

T.C.
ONDOKUZ MAYIS ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ
İSTATİSTİK ANA BİLİM DALI



**DARALTILMIŞ KOVARYANS TAHMİNCİLERİ İLE ÇOK
DEĞİŞKENLİ İNDİRGENMİŞ RANK REGRESYON YÖNTEMİ**

Yüksek Lisans Tezi

Salih ARPAGUŞ

Danışman

Doç. Dr. Emre DÜNDER

SAMSUN

2022

TEZ KABUL VE ONAYI

Salih ARPAGUŞ tarafından, **Doç. Dr. Emre DÜNDER** danışmanlığında hazırlanan “**Daraltılmış Kovaryans Tahmincileri ile Çok Değişkenli İndirgenmiş Rank Regresyon Yöntemi**” başlıklı bu çalışma, jürimiz tarafından 29.4.2022 tarihinde yapılan sınav sonucunda oy birliği ile başarılı bulunarak Yüksek Lisans Tezi olarak kabul edilmiştir.

| | Unvanı Adı Soyadı Üniversitesi Ana Bilim/Ana Sanat Dalı | İmza | Sonuç |
|---------------|---|-------------|--|
| Başkan | Prof. Dr. Mehmet Ali Cengiz Ondokuz Mayıs Üniversitesi İstatistik Anabilim Dalı | | <input type="checkbox"/> Kabul <input type="checkbox"/> Ret |
| Üye | Doç. Dr. Emre DÜNDER Ondokuz Mayıs Üniversitesi İstatistik Anabilim Dalı | | <input type="checkbox"/> Kabul <input type="checkbox"/> Ret |
| Üye | Doç. Dr. Tolga Zaman Çankırı Karatekin Üniversitesi İstatistik Anabilim Dalı | | <input type="checkbox"/> Kabul <input type="checkbox"/> Ret |

Bu tez, Enstitü Yönetim Kurulunca belirlenen ve yukarıda adları yazılı jüri üyeleri tarafından uygun görülmüştür.

ONAY
... / ... / ...
Prof. Dr. Ali BOLAT
Enstitü Müdürü

BİLİMSEL ETİĞE UYGUNLUK BEYANI

Hazırladığım Yüksek Lisans tezinin bütün aşamalarında bilimsel etiğe ve akademik kurallara riayet ettiğimi, çalışmada doğrudan veya dolaylı olarak kullandığım her alıntıya kaynak gösterdiğimi ve yararlandığım eserlerin Kaynaklarda gösterilenlerden oluştuğunu, her unsurun enstitü yazım kılavuzuna uygun yazıldığını ve TÜBİTAK Araştırma ve Yayın Etiği Kurulu Yönetmeliği'nin 3. Bölüm 9. maddesinde belirtilen durumlara aykırı davranılmadığını taahhüt ve beyan ederim.

Etik Kurul Gerekli mi?

Evet (Gerekli ise ekler kısmına ekleyiniz)

Hayır

30/12/2021

Salih ARPAGUŞ

TEZ ÇALIŞMASI ÖZGÜNLÜK RAPORU BEYANI

Tez Başlığı: Daraltılmış Kovaryans Tahmincileri ile Çok Değişkenli İndirgenmiş Rank Regresyon Yöntemi

Yukarıda başlığı belirtilen tez çalışması için şahsım tarafından 30.12.2021 tarihinde intihal tespit programından alınmış olan özgünlük raporu sonucunda;

Benzerlik oranı : % 6

Tek kaynak oranı : % 1 çıkmıştır.

İmza

30 / 12 / 2021

Doç. Dr. Emre DÜNDER

TEŐEKKÜR

2019 yılı itibariyle baŐlamıŐ olduĐum yŐksek lisans eĐitimimi tamamlamıŐ bulunmaktayım. Tarihe bir not olarak dŐŐmesi hasebiyle bu tez alıŐmamı, Covid-19 salgın hastalıĐının baŐ gŐsterdiĐi dŐnem ierisinde tamamlıyorum. Salgın hastalık sebebiyle sıkıntılı ve zorlu bir sŐre yaŐadıĐımız bu zaman zarfında, tez alıŐmamı evde geirerek tamamladım. Bu baĐlamda salgın hastalık sŐrecinde alıŐmamda bana yardımcı olan Sayın Do. Dr. Emre DŐNDER hocama ve bu sŐrete beni yalnız bırakmayan ve desteĐini esirgemeyen sevgili aileme teŐekkŐr ederim.

Salih ARPAGUŐ

ÖZET

DARALTILMIŞ KOVARYANS TAHMİNCİLERİ İLE ÇOK DEĞİŞKENLİ İNDİRGENMİŞ RANK REGRESYON YÖNTEMİ

Salih ARPAGUŞ
Ondokuz Mayıs Üniversitesi
Lisansüstü Eğitim Enstitüsü
İstatistik Ana Bilim Dalı
Yüksek Lisans, Ocak/2022
Danışman: Doç. Dr. Emre DÜNDER

Çok değişkenli indirgenmiş rank regresyon yöntemi için büyük ölçekli veri kümeleri ile ilgilenecek, veri kümelerinden elde edilecek kovaryans matrislerinin durumlarını inceledik. Burada, kovaryans matrislerinin özel durumlarını ele alarak, regresyon yöntemlerinde karşılaştığımız sorunları çözmek için daraltılmış kovaryans tahmin edici yöntemlerini kullandık. Daraltılmış kovaryans tahmin edici yöntemlerini, kovaryans matrisleri üzerinde ilk kez çok değişkenli indirgenmiş rank regresyon yönteminde kullandık. Daha önce bahsedildiği gibi, çok değişkenli indirgenmiş rank regresyon yönteminde veya diğer regresyon yöntemlerinde karşılaştığımız çoklu bağlantı sorunu için daraltılmış kovaryans tahmin edici yöntemlerinin çok değişkenli veri setleri üzerinde etkili olup olmadığını inceledik.

Amacımız; çoklu doğrusallık gibi problemlerin olduğu bir veri seti için kovaryans matrisleri üzerinde kötü koşullandırılmış kovaryans matrisi veya daraltılmış kovaryans tahmin edici yöntemleri kullanarak hem çoklu doğrusallık problemini en aza indirmek hem de kötü şartlandırılmış kovaryans matrisini daha iyi hale getirmektir. Ayrıca bu yöntemi uygulayarak çok değişkenli veri setleri için rank değerlerine baktık ve tahmin edici yöntemlerin hangi rank değerinde etkili olup olmadığını araştırdık. Diğer bir amaç ise, çok değişkenli indirgenmiş rank regresyon yöntemlerinde elde edilen veri setleri için kovaryans matrislerinin tersini dikkate alarak koşul indeksi değerlerini incelemek ve çoklu doğrusallık problemlerini en aza indiren tahmin edici yöntemleri belirlemektir. Yukarıda bahsedildiği gibi, çok değişkenli indirgenmiş rank regresyon yöntemleri için elde edilen veri setlerini incelerken, daraltılmış kovaryans tahmin edici yöntemlerini (MLE/EB, SRE, SDE, BCSE, CSE ve OAS) kullandık. Ayrıca veri kümeleri üzerinde hangi yöntemin daha etkili olduğunu araştırmak için bilgi kriterlerini kullandık. Kovaryans matrisinin tersinin hesaplanabildiği veri setleri için koşul indeks değerlerine bakarak çoklu doğrusallık problemini en aza indiren bir yöntemin daraltılmış kovaryans tahmin edici yöntemlerinden daha etkili olduğuna karar verdik.

