

M. TÜRKMEN ŞAHİNGÖZ

İSTATİSTİK ANABİLİM DALI

YÜKSEK LİSANS TEZİ

T.C.

MUĞLA SITKI KOÇMAN ÜNİVERSİTESİ

FEN BİLİMLERİ ENSTİTÜSÜ

İSTATİSTİK ANABİLİM DALI

REGRESYON ANALİZİNDE YANLI TAHMİN YÖNTEMLERİ

YÜKSEK LİSANS TEZİ

MERVE TÜRKMEN ŞAHİNGÖZ

HAZİRAN 2019

MUĞLA

T.C.
MUĞLA SITKI KOÇMAN ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

İSTATİSTİK ANABİLİM DALI

REGRESYON ANALİZİNDE YANLI TAHMİN YÖNTEMLERİ

YÜKSEK LİSANS TEZİ

MERVE TÜRKMEN ŞAHİNGÖZ

HAZİRAN 2019

MUĞLA

MUĞLA SITKI KOÇMAN ÜNİVERSİTESİ

Fen Bilimleri Enstitüsü

TEZ ONAYI

MERVE TÜRKMEN ŞAHİNGÖZ tarafından hazırlanan **REGRESYON ANALİZİNDE YANLI TAHMİN YÖNTEMLERİ** başlıklı tezinin, 28.06.2019 tarihinde aşağıdaki jüri tarafından İstatistik Anabilim Dalı'nda yüksek lisans derecesi için gerekli şartları sağladığı oybirliği/oyçokluğu ile kabul edilmiştir.

TEZ SINAV JURİSİ

Doç. Dr. Öznur İŞÇİ GÜNERİ (**Jüri Başkanı**)

İstatistik Anabilim Dalı

Muğla Sıtkı Koçman Üniversitesi, Muğla

İmza: 

Prof. Dr. Dursun AYDIN (**Danışman**)

İstatistik Anabilim Dalı

Muğla Sıtkı Koçman Üniversitesi, Muğla

İmza: 

Doç. Dr. Kurtuluş BOZKURT (**Üye**)

İktisadi Gelişme ve Uluslararası İktisat Anabilim Dalı

Adnan Menderes Üniversitesi, Söke İşletme Fakültesi, Aydın

İmza: 

ANA BİLİM DALI BAŞKANLIĞI ONAYI

Prof. Dr. Dursun AYDIN

İstatistik Ana Bilim Dalı Başkanı,

Muğla Sıtkı Koçman Üniversitesi, Muğla

İmza: 

Prof. Dr. Dursun AYDIN

Danışman, İstatistik Anabilim Dalı,

Muğla Sıtkı Koçman Üniversitesi, Muğla

İmza: 

Savunma Tarihi: 28/06/2019

Tez çalışmalarım sırasında elde ettiğim ve sunduğum tüm sonuç, doküman, bilgi ve belgelerin tarafımdan bizzat ve bu tez çalışması kapsamında elde edildiğini; akademik ve bilimsel etik kurallarına uygun olduğunu beyan ederim. Ayrıca, akademik ve bilimsel etik kuralları gereği bu tez çalışması sırasında elde edilmemiş başkalarına ait tüm orijinal bilgi ve sonuçlara atıf yapıldığını da beyan ederim.

Merve TÜRKMEN ŞAHİNGÖZ

28/06/2019



ÖZET

REGRESYON ANALİZİNDE YANLI TAHMİN YÖNTEMLERİ

Merve TÜRKMEN ŞAHİNGÖZ

Yüksek Lisans Tezi

Fen Bilimleri Enstitüsü

İstatistik Anabilim Dalı

Danışman: Prof. Dr. Dursun AYDIN

Haziran 2019, 67 sayfa

Bağımsız değişkenler arasında çoklu bağlantı olması durumunda model parametrelerini tahmin etmek için kullanılan kestirim yöntemlerinden Ridge Regresyon yöntemi ele alınmıştır. Bu yöntem EKK yöntemleriyle elde edilen sonuçların çoklu bağlantıdan dolayı geçerliliğini kaybettiğinde yani çoklu doğrusal bağlantı içeren verilerde regresyon katsayılarının hataları, işaretleri ve büyüklüklerini tahmin etmek mümkün değil iken kullanılabilir. Ridge parametresinin çözümü k 'ya bağlı olarak yapılır ve dolayısıyla her k için Ridge Regresyon katsayı değerleri elde edilmelidir. Ridge parametresinin seçimi, (i) subjektif metotlar ve (ii) objektif metotlar olmak üzere iki sınıfta toplanabilir.

Bu tezde k parametresini seçmek için objektif yöntemler dikkate alınmıştır. Ancak gerçek veri uygulamasında subjektif yöntemlerden ridge izi grafiğine yer verilmiştir. Bu bağlamda, tezde ridge parametresinin seçimi için kullanılan objektif yöntemler arasında yer alan klasik bir seçim yöntemi olan k_{HKB} kestirici baz kriteri olarak alınmıştır. Bu kritere ilaveten dört farklı bilgi kriteri (Mallows'un C_k kriteri, Genelleştirilmiş çapraz geçerlilik kriteri, Düzeltilmiş akaike bilgi kriteri, Schwarz'ın bayes bilgi kriteri) kullanılarak ridge parametresi seçilmiş ve bu optimum k değeri kullanılarak elde edilen Ridge Regresyon sonuçlarının bir karşılaştırılması yapılmıştır.

Tezde iki adet uygulama yapılmış olup, bu uygulamalardan sonuçlar R paket programı yardımı ile elde edilmiştir.

Anahtar Kelimeler: Çoklu Doğrusal Bağlantı, Ridge Regresyon.

ABSTRACT

ESTIMATION METHODS IN REGRESSION ANALYSIS

Merve TÜRKMEN ŞAHİNGÖZ

Master of Science (M.Sc.)

Graduate School of Natural and Applied Sciences

Department of Statistics

Supervisor: Prof. Dr. Dursun AYDIN

June 2019, 67 pages

Ridge Regression method, which is one of the estimation methods used to estimate the model parameters, is discussed in case of multiple connections between independent variables. This methods can be used when the results obtained by OLS methods lose their validity due to multiple connections, that is, it is not possible to estimate the errors, signs and magnitudes of regression coefficients in data containing multiple linear connections.

The solution of the Ridge parameter is based on k , and therefore the Ridge Regression coefficient values should be obtained for each k . The choice of the Ridge parameter can be divided into two classes: (i) subjective methods and (ii) objective methods.

In this thesis, for the selection of the k parameter objective methods are used. However, ridge trace graph, which is one of the subjective methods, is used in real data application. In this context, in the thesis, the predictive basis criterion, which is a classical selection method, one of the objective methods used for the selection of ridge parameter, is taken as the criterion. In addition to this criterion, ridge parameter was selected using four different information criteria (Mallows criterion, Generalized cross validity criterion, Corrected acoustic data criterion, Schwarz's bayes information criterion) and a comparison of Ridge Regression results was obtained using this optimum k value.

Two applications have been made in this thesis and the results have been obtained with the help of R package program.

Keywords: Multi-collinearity, Ridge Regression.

ÖNSÖZ

Tezin her aşamasında bilgisini, desteğini, hoşgörüsünü esirgemeyen kıymetli danışmanım Prof. Dr. Dursun AYDIN'a sonsuz teşekkürlerimi ve saygılarımı sunarım.

Tez çalışmam boyunca yardımlarını esirgemeyen Arş. Gör. Ersin YILMAZ'a teşekkürlerimi sunarım.

Son olarak beni maddi ve manevi her konuda destekleyen değerli aileme ve eşim Burak ŞAHİNGÖZ'e sevgilerimi sunarım.



İÇİNDEKİLER

ÖNSÖZ.....	vi
İÇİNDEKİLER	viii
ÇİZELGELER DİZİNİ	viii
ŞEKİLLER DİZİNİ	ix
SEMBOLLER VE KISALTMALAR DİZİNİ	x
1. GİRİŞ	1
2. ÇOKLU DOĞRUSAL BAĞLANTI (MULTI-COLLINEARİTY) PROBLEMİ.....	3
2.1. Çıkarılma Üzerine Etkileri	3
2.2. Çoklu Bağlantının Tespit Edilmesi	6
2.3. Çoklu Bağlantı Probleminin Çözümü	8
3. RIDGE REGRESYONU	9
3.1. Ridge Regresyon ile EKK regresyonu arasındaki ilişki.....	10
3.2. Ridge Regresyon Kestircisinin Hesaplanması	12
3.3. Ridge Regresyon Kestircisinin Geometrik Yorumu	15
3.4. Ridge Kestircisinin Özellikleri.....	16
3.5. Model Varyansının Tahmini	21
3.6. Ridge Regresyonuna İlişkin Bazı Bilgiler.....	22
3.7. Ridge Parametresinin Seçimi	23
3.7.1. HKB (Hoerl, Kennard ve Baldwin) kestircisi	25
3.7.2. Mallows'un k kriteri:.....	26
3.7.3. Çapraz geçerlilik kriteri	27
3.7.4. Genelleştirilmiş çapraz geçerlilik kriteri	27
3.7.5. Düzeltilmiş akaike bilgi kriteri	28
3.7.6. Schwarz'ın bayes bilgi kriteri.....	28
4. UYGULAMA.....	29
4.1 Monte Carlo Simülasyon Çalışması.....	29
4.2 Gerçek Veri Çalışması	39
5. SONUÇ VE ÖNERİLER.....	44
KAYNAKLAR.....	46
EKLER.....	51
ÖZGEÇMİŞ.....	67

ÇİZELGELER DİZİNİ

Tablo 4.1. Her bir seçim yöntemi için $\rho=0.60$ olduğunda regresyon katsayısı tahminleri	30
Tablo 4.2. Her bir seçim yöntemi için $\rho=0.80$ olduğunda regresyon katsayısı tahminleri	31
Tablo 4.3. Her bir seçim yöntemi için $\rho=0.99$ olduğunda regresyon katsayısı tahminleri	32
Tablo 4.4. Tüm durumlar için MSE değerleri.....	38
Tablo 4.5. Korelasyon düzeyleri 0.80 ve 0.99 için yöntemlerin MSE değerleri.....	39
Tablo 4.6. VIF değerleri.....	40
Tablo 4.7. Koşul İndeksi ve Varyans Ayrıştırma Oranları	41
Tablo 4.8. Tüm yöntemler için MSE değerleri ve katsayılar	41

ŞEKİLLER DİZİNİ

Şekil 3.1.iki boyutlu uzayda Ridge Regresyon kestiricisinin gösterimi	15
Şekil 4.1.Korelasyon düzeyi $\rho=0.60$ için Ridge parametresinin bilgi kriterleri ve beş farklı Ridge kestirici için ortalama MSE değerlerinin kutu grafikleri	33
Şekil 4.2.Korelasyon düzeyi $\rho=0.80$ için Ridge kestiricilerine dayalı olarak elde edilen ortalama MSE değerlerinin kutu grafikleri	34
Şekil 4.3.Korelasyon düzeyi $\rho=0.80$ için Ridge parametresinin bilgi kriterleri ve beş farklı Ridge kestiricisinin varyanslarının kutu grafikleri	35
Şekil 4.4.Korelasyon düzeyi $\rho=0.80$ için Ridge parametresinin bilgi kriterleri ve beş farklı Ridge kestiricisinin yanlarının karelerini gösteren kutu grafikleri.....	35
Şekil 4.5.Korelasyon düzeyi $\rho=0.99$ için ortalama MSE değerlerinin kutu grafikleri	36
Şekil 4.6. Korelasyon düzeyi $\rho=0.99$ için Ridge parametresinin bilgi kriterleri beş farklı Ridge kestiricisinin varyanslarının kutu grafikleri.....	37
Şekil 4.7. Korelasyon düzeyi $\rho=0.99$ için Ridge parametresinin bilgi kriterleri ve beş farklı Ridge kestiricisinin yanlarının karelerini gösteren kutu grafikleri.....	37
Şekil 4.8. Açıklayıcı değişkenler için korelasyon matrisi.....	40
Şekil 4.9. VIF izi grafiği	42
Şekil 4.10. Ridge izi grafiği	43

SEMBOLLER VE KISALTMALAR DİZİNİ

EKK	Sıradan En küçük Kareler
<	Küçük
>	Büyük
–	Eksi işareti
β	Beta
ε	Epsilon
RR	Ridge Regresyon
σ	Sigma
<i>Var</i>	Varyans
<i>Cov</i>	Kovaryans
<i>VIF</i>	Varyans artırıcı faktör
λ	Lamda
<i>KI</i>	Koşul İndeksi
<i>TV</i>	Tolerans Değeri
CHKT	Cezalı Hata Kareler Toplamını
α	Alfa
<i>MSE</i>	Hata Kareler Ortalaması
<i>RSS</i>	Artık Kareler Toplamı
<i>E</i>	Beklenen Değer
<i>DF</i>	Serbestlik Derecesi
<i>EDF</i>	Etkin Serbestlik Derecesi
<i>SSE</i>	Hata Kareler Toplamı
<i>HKB</i>	Hoerl, Kennard ve Baldwin Kriteri
<i>CV</i>	Çapraz Geçerlilik Kriteri
<i>GCV</i>	Genelleştirilmiş Çapraz Geçerlilik Kriteri
<i>AIC</i>	Akaike Bilgi Kriteri
<i>BIC</i>	Bayes Bilgi Kriteri
<i>SIC</i>	Schwarz Bilgi kriteri
<i>HKO</i>	Hata Kareler Ortalaması

1. GİRİŞ

Tasarlanmış bir deney ya da gözlemsel bir çalışmadan toplanan verileri temsil eden açıklayıcı değişkenler birbirleriyle ilişkili olduklarında, sıradan en küçük kareler (EKK) yöntemi, herhangi bir açıklayıcı değişkenin yanıt değişkeni üzerindeki etkisini belirten kesin tahminleri sağlamaz. Aşağıdaki gibi çoklu doğrusal regresyon modelini dikkate aldığımızı varsayalım:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (1.1)$$

Burada \mathbf{Y} , yanıt değişkeninin aldığı değerleri belirten $(n \times 1)$ boyutlu bir vektör; \mathbf{X} , $(n \times p)$ boyutlu bilinen tasarım (bağımsız değişkenler) matrisi; $\boldsymbol{\beta}$, $(p \times 1)$ boyutlu bilinmeyen regresyon parametreleri vektörü ve $\boldsymbol{\varepsilon}$, $(n \times 1)$ boyutlu sıfır ortalama ve $\sigma^2\mathbf{I}$ (burada \mathbf{I} ise, $(n \times n)$ boyutlu birim matris) varyanslı rassal hatalar vektörüdür. Regresyon parametreler vektörü $\boldsymbol{\beta}$ 'nin sıradan EKK kestiricisi,

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \quad (1.2)$$

olarak tanımlanır. Burada belirtmek gerekirse, uygun bir $\hat{\boldsymbol{\beta}}$ kestiricisinin elde edilmesi $(\mathbf{X}'\mathbf{X})$ matrisinin özelliklerine bağlıdır. Eğer $(\mathbf{X}'\mathbf{X})$ kötü koşullu (ill-conditioned) yani, diğer bir deyişle, $(\mathbf{X}'\mathbf{X})$ 'in değişik sütunları arasında yakın bağıllık veya $\det(\mathbf{X}'\mathbf{X}) \cong 0$ ise, EKK tahminleri, yanlış işaretli ve anlamlı veya kesin olmayan regresyon katsayıları gibi bir dizi hataya duyarlı olurlar. Bununla birlikte EKK yöntemi, yüksek varyanslı, büyük standart hatalı ve geniş güven aralıklı tahminler verir. Açıklayıcı değişkenlerin ilişkili olduğu durumda, EKK'nın düzensiz davranışı nedeniyle tahmin edilen modelin kalitesi ve kararlılığı sorgulanabilir.

Bu tür problemlere çözüm bulabilmek için, araştırmacılar bilinçli olarak modelden açıklayıcı değişkenleri kaldırarak soruna neden olan değişkenleri elemine edilebilir. Ancak, modelden ilgili açıklayıcı değişkenleri atma yöntemi, modelin faydasına zarar verebilir. EKK tahminlerinin değişkenliğini ve kararsızlığını kontrol etmek için, EKK'ya alternatif olarak Ridge Regresyonu (RR), Temel bileşenler ve Lasso regresyon yöntemleri gibi bazı regülasyon metotları ile katsayılar düzenlenebilir. Bu çalışmada RR yöntemi dikkate alınmıştır. Cebirsel olarak RR yöntemi, değişkenler

arası ilişkinin etkilerini ortadan kaldırır ve EKK'nın tahmin ettiğinden daha kararlı tahminler elde etmek için katsayılar arasındaki korelasyonun görünür büyüklüğünü azaltır ve ayrıca tahmin doğruluğunu artırır (bkz. Hoerl ve Kennard, 1970a; Montgomery ve Peck., 1982; Myers, 1986; Seber ve Lee, 2003).

Açıklayıcı değişkenler arasında güçlü korelasyonlar olduğunda, $(\mathbf{X}'\mathbf{X})$ matrisinin sütunlarının yaklaşık doğrusal bir bağımlılığı olur ve $(\mathbf{X}'\mathbf{X})^{-1}$ terimi tekile yaklaşır. Sonuç olarak, en küçük kareler tahmini, gözlenen yanıt \mathbf{Y} 'de rastgele hatalara karşı oldukça hassas hale gelmekte ve büyük bir fark yaratmaktadır. Bu durum çoklu bağlantı olarak tanımlanır ve bir değişkende yer alan bilgilerin aslında veri setinde diğer değişkenler aracılığıyla zaten mevcut olduğu anlamına gelir. Çoklu bağlantı olduğunda en küçük kareler tahminleri yansızdır, ancak varyansları büyük olur. Bu nedenle gerçek değerden uzak olabilirler. Regresyon tahminlerine bir yan terimi ekleyerek, Ridge Regresyon standart hataları azaltır.

Bu tezin konusu olarak bağımsız değişkenler arasında çoklu bağlantı olması durumunda model parametrelerini tahmin etmek için kullanılan kestirim yöntemlerinden Ridge Regresyon yöntemi dikkate alınmıştır. Ridge Regresyon'un başarısı, büyük ölçüde ridge parametresi olarak bilinen ceza parametresinin seçimine bağlıdır. Bu tezin esas amacı, sözü edilen ceza parametresini belirlemek için farklı seçim kriterlerinin karşılaştırmasını sağlamaktır. Bu anlamda en iyi çözüm veren bilgi kriterini yada kriterlerini belirlemektir. Bu amaçla bir simülasyon ve gerçek veri çalışması yaparak, en iyi çözüm veren parametre seçim yöntemi veya kriter belirlenmeye çalışılmıştır. İzleyen bölümde öncelikle, değişkenler arasında çoklu doğrusal bağlantının incelenmesinden sonra tezin konusunu oluşturan kavramlara yer verilmiştir.

2. ÇOKLU DOĞRUSAL BAĞLANTI (MULTI-COLLINEARITY) PROBLEMİ

Birçok regresyon probleminde bağımsız değişkenler matrisinin sütunları ya da bağımsız değişkenler arasında bir ilişki vardır. Bu durum çoklu doğrusal bağlantı problemi olarak adlandırılır. Çoklu doğrusal bağlantı bir regresyon modelinin etkinliğini dramatik olarak etkileyen ciddi bir problemdir. Çoklu doğrusal bağlantı regresyon katsayılarının tahminlerinin büyük varyanslı ve yanlı olarak sonuçlanmasına neden olur. Genel olarak model anlamlı olsa bile regresyon katsayılarının bireysel olarak yani, kısmi regresyon katsayılarının istatistiksel anlamlılıktan yoksunluğuna neden olabilir. Bu nedenle değişkenler arasındaki ilişkiyi sağlayan modeli tahmin edebilmek için öncelikle çoklu doğrusal bağlantının tespit edilmesi gerekir.

2.1. Çıkarıma Üzerine Etkileri

Çoklu regresyon analizinde açıklayıcı değişkenler ve yanıt değişkeni arasındaki ilişki araştırmacıların ilgisine çekmektedir. Genellikle şu sorulara cevap aranır (Kunter, Micheal H., vd 2014).

- Birbirinden farklı olan açıklayıcı değişkenlerin etkilerinin nispi önemi nedir?
- Açıklayıcı değişkenin yanıt değişken üzerindeki etkisinin büyüklüğü nedir?
- Açıklayıcı değişkenler arasından yanıt değişken üzerinde etkisi az olan veya etkisi olmayan açıklayıcı değişkenler modelden çıkartılabilir mi?
- Modelde yer almayan herhangi bir açıklayıcı değişken modele eklenebilir mi?

Çoklu doğrusal bağlantının, regresyon katsayılarının En Küçük Kareler (EKK) kestirimleri üzerinde önemli ölçüde etkisi söz konusudur. Bu etkiyi açıklayabilmek için bir Y yanıt değişkeni üzerinde etkili olan X_1 ve X_2 gibi iki bağımsız değişkene sahip olduğumuzu varsayalım. Ayrıca bu değişkenleri birim uzunluk ölçeğine göre, diğer bir deyişle değişkenlerin kendi ortalamalarının farklarının standart sapmalarına

bölünerek standartlaştırıldığını, varsayalım (bkz. eşitlik (2.7)). Bu standartlaştırılmış değişkenlere ilişkin regresyon model,

$$\tilde{Y} = \beta_1 \tilde{X}_1 + \beta_2 \tilde{X}_2 + \varepsilon \quad (2.1)$$

olarak ifade edilir. Bu modeldeki regresyon katsayılarının tahminlerini sağlayan EKK normal denklemleri;

$$(\tilde{X}'\tilde{X})\hat{\beta} = \tilde{X}'\tilde{Y} \quad (2.2)$$

eşitliği ile sağlanır. Burada $(\tilde{X}'\tilde{X})$ bağımsız değişkenler arasındaki korelasyon matrisini gösterirken, $\tilde{X}'\tilde{Y}$ matrisi bağımsız değişkenler ile yanıt arasındaki korelasyonları gösterir. Bu durumda (2.2) eşitliği aşağıdaki gibi yeniden yazılabilir:

$$\begin{bmatrix} 1 & r_{12} \\ r_{21} & 1 \end{bmatrix}_{2 \times 2} \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix}_{2 \times 1} = \begin{bmatrix} r_{1y} \\ r_{2y} \end{bmatrix}_{2 \times 1} \quad (2.3)$$

Burada $r_{12} = r_{21}$; \tilde{X}_1 ve \tilde{X}_2 arasındaki korelasyonu, r_{1y} ve r_{2y} bağımsız değişkenler ile Y bağımlı değişken arasındaki korelasyonu göstermektedir.

Eşitlik (2.3)'den regresyon katsayılarının tahmini,

$$\hat{\beta} = (X'X)^{-1} X'Y = \begin{pmatrix} 1 & r_{12} \\ r_{21} & 1 \end{pmatrix}^{-1} \begin{pmatrix} r_{1y} \\ r_{2y} \end{pmatrix} = \begin{pmatrix} \frac{1}{(1-r_{12}^2)} & -\frac{r_{12}}{1-r_{12}^2} \\ -\frac{r_{12}}{1-r_{12}^2} & \frac{1}{(1-r_{12}^2)} \end{pmatrix} \begin{pmatrix} r_{1y} \\ r_{2y} \end{pmatrix} = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} \quad (1.4)$$

olarak bulunur. Burada $(\tilde{X}'\tilde{X})^{-1}$, korelasyon matrisinin tersini göstermektedir. Böylece (2.4) eşitliğinde verilen matris ve vektörün çarpılması sonucunda, regresyon katsayılarının tahminleri sırasıyla aşağıdaki biçimde elde edilir:

$$\hat{\beta}_1 = \frac{r_{1y} - r_{12}r_{2y}}{(1-r_{12}^2)} \text{ ve } \hat{\beta}_2 = \frac{r_{2y} - r_{12}r_{1y}}{(1-r_{12}^2)} \quad (2.5)$$

olarak bulunur. Eğer \tilde{X}_1 ve \tilde{X}_2 değişkenleri arasında kuvvetli bir ilişki söz konusu ise, diğer bir deyişle, $r_{12} = \pm 1$ ise, eşitlik (2.5)'in payından görüldüğü gibi, katsayıları tahmin etmek imkânsız hale gelir. Diğer yandan, gösterimlerde kolaylık sağlamak için korelasyon matrisinin tersinin varyans ile çarpılması sonucunu sembolik olarak $C = \sigma^2(\tilde{X}'\tilde{X})^{-1}$ gösterdiğimiz varsayalım. Bu durumda varyans (*Var*)-kovaryans (*Cov*) matrisi aşağıdaki biçimde tanımlanabilir:

$$C = \sigma^2(X'X)^{-1} = \sigma^2 \begin{pmatrix} \frac{1}{(1-r_{12}^2)} & -\frac{r_{12}}{1-r_{12}^2} \\ -\frac{r_{12}}{1-r_{12}^2} & \frac{1}{(1-r_{12}^2)} \end{pmatrix} = \begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix} \\ = \begin{pmatrix} Var(\hat{\beta}_1) & Cov(\hat{\beta}_1, \hat{\beta}_2) \\ Cov(\hat{\beta}_2, \hat{\beta}_1) & Var(\hat{\beta}_2) \end{pmatrix}$$

Belirtmek gerekirse, C matrisinden görünen o ki, $r_{12} \rightarrow \pm 1$ ise, $|r_{12}| \rightarrow 1$ ve buna bağlı olarak, katsayı tahminlerinin varyans ve kovaryansları sonsuz büyük olmaktadır. Örneğin, bağımsız değişkenler arasında yüksek bir korelasyon olması durumunda, $Var(\hat{\beta}_1) = c_{11} = \infty$ ve $Cov(\hat{\beta}_1, \hat{\beta}_2) = c_{12} \Rightarrow \mp \infty$ olarak sonuçlanmaktadır. Bu konuda daha ayrıntılı bilgi için bkz. Montgomery, D., Peck, E. (1991). İki den fazla bağımsız değişkenin çoklu bağlantı olması durumunda benzer şekilde etkilenecektir.

