminin

çözümüne (6.14)'de verilen log-olabilirlik fonksiyonunun parametrelere göre türevlerinin alınıp sıfıra eşitlenmesi ile ulaşılır:

$$\frac{\partial \ell}{\partial \boldsymbol{\beta}} = \sum_{i=1}^n \frac{\mathbf{x}_i (y_i - \mathbf{x}_i^T \boldsymbol{\beta})}{\mathbf{e}^{z_i^T \boldsymbol{\gamma}}} = \mathbf{0} \quad (6.15)$$

$$\frac{\partial \ell}{\partial \boldsymbol{\gamma}} = \frac{1}{2} \sum_{i=1}^n z_i \left( \frac{(y_i - \mathbf{x}_i^T \boldsymbol{\beta})^2}{\mathbf{e}^{z_i^T \boldsymbol{\gamma}}} - 1 \right) = \mathbf{0} \quad (6.16)$$

(6.15) eşitliğinin çözülmesiyle  $\boldsymbol{\beta}$  konum parametre vektörünün ML tahmin edicisine ilişkin

$$\hat{\boldsymbol{\beta}} = \left( \sum_{i=1}^n \frac{\mathbf{x}_i \mathbf{x}_i^T}{\mathbf{e}^{z_i^T \hat{\boldsymbol{\gamma}}}} \right)^{-1} \left( \sum_{i=1}^n \frac{y_i \mathbf{x}_i}{\mathbf{e}^{z_i^T \hat{\boldsymbol{\gamma}}}} \right) \quad (6.17)$$

eşitliği elde edilir.

Gösterim kolaylığı için  $\epsilon_i = y_i - \mathbf{x}_i^T \boldsymbol{\beta}$  olmak üzere

$$\mathbf{R}_{11} = \text{diag} \left\{ \frac{1}{\sigma_i^2} \right\}_{i=1}^n \quad (6.18)$$

$$\mathbf{R}_{12} = \text{diag} \left\{ \frac{\epsilon_i}{\sigma_i^2} \right\}_{i=1}^n \quad (6.19)$$

$$\mathbf{R}_{22} = \text{diag} \left\{ \frac{1}{2} \frac{\epsilon_i^2}{\sigma_i^2} \right\}_{i=1}^n \quad (6.20)$$

köşegen matrisleri tanımlansın. Bu durumda

$$\hat{\boldsymbol{\beta}} = \left( \mathbf{X}^T \hat{\mathbf{R}}_{11} \mathbf{X} \right)^{-1} \left( \mathbf{X}^T \hat{\mathbf{R}}_{11} \mathbf{y} \right) \quad (6.21)$$

yazılır.

$\boldsymbol{\gamma}$  ölçek parametre vektörünün ML tahmini

$$\sum_{i=1}^n z_i \left( \frac{(y_i - \mathbf{x}_i^T \hat{\boldsymbol{\beta}})^2}{\mathbf{e}^{z_i^T \boldsymbol{\gamma}}} - 1 \right) = \mathbf{0} \quad (6.22)$$

denkleminin sayısal yöntemlerle çözülmesi ile elde edilebilir.

Log-olabilirlik fonksiyonunun parametre vektörlerine göre kısmi türevleri aşağıdaki gibi elde edilir:

$$\frac{\partial^2 \ell}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} = -\mathbf{X}^T \mathbf{R}_{11} \mathbf{X}$$

$$\begin{aligned}\frac{\partial^2 \ell}{\partial \boldsymbol{\beta} \partial \boldsymbol{\gamma}^T} &= -\mathbf{X}^T \mathbf{R}_{12} \mathbf{Z} \\ \frac{\partial^2 \ell}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\beta}^T} &= -\mathbf{Z}^T \mathbf{R}_{12} \mathbf{X} \\ \frac{\partial^2 \ell}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\gamma}^T} &= -\mathbf{Z}^T \mathbf{R}_{22} \mathbf{Z}\end{aligned}$$

Buradan Fisher bilgi matrisi

$$\mathcal{F}(\boldsymbol{\beta}, \boldsymbol{\gamma}) = -E \left( \begin{bmatrix} \frac{\partial^2 \ell}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} & \frac{\partial^2 \ell}{\partial \boldsymbol{\beta} \partial \boldsymbol{\gamma}^T} \\ \frac{\partial^2 \ell}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\beta}^T} & \frac{\partial^2 \ell}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\gamma}^T} \end{bmatrix} \right) = E \left( \begin{bmatrix} \mathbf{X}^T \mathbf{R}_{11} \mathbf{X} & \mathbf{X}^T \mathbf{R}_{12} \mathbf{Z} \\ \mathbf{Z}^T \mathbf{R}_{12} \mathbf{X} & \mathbf{Z}^T \mathbf{R}_{22} \mathbf{Z} \end{bmatrix} \right) \quad (6.23)$$

biçiminde elde edilir.

$E(\epsilon_i^2) = \sigma_i^2$  olduğu göz önüne alınırsa (6.19) ve (6.20) eşitliklerinden  $E(\mathbf{R}_{12}) = \mathbf{0}$  ve  $E(\mathbf{R}_{22}) = \frac{1}{2} \text{diag}\{1\}$  olacağından Fisher bilgi matrisinin bileşenleri

$$E(\mathbf{X}^T \mathbf{R}_{11} \mathbf{X}) = \mathbf{X}^T \mathbf{R}_{11} \mathbf{X} \quad (6.24)$$

$$E(\mathbf{X}^T \mathbf{R}_{12} \mathbf{Z}) = \mathbf{0} \quad (6.25)$$

$$E(\mathbf{Z}^T \mathbf{R}_{12} \mathbf{X}) = \mathbf{0} \quad (6.26)$$

$$E(\mathbf{Z}^T \mathbf{R}_{22} \mathbf{Z}) = \frac{1}{2} \mathbf{Z}^T \mathbf{Z} \quad (6.27)$$

bulunur. (6.24)-(6.27) eşitliklerinin (6.23) ifadesinde dikkate alınmasıyla normal dağılımın ortak konum-ölçek modelinin parametreleri için Fisher bilgi matrisi

$$\mathcal{F}(\boldsymbol{\beta}, \boldsymbol{\gamma}) = \begin{bmatrix} \mathbf{X}^T \mathbf{R}_{11} \mathbf{X} & \mathbf{0} \\ \mathbf{0}^T & \frac{1}{2} \mathbf{Z}^T \mathbf{Z} \end{bmatrix} \quad (6.28)$$

şeklinde elde edilir. Burada  $\mathbf{0}$ ,  $p \times q$  boyutlu sıfır matrisidir. Buna göre parametrelerin ML tahmin edicilerinin kovaryans matrisi

$$\mathcal{F}^{-1} = \begin{bmatrix} (\mathbf{X}^T \mathbf{R}_{11} \mathbf{X})^{-1} & \mathbf{0} \\ \mathbf{0}^T & 2 (\mathbf{Z}^T \mathbf{Z})^{-1} \end{bmatrix}$$

ve kovaryans matrisi tahmini,

$$\widehat{\mathcal{F}}^{-1} = \begin{bmatrix} (\mathbf{X}^T \widehat{\mathbf{R}}_{11} \mathbf{X})^{-1} & \mathbf{0} \\ \mathbf{0}^T & 2 (\mathbf{Z}^T \mathbf{Z})^{-1} \end{bmatrix}$$

biçimindedir. Bazı matrissel işlemlerden sonra

$$\begin{aligned} \text{tr}(\hat{\mathcal{F}}^{-1}) &= \text{tr}\left(\left(\mathbf{X}^T \hat{\mathbf{R}}_{11} \mathbf{X}\right)^{-1}\right) + 2\text{tr}\left((\mathbf{Z}^T \mathbf{Z})^{-1}\right) \\ |\hat{\mathcal{F}}^{-1}| &= \left|\left(\mathbf{X}^T \hat{\mathbf{R}}_{11} \mathbf{X}\right)^{-1}\right| + 2^q \left|(\mathbf{Z}^T \mathbf{Z})^{-1}\right| \end{aligned}$$

olduğundan  $s = p + q$  olmak üzere,  $\hat{\mathcal{F}}^{-1}$  için maksimal karmaşıklık bilgisi

$$\begin{aligned} C_{1N}(\hat{\mathcal{F}}^{-1}) &= \frac{s}{2} \log \left[ \frac{\text{tr} \hat{\mathcal{F}}^{-1}}{s} \right] - \frac{1}{2} \log |\hat{\mathcal{F}}^{-1}| \\ &= \frac{s}{2} \log \left( \frac{\text{tr} \left( \left( \mathbf{X}^T \hat{\mathbf{R}}_{11} \mathbf{X} \right)^{-1} \right) + 2\text{tr} \left( (\mathbf{Z}^T \mathbf{Z})^{-1} \right)}{s} \right) \\ &\quad - \frac{1}{2} \log \left( \left| \left( \mathbf{X}^T \hat{\mathbf{R}}_{11} \mathbf{X} \right)^{-1} \right| + 2^q \left| (\mathbf{Z}^T \mathbf{Z})^{-1} \right| \right) \end{aligned} \quad (6.29)$$

elde edilir.

$\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\gamma})$  parametre vektörünün ML tahmin edicisi  $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}})$ , parametrelerin ML tahmin edicisinin tahmini kovaryans matrisi  $\hat{\mathcal{F}}^{-1}$  ve  $s = p + q$  olmak üzere normal dağılımın ortak konum-ölçek modeli için *ICOMP*

$$\begin{aligned} \text{ICOMP}_{JLSM-N} &= -2 \log L(\hat{\boldsymbol{\theta}}) + 2 \left( \frac{s}{2} \log \left[ \frac{\text{tr} \hat{\mathcal{F}}^{-1}}{s} \right] - \frac{1}{2} \log |\hat{\mathcal{F}}^{-1}| \right) \\ &= -2 \log L(\hat{\boldsymbol{\theta}}) \\ &\quad + s \log \left( \frac{\text{tr} \left( \left( \mathbf{X}^T \hat{\mathbf{R}}_{11} \mathbf{X} \right)^{-1} \right) + 2\text{tr} \left( (\mathbf{Z}^T \mathbf{Z})^{-1} \right)}{s} \right) \\ &\quad - \log \left( \left| \left( \mathbf{X}^T \hat{\mathbf{R}}_{11} \mathbf{X} \right)^{-1} \right| + 2^q \left| (\mathbf{Z}^T \mathbf{Z})^{-1} \right| \right) \end{aligned} \quad (6.30)$$

formunda elde edilir. Kritere göre en iyi model, en küçük  $\text{ICOMP}_{JLSM-N}$  değerini veren modeldir.

### 6.1.5 Student $t$ dağılımının ortak konum-ölçek modelinde değişken seçimi

Bu kısımda Student  $t$  dağılımı ele alınmış ve Student  $t$  dağılımının ortak konum-ölçek modeli için *ICOMP* bilgi kriterinin formu elde edilmiştir.

Bir konum ölçek dağılımı olan Student  $t$  dağılımına ilişkin konum-ölçek modeli

$$y_i \sim t(\mu_i, \sigma_i^2, \nu) \quad (6.31)$$

$$\mu_i = \mathbf{x}_i^T \boldsymbol{\beta} \quad (6.32)$$

$$\log \sigma_i^2 = \mathbf{z}_i^T \boldsymbol{\gamma} \quad (6.33)$$

$$i = 1, 2, \dots, n$$

biçiminde yazılır. Burada  $\mathbf{x}_i = [x_{i1}, x_{i2}, \dots, x_{ip}]^T$  ve  $\mathbf{z}_i = [z_{i1}, z_{i2}, \dots, z_{iq}]^T$  gözlem vektörlerini,  $\boldsymbol{\beta} = [\beta_1, \beta_2, \dots, \beta_p]^T$  ve  $\boldsymbol{\gamma} = [\gamma_1, \gamma_2, \dots, \gamma_q]^T$  sırasıyla konum ve ölçek modelinin bilinmeyen parametre vektörlerini göstermektedir.

Student  $t$  dağılımının ortak konum-ölçek modelinde *ICOMP* bilgi kriteri ile değişken seçimi yapabilmek için öncelikle parametrelerin ML tahmin edicileri elde edilmelidir:

$(y_i, \mathbf{x}_i, \mathbf{z}_i)$  Student  $t$  dağılımının ortak konum-ölçek modelinden bir rasgele örneklem olmak üzere olabilirlik ve log-olabilirlik fonksiyonları sırasıyla

$$\begin{aligned} L(\boldsymbol{\beta}, \boldsymbol{\gamma} | \mathbf{y}, \mathbf{X}, \mathbf{Z}) &= \prod_{i=1}^n \frac{c_\nu}{\sigma_i} \left( \nu + \frac{(y_i - \mu_i)^2}{\sigma_i^2} \right)^{-\frac{\nu+1}{2}} \\ &= (c_\nu)^n \left( \prod_{i=1}^n \frac{1}{\sigma_i^2} \right)^{1/2} \left( \prod_{i=1}^n \left( \nu + \frac{(y_i - \mu_i)^2}{\sigma_i^2} \right) \right)^{-\frac{\nu+1}{2}} \\ \ell(\boldsymbol{\beta}, \boldsymbol{\gamma} | \mathbf{y}, \mathbf{X}, \mathbf{Z}) &= n \log(c_\nu) - \frac{1}{2} \sum_{i=1}^n \log \sigma_i^2 - \frac{\nu+1}{2} \sum_{i=1}^n \log \left( \nu + \frac{(y_i - \mu_i)^2}{\sigma_i^2} \right) \\ &= n \log(c_\nu) - \frac{1}{2} \sum_{i=1}^n \mathbf{z}_i^T \boldsymbol{\gamma} \\ &\quad - \frac{\nu+1}{2} \sum_{i=1}^n \log \left( \nu + \frac{(y_i - \mathbf{x}_i^T \boldsymbol{\beta})^2}{\mathbf{e}^{\mathbf{z}_i^T \boldsymbol{\gamma}}} \right) \end{aligned} \quad (6.34)$$

şeklinde yazılabilir. Burada  $c_\nu = \frac{\Gamma(\frac{\nu+1}{2}) \nu^{\nu/2}}{\sqrt{\pi} \Gamma(\frac{\nu}{2})}$  sabittir.