Bu tezde iki simülasyon çalışması ve gerçek veri setlerinden elde edilen analizler, çok değişkenli indirgenmiş rank regresyon yöntemi için değerlendirilmiştir. Son olarak üzerinde çalıştığım bu tez için R programını kullandık.

Anahtar Kelimeler: Regresyon, Çok değişkenli indirgenmiş rank regresyon yöntemi, Bilgi kriterleri, Daraltılmış kovaryans tahmin edici yöntemler, R

ABSTRACT

MULTIVARIATE REDUCED RANK REGRESSION METHOD WITH SHRINKAGE COVARIANCE ESTIMATING METHODS

Salih ARPAGUŞ

Ondokuz Mayıs University
Institute of Graduate Studies
Department of Statistics

Master, January/2022

Supervisor: Assoc. Prof. Dr. Emre DÜNDER

For the multivariate reduced rank regression method, we examined the states of the covariance matrices to be obtained from the datasets by dealing with large-scale datasets. Here, we used shrinkage covariance estimation methods to solve the problems we encounter in regression methods, taking into account the special cases of covariance matrices. We used shrinkage covariance estimation methods for the first time on covariance matrices in multivariate reduced rank regression method. As mentioned before, we examined whether shrinkage covariance estimation methods are effective on multivariate datasets for the multicollinearity problem we encountered in multivariate reduced rank regression method or other regression methods.

Our aim; For a data set with problems such as multicollinearity, using ill-conditioned covariance matrix or shrinkage covariance estimation methods on covariance matrices is to both minimize the multicollinearity problem and make the ill-conditioned covariance matrix better. In addition, by applying this method, we looked at the rank values for multivariate data sets and investigated whether the estimator methods were effective at which rank value. Another aim is to examine the condition index values by considering the inverse of the covariance matrices for the data sets obtained in the multivariate reduced rank regression methods and to determine the estimator methods that minimize the multicollinearity problems. As mentioned above, we used shrinkage covariance estimation methods (MLE/EB, SRE, SDE, BCSE, CSE ve OAS) when examining datasets obtained for multivariate reduced rank regression methods. We also used information criteria to investigate which method was more effective on the datasets. By looking at the condition index values for data sets where the inverse of the covariance matrix can be calculated, we decided that a method that minimizes the multicollinearity problem is more effective than shrinkage covariance estimation methods.

In this thesis, two simulation studies and analyzes obtained from real data sets were evaluated for the multivariate reduced rank regression method. Finally, we used the R program for this thesis I was working on.

Keywords: Regression, Multivariate reduced rank regression method, Information criteria, Shrinkage covariance estimation methods, R

İÇİNDEKİLER

| | |
|---|-----|
| ÖZET | iii |
| ABSTRACT | iv |
| İÇİNDEKİLER | v |
| ŞEKİLLER DİZİNİ..... | vi |
| TABLolar DİZİNİ..... | vii |
| 1. GİRİŞ..... | 1 |
| 2. LİTERATÜR | 4 |
| 3. ÇOK DEĞİŞKENLİ İNDİRGENMİŞ RANK REGRESYON MODELİ..... | 8 |
| 3.1. Çok Değişkenli İndirgenmiş Rank Regresyon Modeli Nedir ve Ne Zaman Kullanılmaktadır? | 8 |
| 3.2. Çok Değişkenli İndirgenmiş Rank Regresyon Modelinin Parametreleri ve Parametrelerin Elde Ediliş Yöntemleri | 13 |
| 3.3. Çok Değişkenli İndirgenmiş Rank Regresyon Modelinde Kovaryans Matrisinin Önemi..... | 15 |
| 3.4. Çok Değişkenli İndirgenmiş Rank Regresyon Modelinde Kovaryans Matrisi için Model Seçimi | 21 |
| 4. UYGULAMA | 25 |
| 4.1. Benzetim Çalışması – 1 | 26 |
| 4.2. Benzetim Çalışması – 2 | 26 |
| 4.3. Gerçek Veri Seti Uygulamaları..... | 53 |
| 5. SONUÇLAR ve ÖNERİLER | 68 |
| KAYNAKÇA..... | 71 |
| ÖZGEÇMİŞ | 74 |

ŞEKİLLER DİZİNİ

| | | |
|------------|---|----|
| Şekil 4.1. | R = 3 ve c = (0.3 ve 0.5) değerleri için daraltılmış kovaryans tahmin edici yöntemlerinin ortalaması alınarak elde edilmiş olan determinant değerlerinin kutu grafikleri..... | 30 |
| Şekil 4.2. | R = 3, c = (0.3 ve 0.5) değerleri için eğitim-test modeli üzerinden daraltılmış kovaryans tahmin edici yöntemlerinin ortalaması alınarak elde edilmiş olan determinant değerlerinin kutu grafikleri | 36 |
| Şekil 4.3. | R = 4, c = 0.7 ve 0.9 değerleri için daraltılmış kovaryans tahmin edici yöntemlerinin ortalaması alınarak elde edilmiş olan determinant değerlerinin kutu grafikleri | 40 |
| Şekil 4.4. | R = 4, c = (0.7 ve 0.9) değerleri için eğitim – test modeli üzerinden daraltılmış kovaryans tahmin edici yöntemlerinin ortalaması alınarak elde edilmiş olan determinant değerlerinin kutu grafikleri | 44 |
| Şekil 4.5. | Gözlem sayılarının 100, 500 ve 1000 değerlerine sahip olduğu bir veri seti için r = (3 ve 4) değerlerinin kullanılmasıyla elde edilmiş olan daraltılmış kovaryans tahmin edici yöntemlerinin kutu grafikleri..... | 48 |
| Şekil 4.6. | Gözlem sayılarının 100, 500 ve 1000 değerlerine sahip olduğu bir veri seti için r = 3 ve 4 değerlerinin kullanılmasıyla elde edilmiş olan daraltılmış kovaryans tahmin edici yöntemlerinin kutu grafikleri..... | 53 |
| Şekil 4.7. | Epilepsi veri seti için rank değerlerinin 1 ve 2 olduğu durumlarda daraltılmış kovaryans tahmin edici yöntemlerinin etkisiyle oluşturulmuş olan koşul indeks değerlerinin kutu grafikleri..... | 61 |
| Şekil 4.8. | Laseri veri seti için rank değerlerinin 1, 3, 5 ve 7 olduğu durumlarda daraltılmış kovaryans tahmin edici yöntemlerinin etkisiyle oluşturulmuş olan koşul indeks değerlerinin kutu grafikleri..... | 67 |