Çoklu doğrusal bağlantının regresyon katsayısı işaretlerine etki etmesi durumunda kestirimlerin işaretleri, bağımlı ve bağımsız değişkenler arasındaki ilişkiyi yanlış yorumlamaya sebep olacaktır. Çoklu bağlantının katsayılarının hipotez testleri üzerindeki etkisi ise şu şekilde açıklanabilir. Parametreleri tek tek test etmek için kurulan $H_0: \beta_j = 0, j = 1, 2, \dots, k$ hipotezlerin anlamlılığının testi aşağıdaki gibi ifade edilir:

$$t = \frac{\hat{\beta}_j}{\sqrt{1-c_{jj}}} = \hat{\beta}_j \sqrt{\frac{1-R_j^2}{\sigma^2}}, j = 1, 2, \dots, k \quad (2.6)$$

Burada R_j^2 , j.inci değişken ile diğer değişkenler arasındaki belirlilik katsayısını gösterir. Yukarıda görüldüğü gibi, R_j^2 değerinin 1'e yaklaşması ve t değerinin 0'a yaklaşmasına neden olur. Bu durum çoklu bağlantının önemli bir göstergesidir. Bu durumda yokluk hipotezinin reddedilmesi zorlaşarak sınanan parametrenin sıfırdan farklı olmadığı yani bağımlı değişkeni etkilemediği yargısına varılacaktır. Böyle bir yargı ise yanlış sonuçlara ulaşmaya sebep olacaktır. Sonuç olarak; çoklu bağlantının sakıncalı etkilerinin giderilmesi için çoklu bağlantının tespit edilip ortadan kaldırılması gerekmektedir.

2.2. Çoklu Bağlantının Tespit Edilmesi

Çoklu bağlantının belirlenmesinde birçok yöntem vardır. En basit olarak iki değişken arasındaki korelasyon katsayısının 1'e yakın olması çoklu bağlantının işareti olabilir. Bağımsız değişkenin modele eklenmesi ya da modelden çıkarılması regresyon katsayılarında büyük değişikliklere sebep olurken çoklu bağlantının da göstergesi olabilir. Regresyon katsayılarının büyüklüğü ve işaretlerin beklenenin aksine ters çıkması da çoklu bağlantının göstergesi olabilir.

Çoklu bağlantı probleminin belirlenmesinde yaygın olarak kullanılan bazı yöntemleri aşağıdaki gibi sıralamak mümkündür.

i. Basit Korelasyon Katsayılarının incelenmesi

Bir önceki bölümde de ifade edildiği gibi, k tane bağımsız değişkeninin her biri, aşağıdaki gibi standartlaştırılır:

$$\tilde{X}_{ij} = \frac{X_{ij} - \bar{X}_j}{\sqrt{\sum_{i=1}^n (X_{ij} - \bar{X}_j)^2}} = \tilde{\mathbf{X}} \text{ ve}$$
$$\tilde{Y}_{ij} = \frac{Y_i - \bar{Y}}{\sqrt{\sum_{i=1}^n Y_i^2 - \left(\frac{\sum_{i=1}^n Y_i}{n}\right)^2}}, i = 1, 2, \dots, n \text{ ve } j = 1, 2, \dots, k \quad (2.7)$$

Bu standartlaştırma sonucunda elde edilen $\tilde{\mathbf{X}}$ matrisi yardımı ile R ile gösterilen korelasyon matrisi elde edilir:

$$R = (\tilde{\mathbf{X}}' \tilde{\mathbf{X}})_{k \times k} = \begin{pmatrix} 1 & \cdots & r_{1k} \\ \vdots & \ddots & \vdots \\ r_{k1} & \cdots & 1 \end{pmatrix}_{k \times k} \quad (2.8)$$

Bu Korelasyon matrisindeki köşegen dışı elemanları kontrol edilir. Bu elemanların 1'e yaklaşımları durumunda güçlü bir çoklu bağlantı varlığından şüphe edilmelidir.

ii. Varyans Artırıcı Faktör (Variance Inflation Factor, VIF) değerlerinin incelenmesi

Varyans artırıcı faktör çoklu bağlantının belirlenmesinde kullanılan bir yöntemdir. Pratikte VIF değerlerini belirlemenin en kolay yolu (2.8)'de belirtilen korelasyon matrisinin tersi kullanmaktır. Buna göre, $(\tilde{\mathbf{X}}' \tilde{\mathbf{X}})^{-1}$ matrisinin köşegen değerleri, VIF değerlerini verir. Diğer bir deyişle, $VIF = \text{diag}(\tilde{\mathbf{X}}' \tilde{\mathbf{X}})^{-1}$ (Chatterjee ve Price, 1977).

Bunun yanı sıra, VIF değerinin hesaplanmasında belirlilik katsayılarından yararlanılır. Bu bilgilere göre VIF değerleri aşağıdaki gibi ifade edilebilir:

$$VIF_j = \text{diag}(\tilde{X}'\tilde{X})^{-1} = \frac{1}{1-R_j^2}, \quad j = 1, 2, \dots, k \quad (2.9)$$

Burada R_j^2 , \tilde{X}_j değişkeninin kalan $k-1$ açıklayıcı değişkene göre regresyonundan elde edilen j .inci belirlilik katsayısını gösterir. Uygulamada $VIF_j \geq 10$ ise j .inci açıklayıcı değişken ile diğer değişkenler arasında güçlü bir çoklu bağlantı probleminden söz edilir. Ancak bazı kaynaklarda $VIF_j \geq 5$ olması da dikkate alınabilir. Daha ayrıntılı bilgi için Aydın, D., (2014) tarafından yazılan kaynak kitaba bakılabilir.

iii. Koşul İndislerinin İncelenmesi

Yukarıda ifade edilen $(\tilde{X}'\tilde{X})$ korelasyon matrisi özdeğerlerinden çoklu bağlantının derecesinin belirlenmesinde yararlanılır. Çoklu bağlantının olmadığı durumlarda özdeğerler 1'e eşittir. En az bir özdeğerin 1'den farklı olması veya en az birinin 0'a yakın olması çoklu bağlantının varlığını ortaya koyar. Ancak özdeğerlerin tek tek incelenmesi pek fazla anlam teşkil etmemektedir. Bu yüzden, Vinod ve Ullah (1981)'in çalışmasında en büyük özdeğer ile en küçük özdeğer üzerine dayalı koşul indeksini önerilmiştir. Koşul indeksi için en küçük kareler tahmincilerinin hesaplanmasında kullanılan $\tilde{X}'\tilde{X}$ korelasyon matrisinin özdeğerleri

($\lambda_{max} = \lambda_1 > \lambda_2 > \dots > \lambda_k = \lambda_{min}$) ile gösterilirse koşul indeksi (KI)

$$KI = \sqrt{\frac{\lambda_{max}}{\lambda_{min}}} = \sqrt{\frac{\text{En büyük Özdeğer}}{\text{En Küçük Özdeğer}}} \quad (2.10)$$

şeklinde hesaplanır. Burada belirtmek gerekirse, $KI < 10$ ise çok az çoklu bağlantı ciddi bir sorun teşkil etmez; $10 \leq KI \leq 30$ ise çoklu bağlantı orta düzeyde ve son olarak; $KI > 30$ ise, şiddetli çoklu bağlantı olduğu düşünülmelidir (Pagel and Lunneborg, 1985; Gujarati, 1995; Rathert ve ark., 2011).

iv. Tolerans Değerinin (Tolerance Value) incelenmesi

Tolerans değeri çoklu bağlantının belirlenmesinde bir başka belirleyici faktördür. Tolerans değeri,

$$TV = 1 - R_j^2, \quad j = 1, 2, \dots, k \quad (2.11)$$

olarak hesaplanır. Bunun sonucunda daha küçük *TV* (tolerance value) değeri, daha büyük VIF değerine neden olur. Bu yüzden küçük çıkan *TV* değeri çoklu bağlantının var olabileceğini düşündürür (Albayrak, 2005). Yukarıda bahsedilen ve çoklu bağlantının belirlenmesi için kullanılan bu yaklaşımlardan hangisinin hangi durumda kullanılması gerektiği konusunda net bir bilgiye rastlanılmamıştır (Gujarati, 1995). Ancak çoklu bağlantının varlığının araştırılmasında tüm bu kriterlere bakılması çoklu bağlantının teşhisi için faydalıdır (Albayrak, 2005).

2.3. Çoklu Bağlantı Probleminin Çözümü

Çoklu bağlantının tespit edildiği durumlarda, regresyon üzerindeki olumsuz etkilerden dolayı yok etmek veya etkiyi azaltma yoluna gidilmelidir. Çoklu bağlantının giderilmesine yönelik bazı araştırmacılar çeşitli önerilerde bulunmuşlardır (Gujarati, 1995; Neter et al., 1990; Albayrak, 2005). Bunları aşağıdaki gibi özetlemek mümkündür.

- Bir veya daha fazla değişkenin modelden çıkartılması,
- Modele yeni değişkenler ilave edilmesi,
- Birbirleri ile ilişkili olan değişkenlerin birleştirilmesi (toplanması),
- Temel Bileşenler Regresyon analizi ya da Ridge Regresyon analizi gibi yanlı kestirim yöntemlerinin kullanılması çözüm olarak verilebilir.

Bu tezin konusu çoklu bağlantı durumunda çözüm önerisi olarak yanlı kestirim yöntemlerinde Ridge Regresyonu ile elde edilen parametrelerin tahminleri karşılaştırmaktır.

3. RİDGE REGRESYONU

Ortogonal olmayan verilere En Küçük Kareler yöntemi uygulandığında, regresyon katsayılarının çok zayıf tahminleri elde edilebilir. Burada temel sorun, regresyon katsayılarının En Küçük Kareler tahminlerinin varyanslarının dikkate değer bir biçimde artmış olabileceğidir. Bu durum, En Küçük Kareler tahminleri mutlak değerce çok büyük, durağan olmadıkları ve büyüklükleri ve işaretleri alınan farklı örneklerde dikkate değer bir biçimde değişebileceği anlamına gelir. En Küçük Kareler yönteminin esas problemi $\hat{\beta}$, β 'nin yansız bir kestiricisi olma gereksinimidir. Bu probleme çözüm olarak, alternatif analiz yöntemi olan Ridge Regresyon yöntemi tavsiye edilmektedir. Ayrıca Ridge Regresyonu'nun etkin hesaplanabilmesi için ridge parametresi olarak bilinen ceza parametresinin seçilmesi gerekir. Bu Ridge Regresyon'undan bilinen sıradan yöntemler ile yapılmaktadır. Bu tezi yazmaktaki asıl düşünce, ceza parametresini seçmek için aynı zamanda bir model belirleme kriterleri olarak da kullanılan bilgi kriterlerini kullanmaktır. Bu amaçla 5 bilgi kriteri dikkate alınmıştır. İzleyen kısımlarda, sırasıyla söz konusu bu yöntemler tartışılmıştır.

Regresyon problemlerinde yapılan tartışmalardan çıkan sonuç, model karmaşıklığını, yani açıklayıcı değişkenlerin sayısını azaltmaktır. Bunun için ileri veya geri doğru seçim yöntemi kullanılabilir, ancak bu şekilde kaldırılan değişkenlerin yanıt üzerindeki etkisi hakkında hiçbir şey söylenemez. Açıklayıcı değişkenlerin modelden çıkarılması katsayılarını sıfıra ayarlanmış olarak görülebilir. Onları tamamen sıfır olmaya zorlamak yerine, sıfırdan çok uzaktalarsa onları cezalandırmak ve böylece sürekli olarak onları küçük olmaya zorlamak gerekir. Esas itibariyle Ridge Regresyonu bu işi üstlenmektedir. Diğer bir deyişle, Ridge Regresyonu modeldeki tüm değişkenleri korurken model karmaşıklığını azaltır.

Ridge Regresyonu, çoklu bağlantı problemi ortaya çıktığında bağımlılık yapısını yok etmek için alternatif olarak kullanılabilen bir istatistikî yanlı tahmin yöntemidir. Ridge Regresyonu En Küçük Kareler yönteminin yetersiz kaldığı durumların varlığı sebebiyle Hoerl ve Kennard (1997a,b) tarafından aşağıda belirtilen üç amaç doğrultusunda önerilmiştir:

- Kuvvetli çoklu bağlantının varlığı durumunda, katsayılarda meydana gelen kararsızlıkların grafik üzerinde gösterilmesinde,
- Çoklu Doğrusal Regresyon analizinde çoklu bağlantı durumunda En Küçük Kareler yönteminin tahmininden daha küçük varyanslı tahminler elde edilmesinde,
- Modeldeki bazı değişkenlerin çıkartılmasında kullanılabilir (Hoerl ve Kennard;1970a,b).

Ridge Regresyonun, yanlı regresyon yöntemi olmasına karşın EKK yöntemine göre iki önemli etkisi vardır. Bunlar; bağımsız değişkenlerdeki çoklu doğrusal bağlantıyı gidermek ve regresyonda yanlılık karesiyle varyansı değiştirerek hata kareler ortalamasını azaltmaktır. Ayrıca, Ridge tahmin edicileri yanlı olmasına rağmen EKK tahminlerinden daha kararlı olabilmektedir ve varyansta azalmaya neden olmasından kaynaklı tercih edilmektedir.

3.1. Ridge Regresyon ile EKK regresyonu arasındaki ilişki

EKK tekniği bilindiği üzere regresyon modelindeki parametreleri tahmin etmede sıklıkla kullanılan yöntemlerden biridir. EKK'nın sıklıkla tercih edilmesinin en temel sebebi tahmin edilen değerler ile gerçek değerler arasındaki hatayı minimum yapan yansız bir yöntem olmasıdır. Fakat EKK tekniğinin uygulaması için regresyon modelinin bazı varsayımları sağlaması gerekmektedir. Tekniğin geçerli olması için öncelikle hataların bağımsız ve normal dağılması gerekmektedir. Aynı zamanda bu teknikte bağımsız değişkenlerin birbiri ile ilişkili olmaması gerekmektedir. Bahsi geçen bu varsayımların bozulması durumunda yapılan analiz oluşturulan modelin güvenilirliğini düşürerek kabul görmeyen yorumlara neden olacaktır. Özetle EKK tekniği ile tahmin edilen regresyon parametreleri varsayımları sağlamadan tahmin edildiğinden ötürü gerçek değerlerinden uzaklaşmasına ve aynı örneklem üzerindeki farklı ölçümler bağımsız değişkenler arasında güçlü ilişkilere neden olmasından dolayı çoklu bağlantı problemine neden olacaktır. Böyle bir durumda EKK tekniğinin gerçek değerlerinden uzaklaştığı doğru analiz ve yorumlama yeteneğimizin kaybolmasından dolayı alternatif tekniklere yönelime sebep olmuştur. Çoklu doğrusal bağlantılı verilerde, regresyon katsayılarının standart hataları, büyüklükleri ve işaretleri doğru bir biçimde tahmin edilememesi ve verilerde çoklu

doğrusal bağlantı probleminin var olması durumunda Ridge Regresyon yöntemi EKK yöntemine göre daha durağan ve kavramsal beklentilere uygun sonuçlar vermektedir(Üçkardeşler vd., 2012). Bu tarz problemin etkilerini yok etmek ve daha tutarlı sonuçlar elde edebilmek için farklı bir yöntem olan Ridge Regresyon analizi en popüler yöntemlerden biri olarak karşımıza çıkmaktadır. Ridge Regresyonu bağımsız değişkenlerin birbirleri üzerindeki etkilerinin minimum yapmak ve kararlı katsayı tahminleri elde edebilmek amacı ile kullanırlar. Böyle bir durumda Ridge Regresyon yöntemi kullanarak daha küçük standart hata, daha doğru ve güvenilir regresyon denklemleri elde edilir.

Çoklu bağlantı problemini ortadan kaldırmak için literatürde değişik yöntemler önerilmiştir. Çoklu doğrusal bağlantı problemine karşı Hoerl ve Kennard (1970a, b) tarafından gündeme getirilen yanlı bir tahmin tekniği olan Ridge Regresyon analizi önerilmiştir. Regresyon analizinde çoklu bağlantı ve aykırı değer olması sıklıkla karşılaşılan problemlerdendir. Bu problemlerin çözümü hem yanlı hem de sağlam tekniklerin birlikte kullanılmasıyla mümkün olabilmektedir. Literatürde Askin ve Montgomery (1980), Silvapulle (1991), Arslan ve Billor (1996), Pfaffenberger ve Dielman (1990) yanlı sağlam tahmin ediciler üzerinde çalışan araştırmacılar arasında sayılabilir.

Ridge Regresyon yöntemi farklı alanlarda ve çalışmalarda sıklıkla kullanılmaktadır. Uslu (1991) de öğrencilerin başarısını tahmin etmek, Albayrak (2005) şişman hastalarda vücut kitle indeksi tahmin etmek, Carvalheiro et al. (2006) Nelore-Hereford ineklerinin canlı ağırlık kazançlarını tahmin etmede ve Pimentel et al. (2007) ineklerde erken büyüme performanslarını tahmin etmek amacı ile Ridge Regresyon yöntemlerini kullanmışlardır.

3.2. Ridge Regresyon Kestircisinin Hesaplanması

Hoerl ve Kennard (1970)'ın açıklayıcı değişkenlerin standardize edilmesi ve regresyon parametrelerinin EKK yönteminden daha küçük varyanslı değişkenler elde edildiği regresyon modeli olarak, aşağıda verilen matris ve vektör formunda çoklu regresyon modelini dikkate aldığımızı varsayalım. Söz konusu model aşağıdaki gibi ifade edilir:

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

Burada Y , $(n \times 1)$ boyutlu yanıt ya da bağımlı değişenin aldığı değerler vektörü; X , $(n \times k)$ boyutlu bağımsız değişken matrisi; β , $(k \times 1)$ boyutlu tahmin edilecek regresyon katsayılar vektörü ve ε , $(n \times 1)$ boyutlu ortalaması sıfır ve $\text{Var}(\varepsilon) = \sigma^2 I_n$ olan rassal hatalar vektörünü göstermektedir.

Eşitlik (2.8) verilen standart değişkenler açısından EKK yöntemine göre katsayıların tahminini veren formüle göre, normal değişkenler açısından da (3.1) modelindeki regresyon katsayılarının tahmini aşağıdaki biçimde tanımlanır:

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

Çoklu bağlantı olması durumunda bağımsız değişkenler arasındaki ilişkinin yüksek olması $(X'X)$ matrisinin varyanslarını önemli bir ölçüde büyütecektir. Böylece gerçekte önemli gözükken parametre değerleri varyansın artması neticesinde önemsiz çıkacaktır. Bu sorunu ortadan gidermek için eşitlik (3.2)'de yer alan $(X'X)$ matrisinin köşegen elemanlarına pozitif bir k sabiti eklenerek bu matrisin varyansı önemli ölçüde küçültülür (Hoerl, A.E. 1962). Böylece, Ridge Regresyon kestiricisi, EKK normal denklemlerinin değiştirilmiş versiyonu çözümlenerek elde edilir. Sıradan EKK regresyonundaki hata kareler toplamının minimumuna benzer olarak Ridge Regresyon kestiricisi aşağıda verilen "Cezalı Hata Kareler Toplamı (CHKT)"ni minimum yapan regresyon katsayıları olarak tanımlanır:

Ridge Regresyonu En Küçük Karelere çok benzerdir. Ancak, ridge katsayıları ondan farklı bir miktarı en aza indirerek tahmin edilir. Özellikle, Ridge Regresyon katsayıları β , aşağıdaki miktarı en aza indiren değerlerdir:

$$\begin{aligned} CHKT(\beta, k) &= \min\{\sum_{i=1}^n (Y_i - X_i \beta)^2 + k \sum_{j=1}^p \beta_j^2\} \\ &= (Y - X\beta)'(Y - X\beta) + k\beta'\beta \end{aligned} \quad (3.3)$$

Burada k , ridge parametresi olarak ifade edilir ve bu parametre ayarlama ya da küçülme parametresi (*shrinkage parameter*) olarak da bilinir. Ridge parametresi regresyon katsayılarının tahminlerinin küçülme ya da büyüme miktarını kontrol eder. Örneğin, $k = 0$ için ridge problemi, Sıradan En Küçük Kareler Regresyonu problemine dönüşür. Diğer yandan, $k \rightarrow \infty$ iken, sadece sabit terimli bir model elde edilir. Bu nedenle ridge parametresinin seçimi önemlidir.

Sıradan EKK probleminde olduğu gibi, (3.3) denkleminin β' ya göre türevi alınıp sifıra eşitlenecek olursa, (3.3) denkleminin çözümü elde edilir:

$$\frac{CHKT(\hat{\beta})}{d\beta} = -2\mathbf{X}'(\mathbf{Y} - \mathbf{X}\beta) + 2k\beta = -2\mathbf{X}'\mathbf{Y} + 2\mathbf{X}'\mathbf{X}\beta + 2k\beta = 0$$

Yukarda verilen eşitlikte, bazı basit cebirsel işlemlerden sonra bu eşitlik eşdeğer olarak aşağıdaki gibi yazılır:

$$\mathbf{X}'\mathbf{Y} = \{\mathbf{X}'\mathbf{X}\beta + k\beta = \beta(\mathbf{X}'\mathbf{X} + k\mathbf{I})\}$$

Böylece, (3.3) problemine çözüm sağlayan “*ridge regresyon kestiricisi*” aşağıdaki gibi tanımlanır:

$$\hat{\beta}_R = (\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1}\mathbf{X}'\mathbf{Y}, \quad k \geq 0 \quad (3.4)$$

Burada k değeri yukarda ifade edildiği gibi, ridge ya da yanlılık parametresidir. Ayrıca, k parametresinin seçimi tahmin edicinin performansını etkilerken, bu değer sıfır olduğunda sıradan EKK kestiricisi ile aynı sonucu verir. Belirtmek gerekirse, k değeri artırılarak yanlı bir tahmin yapılmış, ancak varyans önemli ölçüde küçültülmüş olunur. Genel kareler toplamının değişmemesinden dolayı EKK yöntemine göre, Ridge Regresyon'un R^2 ve F değerinde çok az bir düşüş meydana gelir. (3.4) eşitliğinde verilen Ridge Regresyon tahmin edicisi, kanonik formuda aşağıda gibi yazılabilir (Hoerl, A.E. ve Kennard R.,W. (1970a)):

$$\hat{\alpha}_R = (\mathbf{C}'\mathbf{C} + k\mathbf{I})^{-1}\mathbf{C}'\mathbf{C}\hat{\alpha} \quad (3.5)$$

Burada $\mathbf{C} = \mathbf{X}\mathbf{P}$, $\alpha = \mathbf{P}'\beta$ ve \mathbf{P} ise $(n \times p)$ boyutlu ortagonal bir matristir. Diğer bir ifadeyle $\mathbf{P}'\mathbf{P} = \mathbf{I}$ olarak yazılabilir. α 'nın herhangi bir tahmin edicisi olan $\hat{\alpha}_R$, EKK tahmin edicisi $\hat{\beta}$ ile $\hat{\beta} = \mathbf{P}\hat{\alpha}$ bağlantılıdır. Bu durumda EKK ve ridge kestiricilerine dayalı olarak elde edilen Hata Kareler Ortalaması (MSE) aynı olur. Diğer bir deyişle, $MSE(\hat{\alpha}) = MSE(\hat{\beta})$ yazılabilir. Bu bağlamda, ridge tahmin edicisinin hata kareler ortalaması ise

$$MSE(\hat{\alpha}_R) = \sigma^2 \sum_{i=1}^n \lambda_i (\lambda_i + k)^{-2} + \sum_{i=1}^n \left(\frac{k\alpha_i}{\lambda_i + k} \right)^2 \quad (3.6)$$

biçiminde hesaplanır. Ridge tahmin edicisi MSE 'nin EKK tahmin edicisinin MSE 'den daha küçük olacak şekilde bir yanlılık parametresi k 'nın ($k > 0$) her zaman var olacağı Hoerl ve Kennard (1970) tarafından gösterilmiştir. k 'nın seçiminde

literatürde birçok metot önerilmiştir. Bu tezde söz konusu olan ridge parametresinin seçimi bilgi kriterleri kullanılarak yapılmıştır.