Parametrelerin ML tahmin edicileri eşitlik (6.34)'de verilen  $\arg \max_{\beta, \gamma} \ell(\beta, \gamma | \mathbf{y}, \mathbf{X}, \mathbf{Z})$  maksimizasyon probleminin çözümü ile elde edilir. Bu maksimizasyon probleminin çözümüne log-olabilirlik fonksiyonunun parametrelere göre türevlerinin alınıp sıfıra eşitlenmesi ile ulaşılır:

$$\frac{\partial \ell}{\partial \beta} = \sum_{i=1}^n \omega_i \frac{\mathbf{x}_i (y_i - \mathbf{x}_i^T \beta)}{\mathbf{e}^{\mathbf{z}_i^T \gamma}} = \mathbf{0} \quad (6.35)$$

$$\frac{\partial \ell}{\partial \gamma} = \frac{1}{2} \sum_{i=1}^n \left( \omega_i \frac{(y_i - \mathbf{x}_i^T \beta)^2}{\mathbf{e}^{\mathbf{z}_i^T \gamma}} - 1 \right) \mathbf{z}_i = \mathbf{0} \quad (6.36)$$

Burada  $\omega_i$  ağırlıkları  $\omega_i = \frac{\nu+1}{\nu + \frac{(y_i - \mathbf{x}_i^T \beta)^2}{\mathbf{e}^{\mathbf{z}_i^T \gamma}}}$  eşitliği ile tanımlıdır. (6.35) eşitliğinin çözülmesiyle  $\beta$  konum parametre vektörünün ML tahmin edicisine ilişkin

$$\hat{\beta} = \left( \sum_{i=1}^n \frac{\omega_i \mathbf{x}_i \mathbf{x}_i^T}{\mathbf{e}^{\mathbf{z}_i^T \gamma}} \right)^{-1} \left( \sum_{i=1}^n \frac{\omega_i y_i \mathbf{x}_i}{\mathbf{e}^{\mathbf{z}_i^T \gamma}} \right) \quad (6.37)$$

eşitliği elde edilir. Gösterim kolaylığı için  $\epsilon_i = y_i - \mathbf{x}_i^T \beta$  olmak üzere

$$\mathbf{R}_{11}^* = \text{diag} \left\{ \frac{\omega_i}{\sigma_i^2} \right\}_{i=1}^n \quad (6.38)$$

$$\mathbf{R}_{12}^* = \text{diag} \left\{ \frac{\omega_i \epsilon_i}{\sigma_i^2} \right\}_{i=1}^n \quad (6.39)$$

$$\mathbf{R}_{22}^* = \text{diag} \left\{ \frac{1}{2} \frac{\omega_i \epsilon_i^2}{\sigma_i^2} \right\}_{i=1}^n \quad (6.40)$$

$$\mathbf{\Omega} = \text{diag} \{ \omega_i \}_{i=1}^n \quad (6.41)$$

köşegen matrisleri tanımlansın. Bu durumda

$$\hat{\beta} = \left( \mathbf{X}^T \hat{\mathbf{R}}_{11}^* \mathbf{X} \right)^{-1} \left( \mathbf{X}^T \hat{\mathbf{R}}_{11}^* \mathbf{y} \right) \quad (6.42)$$

yazılır.  $\gamma$  ölçek parametre vektörünün ML tahmin edicisi  $\mathbf{1}_{n \times n} = \text{diag} \{1\}_{i=1}^n$  olmak üzere

$$\mathbf{Z}^T (\mathbf{R}_{22}^* - \mathbf{1}_{n \times n}) = \mathbf{0} \quad (6.43)$$

denkleminin sayısal yöntemlerle çözülmesi ile elde edilebilir. Log-olabilirlik fonksiyonunun parametre vektörlerine göre kısmi türevleri aşağıdaki gibi elde edilir:

$$\frac{\partial^2 \ell}{\partial \beta \partial \beta^T} = \mathbf{X}^T \left( \frac{2}{(\nu+1)} (\mathbf{R}_{12}^*)^2 - \mathbf{R}_{11}^* \right) \mathbf{X}$$

$$\frac{\partial^2 \ell}{\partial \beta \partial \gamma^T} = -\mathbf{X}^T \mathbf{R}_{12}^* \left( 1 - \frac{1}{\nu+1} \mathbf{R}_{22}^* \right) \mathbf{Z}$$

$$\begin{aligned}\frac{\partial^2 \ell}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\beta}^T} &= -\mathbf{Z}^T \mathbf{R}_{12}^* \left(1 - \frac{1}{\nu+1} \mathbf{R}_{22}^*\right) \mathbf{X} \\ \frac{\partial^2 \ell}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\gamma}^T} &= -\mathbf{Z}^T \mathbf{R}_{22}^* \left(1 - \frac{2}{\nu+1} \mathbf{R}_{22}^*\right) \mathbf{Z}\end{aligned}$$

Buradan Fisher bilgi matrisi

$$\begin{aligned}\mathcal{F}(\boldsymbol{\beta}, \boldsymbol{\gamma}) &= -E \left( \begin{bmatrix} \frac{\partial \ell}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} & \frac{\partial^2 \ell}{\partial \boldsymbol{\beta} \partial \boldsymbol{\gamma}^T} \\ \frac{\partial^2 \ell}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\beta}^T} & \frac{\partial^2 \ell}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\gamma}^T} \end{bmatrix} \right) \\ &= E \left( \begin{bmatrix} \mathbf{X}^T \left( \frac{2}{(\nu+1)} (\mathbf{R}_{12}^*)^2 - \mathbf{R}_{11}^* \right) \mathbf{X} & \mathbf{X}^T \mathbf{R}_{12}^* \left(1 - \frac{1}{\nu+1} \mathbf{R}_{22}^*\right) \mathbf{Z} \\ \mathbf{Z}^T \mathbf{R}_{12}^* \left(1 - \frac{1}{\nu+1} \mathbf{R}_{22}^*\right) \mathbf{X} & \mathbf{Z}^T \mathbf{R}_{22}^* \left(1 - \frac{2}{\nu+1} \mathbf{R}_{22}^*\right) \mathbf{Z} \end{bmatrix} \right)\end{aligned}$$

eşitliği ile elde edilir.

Fisher bilgi matrisinin elemanlarının açık formlarının elde edilmesi için ihtiyaç duyulan bazı hesaplamalar Lemma 6.1’de verilmiştir.

**Lemma 6.1**  $y \sim t(\mathbf{x}_i^T \boldsymbol{\beta}, \sigma^2, \nu)$  olmak üzere  $z = \frac{y - \mathbf{x}_i^T \boldsymbol{\beta}}{\sigma}$ ,  $a_k = E[(\nu + z^2)^{-k}]$ ,  $b_q = E[(y - \mathbf{x}_i^T \boldsymbol{\beta})^q (\nu + z^2)^{-1}]$  ve  $c_q = E[(y - \mathbf{x}_i^T \boldsymbol{\beta})^q (\nu + z^2)^{-2}]$  olarak tanımlansın Bu durumda aşağıdaki eşitlikler sağlanır.

$$(i) \quad \nu > 1 \text{ için } E(z) = 0, \nu > 2 \text{ için } E(z^2) = \frac{\nu}{\nu-2}, \nu > 3 \text{ için } E(z^3) = 0, \nu > 2 \text{ için } E(z^4) = \frac{3\nu^2}{(\nu-2)(\nu-4)}$$

$$(ii) \quad a_k = E[(\nu + z^2)^{-k}] = \left(\frac{\nu+2k}{\nu}\right)^{(\nu+2k+1)/2} \frac{c(\nu, \sigma^2)}{c(\nu+2k, \frac{\nu\sigma^2}{\nu+2k})}. \quad \text{Özel olarak } a_1 = \frac{1}{\nu+1}, \\ a_2 = \frac{\nu+2}{\nu(\nu+1)(\nu+3)} \text{’dir.}$$

$$(iii) \quad b_q = E[(y - \mathbf{x}_i^T \boldsymbol{\beta})^q (\nu + z^2)^{-1}] = a_1 E[(y - \mathbf{x}_i^T \boldsymbol{\beta})^q]$$

$$(iv) \quad c_q = E[(y - \mathbf{x}_i^T \boldsymbol{\beta})^q (\nu + z^2)^{-2}] = a_2 E[(y - \mathbf{x}_i^T \boldsymbol{\beta})^q].$$

Lemma 6.1 (ii) ve (iii) şıklarından

$$E\left(\frac{2}{(\nu+1)} (\mathbf{R}_{12}^*)^2 - \mathbf{R}_{11}^*\right) = \frac{\nu+1}{\nu+3} \mathbf{R}_{11}$$

elde edilir. Buradan  $E\left(\mathbf{X}^T \left(\frac{2}{(\nu+1)} (\mathbf{R}_{12}^*)^2 - \mathbf{R}_{11}^*\right) \mathbf{X}\right) = \frac{\nu+1}{\nu+3} \mathbf{X}^T \mathbf{R}_{11} \mathbf{X}$  bulunur.

Lemma 6.1 (iii) ve (iv)den  $E \left( \mathbf{R}_{12}^* \left( 1 - \frac{1}{\nu+1} \mathbf{R}_{22}^* \right) \right) = \mathbf{0}$  olduğundan

$$E \left( \mathbf{X}^T \mathbf{R}_{12}^* \left( 1 - \frac{1}{\nu+1} \mathbf{R}_{22}^* \right) \mathbf{Z} \right) = \mathbf{0}$$

$$E \left( \mathbf{Z}^T \mathbf{R}_{12}^* \left( 1 - \frac{1}{\nu+1} \mathbf{R}_{22}^* \right) \mathbf{X} \right) = \mathbf{0}^T$$

elde edilir. Burada  $\mathbf{0}$ ,  $p \times q$  boyutlu sıfır matrisidir.

Lemma 6.1 (iii) ve (iv)'den

$$E \left( \mathbf{R}_{22}^* \left( 1 - \frac{2}{\nu+1} \mathbf{R}_{22}^* \right) \right) = \frac{1}{2} \frac{v}{v+3} \mathbf{R}_{11}$$

elde edilir.

Buradan Student  $t$  dağılımının ortak konum-ölçek modelinin parametreleri için elde edilen Fisher bilgi matrisi

$$\mathcal{F}(\boldsymbol{\beta}, \boldsymbol{\gamma}) = \begin{bmatrix} \mathbf{X}^T \mathbf{R}_{11} \mathbf{X} & \mathbf{0} \\ \mathbf{0}^T & \frac{1}{2} \left( \frac{v}{v+3} \right) \mathbf{Z}^T (\mathbf{R}_{11})^2 \mathbf{Z} \end{bmatrix} \quad (6.44)$$

biçiminde yazılabilir. Buna göre parametrelerin ML tahmin edicilerinin kovaryans matrisinin tahmini

$$\widehat{\mathcal{F}}^{-1} = \begin{bmatrix} \left( \mathbf{X}^T \widehat{\mathbf{R}}_{11} \mathbf{X} \right)^{-1} & \mathbf{0} \\ \mathbf{0}^T & 2 \left( \frac{v+3}{v} \right) \left( \mathbf{Z}^T (\widehat{\mathbf{R}}_{11})^2 \mathbf{Z} \right)^{-1} \end{bmatrix}$$

biçimindedir.

Bazı matrissel işlemlerden sonra

$$tr \left( \widehat{\mathcal{F}}^{-1} \right) = tr \left( \left( \mathbf{X}^T \widehat{\mathbf{R}}_{11} \mathbf{X} \right)^{-1} \right) + 2 \frac{(v+3)}{v} tr \left( \left( \mathbf{Z}^T (\widehat{\mathbf{R}}_{11})^2 \mathbf{Z} \right)^{-1} \right) \quad (6.45)$$

$$\left| \widehat{\mathcal{F}}^{-1} \right| = \left| \left( \mathbf{X}^T \widehat{\mathbf{R}}_{11} \mathbf{X} \right)^{-1} \right| + 2^q \left( \frac{v+3}{v} \right)^q \left| \left( \mathbf{Z}^T (\widehat{\mathbf{R}}_{11})^2 \mathbf{Z} \right)^{-1} \right| \quad (6.46)$$

eşitlikleri elde edilir.

Eşitlik (6.45) ve (6.46)'den  $\widehat{\mathcal{F}}^{-1}$  için maksimal karmaşıklık bilgisi

$$\begin{aligned}
C_{1t}(\widehat{\mathcal{F}}^{-1}) &= \frac{s}{2} \log \left[ \frac{tr \widehat{\mathcal{F}}^{-1}}{s} \right] - \frac{1}{2} \log |\widehat{\mathcal{F}}^{-1}| \\
&= \frac{s}{2} \log \left( \frac{tr \left( (\mathbf{X}^T \widehat{\mathbf{R}}_{11} \mathbf{X})^{-1} \right) + 2 \left( \frac{v+3}{v} \right) tr \left( (\mathbf{Z}^T (\widehat{\mathbf{R}}_{11})^2 \mathbf{Z})^{-1} \right)}{s} \right) \\
&\quad - \frac{1}{2} \log \left( \left| (\mathbf{X}^T \widehat{\mathbf{R}}_{11} \mathbf{X})^{-1} \right| + 2^q \left( \frac{v+3}{v} \right)^q \left| (\mathbf{Z}^T (\widehat{\mathbf{R}}_{11})^2 \mathbf{Z})^{-1} \right| \right)
\end{aligned} \tag{6.47}$$

elde edilir.

$\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\gamma})$  parametre vektörünün ML tahmin edicisi  $\widehat{\boldsymbol{\theta}} = (\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\gamma}})$ , parametrelerin ML tahmin edicisinin tahmini kovaryans matrisi  $\widehat{\mathcal{F}}^{-1}$  ve  $s = p + q$  olmak üzere Student  $t$  dağılımının ortak konum-ölçek modeli için *ICOMP*

$$\begin{aligned}
ICOMP_{JLSM-t} &= -2 \log L(\widehat{\boldsymbol{\theta}}) + 2 \left( \frac{s}{2} \log \left[ \frac{tr \widehat{\mathcal{F}}^{-1}}{s} \right] - \frac{1}{2} \log |\widehat{\mathcal{F}}^{-1}| \right) \\
&= -2 \log L(\widehat{\boldsymbol{\theta}}) \\
&\quad + s \log \left( \frac{tr \left( (\mathbf{X}^T \widehat{\mathbf{R}}_{11} \mathbf{X})^{-1} \right) + 2 \frac{v+3}{v} tr \left( (\mathbf{Z}^T (\widehat{\mathbf{R}}_{11})^2 \mathbf{Z})^{-1} \right)}{s} \right) \\
&\quad - \log \left( \left| (\mathbf{X}^T \widehat{\mathbf{R}}_{11} \mathbf{X})^{-1} \right| + 2^q \left( \frac{v+3}{v} \right)^q \left| (\mathbf{Z}^T (\widehat{\mathbf{R}}_{11})^2 \mathbf{Z})^{-1} \right| \right)
\end{aligned} \tag{6.48}$$

formunda elde edilir.

Kritere göre en iyi model, en küçük  $ICOMP_{JLSM-t}$  değerini veren modeldir.

## 6.2 Ortak Konum-Ölçek-Çarpıklık Modelinde *ICOMP* ile Değişken Seçimi

### 6.2.1 Çarpık simetrik dağılımlar ailesi

Gerçek hayatta birçok uygulamada veri seti simetrik olmayıp, fazlasıyla asimetrik gözlemler içerebilir. Bu gibi durumlarda simetrik bir dağılım varsayımı altında yapılan istatistiksel çıkarımlar yanıltıcı olabilir. Normal dağılımdan farklı esnek parametrik dağılım ailelerinin oluşturulması, bu gibi durumlar için önerilen popüler yöntemlerin başında gelmektedir. Literatürde birçok çarpıklaştırma yöntemi ve çarpık dağılım tanımlanmıştır. Bu kısımda oldukça geniş bir aile olan çarpık simetrik aileler ele alınmıştır. Bir çarpıklık fonksiyonunun kullanılmasıyla inşa edilen çarpık simetrik dağılımlar ailesi Azzalini ve Capitanio (2003) ve Wang vd. (2004) tarafından bağımsız olarak tanımlanmıştır.

**Tanım 6.4**  $f$  fonksiyonu  $\mathbf{0}$  etrafında simetrik olasılık yoğunluk fonksiyonu,  $G'$  sıfır etrafında simetrik bir olasılık yoğunluk fonksiyonu olacak şekilde  $G$  mutlak sürekli dağılım fonksiyonu,  $\omega : \mathbb{R}^n \rightarrow \mathbb{R}$  tek fonksiyon ve  $\boldsymbol{\xi} \in \mathbb{R}^n$  olmak üzere

$$h(\mathbf{x}) = 2f(\mathbf{x} - \boldsymbol{\xi}) G(\omega(\mathbf{x} - \boldsymbol{\xi})), \mathbf{x} \in \mathbb{R}^n$$

şeklinde tanımlanan  $h$  fonksiyonu bir olasılık yoğunluk fonksiyonudur (Azzalini ve Capitanio 2003).

**Tanım 6.5**  $f$  fonksiyonu  $\mathbf{0}$  etrafında simetrik olasılık yoğunluk fonksiyonu,  $\pi : \mathbb{R}^n \rightarrow [0, 1]$ ,  $\pi(-\mathbf{x}) = 1 - \pi(\mathbf{x})$  özelliğini sağlayan çarpıklaştırma fonksiyonu ve  $\boldsymbol{\xi} \in \mathbb{R}^n$  olmak üzere

$$h(\mathbf{x}) = 2f(\mathbf{x} - \boldsymbol{\xi}) \pi(\mathbf{x} - \boldsymbol{\xi}), \mathbf{x} \in \mathbb{R}^n$$

şeklinde tanımlanan  $h$  fonksiyonu bir olasılık yoğunluk fonksiyonudur ve  $h$  fonksiyonuna  $\boldsymbol{\xi}$ 'ya göre çarpık simetrik dağılım denir (Wang vd. 2004).