TABLolar DİZİNİ

| | | |
|-------------|--|----|
| Tablo 4.1. | Simülasyon çalışmasında yer alacak olan daraltılmış kovaryans tahmin edici yöntemleri..... | 27 |
| Tablo 4.2. | Gözlem sayılarının 100, 500 ve 1000 değerlerini aldığı bir veri analizi için korelasyon parametrelerinin de 0.3, 0.5, 0.7 ve 0.9 gibi katsayılara sahip olduğu bir durumda $r = 3$ için daraltılmış kovaryans tahmin edici yöntemlerinin etkisiyle β katsayıları üzerinden elde edilmiş olan hata kareler ortalama değerleri | 28 |
| Tablo 4.3. | Gözlem sayılarının 100, 500 ve 1000 değerlerini aldığı bir veri seti için korelasyon parametrelerinin de 0.3, 0.5, 0.7 ve 0.9 gibi katsayılara sahip olduğu bir durumda daraltılmış kovaryans tahmin edici yöntemlerinin ortalama yönünden determinant değerleri | 29 |
| Tablo 4.4. | Daraltılmış kovaryans tahmin edici yöntemleri kullanılarak $n = 100$, $r = 3$ ve $c = 0.3$ değerleri için bilgi kriterlerinin ortalama yönünden karşılaştırılması..... | 31 |
| Tablo 4.5. | Daraltılmış kovaryans tahmin edici yöntemleri kullanılarak $n = 100$, $r = 3$ ve $c = 0.5$ değerleri için bilgi kriterlerinin ortalama yönünden karşılaştırılması..... | 31 |
| Tablo 4.6. | Daraltılmış kovaryans tahmin edici yöntemleri kullanılarak $n = 500$, $r = 3$ ve $c = 0.3$ değerleri için bilgi kriterlerinin ortalama yönünden karşılaştırılması..... | 32 |
| Tablo 4.7. | Daraltılmış kovaryans tahmin edici yöntemleri kullanılarak $n = 500$, $r = 3$ ve $c = 0.5$ değerleri için bilgi kriterlerinin ortalama yönünden karşılaştırılması..... | 33 |
| Tablo 4.8. | Gözlem sayılarının 100, 500 ve 1000 değerlerinin alındığı bir veri için korelasyon parametrelerinin de 0.3, 0.5, 0.7 ve 0.9 olduğu bir durumda eğitim – test modelinde $r = 3$ için β katsayıları üzerinden elde edilen hata kareler ortalama değerleri..... | 33 |
| Tablo 4.9. | Gözlem sayılarının 100, 500 ve 1000 değerlerini aldığı bir veri seti için korelasyon parametrelerinin de 0.3, 0.5, 0.7 ve 0.9 gibi katsayılara sahip olduğu bir durumda $r = 3$ için eğitim – test modeli üzerinde uygulanmış olan daraltılmış kovaryans tahmin edici yöntemlerinin ortalama yönünden determinant değerleri | 34 |
| Tablo 4.10. | Gözlem sayılarının 100, 500 ve 1000 değerlerini aldığı bir veri analizi için korelasyon parametrelerinin de 0.3, 0.5, 0.7 ve 0.9 gibi katsayılara sahip olduğu bir durumda $r = 4$ için daraltılmış kovaryans tahmin edici yöntemlerinin etkisiyle β katsayıları üzerinden elde edilmiş olan hata kareler ortalama değerleri | 37 |
| Tablo 4.11. | Gözlem sayısının 100, 500 ve 1000 değerlerini alarak oluşturulmuş bir veri için korelasyon parametrelerinin de 0.3, 0.5, 0.7 ve 0.9 değerlerini aldığı bir durumda daraltılmış kovaryans tahmin edici yöntemlerinin $r = 4$ için ortalama yönünden determinant değerleri | 38 |
| Tablo 4.12. | Daraltılmış kovaryans tahmin edici yöntemleri kullanılarak $n = 1000$, $r = 4$ ve $c = 0.5$ değerleri için bilgi kriterlerinin ortalama yönünden karşılaştırılması..... | 40 |
| Tablo 4.13. | Daraltılmış kovaryans tahmin edici yöntemleri kullanılarak $n = 1000$, $r = 4$ ve $c = 0.9$ değerleri için bilgi kriterlerinin ortalama yönünden karşılaştırılması..... | 41 |
| Tablo 4.14. | Gözlem sayılarının 100, 500 ve 1000 değerlerinin alındığı bir veri için korelasyon parametrelerinin de 0.3, 0.5, 0.7 ve 0.9 olduğu bir durumda eğitim – test modelinde $r = 4$ için β katsayıları üzerinden elde edilen hata kareler ortalama değerleri..... | 42 |