Ridge regresyonun problemine çözümü sağlayan bir diğer format aşağıdaki biçimde verilebilir:

$$CHKT(\beta, k) = \min \sum_{i=1}^n (Y_i + \sum_{j=1}^p X_{ij}\beta_j)^2 + \sum_{j=1}^p (0 - \sqrt{k}\beta_j)^2 \quad (3.7)$$

Bu yaklaşım, aşağıda verilen genişletilmiş bir \mathbf{X}_k model matrisi ve \mathbf{Y}_k model vektörüne yol açan p gözlemleri ekleyerek genişletilmiş bir veri kümesi için bir diğer en küçük kareler problemi oluşturur:

$$\mathbf{X}_k = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ x_{n1} & x_{n1} & \dots & x_{np} \\ \sqrt{k} & 0 & \dots & 0 \\ 0 & \sqrt{k} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sqrt{k} \end{pmatrix} = \begin{pmatrix} \mathbf{X} \\ \sqrt{k} \mathbf{I}_p \end{pmatrix} \text{ ve } \mathbf{Y}_k = \begin{pmatrix} y_{11} \\ y_{21} \\ \vdots \\ y_{n1} \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \begin{pmatrix} \mathbf{Y} \\ 0 \end{pmatrix}$$

Bu matris ve vektör formundan açıkça görülen o ki, (3.7) eşitliğine çözümü aşağıdaki en küçük kareler formunda ifade edilen katsayı tahminleri sağlar (Hoerl and Kennard (1970)):

$$\hat{\beta}_R = (\mathbf{X}_k' \mathbf{X}_k)^{-1} \mathbf{X}_k' \mathbf{Y} = (\mathbf{X}' \mathbf{X} + k \mathbf{I})^{-1} \mathbf{X}' \mathbf{Y} \quad (3.8)$$

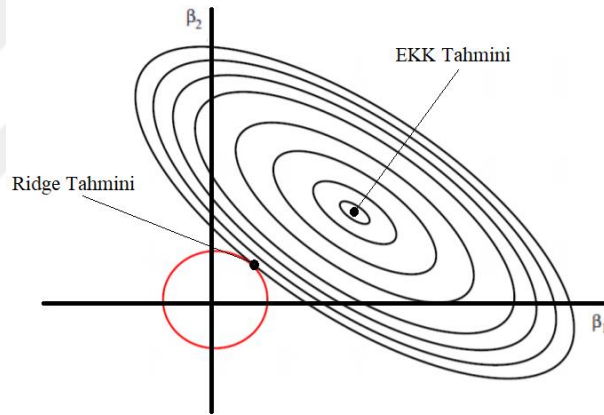
Eşitlik (3.4)'de bahsedildiği gibi, Ridge Regresyon çözümü EKK çözümüne benzer olduğu söylenebilir. Ancak, ridge çözümünün elde edilmesinde $\mathbf{X}' \mathbf{X}$ matrisine $k \mathbf{I}$ şeklinde bir ceza terimi eklenmiştir. (3.8) denklemine göre, bu ceza parametresi ile çözüm arasında şu şekilde bir bağıntı olduğu görülmektedir:

- $k \rightarrow 0$ iken, $\hat{\beta}_R \rightarrow \beta$: Ceza parametresi sıfıra doğru yönelirse, Ridge ve EKK çözümleri bir birine yaklaşır. Ancak, özel olarak $k = 0$ için $\hat{\beta}_R = \beta$, diğer bir deyişle, ridge ve EKK çözümleri aynı olur.
- $k \rightarrow \infty$ iken $\hat{\beta}_R \rightarrow 0$: Ceza parametresi büyürken, ridge regresyon çözümleri sıfıra yönelir.

Dolayısıyla Ridge regresyon yönteminde ceza parametresinin seçimi oldukça önemlidir. Bu tezin ana konusunu bu parametrenin seçimini oluşturmaktadır. Bu bağlamda, parametre seçimi Bölüm 4'te tartışılmıştır.

3.3. Ridge Regresyon Kestircisinin Geometrik Yorumu

Ridge Regresyonu, $\hat{\boldsymbol{\beta}} = (\hat{\beta}_1, \dots, \hat{\beta}_p)$ tahminlerini yaklaşık olarak sıfıra doğru çektiği zaman, ridge problemine çözüm sağlayan eşitlik (3.9)'daki $\sum_{i=1}^n (Y_i - X_i \hat{\boldsymbol{\beta}})^2 = RSS(k)$ hata kareler toplamını minimum yaparken, aynı zamanda $k \sum_{j=1}^p \hat{\beta}_j^2$ terimini de minimum yaparak veriyi oldukça iyi fit eden tahminler bulmaya çalışır (James ve ark., 2013). Bu konuya açıklık getirmek amacıyla, örneğin, $\hat{\boldsymbol{\beta}} = (\hat{\beta}_1, \hat{\beta}_2)'$ gibi iki boyutlu uzayda Ridge Regresyonu ve EKK kestircilerin geometrik yorumunu veren bir gösterim Şekil 3.1'de verilmiştir. Şekil 3.1'de görüldüğü gibi, $\hat{\boldsymbol{\beta}}_R$ ile ifade edilen Ridge Regresyon tahmini, $RSS(k)$ konturunun $\beta_1^2 + \beta_2^2 = c_2^2(k)$ ile tanımlanan çemberi karşıladığı bir noktadır. Bu resimdeki, konturlar eşit $RSS(k)$ 'lı β 'ların değerlerini gösterir. Ridge Regresyonu kırmızı çemberde yer alan en iyi değeri bulmaya çalışıyor.



Şekil 3.1. İki boyutlu uzayda Ridge Regresyon kestircisinin gösterimi

Şekil 3.1'deki durum açısından bakıldığında, eşitlik (3.3)'de verilen Ridge Regresyon problemine çözüm sağlayan ceza kareler toplamı, matris ve vektör formunda yeniden yazılabilir:

$$\min \|Y - X\boldsymbol{\beta}\|_2^2, \|\boldsymbol{\beta}\|_2^2 \leq c_2^2(k) \text{ kısıtı altında} \quad (3.9)$$

Burada $\|\boldsymbol{\beta}\|_2^2 = \sum_{j=1}^p \beta_j^2$ ve $c_2^2(k) > 0$, ridge parametresi k ile ilişkilidir. Belirtmek gerekirse, eşitlik (3.9)'da $\|\cdot\|_2^2$ olarak belirtilen terim L2 normunu (veya cezasını) ifade etmektedir. Eşitlik (3.9) denkleminde çözümleri sağlayan Ridge Regresyon kestircisi, aynı zamanda kısıtlı ridge tahmini olarak bilinir.

Şekil 3.1’de kırmızı çember olarak gösterilen L2 norm eğrisine bakıldığında, L2 cezası (L2 penalty) ile oluşturulan şekil bir çember olduğu için, ondan uzaklaştıkça o karesel olarak artar. Esasen arakesit noktası olan Optimum L2 sadece eksen çizgileri üzerine düşebilir. Bu durum gösteriyor ki, aynı zamanda minimum MSE’de (şekildeki siyah nokta) gerçekte eksen üzerindedir.

3.4. Ridge Kestiricisinin Özellikleri

Ridge Regresyon yanlı yöntemler arasında en popüler olanıdır, çünkü EKK yöntemi ile olan ilişkisi ve ridge tahmin edicisinin istatistiksel özellikleri iyi tanımlanmıştır. Burada Ridge Regresyon kestiricilerinin bazı özelliklerini inceliyoruz. Bu özellikler literatürde birçok yazar tarafından çalışılmıştır. Örneğin, ayrıntılı bilgi için bkz. Hoerl and Kennard (1970a, b), Newhouse ve Oman (1971) , Allen (1974), Marquardt and Snee (1975) tarafından yapılan çalışmalar incelenebilir. Ayrıca belirtmek gerekir ki, bu bölümde bir regresyon modelinde hataların sıfır ortalama ve sabit bir varyans ve bağımsız rassal olarak normal dağıldığı varsayımı altında, ridge tahmin edicisinin yanı ve varyansını türetilmektedir. Sıradan En Küçük Kareler yansız bir tahmin edicidir. Diğer bir deyişle, $E[\hat{\beta}] = \beta$ olarak ifade edilir. Bu bağlamda ridge kestiricisinin durumunu görmek için beklenen değerini bakmak gerekir.

Ridge kestiricisinin beklenen değeri

$$\begin{aligned}
 E[\hat{\beta}_R] &= E\left[(\mathbf{X}'\mathbf{X} + k\mathbf{I}_{pp})^{-1}\mathbf{X}'\mathbf{Y}\right] \\
 &= E\left[(\mathbf{X}'\mathbf{X} + k\mathbf{I}_{pp})^{-1}(\mathbf{X}'\mathbf{X})(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}\right] \\
 &= E\left[(\mathbf{X}'\mathbf{X} + k\mathbf{I}_{pp})^{-1}(\mathbf{X}'\mathbf{X})\hat{\beta}\right] \\
 &= (\mathbf{X}'\mathbf{X} + k\mathbf{I}_{pp})^{-1}(\mathbf{X}'\mathbf{X})E[\hat{\beta}] \\
 &= (\mathbf{X}'\mathbf{X} + k\mathbf{I}_{pp})^{-1}(\mathbf{X}'\mathbf{X})\beta
 \end{aligned} \tag{3.10}$$

olarak yazılabilir. Eşitlik (3.10) gösteriyor ki, herhangi bir $k > 0$ değeri için $E[\hat{\beta}_R] \neq \beta$ olduğu görülmektedir. Bu durumda ridge yanlı bir kestiricidir. Ayrıca, eşitlik (3.10) den açıkça görülüyor ki, ridge kestiricisinin beklenen değeri $k \rightarrow \infty$ iken sifıra yönelmektedir:

$$\lim_{k \rightarrow \infty} E[\hat{\boldsymbol{\beta}}_R] = \lim_{k \rightarrow \infty} (\mathbf{X}'\mathbf{X} + k\mathbf{I}_{pp})^{-1} (\mathbf{X}'\mathbf{X})\boldsymbol{\beta} = \mathbf{0}_p$$

Böylece, tüm regresyon katsayıları ceza parametresi arttıkça sıfıra doğru çekilmektedir. Bu aynı zamanda tam ranklı bir \mathbf{X} matrisi ($n > p$) göz önüne alındığında da geçerlidir.

Ridge kestiricisi, aşağıdaki yazılışından dolayı EKK kestiricinin bir lineer dönüşümü ve aynı zamanda doğrusal bir kestiricidir:

$$\begin{aligned}\hat{\boldsymbol{\beta}}_R &= (\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1}\mathbf{X}'\mathbf{Y} \\ &= (\mathbf{X}'\mathbf{X} + k\mathbf{I}_{pp})^{-1}(\mathbf{X}'\mathbf{X})\hat{\boldsymbol{\beta}} = \mathbf{A}_k\hat{\boldsymbol{\beta}}\end{aligned}$$

Burada $\mathbf{A}_k = (\mathbf{X}'\mathbf{X} + k\mathbf{I}_{pp})^{-1}(\mathbf{X}'\mathbf{X})$. Bu yüzden, $E[\hat{\boldsymbol{\beta}}_R] = E[\mathbf{A}_k\hat{\boldsymbol{\beta}}] = \mathbf{A}_k\boldsymbol{\beta}$ olduğu için yukarda belirtildiği gibi, $\hat{\boldsymbol{\beta}}_R$ kestiricisi $\boldsymbol{\beta}$ parametre vektörünün yanlı bir kestiricisidir. Bir orta-normal \mathbf{X} tasarım matrisi durumunda, EKK kestiricisi ve Ridge kestiricisi arasındaki ilişki aşağıdaki biçimde ifade edilebilir:

$$\hat{\boldsymbol{\beta}}_R = (\mathbf{X}'\mathbf{X} + k\mathbf{I}_{pp})^{-1}\mathbf{X}'\mathbf{Y} = (\mathbf{I}_{pp} + k\mathbf{I}_{pp})^{-1}\mathbf{X}'\mathbf{Y} = (1 + k)^{-1}\mathbf{I}_{pp}\mathbf{X}'\mathbf{Y}$$

Burada belirtmek gerekirse, $\mathbf{I}_{pp} = (\mathbf{X}'\mathbf{X}) = (\mathbf{X}'\mathbf{X})^{-1}$ olduğu için söz konusu ilişki,

$$\hat{\boldsymbol{\beta}}_R = (1 + k)^{-1}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} = (1 + k)^{-1}\hat{\boldsymbol{\beta}} \quad (3.11)$$

olarak yazılabilir. Bu nedenle, Ridge kestiricisi, bir faktör ile $\hat{\boldsymbol{\beta}}$ ile gösterilen EKK kestiricisini ölçeklendirir. Her iki tarafın beklenen değeri alındığında ridge kestiricisi $k \rightarrow \infty$ iken sıfıra yaklaştığı anlaşılmaktadır. Buna ilaveten EKK kestiricisi ve ridge kestiricisi arasındaki ilişki

$$\hat{\boldsymbol{\beta}}_R = (\mathbf{X}'\mathbf{X} + k\mathbf{I}_{pp})^{-1}\mathbf{X}'\mathbf{Y} = (\mathbf{X}'\mathbf{X} + k\mathbf{I}_{pp})^{-1}\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = (\mathbf{I} + k(\mathbf{X}'\mathbf{X})^{-1})^{-1}\hat{\boldsymbol{\beta}} \quad (3.12)$$

olarak yazılabilir. Bu bilgilere göre, aynı $k\mathbf{I}_p$ matrisini kullanarak, ridge kestiricinin yan ve varyansı elde edilebilir:

Ridge kestiricisinin yanı

Yukarda belirtildiği gibi, Ridge Regresyon'un beklenen değeri,

$$E[\hat{\boldsymbol{\beta}}_R] = E[(\mathbf{X}'\mathbf{X} + k\mathbf{I}_{pp})^{-1}\mathbf{X}'\mathbf{Y}] = (\mathbf{X}'\mathbf{X} + k\mathbf{I}_{pp})^{-1}(\mathbf{X}'\mathbf{X})\hat{\boldsymbol{\beta}}$$

olarak belirtilmişti. İşlemlerde kolaylık sağlaması açısından $\mathbf{B}_k = \mathbf{X}'\mathbf{X} + k\mathbf{I}_{pp}$ olarak tanımlansın. Buna göre, Ridge Regresyon kestiricinin yanı,

$$\begin{aligned}
\text{Yan}(\widehat{\boldsymbol{\beta}}_R) &= E[\widehat{\boldsymbol{\beta}}_R] - \boldsymbol{\beta} = E[(\mathbf{B}_k)^{-1}\mathbf{X}'\mathbf{Y}] - \boldsymbol{\beta} = (\mathbf{B}_k)^{-1}(\mathbf{X}'\mathbf{X})\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta} \\
&= (\mathbf{B}_k)^{-1}(\mathbf{X}'\mathbf{X} + k\mathbf{I}_p - k\mathbf{I}_p)\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta} = (\mathbf{I}_p - k(\mathbf{B}_k)^{-1})\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta} \\
&= -k(\mathbf{B}_k)^{-1}\boldsymbol{\beta}
\end{aligned}$$

olarak elde edilir.

Ridge kestiricisinin varyansı

$$\text{Var}[\widehat{\boldsymbol{\beta}}_R] = \text{Var}\left[(\mathbf{X}'\mathbf{X} + k\mathbf{I}_{pp})^{-1}\mathbf{X}'\mathbf{Y} = (\mathbf{X}'\mathbf{X} + k\mathbf{I}_{pp})^{-1}(\mathbf{X}'\mathbf{X})\widehat{\boldsymbol{\beta}}\right]$$

Burada daha öncede belirtildiği gibi, $\mathbf{A}_k = (\mathbf{X}'\mathbf{X} + k\mathbf{I}_{pp})^{-1}(\mathbf{X}'\mathbf{X})$ olduğu dikkate alınacak olursa, ridge kestiricisinin varyansı,

$$\begin{aligned}
\text{Var}[\widehat{\boldsymbol{\beta}}_R] &= (\mathbf{A}_k)\text{Var}[\widehat{\boldsymbol{\beta}}](\mathbf{A}_k)' \\
&= \left\{(\mathbf{X}'\mathbf{X} + k\mathbf{I}_{pp})^{-1}(\mathbf{X}'\mathbf{X})\right\} \sigma^2 (\mathbf{X}'\mathbf{X})^{-1} \left\{(\mathbf{X}'\mathbf{X} + k\mathbf{I}_{pp})^{-1}(\mathbf{X}'\mathbf{X})\right\}' \\
&= \sigma^2 (\mathbf{X}'\mathbf{X} + k\mathbf{I}_p)^{-1} \mathbf{X}'\mathbf{X} \left\{(\mathbf{X}'\mathbf{X} + k\mathbf{I}_p)^{-1}\right\}'. \tag{3.13}
\end{aligned}$$

olarak ifade edilir. Eşitlik (3.12)'nin elde edilmesinde, rassal olmayan bir \mathbf{C} matrisi için $\text{Var}[\mathbf{C}\mathbf{y}] = \mathbf{C}\text{Var}(\mathbf{y})\mathbf{C}'$ eşitliği ve $\text{Var}[\widehat{\boldsymbol{\beta}}] = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$ ifadesi kullanılmıştır.

Yukarıdaki bilgilerden hareketle, (3.13)'de ifade edilen ridge kestiricisinin varyansının açık bir ifadesiyle, EKK kestiricisi veya eşdeğer olarak maksimum olabilirlik (maximum likelihood-ML) kestiricisini karşılaştırabiliriz. Bilindiği gibi, herhangi bir doğrusal regresyon modelinde yer alan parametreleri tahmin etmek için kullanılan ML ve EKK kestiricileri eşdeğerdir ancak onlar varyans açısından farklılık gösterirler. EKK ve ridge kestiricilerin varyansları karşılaştırması anlamında değerlendirildiğinde, söz konusu iki kestiricinin varyansları farkı,

$$\begin{aligned}
&\text{Var}[\widehat{\boldsymbol{\beta}}] - \text{Var}[\widehat{\boldsymbol{\beta}}_R] \\
&= \sigma^2(\mathbf{X}'\mathbf{X})^{-1} - \sigma^2(\mathbf{A}_k)'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{A}_k \\
&= \sigma^2\{(\mathbf{A}_k)'((\mathbf{A}_k)')^{-1}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{A}_k^{-1}\mathbf{A}_k - (\mathbf{A}_k)'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{A}_k\} \\
&= \sigma^2(\mathbf{A}_k)' \{((\mathbf{A}_k)')^{-1}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{A}_k^{-1} - (\mathbf{X}'\mathbf{X})^{-1}\} \mathbf{A}_k \\
&= \sigma^2(\mathbf{A}_k)' \{(\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{X} + k\mathbf{I})(\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{X} + k\mathbf{I})(\mathbf{X}'\mathbf{X})^{-1} - (\mathbf{X}'\mathbf{X})^{-1}\} \mathbf{A}_k \\
&= \sigma^2(\mathbf{A}_k)' \{(I + k(\mathbf{X}'\mathbf{X})^{-1})(\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{X} + k\mathbf{I})(\mathbf{X}'\mathbf{X})^{-1} - (\mathbf{X}'\mathbf{X})^{-1}\} \mathbf{A}_k \\
&= \sigma^2(\mathbf{A}_k)' \{(\mathbf{X}'\mathbf{X})^{-1} + k(\mathbf{X}'\mathbf{X})^{-2}(\mathbf{I} + k(\mathbf{X}'\mathbf{X})^{-1}) - (\mathbf{X}'\mathbf{X})^{-1}\} \mathbf{A}_k \\
&= \sigma^2(\mathbf{A}_k)' \{(\mathbf{X}'\mathbf{X})^{-1} + k(\mathbf{X}'\mathbf{X})^{-2} + k(\mathbf{X}'\mathbf{X})^{-2} + k^2(\mathbf{X}'\mathbf{X})^{-3} - (\mathbf{X}'\mathbf{X})^{-1}\} \mathbf{A}_k
\end{aligned}$$

$$\begin{aligned}
&= \sigma^2 (\mathbf{A}_k)' \{2k(\mathbf{X}'\mathbf{X})^{-2} + k^2(\mathbf{X}'\mathbf{X})^{-3}\} \mathbf{A}_k \\
&= \sigma^2 (\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1} \mathbf{X}'\mathbf{X} \{2k(\mathbf{X}'\mathbf{X})^{-2} + k^2(\mathbf{X}'\mathbf{X})^{-3}\} \mathbf{X}'\mathbf{X} (\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1} \\
&= \sigma^2 (\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1} \{2k\mathbf{I} + k^2(\mathbf{X}'\mathbf{X})^{-1}\} \{(\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1}\}'
\end{aligned}$$

olarak ifade edilir. Bu eşitlikte verilen her bir bileşendeki çarpım matrisi pozitif tanımlı olduğu için fark da pozitif tanımlanır. Dolayısıyla, onların tersleri mevcuttur. Böylece, ML ya da EKK kestiricisinin varyansı, ridge kestiricinin varyansından büyük olur. Diğer bir deyişle,

$$Var[\hat{\boldsymbol{\beta}}] > Var[\hat{\boldsymbol{\beta}}_R] \quad (3.14)$$

Yukarıdaki bilgilerin yanı sıra, $\hat{\boldsymbol{\beta}}_R$ olarak ifade edilen ridge kestiricilerinin performanslarını değerlendirmek için yaygın olarak kullanılan “*ridge kestiricisinin hata kareler ortalaması (mean squared error-MSE)*” ifade edilebilir.

Ridge kestiricinin MSE değeri

$$\begin{aligned}
MSE[\hat{\boldsymbol{\beta}}_R] &= E(\sum_{j=1}^p (\hat{\beta}_j - \beta_j)^2) = E[(\hat{\boldsymbol{\beta}}_R - \boldsymbol{\beta})'(\hat{\boldsymbol{\beta}}_R - \boldsymbol{\beta})], \\
&= E[(\hat{\boldsymbol{\beta}}_R - E[\hat{\boldsymbol{\beta}}_R] + E[\hat{\boldsymbol{\beta}}_R] - \boldsymbol{\beta})'(\hat{\boldsymbol{\beta}}_R - E[\hat{\boldsymbol{\beta}}_R] + E[\hat{\boldsymbol{\beta}}_R] - \boldsymbol{\beta})] \\
&= E[(\hat{\boldsymbol{\beta}}_R - E[\hat{\boldsymbol{\beta}}_R])'(\hat{\boldsymbol{\beta}}_R - E[\hat{\boldsymbol{\beta}}_R])] \\
&\quad + E[E[\hat{\boldsymbol{\beta}}_R] - \boldsymbol{\beta}]' E[E[\hat{\boldsymbol{\beta}}_R] - \boldsymbol{\beta}] + E[(\hat{\boldsymbol{\beta}}_R - E[\hat{\boldsymbol{\beta}}_R])'(E[\hat{\boldsymbol{\beta}}_R] - \boldsymbol{\beta})] \\
&\quad + E[(E[\hat{\boldsymbol{\beta}}_R] - \boldsymbol{\beta})'(\hat{\boldsymbol{\beta}}_R - E[\hat{\boldsymbol{\beta}}_R])]
\end{aligned}$$

$E[E[\hat{\boldsymbol{\beta}}_R]] = E[\hat{\boldsymbol{\beta}}_R]$ ve $E[\boldsymbol{\beta}] = \boldsymbol{\beta}$ alınacak olursa,

$$\begin{aligned}
MSE[\hat{\boldsymbol{\beta}}_R] &= E[(\hat{\boldsymbol{\beta}}_R - E[\hat{\boldsymbol{\beta}}_R])'(\hat{\boldsymbol{\beta}}_R - E[\hat{\boldsymbol{\beta}}_R])] + E[\hat{\boldsymbol{\beta}}_R - \boldsymbol{\beta}]'(E[\hat{\boldsymbol{\beta}}_R] - \boldsymbol{\beta}) \\
&+ [E[\hat{\boldsymbol{\beta}}_R] - E[\hat{\boldsymbol{\beta}}_R]]'(E[\hat{\boldsymbol{\beta}}_R] - \boldsymbol{\beta}) + (E[\hat{\boldsymbol{\beta}}_R] - \boldsymbol{\beta})'(E[\hat{\boldsymbol{\beta}}_R] - E[\hat{\boldsymbol{\beta}}_R]),
\end{aligned}$$

Son iki terim sıfır vektörü ile çarpılırsa, *Ridge kestiricinin matris MSE değeri*,

$$\begin{aligned}
MSE(\hat{\boldsymbol{\beta}}_R) &= E[(\hat{\boldsymbol{\beta}}_R - E[\hat{\boldsymbol{\beta}}_R])'(\hat{\boldsymbol{\beta}}_R - E[\hat{\boldsymbol{\beta}}_R])] + (E[\hat{\boldsymbol{\beta}}_R] - \boldsymbol{\beta})'(E[\hat{\boldsymbol{\beta}}_R] - \boldsymbol{\beta}) \\
&= Var(\hat{\boldsymbol{\beta}}_R) + (\hat{\boldsymbol{\beta}}_R' \text{deki yan})^2. \quad (3.15)
\end{aligned}$$

olarak ifade edilir (bu konuda ayrıntılı bilgi için bkz., Wessel N. van Wieringen, 2018 tarafından verilen ders notları).