Wang vd. (2004) tarafından Tanım 6.4-6.5 ile aynı dağılımlar ailesinin elde edileceği gösterilmiştir. Bu tanımlar doğrultusunda tüm dağılımların bir çarpık simetrik gösterimi olduğu söylenebilir. Literatürde tanımlanmış çarpık  $t$  (Branco ve Dey 2001,

Azzalini ve Capitanio 2003, Jones ve Faddy 2003, Sahu vd. 2003), çarpık Cauchy (Arnold ve Beaver 2000) dağılımları ve diğer çarpık eliptik dağılımlar (Azzalini ve Capitanio 1999, Branco ve Dey 2001, Sahu vd. 2003) uygun  $f$  ve  $\pi$  fonksiyonlarının seçimiyle çarpık simetrik dağılım formunda yazılabilir (Wang vd. 2004).

Tanım 6.4'te  $\omega(\mathbf{x}) = \lambda \mathbf{x}$  ve  $\Phi$  normal dağılımın dağılım fonksiyonunu göstermek üzere  $G = \Phi$  alınmasıyla elde edilen aileye ise çarpık simetrik normal (skew symmetric normal) dağılımlar ailesi denir. Bu kısımda çarpık simetrik normal dağılımlar ailesi ele alınacaktır.

### 6.2.2 Ortak konum-ölçek-çarpıklık modeli

Simetrik dağılımlara robust alternatif olarak çarpık dağılımlar ailelerinin tanımlanmasıyla heteroskedastik verilerin modellenmesinde çarpıklığın da tahmin edilmesi, modellenmesi problemi ön plana çıkmıştır. Çarpıklığın tahmin edilmesi, modellenmesi üzerine birçok çalışma (bkz. Wu vd. 2013b) olmasına rağmen çarpıklığın modellenmesinde değişken seçimi ile ilgili az sayıda çalışma bulunmaktadır. Li vd. (2017) tarafından önerilen çarpık normal dağılımın ortak konum-ölçek-çarpıklık modeli için cezalandırılmış ML yöntemine dayalı bir değişken seçim yöntemi bu sınırlı sayıda çalışmadan biridir. Bu nedenle bu kısımda ortalama ve varyansla beraber çarpıklığa da ilişkin en iyi modelin belirlenmesi problemi ele alınmıştır. Çarpık simetrik dağılımların ortak konum-ölçek-çarpıklık modelinde değişken seçimi yapılmış ve çarpık simetrik dağılımlara örnek olarak çarpık normal dağılım ele alınmıştır.

Ortak konum-ölçek modeline benzer şekilde ve konum-ölçek modeli ile beraber çarpıklığında modellenmesiyle ortak konum-ölçek-çarpıklık modeli aşağıdaki gibi tanımlanmaktadır:

$$y \sim f(\mu, \sigma, \lambda) \quad (6.49)$$

$$\mu = \mathbf{x}^T \boldsymbol{\beta} \quad (6.50)$$

$$\log \sigma^2 = \mathbf{z}^T \boldsymbol{\gamma} \quad (6.51)$$

$$\lambda = \mathbf{v}^T \boldsymbol{\alpha}$$

Burada  $\mathbf{x} = [x_1, x_2, \dots, x_p]^T$ ,  $\mathbf{z} = [z_1, z_2, \dots, z_q]^T$  ve  $\mathbf{v} = [v_1, v_2, \dots, v_r]^T$  değişken vektörlerini,  $\boldsymbol{\beta} = [\beta_1, \beta_2, \dots, \beta_p]^T$  konum modelinin,  $\boldsymbol{\gamma} = [\gamma_1, \gamma_2, \dots, \gamma_q]^T$  ölçek modelinin ve  $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \dots, \alpha_r]^T$  çarpıklık modelinin bilinmeyen parametre vektörünü göstermektedir.

### 6.2.3 Çarpık simetrik dağılımların ortak konum-ölçek-çarpıklık modelinde değişken seçimi

$(y_i, \mathbf{x}_i, \mathbf{z}_i, \mathbf{v}_i)$  çarpık simetrik normal dağılımından  $n$  birimlik rasgele bir örneklem olsun.  $(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\alpha})$  parametrelerine ilişkin olabirlik ve log-olabirlik fonksiyonu

$$\begin{aligned} L(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\alpha} | \mathbf{y}, \mathbf{X}, \mathbf{Z}, \mathbf{V}) &= \prod_{i=1}^n \frac{2}{\sigma_i} f\left(\frac{y_i - \mu_i}{\sigma_i}\right) \Phi\left(\lambda_i \left(\frac{y_i - \mu_i}{\sigma_i}\right)\right) \\ &= 2^n \left(\prod_{i=1}^n \mathbf{e}^{\mathbf{z}_i^T \boldsymbol{\gamma}}\right)^{-1/2} \prod_{i=1}^n f\left(\frac{y_i - \mathbf{x}_i^T \boldsymbol{\beta}}{\mathbf{e}^{\frac{1}{2} \mathbf{z}_i^T \boldsymbol{\gamma}}}\right) \\ &\quad \times \prod_{i=1}^n \Phi\left(\mathbf{v}_i^T \boldsymbol{\alpha} \left(\frac{y_i - \mathbf{x}_i^T \boldsymbol{\beta}}{\mathbf{e}^{\frac{1}{2} \mathbf{z}_i^T \boldsymbol{\gamma}}}\right)\right) \end{aligned}$$

$$\begin{aligned} \ell(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\alpha} | \mathbf{y}, \mathbf{X}, \mathbf{Z}, \mathbf{V}) &= n \log 2 - \frac{1}{2} \sum_{i=1}^n \mathbf{z}_i^T \boldsymbol{\gamma} + \sum_{i=1}^n \log f\left(\frac{y_i - \mathbf{x}_i^T \boldsymbol{\beta}}{\mathbf{e}^{\frac{1}{2} \mathbf{z}_i^T \boldsymbol{\gamma}}}\right) \\ &\quad + \sum_{i=1}^n \log \Phi\left(\mathbf{v}_i^T \boldsymbol{\alpha} \left(\frac{y_i - \mathbf{x}_i^T \boldsymbol{\beta}}{\mathbf{e}^{\frac{1}{2} \mathbf{z}_i^T \boldsymbol{\gamma}}}\right)\right) \end{aligned} \quad (6.52)$$

şeklinde elde edilir.

Parametrelerin ML tahmin edicileri  $t_i = \frac{y_i - \mathbf{x}_i^T \boldsymbol{\beta}}{\mathbf{e}^{\frac{1}{2} \mathbf{z}_i^T \boldsymbol{\gamma}}}$  ve  $f = f(t_i)$ ,  $f' = f'(t_i)$ ,  $\phi = \phi(\mathbf{v}_i^T \boldsymbol{\alpha} t_i)$ ,  $\Phi = \Phi(\mathbf{v}_i^T \boldsymbol{\alpha} t_i)$  olmak üzere

$$\frac{\partial \ell}{\partial \boldsymbol{\beta}} = - \sum_{i=1}^n \frac{1}{\mathbf{e}^{\frac{1}{2} \mathbf{z}_i^T \boldsymbol{\gamma}}} \left( \frac{f'}{f} + \mathbf{v}_i^T \boldsymbol{\alpha} \frac{\phi}{\Phi} \right) \mathbf{x}_i = \mathbf{0} \quad (6.53)$$

$$\frac{\partial \ell}{\partial \boldsymbol{\gamma}} = - \frac{1}{2} \sum_{i=1}^n \left( 1 + t_i \frac{f'}{f} + \mathbf{v}_i^T \boldsymbol{\alpha} t_i \frac{\phi}{\Phi} \right) \mathbf{z}_i = \mathbf{0} \quad (6.54)$$

$$\frac{\partial \ell}{\partial \boldsymbol{\alpha}} = \sum_{i=1}^n \frac{\phi}{\Phi} t_i \mathbf{v}_i = \mathbf{0} \quad (6.55)$$

denklem sisteminin çözülmesiyle elde edilir.

Diğer taraftan Fisher bilgi matrisi

$$\mathcal{F}(\beta, \gamma, \alpha) = -E \begin{bmatrix} \frac{\partial^2 \ell}{\partial \beta \partial \beta^T} & \frac{\partial^2 \ell}{\partial \beta \partial \gamma^T} & \frac{\partial^2 \ell}{\partial \beta \partial \alpha^T} \\ \frac{\partial^2 \ell}{\partial \gamma \partial \beta^T} & \frac{\partial^2 \ell}{\partial \gamma \partial \gamma^T} & \frac{\partial^2 \ell}{\partial \gamma \partial \alpha^T} \\ \frac{\partial^2 \ell}{\partial \alpha \partial \beta^T} & \frac{\partial^2 \ell}{\partial \alpha \partial \gamma^T} & \frac{\partial^2 \ell}{\partial \alpha \partial \alpha^T} \end{bmatrix} \quad (6.56)$$

eşitliği ile hesaplanır. Bu matrisin bileşenleri  $t = \frac{y - \mathbf{x}^T \beta}{\mathbf{e}^{\mathbf{z}^T \gamma/2}}$  ve  $f = f(t)$ ,  $f' = f'(t)$ ,  $\phi = \phi(\mathbf{v}^T \alpha t)$ ,  $\Phi = \Phi(\mathbf{v}^T \alpha t)$  olmak üzere

$$\frac{\partial^2 \ell}{\partial \beta \partial \beta^T} = \left[ \frac{1}{\mathbf{e}^{\mathbf{z}^T \gamma/2}} \int \left( \frac{f'}{f} + \mathbf{v}^T \alpha \frac{\phi}{\Phi} \right)^2 f dy \right] \mathbf{x} \mathbf{x}^T$$

$$\frac{\partial^2 \ell}{\partial \beta \partial \gamma^T} = \left[ \frac{1}{2\mathbf{e}^{\mathbf{z}^T \gamma/2}} \int \left( \frac{f'}{f} + \mathbf{v}^T \alpha \frac{\phi}{\Phi} \right) \left( 1 + t \frac{f'}{f} + \mathbf{v}^T \alpha t \frac{\phi}{\Phi} \right) f dy \right] \mathbf{x} \mathbf{z}^T$$

$$\frac{\partial^2 \ell}{\partial \beta \partial \alpha^T} = \left[ \frac{-1}{\mathbf{e}^{\mathbf{z}^T \gamma/2}} \int \left( \frac{f'}{f} + \mathbf{v}^T \alpha \frac{\phi}{\Phi} \right) \left( \frac{\phi}{\Phi} t \right) f dy \right] \mathbf{x} \mathbf{v}^T$$

$$\frac{\partial^2 \ell}{\partial \gamma \partial \beta^T} = \left[ \frac{1}{2\mathbf{e}^{\mathbf{z}^T \gamma/2}} \int \left( \frac{f'}{f} + \mathbf{v}^T \alpha \frac{\phi}{\Phi} \right) \left( 1 + t \frac{f'}{f} + \mathbf{v}^T \alpha t \frac{\phi}{\Phi} \right) f dy \right] \mathbf{z} \mathbf{x}^T$$

$$\frac{\partial^2 \ell}{\partial \gamma \partial \gamma^T} = \left[ \frac{1}{4} \int \left( 1 + t \frac{f'}{f} + \mathbf{v}^T \alpha t \frac{\phi}{\Phi} \right)^2 f dy \right] \mathbf{z} \mathbf{z}^T$$

$$\frac{\partial^2 \ell}{\partial \gamma \partial \alpha^T} = \left[ -\frac{1}{2} \int \left( 1 + t \frac{f'}{f} + \mathbf{v}^T \alpha t \frac{\phi}{\Phi} \right) \frac{\phi}{\Phi} t f dy \right] \mathbf{z} \mathbf{v}^T$$

$$\frac{\partial^2 \ell}{\partial \alpha \partial \beta^T} = \left[ \int \left( \frac{f'}{f} + \mathbf{v}^T \alpha \frac{\phi}{\Phi} \right) \frac{\phi}{\Phi} t f dy \right] \mathbf{v} \mathbf{x}^T$$

$$\frac{\partial^2 \ell}{\partial \alpha \partial \gamma^T} = \left[ -\frac{1}{2} \int \frac{\phi}{\Phi} t \left( 1 + t \frac{f'}{f} + \mathbf{v}^T \alpha t \frac{\phi}{\Phi} \right) f dy \right] \mathbf{v} \mathbf{z}^T$$

$$\frac{\partial^2 \ell}{\partial \alpha \partial \alpha^T} = \left[ \int \left( \frac{\phi}{\Phi} \mathbf{v}^T \alpha t - \frac{\phi^2}{\Phi^2} \right) t^2 f dy \right] \mathbf{v} \mathbf{v}^T$$

şeklindedir.

### 6.2.4 Çarpık normal dağılımın ortak konum-ölçek-çarpıklık modelinde değişken seçimi

Bu kısımda çarpık simetrik dağılımlara örnek olarak çarpık normal dağılımın ortak konum-ölçek-çarpıklık modeli ele alınmış, bu model için *ICOMP* kullanılarak değişken seçimi yapılmıştır. Çarpık normal dağılımın ortak konum-ölçek modeli aşağıdaki gibi tanımlanmaktadır:

$$y_i \sim SN(\mu_i, \sigma_i^2, \lambda_i) \quad (6.57)$$

$$\mu_i = \mathbf{x}_i^T \boldsymbol{\beta} \quad (6.58)$$

$$\log \sigma_i^2 = \mathbf{z}_i^T \boldsymbol{\gamma} \quad (6.59)$$

$$\lambda_i = \mathbf{v}_i^T \boldsymbol{\alpha} \quad (6.60)$$

$$i = 1, 2, \dots, n$$

Çarpık normal dağılıma ilişkin konum-ölçek-çarpıklık modelinde de diğer modellerde olduğu gibi *ICOMP* bilgi kriteri ile değişken seçimi yapabilmek için öncelikle parametrelerin ML tahmin edicileri elde edilmelidir:

$(y_i, \mathbf{x}_i, \mathbf{z}_i, \mathbf{v}_i)$  çarpık normal dağılımın ortak konum-ölçek-çarpıklık modelinden bir rasgele örneklem olsun. Olabilirlik ve log-olabilirlik fonksiyonları sırasıyla

$$\begin{aligned} L(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\alpha} | \mathbf{y}, \mathbf{X}, \mathbf{Z}, \mathbf{V}) &= \prod_{i=1}^n \frac{2}{\sigma_i} \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{1}{2} \left( \frac{y_i - \mu_i}{\sigma_i} \right)^2 \right\} \Phi \left( \lambda_i \left( \frac{y_i - \mu_i}{\sigma_i} \right) \right) \\ &= \left( \frac{2}{\pi} \right)^{n/2} \left( \prod_{i=1}^n \frac{1}{\mathbf{e}^{\mathbf{z}_i^T \boldsymbol{\gamma}}} \right)^{1/2} \exp \left( -\frac{1}{2} \sum_{i=1}^n \frac{(y_i - \mathbf{x}_i^T \boldsymbol{\beta})^2}{\mathbf{e}^{\mathbf{z}_i^T \boldsymbol{\gamma}}} \right) \\ &\quad \times \prod_{i=1}^n \Phi \left( \mathbf{v}_i^T \boldsymbol{\alpha} \frac{(y_i - \mathbf{x}_i^T \boldsymbol{\beta})}{\mathbf{e}^{\frac{1}{2} \mathbf{z}_i^T \boldsymbol{\gamma}}} \right) \\ l(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\alpha} | \mathbf{y}, \mathbf{X}, \mathbf{Z}, \mathbf{V}) &= \frac{n}{2} \log \frac{2}{\pi} - \frac{1}{2} \sum_{i=1}^n \mathbf{z}_i^T \boldsymbol{\gamma} - \frac{1}{2} \sum_{i=1}^n \frac{(y_i - \mathbf{x}_i^T \boldsymbol{\beta})^2}{\mathbf{e}^{\mathbf{z}_i^T \boldsymbol{\gamma}}} \\ &\quad + \sum_{i=1}^n \log \Phi(t_i) \end{aligned} \quad (6.61)$$

şeklinde yazılabilir. Burada  $t_i = \mathbf{v}_i^T \boldsymbol{\alpha} \frac{(y_i - \mathbf{x}_i^T \boldsymbol{\beta})}{\mathbf{e}^{\frac{1}{2} \mathbf{z}_i^T \boldsymbol{\gamma}}}$  ve  $\Phi(\cdot)$  fonksiyonu standart normal dağılıma ilişkin dağılım fonksiyonudur.