| | | |
|-------------|---|----|
| Tablo 4.15. | Gözlem sayılarının 100, 500 ve 1000 değerlerini aldığı bir veri seti için korelasyon parametrelerinin de 0.3, 0.5, 0.7 ve 0.9 gibi katsayılara sahip olduğu bir durumda $r = 4$ için eğitim – test modeli üzerinde uygulanmış olan daraltılmış kovaryans tahmin edici yöntemlerinin ortalama yönünden determinant değerleri | 43 |
| Tablo 4.16. | Daraltılmış kovaryans tahmin edici yöntemleriyle beraber kullanılarak elde edilmiş olan β değerlerinin, gözlem sayısının $n = (100, 500 \text{ ve } 1000)$ olduğu bir veride $r = (3 \text{ ve } 4)$ için meydana gelmiş olduğu hata kareler ortalama değerleri | 45 |
| Tablo 4.17. | Gözlem sayısının $n = (100, 500 \text{ ve } 1000)$ ve rank değerlerinin de 3 ve 4 olduğu bir veride daraltılmış kovaryans tahmin edici yöntemlerinin determinant değerlerinin ortalama yönünden karşılaştırılması..... | 46 |
| Tablo 4.18. | $n = 100$ ve $r = 3$ olarak alındığı bir durum için bilgi kriterlerinin daraltılmış kovaryans tahmin edici yöntemlerine etkisi..... | 48 |
| Tablo 4.19. | $n = 100$ ve $r = 4$ olarak alındığı bir durum için bilgi kriterlerinin daraltılmış kovaryans tahmin edici yöntemlerine etkisi..... | 49 |
| Tablo 4.20. | $n = 1000$ ve $r = 3$ olarak alındığı bir durum için bilgi kriterlerinin daraltılmış kovaryans tahmin edici yöntemlerine etkisi..... | 49 |
| Tablo 4.21. | $n = 1000$ ve $r = 4$ olarak alındığı bir durum için bilgi kriterlerinin daraltılmış kovaryans tahmin edici yöntemlerine etkisi..... | 50 |
| Tablo 4.22. | Eğitim – test modeli kullanılarak oluşturulmuş olan bir veri için daraltılmış kovaryans tahmin edici yöntemleriyle elde edilmiş olan β değerlerinin, gözlem sayısının $n = (100, 500 \text{ ve } 1000)$ olduğu bir veride $r = (3 \text{ ve } 4)$ için meydana gelmiş olduğu hata kareler ortalama değerleri | 51 |
| Tablo 4.23. | Gözlem sayısının $n = (100, 500 \text{ ve } 1000)$ ve rank değerlerinin de 3 ve 4 olduğu bir veride eğitim – test modeli kullanılarak elde edilmiş olan daraltılmış kovaryans tahmin edici yöntemlerinin determinant değerlerinin ortalama yönünden karşılaştırılması..... | 52 |
| Tablo 4.24. | Epilepsi veri setine ait değişkenlerin sözel ifadeleri..... | 54 |
| Tablo 4.25. | Laseri veri setine ait değişkenlerin sözel ifadeleri..... | 55 |
| Tablo 4.26. | Epilepsi veri seti üzerinden elde edilen, rank değerlerinin 1 ve 2 değerlerini aldığı, daraltılmış kovaryans tahmin edici yöntemlerinin etkisiyle oluşturulmuş bilgi kriterleri değerleri | 56 |
| Tablo 4.27. | Epilepsi veri seti için rank değerlerinin 1 ve 2 değerlerini aldığı durumlarda daraltılmış kovaryans tahmin edici yöntemleri kullanılarak elde edilen hata kareler ortalama değerleri..... | 57 |
| Tablo 4.28. | Epilepsi veri seti için rank değerlerinin 1 ve 2 değerlerini aldığı durumlarda, daraltılmış kovaryans tahmin edici yöntemleri kullanılarak elde edilmiş olan determinant değerleri | 59 |
| Tablo 4.29. | Epilepsi veri seti için rank değerlerinin 1 ve 2 olduğu durumlarda daraltılmış kovaryans tahmin edici yöntemlerinin etkisiyle oluşturulmuş olan koşul indeks değerleri..... | 60 |
| Tablo 4.30. | Laseri veri seti üzerinden elde edilmiş olan, rankın 1, 2, 3, 4, 5, 6 ve 7 değerlerini aldığı, daraltılmış kovaryans tahmin edici yöntemlerinin etkisiyle oluşturulmuş bilgi kriterleri değerleri | 62 |
| Tablo 4.31. | Laseri veri seti için rankın 1, 2, 3, 4, 5, 6 ve 7 değerlerini aldığı durumlarda daraltılmış kovaryans tahmin edici yöntemleri kullanılarak elde edilen hata kareler ortalama değerleri..... | 64 |
| Tablo 4.32. | Laseri veri seti için rank değerlerinin 1, 2, 3, 4, 5, 6 ve 7 olduğu durumlarda daraltılmış kovaryans tahmin edici yöntemlerinin etkisiyle oluşturulmuş olan koşul indeks değerleri | 66 |

1. GİRİŞ

Çok değişkenli istatistik, çok sayıda bağımlı ve bağımsız değişkenlerden oluşan veri setlerinin analizinde kullanılan bir yöntem olup, karmaşık veri setlerinin analiz edilmesini kolaylaştırmaktadır (Izenman, 2008). Bir başka ifadeyle çok değişkenli istatistik, karmaşık veri kümelerini analiz eden, çok sayıda bağımlı değişkenin veya bağımsız değişkenlerin olduğu durumlarda analizi sağlamaktadır (Tabachnick, vd., 2007).

Regresyon kelimesi batı dillerinde “gerileme, geriye gitme” anlamında kullanılmaktadır. TDK’deki Fransızca terim anlamı olarak, “başka bir olayın belirli bir büyüklüğüne karşılık bulan bir olayın, yaklaşık büyüklüğünü bulma amacı güden işlem” olarak tanımlanmaktadır. Yine TDK’nin regresyon terimi için ifade ettiği anlam Y değişkeninin X bağımsız değişkenine matematiksel bir ifadeyle bağımlı olması olarak tanımlanmaktadır (İng.). 1805 yılında regresyon yöntemi, en ilkel biçimde ve ilk kullanım şekli olan en küçük kareler yöntemi olarak Fransız Matematikçi Andrien Marie Legendre tarafından ortaya çıkarılmıştır. Daha sonra ise 1880’lerin ortalarında, Francis Galton tarafından regresyon ve korelasyon fikirleri, “Kalıtımın Önemi” çalışmasında regresyon yöntemini geliştirerek ebeveynlerinin ve çocuklarının boylarını karşılaştırmak için uygulamıştır. Bunun üzerine bilimsel bir model olarak regresyon modeli, 1885 yılında ortaya çıkmıştır (Izenman, 2008). 1897 yılında George Udny Yule, regresyonda bulunan bir Gauss hata eğrisi varsayım sonucunu, değişkenlerin doğrusal olarak ilişkili olduğu varsayımıyla değiştirebileceğini ve bunun üzerine en küçük kareler yönteminin uygulanabileceğini gösteren kişi olmuştur (Izenman, 2008:108).

Regresyon yöntemi veri analizleri için en yaygın kullanılan yöntemdir. Bu yöntem 1897’den sonra doğrusal ve doğrusal olmayan regresyon yöntemleri ile parametrik ve parametrik olmayan regresyon yöntemleri olarak ayrılmıştır. Bu bağlamda doğrusal regresyon yöntemleri, bağımlı ve bağımsız değişkenlerin sayısına bağlı olarak basit, çoklu ve çok değişkenli regresyon yöntemleri olarak ayrılmaktadır (Izenman, 2008).

