

**UKUROVA NİVERSİTESİ  
FEN BİLİMLERİ ENSTİTÜSÜ**

**DOKTORA TEZİ**

**Fela ZBEY**

**OKLU İ İLİŐKİ OLMASI DURUMUNDA LİNEER REGRESYON  
MODELLERİNİN KULLANIMI İLE KESTİRİM VE NGÖRÜ**

**İSTATİSTİK ANABİLİM DALI**

**ADANA, 2012**

**ÇUKUROVA ÜNİVERSİTESİ  
FEN BİLİMLERİ ENSTİTÜSÜ**

**ÇOKLU İÇ İLİŞKİ OLMASI DURUMUNDA LİNEER REGRESYON  
MODELLERİNİN KULLANIMI İLE KESTİRİM VE ÖNGÖRÜ**

**Fela ÖZBEY**

**DOKTORA TEZİ**

**İSTATİSTİK ANABİLİM DALI**

Bu Tez .../.../.... Tarihinde Aşağıdaki Jüri Üyeleri Tarafından  
Oybirliği/Oyçokluğu ile Kabul Edilmiştir.

.....  
Prof. Dr. Selahattin  
KAÇIRANLAR  
DANIŞMAN

.....  
Prof. Dr. Fikri AKDENİZ  
ÜYE

.....  
Prof. Dr. H.Altan ÇABUK  
ÜYE

.....  
Prof. Dr. Hamza EROL  
ÜYE

.....  
Prof. Dr. Sadullah SAKALLIOĞLU  
ÜYE

Bu Tez Enstitümüz İstatistik Anabilim Dalında hazırlanmıştır.

**Kod No:**

**Prof. Dr. Selahattin SERİN  
Enstitü Müdürü**

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

**ÖZ**

**DOKTORA TEZİ**

**ÇOKLU İÇ İLİŞKİ OLMASI DURUMUNDA LİNEER REGRESYON  
MODELLERİNİN KULLANIMI İLE KESTİRİM VE ÖNGÖRÜ**

**Fela ÖZBEY**

**ÇUKUROVA ÜNİVERSİTESİ  
FEN BİLİMLERİ ENSTİTÜSÜ  
İSTATİSTİK ANABİLİM DALI**

Danışman : Prof. Dr. Selahattin KAÇIRANLAR  
Yıl: 2012, Sayfa: 57

Jüri : Prof. Prof. Dr. Selahattin KAÇIRANLAR  
: Prof. Dr. Fikri AKDENİZ  
: Prof. Dr. H.Altan ÇABUK  
: Prof. Dr. Hamza EROL  
: Prof. Dr. Sadullah SAKALLIOĞLU

Lineer regresyon modellerinin tahmininde en küçük kareler (EKK) tahmin edicisi yaygın olarak kullanılmaktadır. Ancak, açıklayıcı değişkenlerin birbirleriyle ilişkili olduğu durumlarda EKK tahmin edicisi istikrarsızlaşır. Bu nedenle, çoklu iç ilişkinin varlığı durumunda EKK tahmin edicisine alternatif olarak yanlı tahmin ediciler önerilmektedir. Birçok alanda, gelecek zamana ait verileri kestirmek/öngörmek büyük önem taşır çünkü öngörü, gelecekteki potansiyel olaylar ve onların sonuçları hakkında belli bilgiler ortaya koymaktadır. Bu da politika belirleyicinin (veya yöneticinin) önemli kararları daha güvenli bir şekilde almasını sağlamaktadır. Yanıt değişkenin bilinmeyen değerlerinin kestirimi ile ilgilendiğimizde, regresyon modelinin uygun bir kestirim denklemi üretebilmesi öncelikli gereksinimdir. Bu nedenle, bu çalışmada, çoklu iç ilişkinin mevcut olduğu durumlarda bazı yanlı tahmin edicilerin kestirim/öngörü performanslarını iyileştirecek yöntemler önerilmiş ve bu yöntemlere göre oluşturulmuş kestirim denklemleri, gerçek veriler kullanılarak kendi aralarında karşılaştırılmıştır.

**Anahtar Kelimeler:** Yanlı tahmin edici, Kestirimin Hata Kareleri Ortalaması, PRESS, Kestirim, Öngörü



## ABSTRACT

### PhD THESIS

# PREDICTION AND FORECASTING WITH LINEAR REGRESSION MODELS UNDER MULTICOLLINEARITY

Fela ÖZBEY

ÇUKUROVA UNIVERSITY  
INSTITUTE OF NATURAL AND APPLIED SCIENCES  
DEPARTMENT STATISTICS

Supervisor : Prof. Dr. Selahattin KAÇIRANLAR  
Year: 2012, Pages: 57

Jury : Prof. Dr. Selahattin KAÇIRANLAR  
: Prof. Dr. Fikri AKDENİZ  
: Prof. Dr. H.Altan ÇABUK  
: Prof. Dr. Hamza EROL  
: Prof. Dr. Sadullah SAKALLIOĞLU

The Ordinary Least Squares (OLS) estimator is the widely used technique for estimating linear regression models. But the OLS estimator can be highly variable in certain directions, especially when the explanatory variables are collinear. Therefore, in the presence of multicollinearity, biased estimation techniques are often suggested as alternatives to the OLS. In many areas, the prediction/forecasting of the future values is very important because, forecasting provides information about the potential future events and their consequences. Thus, it increases the confidence of the policy maker (or the manager) to make important decisions. When a multiple linear regression model is used in predicting/forecasting unknown values of the response variable, its ability to produce an adequate prediction equation is of prime importance. In this study, some techniques are suggested to improve the prediction/forecasting performances of alternative biased estimators. Prediction equations formed by using these techniques are compared on real data sets.

**Keywords:** Biased Estimator, Prediction Mean Squared Error, PRESS, Prediction, Forecasting



## TEŞEKKÜR

Beni İstatistik Doktora eğitimi almaya yönlendiren Sayın Prof. Dr. Bilal VATANSEVER'e desteği ve yardımlarından dolayı gönülden teşekkür ederim.

Doktora eğitimim boyunca bana çok şey kazandıran Sayın Hocalarım Prof. Dr. Sadullah SAKALLIOĞLU, Prof. Dr. Hamza EROL ve Prof. Dr. Olcay ARSLAN'a çok teşekkür ederim. Bana büyük yardımları ve katkıları olmuştur.

Doktora tez danışmanım Sayın Prof. Dr. Selahattin KAÇIRANLAR'a bana verdiği emeklerinden dolayı çok teşekkür ederim.

Her başım sıkıştığında kapısını çaldığım ve yardımlarını benden esirgemeyen değerli arkadaşım Öğr. Gör. Dr. Hüseyin GÜLER'e, zor günlerimde bana destek olan Sayın Hocalarım Yrd. Doç. Dr. Ebru ÖZGÜR GÜLER ve Doç. Dr. Mehmet ÖZMEN'e sonsuz şükranlarımı sunarım.

Her zaman büyük bir sevgi ve minnetle anacağım Sayın Prof. Dr. Fikri AKDENİZ ve Sayın Prof. Dr. Hasan Altan ÇABUK'a bilimsel ve manevi desteklerinden dolayı gönülden teşekkür ederim.

En zor anlarıma şahit olan ve beni teselli eden çok değerli iki insana sevgili Senay AKALIN ve sevgili Hale YAŞAR'a çok teşekkür ederim.

Hayatım boyunca bana hep destek olan annem, babam ve kardeşime sonsuz sevgilerimi ve şükranlarımı sunarım.





## İÇİNDEKİLER

## SAYFA

|                                                                                                                                                            |      |
|------------------------------------------------------------------------------------------------------------------------------------------------------------|------|
| ÖZ .....                                                                                                                                                   | I    |
| ABSTRACT.....                                                                                                                                              | III  |
| TEŞEKKÜR.....                                                                                                                                              | V    |
| İÇİNDEKİLER .....                                                                                                                                          | VII  |
| TABLolar DİZİNİ .....                                                                                                                                      | IX   |
| ŞEKİLLER DİZİNİ .....                                                                                                                                      | XI   |
| SİMGELER VE KISALTMALAR.....                                                                                                                               | XIII |
| 1.GİRİŞ .....                                                                                                                                              | 1    |
| 2. LİU TAHMİN EDİCİNİN KESTİRİM PERFORMANSININ<br>DEĞERLENDİRİLMESİ .....                                                                                  | 4    |
| 2.1. Giriş.....                                                                                                                                            | 4    |
| 2.2. Kestirim Hata Kareleri Ortalamalarının Elde Edilişi.....                                                                                              | 7    |
| 2.3. Kestirim Hata Kareleri Ortalamalarının İki Boyutlu Uzayda Karşılaştırılması                                                                           | 11   |
| 2.4. Sayısal Örnek .....                                                                                                                                   | 24   |
| 2.5. Sonuç.....                                                                                                                                            | 27   |
| 3. R-K VE R-D SINIF TAHMİN EDİCİLERİN KESTİRİM<br>PERFORMANSLARININ İYİLEŞTİRİLMESİ .....                                                                  | 29   |
| 3.1. Giriş.....                                                                                                                                            | 29   |
| 3.2. Tahmin Ediciler ve PRESS İstatistikleri.....                                                                                                          | 30   |
| 3.3. Sayısal Örnek .....                                                                                                                                   | 33   |
| 3.3.1. Mil Başına Benzin Tüketimi Veri Seti.....                                                                                                           | 33   |
| 3.4. Sonuç.....                                                                                                                                            | 39   |
| 4. LİU TİPİ VE İKİ PARAMETRELİ RIDGE TAHMİN EDİCİLERİNİN<br>KESTİRİM PERFORMANSLARININ PRESS VE GCV İSTATİSTİKLERİ<br>KULLANILARAK KARŞILAŞTIRILMASI ..... | 41   |
| 4.1. Giriş.....                                                                                                                                            | 41   |
| 4.2. Tahmin Ediciler ve Kestirim İyiliği İstatistikleri.....                                                                                               | 42   |
| 4.3. Sayısal Örnek .....                                                                                                                                   | 45   |
| 4.3.1. Mil Başına Benzin Tüketimi Veri Seti.....                                                                                                           | 45   |

|                                        |    |
|----------------------------------------|----|
| 4.3.2. Portland Çimento Veri Seti..... | 47 |
| 4.4. Sonuç.....                        | 49 |
| 5. SONUÇLAR VE ÖNERİLER.....           | 50 |
| KAYNAKÇA.....                          | 53 |
| ÖZGEÇMİŞ .....                         | 57 |

## TABLÖLAR DİZİNİ

|                                                                                                                                                                                                                                                                                                                    |    |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| Tablo 3.1. Mil Başına Benzin Tüketimi verisi.....                                                                                                                                                                                                                                                                  | 34 |
| Tablo 3.2a. Mil Başına Benzin Tüketimi veri seti için Ridge, r-k sınıf, EKK ve TB tahmin edicileri ile tahmin edilen regresyon modeli için parametre tahminleri, $R^2$ , HKT, HKO( $\hat{\beta}$ ) ve PRESS istatistikleri ...                                                                                     | 35 |
| Tablo 3.2b. Mil Başına Benzin Tüketimi verisi için Liu, r-d sınıf, EKK ve TB tahmin edicileri ile tahmin edilen regresyon modeli için parametre tahminleri, $R^2$ , HKT, HKO( $\hat{\beta}$ ) ve PRESS istatistikleri .....                                                                                        | 36 |
| Tablo 3.3. Mil Başına Benzin Tüketimi verisi için EKK, TB ve yanlılık parametreleri PRESS istatistikleri minimize edilerek seçilen Ridge, r-k sınıf, Liu, r-d sınıf tahmin edicileri ile tahmin edilen regresyon modeli için parametre tahminleri, $R^2$ , HKT, HKO( $\hat{\beta}$ ) ve PRESS istatistikleri ..... | 37 |
| Tablo 4.1a. Mil Başına Benzin Tüketimi veri seti için EKK, 2 parametrelili Ridge ve Liu tipi tahmin edicileri için parametre tahminleri, HKO ve PRESS istatistikleri .....                                                                                                                                         | 45 |
| Tablo 4.1b. Mil Başına Benzin Tüketimi veri seti için EKK, 2 parametrelili Ridge ve Liu tipi tahmin edicileri için parametre tahminleri, HKO ve GCV istatistikleri.....                                                                                                                                            | 46 |
| Tablo 4.2. Portland Çimento verisi .....                                                                                                                                                                                                                                                                           | 47 |
| Tablo 4.3a. Portland Çimento veri seti için EKK, 2 parametrelili Ridge ve Liu tipi tahmin edicileri için parametre tahminleri, HKO ve PRESS istatistikleri.....                                                                                                                                                    | 48 |
| Tablo 4.3b. Portland Çimento veri seti için EKK, 2 parametrelili Ridge ve Liu tipi tahmin edicileri için parametre tahminleri, HKO ve GCV istatistikleri.....                                                                                                                                                      | 48 |



## ŞEKİLLER DİZİNİ

|                                                                                                                                                                                    |    |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| Şekil 2.1. Liu ve EKK tahmin edicilerin KHKO'larının karşılaştırılması.....                                                                                                        | 25 |
| Şekil 2.2. Liu ve TB tahmin edicilerin KHKO'larının karşılaştırılması.....                                                                                                         | 26 |
| Şekil 2.3. Liu ve Ridge tahmin edicilerin KHKO'larının karşılaştırılması .....                                                                                                     | 27 |
| Şekil 3.1a. Mil Başına Benzin Tüketimi verisi için Ridge, Liu, r-d sınıf ve r-k sınıf tahmin edicilerin yanlılık parametreleri ve onlara karşılık gelen PRESS istatistikleri. .... | 38 |
| Şekil 3.1b. Mil Başına Benzin Tüketimi verisi için Ridge ve r-k sınıf tahmin edicilerin yanlılık parametreleri ve onlara karşılık gelen PRESS istatistikleri.....                  | 38 |
| Şekil 3.1c. Mil Başına Benzin Tüketimi verisi için Liu ve r-d sınıf tahmin edicilerin yanlılık parametreleri ve onlara karşılık gelen PRESS istatistikleri.....                    | 39 |



## SİMGELER VE KISALTMALAR

Regresyon modeli için kullanılan simge ve kısaltmalar

|               |                                                                     |
|---------------|---------------------------------------------------------------------|
| $n$           | : Gözlem sayısı                                                     |
| $p$           | : Parametre sayısı                                                  |
| $y$           | : Yanıt değişken vektörü                                            |
| $X$           | : Açıklayıcı değişkenler matrisi                                    |
| $\beta$       | : Bilinmeyen parametreler vektörü                                   |
| $\varepsilon$ | : Hata terimleri vektörü                                            |
| $Z$           | : Ortogonalleştirilmiş açıklayıcı değişkenler matrisi               |
| $T$           | : $X'X$ matrisinin özvektörlerinden oluşan matris                   |
| $\Lambda$     | Köşegen elemanları $X'X$ matrisinin özdeğerleri olan köşegen matris |

Tahmin ediciler için kullanılan simge ve kısaltmalar

|                     |                                 |
|---------------------|---------------------------------|
| EKK                 | : En Küçük Kareler              |
| TB                  | : Temel Bileşenler              |
| LT                  | : Liu Tipi                      |
| R                   | : Ridge                         |
| 2PR                 | : İki Parametrelili Ridge       |
| $\hat{\beta}_{EKK}$ | : En Küçük Kareler tahmin edici |
| $\hat{\beta}_{TB}$  | : Temel Bileşenler tahmin edici |
| $\hat{\beta}_R$     | : Ridge tahmin edici            |
| $\hat{\beta}_{Liu}$ | : Liu tahmin edici              |
| $\hat{\beta}_{r-k}$ | : r-k sınıf tahmin edici        |
| $\hat{\beta}_{r-d}$ | : r-d sınıf tahmin edici        |
| $\hat{\beta}_{LT}$  | : Liu Tipi tahmin edici         |

$\hat{\beta}_{2PR}$  : İki Parametrel Ridge tahmin edici

Matrisler için kullanılan simge ve kısaltmalar

$\mathbf{I}_r$  :  $r \times r$  boyutlu birim matris

$\mathbf{j}$  : birler vektörü

Tahmin edicilerin istatistiksel özellikleri için kullanılan simge ve kısaltmalar

HKO : Hata kareleri ortalaması

KHKO : Kestirimin hata kareleri ortalaması

PRESS : Kestirimin hata kareleri toplamı (Prediction error sum of squares)

GCV : Genelleştirilmiş çapraz geçerlilik (Generalized cross validation)

$iid(0, s^2)$  : Sıfır ortalamalı,  $s^2$  varyanslı, bağımsız ve özdeş dağılmış (Independent and identically distributed)

$E(.)$  : Beklenen değer operatörü

$J$  : Kestirimin hata kareleri ortalaması

$B$  : Yanlılığın karesi

$V$  : Varyans

$k$  : Koşul sayısı



## 1.GİRİŞ

Ampirik çalışmalarda, değişkenler arasındaki ilişkiyi tanımlamak için en sık kullanılan yöntem çoklu doğrusal regresyon modelidir. Bunun başlıca sebepleri basit yapıda olması ve çoğu zaman, bu basit yapıya rağmen, gerçek dünyadaki verileri modellemede oldukça başarılı olmasıdır.

Çoklu doğrusal regresyon modelleri iki temel amaç için kullanılmaktadır. Bunlardan ilki modelin eldeki verilere uyumunu optimal şekilde sağlamak (tahmin); diğeri ise bu modeli yanıt değişkeninin henüz gerçekleşmemiş veya gözlemlenememiş değerlerine adapte etmektir (kestirim).

Kestirim, eldeki verileri kullanarak yanıt değişkeninin gözlemlenememiş değerlerini tahmin etme işlemi olarak tanımlanır. Zaman serisi verilerinin analizlerinde kestirilmeye çalışılan veriler gelecek ile ilgili olduklarında, yapılan kestirim özel olarak öngörü diye adlandırılır. Yani öngörü, geçmiş ve şimdiki zamana ait gözlemler kullanılarak gelecek ile ilgili verilerin kestirilmesidir.

Birçok alanda, gelecek zamana ait verileri kestirmek/öngörmek büyük önem taşır çünkü öngörü, gelecekteki potansiyel olaylar ve onların sonuçları hakkında belli bilgiler ortaya koymaktadır. Her ne kadar öngörü gelecekte gerçekleşecek olayların karmaşıklığını ve belirsizliğini azaltmasa da, politika belirleyicinin (veya yöneticinin) önemli kararları daha güvenli bir şekilde almasını sağlamaktadır.