Parametrelerin ML tahmin edicileri  $\omega_i = \frac{\phi(t_i)}{\Phi(t_i)}$  ağırlıklar olmak üzere

$$\frac{\partial \ell}{\partial \boldsymbol{\beta}} = \sum_{i=1}^n \frac{(y_i - \mathbf{x}_i^T \boldsymbol{\beta}) x_i}{\mathbf{e}^{z_i^T \boldsymbol{\gamma}}} - \sum_{i=1}^n \omega_i \frac{\mathbf{v}_i^T \boldsymbol{\alpha} \mathbf{x}_i}{\mathbf{e}^{\frac{1}{2} z_i^T \boldsymbol{\gamma}}} = \mathbf{0} \quad (6.62)$$

$$\frac{\partial \ell}{\partial \boldsymbol{\gamma}} = \frac{1}{2} \sum_{i=1}^n z_i \left( \frac{(y_i - \mathbf{x}_i^T \boldsymbol{\beta})^2}{\mathbf{e}^{z_i^T \boldsymbol{\gamma}}} - 1 \right) - \frac{1}{2} \sum_{i=1}^n \omega_i t_i z_i = \mathbf{0} \quad (6.63)$$

$$\frac{\partial \ell}{\partial \boldsymbol{\alpha}} = \sum_{i=1}^n \omega_i \left( \frac{y_i - \mathbf{x}_i^T \boldsymbol{\beta}}{\mathbf{e}^{\frac{1}{2} z_i^T \boldsymbol{\gamma}}} \right) \mathbf{v}_i = \mathbf{0} \quad (6.64)$$

denklem sisteminin çözülmesiyle elde edilir. Gösterim kolaylığı için  $\epsilon_i = y_i - \mathbf{x}_i^T \boldsymbol{\beta}$  olmak üzere

$$\mathbf{R}_{11} = \text{diag} \left\{ \frac{1}{\sigma_i^2} \right\}_{i=1}^n \quad (6.65)$$

$$\mathbf{R}_{12} = \text{diag} \left\{ \frac{\epsilon_i}{\sigma_i^2} \right\}_{i=1}^n \quad (6.66)$$

$$\mathbf{R}_{22} = \text{diag} \left\{ \frac{1}{2} \frac{\epsilon_i^2}{\sigma_i^2} \right\}_{i=1}^n \quad (6.67)$$

$$\mathbf{P}_{11} = \text{diag} \left\{ \frac{\lambda_i}{\sigma_i} \right\}_{i=1}^n \quad (6.68)$$

matrisleri tanımlansın. Bu durumda  $\omega_i = \frac{\phi(t_i)}{\Phi(t_i)}$  ve  $\mathbf{W} = [\omega_1, \omega_2, \dots, \omega_n]^T$  olmak üzere  $\boldsymbol{\beta}$  konum parametresinin ML tahmini

$$\hat{\boldsymbol{\beta}} = \left( \mathbf{X}^T \hat{\mathbf{R}}_{11} \mathbf{X} \right)^{-1} \left( \mathbf{X}^T \hat{\mathbf{R}}_{11} \mathbf{y} - \mathbf{X}^T \hat{\mathbf{P}}_{11} \hat{\mathbf{W}} \right) \quad (6.69)$$

eşitliği ile elde edilir.  $\boldsymbol{\gamma}$  ölçek parametre vektörünün ML tahmin edicisi ise ancak

$$\sum_{i=1}^n z_i \left( \frac{(y_i - \mathbf{x}_i^T \hat{\boldsymbol{\beta}})^2}{\mathbf{e}^{z_i^T \boldsymbol{\gamma}}} - \sum_{i=1}^n \frac{\phi(t_i)}{\Phi(t_i)} t_i - 1 \right) = \mathbf{0} \quad (6.70)$$

denkleminin sayısal yöntemlerle çözülmesi ile elde edilebilir.  $\boldsymbol{\alpha}$  çarpıklık parametresi ise  $\mathbf{1}_{n \times 1} = [1, 1, \dots, 1]^T$  olmak üzere

$$\mathbf{v} \mathbf{W}^T \widehat{\mathbf{R}}_{22} \mathbf{1} = \mathbf{0} \quad (6.71)$$

eşitliğinin çözülmesiyle elde edilir. Log-olabilirlik fonksiyonunun parametre vektörlerine göre kısmi türevleri aşağıdaki gibi elde edilir:

$$\frac{\partial^2 \ell}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} = - \sum_{i=1}^n \frac{\left( 1 + (\mathbf{v}_i^T \boldsymbol{\alpha})^2 \omega_i^2 + (\mathbf{v}_i^T \boldsymbol{\alpha})^2 \omega_i t_i \right)}{\mathbf{e}^{z_i^T \boldsymbol{\gamma}}} \mathbf{x}_i \mathbf{x}_i^T \quad (6.72)$$

$$\frac{\partial^2 \ell}{\partial \beta \partial \gamma^T} = \left[ -\sum_{i=1}^n \frac{(y_i - \mathbf{x}_i^T \beta)}{\mathbf{e}^{\mathbf{z}_i^T \gamma}} - \frac{1}{2} \sum_{i=1}^n \frac{\mathbf{v}_i^T \alpha \omega_i (\omega_i t_i + t_i - 1)}{\mathbf{e}^{\frac{1}{2} \mathbf{z}_i^T \gamma}} \right] \mathbf{x}_i \mathbf{z}_i^T \quad (6.73)$$

$$\frac{\partial^2 \ell}{\partial \beta \partial \alpha^T} = \left[ -\sum_{i=1}^n \frac{\omega_i}{\mathbf{e}^{\frac{1}{2} \mathbf{z}_i^T \gamma}} + \sum_{i=1}^n \frac{(y_i - \mathbf{x}_i^T \beta)}{\mathbf{e}^{\mathbf{z}_i^T \gamma}} \mathbf{v}_i^T \alpha \omega_i (\omega_i + t_i) \right] \mathbf{x}_i \mathbf{v}_i^T \quad (6.74)$$

$$\frac{\partial^2 \ell}{\partial \gamma \partial \beta^T} = \left[ -\sum_{i=1}^n \frac{(y_i - \mathbf{x}_i^T \beta)}{\mathbf{e}^{\mathbf{z}_i^T \gamma}} - \frac{1}{2} \sum_{i=1}^n \frac{\mathbf{v}_i^T \alpha \omega_i (\omega_i t_i + t_i - 1)}{\mathbf{e}^{\frac{1}{2} \mathbf{z}_i^T \gamma}} \right] \mathbf{z}_i \mathbf{x}_i^T \quad (6.75)$$

$$\frac{\partial^2 \ell}{\partial \gamma \partial \gamma^T} = \left[ -\frac{1}{2} \sum_{i=1}^n \frac{(y_i - \mathbf{x}_i^T \beta)^2}{\mathbf{e}^{\mathbf{z}_i^T \gamma}} - \frac{1}{4} \sum_{i=1}^n \omega_i t_i (\omega_i t_i + t_i - 1) \right] \mathbf{z}_i \mathbf{z}_i^T \quad (6.76)$$

$$\frac{\partial^2 \ell}{\partial \gamma \partial \alpha^T} = \left[ -\frac{1}{2} \sum_{i=1}^n \frac{\omega_i (y_i - \mathbf{x}_i^T \beta)}{\mathbf{e}^{\frac{1}{2} \mathbf{z}_i^T \gamma}} + \frac{1}{2} \sum_{i=1}^n \frac{(y_i - \mathbf{x}_i^T \beta)}{\mathbf{e}^{\frac{1}{2} \mathbf{z}_i^T \gamma}} \omega_i t_i (\omega_i + t_i) \right] \mathbf{z}_i \mathbf{v}_i^T \quad (6.77)$$

$$\frac{\partial^2 \ell}{\partial \alpha \partial \beta^T} = \left[ -\sum_{i=1}^n \frac{\omega_i}{\mathbf{e}^{\frac{1}{2} \mathbf{z}_i^T \gamma}} + \sum_{i=1}^n \frac{(y_i - \mathbf{x}_i^T \beta)}{\mathbf{e}^{\mathbf{z}_i^T \gamma}} \mathbf{v}_i^T \alpha \omega_i (\omega_i + t_i) \right] \mathbf{v}_i \mathbf{x}_i^T \quad (6.78)$$

$$\frac{\partial^2 \ell}{\partial \alpha \partial \gamma^T} = \left[ -\frac{1}{2} \sum_{i=1}^n \frac{\omega_i (y_i - \mathbf{x}_i^T \beta)}{\mathbf{e}^{\frac{1}{2} \mathbf{z}_i^T \gamma}} + \frac{1}{2} \sum_{i=1}^n \frac{(y_i - \mathbf{x}_i^T \beta)}{\mathbf{e}^{\frac{1}{2} \mathbf{z}_i^T \gamma}} \omega_i t_i (\omega_i + t_i) \right] \mathbf{v}_i \mathbf{z}_i^T \quad (6.79)$$

$$\frac{\partial^2 \ell}{\partial \alpha \partial \alpha^T} = \left[ -\sum_{i=1}^n (\omega_i^2 + \omega_i t_i) \frac{(y_i - \mathbf{x}_i^T \beta)^2}{\mathbf{e}^{\mathbf{z}_i^T \gamma}} \right] \mathbf{v}_i \mathbf{v}_i^T \quad (6.80)$$

Buradan Fisher matrisi

$$\mathcal{F}(\beta, \gamma, \alpha) = -E \begin{bmatrix} \frac{\partial^2 \ell}{\partial \beta \partial \beta^T} & \frac{\partial^2 \ell}{\partial \beta \partial \gamma^T} & \frac{\partial^2 \ell}{\partial \beta \partial \alpha^T} \\ \frac{\partial^2 \ell}{\partial \gamma \partial \beta^T} & \frac{\partial^2 \ell}{\partial \gamma \partial \gamma^T} & \frac{\partial^2 \ell}{\partial \gamma \partial \alpha^T} \\ \frac{\partial^2 \ell}{\partial \alpha \partial \beta^T} & \frac{\partial^2 \ell}{\partial \alpha \partial \gamma^T} & \frac{\partial^2 \ell}{\partial \alpha \partial \alpha^T} \end{bmatrix} \quad (6.81)$$

eşitliği ile elde edilir.

$\theta = (\beta, \gamma, \alpha)$  parametre vektörünün ML tahmin edicisi  $\hat{\theta} = (\hat{\beta}, \hat{\gamma}, \hat{\alpha})$ , parametrelerin ML tahmin edicisinin tahmini kovaryans matrisi  $\hat{\mathcal{F}}^{-1}$  ve  $s = p + q + r$  olmak üzere normal dağılımın ortak konum-ölçek modeli için *ICOMP*

$$ICOMP_{JLSSM-SN} = -2 \log L(\hat{\theta}) + 2 \left( \frac{s}{2} \log \left[ \frac{tr \hat{\mathcal{F}}^{-1}}{s} \right] - \frac{1}{2} \log |\hat{\mathcal{F}}^{-1}| \right)$$

formundadır. Burada, Fisher bilgi matrisi (6.72)-(6.80) eşitliklerinin (6.81) eşitliğinde dikkate alınmasıyla elde edilir. Kritere göre en iyi model, en küçük  $ICOMP_{JLSSM-SN}$  değerini veren modeldir.

### 6.2.5 Çarpık $t$ dağılımın ortak konum-ölçek-çarpıklık modelinde değişken seçimi

Bu kısımda da çarpık simetrik dağılımlara bir başka örnek olarak çarpık  $t$  dağılımın ortak konum-ölçek-çarpıklık modeli ele alınmış, bu model için *ICOMP* kullanılarak değişken seçimi yapılmıştır. Çarpık  $t$  dağılımın ortak konum-ölçek-çarpıklık modeli aşağıdaki gibi tanımlanmaktadır:

$$y_i \sim St(\mu_i, \sigma_i^2, \lambda_i, \nu) \quad (6.82)$$

$$\mu_i = \mathbf{x}_i^T \boldsymbol{\beta} \quad (6.83)$$

$$\log \sigma_i^2 = \mathbf{z}_i^T \boldsymbol{\gamma} \quad (6.84)$$

$$\lambda_i = \mathbf{v}_i^T \boldsymbol{\alpha} \quad (6.85)$$

$$i = 1, 2, \dots, n$$

Çarpık  $t$  dağılımına ilişkin konum-ölçek-çarpıklık modelinde de diğer modellerde olduğu gibi *ICOMP* bilgi kriteri ile değişken seçimi yapabilmek için öncelikle parametrelerin ML tahmin edicileri elde edilmelidir:

$(y_i, \mathbf{x}_i, \mathbf{z}_i, \mathbf{v}_i)$  çarpık  $t$  dağılımın ortak konum-ölçek-çarpıklık modelinden bir rasgele örneklem olsun. Olabilirlik ve log-olabilirlik fonksiyonları sırasıyla

$$\begin{aligned} L(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\alpha} | \mathbf{y}, \mathbf{X}, \mathbf{Z}, \mathbf{V}) &= \prod_{i=1}^n \frac{c_\nu}{\sigma_i} \left( \nu + \frac{(y_i - \mu_i)^2}{\sigma_i^2} \right)^{-\frac{\nu+1}{2}} \\ &\quad \times T_{\nu+1} \left( \lambda_i \left( \frac{y_i - \mu_i}{\sigma_i} \right) \sqrt{\frac{\nu+1}{\nu + \frac{(y_i - \mu_i)^2}{\sigma_i^2}}} \right) \\ &= (c_\nu)^n \left( \prod_{i=1}^n \frac{1}{\mathbf{e}^{\mathbf{z}_i^T \boldsymbol{\gamma}}} \right)^{1/2} \left( \prod_{i=1}^n \left( \nu + \frac{(y_i - \mathbf{x}_i^T \boldsymbol{\beta})^2}{\mathbf{e}^{\mathbf{z}_i^T \boldsymbol{\gamma}}} \right) \right)^{-\frac{\nu+1}{2}} \\ &\quad \times \prod_{i=1}^n T_{\nu+1} \left( \mathbf{v}_i^T \boldsymbol{\alpha} \frac{(y_i - \mathbf{x}_i^T \boldsymbol{\beta})}{\mathbf{e}^{\frac{1}{2} \mathbf{z}_i^T \boldsymbol{\gamma}}} \sqrt{\frac{\nu+1}{\nu + \frac{(y_i - \mathbf{x}_i^T \boldsymbol{\beta})^2}{\mathbf{e}^{\mathbf{z}_i^T \boldsymbol{\gamma}}}}} \right) \\ \ell(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\alpha} | \mathbf{y}, \mathbf{X}, \mathbf{Z}, \mathbf{V}) &= n \log(c_\nu) - \frac{1}{2} \sum_{i=1}^n \mathbf{z}_i^T \boldsymbol{\gamma} - \frac{\nu+1}{2} \sum_{i=1}^n \log \left( \nu + \frac{(y_i - \mathbf{x}_i^T \boldsymbol{\beta})^2}{\mathbf{e}^{\mathbf{z}_i^T \boldsymbol{\gamma}}} \right) \\ &\quad + \sum_{i=1}^n \log T_{\nu+1} \left( t_i \sqrt{\frac{\nu+1}{\nu + \frac{(y_i - \mathbf{x}_i^T \boldsymbol{\beta})^2}{\mathbf{e}^{\mathbf{z}_i^T \boldsymbol{\gamma}}}}} \right) \end{aligned} \quad (6.86)$$

şeklinde yazılabilir. Burada  $c_\nu = \frac{\Gamma(\frac{\nu+1}{2})\nu^{\nu/2}}{\sqrt{\pi}\Gamma(\frac{\nu}{2})}$ ,  $t_i = \mathbf{v}_i^T \boldsymbol{\alpha} \frac{(y_i - \mathbf{x}_i^T \boldsymbol{\beta})}{\mathbf{e}^{\frac{1}{2} \mathbf{z}_i^T \boldsymbol{\gamma}}}$  ve  $T_{\nu+1}(\cdot)$  fonksiyonu  $\nu + 1$  serbestlik derecesine sahip Student  $t$  dağılıma ilişkin dağılım fonksiyonudur.