Çok değişkenli regresyon yöntemi, tek değişkenli regresyon yönteminin aynı özellik grubuna veya öngörücülerine bağlı olarak birden fazla bağımlı değişkenleri tahmin eden durumun basit bir genellemesidir (Mukherjee, 2013). Çoklu doğrusal regresyonun doğal bir uzantısı olan çok değişkenli regresyon yöntemi, bağımlı ve

bağımsız değişkenler arasındaki doğrusal ilişkileri yorumlamaktadır (Izenman, 2008). Bu yöntem, bağımlı değişken olan Y'lerin, bağımsız değişken olan X'ler tarafından ne derecede ölçüldüğünün tespiti ile ilgilidir (Izenman, 2008). Tek değişkenli regresyon yönteminin, q boyutlu bağımlı değişkenine ve p boyutlu bağımsız değişkenine sahip olduğu bilinmektedir. Bundan dolayıdır ki, çok değişkenli regresyon yöntemi, tek değişkenli regresyon yönteminde yer alan bağımlı değişkenlerin birden büyük olduğu durumların genişletilmesi olarak ifade edilmektedir (Mukherjee, 2013). Çok değişkenli regresyon yöntemlerinde ve diğer regresyon yöntemlerinde de karşımıza çıkan birtakım sorunlar bulunmaktadır. Bu sorunlardan bazıları; çoklu bağlantı problemleri, hataların bağımsız olmaması, çoklu bağlantının olduğu yerde kovaryans matrisinin determinantının hesaplanamaması ve çok büyük boyutlu matrislerde ise matrisin tersinin alınmadığı gibi durumlar yer almaktadır. Böylelikle çok değişkenli regresyon yönteminde ve bu çalışmada yer alan çok değişkenli indirgenmiş rank regresyon yönteminde de karşılaşmakta olduğumuz kovaryans matrislerinde yer alan sorunların giderilmesi amaçlanmaktadır.

İstatistik literatüründe çok değişkenli indirgenmiş rank regresyon yöntemi, Anderson (1951) tarafından geliştirmiştir ve 1975 yılında ise Izenman tarafından adlandırılarak günümüzde kullanılan isimle bilinmektedir (Schmidli, 2013:52). Burada, İndirgenmiş rank regresyon yöntemi kullanılarak, çok büyük veriler için meydana gelen sorunların çözümünde kovaryans matrislerinden yararlanılması amaçlanmaktadır. Buradaki yöntem boyut indirgeme ile ilgili olup çok değişkenli regresyon yöntemini daha kolay bir şekilde yorumlamak ve az sayıda değişkenler ile çalışarak daha iyi bir sonuç elde edilmek istenmektedir. Bu konu ile ilgili 1951 yılında Anderson, 1975 yılında Izenman ve daha sonra ise Joreskog ve Goldberger gibi birçok bilim insanların çalışmaları bulunmaktadır.

Mevcut konuyla ilgili birinci bölümde, indirgenmiş rank regresyon yöntemi hakkında genel bir bilgi verilmiştir. Burada yöntemin ne olduğu ve nerelerde kullanıldığından bahsedilerek yöntem hakkında bilgi sahibi olunması hedeflenmektedir. İkinci bölümde ise; çok değişkenli indirgenmiş rank regresyon yönteminin, parametreleri hakkında bilgi verilmiştir ve parametrelerin elde edilmiş yöntemleri ifade edilmiştir. Üçüncü bölümde ise, tez konusunun amacından bahsederek, çalıştığımız yöntemler üzerinde uygulayacağımız çok değişkenli indirgenmiş rank regresyon yöntemi uygulamaları için, kovaryans matrisinin önemine

deđinilmektedir. Son blmde ise, ok deđiřkenli indirgenmiř rank regresyon yntemi uygulamaları iin daraltılmıř (*shrinkage*) kovaryans tahmin edici yntemlerinden yararlanılmıř olup, model seimi ile ilgili olarak da tahminlerde bulunulmuřtur. Son olarak, nerilere ve sonulara deđinilerek alıřmanın tamamlanması hedeflenmiřtir.

2. LİTERATÜR

Mohebbi (2019) tarafından yazılan makalede, büyük boyutlu veri setleri ile ilgilenilmiştir. Burada büyük boyutlu veri setleri için değişken seçimi üzerinde “Elastik Ağ Tekniği” kullanılmaktadır. Ayrıca bu çalışmada verinin boyutu azaltılarak az sayıda değişkenler ile tahmine dayalı bir model geliştirilmiştir. Bu çalışmada kovaryans matrislerinin sorunları üzerinde durulmuş olup çalıştığım tez konusu için fayda sağlayacak daraltılmış kovaryans tahmin edici yöntemleri yer almaktadır. Bu makalede matematiksel ifade olarak tanımlanan daraltılmış kovaryans tahmin edici yöntemleri çalıştığım tez konusu için faydalı olmuştur.

Pamukçu (2017) tarafından yazılan makalede, büyük boyutlu veri setleri için elde edilmiş olan değişken sayılarının gözlem sayılarından büyük olduğu durumlar incelenmektedir. Bunun için ise bilgi kriterlerinden yararlanılarak en uygun kovaryans yapılar seçilmiştir. Makalenin amacı, aşırı derecede küçük örneklem problemleri ile ilgilenmektir. Bu makalede problemin çözümü için “kötü koşullandırılmış (*ill-conditioned*)” kovaryans matrislerine iz (*ridge*) düzeltmesi uygulanmıştır. Zhang ve İbrahim (2005) tarafından yazılan makalede, çoklu bağlantı problemleri için iz (*ridge*) regresyon yönteminin uygulanabileceğinden bahsetmektedir. Burada, çoklu bağlantının olduğu bir veri seti üzerinde en küçük kareler yönteminin yerine iz (*ridge*) regresyon yönteminin, sorunu giderebileceği görüşü savunulmuştur. Bu makale çalışmasında, SPSS paket programı kullanılarak oluşturulmuş olan simülasyon çalışması yer almaktadır. Makalede çıkan sonuçlarda, iz (*ridge*) regresyon yöntemi ile en küçük kareler yöntemi tahmin edilmiştir. Ayrıca bu makalede elde edilen sonuçlar bakımından iz (*ridge*) regresyon yöntemi için, hata kareler ortalama değerlerinin klasik hata kareler ortalama değerlerinden daha küçük olduğuna karar verilmiştir ve bazı durumlarda ise klasik hata kareler ortalama değerlerinin de arttığı gözlemlenmiştir. Ancak burada büyük boyutlu verilerde meydana gelecek olan çoklu bağlantı probleminde iz (*ridge*) regresyon yönteminin en küçük kareler yönteminden daha iyi performans gösteremediği açıklanmaktadır (Zhang ve İbrahim, 2005).

Fisher ve Sun (2011) tarafından yazılan makalede, büyük boyutlu veri setleri üzerinden elde edilmiş olan kötü koşullandırılmış kovaryans matrislerine Stein tipi daraltılmış tahmin edici yöntemi uygulanmıştır. Makalede normallik varsayımı altında daraltma şiddeti için önerilen tahmin edicilerin kullanılabileceği vurgulanmaktadır. Burada simülasyon çalışmaları ile gerçek veri setleri üzerinden yapılan çalışmalarda

daraltılmış kovaryans tahmin edicilerin kovaryans matrisini iyileştirdiği gözlemlenmiştir (Fisher ve Sun, 2011).

Koç ve Bozdoğan (2015) tarafından yayınlanan makalede, MARS model seçimi ile ICOMP bilgi kriteri incelenmektedir. ICOMP bilgi kriteri ile bazı bilgi kriterleri karşılaştırılarak bu makalede önerilen ICOMP bilgi kriterinin performansı ölçülmüştür. Burada yer alan Monte Carlo simülasyon çalışması, çok değişkenli indirgenmiş rank regresyon yöntemine uygun bir örnek olduğu için üzerinde çalışmaya karar verdik (Koç ve Bozdoğan, 2015).