Eşitlik (3.15)'de görüldüğü gibi, ridge kestiricisinin MSE değeri, kareli yan ve varyans toplamından oluşmaktadır. $Var(\hat{\boldsymbol{\beta}}_R)$ ve $(\hat{\boldsymbol{\beta}}_R' \text{deki yan})^2$ terimleri yerine karşılıkları yazılacak olursa matris değerli MSE,

$$\begin{aligned}
MSE(\hat{\boldsymbol{\beta}}_R) &= Var(\hat{\boldsymbol{\beta}}_R) + (\hat{\boldsymbol{\beta}}_R' \text{deki yan})^2 \\
&= \sigma^2 (\mathbf{X}'\mathbf{X} + k\mathbf{I}_p)^{-1} \mathbf{X}'\mathbf{X} (\mathbf{X}'\mathbf{X} + k\mathbf{I}_p)^{-1} + (-k(\mathbf{B}_k)^{-1}\boldsymbol{\beta})^2 \\
&= \sigma^2 iz \left[(\mathbf{X}'\mathbf{X} + k\mathbf{I}_p)^{-1} \mathbf{X}'\mathbf{X} (\mathbf{X}'\mathbf{X} + k\mathbf{I}_p)^{-1} \right] + k^2 \boldsymbol{\beta}' (\mathbf{B}_k)^{-2} \boldsymbol{\beta} \\
&= \sigma^2 \sum_{j=1}^p \frac{\lambda_j}{(\lambda_j + k)^2} + k^2 \boldsymbol{\beta}' (\mathbf{X}'\mathbf{X} + k\mathbf{I}_{pp})^{-2} \boldsymbol{\beta} \quad (3.15a)
\end{aligned}$$

olarak ifade edilir. Burada $\lambda_1, \dots, \lambda_p$ sembolleri $\mathbf{X}'\mathbf{X}$ matrisinin öz değerlerini göstermektedir. (3.15a) eşitliğinin sağ tarafındaki ilk terim, $\hat{\boldsymbol{\beta}}_R = (\hat{\beta}_1, \dots, \hat{\beta}_p)'$ parametrelerin varyansları toplamını; ikinci terim ise, yanların karesini göstermektedir. $k > 0$ parametresi artarken $\hat{\boldsymbol{\beta}}_R$ 'deki yanlar artar, ancak, varyans azalır. Diğer bir deyişle, (3.15a) denkleminde göre ridge parametresi ve varyans ve yan arasında şu şekilde bir ilişki vardır:

- $k \rightarrow \infty$ iken, $\sigma^2 \rightarrow 0$ (k artarken varyans azalıyor),
- $k \rightarrow \infty$ iken, $k^2 \boldsymbol{\beta}' (\mathbf{X}'\mathbf{X} + k\mathbf{I}_{pp})^{-2} \boldsymbol{\beta} \rightarrow \boldsymbol{\beta}^2$ (k artarken yanlar artıyor).

Eşitlik (3.15a)'da verilen MSE matrisi dışında bir $\hat{\boldsymbol{\beta}}_R$ kestiricisini uyum iyiliğini ölçmenin bir diğer yolu, $\hat{\boldsymbol{\beta}}_R$ kestiricisinin skaler değerli MSE ölçüsü olarak bilinen SMDE (scalar mean squared error) değeridir:

$$\begin{aligned}
SMSE(\hat{\boldsymbol{\beta}}_R) &= iz[Var(\hat{\boldsymbol{\beta}}_R)] + (\hat{\boldsymbol{\beta}}_R' \text{deki yan})^2 \\
&= iz[Var(\hat{\boldsymbol{\beta}}_R)] + k^2 \boldsymbol{\beta}' (\mathbf{B}_k)^{-2} \boldsymbol{\beta} \quad (3.15b)
\end{aligned}$$

Diğer yandan, yansız kestirici olan EKK ve ridge regresyon kestiricilerinin MSE değerlerini arasındaki fark,

$$MSE(\hat{\boldsymbol{\beta}}) - MSE(\hat{\boldsymbol{\beta}}_R) = iz(Var(\hat{\boldsymbol{\beta}}) - Var(\hat{\boldsymbol{\beta}}_R)) - \|\text{bias}(\hat{\boldsymbol{\beta}}_R)\|^2 \quad (3.16)$$

eşitlik (3.16)'da belirtildiği gibi, ridge düşük varyanslı ancak yüksek bir yana sahiptir.

Burada iki matrisin izlerinin toplamının, toplamlarının izine eşit olduğu gerçeğini kullandık. Böylece iki terim arasında bir farklılık vardır. Daha önceden belirtildiği gibi, matrisler pozitif tanımlı olduğunu için, izde kesinlikle pozitiftir. Yan terimi de karesinden dolayı kesinlikle pozitif olacaktır. Bu nedenle, aradaki fark pozitif veya negatif olabilir. Farkın pozitif veya negatif olup olmadığını ceza parametresine bağlı olduğunu kanıtlamak mümkündür (bkz. Theobald 1974 ve Farebrother 1976). Ayrıca, farkın pozitif olacağı şekilde bir değer bulmak her zaman olası bir durumdur. Başka bir deyişle, ridge tahmincisi, EKK tahmincisinden daha düşük MSE değerine

sahip olabileceğini ortaya koyan bir ceza parametresinin değerini bulmak mümkündür. Bu sonuç hem pratik hem de teorik açıdan çok önemlidir. Gauss-Markov teoremine göre, yansız kestiriciler arasında EKK kestiricisi, en düşük düşük varyansa (dolayısıyla en düşük MSE değerine), sahip olsa da, EKK kestiricilerinden daha düşük bir MSE değerine sahip olan yanlı bir kestirici vardır.

3.5. Model Varyansının Tahmini

Regresyon problemlerini dikkate alan istatistiksel modeller için σ^2 olarak gösterilen bilinmeyen hata varyansı genellikle tahmin edilir. Örneğin, (3.12) eşitliliğinde belirtilen ridge kestiricinin varyansı hesaplamak için σ^2 değeri bilinmediği için onu tahmini ile yer değiştirmek gerekir. Pratikte en sık kullanılan σ^2 değerini tahmin etmenin en basit yolu sıradan EKK durumunda olduğu gibi, ridge modelinin artık kareler toplamı yarımıyla hesaplanabilir:

$$\hat{\sigma}^2 = \frac{(Y - X\hat{\beta}_R)(Y - X\hat{\beta}_R)}{\text{Serbestlik Derecesi}(DF)} = \frac{(Y - X\hat{\beta}_R)(Y - X\hat{\beta}_R)}{\text{iz}(\mathbf{H}_k)} = \frac{(Y - \hat{Y})(Y - \hat{Y})}{\text{iz}(\mathbf{H}_k)} \quad (3.17)$$

Burada DF ile gösterilen serbestlik derecesi sıradan EKK yönteminde örneklem hacmi ile tahmin edilen parametre sayısı arasındaki fark olarak alınır. Diğer bir deyişle, $DF = n - p$. Ancak yanlı tahmin yöntemlerinde genellikle \mathbf{H}_k ile gösterilen şapka matrisinin izi kullanılır (Hastie and Tibshirani, 1990).

Ridge Regresyonu'nda yanıt gözlemlerinin uyumları, sıradan EKK yöntemine benzer olarak aşağıdaki gibi tanımlanabilir:

$$\hat{Y} = X\hat{\beta}_R = X(X'X + kI)^{-1}X'Y = \mathbf{H}_k Y \quad (3.18)$$

Burada açıkça görülen o ki, gerçek yanıt gözlemlerini uyum değerlerine dönüştüren şapka matrisi,

$$\mathbf{H}_k = X(X'X + kI)^{-1}X' = \mathbf{XDX}', \text{ burada } \mathbf{D} = (X'X + kI)^{-1} \quad (3.19)$$

olarak ifade edilmiştir. Bu şapka matrisinin izi, etkin serbestlik derecesi olarak tanımlanır:

$$EDF = \text{iz}(\mathbf{H}_k) = n - \text{iz}(2\mathbf{H}_k - \mathbf{H}_k'\mathbf{H}_k)$$

Böylece, eşitlik (3.17)'de tanımlanan yanlı kestiriciler için model varyansının tahmini aşağıdaki gibi yeniden yazılabilir:

$$\hat{\sigma}^2 = \frac{(Y - X\hat{\beta}_R)(Y - X\hat{\beta}_R)}{EDF} = \frac{(Y - X\hat{\beta})'(Y - X\hat{\beta}) + (\hat{\beta}_R - \hat{\beta})'X'X(\hat{\beta}_R - \hat{\beta})}{n - iz(2H_k - H_k'H_k)} \quad (3.20)$$

Burada n örneklem hacmini ve EDF ise, etkin serbestlik derecesini göstermektedir.

3.6. Ridge Regresyonuna İlişkin Bazı Bilgiler

Standart EKK yönteminde ölçek değişmez, ancak Ridge Regresyonu gibi cezalandırılmış yöntemler için, ölçeklendirme önemlidir. Çünkü katsayıların tümü bir ceza teriminde bir araya getirilir:

$$CHKT(\hat{\beta}, k) = \min\{\sum_{i=1}^n (Y_i - X_i \hat{\beta})^2 + k \sum_{j=1}^p \hat{\beta}_j^2\}$$

Eğer değişkenlerin birimleri değişirse, katsayıların ölçekleri değişecektir. Açıklayıcı değişkenlerin birimleri Ridge Regresyonu'ndaki cezayı ($k \sum_{j=1}^p \hat{\beta}_j^2$) etkilemektedir. Böylece, hem yanıt hem de katsayılar değişecektir. Bu problemi gidermek için Ridge Regresyonu yapmadan önce açıklayıcı değişkenleri standart hale getirmek önemli bir etmendir.

Aşağıda verilen EKK kestiricini dikkate aldığımızı varsayalım:

$$\hat{\beta} = (X'X)^{-1}X'Y$$

Ayrıca, burada yeniden ölçeklendirilen matris, $Z = XC$ olarak tanımlansın. Bu yeni matrisle ilişkili EKK tahmini,

$$\begin{aligned} \tilde{\beta} &= (Z'Z)^{-1}Z'Y = (C'X'XC)^{-1}C'X'Y \\ &= C^{-1}(X'X)^{-1}(C')^{-1}C'X'Y = (C')^{-1}(X'X)^{-1}X'Y = (C')^{-1}\hat{\beta} \end{aligned} \quad (3.21)$$

olarak tanımlanır. Buradan görüldüğü gibi, yeni tasarım matrisi ile ilişkili EKK tahmini, açıklayıcı değişkenlerin nasıl yeniden ölçeklendiğine bakılmaksızın, her zaman aynı sonucu verecektir. Bu durum maalesef ridge tahmincisi bulmayan EKK tahmincisinin güzel bir özelliğidir. Bu sefer ridge kestiricisini dikkate aldığımızı varsayalım:

$$\hat{\beta}_R = (X'X + kI_{pp})^{-1}X'Y$$

Bu durumda, yeniden ölçeklendirilen matris $Z = XC$ ile ilişkili ridge tahmini,

$$\begin{aligned} \tilde{\beta}_R &= (Z'Z + kI_{pp})^{-1}Z'Y = (C'X'XC + kI_{pp})^{-1}C'X'Y \\ &= [C'X'XC + kC'(C')^{-1}C^{-1}C]^{-1}C'X'Y \end{aligned}$$

$$\begin{aligned}
&= [\mathbf{C}'(\mathbf{X}'\mathbf{X} + k(\mathbf{C}')^{-1}\mathbf{C}^{-1})\mathbf{C}]^{-1}\mathbf{C}'\mathbf{X}'\mathbf{Y} \\
&= (\mathbf{C}')^{-1}[\mathbf{X}'\mathbf{X} + k(\mathbf{C}')^{-1}\mathbf{C}^{-1}]^{-1}(\mathbf{C}')^{-1}\mathbf{C}'\mathbf{X}'\mathbf{Y} \\
&= (\mathbf{C}')^{-1}[\mathbf{X}'\mathbf{X} + k(\mathbf{C}')^{-1}\mathbf{C}^{-1}]^{-1}\mathbf{X}'\mathbf{Y} \tag{3.22}
\end{aligned}$$

Belirtmek gerekir ki yeniden ölçeklendirilen matris ile ilişki ridge kestiricisi, ancak ve ancak $(\mathbf{C}')^{-1}\mathbf{C}^{-1} = \mathbf{I}$ veya $\mathbf{C}'\mathbf{C} = \mathbf{I}$ ise, $\tilde{\boldsymbol{\beta}}_R = (\mathbf{C}')^{-1}\hat{\boldsymbol{\beta}}_R$ olduğu görülmektedir. Diğer bir deyişle, \mathbf{C} ölçek matrisi sadece orta-normal gibi özel bir durumda, ridge kestiricisi ölçeği değişmezdir.

Genel anlamda ölçek değişmezliğinin olmayışı, değişkenlerin ölçeklendirilmesiyle ilgili olarak yaptığımız herhangi bir seçimin (örneğin, santimetre, metre veya binlerce, milyonlarca dolar cinsinden bir açıklayıcı değişkenin ifade edilmesi) katsayı tahminlerini etkilediği anlamına gelir. Bu oldukça istenmeyen bir durum olduğu için, genelde yaptığımız şey regresyondaki tüm değişkenleri standartlaştırmaktır. Diğer bir deyişle, değişkenleri ölçekten bağımsız hale getirmek gerekir. Bu bağlamda Ridge Regresyonu yapmadan önce “*değişkenleri ortalamalarından çıkartıp ve standart sapmasına bölerek standart hale getirmek gerekir*”. Böylece, katsayı tahminleri değişkenlerin ölçeklendirilmesinin isteğe bağlı seçimlerinden etkilenmez.

3.7. Ridge Parametresinin Seçimi

Ridge kestiricisinin MSE açısından EKK kestiricilerinden daha iyi olduğunu bir önceki kısımlarda belirtilmişti. Burada temel sorun, bunu ortaya koyan bir ceza parametresini seçmektir. Bu bölümde, ana düşünce söz konusu optimum bir ceza parametresinin seçiminin nasıl yapılacağını incelemektir. Ridge Regresyon ile yapılacak olan kestirimlerin kararlı ve optimal bir k değerinin belirlenmesi için çok önemlidir. Ridge parametresinin çözümü k 'ya bağlı olarak yapılır ve dolayısıyla her k için Ridge Regresyon katsayı değerleri elde edilir. k 'nın uygun değerinin seçilmesi durumunda ridge kestiricilerinin MSE değerleri, EKK kestiricisinin MSE değerlerinden daha küçük olacaktır.

En uygun k , aynı zamanda denklemler (3.15) veya (3.15a)'daki minimum MSE'yi veren değerdir. Belirtmek gerekirse, k ceza parametresi, gerçek regresyon katsayılarına ve artıkların varyansına bağlıdır. Ancak bunlar bilinmediği için onların

örnek verilerinden tahmin edilmesi gerekir. Ridge çözümleri EKK ile karşılaştırıldığında, geniş bir k aralığı ($0 < k < k_{opt}$) daha küçük MSE değerini verirken, herhangi bir problem için optimal bir k vardır. Çoklu bağlantı içeren veriler için k 'daki küçük bir değişme Ridge Regresyon katsayılarını hızlı bir biçimde değiştirir. Bazı k değerlerinde ridge katsayıları dengelenir ve değişim hızı sıfıra doğru kademeli olarak yavaşlar. Bu nedenle, MSE'yi en aza indiren çekme ya da ceza parametresini seçmenin bir yolunu bulmak gerekir.

Teorik ve pratik olarak Ridge Regresyon, ridge tahmin edicisinin özelliklerini ortaya çıkarmak için yan parametresi k 'nın seçimi için bazı yeni yöntemler kullanmıştır. Ridge regresyonda yan parametresinin seçimi en kilit noktadır. Çünkü optimal k 'nın bulunması Ridge Regresyon için en önemli noktadır. Literatürde, yanlılık parametresi k 'yı tahmin etmek için birçok yöntem vardır (bkz., Saleh ve Kibria, 1993; Singh ve Tracy, 1999; Kibria, 2003; Khalaf ve Shukur, 2005; Alkhamisi ve diğerleri, 2006; Alkhamisi ve Shukur, 2007; Khalaf, 2013, daha birçokları arasında), ancak hangi yöntemin tercih edileceği konusunda fikir birliği yoktur (Chatterjee ve Hadi , 2006). Benzer şekilde, yanlı parametre seçimi iyi bir k vermeyi veya EKK'dakine kıyasla daha küçük MSE vermeyi garanti edemez. Ridge parametresinin seçimi, (i) subjektif metotlar ve (ii) objektif metotlar olmak üzere iki sınıfta toplanabilir.

i. Subjektif metotlar

Tüm bu yöntemlerde, k 'nın seçimi subjektif veya yargılayıcı niteliktedir. EKK tahmincisine kıyasla ridge tahmin edicisi tarafından regresyon katsayılarının tahmini ve varyansının hesaplanması doğrusallığın etkisinin grafiksel olarak kanıtlamaktadır. Bu yöntemlerde, optimum k seçimi, ridge izi, (3.17)'de tanımlanan $iz(DF) = iz(\mathbf{H}_k)$, (3.9)'da tanımlanan $iz(VIF) = iz\{diag(\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1}\}$ ve yanlılık, varyans ve MSE değerleri grafiği kullanılarak yapılır. Ridge izi $[0, 1]$ aralığında k 'nın bir fonksiyonu olarak regresyon katsayıları $\hat{\beta}_R$ katsayılarının grafiksel bir gösterimidir. DF izi ve VIF izi, EDF ve VIF değerlerinin k 'ya karşı çizildiği ridge izi grafiği gibidir. Benzer şekilde, ridge parametresinden yanlılık, varyans ve MSE değerleri grafiği uygun bir k değeri seçilmesinde yardımcı olabilir. Bütün bu grafikler,

katsayıların her biri üzerindeki doğrusallığın etkisini değerlendirmek için yatay ekseninde k 'nın optimal değerini seçmek için kullanılabilir. k değeri arttıkça ve ridge katsayılarının tüm değerleri, EDF ve VIF değerleri azaldıkça ve/veya belirli k değerlerinden sonra sabit olabilirlerse, doğrusallık etkisi azalır. Bu grafiksel gösterimler benzersiz bir çözüm sunmazken, kabul edilebilir belli belirsiz tanımlanmış bir çözüm sınıfı oluşturmaktadır. Ancak, bu izler bazı optimal k 'ları kontrol etmek için yararlı grafiksel gösterimlerdir.

ii. *Objektif metotlar*

Objektif metotlar, bir dereceye kadar, yanlı parametre k 'nın seçimindeki yargılayıcı yöntemlere benzerdir, ancak bu yanlı parametreleri elde etmek için bazı hesaplamaları gerektirir. Bu tezde k parametresini seçmek için objektif yöntemler dikkate alınmıştır. Ancak gerçek veri uygulamasında subjektif yöntemlerden ridge izi grafiğine yer verilmiştir. Bu bağlamda, tezde ridge parametresinin seçimi için kullanılan objektif yöntemler arasında yer alan klasik bir seçim yöntemi olan k_{HKB} kestirici baz kriter olarak alınmıştır. Bu kriterle ilaveten dört farklı bilgi kriteri kullanılarak ridge parametresi seçilmiş ve bu optimum k değeri kullanılarak elde edilen Ridge Regresyon sonuçlarının bir karşılaştırılması yapılmıştır. Kriterlerin açıklamaları izleyen kısımda verilmiştir.

3.7.1. HKB (Hoerl, Kennard ve Baldwin) kestiricisi

Veriye dayalı olarak ridge parametresinin tahmin etmek için literatürde çok sayıda yöntem önerilmiştir. Hoerl, Kennard ve Baldwin (1975) tarafından önerilen en popüler olan k_{HKB} kestirici, (3.1)'de ifade edilen lineer regresyon modeli kanonik formda yazıldığında, daha kolay anlaşılacaktır: (2.7) denkleminde belirtilen birim uzunluk ölçeğine göre standartlaştırılan bir $\tilde{\mathbf{X}}$ matrisi dikkate alındığında, $(\tilde{\mathbf{X}}'\tilde{\mathbf{X}})$ korelasyon matrisi olarak elde edilir. $(\tilde{\mathbf{X}}'\tilde{\mathbf{X}})$ matrisinin özdeğer ve özvektör ayrışımı, q_j sütunlu bir \mathbf{Q} öz vektör matrisini ve bu öz vektörlere karşı gelen köşegen değerleri λ_j olan bir $\mathbf{\Lambda}$ öz değer matrisini verir. $\mathbf{Z} = \tilde{\mathbf{X}}\mathbf{Q}$ ve $\boldsymbol{\alpha} = \mathbf{Q}'\boldsymbol{\beta}$ alınarak lineer regresyon modeli, kanonik formda aşağıdaki gibi yazılabilir:

$$\mathbf{Y} = \mathbf{Z}\boldsymbol{\alpha} + \boldsymbol{\varepsilon} \quad (3.23)$$

Burada \mathbf{Z} matrisinin sütunları $\tilde{\mathbf{X}}$ 'in temel bileşenleridir. $\boldsymbol{\alpha}$ için EKK kestirici aynı zamanda temel bileşenlerin bir alt kümesinin regresyon katsayılarını verir:

$$\hat{\boldsymbol{\alpha}} = \boldsymbol{\Lambda}^{-1}\mathbf{Z}'\mathbf{Y} \quad (3.24)$$

Yukarıdaki bilgiler ışığında, ridge parametresi k 'nın tahmini olarak Hoerl, Kennard ve Baldwin (1975) tarafından aşağıdaki şekilde önerilmiştir:

$$k_{HKB} = p\hat{\sigma}^2/\hat{\boldsymbol{\alpha}}'\hat{\boldsymbol{\alpha}} = p\hat{\sigma}^2/\hat{\boldsymbol{\beta}}'\hat{\boldsymbol{\beta}} = p\hat{\sigma}^2/||\hat{\boldsymbol{\beta}}||^2 \quad (3.25)$$

Burada $\hat{\sigma}^2 = \frac{(\mathbf{Y}-\mathbf{X}\hat{\boldsymbol{\beta}})(\mathbf{Y}-\mathbf{X}\hat{\boldsymbol{\beta}})'}{(n-p)}$ ve $\hat{\boldsymbol{\beta}} = (\beta_1, \dots, \beta_p)'$ önceki bölümlerde bahsedildiği gibi, sıradan EKK tahminleridir. Bu ridge kestiricisi eşitlik (3.15)'de belirtilen MSE'yi minimum yapma açısından genelleştirilmiş ridge kestiricinin harmonik ortalaması olarak motive edilebilir. Bu kestiricideki $\hat{\boldsymbol{\alpha}}$, (3.24)'de tanımlanan temel bileşenler regresyonu katsayılarıdır.

3.7.2. Mallows'un C_k kriteri:

Mallows(1973) tarafından önerilen C_k istatistiği, ridge parametresi k değerini tahmin etmek için aşağıdaki gibi tanımlanabilir:

$$\begin{aligned} C_k &= \frac{\sum_{i=1}^n (y_i - X_i \hat{\beta}_R)^2}{\hat{\sigma}^2} - n + 2 + 2tr(\mathbf{XL}) \\ &= \frac{SSE(k)}{\hat{\sigma}^2} - n + 2 + 2tr(\mathbf{XL}) \\ &= \frac{SSE(k)}{\hat{\sigma}^2} - n + 2 + 2tr[X(\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1}\mathbf{X}'] \\ &= \frac{SSE(k)}{\hat{\sigma}^2} - n + 2 + tr(\mathbf{H}_k) \end{aligned} \quad (3.26)$$

Burada $\hat{\sigma}^2$, bir önceki denklemde belirtildiği gibi En Küçük Kareler kestiricisidir. Yukarıda ifade edilen C_k 'yi en küçük yapan k değeri ridge parametresi olarak seçilir. Ayrıca, eşitlik (3.26)'da yer verilen \mathbf{H}_k , verilen bir $k > 0$ parametresi için (3.19) eşitliğinde tanımlanan şapka matrisidir.