Yanıt değişkenin bilinmeyen değerlerinin kestirimi ile ilgilendiğimizde, regresyon modelinin uygun bir kestirim denklemi üretebilmesi öncelikli gereksinimdir. Bu çalışmanın amacı, çoklu iç ilişkinin mevcut olduğu durumlarda bazı tahmin edicilerin kestirim/öngörü performanslarının değerlendirilmesidir.

En Küçük Kareler (EKK) tahmin edicisi lineer regresyon modellerinin tahmininde oldukça yaygın olarak kullanılmaktadır. EKK tahmin edicinin kullanımına yönelim, çoğunlukla yansız oluşundan ve yansız tahmin ediciler arasında en küçük varyansa sahip oluşundandır. Ancak, EKK tahmin edicisi, özellikle açıklayıcı değişkenler arasında doğrusal bir ilişki var ise, belli yönlerde oldukça değişken olabilir. Aslında, yansızlık gerekli temel bir özellik değildir. Genel olarak tek bir örnekleme sahibiz ve verilen bir örneklem büyüklüğü için, bir tahmin

ediciinin kalitesi onun Hata Kareleri Ortalaması (HKO) ile ölçülür. Normalde yansız olan bir tahmin ediciye bir miktar yanlılık eklenmesi, onun varyansında kayda değer bir azalmaya neden olacaktır; öyle ki, HKO'su küçülecektir. Bu nedenle çoklu iç ilişkinin varlığı durumunda, çoğunlukla yanlı tahmin yöntemleri EKK'ya alternatif olarak önerilmektedir.

Çoklu iç ilişki sorunu ile baş etmek için kullanılan en eski tekniklerden birisi Temel Bileşenler (TB) regresyonudur (Massy, 1965). Bu yöntem, açıklayıcı değişkenlerin kolon uzayının bir alt kümesini seçip kullanmaya ve yanıt değişkenin bu alt küme üzerine izdüşümüne dayanmaktadır. Bir diğer popüler örnek Hoerl ve Kennard (1970) tarafından önerilen Ridge tahmin edicidir. Ridge tahmin edici, yansız bir tahmin ediciye, parametre tahminlerinin ve kestirimlerinin HKO'sunu küçültmek amacıyla yanlılık eklenilmesinin en bilinen örneğidir. Liu (1993) tarafından önerilen yanlı tahmin edici, Ridge tahmin edicisinin özel bir haline Stein tahmin edicinin (Stein tahmin edici için bkz: Stein, 1956) eklenmesi ile elde edilmiş bir modifikasyondur. Bu tahmin edicinin, HKO'sunun yanlılık parametresinin doğrusal bir fonksiyonu olması nedeniyle, yanlılık parametresinin seçimi daha kolaydır; bu bakımdan da Ridge tahmin edicisine göre üstündür. Bu çalışmanın ikinci bölümünde, Liu tahmin edicinin kestirim performansı, EKK, TB ve Ridge tahmin edici ile karşılaştırılmıştır.

Baye ve Parker (1984), yukarıda adı geçen TB ve Ridge tahmin edicilerini birleştirerek, r-k sınıf tahmin edicisini oluşturmuşlardır. Bu tahmini edici, EKK, TB ve Ridge tahmin edicilerini ihtiva eden daha geniş kapsamlı bir sınıfı oluşturmaktadır. Liu tahmin edicinin Ridge tahmin ediciye üstünlüğünü göz önünde bulundurarak Kaçıranlar ve Sakallıoğlu (2001) TB tahmin ediciyi Liu tahmin edici ile birleştirerek r-k sınıf tahmin ediciye alternatif olarak r-d sınıf tahmin ediciyi önermişlerdir. Bu tahmin edici de EKK, TB ve Liu tahmin edicilerini ihtiva etmektedir. Üçüncü bölümde bu sınıf tahmin edicilerin kestirim performansları kendi aralarında karşılaştırılırken, ihtiva ettikleri tahmin edicilere üstünlükleri de incelenmiştir.

Liu (2003), Liu tahmin edicisinin performansını arttırmak için ona ikinci bir yanlılık parametresi ekleyerek Liu Tipi (LT) tahmin ediciyi oluşturmuştur. Bu tahmin

edicide, yanlılık parametrelerinden biri tasarım matrisinin kötü koşulluluğunu ortadan kaldırma amaçlı kullanılırken, diğer yanlılık parametresi de modelin uyum iyiliğini arttırmaya hizmet etmektedir. Benzer bir amaç doğrultusunda, Lipovetsky ve Conklin (2005) Ridge tahmin ediciye uyumun iyiliğini arttıracak ikinci bir parametre ekleyerek İki Parametrelili Ridge (2PR) tahmin edicisini tanımlamışlardır. Dördüncü bölümde bu iki tahmin edicinin kestirim performansları değerlendirilmiştir.

## **2. LİU TAHMİN EDİCİNİN KESTİRİM PERFORMANSININ** **DEĞERLENDİRİLMESİ**

### **2.1. Giriş**

Çoklu doğrusal regresyon modeli

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (2.1)$$

olarak tanımlanır. Burada  $\mathbf{y}$ ,  $n \times 1$  boyutlu yanıt değişkenler vektörü,  $\mathbf{X}$ ,  $n \times p$  boyutlu stokastik olmayan, önceden belirlenmiş, açıklayıcı değişkenlerin tam kolon ranklı matrisi,  $\boldsymbol{\beta}$ ,  $p \times 1$  boyutlu bilinmeyen parametreler vektörü ve  $\boldsymbol{\varepsilon}$ ,  $n \times 1$  boyutlu,  $iid(0, s^2)$  dağılımlı rassal hatalardır. Değişkenlerin,  $\mathbf{X}'\mathbf{X}$  matrisinin açıklayıcı değişkenler arasındaki korelasyonların matrisi ve  $\mathbf{X}'\mathbf{y}$  vektörünün her bir açıklayıcı değişken ve yanıt değişken arasındaki korelasyonların vektörü olacak şekilde merkezileştirilmesi ve ölçeklendirilmesi önerilmektedir. Bu nedenle bu çalışmada, değişkenlerin bu şekilde standartlaştırıldığı varsayılacaktır.

(2.1) ile verilen eşitlikteki  $\boldsymbol{\beta}$  'nın EKK tahmin edicisi

$$\hat{\boldsymbol{\beta}}_{EKK} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y} \quad (2.2)$$

dir. EKK tahmin edicisi yansızdır ve tüm yansız tahmin ediciler arasında en küçük varyansa sahiptir, ancak, belli yönlerde oldukça değişken olabilmektedir. Klasik doğrusal regresyon modelinin varsayımlarından biri açıklayıcı değişkenler arasında doğrusal bir ilişki olmadığıdır. Ne yazık ki, uygulamada, çoğu açıklayıcı değişken arasında yüksek seviyede doğrusal ilişki vardır. Çoklu iç ilişkinin varlığı durumunda, parametrelerin EKK tahminleri büyük standart hatalara sahip olurlar, bu nedenle de

EKK kullanılarak, katsayılar büyük bir kesinlik veya doğrulukla tahmin edilemez. Bir tahmin ediciden gerçekte beklenen küçük bir HKO değerine sahip olmasıdır ve bu, tahmin edicinin yansızlığını gerektirmez. Yansızlık koşulunu gevşetmek, çoklu iç ilişkinin varlığı durumunda daha iyi özelliklere sahip (oldukça küçük HKO gibi) yanlı tahmin edicileri tartışmaya açmaktadır.

Çoklu iç ilişki sorunu ile baş etmek için kullanılan en eski tekniklerden birisi TB regresyonudur (Massy, 1965). Bu yöntem,  $\mathbf{X}$  'in kolon uzayının bir alt kümesini seçip kullanmaya ve  $\mathbf{y}$  yanıt değişkeninin bu alt küme üzerine izdüşümüne dayanmaktadır. (2.1)'deki bilinmeyen parametrelerin TB tahmin edicisi

$$\hat{\beta}_{TB} = \mathbf{T}_r (\mathbf{T}_r' \mathbf{X}' \mathbf{X} \mathbf{T}_r)^{-1} \mathbf{T}_r' \mathbf{X}' \mathbf{y} = \mathbf{T}_r \mathbf{\Lambda}_r^{-1} \mathbf{T}_r' \mathbf{X}' \mathbf{y} \quad (2.3)$$

şeklinde tanımlanır. Burada,  $\mathbf{T} = [\mathbf{t}_1, \mathbf{t}_2, \mathbf{K}, \mathbf{t}_p]$  ,  $\mathbf{T}' \mathbf{X}' \mathbf{X} \mathbf{T} = \mathbf{\Lambda}$  olacak şekilde ortogonal bir matristir. Yani  $\mathbf{T}$  ,  $\mathbf{X}' \mathbf{X}$  matrisinin  $p \times p$  boyutlu öz vektörleri matrisidir.  $\mathbf{\Lambda} = \text{diag}(l_1, l_2, \mathbf{K}, l_p)$  ,  $\mathbf{X}' \mathbf{X}$  matrisinin öz değerleri matrisidir ve öz değerler azalan sıradadır.  $\mathbf{T}_r$  matrisi,  $\mathbf{T}$  matrisinin  $p-r$  tane kolonu silindikten sonra kalan kısmı ve  $\mathbf{\Lambda}_r = \text{diag}(l_1, l_2, \mathbf{K}, l_r)$  dir.

Hoerl ve Kennard (1970) tarafından önerilen Ridge tahmin edici, yansız bir tahmin ediciye, parametre tahminlerinin ve kestirimlerinin HKO'sunu küçültmek amacıyla yanlılık eklenilmesinin en bilinen örneğidir. Bu tahmin edici

$$\hat{\beta}_R = (\mathbf{X}' \mathbf{X} + k \mathbf{I})^{-1} \mathbf{X}' \mathbf{y}, \quad k \geq 0, \quad (2.4)$$

olarak tanımlanmıştır. Burada  $k$  , EKK tahminleri ile ortaya çıkan zorlukları önlemeye yardımcı olmak için kullanılan yanlılık parametresidir. Böylece,  $\mathbf{X}' \mathbf{X}$  matrisi birim matrise yakın bir matris olmadığı durumlarda,  $\mathbf{X}' \mathbf{X}$  matrisinin köşegen elemanlarına  $k$  eklenerek EKK tahminlerinin duyarlılığının azaltılması hedeflenmiştir.

Liu (1993) tarafından önerilen yanlı tahmin edici, Ridge tahmin edicisinde  $k=1$  iken  $\mathbf{X}'\mathbf{y}$  vektörüne Stein tahmin edicinin (Stein tahmin edici için bkz: Stein, 1956) eklenmesi ile elde edilmiş bir modifikasyondur:

$$\begin{aligned}\hat{\boldsymbol{\beta}}_{Liu} &= (\mathbf{X}'\mathbf{X} + \mathbf{I})^{-1}(\mathbf{X}'\mathbf{y} + d\hat{\boldsymbol{\beta}}_{EKK}) \\ &= (\mathbf{X}'\mathbf{X} + \mathbf{I})^{-1}(\mathbf{X}'\mathbf{X} + d\mathbf{I})\hat{\boldsymbol{\beta}}_{EKK}, \quad 0 < d < 1;\end{aligned}\tag{2.5}$$

$d$  yanlılık parametresidir.  $d=1$  iken  $\hat{\boldsymbol{\beta}}_{Liu} = \hat{\boldsymbol{\beta}}_{EKK}$ . Bu tahmin edicinin, HKO bakımından EKK tahmin edicisinden daha iyi olduğu ve  $d$ 'nin doğrusal bir fonksiyonu olması nedeniyle, yanlılık parametresinin seçimi daha kolay olduğu, bu bakımdan da Ridge tahmin edicisine göre üstün olduğu Liu (1993) tarafından gösterilmiştir. Bu tahmin edici, Akdeniz ve Kaçıranlar (1995) ve Gruber (1998, s.43) tarafından Liu tahmin edici olarak adlandırılmıştır. Sakallıoğlu ve ark. (2001), Liu tahmin ediciyi MSE bakımından Ridge ve İteratif tahmin edicileriyle karşılaştırmıştır.

Yanlı tahmin ediciler, ortalama olarak, tahmin edilen parametreye EKK tahmin edicisinden daha yakındır. Çünkü yansız bir tahmin ediciye bir miktar yanlılık eklenmesi, onun varyansını kayda değer bir şekilde azaltıp HKO'yu küçültmektedir.

Bir tahmin edicinin HKO'su ( $\text{HKO}(\hat{\boldsymbol{\beta}})$ ), bu tahmin edicinin, tahmin edilecek parametreye ortalama uzaklığıdır.

$$\text{HKO}(\hat{\boldsymbol{\beta}}) = E(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^2.\tag{2.6}$$

Açıktır ki, tahmin edici yansız olduğunda, yani  $E(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$  olduğunda, bu tahmin edicinin HKO'su varyansına eşittir. Tahmin edici yanlı olduğunda, bu tahmin edicinin HKO'su yanlılığın karesi ile varyansının toplamına eşittir.

Çoklu doğrusal regresyon modelleri sıklıkla  $\mathbf{y}$  yanıt değişkeninin bilinmeyen değerlerini kestirmek için kullanılmaktadır. Gözlemlenen örneklemdeki korelasyon

yapısı, kestirilecek kitlede de aynı kalması durumunda, çoklu iç ilişki EKK'nın kestirim gücünü azaltmaz (nokta tahmini bakımından); ancak, gözlenen korelasyon yapısının istikrarlı olmadığı veya kestirilecek kitle için aynı kalmadığı bir sistemin kestiriminde oldukça yanıltıcı sonuçlar vermektedir. Bu gibi durumlarda, bir regresyon modelinin uygun kestirim denklemini üretebilmesi temel önceliktir. Tahmin edicilerin performanslarının değerlendirilmesinde genellikle, tahmin edilen regresyon katsayılarının HKO özelliklerine veya kestiricinin tasarım noktalarındaki HKO'suna odaklanılmaktadır. Ancak, amaç  $y$ 'nin bilinmeyen değerlerini kestirmek olduğunda, kriter olarak kestirimin hata kareleri ortalamasının (KHKO) kullanılması daha uygundur.

Gunst ve Mason (1979) çoklu doğrusal modeli için tahminin bütünleşik HKO'sunu geliştirmişler ve bazı regresyon tahmin edicileri için varyansı, yanlılığı ve HKO'yu elde etmişlerdir. Friedman ve Montgomery (1985), aynı tahmin ediciler üzerine benzer bir yaklaşımı, yanıt değişkeninin ortalamasının tahmini yerine, yeni bir değerinin kestirimine odaklanarak adapte etmişlerdir.  $\mathbf{x}'_0 = [1, X_{01}, X_{02}, \dots, X_{0k}]$  noktasındaki kestirim ölçüsü  $y_0 - \hat{y}_0$ 'dır. Burada  $\hat{y}_0 = \mathbf{x}'_0 \hat{\boldsymbol{\beta}}$ . Bu, Gunst ve Mason (1979) tarafından elde edilen denklemlerin genelleştirilmiş halidir.

Burada, Friedman ve Montgomery (1985) tarafından elde edilen KHKO'ları ve karşılaştırmaları yeniden ele alınarak bazı ilave yorumlarda bulunulacaktır. Daha sonra, aynı yaklaşım kullanılarak, Liu tahmin edicisinin kestirim performansı EKK, TB ve Ridge tahmin edicilerinin kestirim performansları ile karşılaştırılacaktır. Teorik sonuçlar, sayısal bir örnek üzerinde gösterilecektir.

## **2.2. Kestirim Hata Kareleri Ortalamalarının Elde Edilişi**

İlk olarak, Friedman ve Montgomery (1985) tarafından EKK, TB ve Ridge tahmin edicileri için elde edilen KHKO'lar bu çalışmadaki notasyona uygun olarak yeniden düzenlenerek verilecektir. Daha sonra Liu tahmin edicinin KHKO'su elde edilecektir.

KHKO, bir kestirimin, kestirilecek olan yanıt değişkenin gerçek değerine yakınlığının ölçüsüdür:

$$\text{KHKO} = E(y_0 - \hat{y}_0)^2. \quad (2.7)$$

KHKO'yu  $J$  ile gösterelim.  $J$ , varyansın ( $V$ ) ve yanlılığın karesinin ( $B$ ) toplamıdır:

$$J = V + B. \quad (2.8)$$

Eğer  $y_0$  kestirilecek değer ve  $\hat{y}_0$  bu değerın kestirimi ise, kestirim hatasının varyansı ve yanlılığı sırasıyla

$$V(y_0 - \hat{y}_0) = V(y_0) + V(\hat{y}_0) \quad (2.9)$$

ve

$$\text{Yanlılık} = E(y_0 - \hat{y}_0) \quad (2.10)$$

olarak ifade edilir. İşlemlerde kolaylık sağlaması için (2.1) eşitliğinin ortogonal formu

$$\mathbf{y} = \mathbf{Z}\mathbf{a} + \boldsymbol{\varepsilon} \quad (2.11)$$

kullanılacaktır. Burada  $\mathbf{Z} = \mathbf{X}\mathbf{T}$  ve  $\mathbf{a} = \mathbf{T}'\boldsymbol{\beta}$ . Bu durumda (2.11)'deki  $\mathbf{a}$ 'nın EKK tahmin edicisi

$$\hat{\mathbf{a}}_{\text{EKK}} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y} = \boldsymbol{\Lambda}^{-1}\mathbf{Z}'\mathbf{y} \quad (2.12)$$



dir. Eğer  $\mathbf{z}_0$ ,  $\hat{y}_0$  kestiriminin yapılacağı ortonormalleştirilmiş nokta ise EKK tahmin edicisinin KHKO'su

$$J_{EKK} = s^2 \left( 1 + \sum_{i=1}^p \frac{z_{0i}^2}{l_i} \right) \quad (2.13)$$

olarak elde eldir.