Heinen ve Rengifo (2008) tarafından yazılan makalede, Gauss dağılımlı olmayan veri setleri ile ilgilenilmiştir. Burada çok değişkenli indirgenmiş ranklı bir dağılım modeli gösterilmekte olup Poisson, Negatif Binom, Gamma ve Normal dağılım gibi kesikli ve sürekli dağılımları içeren dağılım modellerinin genelleştirilebileceği gösterilmiştir. Makalede yer alan uygulama çalışması 1999 yılına ait ABD’de yer alan menkul kıymetler borsasındaki hisse senetleridir ve çok değişkenli regresyon yöntemi için “Copula Teorisi”ni uygulamaya çalışmışlardır. Beş yüz kez tekrarlanan simülasyon çalışması sonucunda farklı dağılım modelleri üzerindeki parametre tahminleri ile hata kareler ortalama değerleri incelenmiştir. Sonuç olarak bu çalışmada dağılım modelleri üzerinden geliştirilen yeni teknikler karşılaştırılmıştır (Heinen ve Rengifo, 2008).

Perez-Melo ve Kibria (2020) tarafından yayınlanan makalede çoklu bağlantı problemini çözmek için iz (*ridge*) regresyon yönteminin uygulanabileceği önerilmektedir. Bu çalışmada farklı T – testleri ile analizler yapılarak daha kapsamlı bir şekilde karşılaştırma yapılmıştır. Bunun için ise hangi test istatistiklerini kullanacaklarına karar vermek için farklı şekilde oluşturulmuş olan iz (*ridge*) regresyon yöntemleri kullanılmıştır. Makale kötü koşullandırılmış kovaryans matrisleri için bazı çözüm yöntemleri önermektedir. Bu çözümlerden biri iz (*ridge*) parametresinin eklenmesi iken bir diğer yöntemde çalıştığım tezin konusunu oluşturacak olan daraltılmış kovaryans tahmin edici yöntemlerinin eklenmesiyle elde edilen çözüm yöntemidir (Perez-Melo ve Kibria, 2020).

Bura ve Cook (2003) tarafından yazılan makalede, çok değişkenli regresyon yönteminde rank tahmini için bir teori geliştirilmiştir. Geliştirdikleri teoriye göre rank tahmini için en küçük kareler yönteminden yararlanılmıştır. Bu makale çalışmasında kanonik korelasyon analizi ile indirgenmiş rank regresyon yöntemi arasındaki ilişki

incelenerek “Bartlett testi” ile indirgenmiş rank regresyon yöntemi tahmin edilmeye çalışılmaktadır. Ayrıca bu makalede Minneapolis okulundaki öğrencilerin performanslarını gösteren bir veri seti kullanılmaktadır (Bura ve Cook, 2003).

Izenman (2008) tarafından yazılan “Modern Çok Değişkenli İstatistiksel Teknikler (*Modern Multivariate Statistical Techniques*)” adlı eserinde, büyük boyutlu veri setleri için ayrıntılı bir şekilde istatistiksel araçların nasıl kullanılacağı açıklanmaktadır. Altmışın üzerinde veri seti olan bu eserde kovaryans matrislerinin özel durumlarını ele alan çalışmalarda bulunmaktadır. Regresyon yöntemi için karşımıza çıkacak problemlerden biri de çoklu bağlantı sorunu olup bu konuda da uygulamalar yer almaktadır. Bu eser çalıştığım tez konuma yardımcı olması bakımından yararlandığım kaynaklardan biri olmuştur (Izenman, 2008).

Braak ve Looman (1994) tarafından yazılan makalede, indirgenmiş rank regresyon yöntemi için ikili çizimlerin (*biplot*) nasıl görselleştirilebileceği amaçlanmaktadır. Makalede halk sağlığı verileri kullanılmış olup elde edilen sonuçlar bağlamında ikili çizimlerin boyut indirgeme ile gerçek bir kazanım sağladığı ve indirgenmiş rank regresyon yöntemi için alternatif bir yol olarak kullanılabilirliği vurgulanmıştır (Braak ve Looman, 1994).

Cook ve arkadaşları (2015) tarafından yazılan makalede indirgenmiş rank regresyon yönteminin geliştirilmesiyle indirgenmiş zarf modeli yöntemi bulunmuştur. Burada toplam parametre sayısını, az sayıda değişkenler ile elde ettikleri tahminler sonucunda, verinin etkinliğini sağlamak ve doğru tahmin etmeyi amaçlamışlardır. Asıl amaçları ise, β tahmini için önemsiz olan ve tahminlerde gereksiz değişkenlere sebep olacak bağımlı değişkenleri ve katsayı tahminlerini belirlemektir (Cook, vd., 2015).

Fan ve arkadaşları (2008) tarafından yazılan makalede, örneklem boyutunun arttığı durumlarda kovaryans matrisleri üzerinden elde edilen tahminlerin etkileri araştırılmıştır (Fan, vd., 2008).

Mukherjee (2013) tarafından yazılan makalede, bağımlı değişkenler arasında boyut indirgeme yöntemi kullanılarak, yorumlanabilen modellerin elde edilmesi amaçlanmaktadır. Burada çoklu bağlantı sorunu üzerinde durularak, çözüm için ceza katsayısı olan iz (*ridge*) cezası kullanılmıştır (Mukherjee, 2013).

Velu ve Reinsel (2013) tarafından yazılan “Çok Değişkenli İndirgenmiş Rank Regresyon (*Multivariate Reduced Rank Regression*)” adlı eserde, indirgenmiş rank

regresyon hakkında kapsamlı bilgiler sunmaktadır. Ayrıca tez çalışmam için yaralandığım bu ikinci eserde, açıklayıcı değişkenler için matematiksel ifadeler ortaya konulmuş olup konunun daha anlaşılır olması sağlanmıştır (Velu ve Reinsel, 2013).

Huang ve arkadaşları (2006) tarafından yazılan makalede, az sayıda değişkenler ile çalışarak büyük boyutlu kovaryans matrisler için istatistiksel anlamda bir tahmin edici üreterek, parametrik olmayan bir model önerilmiştir. Burada “Cholesky Ayrıştırması” kullanılarak model seçiminde bir tahminde bulunulmuştur (Huang, vd., 2006).

Bernardini ve Cubadda (2015) tarafından yazılan makalede, orta büyüklükte olan otoregresif modeller için indirgenmiş rank regresyon yönteminde bir strateji önerilmiştir. Burada büyük boyutlu kovaryans matrislerinin tersinin alınması durumunda, kanonik korelasyon yöntemine alternatif ve asimptotik olarak eşdeğer olan daraltılmış tahmin edici yöntemler önerilmektedir (Bernardini ve Cubadda, 2015).