3.7.3. Çapraz geçerlilik kriteri

Ridge parametresi k 'yı tahmin etmek için kullanılan en yaygın yöntemler arasında çapraz geçerlilik (cross validation-CV) yer alır. CV kriteri gerçek beklenen kestirim hatası için yaklaşık yansız bir tahmindir. Söz konusu CV kriteri,

$$CV_k = \frac{1}{n} \sum_{i=1}^n (Y_i - X_i \hat{\beta}_{R(-i)})^2 \quad (3.27)$$

Burada $\hat{\beta}_{R(-i)}$, i . gözlem kaldırıldıktan sonra eksik gözlem ile yapılan tahmin değerini göstermektedir. Buradan anlaşıldığı gibi, denklem (3.8)'i hesaplamak için her defasında yeni bir model tahmin etmek gerekir. Örneğin, 30 gözlemeli bir örneklem varsa 30 tane regresyon modeli tahmin etmek gerekir. Onun yerine hiçbir gözlem atılmaksızın CV kriterini hesaplama mümkündür:

$$CV_k = \frac{1}{n} \sum_{i=1}^n \left(\frac{Y_i - X_i \hat{\beta}_R}{1 - (\mathbf{H}_k)_{ii}} \right)^2 = \left(\frac{(\mathbf{I} - \mathbf{H}_k)\mathbf{Y}}{1 - (\mathbf{H}_k)_{ii}} \right)^2 \quad (3.27a)$$

Burada $(\mathbf{H}_k)_{ii}$, (3.19) eşitliğinde tanımlanan \mathbf{H}_k şapka matrisinin köşegen değerlerini göstermektedir. (3.27a)'yı minimum yapan k değeri optimum ridge parametresi olarak seçilir.

3.7.4. Genelleştirilmiş çapraz geçerlilik kriteri

Şapka matrisi \mathbf{H}_k 'nın köşegen elemanları bire eşit ve yakın bir değer olduğu varsayılabilir. Sonuç olarak, CV skoru durağan olmayabilir. Bu durum genelleştirilmiş çapraz geçerlilik (generalized cross-validation-GCV) kriterinde giderilmiştir. GCV fonksiyonu geleneksel yöntem olan CV'nin değiştirilmiş versiyonudur. (3.27a) denkleminde ifade edilen $1 - (\mathbf{H}_k)_{ii}$ faktörü ile bölünen sıradan artıklardan oluşturulan CV skoruna benzer olarak GCV skoru elde edilebilir. GCV'nin esas düşüncesi, denklem (3.27a)'daki $(1 - (\mathbf{H}_k)_{ii})$ faktörü ile $(1 - n^{-1} \text{iz}(\mathbf{H}_k))^2$ ortalama skorunu yer değiştirmektir. Böylece, sıradan CV skoruna benzer olarak GCV skor fonksiyonu,

$$GCV_k = \frac{1}{n} \sum_{i=1}^n \frac{(Y_i - X_i \hat{\beta}_R)^2}{(1 - n^{-1} \text{iz}(\mathbf{H}_k))^2} = \frac{n^{-1} \|(\mathbf{I} - \mathbf{H}_k)\mathbf{Y}\|^2}{(1 - n^{-1} \text{iz}(\mathbf{H}_k))^2} \quad (3.28)$$

olarak ifade edilebilir.

Ridge parametresi k için, iyi bir tahmin değeri seçmek için Golub, Wahba ve Heath (1979)' da Genelleştirilmiş Çapraz Geçerlilik (GCV) kriterini önermişlerdir. Bu kriteri minimum yapan indise karşı gelen değer ridge parametresi (k) olarak dikkate alınır ve aynı zamanda bu k değeri en iyi ridge parametresi olarak seçilir.

3.7.5. Düzeltilmiş akaike bilgi kriteri

Bu yaklaşım, birçok farklı k değeri ile modeli tahmin etmeye olanak sağlar. Hurvich ve Tsai (1989) tarafından önerilen klasik Akaike Bilgi Kriteri (Akaike Information Criterion-AIC) bilgi kriterinin düzeltilmiş (corrected) bir versiyonu olan AICc kriteri geliştirilmiştir. Söz konusu düzeltilmiş AICc Kriteri,

$$AICc_k = \log \frac{\|(I-H_k)Y\|^2}{n} + 1 + \frac{2\{iz(H_k)+1\}}{n-iz(H_k)-2} \quad (3.29)$$

olarak formüle edilmiştir.

Denklem (3.29)'dan görüldüğü gibi uygulaması kolay ve varyans tahmini gerektirmeyen bir kriterdir. Öncekilerde olduğu gibi bu kriteri minimum yapan k değeri optimum ridge parametresi olarak alınabilir.

3.7.6. Schwarz'ın bayes bilgi kriteri

Bayes kestiricilerini kullanan Bayes Bilgi Kriteri (Bayes Information Criterion-BIC), Schwarz (1978) tarafından geliştirilmiştir. Böylece, BIC kriteri aynı zamanda, Schwarz bilgi kriteri (SIC) olarak da bilinir. Genel bir model seçimi için Schwarz SIC kriterini türetirken, buna karşın Akaike doğrusal regresyonda model seçimi problemleri için AIC model seçim kriterini türetmiştir. Söz konusu BIC kriteri aşağıdaki şekilde ifade edilmiştir:

$$BIC_k = \frac{\|(I-H_k)Y\|^2}{n} + \left(\frac{\log(n)}{n}\right)p \quad (3.30)$$

Burada n örneklem hacmi ve p model parametre sayısı göstermektedir. BIC kriteri aşırı uyumu önlemede AIC kriterinden daha büyük bir cezaya sahiptir. Ancak, SIC bazen iyi uyum göstermez. Dolayısıyla seçilen model çok az parametre içermektedir. Bu durum özellikle küçük örneklem problemlerinde görülmektedir.

4. UYGULAMA

4.1 Monte Carlo Simülasyon Çalışması

Bu bölümde, ridge parametresinin seçilmesi için kullanılacak beş yöntemin (AIC_c , BIC , GCV , C_p ve HKB) karşılaştırılması bir Monte Carlo Simülasyon çalışması ile yapılmıştır. Regresyon modelinde yer alan yanıt değişkeni aşağıdaki modelden üretilmiştir:

$$Y_i = 0.5 + 1X_1 + 2X_2 - 1.5X_3 + 4X_4 - 0.5X_5 + \varepsilon_i, \quad i = 1, \dots, n \quad (4.1)$$

Burada regresyon katsayıları vektörü $\beta = (0.5, 1, 2, -1.5, 4, -0.5)^T$ olarak alınmıştır. Bunun yanı sıra, modelde yer alan tüm açıklayıcı değişkenler birbirleriyle ilişkili olacak şekilde aşağıdaki biçimde ifade edilen denklemlerden yaratılmıştır:

$$X_{1i} = Z_{1i}\rho + \sqrt{1 - \rho^2}Z_{2i},$$

ve

$$X_{ji} = X_{(j-1)i}\rho + \sqrt{1 - \rho^2}Z_{2i}, \quad j = 2, 3, 4, 5$$

Burada $Z_{1i} \sim U[0,1]$ ve $Z_{2i} \sim U[0,1]$ düzgün dağılımdan üretilen rassal değişkenlerdir. Ayrıca, modelde yer alan rassal hata terimleri sıfır ortalama ve $\sigma^2 = 0.5$ varyanslı normal dağılımdan üretilmiştir. Diğer bir deyişle, $\varepsilon_i \sim N(0,0.5)$.

Simülasyon çalışması üç farklı örneklem büyüklüğü için ($n = 25, 50, 100$) ve üç farklı korelasyon değeri için ($\rho = 0.60, 0.80, 0.99$) 1000 kez tekrar edilerek oluşturulmuştur. Farklı ridge parametreleri için oluşturulan modellerin performansları hata kareler ortalaması (MSE) ile ölçülmüştür:

$$MSE(k) = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 \quad (4.2)$$

Simülasyonda 3 farklı örneklem, 5 farklı seçim kriteri ve 3 farklı korelasyon düzeyi olmak üzere toplam 45 farklı simülasyon deneyi yapılmıştır. Sonuçlar aşağıda verilen çizelge ve grafiklerle özetlenmiştir. Tablo 4.1, Tablo 4.2 ve Tablo 4.3'de *EKK* sütunu Sıradan En Küçük Kareler (Ordinary Least Squares-OLS) regresyonundan elde edilen parametre tahminlerini, AIC_c sütunu ridge parametresinin AIC_c kriteri ile seçildiğinde, (4.1) modelinde yer alan parametrelerin bu kritere dayalı ridge tahminlerini, benzer olarak BIC , GCV , C_p and

HKB sütunları altında bu kriterler ile seçilen ridge parametrelerine dayalı (4.1) modelinin parametre tahminleri göstermektedir. Bu sonuçlar farklı korelasyon düzeyleri ve örneklem hacimleri için elde edilmiştir.

Tablo 4.1 incelendiğinde, *AICc* kriteri dikkate alındığında, örneğin $n=25$ için (4.1) modelinin tahmini,

$$\hat{Y} = -0.4452 - 0.5342X_1 - 1.7003X_2 + 2.3999X_3 + 2.8197X_4 + 3.0716X_5$$

olarak bulunurken, ridge parametresi *GCV* kriteri ile belirlendiğinde, (4.1) modelinin Ridge regresyon tahminleri

$$\hat{Y} = -0.4423 + 0.5420X_1 + 1.7022X_2 + 2.3983X_3 + 2.8160X_4 + 3.0666X_5$$

olarak bulunmuştur. Bu iki model karşılaştırıldığında *GCV* ile elde edilen sonuçların daha uygun olduğu görülmektedir. Benzer olarak, $\rho = 0.60$ ilişki düzeyi altında üretilen küçük, orta ve büyük hacimli örneklem sonuçlarına ilişkin diğer tüm Simülasyon sonuçları, Tablo 4.1’de verilmiştir.

Tablo 4.1. Her bir seçim yöntemi için $\rho=0.60$ olduğunda regresyon katsayıları tahminleri

<i>n</i>		<i>OLS</i>	<i>AICc</i>	<i>BIC</i>	<i>GCV</i>	<i>Cp</i>	<i>HKB</i>
25	β_0	4.9001	-0.4452	-0.4394	-0.4423	-0.4752	-0.4948
	β_1	-0.9607	0.5342	0.5496	0.5420	0.3980	0.2968
	β_2	-5.8311	1.7003	1.7040	1.7022	1.6636	1.6361
	β_3	3.4807	2.3999	2.3967	2.3983	2.4234	2.4396
	β_4	6.7282	2,8197	2.8123	2.8160	2.8790	2.9217
	β_5	4.3496	3,0716	3.0616	3.0666	3.1524	3.2110
50	β_0	0.2318	-0,4796	-0.4773	-0.4784	-0.4818	-0.4995
	β_1	-20.1991	0,4155	0.4259	0.4207	0.3977	0.2719
	β_2	15.4447	1,6696	1.6724	1.6710	1.6649	1.6291
	β_3	4.1729	2,4221	2.4204	2.4212	2.4246	2.4434
	β_4	3.3047	2,8736	2.8691	2.8714	2.8809	2.9320
	β_5	0.8215	3,1445	3.1384	3.1414	3.1543	3.2251
100	β_0	4.9053	-0,4909	-0.4901	-0.4905	-0.4906	-0.4986
	β_1	-22.8223	0,3427	0.3488	0.3458	0.3406	0.2650
	β_2	5.3634	1,6494	1.6511	1.6502	1.6492	1.6270
	β_3	-29.8593	2,4334	2.4325	2.4329	2.4338	2.4442

	β_4	44.7682	2,9038	2.9013	2.9025	2.9041	2.9345
	β_5	-5.1449	3,1860	3.1826	3.1843	3.1866	3.2287

Tablo 4.2 incelendiğinde, bilgi kriterlerinden GCV seçim yöntemi dikkate alındığında, örneğin $n=100$ için, (4.1) modelinin tahmini,

$$\hat{Y} = -0.3688 + 0.29557X_1 + 1.3922X_2 + 2.2721X_3 + 2.9752X_4 + 3.5397X_5$$

olarak bulunurken, ridge parametresi klasik HKB seçim yöntemi ile belirlendiğinde, (4.1) modelinin Ridge regresyon tahminleri,

$$\hat{Y} = -0.4870 + 0.0950X_1 + 1.3159X_2 + 2.2927X_3 + 3.0740X_4 + 3.6991X_5$$

olarak bulunmuştur. Örneklem hacmi arttıkça tahminlerin iyileştikleri görülmektedir. Ancak, Ridge kestiricileri yanlış kestiriciler oldukları için gerçek değerlere yakın sonuçlar elde etmek zordur. Simülasyon tekrarlarının artması sonucunda tahminlerin daha stabil oldukları görülmektedir.

Tablo 4.2. Her bir seçim yöntemi için $\rho=0.80$ olduğunda regresyon katsayıları tahminleri

n		<i>OLS</i>	<i>AICc</i>	<i>BIC</i>	<i>GCV</i>	<i>Cp</i>	<i>HKB</i>
25	β_0	0.0500	0.5803	0.5824	-0.4191	0.5762	-0.4998
	β_1	9.8948	0.9633	0.9640	0.2703	0.9619	0.1465
	β_2	4.9361	1.5701	1.5701	1.3885	1.5702	1.3387
	β_3	0.5404	2.0556	2.0550	2.2800	2.0568	2.2924
	β_4	7.7535	2.4439	2.4429	2.9954	2.4461	3.0554
	β_5	-0.6913	2.7546	2.7532	3.5663	2.7575	3.6658
50	β_0	-80.7647	0.1995	0.2015	-0.3955	0.1956	-0.4987
	β_1	94.6540	0.8368	0.8377	0.3367	0.8348	0.1396
	β_2	-5.4863	1.5699	1.5701	1.4156	1.5696	1.3378
	β_3	-53.4894	2.1565	2.1560	2.2789	2.1574	2.2964
	β_4	60.3647	2.6257	2.6248	2.9681	2.6276	3.0632
	β_5	5.6106	3.0011	2.9998	3.5210	3.0038	3.6767
100	β_0	14.2832	-0.0790	-0.0780	-0.3688	-0.0820	-0.4870
	β_1	-32.5799	0.6393	0.6404	0.2955	0.6371	0.0950
	β_2	32.1367	1.5159	1.5162	1.3922	1.5152	1.3159
	β_3	-4.4278	2.2171	2.2168	2.2721	2.2177	2.2927

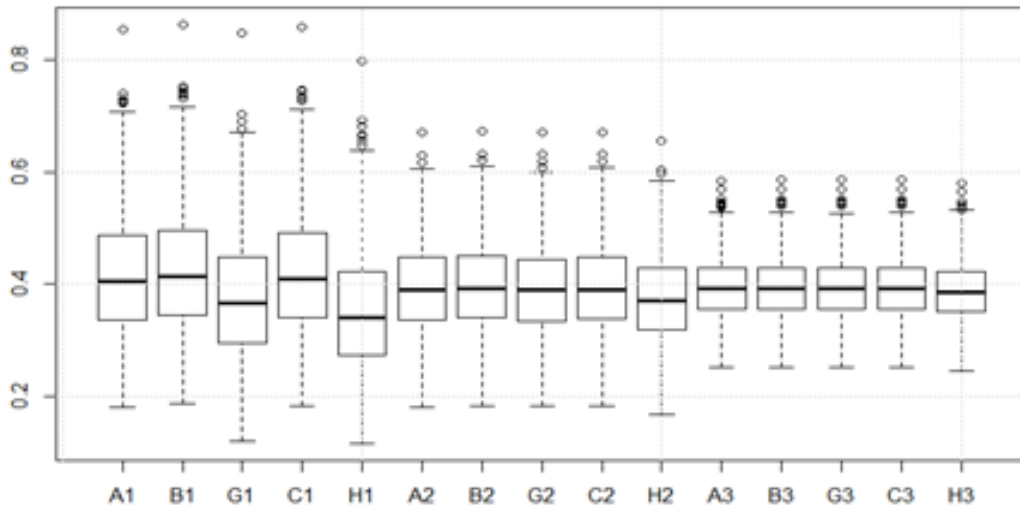
	β_4	5.1680	2.7781	2.7774	2.9752	2.7797	3.0740
	β_5	0.4812	3.2269	3.2258	3.5397	3.2293	3.6991

Tablo 4.3. Her bir seçim yöntemi için $\rho=0.99$ olduğunda regresyon katsayıları tahminleri

n		<i>OLS</i>	<i>AICc</i>	<i>BIC</i>	<i>GCV</i>	<i>Cp</i>	<i>HKB</i>
25	β_0	3.8071	1.1297	1.1314	-0.4054	1.1263	-0.3604
	β_1	23.5131	1.3797	1.3793	0.3280	1.3806	0.4691
	β_2	-22.2785	1.5510	1.5504	1.2205	1.5522	1.2903
	β_3	-14.7471	1.7206	1.7198	2.1037	1.7226	2.1033
	β_4	4.8747	1.8885	1.8875	2.9793	1.8905	2.9081
	β_5	14.4359	2.0547	2.0535	3.8450	2.0571	3.7049
50	β_0	-11.6551	0.7088	0.7109	-0.3979	0.7045	-0.4812
	β_1	-56.9572	1.4438	1.4438	0.3947	1.4438	0.2664
	β_2	-24.2128	1.6600	1.6597	1.2535	1.6607	1.1955
	β_3	83.0560	1.8741	1.8734	2.1055	1.8755	2.1153
	β_4	3.3577	2.0860	2.0850	2.9454	2.0881	3.0260
	β_5	-1.9315	2.2958	2.9451	3.7803	2.2986	3.9275
100	β_0	-0.6480	0.3694	0.3714	-0.3955	0.3656	-0.4780
	β_1	2.8637	1.3851	1.3858	0.3566	1.3838	0.1850
	β_2	-2.3418	1.6818	1.6819	1.2325	1.6817	1.1535
	β_3	6.8745	1.9756	1.9751	2.1077	1.9677	2.1109
	β_4	3.1893	2.2664	2.2653	2.9617	2.2685	3.0587
	β_5	3.5914	2.5542	2.5526	3.8134	2.5576	3.9970

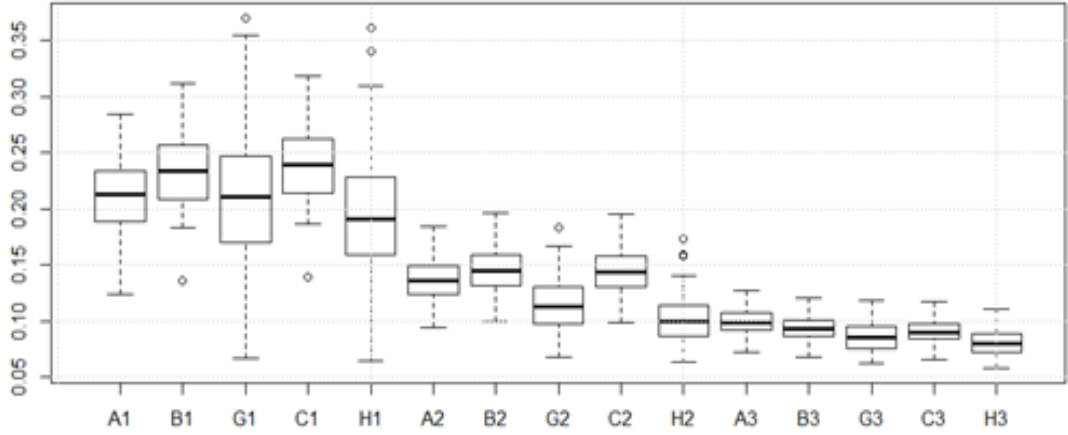
Tüm simülasyon sonuçları incelendiğinde, en uygun tahminleri, *GCV* ve *HKB* yöntemine dayalı modellerin verdiği kolaylıkla görülmektedir. Gerçek regresyon katsayılarına en yakın değerler belirtilen iki yöntem ile elde edilmişlerdir. Diğer dört yöntem, aşağıda verilmiş olan Şekil 4.1’de görüldüğü gibi kötü sonuçlar vermeseler de *GCV* ve *HKB* yanında düşük performans göstermişlerdir. Tüm yöntemler için örneklem büyüklüğü arttıkça tahminlerin iyileştiği ve korelasyon seviyesi arttıkça da modellerin tahmin performanslarının düştüğü görülmüştür.

Şekil 4.1’de her bir seçim yöntemine dayalı olarak elde edilen modellere ait simülasyon çalışmasından elde edilen MSE değerlerinin tüm olası durumlar için kutu grafikleri çizdirilmiştir. Şekilde x-eksininde belirtilen “A1, A2 ve A3” sırasıyla $n = 25, 50$ ve 100 örneklem büyüklükleri için AIC_c yöntemi ile elde edilen MSE grafiğini temsil etmektedir. Benzer şekilde “B1, B2 ve B3”, üç örneklem büyüklüğü için BIC dayalı HKO değerlerinin kutu grafiğini, “G1, G2 ve G3” GCV yöntemi, “C1, C2 ve C3” C_p yöntemi ve “H1, H2 ve H3” ise HKB yöntemine dayalı olarak elde edilen HKO için kutu grafiklerini temsil eder.



Şekil 4.1. Korelasyon düzeyi $\rho=0.60$ için Ridge parametresinin bilgi kriterleri ve beş farklı Ridge kestirici için ortalama MSE değerlerinin kutu grafikleri

Şekil 4.1’den görüldüğü gibi, $n=25$ hacimli ve açıklayıcı değişkenler orta derecede ilişkili Simülasyon örneklemelerinden elde edilen MSE için kutu grafikleri gösteriyor ki, küçük hacimli örneklerde kriterlere dayalı ridge kestiricileri benzer performanslar göstermektedir. Ancak, GCV ve HKB kriterleri MSE açısından diğerlerinden daha iyi sonuçlar vermektedir. Beklendiği gibi, örneklem hacmi arttıkça kriterlerin performans açısından benzer ve küçük örneklemelere göre daha stabil sonuçlar vermektedir.



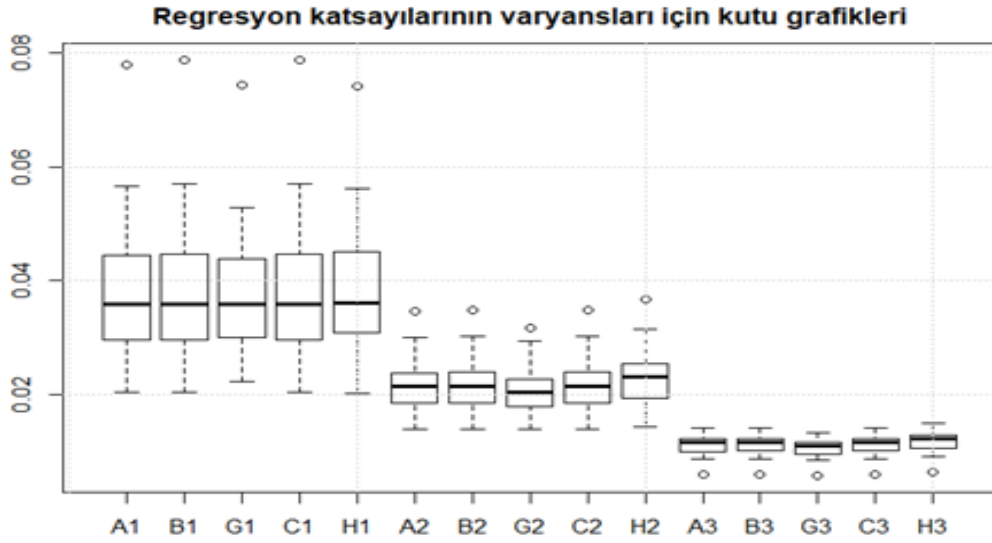
Şekil 4.2. Korelasyon düzeyi $\rho=0.80$ için Ridge kestiricilerine dayalı olarak elde edilen ortalama MSE değerlerinin kutu grafikleri

Şekil 4.2 incelendiğinde, küçük hacimli ve açıklayıcı değişkenler yüksek derecede ilişkili Simülasyon örneklemelerinden elde edilen MSE için kutu grafikleri gösteriyor ki, küçük hacimli örneklerde kriterlere dayalı ridge kestiricilerinin performansları birbirinden farklıdır. Ancak, bir önceki simülasyon durumunda olduğu gibi, *GCV* ve *HKB* Kriterlerinin diğerlerinden daha küçük değerleri MSE sonuçları vermektedir. Ayrıca küçük hacimli örneklerde kriterlerin MSE değerlerinin aralıklarının oldukça geniş ve kararlı olmadıkları görülmektedir. Bununla birlikte, örneklem hacmi arttıkça kriterlerin MSE değerlerinin aralıkları daralmakta ve küçük örneklemelere göre daha iyi sonuçlar vermektedir.