Böylelikle parametrelerin ML tahmin edicileri  $\omega_i = \frac{\nu+1}{\left[\nu + \frac{(y_i - \mathbf{x}_i^T \hat{\boldsymbol{\beta}})^2}{\hat{\sigma}^2}\right]}$  ağırlıklar olmak üzere

$$\frac{\partial \ell}{\partial \boldsymbol{\beta}} = \sum_{i=1}^n \omega_i \frac{\mathbf{x}_i (y_i - \mathbf{x}_i^T \boldsymbol{\beta})}{\mathbf{e}^{\mathbf{z}_i^T \boldsymbol{\gamma}}} - \frac{\nu}{\nu+1} \sum_{i=1}^n \mathbf{v}_i^T \boldsymbol{\alpha} \frac{t_{\nu+1}(t_i \sqrt{\omega_i})}{T_{\nu+1}(t_i \sqrt{\omega_i})} \frac{\omega_i^{3/2}}{\mathbf{e}^{\frac{1}{2} \mathbf{z}_i^T \boldsymbol{\gamma}}} \mathbf{x}_i = \mathbf{0} \quad (6.87)$$

$$\begin{aligned} \frac{\partial \ell}{\partial \boldsymbol{\gamma}} &= \frac{1}{2} \sum_{i=1}^n \left( \omega_i \frac{(y_i - \mathbf{x}_i^T \boldsymbol{\beta})^2}{\mathbf{e}^{\mathbf{z}_i^T \boldsymbol{\gamma}}} - 1 \right) \mathbf{z}_i \\ &\quad - \frac{\nu}{2(\nu+1)} \sum_{i=1}^n \frac{v_i^T \boldsymbol{\alpha}}{\mathbf{e}^{\frac{1}{2} \mathbf{z}_i^T \boldsymbol{\gamma}}} \frac{t_{\nu+1}(t_i \sqrt{\omega_i})}{T_{\nu+1}(t_i \sqrt{\omega_i})} \mathbf{x}_i \mathbf{z}_i \omega_i^{3/2} = \mathbf{0} \end{aligned} \quad (6.88)$$

$$\frac{\partial \ell}{\partial \boldsymbol{\alpha}} = \sum_{i=1}^n \frac{t_{\nu+1}(t_i \sqrt{\omega_i})}{T_{\nu+1}(t_i \sqrt{\omega_i})} \frac{(y_i - \mathbf{x}_i^T \boldsymbol{\beta})}{\mathbf{e}^{\frac{1}{2} \mathbf{z}_i^T \boldsymbol{\gamma}}} \sqrt{\omega_i} = \mathbf{0} \quad (6.89)$$

denklem sisteminin kökleri ile elde edilir. Gösterim kolaylığı için  $\epsilon_i = y_i - \mathbf{x}_i^T \boldsymbol{\beta}$  olmak üzere

$$\mathbf{R}_{11} = \text{diag} \left\{ \frac{1}{\sigma_i^2} \right\}_{i=1}^n \quad (6.90)$$

$$\mathbf{R}_{12} = \text{diag} \left\{ \frac{\epsilon_i}{\sigma_i^2} \right\}_{i=1}^n \quad (6.91)$$

$$\mathbf{R}_{22} = \frac{1}{2} \text{diag} \left\{ \frac{\epsilon_i^2}{\sigma_i^2} \right\}_{i=1}^n \quad (6.92)$$

$$\mathbf{P}_{11} = \text{diag} \left\{ \frac{\lambda_i}{\sigma_i} \right\}_{i=1}^n \quad (6.93)$$

matrisleri tanımlansın. Bu durumda  $\delta_i = \frac{\nu}{\nu+1} \frac{t_{\nu+1}(\lambda z_i \sqrt{\omega_i})}{T_{\nu+1}(\lambda z_i \sqrt{\omega_i})}$  ve  $\boldsymbol{\Delta} = \text{diag} \{\delta_i\}_{i=1}^n$   $\mathbf{W} = \text{diag} \{\omega_i\}_{i=1}^n$  ve  $\mathbf{W}^{(3/2)} = \text{diag} \{\omega_i^{3/2}\}_{i=1}^n$  olmak üzere  $\boldsymbol{\beta}$  konum parametresinin ML tahmini

$$\hat{\boldsymbol{\beta}} = \left( \mathbf{X}^T \hat{\mathbf{R}}_{11} \hat{\mathbf{W}} \mathbf{X} \right)^{-1} \left( \mathbf{X}^T \hat{\mathbf{R}}_{11} \hat{\mathbf{W}} \mathbf{Y} - \mathbf{X}^T \hat{\mathbf{P}}_{11} \hat{\mathbf{W}}^{(3/2)} \hat{\boldsymbol{\Delta}} \right) \quad (6.94)$$

eşitliği ile elde edilir.  $\boldsymbol{\gamma}$  ölçek ve  $\boldsymbol{\alpha}$  çarpıklık parametre vektörünün ML tahmin edicileri (6.88) ve (6.89) denklemlerinin çözülmesiyle elde edilir.

Log-olabilirlik fonksiyonunun ikinci kısmi türevleri

$$\begin{aligned}
\ell_u &= -u_i \omega_i + \frac{\lambda_i \omega_i^{1/2} \nu \delta_i}{\nu + u_i^2} \\
\ell_{uu} &= \frac{2\omega_i u_i^2}{\nu + u_i^2} - \omega_i - \frac{3\lambda_i \omega_i^{1/2} \nu u_i \delta_i}{(\nu + u_i^2)^2} + \frac{\lambda_i \omega_i^{1/2} \nu \gamma_{ui}}{\nu + u_i^2} \\
\ell_{u\lambda} &= \frac{\nu \omega_i^{1/2} (\delta_i + \lambda_i \delta_{\lambda i})}{\nu + u_i^2} \\
\delta_{ui} &= \frac{\partial \delta_i}{\partial u_i} = -\frac{\nu (\nu + 2) \lambda^2 u_i \delta_i}{(\nu + u_i^2 + \lambda^2 u_i^2) (\nu + u_i^2)} - \frac{\nu \lambda \omega_i^{1/2}}{\nu + u_i^2} \\
\delta_{\lambda i} &= \frac{\partial \delta_i}{\partial \lambda_i} = -\frac{(\nu + 2) \lambda_i u_i^2 \delta_i}{(\nu + u_i^2 + \lambda_i^2 u_i^2)} - u_i \omega_i^{1/2} \delta_i^2
\end{aligned}$$

eşitlikleri yardımıyla

$$\ell_{\beta\beta} = \sum_{i=1}^n \frac{\ell_{uu}}{\sigma_i^2} \mathbf{x}_i \mathbf{x}_i^T \quad (6.95)$$

$$\ell_{\sigma\sigma} = \frac{1}{4} \left( \frac{1}{\sigma_i^2} + \frac{u_i^2}{\sigma_i^2} l_{zz} + \frac{2u_i}{\sigma_i^2} l_u \right) \sigma_i \mathbf{z}_i \mathbf{z}_i^T \quad (6.96)$$

$$\ell_{\lambda\lambda} = u_i \omega_i^{1/2} \delta_{\lambda i} \boldsymbol{\alpha}_i \boldsymbol{\alpha}_i^T \quad (6.97)$$

$$\ell_{\beta\sigma} = \frac{1}{2} \left( \frac{z_i}{\sigma_i^2} I_{uu} + \frac{1}{\sigma_i^2} I_u \right) \sigma_i \mathbf{x}_i \mathbf{z}_i^T \quad (6.98)$$

$$\ell_{\beta\lambda} = -\frac{1}{\sigma} I_{u\lambda} \mathbf{x}_i \boldsymbol{\alpha}_i^T \quad (6.99)$$

$$\ell_{\sigma\lambda} = -\frac{1}{2} \frac{u_i}{\sigma} I_{u\lambda} \sigma_i \mathbf{z}_i \boldsymbol{\alpha}_i^T \quad (6.100)$$

biçiminde elde edilir. Buradan Fisher bilgi matrisi

$$\mathcal{F}(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\alpha}) = -E \begin{bmatrix} \frac{\partial^2 \ell}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} & \frac{\partial^2 \ell}{\partial \boldsymbol{\beta} \partial \boldsymbol{\gamma}^T} & \frac{\partial^2 \ell}{\partial \boldsymbol{\beta} \partial \boldsymbol{\alpha}^T} \\ \frac{\partial^2 \ell}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\beta}^T} & \frac{\partial^2 \ell}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\gamma}^T} & \frac{\partial^2 \ell}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\alpha}^T} \\ \frac{\partial^2 \ell}{\partial \boldsymbol{\alpha} \partial \boldsymbol{\beta}^T} & \frac{\partial^2 \ell}{\partial \boldsymbol{\alpha} \partial \boldsymbol{\gamma}^T} & \frac{\partial^2 \ell}{\partial \boldsymbol{\alpha} \partial \boldsymbol{\alpha}^T} \end{bmatrix} \quad (6.101)$$

eşitliğinden elde edilir.

$\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\alpha})$  parametre vektörünün ML tahmin edicisi  $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\alpha}})$ , parametrelerin ML tahmin edicisinin tahmini kovaryans matrisi  $\hat{\mathcal{F}}^{-1}$  ve  $s = p + q + r$  olmak üzere normal dağılımın ortak konum-ölçek modeli için *ICOMP*

$$ICOMP_{JLSSM-St} = -2 \log L(\hat{\boldsymbol{\theta}}) + 2 \left( \frac{s}{2} \log \left[ \frac{tr \hat{\mathcal{F}}^{-1}}{s} \right] - \frac{1}{2} \log |\hat{\mathcal{F}}^{-1}| \right)$$

formundadır. Kriteria göre en iyi model, en küçük  $ICOMP_{JLSSM-St}$  değerini veren modeldir. Burada Fisher bilgi matrisi (6.95)-(6.100) eşitliklerinin (6.101) e şitliğinde dikkate alınmasıyla elde edilir. Bu tezde  $v$  serbestlik derecesinin bilindiği varsayılmaktadır. Aksi halde  $v$  parametresi de tahmin edilmelidir ve Fisher bilgi matrisinde serbestlik derecesine ilişkin terimler yer almalıdır.

### 6.2.6 Simülasyon çalışması

Bu kısımda ortak konum-ölçek-çarpık modellerinde  $ICOMP$  bilgi kriterinin formları  $ICOMP_{JLSM-N}$ ,  $ICOMP_{JLSM-t}$ ,  $ICOMP_{JLSSM-SN}$  ve  $ICOMP_{JLSSM-St}$  ile değişken seçimine yer verilmiş, kriterlerin performansları karşılaştırılmıştır. Simülasyon çalışması Matlab2013a programlama dili aracılığıyla gerçekleştirilmiştir.

Simülasyon çalışmasında veri

$$y_i \sim SN(\mu_i, \sigma_i, \lambda_i) \quad (6.102)$$

$$\mu_i = \mathbf{x}_i^T \boldsymbol{\beta} \quad (6.103)$$

$$\log \sigma_i^2 = \mathbf{z}_i^T \boldsymbol{\gamma} \quad (6.104)$$

$$\lambda_i = \mathbf{v}_i^T \boldsymbol{\alpha} \quad (6.105)$$

$$i = 1, 2, \dots, n.$$

biçiminde tanımlı olan çarpık normalin ortak konum-ölçek-çarpık modelinden ve

$$y_i \sim St(\mu_i, \sigma_i, \lambda_i, \nu) \quad (6.106)$$

$$\mu_i = \mathbf{x}_i^T \boldsymbol{\beta} \quad (6.107)$$

$$\log \sigma_i^2 = \mathbf{z}_i^T \boldsymbol{\gamma} \quad (6.108)$$

$$\lambda_i = \mathbf{v}_i^T \boldsymbol{\alpha} \quad (6.109)$$

$$i = 1, 2, \dots, n.$$

ile tanımlı çarpık  $t$  dağılımının ortak konum-ölçek-çarpıklık modelinden üretilmiştir.

Burada serbestlik derecesi  $\nu = 3$  alınmıştır.

Öncelikle model parametreleri  $\beta = [1, 1, 0, 0]^T$ ,  $\gamma = [1, 0, 1]^T$  ve  $\alpha = [2, 3, 0]^T$  alınmıştır.  $x_i$ ,  $z_i$  ve  $v_i$  ise  $U(-1, 1)$  dağılımından rasgele üretilip  $y_i$  ise (6.102)-(6.105) modelinden üretilmiştir. Örnek çapı  $n = 30, 50, 100$  ele alınmıştır.

Bu üç ayrı örnek çapı ve yukarıda verilen model için 100 tekrarlı simülasyon gerçekleştirilmiş ve her model için ele alınan model seçim ölçütlerinin doğru modeli seçme sıklıkları hesaplanmıştır. Elde edilen sonuçlar çizelge 6.1’de verilmiştir. Benzer şekilde  $y_i$  (6.106)-(6.109) modelinden üretilerek çarpık  $t$  dağılımının ortak konum-ölçek-çarpıklık modeli için sonuçlar ise çizelge 6.2’de verilmiştir.

Çarpık normal dağılımın ortak konum-ölçek-çarpıklık modelinden üretilen veri seti gözönüne alındığında çizelge 6.1’den  $ICOMP_{JLSSM-SN}$  doğru modelin seçilmesinde en iyi performansa sahiptir. Çarpık  $t$  dağılımının ortak konum-ölçek-çarpıklık modelinden üretilen veri seti gözönüne alındığında çizelge 6.2’den  $ICOMP_{JLSSM-St}$  doğru modelin seçilmesinde en iyi performansa sahiptir.

Sonuç olarak çizelge 6.1 ve 6.2’de görüldüğü üzere verinin üretildiği dağılıma ve modele ilişkin  $ICOMP$  bilgi kriterinin elde edilen formu daha iyi performansa sahiptir. Bununla birlikte, tüm seçim kriterleri konum modelinde değişken seçiminde, ölçek ve çarpıklık modellerinde değişken seçimine göre daha başarılıdır.