3. ÇOK DEĞİŞKENLİ İNDİRGENMİŞ RANK REGRESYON MODELİ

3.1. Çok Değişkenli İndirgenmiş Rank Regresyon Modeli Nedir ve Ne Zaman Kullanılmaktadır?

Birçok araştırmalar sonucunda ve verilerin elde ediliş yöntemlerinin gelişmesi üzerine, ölçüm sistemlerindeki ilerlemeler açısından birbiriyle ilişkili birçok değişken meydana gelmektedir. Buna göre bağımlı değişkenler ile tahmini değişkenler arasındaki yüksek korelasyonun olması durumunda bir takım istatistiksel yöntemler kullanılmıştır ve bu yöntemlerden birisi de çok değişkenli regresyon yöntemidir (Burnham, vd., 1996). Bu bağlamda çok sayıda bağımlı değişkenin ve bağımsız değişkenlerin olduğu durumlarda, veri analizleri için çok değişkenli regresyon yöntemi en çok kullanılan istatistiksel bir yöntem olarak bilinmektedir.

Çok değişkenli regresyon yönteminde iki durum söz konusudur. İlk olarak tüm regresyon katsayılarının tahmini için çok sayıda gözlemin olması, ikincisi ise çok sayıda gözlem olsa da regresyonu yorumlama durumunun zor olmasıdır (Velu ve Reinsel, 2013). Bu yüzden parametre sayısının indirgenmek istenmesi, çok değişkenli regresyon yöntemini hem daha iyi tahmin etmek hem de daha iyi yorumlamak açısından arzu edilmektedir.

Daha sonra bahsedilecek olan çok değişkenli indirgenmiş rank regresyon yöntemi için regresyon katsayısı matrisi, düşük boyutlu iki bileşenli matris çarpımı olarak yazılmaktadır (Velu ve Reinsel, 2013). Çok sayıda bağımlı ve tahmin değişkenleri ile modelleme yapıldığında kısıtlamalar açısından yararlı olmaktadır. Bağımlı değişkenlerin belirli doğrusal kombinasyonları regresyon yöntemleri için göz ardı edilmektedir. Bundan dolayıdır ki, regresyon yönteminde elde edilen kombinasyonların tahmin değişkenleri ile ilişkisi olmadığı görülmüştür (Velu ve Reinsel, 2013).

Regresyon yöntemi için bir başka çıkarım ise modelde yer alan tahmin değişkenleri olup, sadece belirli doğrusal kombinasyonların kullanılmasını sağlayarak tahminde bulunulması olarak ifade edilmektedir. Çünkü kalan doğrusal kombinasyonlar için ilk doğrusal kombinasyon kümeleri verildiğinde bağımlı değişkenler üzerinde etkisi olmadığı görülmüştür. (Velu ve Reinsel, 2013). Buna göre çok değişkenli indirgenmiş rank regresyon yöntemi, regresyon katsayısı matrisi için

daha düşük ranklı varsayım altında çok değişkenli regresyon yöntemini oluşturarak, boyut indirgemesiyle ilgilenilmiştir.

Buradan, klasik çok değişkenli regresyon yöntemini aşağıdaki gibi tanımladığımızda (Velu ve Reinsel, 2013:3);

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

Burada;

Y : Bağımlı değişkenler olup $(q \times n)$ boyutludur ve $Y = [Y_1 Y_2 \dots Y_q]^T$ şeklinde ifade edilmektedir.

X : Bağımsız değişkenler olup $(p \times n)$ boyutludur ve $X = [X_1 X_2 \dots X_p]^T$ şeklinde ifade edilmektedir.

C : Regresyon katsayısı matrisi olup $(q \times p)$ boyutludur. Bu matris β yerine kullanılmaktadır.

ε_k : Hata matrisidir olarak bilinmektedir ve $(q \times p)$ boyutlu olup $\varepsilon = [\varepsilon_1 \varepsilon_2 \dots \varepsilon_q]^T$ şeklinde ifade edilmektedir (Velu ve Reinsel, 2013).

$$[Y_1 Y_2 \dots Y_q]^T = C_{q \times p} [X_1 X_2 \dots X_p]^T + [\varepsilon_1 \varepsilon_2 \dots \varepsilon_q]^T \quad (3.2)$$

Bu modelde regresyon katsayısı matrisi olan "c" ile kovaryans matrisi olan " Σ_{ee} " bilinmeyen parametreler olup hataların normallik varsayımı altında en küçük kareler yöntemi veya maksimum olabilirlik yöntemi ile tahmin edilebilir (Velu ve Reinsel, 2013). Model (3.1)'deki klasik çok değişkenli regresyon yöntemi, bağımlı değişkenlerin birbirleriyle olan ilişkilerini doğrudan kullanmamaktadır. En önemli durum ise, yeteri kadar bağımlı değişkenler olsa bile regresyon matrisindeki parametre sayısının fazla olmasıdır (Velu ve Reinsel, 2013). Bundan dolayıdır ki, parametre sayısının indirgenmesine ihtiyaç duyulmuştur. Bunun için ise, Model (3.1)'de yer alan "c" matrisinin rankı düşürülerek bir varsayımda bulunulmuştur (Velu ve Reinsel, 2013). Buradan hareket edecek olursak, klasik çok değişkenli regresyon yönteminde regresyon katsayısı, tam ranklı ve modeldeki hata terimlerinin bağımsız olması gerektiğini varsaymaktadır. Ancak bazı durumlarda ise regresyon matrisinin rankının tam olmadığı görülmüştür. Bu durumda, regresyon yönteminde yer alan katsayı matrisinin düşük rank varsayımı altında, çok değişkenli regresyon yöntemine göre daha az sayıda parametre ile açıklanabilmesine imkân sağlamak için çok değişkenli indirgenmiş rank regresyon yöntemine ihtiyaç duyulmuştur (Velu ve Reinsel, 1987).

Çok değişkenli indirgenmiş rank regresyon yöntemi, 1940'lı yılların sonlarında ve 1950'li yılların başlarında Ted. W. Anderson tarafından ifade edilmiştir ve birçok uygulamalarda kullanılmaya başlanmıştır. Çok değişkenli indirgenmiş rank regresyon yöntemi, çok değişkenli regresyon yönteminde yer alan katsayı matrisinin tam ranklı olmamış halidir. Burada tespit edilemeyen rank değişkeni, bağımlı değişkenler varsayımı altında tahmin edilmektedir (Bura ve Cook, 2003). Çok değişkenli İndirgenmiş rank regresyon yönteminin geometrik yorumu, tahmin modellerinin grafiksel gösteriminde yardımcı olmaktadır (Schmidli, 2013). Çok değişkenli indirgenmiş rank regresyon yönteminde regresyon katsayı matrisi, rankı sınırlandırarak toplam parametre sayısını azaltır ve tahmindeki verimliliği artırmaktadır (Cook, vd., 2015).