(2.11)'deki  $\boldsymbol{\alpha}$ 'nın TB tahmin edicisi

$$\hat{\boldsymbol{\alpha}}_{TB} = (\mathbf{Z}'_r \mathbf{Z}_r)^{-1} \mathbf{Z}'_r \mathbf{y} = \boldsymbol{\Lambda}_r^{-1} \mathbf{Z}'_r \mathbf{y} \quad (2.14)$$

dir. Burada  $\mathbf{Z}_r = \mathbf{X} \mathbf{T}_r$ . TB tahmin edicinin KHKO'su

$$J_{TB} = s^2 \left( 1 + \sum_{i=1}^r \frac{z_{0i}^2}{l_i} \right) + \left( \sum_{i=r+1}^p z_{0i} a_i \right)^2 \quad (2.15)$$

olarak elde edilir. (2.11)'deki  $\boldsymbol{\alpha}$ 'nın Ridge tahmin edicisi

$$\hat{\boldsymbol{\alpha}}_R = (\mathbf{Z}'\mathbf{Z} + k\mathbf{I})^{-1} \mathbf{Z}'\mathbf{y} = (\boldsymbol{\Lambda} + k\mathbf{I})^{-1} \mathbf{Z}'\mathbf{y} \quad , \quad k \geq 0 \quad (2.16)$$

ve bu tahmin edicinin KHKO'su

$$J_R = s^2 \left( 1 + \sum_{i=1}^p \frac{z_{0i}^2 l_i}{(l_i + k)^2} \right) + k^2 \left( \sum_{i=1}^p \frac{z_{0i} a_i}{l_i + k} \right)^2 \quad (2.17)$$

şeklindedir. (bkz, Friedman ve Montgomery, 1985).

(2.11)'deki  $\alpha$ 'nın Liu tahmin edicisi

$$\begin{aligned}\hat{\alpha}_{Liu} &= (\mathbf{Z}'\mathbf{Z} + \mathbf{I})^{-1}(\mathbf{Z}'\mathbf{y} + d\hat{\alpha}_{EKK}) \\ &= (\mathbf{Z}'\mathbf{Z} + \mathbf{I})^{-1}(\mathbf{Z}'\mathbf{Z} + d\mathbf{I})\hat{\alpha}_{EKK} \quad , \quad 0 < d < 1\end{aligned}\quad (2.18)$$

dir. Liu tahmin edicisinin kestirim hatasının varyansı

$$\begin{aligned}V_{Liu}(y_0 - \hat{y}_0) &= V(y_0) + V_{Liu}(\hat{y}_0) \\ &= s^2 + V(\mathbf{z}_0'\hat{\alpha}_{Liu}) \\ &= s^2 \left( 1 + \sum_{i=1}^p \frac{z_{0i}^2(l_i + d)^2}{l_i(l_i + 1)^2} \right)\end{aligned}\quad (2.19)$$

olarak elde edilir. Liu tahmin edicisinin kestirim hatasının yanlılığı ise

$$\begin{aligned}Yanlılık_{Liu} &= E(y_0 - \hat{y}_0) = \mathbf{z}_0'\alpha - \mathbf{z}_0'E(\hat{\alpha}_{Liu}) \\ &= \sum_{i=1}^p \frac{(1-d)z_{0i}a_i}{l_i + 1}\end{aligned}\quad (2.20)$$

dir. Dolayısıyla yanlılığın karesi

$$B_{Liu} = (Yanlılık_{Liu})^2 = \left( \sum_{i=1}^p \frac{(1-d)z_{0i}a_i}{l_i + 1} \right)^2 \quad (2.21)$$

olarak elde edilir. Varyans ve yanlılığın karesinin toplamı Liu tahmin edicisinin KHKO'sunu verecektir:

$$J_{Liu} = V_{Liu} + B_{Liu}$$

$$= s^2 \left( 1 + \sum_{i=1}^p \frac{z_{0i}^2 (l_i + d)^2}{l_i (l_i + 1)^2} \right) + \left( \sum_{i=1}^p \frac{(1-d)z_{0i} a_i}{l_i + 1} \right)^2. \quad (2.22)$$

### 2.3. Kestirim Hata Kareleri Ortalamalarının İki Boyutlu Uzayda Karşılaştırılması

Gunst ve Mason (1979) EKK, TB ve Ridge tahmin edicilerini bütünleşik HKO kriterine göre iki boyutlu uzayda karşılaştırmışlardır. Friedman ve Montgomery (1985) benzer bir yaklaşımı, aynı tahmin edicilerin KHKO'larını karşılaştırmak için kullanmışlardır.

Bu bölümde, ilk olarak Friedman ve Montgomery (1985) tarafından gerçekleştirilen KHKO karşılaştırmaları ele alınarak bazı değerlendirmeler yapılacaktır. Daha sonra Liu tahmin edicinin kestirim performansı değerlendirilecektir.

Gunst ve Mason (1979) ve Friedman ve Montgomery (1985)'nin çalışmalarında olduğu gibi  $a_1^2$  sıfır kabul edilecektir. Çünkü  $a_1^2$ 'nin sıfırdan farklı değerleri  $J_R$  ve  $J_{Liu}$ 'nin sabitlerini arttırmakta  $J_{EKK}$  ve  $J_{TB}$  nin değerlerini ise değiştirmemektedir.

Friedman ve Montgomery (1985), eğer  $J_{EKK} > J_R$  ise

$$\frac{z_{02}^2}{z_{01}^2} < \left[ \frac{s^2}{l_1} - \frac{s^2 l_1}{(l_1 + k)^2} \right] / \left[ \frac{s^2 l_2}{(l_2 + k)^2} + \frac{k^2 a_2^2}{(l_2 + k)^2} - \frac{s^2}{l_2} \right]. \quad (2.23)$$

olduğunu belirtmişlerdir. Ayrıca,  $a_2$ 'nin sağ taraftaki paydayı negatif yapmayan tüm değerleri için  $J_R$  nin  $J_{EKK}$  ya göre daima üstün olduğu ve bu kriterin,

$$a_2^2 > \frac{s^2 (2l_2 + k)}{k l_2}. \quad (2.24)$$

olması durumunda Ridge regresyonun EKK'ya göre daima üstün olduğu şeklinde ifade edilebileceğini belirtmişlerdir. Ancak, eğer Ridge tahmin edici, KHKO bakımından EKK tahmin ediciden üstün olması için

$$s^2 + \frac{s^2 z_{01}^2}{l_1} + \frac{s^2 z_{02}^2}{l_2} > s^2 + \frac{s^2 z_{01}^2 l_1}{(l_1 + k)^2} + \frac{s^2 z_{02}^2 l_2}{(l_2 + k)^2} + \frac{k^2 z_{02}^2 a_2^2}{(l_2 + k)^2}$$

ifadesi sağlanmalıdır. Bu eşitsizlikten

$$z_{01}^2 \left[ \frac{s^2}{l_1} - \frac{s^2 l_1}{(l_1 + k)^2} \right] > z_{02}^2 \left[ \frac{s^2 l_2}{(l_2 + k)^2} + \frac{k^2 a_2^2}{(l_2 + k)^2} - \frac{s^2}{l_2} \right]$$

elde edilir. Burada  $z_{01}^2$ ,  $z_{02}^2$  ve  $\frac{s^2}{l_1} - \frac{s^2 l_1}{(l_1 + k)^2}$  pozitifdir ama

$\frac{s^2 l_2}{(l_2 + k)^2} + \frac{k^2 a_2^2}{(l_2 + k)^2} - \frac{s^2}{l_2}$  ifadesinin işareti  $a_2^2$ 'nin değerine bağlıdır. Dolayısıyla

eğer (2.24) sağlanıyorsa (2.23) elde edilir. Ancak eğer

$$a_2^2 < \frac{s^2(2l_2 + k)}{kl_2} \quad (2.25)$$

ise, bu durumda Ridge tahmin edicinin KHKO bakımından EKK tahmin edicisine üstün olması koşulu

$$\frac{z_{02}^2}{z_{01}^2} > \left[ \frac{s^2}{l_1} - \frac{s^2 l_1}{(l_1 + k)^2} \right] / \left[ \frac{s^2 l_2}{(l_2 + k)^2} + \frac{k^2 a_2^2}{(l_2 + k)^2} - \frac{s^2}{l_2} \right] \quad (2.26)$$

halini alacaktır. (2.25) ile verilen koşul altında (2.26) nın sol tarafı daima pozitif, sağ tarafı da daima negatif olduğundan Ridge tahmin edici KHKO bakımından daima EKK tahmin edicisinden üstün olacaktır. Dolayısıyla, Ridge tahmin edici KHKO bakımından EKK tahmin edicisinden daima üstün olmasının koşulu (2.25) tir; Friedman ve Montgomery (1985) de belirtildiği gibi (2.24) değildir. (2.24) ile verilen koşul altında ise  $J_R < J_{OLS}$  olması için gerek ve yeter koşul (2.23)'ün sağlanmasıdır.

Friedman ve Montgomery tarafından  $k = 0.1$ ,  $l_1 = 1.95$  ve  $l_2 = 0.05$  şeklinde verilen örneği tekrar ele alalım (buradaki notasyonda en büyük özdeğerin  $l_1$  olduğu unutulmamalıdır). Bu örnekte,  $a_2^2 < 40$  olması durumunda,  $a_2^2$  ve  $z_{02}^2/z_{01}^2$  nin tüm değerleri için Ridge tahmin edici KHKO bakımından EKK tahmin ediciden üstündür (yani,  $J_R < J_{OLS}$ ).  $a_2^2 > 40$  olması durumunda ise, eğer

$$\frac{z_{02}^2}{z_{01}^2} < \frac{0.048811}{0.44444a_2^2 - 17.778}$$

koşulu sağlanıyorsa,  $J_R < J_{OLS}$  dir ( $J_R > J_{OLS}$  değil).

TB ve Ridge tahmin edicilerin KHKO'larının karşılaştırmasında Friedman ve Montgomery (1985), eğer  $J_{TB} > J_R$  ise, o zaman

$$a_2^2 < \frac{s^2}{l_2 + 2k}$$

için

$$\frac{z_{02}^2}{z_{01}^2} < \left[ \frac{s^2}{l_1} - \frac{s^2 l_1}{(l_1 + k)^2} \right] / \left[ \frac{s^2 l_2}{(l_2 + k)^2} + \frac{k^2 a_2^2}{(l_2 + k)^2} - a_2^2 \right] \quad (2.27)$$

olacağını, diğer durumlarda ise  $J_{TB} < J_R$  olacağını belirtmişlerdir.

Ancak, Ridge tahmin edici, KHKO bakımından TB tahmin ediciden üstün olması için

$$s^2 + \frac{s^2 z_{01}^2}{l_1} + z_{02}^2 a_2^2 > s^2 + \frac{s^2 z_{01}^2 l_1}{(l_1 + k)^2} + \frac{s^2 z_{02}^2 l_2}{(l_2 + k)^2} + \frac{k^2 z_{02}^2 a_2^2}{(l_2 + k)^2}$$

koşulu sağlanmalıdır. Bu eşitsizlikten

$$z_{01}^2 \left[ \frac{s^2}{l_1} - \frac{s^2 l_1}{(l_1 + k)^2} \right] > z_{02}^2 \left[ \frac{s^2 l_2}{(l_2 + k)^2} + \frac{k^2 a_2^2}{(l_2 + k)^2} - a_2^2 \right]$$

elde edilir. Burada  $z_{01}^2$ ,  $z_{02}^2$  ve  $\frac{s^2}{l_1} - \frac{s^2 l_1}{(l_1 + k)^2}$  pozitifdir ama

$\frac{s^2 l_2}{(l_2 + k)^2} + \frac{k^2 a_2^2}{(l_2 + k)^2} - a_2^2$  ifadesinin işareti  $a_2^2$  nin değerine bağlıdır. Eğer

$$a_2^2 < \frac{s^2}{l_2 + 2k} \quad (2.28)$$

ise (2.27) ile verilen ifade elde edilir. Ancak eğer

$$a_2^2 > \frac{s^2}{l_2 + 2k} \quad (2.29)$$

ise, bu durumda Ridge tahmin edicinin KHKO bakımından TB tahmin edicisine üstün olması koşulu

$$\frac{z_{02}^2}{z_{01}^2} > \left[ \frac{s^2}{l_1} - \frac{s^2 l_1}{(l_1 + k)^2} \right] / \left[ \frac{s^2 l_2}{(l_2 + k)^2} + \frac{k^2 a_2^2}{(l_2 + k)^2} - a_2^2 \right]. \quad (2.30)$$

halini alacaktır. (2.29) ile verilen koşul altında (2.30) un sol tarafı daima pozitif, sağ tarafı da daima negatif olduğundan, Ridge tahmin edici KHKO bakımından daima TB tahmin edicisinden üstün olacaktır. (2.28) ile verilen koşul altında ise  $J_R < J_{TB}$  olması için gerek ve yeter koşul (2.27) nin sağlanmasıdır. Bu nedenle Ridge tahmin edicinin KHKO bakımından TB tahmin ediciden üstün olması koşulunun şu şekilde ifade edilmesi uygundur:

Eğer  $a_2^2 > \frac{s^2}{l_2 + 2k}$  ise Ridge tahmin edici KHKO bakımından TB tahmin ediciden

daima üstündür. Eğer  $a_2^2 < \frac{s^2}{l_2 + 2k}$  ise Ridge tahmin edici KHKO bakımından TB tahmin ediciden ancak ve ancak (2.27) koşulu sağlandığı durumunda üstündür.

Şimdi, Liu tahmin edicisinin kestirim performansını EKK, TB ve Ridge tahmin edicilerinin kestirim performanslarını karşılaştıralım.

**Teorem 1:**

- a. Eğer  $a_2^2 < \frac{(l_2 + 1)^2 - (l_2 + d)^2}{(1 - d)^2} \cdot \frac{s^2}{l_2}$  ise  $J_{EKK} > J_{Liu}$  dir.
- b. Eğer  $a_2^2 > \frac{(l_2 + 1)^2 - (l_2 + d)^2}{(1 - d)^2} \cdot \frac{s^2}{l_2}$  ise  $J_{EKK} > J_{Liu}$  olması için gerek ve yeter koşul  $\frac{z_{02}^2}{z_{01}^2} < \left[ \frac{s^2}{l_1} - \frac{s^2 (l_1 + d)^2}{l_1 (l_1 + 1)^2} \right] / \left[ \frac{s^2 (l_2 + d)^2}{l_2 (l_2 + 1)^2} + \frac{(1 - d)^2 a_2^2}{(l_2 + 1)^2} - \frac{s^2}{l_2} \right]$  olmasıdır.

**İspat:** Liu tahmin edicisi KHKO bakımından EKK tahmin edicisinden daha iyi ise  $J_{EKK} > J_{Liu}$  olur. İki boyutlu uzayda açık olarak yazılmış hali

$$s^2 + \frac{s^2 z_{01}^2}{l_1} + \frac{s^2 z_{02}^2}{l_2} > \quad (2.31)$$

$$s^2 + \frac{s^2 z_{01}^2 (l_1 + d)^2}{l_1 (l_1 + 1)^2} + \frac{s^2 z_{02}^2 (l_2 + d)^2}{l_2 (l_2 + 1)^2} + \frac{(1-d)^2 z_{02}^2 a_2^2}{(l_2 + 1)^2}$$

dir. Bu eşitsizlikten

$$z_{01}^2 \left[ \frac{s^2}{l_1} - \frac{s^2 (l_1 + d)^2}{l_1 (l_1 + 1)^2} \right] > z_{02}^2 \left[ \frac{s^2 (l_2 + d)^2}{l_2 (l_2 + 1)^2} + \frac{(1-d)^2 a_2^2}{(l_2 + 1)^2} - \frac{s^2}{l_2} \right] \quad (2.32)$$

elde edilir. Burada  $z_{01}^2$ ,  $z_{02}^2$  ve  $\frac{s^2}{l_1} - \frac{s^2 (l_1 + d)^2}{l_1 (l_1 + 1)^2}$  pozitifdir ancak,

$\frac{s^2 (l_2 + d)^2}{l_2 (l_2 + 1)^2} + \frac{(1-d)^2 a_2^2}{(l_2 + 1)^2} - \frac{s^2}{l_2}$  ifadesinin işareti  $a_2^2$ 'nin değerine bağlıdır.

$$\left[ \frac{s^2}{l_1} - \frac{s^2 (l_1 + d)^2}{l_1 (l_1 + 1)^2} \right] \left/ \left[ \frac{s^2 (l_2 + d)^2}{l_2 (l_2 + 1)^2} + \frac{(1-d)^2 a_2^2}{(l_2 + 1)^2} - \frac{s^2}{l_2} \right] \right. \quad (2.33)$$

ifadesini  $a_2^2$ 'nin bir fonksiyonu olarak tanımlayalım ve bu fonksiyonu  $f_1(a_2^2)$  ile gösterelim.  $f_1(a_2^2)$  fonksiyonu

$$\frac{s^2 (l_2 + d)^2}{l_2 (l_2 + 1)^2} + \frac{(1-d)^2 a_2^2}{(l_2 + 1)^2} - \frac{s^2}{l_2} = 0 \quad (2.34)$$

noktasında düşey asimptota sahiptir. Bu eşitlikten  $f_1(a_2^2)$  fonksiyonunu tanımsız yapan  $a_2^2$  değeri



$$a_2^2 = \frac{(l_2 + 1)^2 - (l_2 + d)^2}{(1 - d)^2} \cdot \frac{s^2}{l_2} \quad (2.35)$$

olarak elde edilir. Böylece,

**a.** eğer

$$a_2^2 < \frac{(l_2 + 1)^2 - (l_2 + d)^2}{(1 - d)^2} \cdot \frac{s^2}{l_2} \quad (2.36)$$

ise

$$\frac{z_{02}^2}{z_{01}^2} > f_1(a_2^2) \quad (2.37)$$

eşitsizliği elde edilir.  $z_{02}^2/z_{01}^2$  daima pozitif olduğundan ve (2.36) ile verilen koşul için  $f_1(a_2^2)$  daima negatif olduğundan Liu tahmin edicisi EKK tahmin edicisinden üstündür.

**b.** Eğer

$$a_2^2 > \frac{(l_2 + 1)^2 - (l_2 + d)^2}{(1 - d)^2} \cdot \frac{s^2}{l_2} \quad (2.38)$$

ise,

$$\frac{z_{02}^2}{z_{01}^2} < f_1(a_2^2) \quad (2.39)$$

eşitsizliği elde edilir. (2.38) ile verilen koşul altında  $f_1(a_2^2)$  pozitif olduğundan, ancak (2.39) sağlandığı durumda Liu tahmin edicisi EKK tahmin edicisinden üstündür.