Çizelge 6.1 Çarpık normal dağılımın JLSSM modeli için kriterlerin doğru modeli seçme sıklığı

| Model     | n   | ICOMP <sub>JLSM-N</sub> | ICOMP <sub>JLSM-t</sub> | ICOMP <sub>JLSSM-SN</sub> | ICOMP <sub>JLSSM-St</sub> |
|-----------|-----|-------------------------|-------------------------|---------------------------|---------------------------|
| Konum     | 30  | 0.72                    | 0.70                    | 0.85                      | 0.82                      |
|           | 50  | 0.76                    | 0.73                    | 0.88                      | 0.83                      |
|           | 100 | 0.77                    | 0.75                    | 0.91                      | 0.86                      |
| Ölçek     | 30  | 0.69                    | 0.69                    | 0.75                      | 0.72                      |
|           | 50  | 0.75                    | 0.71                    | 0.85                      | 0.76                      |
|           | 100 | 0.76                    | 0.74                    | 0.89                      | 0.76                      |
| Çarpıklık | 30  | -                       | -                       | 0.79                      | 0.69                      |
|           | 50  | -                       | -                       | 0.81                      | 0.72                      |
|           | 100 | -                       | -                       | 0.83                      | 0.75                      |

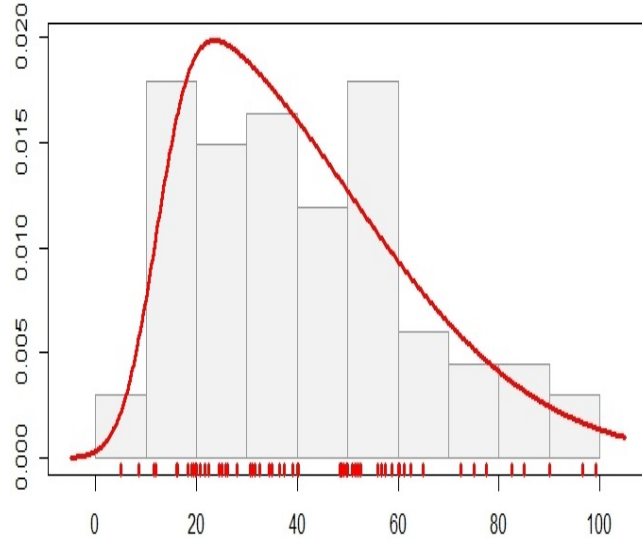
Çizelge 6.2 Çarpık  $t$  dağılımın JLSSM modeli için kriterlerin doğru modeli seçme sıklığı

| Model     | n   | ICOMP <sub>JLSM-N</sub> | ICOMP <sub>JLSM-t</sub> | ICOMP <sub>JLSSM-SN</sub> | ICOMP <sub>JLSSM-St</sub> |
|-----------|-----|-------------------------|-------------------------|---------------------------|---------------------------|
| Konum     | 30  | 0.76                    | 0.75                    | 0.78                      | 0.82                      |
|           | 50  | 0.76                    | 0.79                    | 0.83                      | 0.88                      |
|           | 100 | 0.78                    | 0.79                    | 0.90                      | 0.92                      |
| Ölçek     | 30  | 0.64                    | 0.64                    | 0.75                      | 0.77                      |
|           | 50  | 0.72                    | 0.73                    | 0.85                      | 0.79                      |
|           | 100 | 0.77                    | 0.78                    | 0.80                      | 0.82                      |
| Çarpıklık | 30  | -                       | -                       | 0.67                      | 0.72                      |
|           | 50  | -                       | -                       | 0.77                      | 0.77                      |
|           | 100 | -                       | -                       | 0.78                      | 0.80                      |

### 6.2.7 Arazi kiralari verisi

Bu kısımda yonca dikili tarım arazisi için 1977 yılında ödenen kira varyasyonunu incelemek için Douglas Tiffany tarafından arazi rantı veri seti kullanılmıştır (Weisberg 1985). Veri setine "alr3" R paketinden ulaşılabilir. Bu veri seti 67 ilçede kiralanan tarım arazilerine ilişkin gözlemlerden oluşmaktadır.  $Y$  yonca dikili dönüm başına ortalama kira,  $X_1$  tüm sürülebilir arazi için ödenen ortalama kira,  $X_2$  süt ineklerinin (mil kare başına sayısı) yoğunluğu,  $X_3$  otlak olarak kullanılan tarım arazilerinin kireçlemeyi gerektirdiği durumda  $X_4 = 1$  aksi halde  $X_4 = 0$ 'dır.

Veri seti Hui-qiong ve Wu (2014) tarafından çarpık normal dağılımın ortak konum-ölçek modeline örnek olarak incelenmiştir. Açıklanan değişkenin histogramı şekil 6.1'de verilmiştir. Şekil 6.1'den görüldüğü üzere veri sağa çarpık görülmektedir.



Şekil 6.1 Arazi kiralari veri seti y değişkenine ilişkin histogram

Bu nedenle veri setine çarpık normal dağılımın ortak konum-ölçek modeli uygulanabilir:

$$y_i \sim SN(\mu_i, \sigma_i, \lambda_i), i = 1, 2, \dots, 67. \quad (6.110)$$

$$\mu_i = \beta_0 + x_{i1}\beta_1 + x_{i2}\beta_2 + x_{i3}\beta_3 + x_{i4}\beta_4 \quad (6.111)$$

$$\log \sigma_i^2 = \gamma_0 + x_{i1}\gamma_1 + x_{i2}\gamma_2 + x_{i3}\gamma_3 + x_{i4}\gamma_4 \quad (6.112)$$

$$\lambda_i = \alpha_0 + x_{i1}\alpha_1 + x_{i2}\alpha_2 + x_{i3}\alpha_3 + x_{i4}\alpha_4$$

$ICOMP_{JLSM-N}$  ve  $ICOMP_{JLSSM-SN}$  verilen değişken seçim kriterleri ve  $AIC$  ile elde edilen sonuçlar çizelge 6.3'de özetlenmiştir.

Çizelge 6.3 Arazi kiralari veri seti için çarpık normal dağılımın JLSSM'de değişken seçimi sonuçları

|                    | Konum Modeli    | Ölçek Modeli | Çarpıklık Modeli |
|--------------------|-----------------|--------------|------------------|
| $AIC_{JLSM}$       | $x_1, x_2, x_3$ | $x_2, x_3$   | —                |
| $AIC_{JLSSM}$      | $x_1, x_2$      | $x_2, x_3$   | $x_2$            |
| $ICOMP_{JLSM-N}$   | $x_1, x_2$      | $x_2, x_3$   | —                |
| $ICOMP_{JLSSM-SN}$ | $x_1, x_2$      | $x_3$        | $x_2, x_3$       |

Çizelge 6.3'te görüldüğü üzere, konum modelinde  $x_1, x_2$  ve ölçek modelinde  $x_3$  değişkenleri tüm kriterler tarafından ortak seçilen değişkenlerdir. Konum ve ölçek modelleri için  $ICOMP_{JLSSM}$  ile elde edilen sonuçlar Li vd. (2014) tarafından JLSM kullanılarak elde edilen sonuçlar ile aynıdır. Bununla beraber  $ICOMP_{JLSSM}$  tarafından çarpıklık modeli için  $x_2$  ve  $x_3$  değişkenleri seçilmiştir. Burada  $AIC_{JLSM}$  ve  $AIC_{JLSSM}$  ile ilgili olabilirlik fonksiyonu ile klasik  $AIC$  bilgi kriteri gösterilmiştir.

### 6.2.8 Avustralya Spor Enstitüsü verisi

Bu kısımda dördüncü bölümde ele alınan Avustralya Spor Enstitüsü veri setini modellemek üzere JLSM ve JLSSM modelleri kullanılmıştır. Veri seti çarpık  $t$  dağılımının ortak konum-ölçek-çarpıklık modeline örnek olarak ele alınmıştır.

$$y_i \sim St(\mu_i, \sigma_i, \lambda_i, \nu) \quad (6.113)$$

$$\mu_i = \beta_0 + x_{i1}\beta_1 + x_{i2}\beta_2 + x_{i3}\beta_3 + \cdots + x_{i8}\beta_8 \quad (6.114)$$

$$\log \sigma_i^2 = \gamma_0 + x_{i1}\gamma_1 + x_{i2}\gamma_2 + x_{i3}\gamma_3 + \cdots + x_{i8}\gamma_8 \quad (6.115)$$

$$\lambda_i = \alpha_0 + x_{i1}\alpha_1 + x_{i2}\alpha_2 + x_{i3}\alpha_3 + \cdots + x_{i8}\alpha_8 \quad (6.116)$$

$$i = 1, 2, \dots, 67.$$

$Y$  yağsız vücut kitlesi olmak üzere,  $X_1, X_2, \dots, X_8$  değişkenleri çizelge 6.4'de verilmiştir.

Çizelge 6.4 AIS veri seti değişkenleri

---

|       |                                       |
|-------|---------------------------------------|
| $x_1$ | : rcc (kırmızı hücre sayısı)          |
| $x_2$ | : wcc (beyaz hücre sayısı)            |
| $x_3$ | : Hc (Hematokrit)                     |
| $x_4$ | : Hg (Hemoglobin)                     |
| $x_5$ | : Fe (plazma ferritin konsantrasyonu) |
| $x_6$ | : bmi (vücut kitle indeksi)           |
| $x_7$ | : sff (deri kıvrımlarının toplamı)    |
| $x_8$ | : bfat (vücut yağ yüzdesi)            |

---

$ICOMP_{JLSM-t}$ ,  $ICOMP_{JLSSM-St}$  kriterleri kullanılarak değişken seçimi yapılmıştır. Sonuçlar çizelge 6.5'te verilmiştir.

Çizelge 6.5 AIS veri seti için çarpık  $t$  dağılımın JLSSM'de değişken seçimi sonuçları

|                    | Konum Modeli                       | Ölçek Modeli   | Çarpıklık Modeli    |
|--------------------|------------------------------------|----------------|---------------------|
| $AIC_{JLSM}$       | $\{x_1, x_4, x_5, x_6, x_8\}$      | $\{x_8\}$      | -                   |
| $AIC_{JLSSM}$      | $\{x_1, x_4, x_5, x_6, x_7, x_8\}$ | $\{x_5, x_8\}$ | $\{x_1, x_7, x_8\}$ |
| $ICOMP_{JLSM-t}$   | $\{x_1, x_4, x_5, x_7, x_8\}$      | $\{x_5, x_8\}$ | -                   |
| $ICOMP_{JLSSM-St}$ | $\{x_1, x_4, x_5, x_7, x_8\}$      | $\{x_7, x_8\}$ | $\{x_1, x_4, x_8\}$ |

Çarpıklığın modellendiği durumlarda  $AIC_{JLSSM}$  kriteri,  $ICOMP_{JLSSM-St}$  kriterinin seçtiği değişkenlerle birlikte  $x_6$  değişkenini içeren konum modelini seçmiştir. Ölçek modelinde ise  $x_8$  değişkeni hem  $AIC_{JLSSM}$  hem  $ICOMP_{JLSSM-St}$  kriterine göre etkilidir. Bununla birlikte  $AIC_{JLSSM}$  kriteri ölçek modelinde  $x_5$  değişkenini seçerken,  $ICOMP_{JLSSM-St}$  kriteri  $x_7$  değişkenini seçmektedir. Her iki kriter de çarpıklığın modellenmesinde  $x_1$  ve  $x_8$  değişkenini ortak olarak seçmektedir.  $AIC_{JLSSM}$  kriterinin seçtiği çarpıklık modeli  $x_7$  değişkenini içerirken,  $ICOMP_{JLSSM-St}$  kriterinin seçtiği çarpıklık modeli  $x_4$  değişkenini içermektedir.

Veri setinin modellenmesinde Student  $t$  dağılımının ortak konum-ölçek modeli göz önüne alındığında,  $AIC_{JLSM}$  ve  $ICOMP_{JLSM-t}$  kriterlerine göre seçilen konum modellerinin ortak değişkenleri  $x_1, x_4, x_5$  ve  $x_8$ 'dir. Bununla birlikte  $AIC_{JLSM}$  kriterinin seçtiği konum modeli  $x_6$  ve  $ICOMP_{JLSM-t}$  kriterinin seçtiği konum modeli ise  $x_7$  değişkenini içermektedir.  $AIC_{JLSM}$  kriterine göre sadece  $x_8$  değişkeni,  $ICOMP_{JLSM-t}$  kriterine göre ise hem  $x_5$  hem de  $x_8$  değişkenleri  $y$  değişkeninin ölçeği üzerinde etkilidir.

Çizelge 6.5'te verilen sonuçlara göre  $ICOMP_{JLSM-t}$  ve  $ICOMP_{JLSSM-St}$  kriterlerinin yakın modelleri seçtikleri görülmektedir. Bu iki kriter aynı değişkenlerden oluşan konum modelini seçmektedir. Bununla beraber seçilen ölçek modellerinde

$x_8$  deęiřkeni ortaktır. Çizelge 6.5'te verilen tüm sonuçlar birlikte gözönüne alınacak olursa,  $x_1$  ve  $x_4$  deęiřkenleri deęiřkeninin konum ve çarpıklığı üzerinde etkiye sahip olmasına rağmen ölçeęi üzerinde bir etkiye sahip deęildir.  $ICOMP_{JLSM-t}$  ve  $ICOMP_{JLSSM-St}$  kriterleri ile Li vd. (2017) tarafından yapılan çalışmada elde edilen sonuçlara yakın sonuçlar elde edilmiştir.



## 7. TARTIŞMA VE SONUÇ

İstatistiksel modelleme ve veri analizi için model seçimi problemi en az parametre tahmini kadar önem taşımaktadır. Literatürdeki değişken seçim problemleri ele alındığında bilgi tabanlı model seçim kriterleri AIC bilgi kriterinin tanımlanmasından günümüze kadar artan bir ilgi görmüş ve hala mevcut kriterleri iyileştirme üzerine birçok çalışma yapılmaktadır. Bu çalışmalar incelendiğinde model varsayımlarının bir veya birkaçının sağlanmaması ve/veya veri setinde aykırı değer olması durumları üzerinde hala çalışmaların devam etmekte olduğu görülmektedir.

Çoğu klasik bilgi tabanlı model seçim kriteri model karmaşıklığını yeterli derecede tanımlayamamakla beraber, klasik tahmin yöntemlerine dayalı oldukları için yukarıda bahsedilen problemlerden kolaylıkla etkilenmektedir. Bu nedenle bu tezde Bozdoğan'ın *ICOMP* bilgi kriterinin robust formları farklı dağılım varsayımları altında çoklu regresyon modeli için elde edilmiştir. Normal dağılıma alternatif olarak veri setinin kalın kuyruklu olduğu durumlar için Student  $t$ , çarpık olduğu durumlar için Azzalini tipi çarpık normal ve veri setinin hem kalın kuyruklu hem de çarpık olduğu durumlar için Azzalini tipi çarpık  $t$  dağılım varsayımları ele alınmıştır. Bu dağılım varsayımları altında kovaryans karmaşıklığının yeniden hesaplanmasıyla Student  $t$  dağılımına ilişkin serbestlik derecesi, çarpık normal dağılıma ilişkin çarpıklık parametresi ve çarpık  $t$  dağılımı için çarpıklık parametresi ile serbestlik derecesi model karmaşıklığına dahil edilmiştir. Bu sayede *ICOMP*'ın önerilen formları ile daha iyi sonuçlar elde edilmiştir. Hataların kalın kuyruklu/çarpık olduğu durumlarda model seçimi için orijinal *ICOMP* kriterine yerine  $ICOMP_t$ ,  $ICOMP_{SN}$  ve  $ICOMP_{St}$  kriterlerinin kullanılması önerilmiştir.

Yukarıda da bahsedildiği gibi çoğu klasik bilgi tabanlı model seçim kriteri aykırı gözlemlere karşı hassastır. Bu nedenle model karmaşıklığını daha iyi formülize eden bir robust seçim kriteri elde etmek için robust M, S ve MM tahmin edicilerine dayalı robust *ICOMP* bilgi kriteri önerilmiştir. M tahmin edicileri,  $y$  yönündeki aykırı değerlere karşı robust olmalarına rağmen  $x$  yönündeki aykırı gözlemlerin etkisine karşı

kırılma noktası  $1/n$ 'dir. Bu nedenle kırılma noktası yüksek olan S ve MM tahmin edicileri de göz önüne alınmıştır. Simülasyon sonuçları değerlendirildiğinde beklendiği gibi, orijinal *ICOMP*, aykırı değer olmayan veriler için robust *ICOMP* kriterlerinden ( $RICOMP_M$ ,  $RICOMP_S$ ,  $RICOMP_{MM}$ ) daha iyi performansa sahiptir. Bunun yanı sıra veri setinde aykırı değer olması halinde orijinal *ICOMP*, LS tahmin edicisine dayandığından beklendiği gibi en düşük performansa sahiptir. Bu durum normal dağılım varsayımından kaynaklanmaktadır.

Çoklu doğrusal regresyon analizinde karşılaşılan bir diğer sorun ise çoklu doğrusal bağlantı problemidir. Veri seti hem aykırı değer içerdiğinde hem de çoklu doğrusal bağlantı durumunda klasik yöntemler yalnız başına güvenilir bir analiz için yeterli olmamaktadır. Bu iki problem için önerilen yöntemlerin birleştirilmesi, bu problemler için bir çözüm yolu olabilir. Bu düşünce ile *ICOMP* bilgi kriterinin robust ridge tahmin edicilerine dayalı bir başka genellemesi önerilmiştir. Simülasyon sonuçlarına göre *ICOMP* bilgi kriterinin bu robust genellemesi, veri seti aykırı değer içerdiğinde ve çoklu doğrusal bağlantı problemi olduğunda orijinal *ICOMP* bilgi kriterinden daha başarılı sonuçlar vermektedir. Elde edilen bulgulara göre veri seti aykırı değer içerdiğinde ve/veya çoklu doğrusal bağlantı durumunda dayanıklı *ICOMP* kriterlerinin kullanılması önerilmiştir.