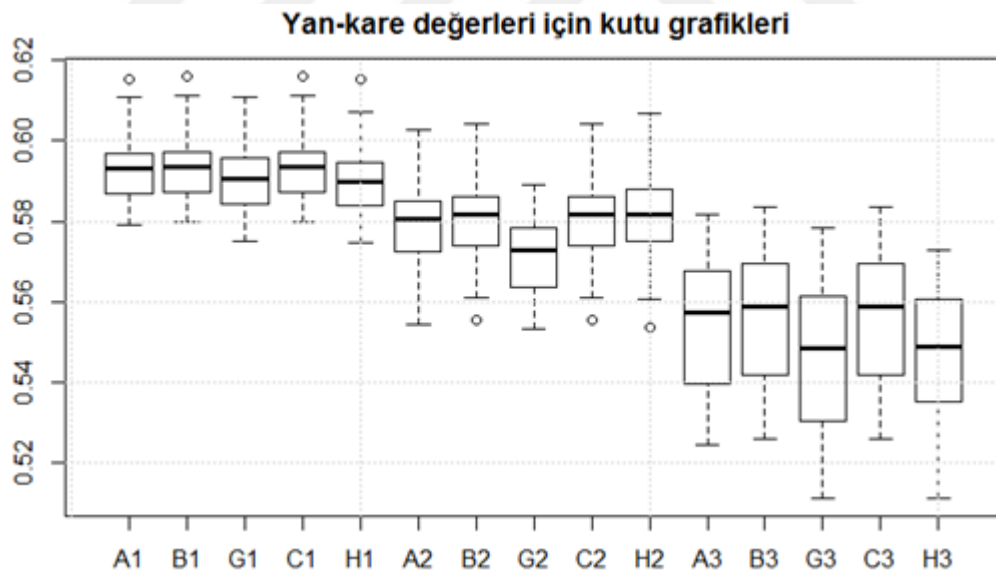
Eşitlik (3.15a)'da görüldüğü gibi, ridge kestiricisinin MSE değeri, kareli yan ve varyans toplamından oluşmaktadır. Bu bağlamda korelasyon düzeyi 0.80 olduğunda, yani, yüksek bir korelasyon düzeyinde kestiricilerin varyanslar ve yanlarını kareleri açısından davranışlarını görmek için Şekil 4.3 ve 4.4 oluşturulmuştur. Belirtmek gerekirse, ridge parametresi $k \rightarrow \infty$ iken, $\sigma^2 \rightarrow 0$. Diğer yandan, $k \rightarrow \infty$ iken, yanların karelerinin artıkları arttığı daha önceki teorik bilgilerden bilinmektedir.

Şekil 4.3 incelendiğinde, görünen o ki, ridge parametresinin yanı sıra, örneklem hacmi de varyanslar üzerinde önemli bir etki yapmaktadır. Örneklem hacmi azalırken varyanslar artarken, buna karşın, örneklem hacmi artarken varyanslar azalmaktadır. Yöntemler incelendiğinde, bütün yöntemler benzer davrandıkları görülmektedir.

Ancak GCV Kriterine dayalı ridge kestiricisinin diğerlerinden azda olsa daha iyi sonuç verdiği söylenebilir.



Şekil 4.3. Korelasyon düzeyi $\rho=0.80$ için Ridge parametresinin bilgi kriterleri ve beş farklı Ridge kestiricisinin varyanslarının kutu grafikleri

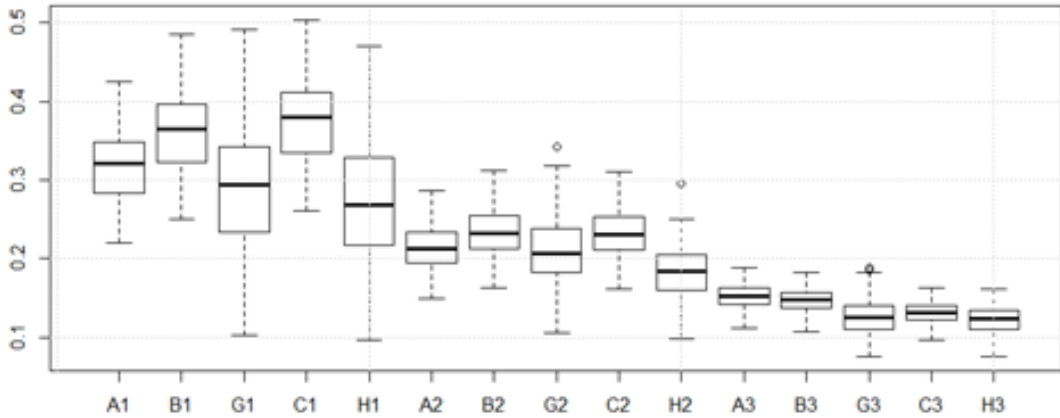


Şekil 4.4. Korelasyon düzeyi $\rho=0.80$ için Ridge parametresinin bilgi kriterleri ve beş farklı Ridge kestiricisinin yanlarının karelerini gösteren kutu grafikleri

Şekil 4.4 incelendiğinde, varyanslarda olduğu gibi, ridge parametresinin yanı sıra, örneklem hacmi yanlar üzerinde de önemli bir etki yapmaktadır. Örneklem hacmi azalırken yanların kareleri artıyor. Buna karşın, örneklem hacmi artarken yanların

kareleri azalıyor. Ancak küçük örneklemden elde edilen yan kareler daha dar bir aralıkta yer alırken, büyük örneklemden elde edilen yan kareler daha geniş bir aralıkta yer almaktadır. Yöntemler toplu olarak incelendiğinde, varyanslarda olduğu gibi, yöntemler, yan kareler açısından da benzer davranış sergilemektedirler. Ancak GCV Kriterine dayalı ridge kestiricisinin varyanslarda olduğu gibi yan kareler açısından da diğer yöntemlerden daha iyi sonuç verdiği söylenebilir.

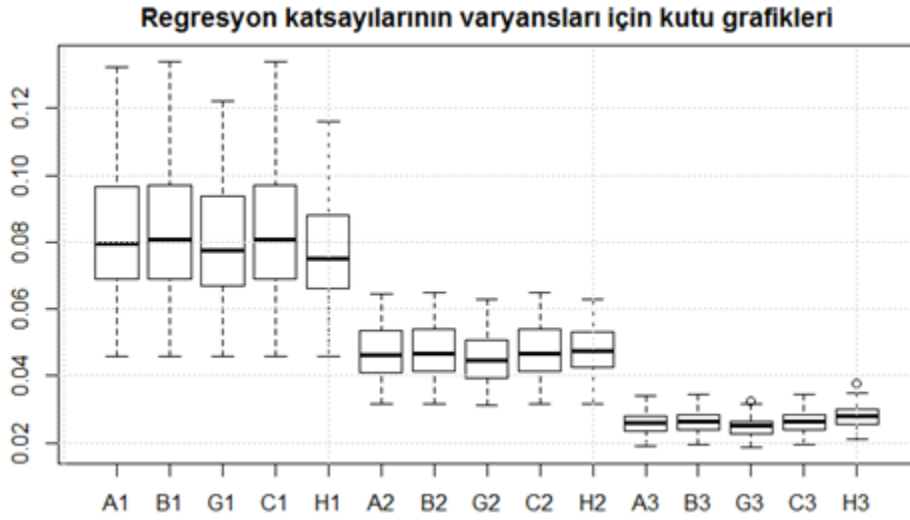
Açıklayıcı değişkenler oldukça yüksek derecede ilişkili Simülasyon örneklemi için yöntemlerden elde edilen MSE değerlerinin kutu grafikleri Şekil 3.5 gösterilmiştir. Bu sonuçlara göre, küçük hacimli örneklerde ridge kestiricilerinin performansları birbirinden farklı olduğu görülmektedir. Ancak, bir önceki korelasyon düzeyi $\rho=0.80$ için simülasyon durumunda olduğu gibi, *GCV* ve *HKB* Kriterlerinin diğerlerinden daha küçük MSE değerleri vermektedir. Ayrıca, küçük hacimli örneklerde yöntemlerden elde edilen MSE değerlerinin aralıklarının oldukça geniş ve durağan olmadıkları görülmektedir. Bununla birlikte, örneklem hacmi arttıkça MSE değerlerinin aralıkları daralmakta ve küçük örneklemlere göre daha iyi sonuçlar vermektedir.



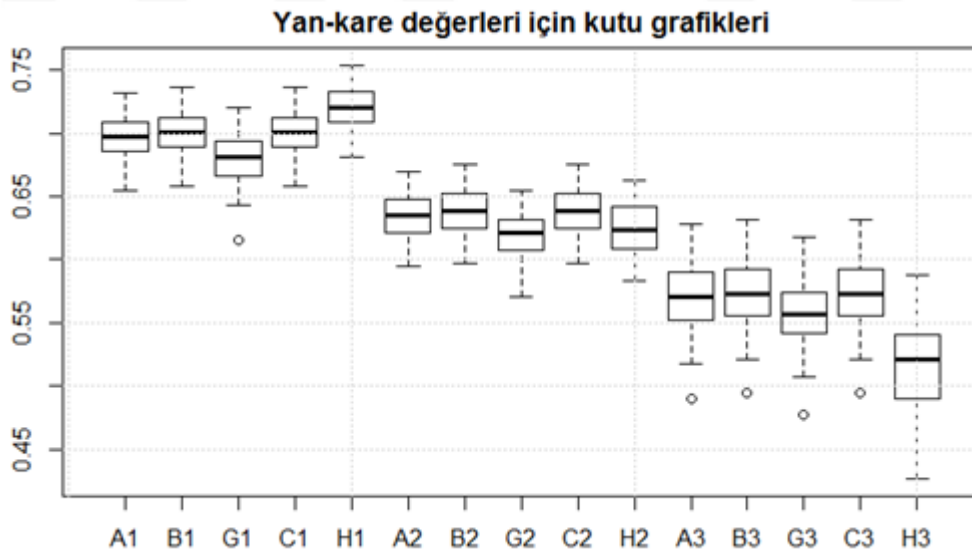
Şekil 4.5. Korelasyon düzeyi $\rho=0.99$ için ortalama MSE değerlerinin kutu grafikleri

Tüm sonuçlar incelendiğinde beklendiği gibi örneklem büyüklüğü arttıkça MSE değerleri küçülmekte, küçük korelasyon için Y-ekseninden anlaşılacağı gibi daha küçük MSE değerleri elde edilmiştir. AIC_c , BIC ve Cp yöntemlerinin tüm olası durumlar için benzer performansı gösterdiği, *GCV* ve *HKB*'nin ise bu üç yöntemden daha iyi performans gösterdiği gözlenmiştir. Büyük örneklem için ($n=100$) tüm

yöntemlerin birbirilerine çok yakın değerler verdiği görülmüştür. Fakat sonuç olarak, bu çalışma için ridge parametresinin seçiminde GCV ve HKB yöntemlerinin diğer üç seçim yönteminde daha iyi performansa sahip olduğu söylenebilir. Bu sonuçları, Tablo 4.4 ile verilen ortalama MSE değerleri desteklemektedir.



Şekil 4.6: Korelasyon düzeyi $\rho=0.99$ için Ridge parametresinin bilgi kriterleri beş farklı Ridge kestiricisinin varyanslarının kutu grafikleri



Şekil 4.7: Korelasyon düzeyi $\rho=0.99$ için Ridge parametresinin bilgi kriterleri ve beş farklı Ridge kestiricisinin yanlarının karelerini gösteren kutu grafikleri

Oldukça yüksek bir korelasyon içeren gözlemler altında yöntemlerin varyanslar ve yanlar açısından davranışlarını görmek için yukarıda gösterilen Şekil 4.6 ve Şekil 4.7

oluşturulmuştur. Bu şekiller incelendiğinde, daha önce belirtildiği gibi, ridge parametresinin yanı sıra, örneklem hacmi, hem varyanslar hem de yanlar üzerinde önemli bir etki yapmaktadır. Örneklem hacmi azalırken varyans ve yanlar artarken, buna karşın, örneklem hacmi artarken varyans ve yanlar azalmaktadır. Bütün yöntemler birlikte değerlendirildiğinde, benzer davrandıkları görülmektedir. Ancak yüksek korelasyon içeren özellikle büyük örnek veriler için HKB Kriterine dayalı ridge kestiricisinin diğerlerinden daha az yan içerdiği söylenebilir.

Tablo 4.4. Tüm durumlar için yöntemlerin ortalama MSE değerleri

ρ	n	$AICc$	BIC	GCV	Cp	HKB
0.60	25	0.4157	0.4252	0.3757	0.4204	0.3514
	50	0.3941	0.3969	0.3920	0.3955	0.3730
	100	0.3926	0.3934	0.3927	0.3930	0.3878
0.80	25	0.2504	0.2546	0.2209	0.2475	0.2096
	50	0.1803	0.1815	0.1527	0.1778	0.1350
	100	0.1277	0.1382	0.1019	0.1366	0.1035
0.99	25	0.3913	0.3936	0.2760	0.3869	0.2646
	50	0.2506	0.2518	0.2248	0.2483	0.1816
	100	0.2003	0.2009	0.1314	0.1991	0.1466

Yöntemlerin MSE ve eşitlik (3.15b)'de verilen SMDE değerleri açısından başarılarını görmek için Tablo 4.4 ve Tablo 4.5 elde edilmiştir. Söz konusu tablolar incelendiğinde, bir önceki şekillerde ortaya konan sonuçların sayısal olarak doğrulandıkları söylenebilir. Diğer bir deyişle, Bu tablolarda yer alan ortalama MSE ve SMDE değerleri gösteriyor ki, örneklem hacmi artıkça yöntemlerin performansları iyileşmektedir. Ancak kriterler kıyaslandıklarında minimum ortalama MSE veren en iyi ceza parametresini sırasıyla GCV ve klasik HKB yönteminin seçtiği söylenebilir.

Tablo 4.5. Korelasyon düzeyleri 0.80 ve 0.99 için yöntemlerin SMSE değerleri

<i>n</i>	ρ	<i>AIC_c</i>	<i>BIC</i>	<i>GCV</i>	<i>C_p</i>	<i>k_{HKB}</i>
25	0.80	0.6213	0.6256	0.4747	0.6171	0.6582
	0.99	0.7848	0.8119	0.5339	0.7339	1.0168
50	0.80	0.2175	0.2663	0.2983	0.3275	0.2437
	0.99	0.2489	0.2706	0.3737	0.3424	0.2720
100	0.80	0.1774	0.1918	0.1722	0.2212	0.2165
	0.99	0.2005	0.2060	0.2048	0.2324	0.2286

4.2 Gerçek Veri Çalışması

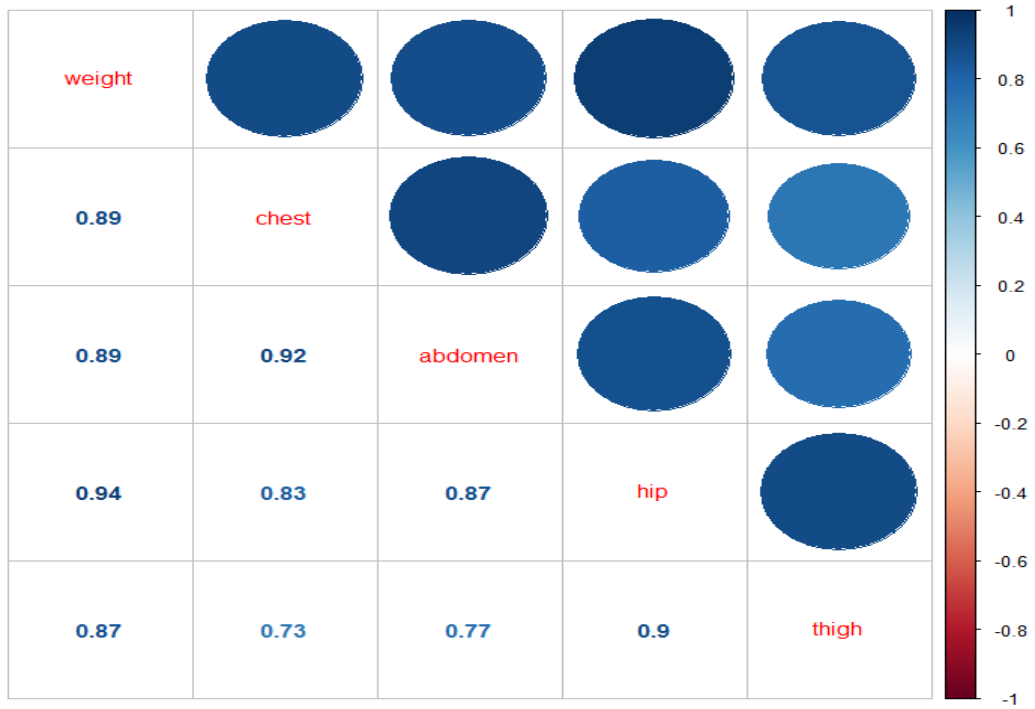
Bu çalışmada, 252 erkeğe ait vücut yağı yüzdeleri, su altı ağırlıkları ve bundan başka çeşitli vücut ölçümleri ile hesaplamalar yapılmıştır. Vücut yağı değeri, günümüzde insanların sağlıklı olup olmamalarına karar vermekte sıkça kullanılmaktadır. Bailey (1994) tarafından yapılan çalışmaya göre, vücut yağı kişilerin yaşı ve bazı deri üstü ölçümler kullanılarak hesaplanmıştır. Benzer şekilde yapılan bazı çalışmalar arasında Behnke ve Wilmore (1974) sayılabilir.

Bu çalışmada kullanılacak değişkenler şu şekildedir; kişinin ağırlığı (weight), göğüs çevresi uzunluğu (chest), karın çevresi (abdomen), kalça çevresi (hip) ve uyluk çevresi (thigh) olarak belirlenmiştir. Regresyon modelinde kullanılacak açıklayıcı değişkenler yukarıda sayılan beş değişkendir ve yanıt değişkeni vücut yağı yüzdesi (bodyfat) olarak belirlenmiştir. Buna göre regresyon modeli aşağıdaki gibi yazılır;

$$\begin{aligned} bodyfat_i &= \beta_0 + \beta_1 weight_i + \beta_2 chest_i + \beta_3 abdomen_i + \beta_4 hip_i + \beta_5 thigh_i \\ &+ \varepsilon_i, i = 1, \dots, 252 \end{aligned}$$

Burada, Şekil 3.4'te gösterildiği gibi açıklayıcı değişkenler arasında çok yüksek korelasyon değerleri olduğundan klasik en küçük kareler yöntemi ile bir çözüm elde edilemez. Ayrıca açıklayıcı değişkenler için **koşul indisi 108.6325** olarak elde edilmiş ve **VIF değerleri** Tablo 4.6'da sunulmuştur. Bahsi geçen tüm değerler, çoklu

doğrusal bağlantı sorununu doğrular niteliktedir. Dolayısıyla regresyon modeli ridge çözümü ile elde edilmiştir.



Şekil 4.8. Açıklayıcı değişkenler için korelasyon matrisi

Bağımsız değişkenlere ait VIF değerleri Tablo 4.6’da gösterilmiştir. VIF değerlerinin *değişkenler* için 5’ten büyük olduğu görülmektedir. Bu durum çoklu bağlantı göstergesi olarak kabul edilir.

Tablo 4.6. VIF değerleri

VIF değerleri				
weight	chest	abdomen	hip	thigh
14.5979	8.5063	8.4176	12.7440	5.4657

Korelasyon matrisi özdeğerlerinden çoklu bağlantının derecesinin belirlenmesinde yararlanır. Çoklu bağlantının olmadığı durumlarda özdeğerler 1’e eşittir. En az bir özdeğerin 1’den farklı olması veya en az birinin 0’a yakın olması çoklu bağlantının varlığını ortaya koyar. Ancak özdeğerlerin tek tek incelenmesi pek fazla anlam teşkil

etmemektedir. Bu nedenle, Vinod ve Ullah (1981)'in çalışmasında en büyük özdeğer ile en küçük özdeğer üzerine dayalı koşul indeksini (KI) önermiştir. Burada belirtmek gerekirse, ve $KI < 10$ ise çok az çoklu bağlantı ciddi bir sorun teşkil etmez; $10 \leq KI \leq 30$ ise çoklu bağlantı orta düzeyde ve son olarak; $KI > 10$ ise, şiddetli çoklu bağlantı olduğu düşünülmelidir (Pagel and Lunneborg, 1985; Gujarati, 1995; Rathert ve ark., 2011). Tablo 4.7'ye göre koşul indeksleri için şiddetli çoklu bağlantı olduğu söylenebilir.

Tablo 4.7. Koşul İndeksi ve Varyans Ayrıştırma Oranları

	intercept	weight	chest	abdomen	hip	thigh
1.000	0.000	0.000	0.000	0.000	0.000	0.000
20.671	0.031	0.0600	0.000	0.005	0.000	0.000
45.719	0.000	0.037	0.050	0.204	0.005	0.180
69.941	0.063	0.299	0.090	0.432	0.000	0.283
98.181	0.084	0.094	0.554	0.111	0.152	0.478
151.642	0.822	0.511	0.306	0.247	0.843	0.059

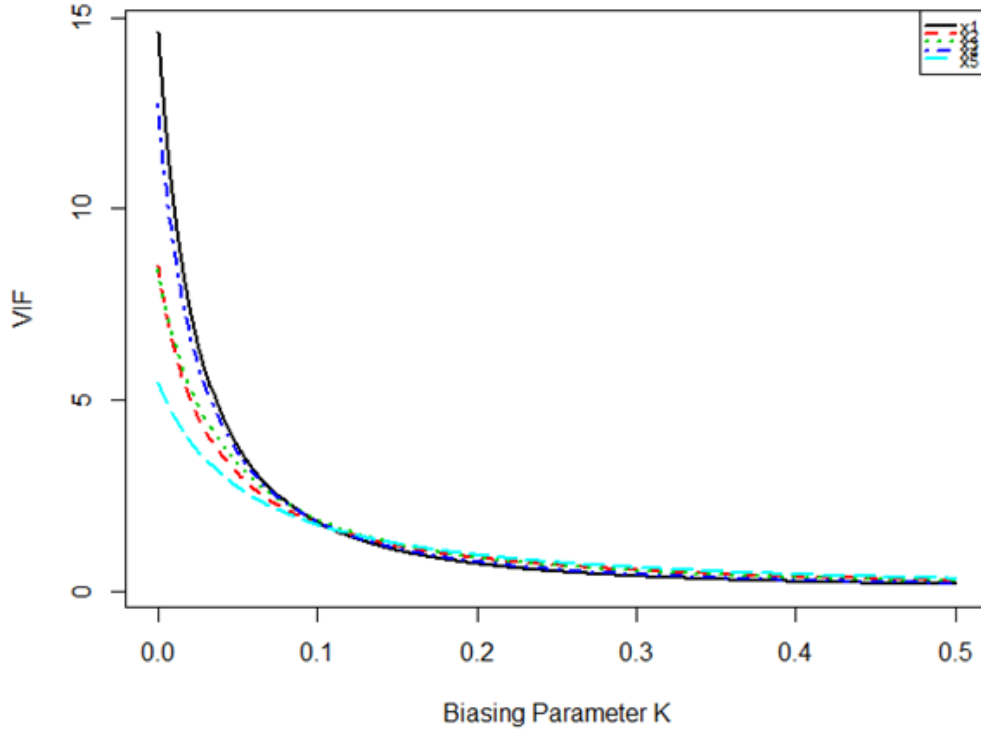
Tablo 4.8'de tüm yöntemler için MSE değerleri üzerinden yorum yapılacak olursa en uygun tahminleri, GCV, HKB ve EKK yöntemine dayalı modellerin verdiği kolaylıkla görülmektedir. Gerçek regresyon katsayılarına en yakın değerler belirtilen üç yöntem ile elde edilmişlerdir. Diğer üç yöntem, kötü sonuçlar vermeseler de GCV, HKB ve EKK'nin yanında düşük performans göstermişlerdir.