**Teorem 2:**

- a. Eğer  $a_2^2 < \frac{(l_2 + d)^2}{(l_2 + 1)^2 - (1 - d)^2} \cdot \frac{s^2}{l_2}$  ise  $J_{TB} > J_{Liu}$  dir.
- b. Eğer  $a_2^2 > \frac{(l_2 + d)^2}{(l_2 + 1)^2 - (1 - d)^2} \cdot \frac{s^2}{l_2}$  ise  $J_{TB} > J_{Liu}$  olması için gerek ve yeter koşul  $\frac{z_{02}^2}{z_{01}^2} < \left[ \frac{s^2}{l_1} - \frac{s^2(l_1 + d)^2}{l_1(l_1 + 1)^2} \right] / \left[ \frac{s^2(l_2 + d)^2}{l_2(l_2 + 1)^2} + \frac{(1 - d)^2 a_2^2}{(l_2 + 1)^2} - a_2^2 \right]$  olmasıdır.

**İspat :** Liu tahmin edicisi KHKO bakımından TB tahmin edicisinden daha iyi ise  $J_{TB} > J_{Liu}$  olur. Bu,

$$s^2 + \frac{s^2 z_{01}^2}{l_1} + z_{02}^2 a_2^2 > s^2 + \frac{s^2 z_{01}^2 (l_1 + d)^2}{l_1 (l_1 + 1)^2} + \frac{s^2 z_{02}^2 (l_2 + d)^2}{l_2 (l_2 + 1)^2} + \frac{(1 - d)^2 z_{02}^2 a_2^2}{(l_2 + 1)^2} \quad (2.40)$$

olmasını gerektirir. Bu ifade yeniden düzenlenirse

$$z_{01}^2 \left[ \frac{s^2}{l_1} - \frac{s^2 (l_1 + d)^2}{l_1 (l_1 + 1)^2} \right] > z_{02}^2 \left[ \frac{s^2 (l_2 + d)^2}{l_2 (l_2 + 1)^2} + \frac{(1 - d)^2 a_2^2}{(l_2 + 1)^2} - a_2^2 \right] \quad (2.41)$$

elde edilir. Bu eşitsizlikte,

$$\frac{s^2(l_2+d)^2}{l_2(l_2+1)^2} + \frac{(1-d)^2 a_2^2}{(l_2+1)^2} - a_2^2 \quad (2.42)$$

dışındaki çarpanlar pozitifdir. (2.42) ile verilen çarpanın işareti ise  $a_2^2$ 'nin değerine bağlıdır.

$$\left[ \frac{s^2}{l_1} - \frac{s^2(l_1+d)^2}{l_1(l_1+1)^2} \right] \left/ \left[ \frac{s^2(l_2+d)^2}{l_2(l_2+1)^2} + \frac{(1-d)^2 a_2^2}{(l_2+1)^2} - a_2^2 \right] \right. \quad (2.43)$$

ifadesi  $a_2^2$ 'nin bir fonksiyonu olarak tanımlansın ve  $f_2(a_2^2)$  ile gösterilsin. Bu fonksiyon,

$$a_2^2 = \frac{(l_2+d)^2}{(l_2+1)^2 - (1-d)^2} \cdot \frac{s^2}{l_2} \quad (2.44)$$

noktasında düşey asimptota sahiptir. Böylece,

**a.** Eğer

$$a_2^2 < \frac{(l_2+d)^2}{(l_2+1)^2 - (1-d)^2} \cdot \frac{s^2}{l_2} \quad (2.45)$$

ise

$$\frac{z_{02}^2}{z_{01}^2} > f_2(a_2^2) \quad (2.46)$$

olur.  $f_2(a_2^2)$ , (2.45) ile verilen koşul altında daima negatif olduğundan (2.46) ile verilen koşul daima sağlanmaktadır. Sonuç olarak, bu alanda Liu tahmin edicisi TB tahmin edicisinden daima üstündür.

**b. Eğer**

$$a_2^2 > \frac{(l_2 + d)^2}{(l_2 + 1)^2 - (1 - d)^2} \cdot \frac{s^2}{l_2} \quad (2.47)$$

ise

$$\frac{z_{02}^2}{z_{01}^2} < f_2(a_2^2) \quad (2.48)$$

eşitsizliği elde edilir.  $f_2(a_2^2)$ , (2.47) ile verilen koşul altında daima pozitif olduğundan, ancak (2.48) ile verilen koşul sağlandığı durumlarda Liu tahmin edicisi TB tahmin edicisinden daha iyidir.

**Teorem 3:**

**a. Eğer**  $a_2^2 < \frac{l_2^2(l_2 + 1)^2 - (l_2 + k)^2(l_2 + d)^2}{(1 - d)^2(l_2 + k)^2 - k^2(l_2 + 1)^2} \cdot \frac{s^2}{l_2}$  ise

**i.**  $l_1^2(l_1 + 1)^2 > (l_1 + d)^2(l_1 + k)^2$  için  $J_R > J_{Liu}$  dir.

**ii.**  $l_1^2(l_1 + 1)^2 < (l_1 + d)^2(l_1 + k)^2$  için ancak ve ancak  $z_{02}^2/z_{01}^2 < f_3(a_2^2)$  ise  $J_R > J_{Liu}$  dir.

**b. Eğer**  $a_2^2 > \frac{l_2^2(l_2 + 1)^2 - (l_2 + k)^2(l_2 + d)^2}{(1 - d)^2(l_2 + k)^2 - k^2(l_2 + 1)^2} \cdot \frac{s^2}{l_2}$  ise

- i.  $l_1^2(l_1+1)^2 < (l_1+d)^2(l_1+k)^2$  için  $J_R > J_{Liu}$  dir.
- ii.  $l_1^2(l_1+1)^2 > (l_1+d)^2(l_1+k)^2$  için ancak ve ancak  $z_{02}^2/z_{01}^2 < f_3(a_2^2)$  ise  $J_R > J_{Liu}$  dir.

Burada

$$f_3(a_2^2) = \frac{\frac{s^2 l_1}{(l_1+k)^2} - \frac{s^2 (l_1+d)^2}{l_1(l_1+1)^2}}{\frac{s^2 (l_2+d)^2}{l_2(l_2+1)^2} + \frac{(1-d)^2 a_2^2}{(l_2+1)^2} - \frac{s^2 l_2}{(l_2+k)^2} - \frac{k^2 a_2^2}{(l_2+k)^2}}. \quad (2.49)$$

**İspat:** : Liu tahmin edicisi KHKO bakımından Ridge tahmin edicisinden daha iyi ise  $J_R > J_{Liu}$  olur. Yani,

$$s^2 + \frac{s^2 z_{01}^2 l_1}{(l_1+k)^2} + \frac{s^2 z_{02}^2 l_2}{(l_2+k)^2} + \frac{k^2 z_{02}^2 a_2^2}{(l_2+k)^2} > \quad (2.50)$$

$$s^2 + \frac{s^2 z_{01}^2 (l_1+d)^2}{l_1(l_1+1)^2} + \frac{s^2 z_{02}^2 (l_2+d)^2}{l_2(l_2+1)^2} + \frac{(1-d)^2 z_{02}^2 a_2^2}{(l_2+1)^2}.$$

Bu eşitsizlik yeniden düzenlendiğinde

$$z_{01}^2 \left[ \frac{s^2 l_1}{(l_1+k)^2} - \frac{s^2 (l_1+d)^2}{l_1(l_1+1)^2} \right] > \quad (2.51)$$

$$z_{02}^2 \left[ \frac{s^2 (l_2+d)^2}{l_2(l_2+1)^2} + \frac{(1-d)^2 a_2^2}{(l_2+1)^2} - \frac{s^2 l_2}{(l_2+k)^2} - \frac{k^2 a_2^2}{(l_2+k)^2} \right]$$

elde edilir. Eğer

$$\frac{s^2 l_1}{(l_1 + k)^2} - \frac{s^2 (l_1 + d)^2}{l_1 (l_1 + 1)^2} \quad (2.52)$$

ve

$$\frac{s^2 (l_2 + d)^2}{l_2 (l_2 + 1)^2} + \frac{(1-d)^2 a_2^2}{(l_2 + 1)^2} - \frac{s^2 l_2}{(l_2 + k)^2} - \frac{k^2 a_2^2}{(l_2 + k)^2} \quad (2.53)$$

aynı işarete sahipse, Liu tahmin edicinin Ridge tahmin ediciden daha iyi olmasının koşulu

$$\frac{z_{02}^2}{z_{01}^2} < f_3(a_2^2) \quad (2.54)$$

olur. Eğer (2.52) ve (2.53) zıt işaretlere sahipse, Liu tahmin edicinin Ridge tahmin ediciden daha iyi olmasının koşulu

$$\frac{z_{02}^2}{z_{01}^2} > f_3(a_2^2) \quad (2.55)$$

dır. Eğer (2.52) ve (2.53) zıt işaretlere sahipse (2.55)'in sağ tarafının negatif olacağı açıktır. Böylece (2.55) daima sağlanır. Sonuç olarak, bu alanda Liu tahmin edici Ridge tahmini ediciden daima üstündür.

(2.52)'nin pozitif olma koşulu

$$\frac{s^2 l_1}{(l_1 + k)^2} - \frac{s^2 (l_1 + d)^2}{l_1 (l_1 + 1)^2} > 0$$

olmasıdır. Bu da

$$l_1^2(l_1+1)^2 > (l_1+d)^2(l_1+k)^2 \quad (2.56)$$

şeklinde yazılabilir. (2.53)'ün pozitif olma koşulu ise

$$\frac{s^2(l_2+d)^2}{l_2(l_2+1)^2} + \frac{(1-d)^2a_2^2}{(l_2+1)^2} - \frac{s^2l_2}{(l_2+k)^2} - \frac{k^2a_2^2}{(l_2+k)^2} > 0$$

olmasıdır. Ya da denk olarak

$$a_2^2 > \frac{l_2^2(l_2+1)^2 - (l_2+k)^2(l_2+d)^2}{(1-d)^2(l_2+k)^2 - k^2(l_2+1)^2} \cdot \frac{s^2}{l_2} \quad (2.57)$$

olmasıdır. (2.52) ve (2.53)'ün negatif olmaları için tam tersi koşulların sağlanması gerekmektedir.  $f_3(a_2^2)$  hiperbolünün düşey asimptotu

$$a_2^2 = \frac{l_2^2(l_2+1)^2 - (l_2+k)^2(l_2+d)^2}{(1-d)^2(l_2+k)^2 - k^2(l_2+1)^2} \cdot \frac{s^2}{l_2} \quad (2.58)$$

noktasından geçmektedir.

## 2.4. Sayısal Örnek

Burada, önceki bölümde elde edilen teorik sonuçlar, Friedman ve Montgomery (1985) tarafından verilen sayısal örnek kullanılarak (yani,  $k = 0.1$  ve  $r_{12} = 0.95$ ) gösterilmeye çalışılacaktır. Ridge tahmin ediciyle aynı yanlılığa sahip olması için Liu tahmin edicinin yanlılık parametresi  $d = 0.9$  olarak seçilmiştir. Bu şekilde, iki tahmin edicinin eşit koşullar altında karşılaştırması hedeflenmiştir.

İlk olarak Liu tahmin edicinin kestirim performansını EKK tahmin edicinin kestirim performansı ile karşılaştıralım. (2.33) ten

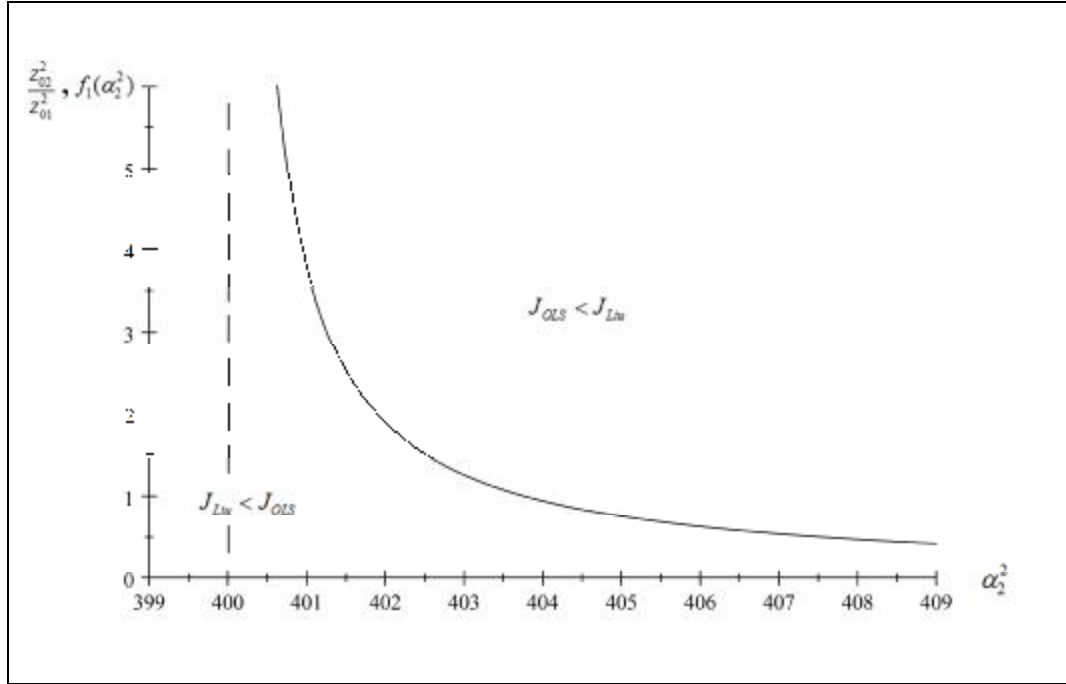
$$f_1(a_2^2) = \frac{0.034178}{0.00907a_2^2 - 3.6281} \quad (2.59)$$

elde edilir. Bu fonksiyon, düşey asiptotu

$$a_2^2 = 400 \quad (2.60)$$

noktasından geçen bir hiperboldür. Hem  $z_{02}^2/z_{01}^2$  hem de  $a_2^2$  pozitif değerler olduğundan, fonksiyonun sadece I. Bölge’de yer alan kısmı ile ilgilenmekteyiz. Şekil 2.1. tahmin edicilerin KHKO bakımından birbirine göre üstün oldukları bölgeleri göstermektedir. Şekilde görüldüğü gibi, bu örnekte  $a_2^2 = 400$  noktasına kadar Liu tahmin edici EKK tahmin ediciye göre daima üstündür. Bu noktadan sonra ise hangi tahmin edicinin üstün olduğu  $f_1(a_2^2)$  foksiyonuna bağlıdır.  $z_{02}^2/z_{01}^2$  oranının küçük değerleri için Liu tahmin edici EKK tahmin ediciye kıyasla KHKO bakımından üstündür. Friedman ve Montgomery (1985) nin de belirttiği gibi  $z_{02}^2/z_{01}^2$  oranı kestirilecek değerlerin alt uzayını belirlemektedir, dolayısıyla özel bir öneme sahiptir. Bu oranın küçük değerleri, temel bileşene bağımlılığı işaret etmektedir.





Şekil 2.1. Liu ve EKK tahmin edicilerin KHKO'larının karşılaştırılması

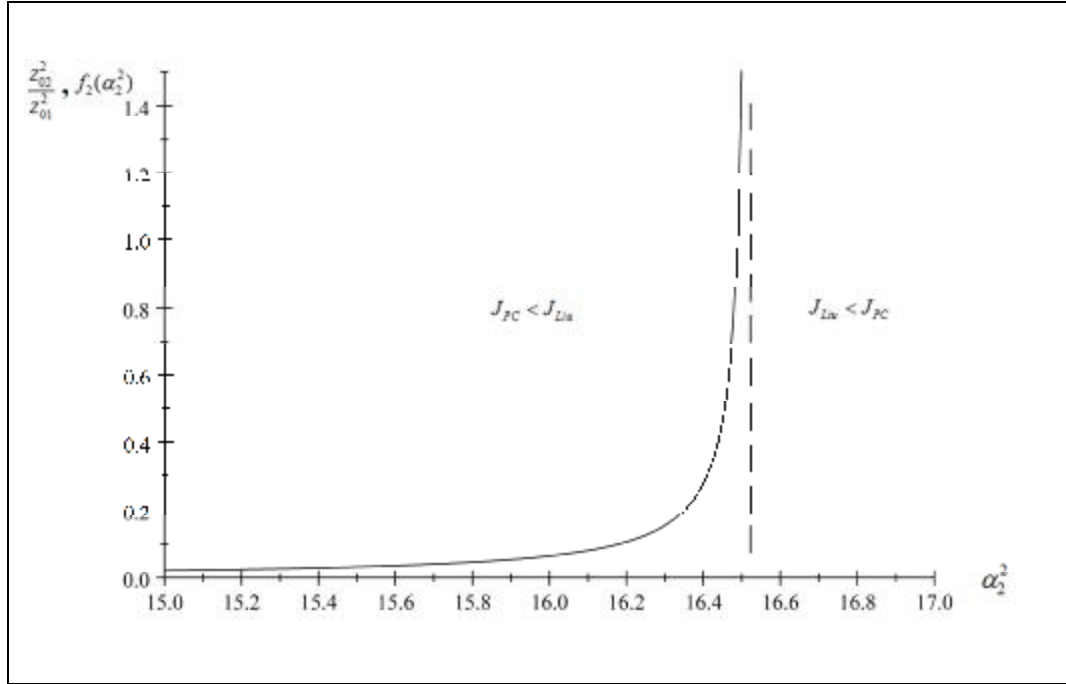
Aynı örnek için Liu ve TB tahmin edicilerinin KHKO kriterine göre kestirim performanslarını değerlendirelim. (2.43) ten

$$f_2(a_2^2) = \frac{0.034178}{16.372 - 0.99093a_2^2} \quad (2.61)$$

fonksiyonu elde edilir. Bu fonksiyonun düşey asimptotu

$$a_2^2 = 16.522 \quad (2.62)$$

noktasından geçmektedir. Şekil 2.2., Liu ve TB tahmin edicilerin KHKO kriterine göre üstün oldukları bölgeleri göstermektedir.  $a_2^2$  nin 16.522'den büyük değerleri için Liu tahmin edici TB tahmin ediciye göre daima üstündür.  $a_2^2$  nin 16.522'den küçük değerleri için ise hangi tahmin edicinin üstün olduğu  $z_{02}^2/z_{01}^2$  oranının değerine bağlıdır. Eğer  $z_{02}^2/z_{01}^2$  oranının değeri  $f_2$  nin değerinden küçükse Liu tahmin edici TB



Şekil 2.2. Liu ve TB tahmin edicilerin KHKO'larının karşılaştırılması

tahmin ediciden üstündür.  $z_{02}^2/z_{01}^2$  oranının değeri  $f_2$  nin değerinden büyükse TB tahmin edici Liu tahmin ediciden üstündür.