Değişken seçimi regresyon analizinde önemli bir konudur. Ancak bilindiği kadarıyla, mevcut değişken seçim yöntemlerinin çoğu yalnızca konum açıklama değişkenleri ile sınırlıdır. Uygulamada, istatistiksel modellerin en yaygın kullanılan varsayımı olan hataların sabit varyanslı olması varsayımı genellikle sağlanmaz. Bu durumda hangi değişkenlerin varyansı yönlendirdiğinin belirlenmesi önemlidir. Bunun yanı sıra veri simetrik değil ise çarpıklığın modellenmesi ihtiyacı da söz konusu olur. Bu modellemeler ile beraber daha iyi tahminler ve istatistiki çıkarımlar elde edilir. Bu nedenle bu tezde konum ölçek dağılımlar ailesinin ortak konum-ölçek modeli ve çarpık simetrik dağılımların ortak konum-ölçek-çarpıklık modeli için *ICOMP* ile değişken seçimi problemi ele alınmıştır. Konum ölçek dağılımlar ailesine örnek olarak normal dağılım ve Student  $t$  dağılım alınmıştır. Bu dağılımlara ilişkin ortak

konum-ölçek modeli için *ICOMP* kriterinin formları elde edilmiştir. Çarpık simetrik dağılımlar ailesine örnek olarak da çarpık normal dağılım ve çarpık  $t$  dağılım ele alınmıştır. Bu dağılımlara ilişkin ortak konum-ölçek-çarpıklık modeli için *ICOMP* kriterinin formları elde edilmiştir. Önerilen formlar simülasyon çalışması ve gerçek veri örnekleri ile karşılaştırılmıştır.

İleriki çalışmalarda bu tezde ele alınan modellerin yanı sıra, biyolojide tepki modeli, finans sektöründeki ayrık seçim modelleri gibi doğrusal regresyon modelinin bir genellemesi olan tekli indeks (single index) modelde *ICOMP* ile değişken seçimi problemi ele alınacaktır. Ayrıca ele alınan modeller için parametrik olmayan bir yaklaşım olan ampirik olabilirlik ile *ICOMP* bilgi kriterinin birlikte kullanılmasının etkileri incelenecektir.

## KAYNAKLAR

- Aas, K. and Haff, I.H. 2006. The Generalized Hyperbolic Skew Student's  $t$ -Distribution. *Journal of Financial Econometrics*, 4(2), 275-309.
- Ahsanullah, M., Kibria, B.M.G. and Shakil, M. 2014. Normal and Student's  $t$  Distributions and Their Applications. *Atlantis Studies in Probability and Statistics*, Atlantis Press, 144, Paris.
- Aitkin, M. 1987. Modelling variance heterogeneity in Normal regression using GLIM. *Appl. Statist.*, 36, 332-339.
- Akaike, H. 1973. Information theory and extension of the maximum likelihood principle. In B.N Petrov and F. Csáki (Eds.), *Second international symposium on information theory*, Akadémiai Kiadó, Budapest, 267-281.
- Akaike, H. 1974. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6), 716-723.
- Alfonso-Reese, L.A., Ashby, F.G. and Brainard, D.H. 2002. What makes a categorization task difficult? *Perception & Psychophysics*, 64 (4), 570-583.
- Andersen, R. 2008. *Modern Methods For Robust Regression*. SAGE Publications, California.
- Andrews, D.F. 1974. A robust method for multiple linear regression. *Technometrics*, 16, 523-531.
- Arellano-Valle, R.B. and Genton, M.G. 2005. On fundamental skew distributions. *J. Multivariate Anal.* 96, 93-116.
- Arellano-Valle, R.B. and Genton, M.G. 2010. Multivariate extended skew- $t$  distributions and related families. *METRON*, 68, 201-234.
- Arellano-Valle, R.B., Contreras-Reyes, J.E. and Genton, M.G. 2010. Shannon Entropy and Mutual Information for Multivariate Skew-Elliptical Distributions. *Scandinavian Journal of Statistics*, 40(1), 42-62.
- Arnold, B.C. and Beaver, R.J. 2000. The skew-Cauchy distribution. *Statist. Probab. Lett.*, 49, 285-290.
- Arslan, O. 1993. *Multivariate Robust Analysis Based on the  $t$ -Distribution and the EM Algorithm*. Ph. D. Thesis, Leeds University, UK.
- Arslan, O. and Billor, N. 1996. Robust ridge regression estimation based on the GM estimators, *Journal of Mathematical and Computational Science*, 9(1), 1-9.
- Arslan, O. 1996. Robustness asymptotic and computational properties of the  $tM$  estimators. *Journal of Institute of Mathematics and Computer Sciences*, 9(1) 119-127.

- Arslan, O. 1997. Asymptotic Rejection Probability of Constrained M estimators. Bulletin of the 51st Session of the International Statistical Institute, Contributed Book 2, 283–284.
- Arslan, O. and Billor, N. 2000. Robust Liu estimator based on an M estimator. Journal of Applied Statistics, 27(1), 39–47.
- Arslan, O., Edlund, O. and Ekblom, H. 2002. Algorithm to compute S and CM Estimates for regression. *Metrika*, 55(1–2), 37–51.
- Arslan, O. 2012. Weighted LAD LASSO method for robust parameter estimation and variable selection in regression. *Computational Statistics & Data Analysis*, 56(6), 1952–1965.
- Arslan, O. 2016. Penalized MM Regression Estimation with  $L_\gamma$  Penalty: A Robust Version of Bridge Regression. *A Journal of Theoretical and Applied Statistics*, 50(6), 1236–1260.
- Askin, G.R. and Montgomery, D.C. 1980. Augmented robust estimators, *Technometrics*, 22, 333–341.
- Azzalini, A. 1985. A class of distributions which includes the normal ones. *Scand. J. Statist.*, 12, 171–178.
- Azzalini, A. 1986. Further results on a class of distributions which includes the normal ones. *Statistica*, 46, 199–208.
- Azzalini, A. and Dalla-Valle, A. 1996. The multivariate skew-normal distribution. *Biometrika*, 83, 715–726.
- Azzalini, A. and Capitanio, A. 1999. Statistical applications of the multivariate skew normal distributions. *J. R. Stat. Soc. Ser. B*, 61, 579–602.
- Azzalini, A. and Capitanio, A. 2003. Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t distribution. *J. Roy. Statist. Soc. Ser. B* 65, 367–389.
- Azzalini, A. 2005. The Skew-normal Distribution and Related Multivariate Families. *Scandinavian Journal of Statistics*, 32(2), 159–188.
- Azzalini, A. and Capitanio, A. 2014. The Skew-Normal and Related Families. *IMS monographs*, Cambridge University Press., 270, Cambridge.
- Beaton, A.E. and Tukey, J.W. 1974. The Fitting of Power Series, Meaning Polynomials, Illustrated on Band-Spectroscopic Data. *Technometrics* 16(2), 147–185.
- Bhansali, R. and Downham, D.Y. 1977. Some Properties of the Order of an Autoregressive Model by a Generalization of Akaike’s FPE Criterion. *Biometrika* 64(3)
- Birnbaum, Z.W. 1950. Effect of linear truncation on a multinormal population. *Ann. Math. Statist.* 21, 272–279.

- Box, G.E.P. 1953. Non-Normality and Tests on Variances. *Biometrika*, 40, 318–335.
- Bozdogan, H. 1987a. Model selection and Akaike’s Information Criterion (AIC): The general theory and its analytical extensions. *Psychometrika*, 52 (3), 345-370.
- Bozdogan, H. and Ramirez, D.E. 1987b. An Expert Model Selection Approach to Determine the “Best” Pattern Structure in Factor Analysis Models. *Multivariate Statistical Modeling and Data Analysis. Theory and Decision Library (Series B: Mathematical and Statistical Methods)*, Bozdogan H., Gupta A.K. (eds), vol 8. Springer, Dordrecht, 35-60.
- Bozdogan, H. 1988. *ICOMP*: A new model selection criterion. *Classification and Related Methods of Data Analysis*. 599-608.
- Bozdogan, H. 1990. On the information-based measure of covariance complexity and its application to the evaluation of multivariate linear models. *Communications in Statistics - Theory and Methods*, 19, 221-278.
- Bozdogan, H. and Haughton, D.M.A. 1998. Information complexity criteria for regression models. *Computational Statistics and Data Analysis*, 28, 51-76
- Bozdogan, H. 2000. Akaike’s information criterion and recent developments in information complexity. *Journal of Mathematical Psychology*, 44, 62-91.
- Bozdogan, H. 2003. Intelligent statistical data mining with information complexity and genetic algorithms. *Statistical Data Mining and Knowledge Discovery. Joint International Summer School JISS-2003*, Vol. II, July 23th – 30th, 2003. Universidade de Lisboa Lisbon, Portugal.
- Bozdogan, H. and Bearse, P. 2003. Information complexity criteria for detecting influential observations in dynamic multivariate linear models using the genetic algorithm. *Journal of Statistical Planning and Inference* 114(1), 31-44.
- Bozdogan, H. 2010. A New Class of Information Complexity (*ICOMP*) Criteria with an Application to Customer Profiling and Segmentation. *Istanbul University Journal of the School of Business Administration*, 39(2), 370-398.
- Branco, M.D. and Dey, D.K. 2001. A general class of multivariate skew-elliptical distributions. *J. Multivariate Anal.* 79, 99–113.
- Casella, G. and Berger, R.L. 2002. *Statistical Inference (Second Edition)*. Duxbury/Thomson Learning, 660, USA.
- Contreras-Reyes, J.E. and Arellano-Valle, R.B. 2012. Kullback–Leibler divergence measure for multivariate skew-normal distributions, *Entropy*, 14, 1606 - 1626.
- Contreras-Reyes, J.E. 2014. Asymptotic form of the Kullback–Leibler divergence for multivariate asymmetric heavy-tailed distributions, *Physica A*, 395, 200 - 208.
- Çetin, M.C. 2000. Sağlam Regresyonda Değişken Seçim Ölçütleri. Doktora Tezi, Hacettepe Üniversitesi, Fen Bilimleri Enstitüsü, İstatistik Anabilim Dalı, Ankara.

- Çetin, M.C. ve Orsoy, A. 2001. Doğrusal Regresyonda Sağlam Tahmin Ediciler ve Bir Uygulama. *Anadolu Üniversitesi Bilim ve Teknoloji Dergisi A - Uygulamalı Bilimler ve Mühendislik*, 2(2), 265-270.
- Çetinkaya, O.Z. 2011. Belirsizliğin Ölçülmesi ve Entropi, *İktisat Fakültesi Mecmuası*, 44, 1-4.
- Dunnett, C.W. and Sobel, M. 1955. Approximations to the probability integral and certain percentage points of a multivariate analogue of Student's t-distribution. *Biometrika*, 42 (1-2), 258 - 267.
- Efron, B., Hastie, T., Johnstone, I. and Tibshirani, R. 2004. Least angle regression (with discussion). *Ann. Statist.*, 32, 407–499.
- Fernandez, C. and Steel, M. 1998. On Bayesian Modelling of Fat Tails and Skewness. *Journal of the American Statistical Association*, 93, 359–371.
- Fisher, R. A. 1922. On the mathematical foundations of theoretical statistics. *Philos. Trans. Roy. Soc. London A.*, 222, 309–368.
- Fisher, R.A. 1925. Applications of student's t distribution. *Metron*, 5, 90–104.
- Fonseca, T.C.O., Ferreira, M.A.R. and Migon, H.S. 2008. Objective Bayesian analysis for the Student-t regression model. *Biometrika*, 95(2), 325–333.
- Frank, I.E. and Friedman, J.H. 1993 A statistical view of some chemometrics regression tools (with discussion). *Technometrics* 35, 109–148.
- Fu, W.J. 1998. Penalized regressions: the bridge versus the lasso. *J. Comput. Graph. Stat.*, 7, 397–416.
- Gao, P.A. 2005. A subadditive property of the digamma function, *RGMIA Res. Rep. Coll.* 8 (3), Web Sites: <http://vuir.vu.edu.au/18103/1/Psi.pdf>.
- Gauss, C.F. 1809. *Theoria Motus Corporum Coelestium in Sectionibus Conicis Solem Ambientium* Frid. Perthes et I H Besser, Hamburgi.
- Genton, M.G. 2004. *Skew-Elliptical Distributions and their Applications: A Journey Beyond Normality*, Chapman & Hall/CRC, Boca Raton, FL.
- Gosset, W.S. 1908. The probable error of a mean. *Biometrika*, 6, 1–25.
- Gujarati, D.N. 2003. *Basic Econometrics*, 4. Ed., McGraw-Hill, 1002, New York.
- Gupta, A.K. 2003. Multivariate skew t-distribution. *A Journal of Theoretical and Applied Statistics*, 37(4), 359-363.
- Hahn, P.R. and Carvalho, C.M. 2014. Decoupling Shrinkage and Selection in Bayesian Linear Models: A Posterior Summary Perspective. *Journal of the American Statistical Association*, 110 (509), 435-448.

- Hampel, F.R. 1971. A General Qualitative Definition of Robustness. *The Annals of Mathematical Statistics* 42(6), 1887-1896.
- Hampel, F.R. 1974. The influence curve and its role in robust estimation, *J. Am. Statist. Assoc.* 69: 383-393.
- Hampel, F.R. 1983. Some aspects of model choice in robust statistics. *Proceedings of the 4th Session of ISI Book Madrid*.
- Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J. and Stahel, W.A. 1986. *Robust Statistics: The Approach Based on Influence Functions*. John Wiley & Sons, New York.
- Hansen, B. 1994. Autoregressive Conditional Density Estimation. *International Economic Review*, 35(3), 705-730.
- Harvey, A.C. 1976. Estimating regression models with multiplicative heteroscedasticity. *Econometrica*, 44, 460-465.
- Hoerl, A.E. and Kennard, R.W. 1970a. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12, 55-67.
- Hoerl, A.E. and Kennard, R.W. 1970b. Ridge regression: Applications to nonorthogonal problems. *Technometrics*, 12, 69-82.
- Hoerl, A.E, Kennard, R.W. and Baldwin, K.F. 1975. Ridge regression: Some simulations. *Communications in Statistical Theory*, 4(2), 105-123.
- Holland, P.W. 1973. Weighted ridge regression: Combining ridge and robust regression methods, *NBER Working Paper Series*, 11, 1-19.
- Holland, P.W. and Welsch, R.E. 1977. Robust regression using iteratively reweighted least-squares. *Communications in Statistics - Theory and Methods*, 6 (9), 813-827.
- Horton, J.A. 1989. Diagnosis by the Bit: A Method for Evaluating the Diagnostic Process. *Annals of Clinical Laboratory Science*, 19(5), 323-331.
- Huber, P.J. 1964. Robust estimation of a location parameter. *The Annals of Mathematical Statistics*, 35, 73-101.
- Huber, P. J. 1967. The Behavior of Maximum Likelihood Estimates Under Nonstandard Conditions, *Proceedings of the 5th Berkeley Symposium*, 1, 221-233.
- Huber, P.J. 1973. Robust Regression: Asymptotics, Conjectures and Monte Carlo. *The Annals of Statistics*, 1-5, 799-821.
- Huber, P.J. 1977. Robust Covariances. *Statistical Decision Theory and Related Topics*, II. S. S. Gupta, D.S. Moore, Eds, Academic Press, New York.