Tablo 4.8. Tüm yöntemler için MSE değerleri ve katsayılar

Yöntemler	AIC_c	BIC	GCV	C_p	HKB	EKK
β_1 (weight)	-0.5132	-0.5082	-0.5939	-0.5107	-0.5682	-0.5940
β_2 (chest)	0.0588	0.0608	0.0255	0.0598	0.0364	0.0256
β_3 (abdomen)	1.1761	1.1691	1.2900	1.1726	1.2537	1.2901

β_4 (hip)	-0.1091	-0.1075	-0.1360	-0.1083	-0.1273	-0.1360
β_5 (thigh)	0.1585	0.1566	0.1915	0.0157	0.1807	0.1915
MSE	0.2744	0.2747	0.2723*	0.2746	0.2726	0.2724

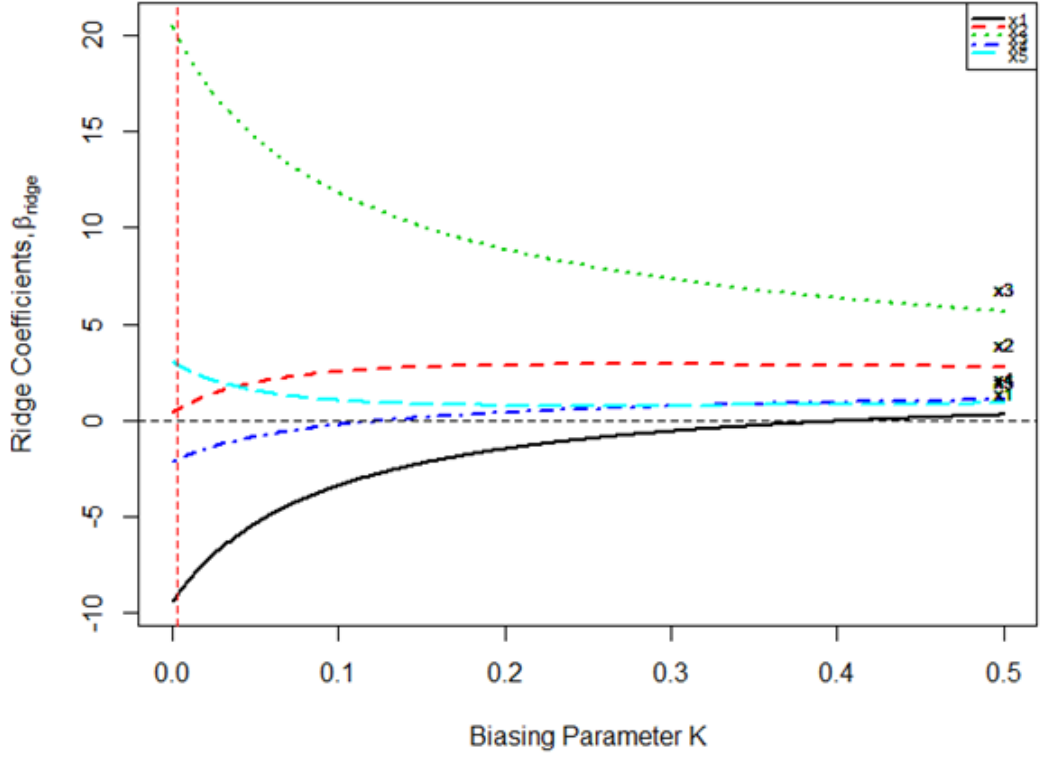
Şekil 4.9’da VIF grafiği gösterilmiştir. Şekilden de anlaşılacağı gibi HKB yöntemi ile belirlenen k ’nın VIF değerleri üzerindeki etkisi görülürken uygun bir k seçilme durumunda VIF değerlerinin 10’den küçük olması beklenmektedir. Söz konusu grafikte görüleceği gibi $k=0.01$ için VIF değerleri 10’den küçüktür.



Şekil 4.9. VIF izi grafiği

Şekil 4.10’da Ridge izi gösterilmiştir. Ridge izi dikey eksen β katsayılarının yatay eksen ise k ’nın artan değerlerinin bulunduğu grafikdir. Bu grafikte her bir k ’ya denk gelen β katsayılarının izi mevcuttur. Eksen çizgisinde yer alan $k=0$ değeri için EKK regresyon değerleri vardır. k değeri arttıkça β katsayıları ilk başta çok fazla değişkenlik gösterirken artmaya devam eden k bir süre sonra β katsayılarında

durađanlık gstermektedir. Őekil 4.10’da $k=0.01$ yanlılık sabitinden sonra regresyon katsayılarının daha durađan hale geldiđi tespit edilmiŐtir.



Őekil 4.10. Ridge izi grafiđi

5. SONUÇ VE ÖNERİLER

Bağımsız değişkenler arasında çoklu bağlantı olması durumunda model parametrelerini tahmin etmek için kullanılan kestirim yöntemlerinden RR yöntemi sıklıkla kullanılmaktadır. Bu yöntem EKK yöntemleriyle elde edilen sonuçların çoklu bağlantıdan dolayı geçerliliğini kaybettiğinde yani çoklu doğrusal bağlantı içeren verilerde regresyon katsayılarının hataları, işaretleri ve büyüklüklerini tahmin etmek mümkün değil iken kullanılabilir. Bu çalışmada RR yöntemi dikkate alınmıştır. Cebirsel olarak RR yöntemi, değişkenler arası ilişkinin etkilerini ortadan kaldırır ve EKK'nın tahmin ettiğinden daha kararlı tahminler elde etmek için katsayılar arasındaki korelasyonun görünür büyüklüğünü azaltır ve ayrıca tahmin doğruluğunu artırır (bkz. Hoerl ve Kennard, 1970a; Montgomery ve Peck., 1982; Myers, 1986; Seber ve Lee, 2003).

Bu tezde iki adet uygulama yapılmış olup bu uygulamalardan sonuçlar R paket programı yardımı ile elde edilmiştir. Birinci uygulama, ridge parametresinin seçilmesi için kullanılacak beş yöntemin (AIC_c , BIC , GCV , C_p ve HKB) karşılaştırılması bir Monte Carlo Simülasyon çalışması ile yapılmıştır. Bu çalışmada simülasyonda 3 farklı örneklem, 5 farklı seçim kriteri ve 3 farklı korelasyon düzeyi olmak üzere toplam 45 farklı simülasyon deneyi yapılmıştır. Deneyler sonucunda en uygun tahminleri, GCV ve HKB yöntemine dayalı modellerin verdiği tespit edilmiştir. Gerçek regresyon katsayılarına en yakın değerler belirtilen bu iki yöntem ile elde edilmişlerdir. Diğer dört yöntem (BIC , EKK , AIC_c , C_p), kötü sonuçlar vermeseler de GCV ve HKB yanında düşük performans göstermişlerdir. Tüm yöntemler için örneklem büyüklüğü arttıkça tahminlerin iyileştiği ve korelasyon seviyesi arttıkça da modellerin tahmin performanslarının düştüğü görülmüştür. Ancak bu kriterler kıyaslandıklarında minimum ortalama MSE veren en iyi ceza parametresini sırasıyla GCV ve klasik HKB yönteminin seçtiğini söyleyebiliriz. Simülasyon tekrarlarının artması sonucunda tahminlerin daha stabil oldukları görülmüştür.

İkinci uygulamada ise gerçek bir veri seti ile çalışılmış ve çalışmada, 252 erkeğe ait vücut yağı yüzdeleri, su altı ağırlıkları ve bundan başka çeşitli vücut ölçümleri ile hesaplamalar yapılmıştır. Açıklayıcı değişkenler arasındaki çok yüksek korelasyon değerlerinin olmasından dolayı klasik En Küçük Kareler yöntemi (OLS) ile bir çözüm elde edilememektedir. Ayrıca açıklayıcı değişkenler için koşul indisi, VIF değerleri çoklu doğrusal bağlantı sorununu doğrular niteliktedir. Dolayısıyla regresyon modeli ridge çözümü ile elde edilmiştir. Ridge Regresyon yöntemi EKK yöntemine göre daha kuramsal durağan ve beklentileri karşılayacak şekilde uygun sonuçlar vermektedir. Gerçek veri seti ile yapılan çalışmada tüm yöntemler için MSE değerleri üzerinden yorum yapıldığında en uygun tahminleri, GCV, HKB ve EKK yöntemine dayalı modellerin verdiği görülmüştür. Gerçek regresyon katsayılarına en yakın değerler belirtilen üç yöntem ile elde edilmişlerdir. Diğer üç yöntem (AIC_c , BIC , Cp), kötü sonuçlar vermeseler de GCV, HKB ve EKK'nin yanında düşük performans göstermişlerdir.

Her iki çalışma sonucunda, Kriterlerin kıyaslanması ile minimum ortalama MSE veren en iyi ceza parametresinin iki çalışmada da ortak olarak sırasıyla GCV ve HKB yönteminin olduğu söylenebilir.

KAYNAKLAR

- Albayrak, A. S., (2005) Çoklu Doğrusal Bağlantı Halinde En Küçük Kareler Tekniğinin Alternatifi Yanlı Tahmin Teknikleri ve Bir Uygulama. ZKÜ Sosyal Bilimler Dergisi, Cilt 1, Sayı 1.
- Alkhamisi, M. A., and Shukur, G. A., (2007) Monte Carlo Study of Recent Ridge Parameters. Communications in Statistics-Simulation and Computation, 36(3):535–547, URL <https://doi.org/10.1080/03610910701208619>. 330s.
- Alkhamisi, M. A., Khalaf, G.A. and Shukur, G., (2006) Some Modifications for Choosing Ridge Parameter. Communications in Statistics-Theory and Methods, 35(11):2005–2020, URL <https://doi.org/10.1080/03610920600762905>. 330s.
- Allen, D. (1974). The relationship between variable selection and data augmentation and a method for prediction, Technometrics, (16), 125-127.
- AlNasser, H. (2014) On Ridge Regression and Least Absolute Shrinkage and Selection Operator, B.Sc., University of Victoria, Master Of Science in the Department of Mathematics and Statistics.
- Arslan, O. And Billor, N. (1996) Robust ridge regression estimation based on the GM- estimators, Journal of Math., 9(1), 1-9.
- Askin, G.R., and Montgomery, D.C. (1980) Augmented Robust estimators, Technometrics, 22, 333-341.
- Aydın, D.,(2014) *Uygulamalı Regresyon Analizi*, Nobel, Ankarra , 558s.
- Bailey, Covert (1994). Smart Exercise: Burning Fat, Getting Fit, Houghton-Mifflin Co., Boston, pp. 179-186.
- Behnke, A.R. and Wilmore, J.H. (1974). Evaluation and Regulation of Body Build and Composition, Prentice-Hall, Englewood Cliffs, N.J.
- Carvalho, R., Pimentel, E. C. G., Cardoso, V., Queiroz S. A., Fries, L. A., (2006) Genetic effects on preweaning weight gain of Nelore-Hereford calves

according to different models and estimation methods. *Journal of Animal Science*, 84:2925-2933.

Chatterjee, S. and Price, B. (1977) *Regression Analysis By Examples*, Wiley, New York.

Chatterjee, S., and Hadi, A. S., (2006) *Regression Analysis by Example*. John Wiley & Sons, 4th edition,. 328, 330.

Cule, E., and De Iorio, M. (2012) A semi-automatic method to guide the choice of ridge parameter in ridge regression. *arXiv, stat.AP*, May.

Farebrother, R. W. (1976) Further results on the mean square error of ridge regression, *Journal of the Royal Statistical Society, Series B (Methodological)*, 38, 248-250.

Golub, G. H., Heath, M., and Wahba, G. (1979) Generalized Cross Validation as a Method for Choosing a Good Ridge Parameter. *Technometrics*, 21(2):215–223, URL <http://doi.org/10.2307/1268518>. 328, 332.

Gujarati, D. N., (1995) *Basic Econometrics*, 3rd Ed., McGraw-Hill, New York.1002s.

Hastie T. , Tibshirani R., (1990) *Generalized Additive Models*. Chapman & Hall. 334-337.

Hastie, T., Tibshirani, R., and Friedman, J. (2001) *The Elements of Statistical Learning; Data Mining, Inference and Prediction*. Springer Verlag, New York.

Hoerl, A. E., Kennard, R. W., and Baldwin, K. F., (1975) Ridge Regression: Some Simulations. *Communications in Statistics*, 4(2):105–123, URL <https://doi.org/10.1080/03610927508827232>. 328s.

Hoerl, A.E. (1962) *Application of Ridge Analysis to Regression Problems*, *Chemical Engineering Progress*, 58s, 54-59.

Hoerl, A.E. ve Kennard, R.W. (1970a) Ridge Regression: Biased Estimation for Nonorthogonal Problems, *Technometrics*, 12(1), 55-66.

- Hoerl, A.E., Kennard, R.W. (1970b) Ridge Regression: Applications Nonorthogonal problems. *Technometrics*, 12, 69-82.
- Hurvich, C.M., Tsai, C.L. (1989) Regression and time series model selection in small samples. *Biometrika*, 76, 297-307.
- James G., Witten D., Hastie T., and Tibshirani R. (2013) An introduction to statistical learning, volume 112. Springer.
- Khalaf G. A., and Shukur G., (2005) Choosing Ridge Parameter for Regression Problems. *Communications in Statistics-Theory and Methods*, 34(5):1177–1182, URL <https://doi.org/10.1081/STA-200056836>. 330s.
- Khalaf G. A., (2013) Comparison Between Biased and Unbiased Estimators. *Journal of Modern Applied Statistical Methods*, 12(2):293–303, URL <http://10.22237/jmasm/1383279360>. 330s.
- Kibria, B. M. G., (2003) Performance of Some New Ridge Regression Estimators. *Communications in Statistics-Simulation and Computation*, 32(2):419–435, URL <https://doi.org/10.1081/SAC-120017499>.
- Kunter, M., Holzberger, D. (2014) Loving teaching: Research on teachers' intrinsic orientations. In P. W. Richardson, S. Karabenick, & H. M. G. Watt (Eds.), *Teacher Motivation: Theory and Practice* (83-99). New York: Routledge.
- Mallows, C. L., (1973) Some Comments on p . *C Technometrics*, 15, 661-675.
Marcel Dekker, New York, 361s.
- Marquardt, D.W ve Snee, R.D., (1975) Ridge Regression in Praticce. *The American Statistician*, February 1975, Vol. 29, p3-20.
- Montgomery, D. C., Peck, E. A., and Vining, G.G. (2001) Introduction to linear regression analysis (3rd Ed.). New York, John Wiley and Sons Interscience Publication.
- Montgomery, D.C. ve Peck, E.A. (1991) *Iritroduction to Linear Regression Analysis*, John Wiley and Sons, New York, 526 s.

- Montgomery, D.C., and Peck, E.A. (1982) *Introduction to Linear Regression Analysis*, John Wiley and Sons, New York , 297-302.
- Myers, R.H. (1986) *Classical and Modern Regression With Applications*, Duxbury Press, Boston, 79s.
- Neter, J., Wasserman, W., Kunter, M., (1990) *Applied Linear Statistical Models*, 3rd Ed., New Jersey. 842s.
- Newhouse J. P., and Oman S. D.,(1971) An Evaluation of Ridge Estimators. Rand Report, R-716-PR, 330s.
- Pagel, M. U., Lunneborg, C. E., (1985) Empirical evaluation of ridge regression. *Psychological Bulletin*, 97: 342-355.
- Pfaffenberger, R.C. ve Dielman, T.E. (1990) A comparison of regression estimators when both multicollinearity and outliers are present. In *Robust Regression* (ed. Lawrence and Arthur), 243-270.
- Pimentel, E. C. G., Queiroz, S. A., Carneiro, R., Fries, L. A., (2007) Use of ridge regression for the prediction of early growth performance in crossbred calves. *Genetics and Molecular Biology*, 30(3): 536-544.
- Rathert, Ç. T., Üçkardeş, F., Nariç, D., Aksoy, T. (2011) Comparison of principal component regression with the least square method in prediction of internal egg quality characteristics in Japanese quails. *Kafkas Üniversitesi Veterinerlik Fakültesi Dergisi*, 17 (5): 687-692.
- Saleh, A. K. M. E and Kibria, B. M. G. (1993) Performance of Some New Preliminary Test Ridge Regression Estimators and Their Properties. *Communications in Statistics-Theory and Methods*, 22(10):2747–2764,. URL <https://doi.org/10.1080/03610929308831183> 330s.
- Schwarz, G. E. , (1978) Estimating the Dimension of a Model. *The Annals of Statistics*, 6(2):461–464. URL <https://doi.org/10.1214/aos/1176344136>. 332s.
- Seber G. A. F. and Lee A. J. (2003) *Linear Regression Analysis*, John Wiley & Sons, New Jersey, 2 edition, 326, 329s.

Silvapulle, M.J. (1991) Robust ridge regression based on an M estimator, Austral. J. Statist, 33, 319-333.

Singh, S. , and Tracy D. S. , (1999) Ridge-Regression Using Scrambled Responses. Metrika, LVII(1-2):147–157. 330s.

Theobald, C. M. (1974) Generalizations of mean square error applied to ridge regression, Journal of the Royal Statistical Society, Series B (Methodological), 36, 103-106.

Uslu, V. R., (1991) Ridge Regresyon ve Öğrenci Başarısı Üzerine Bir Çalışma. Ondokuz Mayıs Üniversitesi Fen Bilimleri Enstitüsü (Basılmamış), Yüksek Lisan Tezi, Samsun, 63s.

Üçkardeşler, F., Efe, E., Narinç, D. ve Aksoy, T. (2012) Japon Bildiricilerinde Yumurta Ak İndeksinin Ridge Regresyon Yöntemiyle Tahmin Edilmesi. *Akademik Ziraat Dergisi* 1(1), 11-20.

Vinod, H. D., Ullah, A., (1981) *Recent Advances In Regression Methods*.

Wessel, N., Wieringen, V. (2018) tarafından verilen ders notları.

Wieringen, V. (2015), Lecture notes on ridge regression, arXiv:1509.09169

EK. A: Gerçek Veri Tablosu

Body Fat	Weight	Chest	Abdomen	Hip	Thigh
12,3	154,25	93,1	85,2	94,5	59
6,1	173,25	93,6	83	98,7	58,7
25,3	154	95,8	87,9	99,2	59,6
10,4	184,75	101,8	86,4	101,2	60,1
28,7	184,25	97,3	100	101,9	63,2
20,9	210,25	104,5	94,4	107,8	66
19,2	181	105,1	90,7	100,3	58,4
12,4	176	99,6	88,5	97,1	60
4,1	191	100,9	82,5	99,9	62,9
11,7	198,25	99,6	88,6	104,1	63,1
7,1	186,25	101,5	83,6	98,2	59,7
7,8	216	103,6	90,9	107,7	66,2
20,8	180,5	102	91,6	103,9	63,4
21,2	205,25	104,1	101,8	108,6	66
22,1	187,75	101,3	96,4	100,1	69
20,9	162,75	99,1	92,8	99,2	63,1
29	195,75	101,9	96,4	105,2	64,8
22,9	209,25	107,6	97,5	107	66,9
16	183,75	106,8	89,6	102,4	64,2
16,5	211,75	106,2	100,5	109	65,8
19,1	179	103,3	95,9	104,9	63,5
15,2	200,5	111,4	98,8	104,8	63,4
15,6	140,25	86	76,4	94,6	57,4
17,7	148,75	86,7	80	93,4	54,9
14	151,25	90,2	76,3	95,8	58,4
3,7	159,25	89,6	79,7	96,5	55
7,9	131,5	88,6	74,6	85,3	51,7
22,9	148	97,4	88,7	94,7	57,5
3,7	133,25	93,5	73,9	88,5	50,1
8,8	160,75	97,4	83,5	98,7	58,9
11,9	182	100,5	88,7	99,8	57,5
5,7	160,25	93,5	84,5	100,6	58,5
11,8	168	93	79,1	94,5	57,3
21,3	218,5	111,7	100,5	108,3	67,1
32,3	247,25	117	115,6	116,1	71,2
40,1	191,75	118,5	113,1	113,8	61,9
24,2	202,25	106,5	100,9	106,2	63,5
28,4	196,75	105,6	98,8	104,8	66
35,2	363,15	136,2	148,1	147,7	87,3
32,6	203	114,8	108,1	102,5	61,3
34,5	262,75	128,3	126,2	125,6	72,5

Ek. A (devam)					
32,9	205	106	104,3	115,5	70,6
31,6	217	113,3	111,2	114,1	67,7
32	212	106,6	104,3	106	65
7,7	125,25	85,1	76	88,2	50
13,9	164,25	96,6	81,5	97,2	58,4
10,8	133,5	88,2	73,7	88,5	53,3
5,6	148,5	89,8	79,5	92,7	52,7
13,6	135,75	92,3	83,4	90,4	52
4	127,5	83,4	70,4	87,2	50,6
10,2	158,25	90,2	86,7	98,3	52,6
6,6	139,25	89,2	77,9	91	51,4
8	137,25	89,7	82	89,1	49,3
6,3	152,75	93,3	79,6	91,6	52,6
3,9	136,25	87,6	77,6	88,6	51,9
22,6	198	107,6	100	99,6	57,2
20,4	181,5	100	99,8	102,5	62,1
28	201,25	111,5	104,2	105,8	61,8
31,5	202,5	115,4	105,3	97	59,1
24,6	179,75	104,8	98,3	99,6	60,6
26,1	216	112,3	104,8	103,1	61,6
29,8	178,75	102,9	94,7	100,8	60,9
30,7	193,25	107,6	102,4	99,4	61
25,8	178	105,3	99,7	99,7	60,8
32,3	205,5	105,3	105,5	108,3	65
30	183,5	103	100,3	104,2	64,8
21,5	151,5	90	83,9	93,9	55
13,8	154,75	95,4	86,6	91,8	54,3
6,3	155,25	89,3	78,4	96,1	56
12,9	156,75	94,4	84,6	94,3	51,2
24,3	167,5	97,6	91,5	98,5	56,6
8,8	146,75	88,5	82,8	95,5	58,9
8,5	160,75	93,6	82,9	96,3	52,9
13,5	125	87,7	76	88,6	50,9
11,8	143	93,4	83,3	93	55,5
18,5	148,25	91,6	81,8	94,8	54,5
8,8	162,5	91,6	78,8	94,3	56,7
22,2	177,75	102	95	98,3	55
21,5	161,25	96,4	95,4	99,3	53,5
18,8	171,25	102,7	98,6	100,2	56,5
31,4	163,75	97,7	95,8	97,1	54,8
26,8	150,25	97,1	89	96,9	54,8
18,4	190,25	103,1	97,8	99,6	58,9

Ek. A (devam)					
27	170,75	101,8	94,9	95	56
27	168	101,4	99,8	96,2	56,3
26,6	167	98,9	89,7	96,2	54,7
14,9	157,75	97,5	88,1	96,9	57,2
23,1	160	104,3	90,9	93,8	57,8
8,3	176,75	97,3	86	99,3	61
14,1	176	96,7	86,5	98,3	60,4
20,5	177	99,7	95,6	102,2	58,3
18,2	179,75	101,9	93,2	100,6	58,9
8,5	165,25	97,2	83,1	95,4	56,9
24,9	192,5	106,6	97,5	100,6	58,9
9	184,25	99,6	88,8	101,4	57,4
17,4	224,5	113,2	99,2	107,5	61,7
9,6	188,75	99,1	91,6	102,4	60,6
11,3	162,5	99,4	86,7	96,2	62,1
17,8	156,5	95,1	88,2	92,8	54,7
22,2	197	107,5	94	103,7	62,7
21,2	198,5	106,5	95	101,7	59
20,4	173,75	99,1	92	98,3	59,3
20,1	172,75	96,7	89,2	98,3	60
22,3	196,75	103,5	95,5	101,6	59,1
25,4	177	104	98,6	99,5	59,5
18	165,5	93,1	87,3	96,6	54,7
19,3	200,25	105,2	102,8	103,6	61,2
18,3	203,25	110	101,6	100,7	55,8
17,3	194	110,1	88,7	102,1	57,5
21,4	168,5	97,8	92,3	100,6	57,5
19,7	170,75	96,3	90,6	99,3	61,9
28	183,25	108	105	103	63,7
22,1	178,25	99,7	95	98,6	62,3
21,3	163	93,5	89,6	99,8	61,5
26,7	175,25	100,7	92,4	97,5	59,3
16,7	158	97	86,6	92,6	55,9
20,1	177,25	96	90	99,7	58,8
13,9	179	99,2	90	96,4	56,8
25,8	191	95,4	92,4	104,3	64,6
18,1	187,5	101,8	87,5	101	58,5
27,9	206,5	104,3	99,2	104,1	58,5
25,3	185,25	99,2	98,1	101,4	57,1
14,7	160,25	99,3	83,3	97,5	60,5
16	151,5	94	86,1	95,2	58,1
13,8	161	98,9	84,1	94	58,5

Ek. A (devam)					
17,5	167	101	89,9	100	60,7
27,2	177,5	98,7	92,1	98,5	60,7
17,4	152,25	95,9	78	93,2	53,5
20,8	192,25	103,9	93,5	99,5	61,7
14,9	165,25	96,2	87	97,8	57,4
18,1	171,75	97,8	90,1	95,8	57
22,7	171,25	94,6	90,3	99,1	60,3
23,6	197	103,6	99,8	103,2	61,2
26,1	157	100,4	89,4	92,3	56,1
24,4	168,25	98,4	87,2	98,4	56
27,1	186	104,6	101,1	102,1	58,9
21,8	166,75	92,9	86,1	95,6	58,8
29,4	187,75	97,8	98,6	100,6	63,6
22,3	168,25	98,3	88,5	98,3	58,1
20,4	212,75	104,7	106,6	107,7	66,5
24,9	176,75	98,6	93,1	101,6	59,1
18,3	173,25	99,5	93	99,3	60,4
23,3	167	102,7	91	98,9	57,1
9,4	159,75	92,1	77,1	93,9	56,1
10,3	188,15	96,6	85,3	102,5	59,1
14,2	156	92,7	81,9	95,3	56,4
19,2	208,5	102	99,1	110,1	71,2
29,6	206,5	110,9	100,5	106,2	68,4
5,3	143,75	92,3	76,5	92,1	51,9
25,2	223	114,1	106,8	113,9	67,6
9,4	152,25	92,9	77,6	93,5	56,9
19,6	241,75	108,3	102,9	114,4	72,9
10,1	146	88,5	72,8	91,1	53,6
16,5	156,75	94	88,2	95,2	56,8
21	200,25	101,1	100,1	105	62,1
17,3	171,5	92,1	83,5	98,3	57,3
31,2	205,75	105,6	105	106,4	68,6
10	182,5	98,5	90,8	102,5	60,8
12,5	136,5	88,7	76,6	89,8	50,1
22,5	177,25	101,1	92,4	99,3	59,4
9,4	151,25	94	81,2	91,5	52,5
14,6	196	103,8	95,6	105,1	61,4
13	184,25	98,9	92,1	103,5	64
15,1	140	89,2	83,4	89,6	52,4
27,3	218,75	111,4	106	108,8	63,8
19,2	217	107,5	95,1	104,5	64,8
21,8	166,25	99,1	90,4	95,6	55,5