Son olarak Liu tahmin edici ile Ridge tahmin ediciyi KHKO bakımından karşılaştıralım. (2.49) dan  $a_2^2$  nin fonksiyonu

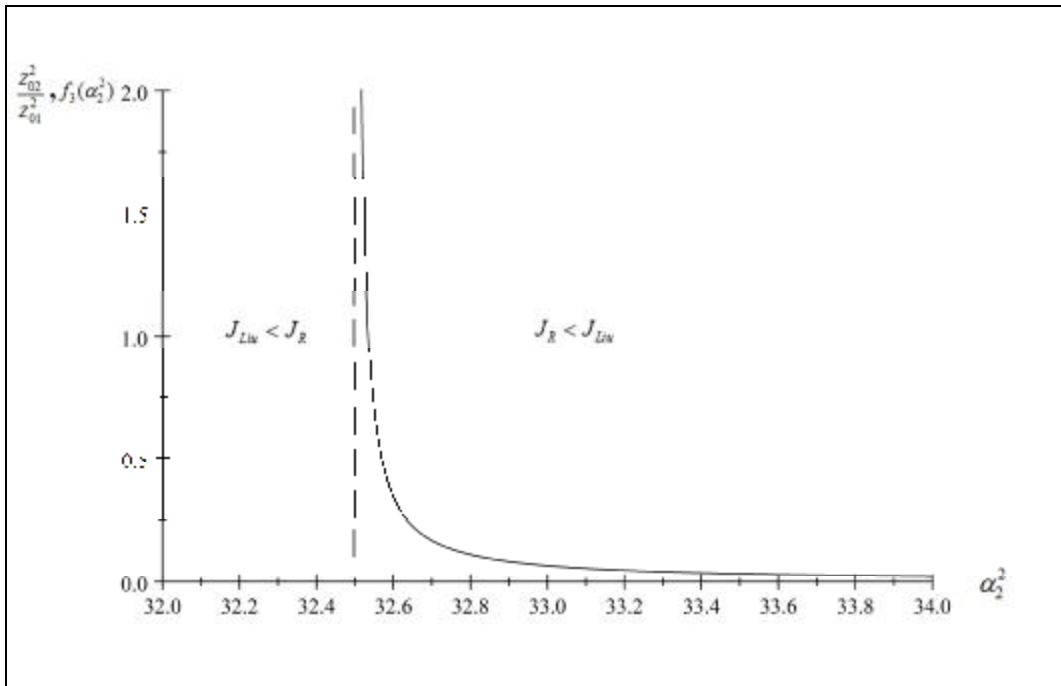
$$f_3(a_2^2) = \frac{0.014633}{0.43537a_2^2 - 14.150} \quad (2.63)$$

olarak elde edilir. Düşey asimptot

$$a_2^2 = 32.5 \quad (2.64)$$

noktasından geçmektedir. Şekil 2.3. te Liu ve Ridge tahmin edicilerin karşılaştırılması görülmektedir.  $a_2^2$  nin 32.5 ten küçük değerleri için Liu tahmin edici

Ridge tahmin ediciye göre daima üstündür.  $a_2^2$  nin 32.5 ten büyük değerleri için ise Liu ve Ridge tahmin edicileri arasında geçişler vardır.  $z_{02}^2/z_{01}^2$  oranının değeri  $f_3$  fonksiyonunun değerinden küçükse Liu tahmin edicinin kestirimi Ridge tahmin edicinin kestiriminden daha iyidir.  $z_{02}^2/z_{01}^2$  oranının değeri  $f_3$  fonksiyonunun değerinden büyük olduğu durumlarda ise Ridge tahmin edicinin kestirim performansı Liu tahmin edicinin kestirim performansından daha iyidir.



Şekil 2.3. Liu ve Ridge tahmin edicilerin KHKO'larının karşılaştırılması

Yukarıda elde edilen sonuçlar, Liu tahmin edicinin diğer üç tahmin ediciye göre üstün olduğu bir bölgenin belirlenebileceğini göstermektedir. Verilen örnekte bu alan,  $z_{02}^2/z_{01}^2$  oranının tüm değerleri ile  $a_2^2$  nin  $[16.522, 32.5]$  aralığındaki değerleri arasında kalan bölgedir.

## 2.5. Sonuç

Bu bölümde Liu tahmin edicinin EKK, TB ve Ridge tahmin edicilerine kıyasla kestirim performansı değerlendirilmiştir. Bu tahmin ediciler arasındaki

karşılaştırmalar, iki boyutlu açıklayıcı değişkenler uzayında belirli bir noktadaki KHKO cinsindendir.

İlk olarak Friedman ve Montgomery (1985) tarafından yapılan bazı karşılaştırmalara değinilmiş ve bu karşılaştırmalar üzerine bazı yorumlar getirilmiştir. Daha sonra Friedman ve Montgomery'nin yaklaşımı kullanılarak, Liu tahmin edicinin kestirim performansı EKK, TB ve Ridge tahmin edicilerinin kestirim performansları ile karşılaştırılmıştır. Bu bağlamda, öncelikle Liu tahmin edicinin KHKO'su elde edilmiş ve devamında üç teorem verilmiştir.

Elde edilen teorik sonuçlar sayısal bir örnek üzerinde gösterilmiş ve adı geçen tahmin edicilerin üstün olduğu bölgeler belirlenmiştir. Hangi tahmin edicinin daha etkili olacağını belirleyen etkenlerden biri  $a_2^2$  nin değerleridir. Örneğin, EKK tahmin edicisi ancak  $a_2^2$  çok büyük değerler aldığı anda etkilidir. Bu tahmin edicilerin etkili oluşlarını belirleyen diğer etken ise kestirilecek noktanın konumudur. Eğer kestirim, daha istikrarlı regresör yönünde ise bu durumda TB tahmin edicisi tercih edilebilir. Dolayısıyla, tahmin edicinin seçimi kestirilecek noktanın konumuna bağlı olduğu gibi, çoklu iç ilişkinin derecesine de bağlıdır.

Verilen sayısal örnekte, Liu tahmin edicinin EKK, TB ve Ridge tahmin edicilerinden daima üstün olduğu bir bölge belirlenmiştir. Bu, böyle bir alanın belirlenmesinin teorik olarak mümkün olduğunu göstermektedir.

### **3. r-k VE r-d SINIF TAHMİN EDİCİLERİN KESTİRİM PERFORMANSLARININ İYİLEŞTİRİLMESİ**

EKK tahmin edicisine alternatif olarak tanımlanan yanlı tahmin edicilerden ikisi de r-k ve r-d sınıf tahmin edicileridir. Bunların hata kareleri ortalaması (HKO) kriterine göre karşılaştırılmaları literatürde ele alınmıştır. Bu bölümde, birbirine alternatif bazı yanlı tahmin edicilerin, yanlılık parametrelerinin PRESS istatistiğini minimize edecek şekilde seçimi önerilmiştir. Daha sonra bu tahmin edicilerin kestirim performansları bir veri seti yardımıyla karşılaştırılmıştır.

#### **3.1. Giriş**

Daha önce de belirtildiği gibi çoklu iç ilişkinin ortaya çıkardığı sorunlarla baş etmek için EKK tahmin edicisine alternatif bazı yanlı tahmin ediciler önerilmiştir. Bir sonraki alt bölümde bu tahmin edicilerden bazıları ele alınacaktır.

Yanlı tahmin edicilerdeki yanlılık parametreleri, amaca uygun olarak, katsayıların tahmininde bir gelişme veya kestirimlerde bir iyileşme olacak şekilde seçilebilir. Tahminin iyiliği, genellikle, hata kareleri toplamı (HKT) veya HKO kriterleri ile ölçülür. Mümkün olduğunca iyi parametre tahminleri elde edilmek istenildiğinde, çoğunlukla, yanlılık parametresi HKO istatistiğini minimum yapacak şekilde seçilir. Literatürde  $C_p$  istatistiğinin minimizasyonu da önerilmektedir (bkz. Mallows, 1973 ve Liu, 1993). Ancak amaç, kestirim olduğunda, yanlılık parametresinin kestirim iyiliğini ölçen kriterlerden birini minimum yapacak şekilde seçilmesi daha uygun olur. Kestirim iyiliğinin ölçümünde kullanılan en yaygın kriterlerden biri, Allen (1971, 1974) tarafından önerilen ve PRESS olarak da bilinen CV (Cross-Validation) istatistiğidir. Bir diğeri ise Golub ve diğerleri (1979) tarafından önerilen GCV (Generalized Cross-Validation) istatistiğidir. GCV istatistiği, CV istatistiğinin rotasyona duyarsızlaştırılmış formudur. Bu kriterler, tanımlarında bilinmeyen parametreler içermediklerinden uygulamalarda kullanımları avantaj sağlar. Ancak, r-k ve r-d sınıf tahmin edicilerinde rotasyonun etkisi

önemlidir. Bu nedenle, bu tahmin edicilerde PRESS istatistiği daha uygun bir kriterdir.

Bir tahmin edicinin PRESS istatistiği ne kadar küçükse kestirim performansı o derece iyidir. Montgomery ve Friedman (1993) Ridge tahmin edicisinin yanlılık parametresi  $k$  nın seçiminde PRESS istatistiğini önermişlerdir. Benzer şekilde Özkale ve Kaçıranlar (2007), ridge tahmin edicisine alternatif olarak tanımlanan, Liu tahmin edicisinin yanlılık parametresi  $d$  nin seçiminde bu yöntemi önermişlerdir.

Baye ve Parker (1984) özel durumlarda EKK, ridge tahmin edici ve temel bileşenler (TB) tahmin ediciyi veren r-k sınıf tahmin ediciyi tanımlamışlardır. r-k sınıf tahmin edicinin, temel bileşenlerden daha küçük HKO değerine sahip olduğunu göstermişlerdir. Bu tahmin edicinin performansı farklı kriterlere göre Nomura ve Ohkubo (1985), Sarkar (1989), Sarkar (1996) ve Özkale ve Kaçıranlar (2008) de ele alınmıştır. Benzer şekilde, Kaçıranlar ve Sakallıoğlu (2001) özel durumlarda EKK, Liu tahmin edici ve TB tahmin ediciyi veren r-d sınıf tahmin ediciyi tanımlamışlardır. r-d sınıf tahmin edicinin, HKO yönünden, EKK, TB ve Liu tahmin edicilerinden daha iyi olduğu Kaçıranlar ve Sakallıoğlu (2001) tarafından gösterilmiştir.

r-k ve r-d sınıf tahmin edicilerinin HKO istatistikleri minimize edildiğinde, diğer tahmin edicilere göre daha iyi tahmin performansı gösterdiği literatürdeki çalışmalarda ele alınmıştır. Biz bu çalışmada, bu tahmin edicilerin PRESS istatistiklerini minimize ederek, daha iyi kestirim performansı verecek şekilde yanlılık parametrelerinin seçimlerini ele alacağız.

### **3.2. Tahmin Ediciler ve PRESS İstatistikleri**

Allen (1971,1974) tarafından önerilen PRESS istatistiği

$$PRESS = \sum_{i=1}^n e_{(i)}^2 = \sum_{i=1}^n \left[ \frac{e_i}{1-h_{ii}} \right]^2 \quad (3.1)$$

şeklinde tanımlanmaktadır. Burada  $e_{(i)}$ ,  $i$ . gözlem değeri  $y_i$  ile  $i$ . gözlem silinerek geriye kalan  $n-1$  gözlemden elde edilen parametre tahminleri ( $\hat{\beta}_{(i)}$ ) kullanılarak yapılan  $i$ . gözlem tahmini ( $\hat{y}_{(i)}$ ) arasındaki farktır:

$$e_{(i)} = y_i - \hat{y}_{(i)} = y_i - \mathbf{x}_i' \hat{\beta}_{(i)}. \quad (3.2)$$

$e_i$ ,  $i$ . gözleme karşılık gelen hata terimi ve  $h_{ii}$ , şapka matrisinin  $i$ . köşegen elemanıdır.

PRESS istatistiğini minimize edecek şekilde yanlılık parametreleri seçilerek Ridge, Liu, r-k sınıf ve r-d sınıf tahmin ediciler karşılaştırılacaktır. Ayrıca bu tahmin ediciler EKK ve TB tahmin edicileri ile de karşılaştırılacaktır.

(2.1) ile verilen çoklu doğrusal regresyon modelinin EKK tahmin edicisi (2.2) ile verilmiştir. EKK tahmin edicisi için PRESS istatistiği:

$$\text{PRESS}_{EKK} = \sum_{i=1}^n \left[ \frac{y_i - \mathbf{x}_i' \hat{\beta}_{EKK}}{1 - \mathbf{x}_i' (\mathbf{X}' \mathbf{X})^{-1} \mathbf{x}_i} \right]^2 \quad (3.3)$$

olarak tanımlanır. (2.3) ile verilen TB tahmin edicinin PRESS istatistiği ise aşağıdaki gibi tanımlanır:

$$\text{PRESS}_{TB} = \sum_{i=1}^n \left[ \frac{y_i - \mathbf{z}_{ri}' \hat{\mathbf{a}}_{TB}}{1 - \mathbf{z}_{ri}' (\mathbf{Z}_r' \mathbf{Z}_r)^{-1} \mathbf{z}_{ri}} \right]^2. \quad (3.4)$$

(2.4) ile tanımlanan Ridge tahmin edicinin PRESS istatistiği Montgomery ve Friedman (1993) tarafından

$$PRESS_R = \sum_{i=1}^n \left[ \frac{y_i - \mathbf{x}_i' \hat{\boldsymbol{\beta}}_R}{1 - \mathbf{x}_i' (\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1} \mathbf{x}_i} \right]^2 \quad (3.5)$$

şeklinde tanımlanmıştır.

Baye ve Parker (1984)'ün ridge ve temel bileşenler tahmin edicilerini birleştirerek önerdiği  $r-k$  sınıf tahmin edicisi:

$$\hat{\boldsymbol{\beta}}_{r-k} = \mathbf{T}_r (\mathbf{Z}_r' \mathbf{Z}_r + k\mathbf{I}_r)^{-1} \mathbf{Z}_r' \mathbf{y} \quad , \quad k \geq 0 \quad (3.6)$$

dir. Burada  $k$  yarıllılık parametresidir. Bu tahmin edicinin PRESS istatistiği:

$$PRESS_{r-k} = \sum_{i=1}^n \left[ \frac{y_i - \mathbf{z}_{ri}' \hat{\boldsymbol{\alpha}}_{r-k}}{1 - \mathbf{z}_{ri}' (\mathbf{Z}_r' \mathbf{Z}_r + k\mathbf{I}_r)^{-1} \mathbf{z}_{ri}} \right]^2 \quad (3.7)$$

olarak elde edilir. Burada  $\hat{\boldsymbol{\alpha}}_{r-k}$ ,  $\hat{\boldsymbol{\alpha}}_{r-k} = (\mathbf{Z}_r' \mathbf{Z}_r + k\mathbf{I}_r)^{-1} \mathbf{Z}_r' \mathbf{y}$  ile elde edilen  $r-k$  sınıf tahmin edicisinin kanonik formudur.  $\mathbf{z}_{ri}$ ,  $\mathbf{Z}_r$  matrisinin  $i$ .satırından oluşan vektördür.

Liu (1993)'ün, Stein tahmin edici ile ridge tahmin ediciyi birleştirerek önerdiği ve Akdeniz ve Kaçıranlar (1995) tarafından Linear Unified (Liu) olarak adlandırılan tahmin edici (2.5) ile tanımlanmıştır. Bu tahmin edicinin PRESS istatistiği:

$$PRESS_{Liu} = \sum_{i=1}^n \left[ \frac{y_i - \mathbf{x}_i' \hat{\boldsymbol{\beta}}_{Liu}}{1 - \mathbf{x}_i' (\mathbf{X}'\mathbf{X} + \mathbf{I})^{-1} (\mathbf{X}'\mathbf{X} + d\mathbf{I})(\mathbf{X}'\mathbf{X})^{-1} \mathbf{x}_i} \right]^2 \quad (3.8)$$



ile elde edilir (Özkale ve Kaçıranlar (2007)).

r-k sınıf tahmin ediciye alternatif olarak, Kaçıranlar ve Sakallıoğlu (2001)'in Liu ve temel bileşenler tahmin edicileri birleştirerek önerdiği r-d sınıf tahmin edicisi:

$$\hat{\beta}_{r-d} = \mathbf{T}_r (\mathbf{Z}'_r \mathbf{Z}_r + \mathbf{I}_r)^{-1} (\mathbf{Z}'_r \mathbf{y} + d \hat{\mathbf{a}}_{TB}) \quad , \quad -\infty < d < \infty \quad (3.9)$$

olarak tanımlanır. Bu tahmin edicinin PRESS istatistiği:

$$\text{PRESS}_{r-d} = \sum_{i=1}^n \left[ \frac{y_i - \mathbf{z}'_{ri} \hat{\mathbf{a}}_{r-d}}{1 - \mathbf{z}'_{ri} (\mathbf{Z}'_r \mathbf{Z}_r + \mathbf{I}_r)^{-1} (\mathbf{Z}'_r \mathbf{Z}_r + d \mathbf{I}_r) (\mathbf{Z}'_r \mathbf{Z}_r)^{-1} \mathbf{z}_{ri}} \right]^2 \quad (3.10)$$

şeklinde tanımlanır.

### 3.3. Sayısal Örnek

Bu uygulamada kullanılan yanlı tahmin edicilerin yanlılık parametreleri, modelin hem HKO istatistiği hem de PRESS istatistiği minimize olacak şekilde hesaplanacak ve PRESS istatistiğindeki iyileşme gözlenecektir. Her tahmin edici için mümkün olan en küçük PRESS istatistiği elde edildikten sonra da bu PRESS istatistikleri kendi aralarında karşılaştırılarak, o veri seti için en iyi tahmini veren tahmin edici seçilecektir.

#### 3.3.1. Mil Başına Benzin Tüketimi Veri Seti

Kullanılacak veri seti Montgomery ve Friedman (1993)'dan alınan Mil Başına Benzin Tüketimi verisidir. Burada  $\mathbf{y}$  arabaların mil başına harcadıkları benzin,  $\mathbf{x}_1$  toplam silindir hacmi,  $\mathbf{x}_2$  beygir gücü,  $\mathbf{x}_3$  tork,  $\mathbf{x}_4$  kompresyon oranı,  $\mathbf{x}_5$  arka aks oranı,  $\mathbf{x}_6$  karbüratör büyüklüğü,  $\mathbf{x}_7$  vites sayısı,  $\mathbf{x}_8$  arabanın uzunluğu,  $\mathbf{x}_9$  arabanın genişliği,  $\mathbf{x}_{10}$  arabanın ağırlığı ve  $\mathbf{x}_{11}$  vites tipidir.