- Huber, P.J. 1981. Robust statistics, New York: John Wiley & Sons, Inc., ISBN 0-471-41805-6, MR 606374. Republished in paperback, 2004. 2nd ed., Wiley, 2009.
- Huber, P.J. and Ronchetti, E.M. 2009. Robust Statistics. John Wiley & Sons, Hoboken, New Jersey.
- Hurvich, C.M. and Tsai, C.L. 1991. Bias of the corrected AIC criterion for underfitted regression and time series models. *Biometrika*, 78(3), 499-509
- Jones, M.C. and Faddy, M.J. 2003. A skew extension of the  $t$  distribution, with applications. *J. Roy. Statist. Soc. Ser. B*, 65, 159–174.
- Jureckova, J. 1971. Nonparametric estimate of regression coefficients, *Ann. Math. Stat.*, 42, 1328-1338.
- Kirtland, K.M. 2017. Outlier Detection and Multicollinearity in Sequential Variable Selection: A Least Angle Regression-Based Approach. *Doktora Tezi*, Cornell University.
- Knight, K. and Fu, W.J. 2000. Asymptotics for Lasso-type estimators. *Ann Stat* 28:1356–1378
- Konishi, S. and Kitagawa, G. 2008. Information Criteria and Statistical Modelling. New York: Springer.
- Kullback, S. 1959. Information Theory. New-York: Wiley.
- Kullback, S. and Leibler, R. 1951. On information and sufficiency. *Ann. Math. Statist.*, 22, 79–86.
- Kullback, S. 1978. Information Theory and Statistics. Phd. Thesis. The George Washington University, Department of Statistics, Washington.
- Lange, K.L., Little, R.J.A. and Taylor, J.M.G. 1989. Robust statistical modeling using the  $t$  distribution. *J. Am. Statist. Assoc.*, 84, 881–96.
- Legendre, A.M. 1805. *Nouvelles Méthodes pour la Détermination des Orbites des Comètes*. Courcier, Paris.
- Li, B., and Yu, Q. 2009. Robust and sparse bridge regression. *Statistics and Its Interface*, 2, 481-491.
- Li, G., Peng, H., and Zhu, L. 2011. Nonconcave Penalized  $m$ -Estimation With a Diverging Number of Parameters. *Statistica Sinica*, 21, 391.
- Li, H. and Wu, L. 2014. Joint modelling of location and scale parameters of the skew-normal distribution. *Appl. Math. J. Chinese Univ.*, 29(3), 265-272.
- Li, H., Wu, L. and Ma, T. 2017. Variable selection in joint location, scale and skewness models of the skew-normal distribution. *Journal of Systems Science and Complexity*, 30(3), 694–709.

- Lin, T.I. and Wang, Y.J. 2009. A robust approach to joint modeling of mean and scale covariance for longitudinal data. *J. Statist Plann Inference*, 139, 3013-3026.
- Machado, J.A. F. 1993. Robust model selection and M-estimation. *Econometric Theory*, 9, 478-493.
- Magnus, J.R. and Neudecker, H. 1999. *Matrix Differential Calculus With Applications in Statistics and Econometrics*. John Wiley and Sons.
- Mallows, C.L. 1964. Choosing variables in a linear regression : A graphical aid, presented at the Central Regional Meeting of the Institute of Mathematical Statistics, Manhattan, Kansas.
- Mallows, C.L. 1973. Some Comments on Cp. *Technometrics*, 15, 661-675.
- Mallows, C.L. 1995. More Comments on Cp. *Technometrics*, 37, 362-372.
- Maronna R.A., Martin, R.D. and Yohai, V.J. 2006. *Robust Statistics: Theory, and Methods*, Wiley.
- Maronna, R.A. 2011. Robust Ridge Regression for High-Dimensional Data. *Technometrics*, 53(1), 44-53.
- Matoušek, J., Mount, D.M. and Netanyahu, N.S. 1998. Efficient randomized algorithms for the repeated median line estimator, *Algorithmica*, 20 (2), 136-150
- McDonald, G.C. and Galarneau, D.I. 1975. A monte carlo evaluation of some ridge type estimators, *Journal of. American Statistics*, 70, 407-416.
- Midi, H. and Zahari M. 2007. A Simulation Study on Ridge Regression Estimators in the Presence of Outliers and Multicollinearity. *Jurnal Teknologi*. 47(C), 59-74.
- Montgomery, D.C., Peck, A.E. and Vining, G.G. 2001. *Introduction to the Linear Regression Analysis*, John Willey&Sons, Canada.
- Muqattash, I. and Yahdi, M. 2006. Infinite family of approximations of the Digamma function. *Math. Comput. Modelling* 43, 1329-1336.
- Müller, S. and Welsh, A. H. 2005. Outlier robust model selection in linear regression. *J. Amer. Statist. Assoc.*, 100, 1297-1310.
- Myung, I.J. 2000. The Importance of Complexity in Model Selection. *Journal of Mathematical Psychology*, 44, 190-204.
- Myung, I.J. and Navarro, D.J. 2004. Information Matrix. In press in B, Everitt & D. Howel (eds.), *Encyclopedia of Behavioral Statistics*. Wiley.
- Nelson, L.S. 1964. The sum of values from a normal and a truncated normal distribution. *Technometrics*, 6, 469-471.

- Nishii, R. 1984. Asymptotic properties of criteria for selection of variables in multiple regression. *Ann. Statist.*, 12, 758-765.
- O'Hagan, A. and Leonard, T. 1976. Bayes estimation subject to uncertainty about parameter constraints. *Biometrika*, 63, 201-202.
- Owen, A.B. 2007. A Robust Hybrid of Lasso and Ridge Regression. *Contemporary Mathematics*, 443, 59-72.
- Park, R.E. 1966. Estimation with heteroscedastic error terms, *Econometrica*, 1966, 34: 888.
- Rao, C.R. 1945. Information and accuracy attainable in the estimation of statistical parameters. *Bull. Calcutta Math. Soc.*, 37, 81.
- Rao, C.R. 1947. Minimum variance and the estimation of several parameters. *Proc. Cam. Phil. Soc.*, 43, 280.
- Rao, C.R. 1948. Sufficient statistics and minimum variance estimates. *Proc. Cam. Phil. Soc.*, 45, 213.
- Rao, C.R., Toutenburg, H., Shalabh, Heumann, C. and Schomaker, M. 2008. *Linear Models and Generalizations: Least Squares and Alternatives*, Springer Verlag, 487, Berlin.
- Roberts, C. 1966. A correlation model useful in the study of twins. *Journal of the American Statistical Association*, 61, 1184-1190.
- Ronchetti, E. 1982. Robust testing in linear models: The infinitesimal approach. *Doktora Tezi*, ETH Zürich, Switzerland.
- Ronchetti, E. 1985. Robust model selection in regression. *Statistics and Probability Letters*, 3, 21-23.
- Ronchetti, E. and Staudte, R.G. 1994. A Robust Version of Mallows's Cp. *J. Amer. Statist. Assoc.*, 89, 550-559.
- Rousseeuw, P. and Yohai, V. 1984. Robust Regression by Means of S-estimators. *Robust and nonlinear time series analysis*, 256-272.
- Rousseeuw, P.J. and Leroy, A.M. 1987. *Robust Regression and Outlier Detection*. John Wiley, 313, New York.
- Sahu, S.K., Dey, D.K. and Branco, M.D. 2003. A new class of multivariate skew distributions with applications to Bayesian regression models. *Can. J. Statist.*, 31, 129-150.
- Schwarz, G.D. 1978. Estimating the Dimension of a Model. *Annals of Statistics*, 6, 461-464.
- Shannon, C.E. 1948. A Mathematical Theory of Communication. *The Bell System Technical Journal*, 27, 379-423, 623-656.

- Shao, J. 2003. *Mathematical Statistics* (Second edition). Springer, 592, New York.
- Shi, P. and Tsai, C.L. 1998. A note on the unification of the Akaike information criterion. *J. R. Statist. Soc. B*, 60(3), 551-558.
- Shibata, R. 1980. Asymptotic efficient selection of the order of the model for estimating parameters of a linear process. *Ann. Statist.*, 8, 147-164.
- Staudte, R.G. and Sheather, S.J. 1990. *Robust Estimation and Testing*. John Wiley and Sons, New York.
- Sugiura, N. 1978. Further analysis of the data by Akaike's information criterion and the finite corrections. *Comm. Statist. A*, 7, 13-26.
- Taylor, J.T. and Verbyla, A.P. 2004. Joint modelling of location and scale parameters of the t distribution. *Stat Model*, 4, 91-112.
- Tharmaratnam, K. and Claeskens, G. 2013. A comparison of robust versions of the AIC based on M-, S- and MM-estimators, *Statistics*, 47(1), 216-235.
- Theodossiou, P. 1998. Financial Data and the Skewed Generalized T Distribution, *Management Science*, 44(12), 1650-1661.
- Thomas, J.A. 2001. *Elements of Information Theory*, John Wiley & Sons, Inc., Publication, 792, USA.
- Tibshirani, R. 1996. Regression shrinkage and selection via the lasso. *J R Stat Soc Ser B*, 58, 267-288.
- Tukey, J.W. 1960. A survey of sampling from contaminated distributions. *Contributions to Probability and Statistics*, edited by I. Olkin, Stanford University Press, 511, Stanford.
- Tukey, J.W. 1962. The Future of Data Analysis, *The Annals of Mathematical Statistics*, 33(1), 1-67.
- Van Emden, M.H. 1971. *An analysis of Complexity*. Mathematical Centre Tracts, 35, Amsterdam.
- Wang, J., Boyer, J. and Genton, M.G. 2004. A skew-symmetric representation of multivariate distributions. *Scandinavian Journal of Statistics*, 31(3), 459-468.
- Wang, H., Li, G., and Jiang, G. 2007. Robust Regression Shrinkage and Consistent Variable Selection Through the Lad-Lasso, *Journal of Business & Economic Statistics*, 25, 347-355.
- Weinstein, M.A. 1964. The sum of values from a normal and a truncated normal distribution. *Technometrics*, 6, 104-105.
- Weisberg, S. 1985. *Applied Linear Regression*. Wiley, New York.

- Wu, L. and Li, H. 2012. Variable selection for joint mean and dispersion models of the inverse Gaussian distribution. *Metrika*, 75(6), 795–808.
- Wu, L.C., MA, T. and Dai, L. 2013a. Maximum Likelihood Estimation for Joint Location and Scale Models of the Skew-t-normal Distribution. *Mathematica Applicata*, 3, 033.
- Wu, L.C., MA, T. and Zhan, J. 2013b. Maximum likelihood estimation for joint location, scale and skewness models of the StN distribution. *Applied Mathematics A Journal of Chinese Universities (Ser. A)*, 4, 006.
- Wu, L.C., Zhang, Z.Z. and Xu, D.K. 2013c. Variable selection in joint location and scale models of the skew-normal distribution, *Journal of Statistical Computation and Simulation*, 2013, 83, 1266–1278.
- Wu, L. 2014. Variable Selection in Joint Location and Scale Models of the Skew-t-Normal Distribution. *Communications in Statistics - Simulation and Computation*, 43, 615–630.
- Xu, D. and Zhang, Z. 2011. Regularized REML for estimation in heteroscedastic regression models. In *Nonlinear Mathematics for Uncertainty and its Applications* (495-502). Springer, Berlin, Heidelberg.
- Yohai, V.J. 1987. High breakdown point and high efficiency robust estimates for regression. *The Annals of Statistics*, 15(20), 642-656.
- Zhang, Z.Z. and Wang, D.R. 2011. Simultaneous variable selection for heteroscedastic regression models. *Science China Mathematics*, 54, 515–530.
- Zou, H. and Hastie, T. 2005. Regularization and variable selection via the elastic net. *J. R. Stat. Soc. Ser B*, 67, 301–320.

## ÖZGEÇMİŞ

Adı Soyadı : Yeşim GÜNEY  
Doğum Yeri : Ankara  
Doğum Tarihi : 28/08/1986  
Medeni Hali : Evli  
Yabancı Dili : İngilizce

### Eğitim Durumu (Kurum ve Yıl)

Lise : Ankara Atatürk Anadolu Lisesi (2004)  
Lisans : Ankara Üniversitesi, Fen Fakültesi, Matematik Bölümü (2009)  
Yüksek Lisans : Ankara Üniversitesi, Fen Bilimleri Enstitüsü, Matematik ABD (2011)

### Çalıştığı Kurum/Kurumlar ve Yıl

Ankara Üniversitesi, Fen Fakültesi, İstatistik Bölümü (2011-Devam Ediyor)

### SCI-Exp

**Güney, Y.**, Tuuç, Y. and Arslan, O. 2017. Marshall--Olkin distribution: parameter estimation and application to cancer data. Journal of Applied Statistics, 44 (12), 2238-2250.

Tuaç, Y., **Güney, Y.**, Şenoğlu, B. and Arslan, O. 2017. Robust Parameter Estimation of Regression Model with AR(p) Error Terms. Communications in Statistics - Simulation and Computation. <https://doi.org/10.1080/03610918.2017.1343839>.

### ESCI

**Güney, Y.** and Arslan, O. 2017. Robust Parameter Estimation for the Marshall-Olkin Extended Burr XII Distribution. Commun. Fac. Sci. Univ. Ank. Series A1, 66(2), 141-161.

Özdemir, Ş., **Güney, Y.**, Tuuç, Y. and Arslan, O. 2018. Maximum Lq-Likelihood Estimation for the Parameters of Marshall-Olkin Extended Burr XII Distribution. Commun. Fac. Sci. Univ. Ank. Series A1, 68(1), 17-34.

### Hakemli Dergiler

Sucu, S., **Güney (Döne), Y.** and İbikli, E. 2014. Approximation Properties of Schurer-Stancu Type Polynomials. Mathematica Balkanica, 28 (1-2), 37-48.

### Uluslararası Kongre

**Güney Y.**, Tuuç Y and Arslan O. "Marshall-Olkin Distribution : Parameter Estimation and Applications". International Statistics Days of Symposium, Side/Antalya, 10-14 Mayıs 2014.

Tuaç Y., **Güney Y.** and Arslan O. "Variable Selection in Regression Models with AR(p) Error Terms Based on Skew Distributions". European Meeting of Statisticians (EMS 2015), VU University Amsterdam, 6-10 Temmuz 2015.

**Güney Y.**, Bozdoğan, H. and Arslan O. “Variable Selection in Single-Index Models Using Information Complexity (ICOMP) Criterion” International Conference on Information Complexity and Statistical Modeling in High Dimensions with Applications (IC-SMHD-2016) Kapadokya/Nevşehir, 18-21 Mayıs 2016.

Özdemir Ş., **Güney Y.** and Arslan O. “Variable Selection in Empirical Likelihood based Regression with Information Complexity (ICOMP)” International Conference on Information Complexity and Statistical Modeling in High Dimensions with Applications (IC-SMHD-2016), Kapadokya /Nevşehir, 18-21 Mayıs 2016.

Tuaç Y., **Güney Y.**, Özdemir Ş. and Arslan O. Parameter estimation for skew normal distribution via the maximum Lq-likelihood estimation. IRSYSC 2017 – 3RD International Researchers, Statisticians and Young Statisticians Congress, Konya, 24-26 Mayıs 2017.

**Güney Y.**, Özdemir Ş., Tuaç Y. and Arslan O. Optimal B-robust estimators for the parameters of the MOEBXII distribution. IRSYSC 2017 – 3RD International Researchers, Statisticians and Young Statisticians Congress, Konya, 24-26 Mayıs 2017.

Özdemir Ş., Tuaç Y., **Güney Y.**, and Arslan O. Maximum Lq-Likelihood Estimation for the Marshall-Olkin Extended Burr XII Distribution. IRSYSC 2017 – 3RD International Researchers, Statisticians and Young Statisticians Congress, Konya, 24-26 Mayıs 2017.

**Güney Y.**, Tuaç Y., Özdemir Ş. and Arslan O. Maximum Lq-likelihood estimation for autoregressive error term regression model. 31st EMS (The European Meeting of Statisticians) Helsinki, 24-28 Temmuz 2017.

Tuaç Y., **Güney Y.**, Özdemir Ş. and Arslan O. Robust variable selection in an autoregressive error term regression model. 31st EMS (The European Meeting of Statisticians) Helsinki, 24-28 Temmuz 2017.

Özdemir Ş., **Güney Y.**, Tuaç Y. and Arslan O. Empirical likelihood estimation for linear regression models with AR(p) error terms. 31st EMS (The European Meeting of Statisticians) Helsinki, 24-28 Temmuz 2017.

**Güney Y.**, Tuaç Y. and Arslan O. Robust Estimation for the Parameter of the Zipfian Distribution Properties and Applications. International Conference on Robust Statistics ICORS (2014), (Poster Presentation).

**Güney Y.**, Arslan O. and Bozdoğan H. Model Selection in Linear Regression Models with Skew Normal Distributed Errors Using ICOMP. European Meeting of Statisticians (EMS 2015), 2015, (Poster Presentation).