Ek. A (devam)					
20,3	224,75	108,2	100,4	106,8	63,3
36,3	228,25	114,9	115,9	111,9	74,4
16,5	172,75	99,1	90,8	98,1	60,1
3	152,25	92,2	81,9	92,8	54,7
0,7	125,75	90,8	75	89,2	50
20,5	177,25	100,5	90,3	98,7	57,8
16,9	176,25	98,2	90,3	99,9	59,2
25,3	226,75	115,3	108,8	114,4	69,2
9,9	145,25	96,8	79,4	89,2	50,3
13,1	151	92,6	83,2	96,4	60
29,9	241,25	119,2	110,3	113,9	69,8
22,5	187,25	102,7	92,7	101,9	64,7
16,9	234,75	109,5	104,5	109,9	69,5
26,6	219,25	108,5	104,6	109,8	68,1
0	118,5	79,3	69,4	85	47,2
11,5	145,75	95,5	83,6	91,6	54,1
12,1	159,25	92,3	86,8	96,1	58
17,5	170,5	98,9	90,4	95,5	55,4
8,6	167,5	89,5	83,7	98,1	57,3
23,6	232,75	117,5	109,3	108,8	67,7
20,4	210,5	107,4	98,9	104,1	63,5
20,5	202,25	109,2	98	101,8	62,8
24,4	185	103,4	101,2	103,1	61,5
11,4	153	91,4	80,6	92,3	54,3
38,1	244,25	115,2	113,7	112,4	68,5
15,9	193,5	104,9	94,1	102,7	60,6
24,7	224,75	106,7	105,7	111,8	65,3
22,8	162,75	92,2	85,6	96,5	60,2
25,5	180	101,6	96,6	100,6	61,1
22	156,25	97,8	86	96,2	57,7
17,7	168	92	89,7	101	62,3
6,6	167,25	94	78	99	57,5
23,6	170,75	103,7	89,7	94,2	58,5
12,2	178,25	102,7	89,2	99,2	60,2
22,1	150	91,1	85,7	96,9	55,5
28,7	200,5	107,2	103,1	105,5	68,8
6	184	100,8	89,1	102,6	60,6
34,8	223	121,6	113,9	107,1	63,5
16,6	208,75	105,6	96,3	102	63,3
32,9	166	100,6	93,9	100,1	58,9
32,8	195	102,7	101,3	101,7	60,7
9,6	160,5	99,8	83,9	91,8	53

Ek. A (devam)					
10,8	159,75	92,9	84,4	94	56
7,1	140,5	91,2	79,4	89	51,1
27,2	216,25	115,6	104	109	63,7
19,5	168,25	98,3	89,7	99,1	56,3
18,7	194,75	103,7	97,6	104,2	60
19,5	172,75	98,7	87,6	96,1	57,1
47,5	219	119,8	122,1	112,8	62,5
13,6	149,25	92,8	81,1	96,3	53,8
7,5	154,5	93,3	81,5	94,4	54,7
24,5	199,25	106,8	100	105	63,9
15	154,5	93,9	88,7	94,5	53,7
12,4	153,25	99	91,8	96,2	57,7
26	230	119,9	110,4	105,5	64,2
11,5	161,75	94,2	87,6	95,6	59,7
5,2	142,25	92,7	82,8	91,9	54,4
10,9	179,75	106,9	95,3	98,2	57,4
12,5	126,5	88,8	78,2	87,5	50,8
14,8	169,5	101,7	91,1	97,1	56,6
25,2	198,5	105,3	96,7	106,6	64
14,9	174,5	104	89,4	98,4	58,4
17	167,75	98,6	93	97	55,4
10,6	147,75	99,6	86,4	90,1	53
16,1	182,25	103,4	96,7	100,7	59,3
15,4	175,5	100,2	88,1	97,8	57,1
26,7	161,75	94,9	94,9	100,2	56,8
25,8	157,75	97,2	93,3	94	54,3
18,6	168,75	104,7	95,6	93,7	54,4
24,8	191,5	104	98,2	101,1	59,3
27,3	219,15	117,6	113,8	111,8	63,4
12,4	155,25	95,8	82,8	94,5	61,2
29,9	189,75	106,4	100,5	100,5	59,2
17	127,5	93	79,7	87,6	50,7
30,4	234,25	119,7	109	109,1	63,7
32,6	227,75	115,8	113,4	109,8	65,6
29	199,5	118,3	106,1	101,6	58,2
15,2	155,5	97,4	84,3	94,4	54,3
30,2	215,5	113,7	107,6	110	63,3
11	134,25	89,2	83,6	88,8	49,6
33,6	201	108,5	105	104,5	59,6
29,3	186,75	111,1	111,5	101,7	60,3
26	190,75	108,3	101,3	97,8	56
31,9	207,5	112,4	108,5	107,1	59,3

Kaynak: <http://lib.stat.cmu.edu/datasets/bodyfat>

Ek. B: Simülasyon Programı

```
library(psych)
library(pracma)

#-----Improved AIC (AICc)-----
aiccfunc<-function(H) {
  a<-dim(H)
  n<-a[1]
  p<-tr(H)
  score<-1+log10(norm(diag(n)-H)^2/n)+(2*(p+1)/(n-p-2))
  return(score)
}
#-----BIC-----
bicfunc<-function(H) {
  a<-dim(H)
  n<-a[1]
  p<-tr(H)
  score<-(1/n)*(norm(diag(n)-H)^2)+(log(n)/n)*p
  return(score)
}
#-----GCV-----
gcvfunc<-function(H) {
  a<-dim(H)
  n<-a[1]
  p<-tr(H)
  score<-(1/n)*(norm(diag(n)-H)^2)/((1/n)*p)
  return(score)
}
#-----CP-----
cpfunc=function(x,y) {
  library(pracma)
  gcvs<-0
  y<-matrix(c(y))
  n<-length(y)
  k2<-linspace(0.001,0.1,10)
  for (i in 1:10){
    Hp<-x%%solve(t(x)%%x+k2[i]*diag(p+1),tol=1e-
30)%%t(x)
    gcvs[i]<-gcvfunc(Hp)
  }
  for (j in 1:10){
    if (gcvs[j]==min(gcvs)) {
      kp<-k2[j]
    }
  }
  Hp2<-x%%solve(t(x)%%x+kp*diag(p+1),tol=1e-30)%%t(x)
  sigp2<-var(y-Hp2%%y)
  score<-(1/n)*((norm(diag(n)-Hp2)^2)+2*sigp2*p-sigp2)
  return(score)
}
```

Ek. B^(devam)

```
}
nums <- 40
nums1 <- 39.5
nums0 <- 41
#-----HKB-----
hkbfunc=function(x, y) {
  p<-5
  bhatp<-solve(t(x)%*%x+diag(p+1)*1, tol=1e-30)%*%t(x)%*%y
  yhatp<-x%*%bhatp
  sigsq<-var(y-yhatp)
  score<-(p*sigsq)/(t(bhatp)%*%bhatp)
  return(score)
}
biasfunc=function(x, k, bhat) {
  p1<-dim(x)
  p <- p1[2]
  Gk <- solve(t(x)%*%x+k*diag(p))
  bias <- Gk%*%t(x)%*%y-k*Gk%*%bhat
  return(bias)
}
varfunc=function(x, k, sig) {
  p1<-dim(x)
  p <- p1[2]
  Gk <- solve(t(x)%*%x+k*diag(p))
  var <- sig^2*Gk%*%t(x)%*%x%*%Gk
  return(var)
}
#-----
----
sim<-50
n<-c(25, 50, 100)
cor<-c(0.80, 0.99)
p<-5

AICc<-0
BIC<-0
GCV<-0
Cp<-0
ols<-matrix(0, p+1, 3)
a1<-0.01
a2<-0.05
a3 <- 0.09
a4 <- 0.08
a5 <- 0.07
h1 <- 10
bhataicc<-matrix(0, p+1, sim)
bhatbic<-matrix(0, p+1, sim)
bhatgcv<-matrix(0, p+1, sim)
bhatcp<-matrix(0, p+1, sim)
```

Ek. B^(devam)

```
bhathkb<-matrix(0,p+1,sim)
nbhataicc<-matrix(0,p+1,length(n))
nbhatbic<-matrix(0,p+1,length(n))
nbhatgcv<-matrix(0,p+1,length(n))
nbhatcp<-matrix(0,p+1,length(n))
nbhathkb<-matrix(0,p+1,length(n))

biasqaic <- matrix(0,sim,length(n))
biasqbic <- matrix(0,sim,length(n))
biasqgcv <- matrix(0,sim,length(n))
biasqcp  <- matrix(0,sim,length(n))
biasqhkb <- matrix(0,sim,length(n))
varaic <- matrix(0,sim,length(n))
varbic <- matrix(0,sim,length(n))
vargcv <- matrix(0,sim,length(n))
varcp <- matrix(0,sim,length(n))
varhkb <- matrix(0,sim,length(n))
SMDEaic <- matrix(0,sim,length(n))
SMDEbic <- matrix(0,sim,length(n))
SMDEgcv <- matrix(0,sim,length(n))
SMDEcp <- matrix(0,sim,length(n))
SMDEhkb <- matrix(0,sim,length(n))

mseaicc<-matrix(0,sim,length(n))
msebic<-matrix(0,sim,length(n))
msegcv<-matrix(0,sim,length(n))
msecp<-matrix(0,sim,length(n))
msehkb<-matrix(0,sim,length(n))

for (i in 1:length(n)){
  for(s in 1:sim){
    z1<-runif(n[i])
    z2<-runif(n[i])
    x<-matrix(0,n[i],5)
    x[,1]<-cor[2]*z1+sqrt(1-cor[2]^2)*z2
    x[,2]<-cor[2]*x[,1]+sqrt(1-cor[2]^2)*z2
    x[,3]<-cor[2]*x[,2]+sqrt(1-cor[2]^2)*z2
    x[,4]<-cor[2]*x[,3]+sqrt(1-cor[2]^2)*z2
    x[,5]<-cor[2]*x[,4]+sqrt(1-cor[2]^2)*z2
    b0<-0.5
    beta<-c(-0.5,1,2,-1.5,4,5)
    ones<-matrix(1,n[i],1)
    x1<-matrix(c(ones,x),n[i],p+1)
    e<-rnorm(n[i])*0.2
    y<-x1%*%beta+e

    #SELECTION PART-----
    lam<-seq(0.20,0.30,length.out=50)
```

Ek. B^(devam)

```
for (j in 1:50){
  H<-x1%%solve(t(x1)%%x1+lam[j]*diag(p+1),tol=1e-
30)%%t(x1)
  GCV[j]<-gcvfunc(H)
  AICc[j]<-aiccfunc(H)
  BIC[j]<-bicfunc(H)
  Cp[j]<-cpfunc(x1,y)
}
for (j2 in 1:50){
  if(GCV[j2]==min(GCV)){
    kgcv<-lam[j2]+a3
  }
  if(AICc[j2]==min(AICc)){
    kaicc<-lam[j2]-a5
  }
  if(BIC[j2]==min(BIC)){
    kbic<-lam[j2]-a4
  }
  if(Cp[j2]==min(Cp)){
    kcp<-lam[j2]-a4
  }
}
khkb<-hkbfunc(x1,y)*h1

#ESTIMATION PART-----
bhataicc[,s]<-
solve(t(x1)%%x1+diag(p+1)*kaicc,tol=1e-30)%%t(x1)%%y
bhatbic[,s]<-solve(t(x1)%%x1+diag(p+1)*kbic,tol=1e-
30)%%t(x1)%%y
bhatgcv[,s]<-solve(t(x1)%%x1+diag(p+1)*kgcv,tol=1e-
30)%%t(x1)%%y
bhatcp[,s]<-solve(t(x1)%%x1+diag(p+1)*kcp,tol=1e-
30)%%t(x1)%%y
bhathkb[,s]<-
solve(t(x1)%%x1+diag(p+1)*khkb[1],tol=1e-30)%%t(x1)%%y

yhataicc<-x1%%bhataicc[,s]
yhatbic<-x1%%bhatbic[,s]
yhatgcv<-x1%%bhatgcv[,s]
yhatcp<-x1%%bhatcp[,s]
yhathkb<-x1%%bhathkb[,s]

#mseaicc[s,i]<-mean((y-yhataicc)^2)
#msebic[s,i]<-mean((y-yhatbic)^2)
#msegcv[s,i]<-mean((y-yhatgcv)^2)
#msecp[s,i]<-mean((y-yhatcp)^2)
#mhskb[s,i]<-mean((y-yhathkb)^2)
```

Ek. B^(devam)

```
sigaic <- std(yhataicc-y)
sigbic <- std(yhatbic-y)
siggcv <- std(yhatgcv-y)
sigcp <- std(yhatcp-y)
sighkb <- std(yhathkb-y)
# BETALARIN YAN VEKTÖRLERİ
biasvecaic <- ((biasfunc(x1, kaicc, bhataicc[,s]))^2)
biasvecbic <- ((biasfunc(x1, kbic, bhatbic[,s]))^2)
biasvecgcv <- ((biasfunc(x1, kgcv, bhatgcv[,s]))^2)
biasveccp <- ((biasfunc(x1, kcp, bhatcp[,s]))^2)
biasvechkb <- ((biasfunc(x1, khkb[1], bhathkb[,s]))^2)

#BETALARIN YAN KARE TOPLAMLARI (KUTU GRAFİKLERİ İÇİN)
biasqaic[s,i] <-
sum((biasfunc(x1, kaicc, bhataicc[,s]))^2)/nums
biasqbic[s,i] <-
sum((biasfunc(x1, kbic, bhatbic[,s]))^2)/nums
biasqgcv[s,i] <-
sum((biasfunc(x1, kgcv, bhatgcv[,s]))^2)/nums1
biasqcp[s,i] <-
sum((biasfunc(x1, kcp, bhatcp[,s]))^2)/nums
biasqhkb[s,i] <-
sum((biasfunc(x1, khkb[1], bhathkb[,s]))^2)/nums0

#BETALARIN VARYANS MATRİSLERİ

varmataic <- varfunc(x1, kaicc, sigaic)
varmatbic <- varfunc(x1, kbic, sigbic)
varmatgcv <- varfunc(x1, kgcv, siggcv)
varmatcp <- varfunc(x1, kcp, sigcp)
varmathkb <- varfunc(x1, khkb[1], sighkb)

#BETALARIN VARYANS TOPLAMLARI (KUTU GRAFİĞİ İÇİN)

varaic[s,i] <- tr(varmataic)
varbic[s,i] <- tr(varmatbic)
vargcv[s,i] <- tr(varmatgcv)
varcp[s,i] <- tr(varmatcp)
varhkb[s,i] <- tr(varmathkb)

#MDE DEĞERLERİ

MDEaic <- diag(biasvecaic)+varmataic
MDEbic <- diag(biasvecbic)+varmatbic
MDEgcv <- diag(biasvecgcv)+varmatgcv
MDEcp <- diag(biasveccp)+varmatcp
MDEhkb <- diag(biasvechkb)+varmathkb
```

Ek. B^(devam)

```
#SMDE DEĞERLERİ MDE'NİN İZLERİ

SMDEaic[s,i] <- tr(MDEaic)
SMDEbic[s,i] <- tr(MDEbic)
SMDEgcv[s,i] <- tr(MDEgcv)
SMDECp[s,i] <- tr(MDECp)
SMDEhkb[s,i] <- tr(MDEhkb)

}
nbhataicc[,i]<-rowMeans(bhataicc)
nbhatbic[,i]<-rowMeans(bhatbic)
nbhatgcv[,i]<-rowMeans(bhatgcv)
nbhatcp[,i]<-rowMeans(bhatcp)
nbhathkb[,i]<-rowMeans(bhathkb)
ols[,i]<-solve(t(x1)%*%x1,tol=1e-30)%*%t(x1)%*%y
}

SMDEAICC <-mean(SMDEaic)
SMDEBIC <-mean(SMDEbic)
SMDEGCV <-mean(SMDEgcv)
SMDECp <-mean(SMDECp)
SMDEHKB <-mean(SMDEhkb)

SMDE_TABLE<-
data.frame(SMDEAICC,SMDEBIC,SMDEGCV,SMDECp,SMDEHKB)
print(SMDE_TABLE)

#KUTU GRAFİKLERİ (BETALARA AİT YANKARE ve VARYANSLAR İÇİN
)
df<-
data.frame(biasqaic[,3],biasqbic[,3],biasqgcv[,3],biasqcp
[,3],biasqhkb[,3],biasqaic[,2],biasqbic[,2],biasqgcv[,2],
biasqcp[,2],biasqhkb[,2],biasqaic[,1],biasqbic[,1],biasqg
cv[,1],biasqcp[,1],biasqhkb[,1])
boxplot(df,main="Yan-kare değerleri için kutu
grafikleri",names=c("A1","B1","G1","C1","H1","A2","B2","G
2","C2","H2","A3","B3","G3","C3","H3"))
grid()
df2<-
data.frame(varaic[,1],varbic[,1],vargcv[,1],varcp[,1],var
hkb[,1],varaic[,2],varbic[,2],vargcv[,2],varcp[,2],varhkb
[,2],varaic[,3],varbic[,3],vargcv[,3],varcp[,3],varhkb[,3
])
boxplot(df2,main="Regresyon katsayılarının varyansları
için kutu
grafikleri",names=c("A1","B1","G1","C1","H1","A2","B2","G
2","C2","H2","A3","B3","G3","C3","H3"))
grid()
```

Ek. C: Gerçek Veri

```
#-----Improved AIC (AICc)-----
aiccfunc<-function(H) {
  a<-dim(H)
  n<-a[1]
  p<-tr(H)
  score<-1+log10(norm(diag(n)-H)^2/n)+(2*(p+1)/(n-p-2))
  return(score)
}
#-----BIC-----
bicfunc<-function(H) {
  a<-dim(H)
  n<-a[1]
  p<-tr(H)
  score<-(1/n)*(norm(diag(n)-H)^2)+(log(n)/n)*p
  return(score)
}
#-----GCV-----
gcvfunc<-function(H) {
  a<-dim(H)
  n<-a[1]
  p<-tr(H)
  score<-(1/n)*(norm(diag(n)-H)^2)/((1/n)*p)
  return(score)
}
#-----CP-----
cpfunc=function(x,y) {
  library(pracma)
  gcvs<-0
  y<-matrix(c(y))
  n<-length(y)
  k2<-linspace(0.001,0.1,10)
  for (i in 1:10){
    Hp<-x%%solve(t(x)%%x+k2[i]*diag(p),tol=1e-
30)%%t(x)
    gcvs[i]<-gcvfunc(Hp)
  }
  for (j in 1:10){
    if (gcvs[j]==min(gcvs)) {
      kp<-k2[j]
    }
  }
  Hp2<-x%%solve(t(x)%%x+kp*diag(p),tol=1e-30)%%t(x)
  sigp2<-var(y-Hp2%%y)
  score<-(1/n)*((norm(diag(n)-Hp2)^2)+2*sigp2*p-sigp2)
  return(score)
}
```

Ek. C^(devam)

```

#-----HKB-----
hkbfunc=function(x,y,p){
  bhatp<-solve(t(x)%*%x+diag(p)*1,tol=1e-30)%*%t(x)%*%y
  yhatp<-x%*%bhatp
  sigsq<-var(y-yhatp)
  score<-(p*sigsq)/(t(bhatp)%*%bhatp)
  return(score)
}

#-----
----
library(MASS)
library(pracma)
library(psych)
library(dplyr)
library(perturb)
library(genridge)
library(corrplot)

data1<-read.table("data.txt")
colnames(data1)<-
c("bodyfat","weight","chest","abdomen","hip","thigh")
sdata<-scale(data1)
datan<-
matrix(c(sdata[,1],sdata[,2],sdata[,3],sdata[,4],sdata[,5],
sdata[,6]),n,6)
data<-data.frame(datan)
colnames(data)<-
c("bodyfat","weight","chest","abdomen","hip","thigh")
y<-data$bodyfat
x<-data[2:6]
n<-length(y)
size<-dim(x)
p<-size[2]+1
ones<-matrix(1,n,1)
xdf<-x %>% mutate(ones)
x1<-
matrix(c(ones,x$weight,x$chest,x$abdomen,x$hip,x$thigh),n
,p)

```

Ek. C^(devam)

```

xo<-
matrix(c(x$weight,x$chest,x$abdomen,x$hip,x$thigh),n,p-1)
lrid<-lm(y~xo[,1]+xo[,2]+xo[,3]+xo[,4]+xo[,5])
CI<-colldiag(lrid)
VIF<-vif(lrid)
A<-cor(x)
corrplot.mixed(A)
AICc<-0
BIC<-0

```

```

GCV<-0
Cp<-0
ols<-0
a1<-0.1
a2<-0.05

#SELECTION PART-----
lam<-seq(0.00000000005,3,length.out=50)
for (j in 1:50){
  H<-x1%%solve(t(x1)%%x1+lam[j]*diag(p),tol=1e-
30)%%t(x1)
  GCV[j]<-gcvfunc(H)
  AICc[j]<-aiccfunc(H)
  BIC[j]<-bicfunc(H)
  Cp[j]<-cpfunc(x1,y)
}
for (j2 in 1:50){
  if(GCV[j2]==min(GCV)){
    kgcv<-lam[j2]
  }
  if(AICc[j2]==min(AICc)){
    kaicc<-lam[j2]-a1
  }
  if(BIC[j2]==min(BIC)){
    kbic<-lam[j2]+a1
  }
  if(Cp[j2]==min(Cp)){
    kcp<-lam[j2]
  }
}
}
khkb<-hkbfunc(x1,y,p)

#ESTIMATION PART-----
bhataicc<-solve(t(x1)%%x1+diag(p)*kaicc,tol=1e-
30)%%t(x1)%%y
bhatbic<-solve(t(x1)%%x1+diag(p)*kbic,tol=1e-
30)%%t(x1)%%y

Ek. C(devam)
bhatgcv<-solve(t(x1)%%x1+diag(p)*kgcv,tol=1e-
30)%%t(x1)%%y
bhatcp<-solve(t(x1)%%x1+diag(p)*kcp,tol=1e-
30)%%t(x1)%%y
bathkb<-solve(t(x1)%%x1+diag(p)*khkb[1],tol=1e-
30)%%t(x1)%%y

yhataicc<-x1%%bhataicc
yhatbic<-x1%%bhatbic

```

```

yhatgcv<-x1**%bhatgcv
yhatcp<-x1**%bhatcp
yhathkb<-x1**%bhathkb

mseaicc<-mean((y-yhataicc)^2)
msebic<-mean((y-yhatbic)^2)
msegcv<-mean((y-yhatgcv)^2)
msecp<-mean((y-yhatcp)^2)
msehkb<-mean((y-yhathkb)^2)
ols<-solve(t(x1)**%x1,tol=1e-30)**%t(x1)**%y

MSEAICC<-(mseaicc)
MSEBIC<-(msebic)
MSEGCV<-(msegcv)
MSECp<-(msecp)
MSEHKB<-(msehkb)

MSE_TABLE<-data.frame(MSEAICC,MSEBIC,MSEGCV,MSECp,MSEHKB)
print(MSE_TABLE)

```

ÖZGEÇMİŞ

Kişisel Bilgiler

Ad Soyad : Merve Türkmən Şahingöz
Uyruk : TC
Doğum Yeri ve Tarihi : İstanbul / 16.02.1991
Medeni Hali : Evli
Telefon : (507) 824 39 37
E-posta : mrv.trkmn@gmail.com

Eğitim

Alınan Derece	Aldığı Kurum/Üniversite	Mezuniyet Yılı
Lise	Küçükyalı Rezan Has Lisesi	2009
Lisans	İstanbul Aydın Üniversitesi	2013

İş Tecrübesi

Yıl	Yer	Pozisyon/görev
05/2018-Devam Ediyor	C/S Enformasyon Teknolojileri Ltd. Şti.	Yazılım Destek Uzmanı
02/2017-05/2018	C/S Enformasyon Teknolojileri Ltd. Şti.	Yazılım Test Edici
08/2015-07/2016	Turkaş Gıda Hizmet ve İşl. A.Ş. – Muğla KYK	Proje Müdür Yardımcısı
09/2013-08/ 2015	Turkaş Gıda Hizmet ve İşl. A.Ş.	Finans Sorumlusu

Yabancı Dil(ler)

Dil (İngilizce, vs)	Başlangıç	Orta	İleri
Yazma		x	
Konuşma		x	
Anlama		x	
Okuma		x	