Tablo 3.1. Mil Başına Benzin Tüketimi verisi

| y     | x <sub>1</sub> | x <sub>2</sub> | x <sub>3</sub> | x <sub>4</sub> | x <sub>5</sub> | x <sub>6</sub> | x <sub>7</sub> | x <sub>8</sub> | x <sub>9</sub> | x <sub>10</sub> | x <sub>11</sub> |
|-------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|-----------------|-----------------|
| 19.73 | 318.0          | 140            | 255            | 8.5            | 2.71           | 2              | 3              | 215.3          | 76.3           | 4370            | 1               |
| 11.20 | 440.0          | 215            | 330            | 8.2            | 2.88           | 4              | 3              | 184.5          | 69.0           | 4215            | 1               |
| 18.25 | 351.0          | 143            | 255            | 8.0            | 3.00           | 2              | 3              | 199.9          | 74.0           | 3890            | 1               |
| 21.47 | 360.0          | 180            | 290            | 8.4            | 2.45           | 2              | 3              | 214.2          | 76.3           | 4250            | 1               |
| 20.30 | 140.0          | 83             | 109            | 8.4            | 3.40           | 2              | 4              | 168.8          | 69.4           | 2700            | 0               |
| 36.50 | 85.3           | 80             | 83             | 8.5            | 3.89           | 2              | 4              | 160.6          | 62.2           | 2009            | 0               |
| 18.90 | 350.0          | 165            | 260            | 8.0            | 2.56           | 4              | 3              | 200.3          | 69.9           | 3910            | 1               |
| 30.40 | 96.9           | 75             | 83             | 9.0            | 4.30           | 2              | 5              | 165.2          | 65.0           | 2320            | 0               |
| 13.90 | 351.0          | 148            | 243            | 8.0            | 3.25           | 2              | 3              | 215.5          | 78.5           | 4540            | 1               |
| 14.89 | 440.0          | 215            | 330            | 8.2            | 2.71           | 4              | 3              | 231.0          | 79.7           | 5185            | 1               |
| 21.50 | 171.0          | 109            | 146            | 8.2            | 3.22           | 2              | 4              | 170.4          | 66.9           | 2655            | 0               |
| 17.80 | 302.0          | 129            | 220            | 8.0            | 3.00           | 2              | 3              | 199.9          | 74.0           | 3890            | 1               |
| 17.80 | 350.0          | 155            | 250            | 8.5            | 3.08           | 4              | 3              | 196.7          | 72.2           | 3910            | 1               |
| 16.41 | 318.0          | 145            | 255            | 8.5            | 2.45           | 2              | 3              | 197.6          | 71.0           | 3660            | 1               |
| 23.54 | 231.0          | 110            | 175            | 8.0            | 2.56           | 2              | 3              | 179.3          | 65.4           | 3050            | 1               |
| 31.90 | 96.9           | 75             | 83             | 9.0            | 4.30           | 2              | 5              | 165.2          | 61.8           | 2275            | 0               |
| 14.39 | 500.0          | 190            | 360            | 8.5            | 2.73           | 4              | 3              | 224.1          | 79.8           | 5290            | 1               |
| 22.12 | 231.0          | 110            | 175            | 8.0            | 2.56           | 2              | 3              | 179.3          | 65.4           | 3020            | 1               |
| 17.00 | 350.0          | 170            | 275            | 8.5            | 2.56           | 4              | 3              | 199.6          | 72.9           | 2860            | 1               |
| 20.00 | 250.0          | 105            | 185            | 8.25           | 2.73           | 1              | 3              | 196.7          | 72.2           | 3510            | 1               |

Hesaplamalara başlamadan önce değişkenler standartlaştırılmıştır.  $X'X$  matrisinin özdeğerleri  $l_1 = 7.8929$ ,  $l_2 = 1.4237$ ,  $l_3 = 0.8371$ ,  $l_4 = 0.3934$ ,  $l_5 = 0.1652$ ,  $l_6 = 0.1373$ ,  $l_7 = 0.0650$ ,  $l_8 = 0.0472$ ,  $l_9 = 0.0265$ ,  $l_{10} = 0.0108$ ,  $l_{11} = 0.0009$  dur. Koşul sayısı  $k = \frac{l_1}{l_{11}} \cong 8769.9$  çoklu doğrusal bağlantının olduğunu gösterir.

Tablo 3.2a. Mil Başına Benzin Tüketimi veri seti için Ridge, r-k sınıf, EKK ve TB tahmin edicileri ile tahmin edilen regresyon modeli için parametre tahminleri,  $R^2$ , HKT, HKO( $\hat{\beta}$ ) ve PRESS istatistikleri

|                      | Ridge                       |                               | r-k sınıf                   |                               | EKK     | TB      |
|----------------------|-----------------------------|-------------------------------|-----------------------------|-------------------------------|---------|---------|
|                      | $\hat{k}_{HKO} =$<br>0.0033 | $\hat{k}_{PRESS} =$<br>0.9784 | $\hat{k}_{HKO} =$<br>1.0689 | $\hat{k}_{PRESS} =$<br>0.2121 |         |         |
| $\hat{b}_1$          | -0.6886                     | -0.1203                       | -0.0917                     | -0.1008                       | -1.5741 | -0.1032 |
| $\hat{b}_2$          | -0.1436                     | -0.0920                       | -0.0846                     | -0.0925                       | -0.4580 | -0.0946 |
| $\hat{b}_3$          | 0.3460                      | -0.1032                       | -0.0911                     | -0.1001                       | 1.4987  | -0.1025 |
| $\hat{b}_4$          | 0.1643                      | 0.0960                        | 0.0550                      | 0.0632                        | 0.0839  | 0.0660  |
| $\hat{b}_5$          | 0.4508                      | 0.0782                        | 0.0811                      | 0.0911                        | 0.6251  | 0.0942  |
| $\hat{b}_6$          | -0.0875                     | -0.0560                       | -0.0465                     | -0.0493                       | -0.0240 | -0.0497 |
| $\hat{b}_7$          | -0.5611                     | 0.0735                        | 0.0882                      | 0.0989                        | -0.6516 | 0.1022  |
| $\hat{b}_8$          | 1.4341                      | -0.0275                       | -0.0859                     | -0.0946                       | 1.5734  | -0.0969 |
| $\hat{b}_9$          | -1.2398                     | -0.1115                       | -0.0796                     | -0.0875                       | -1.3603 | -0.0897 |
| $\hat{b}_{10}$       | -0.2168                     | -0.0900                       | -0.0865                     | -0.0951                       | -0.1897 | -0.0974 |
| $\hat{b}_{11}$       | -0.5021                     | -0.0576                       | -0.0874                     | -0.0976                       | -0.5389 | -0.1007 |
| $R^2$                | 0.8969                      | 0.7437                        | 0.7240                      | 0.7340                        | 0.9020  | 0.7346  |
| HKT                  | 0.1031                      | 0.2563                        | 0.2760                      | 0.2660                        | 0.0980  | 0.2654  |
| HKO( $\hat{\beta}$ ) | 4.2919                      | 10.0167                       | 10.3209                     | 10.3232                       | 14.6368 | 10.3250 |
| PRESS                | 0.36054                     | 0.3488                        | 0.3263                      | 0.3168                        | 0.7102  | 0.3177  |

Veri seti için, adı geçen yanlış tahmin edicilerin yanlışlık parametreleri, modelin hem HKO istatistikleri hem de PRESS istatistikleri minimize olacak şekilde hesaplanmıştır. Sonuçlar Tablo 3.1.2a. ve 3.1.2b.'de verilmiştir. TB, r-k sınıf ve r-d sınıf tahmin edicilerde özdeğerleri 1'den büyük olan bileşenler seçilmiştir. Bu nedenle bileşen sayısı 2'dir.

Tablo 3.2b. Mil Başına Benzin Tüketimi verisi için Liu, r-d sınıf, EKK ve TB tahmin edicileri ile tahmin edilen regresyon modeli için parametre tahminleri,  $R^2$ , HKT, HKO( $\hat{\beta}$ ) ve PRESS istatistikleri

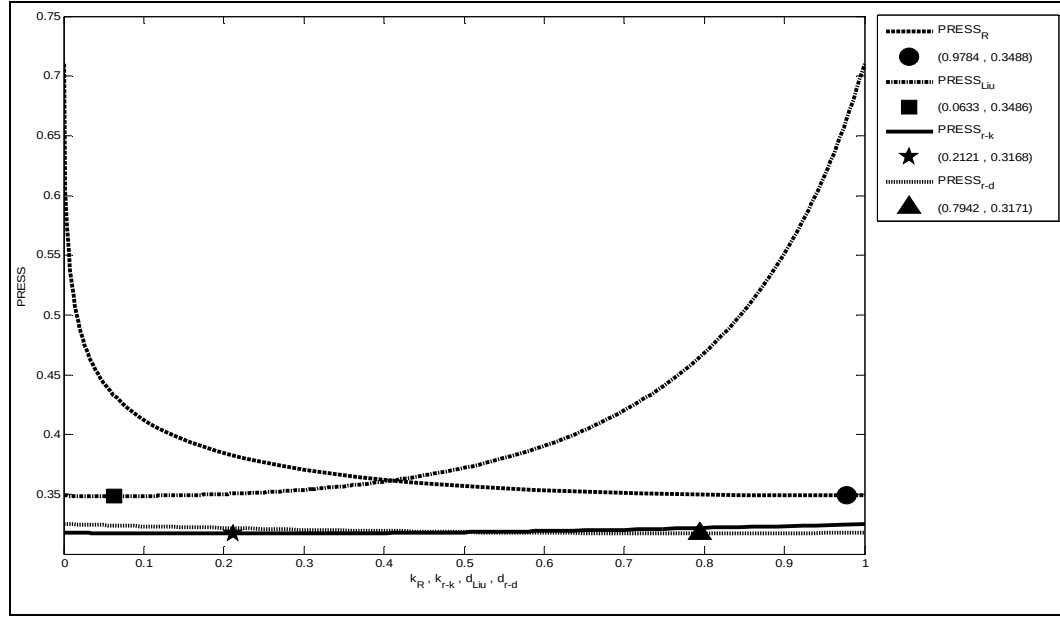
|                      | Liu Tahmin Edici          |                            | r-d sınıf Tahmin Edici   |                            | EKK     | TB      |
|----------------------|---------------------------|----------------------------|--------------------------|----------------------------|---------|---------|
|                      | $\hat{d}_{HKO} = -0.4058$ | $\hat{d}_{PRESS} = 0.0633$ | $\hat{d}_{HKO} = 0.2993$ | $\hat{d}_{PRESS} = 0.7942$ |         |         |
| $\hat{b}_1$          | 0.4706                    | -0.2116                    | -0.0956                  | -0.1010                    | -1.5741 | -0.1032 |
| $\hat{b}_2$          | 0.0568                    | -0.1149                    | -0.0880                  | -0.0927                    | -0.4580 | -0.0946 |
| $\hat{b}_3$          | -0.7525                   | -0.0014                    | -0.0950                  | -0.1003                    | 1.4987  | -0.1025 |
| $\hat{b}_4$          | 0.0999                    | 0.0945                     | 0.0587                   | 0.0639                     | 0.0839  | 0.0660  |
| $\hat{b}_5$          | -0.1440                   | 0.1126                     | 0.0855                   | 0.0916                     | 0.6251  | 0.0942  |
| $\hat{b}_6$          | -0.0688                   | -0.0538                    | -0.0476                  | -0.0491                    | -0.0240 | -0.0497 |
| $\hat{b}_7$          | 0.3678                    | 0.0277                     | 0.0929                   | 0.0995                     | -0.6516 | 0.1022  |
| $\hat{b}_8$          | -0.6784                   | 0.0729                     | -0.0897                  | -0.0948                    | 1.5734  | -0.0969 |
| $\hat{b}_9$          | 0.3965                    | -0.1897                    | -0.0830                  | -0.0877                    | -1.3603 | -0.0897 |
| $\hat{b}_{10}$       | -0.0493                   | -0.0961                    | -0.0902                  | -0.0953                    | -0.1897 | -0.0974 |
| $\hat{b}_{11}$       | 0.1373                    | -0.0883                    | -0.0919                  | -0.0981                    | -0.5389 | -0.1007 |
| $R^2$                | 0.5879                    | 0.7625                     | 0.7300                   | 0.7342                     | 0.9020  | 0.7346  |
| HKT                  | 0.4121                    | 0.2375                     | 0.2700                   | 0.2658                     | 0.0980  | 0.2654  |
| HKO( $\hat{\beta}$ ) | 5.9931                    | 11.3769                    | 10.3207                  | 10.3213                    | 14.6368 | 10.3250 |
| PRESS                | 0.6440                    | 0.3486                     | 0.3203                   | 0.3171                     | 0.7102  | 0.3177  |

Tablo 3.2a. ve 3.2b.'den, yanlışlık parametreleri PRESS istatistiği minimize olacak şekilde seçildiğinde, PRESS istatistiğinin yanı sıra, genel olarak, HKT'lerde de küçülme olduğu gözlemlenmektedir.

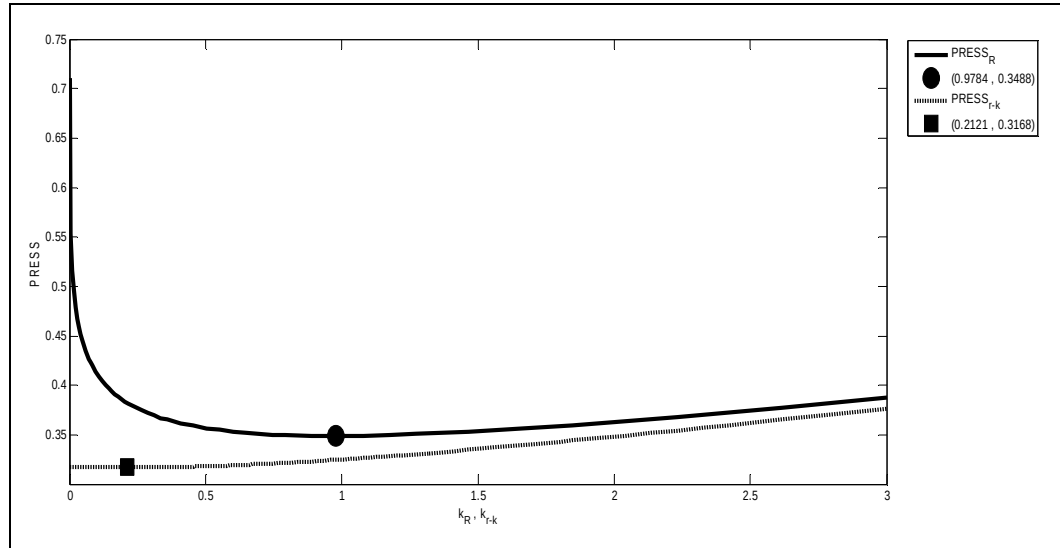
Tablo 3.3. Mil Başına Benzin Tüketimi verisi için EKK, TB ve yanlılık parametreleri PRESS istatistikleri minimize edilerek seçilen Ridge, r-k sınıf, Liu, r-d sınıf tahmin edicileri ile tahmin edilen regresyon modeli için parametre tahminleri,  $R^2$ , HKT, HKO( $\hat{\beta}$ ) ve PRESS istatistikleri

|                      | Ridge                      | r-k sınıf                  | Liu                        | r-d sınıf                  | EKK     | TB      |
|----------------------|----------------------------|----------------------------|----------------------------|----------------------------|---------|---------|
|                      | $\hat{k}_{PRESS} = 0.9784$ | $\hat{k}_{PRESS} = 0.2121$ | $\hat{d}_{PRESS} = 0.0633$ | $\hat{d}_{PRESS} = 0.7942$ |         |         |
| $\hat{b}_1$          | -0.1203                    | -0.1008                    | -0.2116                    | -0.1010                    | -1.5741 | -0.1032 |
| $\hat{b}_2$          | -0.0920                    | -0.0925                    | -0.1149                    | -0.0927                    | -0.4580 | -0.0946 |
| $\hat{b}_3$          | -0.1032                    | -0.1001                    | -0.0014                    | -0.1003                    | 1.4987  | -0.1025 |
| $\hat{b}_4$          | 0.0960                     | 0.0632                     | 0.0945                     | 0.0639                     | 0.0839  | 0.0660  |
| $\hat{b}_5$          | 0.0782                     | 0.0911                     | 0.1126                     | 0.0916                     | 0.6251  | 0.0942  |
| $\hat{b}_6$          | -0.0560                    | -0.0493                    | -0.0538                    | -0.0491                    | -0.0240 | -0.0497 |
| $\hat{b}_7$          | 0.0735                     | 0.0989                     | 0.0277                     | 0.0995                     | -0.6516 | 0.1022  |
| $\hat{b}_8$          | -0.0275                    | -0.0946                    | 0.0729                     | -0.0948                    | 1.5734  | -0.0969 |
| $\hat{b}_9$          | -0.1115                    | -0.0875                    | -0.1897                    | -0.0877                    | -1.3603 | -0.0897 |
| $\hat{b}_{10}$       | -0.0900                    | -0.0951                    | -0.0961                    | -0.0953                    | -0.1897 | -0.0974 |
| $\hat{b}_{11}$       | -0.0576                    | -0.0976                    | -0.0883                    | -0.0981                    | -0.5389 | -0.1007 |
| $R^2$                | 0.7437                     | 0.7340                     | 0.7625                     | 0.7342                     | 0.9020  | 0.7346  |
| HKT                  | 0.2563                     | 0.2660                     | 0.2375                     | 0.2658                     | 0.0980  | 0.2654  |
| HKO( $\hat{\beta}$ ) | 10.0167                    | 10.3232                    | 11.3769                    | 10.3213                    | 14.6368 | 10.3250 |
| PRESS                | 0.3488                     | 0.3168                     | 0.3486                     | 0.3171                     | 0.7102  | 0.3177  |

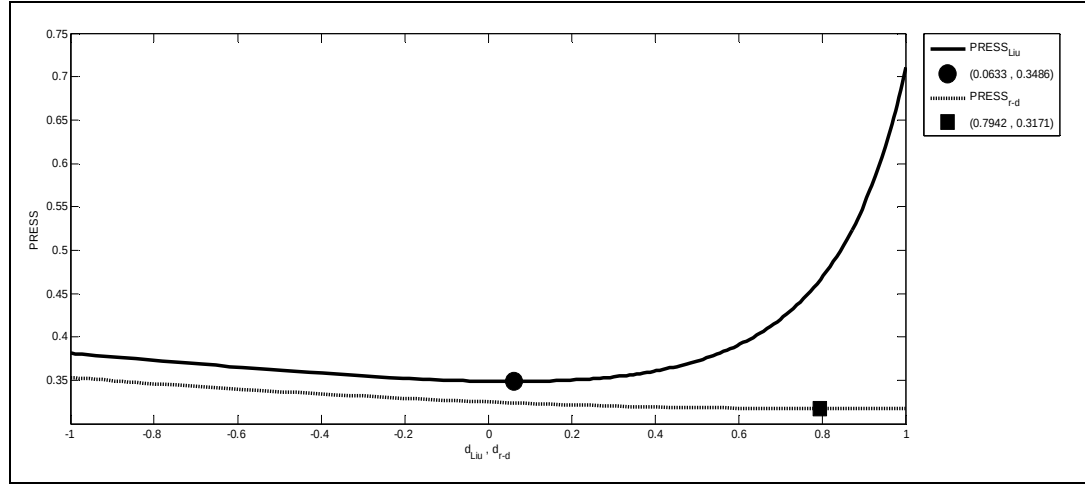
Yukarıdaki tablolardan da açıkça görüldüğü gibi EKK tahmin edicisi en kötü kestirim performansı sergilemektedir. r-k ve r-d sınıf tahmin edicileri en iyi PRESS istatistiklerini vermektedir. TB tahmin edicisi de bu iki tahmin edicinininkine yakın bir PRESS istatistiği vermektedir.



Şekil 3.1a. Mil Başına Benzin Tüketimi verisi için Ridge, Liu, r-d sınıf tahmin edicilerin yanlılık parametreleri ve onlara karşılık gelen PRESS istatistikleri.



Şekil 3.1b. Mil Başına Benzin Tüketimi verisi için Ridge ve r-k sınıf tahmin edicilerin yanlılık parametreleri ve onlara karşılık gelen PRESS istatistikleri.



Şekil 3.1c. Mil Başına Benzin Tüketimi verisi için Liu ve r-d sınıf tahmin edicilerin yanlılık parametreleri ve onlara karşılık gelen PRESS istatistikleri.

Grafikler üzerinde işaretlenen noktalar belirtilen tahmin edicinin PRESS istatistiğinin minimum olduğu nokta ve ona karşılık gelen yanlılık parametresi değeridir.

### 3.4. Sonuç

Montgomery ve Friedman (1993), ridge tahmin edici için, PRESS istatistiği minimize edilerek daha iyi kestirim performansı elde edildiğini göstermişlerdir. Özkale ve Kaçıranlar (2007), kestirim performansını arttırmak için aynı yöntemi Liu tahmin edicisinin yanlılık parametresinin seçiminde önermektedir. Burada biz, en iyi kestirim performansını hedeflediğimizde, r-k sınıf ve r-d sınıf tahmin edicilerde, yanlılık parametresinin PRESS istatistiğini minimize edecek şekilde seçilmesinin, bu yanlı tahmin edicilerin kestirim performansını iyileştirdiğini göstermiş olduk. Ayrıca yanlılık parametreleri PRESS istatistiğini minimum yapacak şekilde seçildiğinde r-k ve r-d sınıf tahmin edicileri ile diğer tahmin edicilere göre daha küçük PRESS istatistikleri elde edilebileceğini de ortaya koyduk. Dolayısıyla, Montgomery ve Friedman (1993) Ridge tahmin edicisi ve Özkale ve Kaçıranlar (2007) Liu tahmin edicisinin yanlılık parametresinin seçiminde önerdikleri yöntemin, r-k sınıf ve r-d sınıf tahmin edicileri için de uygulanmasını önermekteyiz. Böylece hem bu tahmin edicilerin kestirim performansını iyileştirmiş hem de bu tahmin edicilerin kestirim

performanslarını kendi aralarında daha objektif bir şekilde karşılaştırma ve değerlendirme olanağını elde etmiş oluruz.



#### **4. LİU TİPİ VE İKİ PARAMETRELİ RİDGE TAHMİN EDİCİLERİNİN** **KESTİRİM PERFORMANSLARININ PRESS VE GCV İSTATİSTİKLERİ** **KULLANILARAK KARŞILAŞTIRILMASI**

Liu Tipi ve 2 Parametrelili Ridge tahmin edicilerin her ikisi de iki yanlılık parametresi içermektedir. Bunlardan biri  $X'X$  matrisinin kötü koşulluluğunu ortadan kaldırmak, diğeri ise modelin uyumunu sağlamak için kullanılmaktadır. Bu çalışmada, her iki tahmin edicide  $X'X$  matrisinin kötü koşulluluğunu ortadan kaldırmak için kullanılan parametre aynı seçilerek diğeri parametrenin modelin uyumunu sağlamada ne kadar başarılı olduğu araştırılacaktır. Amacımız kestirimin iyiliği olduğundan, modelin uyumunun iyiliğinde kullanılan parametreler, GCV ve PRESS istatistikleri minimize edilerek seçilecektir.

##### **4.1. Giriş**

Yanlı tahmin edicilerdeki yanlılık parametreleri, amaca uygun olarak, katsayıların tahmininde bir gelişme veya kestirimlerde bir iyileşme olacak şekilde seçilebilir. Mümkün olduğunca iyi parametre tahminleri elde edilmek istenildiğinde, çoğunlukla, yanlılık parametresi HKO istatistiğini minimum yapacak şekilde seçilir. Literatürde  $C_p$  istatistiğinin minimizasyonu da önerilmektedir (bkz. Mallows, 1973a,b ve Liu, 1993). Ancak amaç, kestirim olduğunda, yanlılık parametresinin kestirim iyiliğini ölçen kriterlerden birini minimum yapacak şekilde seçilmesi daha uygun olur. Kestirim iyiliğinin ölçümünde kullanılan en yaygın kriterlerden biri, Allen (1971) tarafından önerilen ve PRESS olarak da bilinen CV (Cross-Validation) istatistiğidir. Bir diğeri ise Golub ve diğeri (1979) tarafından önerilen GCV (Generalized Cross-Validation) istatistiğidir. GCV istatistiği, CV istatistiğinin rotasyona duyarlılaştırılmış formudur.

Allen (1974), yanı tahmin ediciler arasından seçim yaparken yanlılık parametrelerini herhangi bir kritere göre belirledikten sonra PRESS istatistiğine bakarak en küçük PRESS istatistiğini veren tahmin edicinin seçilmesini önermiştir.

Montgomery ve Friedman (1993) ise Ridge tahmin edicisinin yanlılık parametresi  $k$  nın seçiminde doğrudan PRESS istatistiğinin kullanımını önermişlerdir. Benzer şekilde, Özkale ve Kaçıranlar (2007) Liu tahmin edicisinin, Özbey ve Kaçıranlar (2010) da  $r-k$  ve  $r-d$  sınıf tahmin edicilerinin yanlılık parametreleri seçiminde bu yöntemi önermişlerdir.

Golub ve diğerleri (1979) GCV istatistiğini geliştirmişlerdir ve onun genel olarak PRESS istatistiğinden daha iyi olduğunu göstermişlerdir. Bu nedenle Liu (1993) ve Lipovetsky (2006) önerdikleri tahmin edicilerin yanlılık parametrelerinin GCV kriterini minimize ederek seçilebileceğini de belirtmişlerdir.

Bu çalışmada iki parametrelili ridge ve Liu tipi tahmin ediciler için 2.yanlılık (yani düzeltme) parametreleri hem PRESS hem de GCV istatistiği kullanılarak seçilecek ve bu iki tahmin edicinin kestirim performansları karşılaştırılacaktır.

#### **4.2. Tahmin Ediciler ve Kestirim İyiliği İstatistikleri**

Liu tahmin edicisi geliştirilerek Liu tipi tahmin edici Liu (2003) tarafından önerilmiştir.

$$\begin{aligned}\hat{\beta}_{LT} &= (\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1}(\mathbf{X}'\mathbf{y} - d\hat{\beta}) \\ &= (\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1}(\mathbf{X}'\mathbf{X} - d\mathbf{I})\hat{\beta} \quad k > 0, \quad -\infty < d < \infty\end{aligned}\tag{4.1}$$

şeklinde tanımlanan bu tahmin edicide  $\hat{\beta}$ ,  $\beta$  'nın herhangi bir tahmin edicisi olabilir. Liu (2003)  $\hat{\beta}$  'nın EKK ve Ridge tahmin edicisi olduğu durumları incelemiştir. Bu çalışmada da  $\hat{\beta}$  EKK tahmin edicisi olarak alınacaktır.  $k$  ve  $d$  ise yanlılık parametreleridir. Liu,  $k$  parametresinin kötü koşullu  $\mathbf{X}'\mathbf{X}$  matrisinin kötü koşulluluğunu ortadan kaldırmak,  $d$  parametresinin ise modeli uyumunu sağlamak için kullanılmasını önermektedir. Bu amaçla  $k$  'nın koşul indeksini 10 yapacak şekilde,  $d$  'nin ise MSE'yi minimum yapacak şekilde seçilebileceğini belirtmektedir.

Ancak daha önceki bir çalışmada (bkz: Liu,1993)  $d$  'nin seçimi için GCV istatistiğinin minimizasyonunu da önermektedir.

Lipovetsky ve Conklin (2005)'in önerdiği 2 Parametrelili Ridge tahmin edici

$$\hat{\beta}_{2PR} = q(\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1}\mathbf{X}'\mathbf{y} \quad (4.2)$$

şeklinde tanımlanmıştır. Burada  $q$  ve  $k$  yanlılık parametreleridir. Hoerl ve Kennard (1970)'ın önerdiği Ridge tahmin edicisini yeni bir yanlılık parametresi olan  $q$  ile çarparak Ridge tahmin edicisinin performansının artırılması hedeflenmektedir. Liu tipi tahmin edicide olduğu gibi,  $k$  'nın kötü koşullu  $\mathbf{X}'\mathbf{X}$  matrisinin kötü koşulluluğunu ortadan kaldırmak,  $q$  parametresinin ise modelin uyumunu sağlamak için kullanılması önerilmektedir. Lipovetsky (2006)  $q$  parametresinin seçiminde çoklu belirleyicilik katsayısı  $R^2$  'nin maksimizasyonunun yanı sıra GCV istatistiğinin minimizasyonunu da önermiştir.

Bu çalışmada her iki tahmin edici için düzeltme parametresi hem PRESS hem de GCV istatistiği minimize edilerek seçilecek ve karşılaştırılmaları ele alınacaktır.

PRESS istatistiğinin genel tanımı (3.1) ile verilmişti. Bu genel tanımdan Liu tipi tahmin edicinin PRESS istatistiği

$$\text{PRESS}_{LT} = \sum_{i=1}^n \left( \frac{y_i - \mathbf{x}_i'(\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1}(\mathbf{X}'\mathbf{X} - d\mathbf{I})(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}}{1 - \mathbf{x}_i'(\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1}(\mathbf{X}'\mathbf{X} - d\mathbf{I})(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i} \right)^2 \quad (4.3)$$

olarak elde edilir. Benzer şekilde iki parametrelili Ridge tahmin edicisinin PRESS istatistiği de

$$\text{PRESS}_{2PR} = \sum_{i=1}^n \left( \frac{y_i - q\mathbf{x}_i'(\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1}\mathbf{X}'\mathbf{y}}{1 - q\mathbf{x}_i'(\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1}\mathbf{x}_i} \right)^2 \quad (4.4)$$

olarak elde edilir.

GCV istatistiği

$$GCV = \frac{n \|(I - H)y\|^2}{(Tr(I - H))^2} \quad (4.5)$$

ile tanımlanır. Buradan Liu tipi tahmin edicinin GCV istatistiği

$$GCV_{LT} = \frac{n \sum_{i=1}^n (y - X(X'X + kI)^{-1}(X'X - dI)(X'X)^{-1}X'y)^2}{(\sum_{i=1}^n (1 - x_i(X'X + kI)^{-1}(X'X - dI)(X'X)^{-1}x'_i))^2} \quad (4.6)$$

olarak elde edilir. Benzer şekilde iki parametrelı Ridge tahmin edicisinin GCV istatistiği de

$$GCV_{2PR} = \frac{n \sum_{i=1}^n (y - qX(X'X + kI)^{-1}X'y)^2}{(\sum_{i=1}^n (1 - qx_i(X'X + kI)^{-1}x'_i))^2} \quad (4.7)$$

şeklinde elde edilir.

Bu çalışmada her iki yanlı tahmin edici için,  $k$  parametresi, Liu (2003) tarafından önerildiği gibi, koşul indeksini 10 yapacak şekilde seçileceğinden  $k$ ,

$$\sqrt{\frac{l_{maks} + k}{l_{min} + k}} = 10 \Rightarrow k = \frac{l_{maks} - 100 l_{min}}{99} \quad (4.8)$$

formülü kullanılarak elde edilecektir. Düzeltme parametreleri ise yukarıda tanımlanan PRESS ve GCV istatistikleri minimize edilerek elde edilecektir. Bu

şartlar altında tahmin edicilerin kestirim performansları, iki farklı veri seti kullanılarak değerlendirilecektir.

### 4.3. Sayısal Örnek

Bu uygulamada kullanılan yanlış tahmin edicilerin yanlışlık parametreleri, modelin hem PRESS istatistiği hem de GCV istatistiği minimize olacak şekilde hesaplanacaktır. Her tahmin edici için mümkün olan en küçük PRESS ve GCV istatistikleri elde edildikten sonra bu istatistikler kendi aralarında karşılaştırılarak, o veri seti için en iyi kestirimi veren tahmin edici seçilecektir.

#### 4.3.1. Mil Başına Benzin Tüketimi Veri Seti

Kullanılacak ilk veri seti 3. Bölümde verilen ve Montgomery ve Friedman (1993)'dan alınan Mil Başına Benzin Tüketimi verisidir.

Tablo 4.1a. Mil Başına Benzin Tüketimi veri seti için EKK, 2 parametrelî Ridge ve Liu tipi tahmin edicileri için parametre tahminleri, HKO ve PRESS istatistikleri

| Tahmin edici | EKK     | Liu Tipi<br>$k = 0.0788, d = 0.0031$ | 2 Parametrelî Ridge<br>$k = 0.0788, q = 0.9411$ |
|--------------|---------|--------------------------------------|-------------------------------------------------|
| $b_1$        | -1.5741 | -0.1150                              | -0.2949                                         |
| $b_2$        | -0.4580 | -0.0906                              | -0.0544                                         |
| $b_3$        | 1.4987  | -0.1077                              | -0.1451                                         |
| $b_4$        | 0.0839  | 0.0953                               | 0.1576                                          |
| $b_5$        | 0.6251  | 0.0763                               | 0.1303                                          |
| $b_6$        | -0.0240 | -0.0559                              | -0.0422                                         |
| $b_7$        | -0.6516 | 0.0758                               | -0.0244                                         |
| $b_8$        | 1.5734  | -0.0334                              | 0.3282                                          |
| $b_9$        | -1.3603 | -0.1067                              | -0.3642                                         |
| $b_{10}$     | -0.1897 | -0.0895                              | -0.0963                                         |
| $b_{11}$     | -0.5389 | -0.0564                              | -0.0395                                         |
| HKO          | 14.6368 | 8.6417                               | 8.1319                                          |
| PRESS        | 0.7102  | 0.4224                               | 0.4169                                          |

**4. LIU TİPİ VE İKİ PARAMETRELİ RIDGE TAHMİN EDİCİLERİNİN  
KESTİRİM PERFORMANSLARININ PRESS VE GCV İSTATİSTİKLERİ  
KULLANILARAK KARŞILAŞTIRILMASI** Fela ÖZBEY

Her iki tahmin edici için  $k$  yanlılık parametresi koşul indeksini 10 yapacak şekilde

$$k = \frac{7.8929 - 100 \times 0.0009}{99} = 0.0788 \quad (4.9)$$

olarak seçilmiştir. Diğer yanlılık parametresi ise modelin hem PRESS istatistikleri hem de GCV istatistikleri minimize olacak şekilde hesaplanmıştır. Sonuçlar Tablo 4.1a. ve Tablo 4.1b. de sunulmuştur.

Tablo 4.1b. Mil Başına Benzin Tüketimi veri seti için EKK, 2 parametrelili Ridge ve Liu tipi tahmin edicileri için parametre tahminleri, HKO ve GCV istatistikleri

| Tahmin edici | EKK     | Liu Tipi<br>$k = 0.0788, d = -0.0298$ | 2 Parametrelili Ridge<br>$k = 0.0788, q = 0.9373$ |
|--------------|---------|---------------------------------------|---------------------------------------------------|
| $b_1$        | -1.5741 | -0.7902                               | -0.2937                                           |
| $b_2$        | -0.4580 | -0.2092                               | -0.0542                                           |
| $b_3$        | 1.4987  | 0.4709                                | -0.1445                                           |
| $b_4$        | 0.0839  | 0.1358                                | 0.1570                                            |
| $b_5$        | 0.6251  | 0.3225                                | 0.1298                                            |
| $b_6$        | -0.0240 | -0.0370                               | -0.0420                                           |
| $b_7$        | -0.6516 | -0.2625                               | -0.0243                                           |
| $b_8$        | 1.5734  | 0.8119                                | 0.3269                                            |
| $b_9$        | -1.3603 | -0.7551                               | -0.3627                                           |
| $b_{10}$     | -0.1897 | -0.1354                               | -0.0959                                           |
| $b_{11}$     | -0.5389 | -0.2299                               | -0.0393                                           |
| HKO          | 14.6368 | 5.5093                                | 8.1387                                            |
| GCV          | 0.2178  | 0.0182                                | 0.0188                                            |

Tablo 4.1a. ve Tablo 4.1b.'den, yanlılık parametreleri PRESS ve GCV istatistikleri minimize olacak şekilde seçildiğinde, bu istatistiklerin yanı sıra, HKO'larda da küçülme olduğu gözlemlenmektedir. Bu sonuçlardan gözlenen diğer önemli bir sonuç şu ki, 2 parametrelili Ridge tahmin edici için hem PRESS hem de GCV istatistiği minimize edildiğinde parametre tahminleri ve MSE fazla değişmemektedir. Ancak Liu tipi tahmin edicide PRESS ve GCV minimizasyonları sonucunda

parametre tahminlerinde ve HKO değerlerinde oldukça büyük bir farklılık gözlenmektedir.

#### 4.3.2. Portland Çimento Veri Seti

Kullanılacak ikinci veri seti uygulamalarda çok yaygın kullanılan Hald (1952)'in Portland Çimento verisidir.

Tablo 4.2. Portland Çimento verisi

| $y$   | $x_1$ | $x_2$ | $x_3$ | $x_4$ |
|-------|-------|-------|-------|-------|
| 78.5  | 7     | 26    | 6     | 60    |
| 74.3  | 1     | 29    | 15    | 52    |
| 104.3 | 11    | 56    | 8     | 20    |
| 87.6  | 11    | 31    | 8     | 47    |
| 95.9  | 7     | 52    | 6     | 33    |
| 109.2 | 11    | 55    | 9     | 22    |
| 102.7 | 3     | 71    | 17    | 6     |
| 72.5  | 1     | 31    | 22    | 44    |
| 93.1  | 2     | 54    | 18    | 22    |
| 115.9 | 21    | 47    | 4     | 26    |
| 83.8  | 1     | 40    | 23    | 34    |
| 113.3 | 11    | 66    | 9     | 12    |
| 109.4 | 10    | 68    | 8     | 12    |

Uygulamada değişkenler standartlaştırılmış olarak kullanılmıştır.  $\mathbf{X}'\mathbf{X}$  matrisinin özdeğerleri  $l_1 = 2.2357$ ,  $l_2 = 1.5761$ ,  $l_3 = 0.188661$ ,  $l_4 = 0.0016237$  'dir. Koşul

sayısı  $k = \frac{l_1}{l_4} \cong 1376.917$  çoklu doğrusal bağlantının olduğunu gösterir.

Bu veri setinde, Liu Tipi ve 2 Parametrelı Ridge tahmin edicileri için  $k$  parametresi

$$k = \frac{2.2357 - 100 \times 0.0016237}{99} = 0.0209$$

**4. LİU TİPİ VE İKİ PARAMETRELİ RIDGE TAHMİN EDİCİLERİNİN  
KESTİRİM PERFORMANSLARININ PRESS VE GCV İSTATİSTİKLERİ  
KULLANILARAK KARŞILAŞTIRILMASI** Fela ÖZBEY

olarak seçilmiştir. Düzeltme amaçlı kullanılan parametreler, hem PRESS istatistikleri hem de GCV istatistikleri minimize olacak şekilde hesaplanmıştır. Sonuçlar aşağıdaki tablolarda sunulmuştur.

Tablo 4.3a. Portland Çimento veri seti için EKK, 2 parametrel Ridge ve Liu tipi tahmin edicileri için parametre tahminleri, HKO ve PRESS istatistikleri

| Tahmin Edici | EKK     | 2 Parametrel Ridge<br>$k = 0.0209, q = 1.0162$ | Liu Tipi<br>$k = 0.0209, d = 0.0158$ |
|--------------|---------|------------------------------------------------|--------------------------------------|
| $b_1$        | 0.6065  | 0.5043                                         | 0.2401                               |
| $b_2$        | 0.5277  | 0.3076                                         | 0.2390                               |
| $b_3$        | 0.0434  | -0.0716                                        | -0.1482                              |
| $b_4$        | -0.1603 | -0.4010                                        | -0.2615                              |
| HKO          | 1.2186  | 0.1474                                         | 0.8862                               |
| PRESS        | 0.0321  | 0.0274                                         | 0.0274                               |

Tablo 4.3b. Portland Çimento veri seti için EKK, 2 parametrel Ridge ve Liu tipi tahmin edicileri için parametre tahminleri, HKO ve GCV istatistikleri

| Tahmin Edici | EKK     | 2 Parametrel Ridge<br>$k = 0.0209, q = 1.0053$ | Liu Tipi<br>$k = 0.0209, d = 0.0712$ |
|--------------|---------|------------------------------------------------|--------------------------------------|
| $b_1$        | 0.6065  | 0.4989                                         | 0.1211                               |
| $b_2$        | 0.5277  | 0.3043                                         | -0.4625                              |
| $b_3$        | 0.0434  | -0.0708                                        | -0.4576                              |
| $b_4$        | -0.1603 | -0.3967                                        | -1.1916                              |
| HKO          | 1.2186  | 0.1474                                         | 14.0094                              |
| GCV          | 0.0255  | 0.0023                                         | 0.0019                               |

Elde edilen sonuçlar incelendiğinde, Liu tipi tahmin edicinin GCV ve PRESS istatistiklerinin minimizasyonu ile elde edilen  $d$  parametrelerinin çok farklı olduğu, bu nedenle de elde edilen parametre tahminleri ve HKO değerleri arasında büyük farklılıklar olduğu gözlemlenmektedir. Hatta GCV minimizasyonu sonucunda elde edilen tahminlerin HKO değeri EKK tahmin edicisinden kat kat daha büyüktür. İki parametrel ridge tahmin edicisi için GCV ve PRESS istatistikleri minimize edilerek elde edilen parametre tahminleri ve HKO değerlerinde ise pek farklılık yoktur. Bu,



iki parametrelili ridge tahmin edicisinin daha istikrarlı tahminler verdiği yönünde bir görüş oluşturmaktadır.

#### **4.4. Sonuç**

Liu Tipi ve 2 Parametrelili Ridge tahmin edicilerinin her ikisi de iki yanlılık parametresi içermektedir. Bunlardan biri  $X'X$  matrisinin kötü koşulluluğunu ortadan kaldırmak, diğeri ise modelin uyumunu sağlamak için kullanılmaktadır. Bu çalışmada, her iki tahmin edicide  $X'X$  matrisinin kötü koşulluluğunu ortadan kaldırmak için kullanılan parametre aynı seçilerek diğeri parametrenin modelin uyumunu sağlamakta ne kadar başarılı olduğu araştırılmıştır. Amacımız kestirimin iyiliği olduğundan da modelin uyumunu sağlamakta kullanılan parametreler GCV ve PRESS istatistikleri minimize edilerek seçilmiştir.

Yapılan nümerik çalışmada, elde edilen sonuçlar, genel olarak Liu tipi tahmin edici ve iki parametrelili ridge tahmin edicinin her ikisi ile elde edilen tahminlerde, sadece GCV ve PRESS istatistiklerinde değil, aynı zamanda HKO değerlerinde de belirgin iyileşmeler elde edilebileceğini göstermektedir. Liu tipi tahmin edicinin GCV ve PRESS istatistiklerinin minimizasyonu ile elde edilen  $d$  parametrelerinin oldukça farklı seçildiği, bu nedenle de elde edilen parametre tahminleri ve HKO değerleri arasında büyük farklılıklar olduğu gözlemlenmiştir. Hatta bazı durumlarda EKK tahmin edicisinden daha büyük HKO değerleri elde edilebileceği görülmüştür. İki parametrelili ridge tahmin edicisi için GCV ve PRESS istatistikleri minimize edilerek elde edilen parametre tahminleri ve HKO değerlerinde ise pek farklılık yoktur. Bu, iki parametrelili ridge tahmin edicisinin daha istikrarlı tahminler verdiği yönünde bir görüş oluşturmaktadır.

## 5. SONUÇLAR VE ÖNERİLER

Bu çalışmada benzer özellikler taşıyan ve birbirlerine alternatif olabilecek bazı yanlı tahmin ediciler, kestirim performansları bakımından birbirleriyle karşılaştırılmıştır.

İlk aşamada Liu tahmin edicinin EKK, TB ve Ridge tahmin edicilerine kıyasla, iki boyutlu açıklayıcı değişkenler uzayında belirli bir noktadaki KHKO cinsinden kestirim performansı değerlendirilmiştir. Bu yöntem, Friedman ve Montgomery (1985) tarafından önerilmiş ve EKK, TB ve Ridge tahmin edicilerinin kestirim performanslarını karşılaştırmak için kullanılmıştır. Çalışmanın bir bölümünde, Friedman ve Montgomery'nin yaptığı karşılaştırmalarda eksik görülen bazı noktalara değinilmiş ve bu karşılaştırmalar üzerine bazı yorumlar getirilmiştir. Daha sonra Liu tahmin edicinin kestirim performansının değerlendirilmesine geçilmiştir. Bu bağlamda, öncelikle Liu tahmin edicinin KHKO'su elde edilmiş; devamında Liu tahmin edicinin EKK, TB ve Ridge tahmin edicilerine göre KHKO anlamında daha iyi olması için gerek ve yeter koşullar üç teorem altında verilmiştir. Elde edilen teorik sonuçlar sayısal bir örnek üzerinde gösterilmiş ve adı geçen tahmin edicilerin üstün olduğu bölgeler belirlenmiştir. Kestirim için uygun olan tahmin edicinin seçimi, kestirilecek noktanın konumuna olduğu gibi çoklu iç ilişkinin derecesine de bağlı olduğu görülmüştür. Verilen sayısal örnekte, Liu tahmin edicinin EKK, TB ve Ridge tahmin edicilerinden daima üstün olduğu bir bölge belirlenmiştir. Bu, böyle bir alanın belirlenmesinin teorik olarak mümkün olduğunu göstermektedir.

İkinci aşamada, en iyi kestirim performansı hedeflendiğinde,  $r$ - $k$  sınıf ve  $r$ - $d$  sınıf tahmin edicilerde, yanlılık parametresinin PRESS istatistiğini minimize edecek şekilde seçilmesinin, bu yanlı tahmin edicilerin kestirim performansını iyileştirdiği gösterilmiştir. Ayrıca yanlılık parametreleri PRESS istatistiğini minimum yapacak şekilde seçildiğinde  $r$ - $k$  ve  $r$ - $d$  sınıf tahmin edicileri ile ihtiva ettikleri diğer tahmin edicilere göre daha küçük PRESS istatistikleri elde edilebileceği de ortaya konulmuştur. Dolayısıyla, amaç iyi bir kestirim elde etmek olduğunda, bu yöntemin  $r$ - $k$  sınıf ve  $r$ - $d$  sınıf tahmin edicileri için uygulanması önerilmektedir. Böylece hem bu tahmin edicilerin kestirim performansları iyileştirilmiş hem de bu tahmin

edicilerin kestirim performanslarının kendi aralarında daha objektif bir şekilde karşılaştırılması ve değerlendirilmesi olanağı elde edilmiş olur.

Son olarak, Liu Tipi ve 2 Parametrelili Ridge tahmin edicilerinin kestirim performansları bakımından karşılaştırılmıştır. Liu Tipi ve 2 Parametrelili Ridge tahmin edicilerinin her ikisi de iki yanlılık parametresi içermektedir. Bunlardan biri  $X'X$  matrisinin kötü koşulluluğunu ortadan kaldırmak, diğeri ise modelin uyumunu sağlamak için kullanılmaktadır. Bu çalışmanın 4.Bölümünde, her iki tahmin edicide  $X'X$  matrisinin kötü koşulluluğunu ortadan kaldırmak için kullanılan parametre aynı seçilerek diğeri parametrenin modelin uyumunu sağlamakta ne kadar başarılı olduğu araştırılmıştır. Amacımız kestirimin iyiliği olduğundan da modelin uyumunu sağlamakta kullanılan parametreler GCV ve PRESS istatistikleri minimize edilerek seçilmiştir. Yapılan nümerik çalışmada, elde edilen sonuçlar, genel olarak Liu tipi tahmin edici ve iki parametrelili ridge tahmin edicinin her ikisi ile elde edilen tahminlerde, sadece GCV ve PRESS istatistiklerinde değil, aynı zamanda HKO değerlerinde de belirgin iyileşmeler elde edilebileceğini göstermektedir. Liu tipi tahmin edicinin GCV ve PRESS istatistiklerinin minimizasyonu ile elde edilen  $d$  parametrelerinin oldukça farklı seçildiği, bu nedenle de elde edilen parametre tahminleri ve HKO değerleri arasında büyük farklılıklar olduğu gözlemlenmiştir. Hatta bazı durumlarda EKK tahmin edicisinden daha büyük HKO değerleri elde edilebileceği görülmüştür. İki parametrelili ridge tahmin edicisi için GCV ve PRESS istatistikleri minimize edilerek elde edilen parametre tahminleri ve HKO değerlerinde ise pek farklılık yoktur. Bu, iki parametrelili ridge tahmin edicisinin daha istikrarlı öngörüler verdiği yönünde bir görüş oluşturmaktadır.



## KAYNAKÇA

- AKDENİZ F, KAÇIRANLAR S., 1995. On the almost unbiased generalized Liu estimator and unbiased estimation of the bias and MSE. *Communications in Statistics-Theory and Methods*, 24, 1789-1797.
- ALLEN, D. M. (1971), Mean Square Error of Prediction as a Criterion for selection of Variables, *Technometrics*, 13, 469-475.
- ALLEN, D. M. (1974), The Relationship Between Variable Selection and Data Augmentation and a Method for Prediction, *Technometrics*, 16, 125-127.
- BAYE, M. R. ve PARKER D. F. (1984), Combining Ridge and Principal Component Regression: A Money Demand Illustration, *Comm. Statist. Theory Methods*, 13 (2), 197-205.
- FRIEDMAN D. J., MONTGOMERY D. C. 1985. Evaluation of the predictive performance of biased regression estimators. *Journal of Forecasting*, 4, 153-163.
- GOLUB, G. H., HEATH, M. ve WAHBA, G., 1979. Generalized Cross-Validation as a Method for Choosing a Good Ridge Parameter. *Technometrics*, 21, 215-223
- GRUBER, M. H. J., 1998. Improving efficiency by shrinkage: the James-Stein and Ridge regression estimators. Marcel Dekker: New York.
- GUNST, R. F ve MASON R. L., 1979. Some considerations in the evaluation of alternate prediction equations. *Technometrics*, 21, 55-63.
- HALD, A., 1952. *Statistical Theory with Engineering Applications*. New York: Wiley.
- HOERL, A. E ve KENNARD, R. W., 1970. Ridge regression: biased estimation for nonorthogonal problems, *Technometrics*, 12 (1), 55-76.
- KAÇIRANLAR, S. VE SAKALLIOĞLU S., (2001), Combining the Liu Estimator and the Principal Component Regression Estimator, *Comm. Statist. Theory Methods*, 30 (12), 2699-2705.

- LIPOVETSTY, S. 2006. Two-Parameter Ridge Regression and its Convergence to the Eventual Pairwise Model. *Mathematical and Computer Modelling*, 44, 304-318
- LIPOVETSKY, S. ve W.M.CONKLIN. 2005. "Ridge Regression in Two-Parameter Solution. *Applied Stochastic Models in Business and Industry*. c.21 sf:525-540.
- LIU,K., 1993. A New Class of Biased Estimate in Linear Regression. *Communications in Statistics – Theory and Methods*., c:22 (2), sf:393-402
- LIU,K., 2003. Using Liu - Type Estimator to Combat Collinearity. *Communications in Statistics – Theory and Methods*., c:32 (5), sf:1009-1020.
- MALLOWS, C.L., 1973a. Data Analysis in a Regression Context. University of Kentucky Conference on Regression with a Large Number of Predictor Variables, Department of Statistics, University of Kentucky.
- MALLOWS, C.L., 1973b. Some Comments on  $C_p$ . *Technometrics*, 15, 661-675
- MASSY WF. 1965. "Principal components regression in exploratory Statistical Research. " *Journal of the American Statistical Association*, 60, 234-256.
- MONTGOMERY, D.C. ve D.J.FRIEDMAN, 1993. Prediction Using Regression Models with Multicollinear Predictor Variables. *IIE Transactions*, 25 (3), 73-85
- NOMURA, M. ve T. OHKUBO, (1985). A Note on Combining Ridge and Principal Component Regression, *Communications in Statistics – Theory and Methods*, 14, 2489-2493.
- ÖZKALE, M. R. ve S. KAÇIRANLAR (2007), A Prediction-Oriented Criterion for Choosing the Biasing Parameter in Liu Estimation, *Communications in Statistics – Theory and Methods*, 36, 1889-1903.
- ÖZKALE, M. R. ve S. KAÇIRANLAR (2008), Comparisons of the r-k Class Estimator to the Ordinary Least Squares Estimator under the Pitman's Closeness Criterion, *Statistical Papers*, 49, sf.503-512
- ÖZBEY, F. ve KAÇIRANLAR, S., 2010. Bazı Yanlı Tahmin Edicilerin Kestirim Performanslarının Karşılaştırılması. X. Ekonometri ve İstatistik Sempozyumu, Sakarya.

- SAKALLIOĞLU S, KAÇIRANLAR S, AKDENİZ F. 2001. Mean squared error comparisons of some biased regression estimator. Communications in Statistics-Theory and Methods, 30, 347-361.
- SARKAR, N. (1989). Comparisions Among Some Estimators in Misspecified Linear Models with Multicollinearity, Ann. Inst. Statist. Math., 41, 717-724.
- SARKAR, N. (1996). Mean Square Error Matrix Comparision of Some Estimators in Linear Regressions with Multicollinearity, Statistics and Probability Letters, 30, 133-138
- STEIN C. 1956. Inadmissibility of usual estimator for the mean of a multivariate normal distribution. Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 1, 197-206.





## ÖZGEÇMİŞ

1978 yılında Bulgaristan’da doğdu. İlk, orta ve lise öğrenimini Adana’da tamamladı. Lisans öğrenimine devam ederken 2001 yılında İstanbul Menkul Kıymetler Borsası’nda bir dönem stajyer olarak bulundu. 2002 yılında Çukurova Üniversitesi İktisadi ve İdari Bilimler Fakültesi Ekonometri Bölümü’nden mezun oldu. Aynı yıl Çukurova Üniversitesi Sosyal Bilimler Enstitüsü Ekonometri Anabilim Dalı’nda Araştırma Görevlisi olarak göreve başladı. 2005 yılında Çukurova Üniversitesi Sosyal Bilimler Enstitüsü Ekonometri Anabilim Dalı’nda Yüksek Lisansını tamamladı ve Çukurova Üniversitesi İktisadi ve İdari Bilimler Fakültesi Ekonometri Bölümü’nde Araştırma Görevlisi olarak çalışmaya başladı. 2006 yılında, Bulgar hükümetinin bursu ile Sofya St. Kliment Ohridski Üniversitesi’nin Slav dilleri eğitim seminerlerine katıldı. Yine 2006 yılında Çukurova Üniversitesi Fen Bilimler Enstitüsü İstatistik Anabilim Dalı’nda Doktora öğrenimine başladı.