İstatistik ve Ekonometri konularında araştırma konusu olan indirgenmiş rank regresyon (*reduced rank regression*) yöntemi çok farklı istatistiksel problemleri analiz eden, boyutsallığı azaltan bir regresyon yöntemidir (Izenman, 2008). Çok değişkenli indirgenmiş rank regresyon yöntemi ise, çok değişkenli regresyon yönteminin kısıtlanmış hali olarak bilinmektedir (Izenman, 2008). Bir başka ifadeyle çok değişkenli indirgenmiş rank regresyon yöntemi, katsayı matrisinin rankının tamamlanmamış olma ihtimalini kabul ederek, katsayı matrisi için doğrusal kısıtlamaların olduğunu varsaydığımız bir durumda, düşük boyutlu iki bileşenli matrislerin bir ürünü olarak yazılabildiği bir model olarak ifade edilmektedir (Heinen ve Rengifo, 2008).

Çok değişkenli indirgenmiş rank regresyon yönteminin özel durumları bulunmaktadır. Buna göre çok değişkenli indirgenmiş rank regresyon yönteminin özel durumları aşağıdaki gibidir (Izenman, 2008:180);

i) Bağımlı değişkenler ile bağımsız değişkenlerin özdeş ($X \equiv Y$ (ve $p = q$) gibi) olduğu bir durumu varsaydığımızda ve kovaryans matrisleri üzerinden özdeğer matrisleri için gamma değerinin ($\Gamma = I_p$) $p \times p$ boyutlu birim matris olarak ayarlandığında, “Harold Hotelling temel bileşen analizi” veya “açıklayıcı faktör” olarak elde edildiği görülmüştür.

ii) Eğer $\Gamma = (\Sigma_{YY})^{-1}$ olarak belirlenirse, “Hotelling’ in Kanonik varyasyonu” ve “Korelasyon Analizi” elde edilmektedir.

iii) Bağımlı değişken bileşenleri 0 ve 1 olarak ayarlanırsa ve indirgenmiş rank regresyon yöntemi için “Kanonik Değişken Analizi” kullanıldığında Fisher’ in doğrusal diskriminant analizi elde edilmektedir.

Bu özel durumlardan yola çıkılarak çok deęişkenli indirgenmiş rank regresyon yöntemi, çok deęişkenli regresyon yöntemi için genel bir yöntem olarak kullanılabilereği veya kullanılabilir olma durumlarını göstermektedir (Izenman, 2008). İndirgenmiş rank regresyon yönteminde esas belirlenecek olan rank deęeri olup, β deęerlerinin gerçek deęerleri ile tahmin edilen β deęerlerinin istatistiksel olarak karşılaştırılması durumu olarak belirlenmektedir (Izenman, 2008). Rank deęerini belirleme sorunu bir seçim problemi olup rank deęerinin, p ve q gibi minimum deęerinden ($r \leq \min(p, q)$) gibi küçük veya küçük eşit olması şartıyla en küçük tam sayı seçilmektedir. Buna göre bağımsız deęişken üzerindeki Y deęerinin, indirgenmiş rank regresyondaki rankına yakın olmaktadır. Örneğin, “ r ” betanın gerçek rankı olsun. Buna göre ($r \leq \min(p, q)$) şartı gereęi, tahmini regresyon katsayı matrisindeki ve hata kovaryans matrislerindeki deęerleri rank deęiştikçe deęişmektedir. Ancak betanın gerçek deęerlerine yaklaştıkça deęerler deęişmeyi bırakır ve stabilize olmaktadır (Izenman, 2008).

Katsayı matrisleri için rank durumu önemli olduğundan indirgenmiş rank regresyon yöntemi, katsayı matrisleri üzerinde rank kısıtlamalarının yapıldığı çok deęişkenli regresyon yöntemi olarak da tanımlanmaktadır (Schmidli, 2013). Burada rank kısıtlamalarının göz ardı edildięi görülmektedir. Ancak rank deęerlerinin şartlandırıldığı durumlarda ve uygun koşulların sağlanamadığı durumlar oluştuğunda, üç ana şart belirlenmiştir. Bu şartlar aşağıdaki gibi ifade edilmektedir (Schmidli, 2013:57).

i) Katsayı matrisinin (c), parametre sayıları indirgenmiş rank regresyon yöntemi kullanıldığında azalacağından, birçok parametre deęerlerinin dikkate alınmadığı görülmektedir. Örneğin; $p=10$, $q=5$ ve $r=1$ olan bir model olsun. Buradan parametre sayısı büyük bir ihtimalle $pq=50$ olması tahmin edilmektedir. Ancak rank deęerinin de kullanılarak birtakım düzeltmeler yapıldığında ($r(p+q-r)$, ... gibi) indirgenmiş rank regresyon yöntemi için parametre sayısının $p \times q$ deęerinden az olduğu görülmektedir (Schmidli, 2013).

ii) Bağımlı deęişkenler için elde edilen tahmin deęişkenleri, çok deęişkenli regresyon yönteminde bazen çelişkili durumlar vermektedir. Bu durum tahmin edilen hatalardan kaynaklanmakta olup, $r=1$ durumunda indirgenmiş rank regresyon yöntemi için yeterli bir durum ise sorunun düzeldięi yani çok deęişkenli regresyon yönteminde görülen çelişkili durumların görülmedięi, modelin yorumlanabileceęi görülmüştür (Schmidli, 2013:57).

iii) Tahmin deęerlerinin daha net olması için indirgenmiř rank regresyon yönteminde daha az sayıda parametrelerin olması, örneklemin büyük olduęu durumlar için kanıtlanabilen bir durum iken, örnekleme deęerlerinin küçük olduęu durumlarda ise makul görülebilen bir durum olmuřtur (Schmidli, 2013:57).

İndirgenmiř rank regresyon yönteminde iki farklı tahmin söz konusudur. Bunlardan birincisi, hatalar ile ilgili duruma baęlı olarak parametre tahmini için kanonik korelasyon analizi veya baęımlı deęiřkenler için temel bileřenler analizinden elde edilme durumudur. İkincisi ise yardımcı bileřenler aracılıęıyla temel bileřenler analizini ya da hataların analizi kullanılarak elde edilme durumudur (Braak ve Looman, 1994). İstatistikte boyut küçültme önemli bir sorun olup bölünmüş ters regresyon, bölünmüş ortalama varyans tahmini ve Hessian yöntemi gibi birçok yöntem bulunmaktadır. Ancak en bilinen yöntem Anderson (1951) tarafından geliştirilmiř ve Izenman (1975) tarafından kullanılan çok deęiřkenli indirgenmiř rank regresyon yöntemidir (Yee ve Hastie, 2003).

Çok deęiřkenli indirgenmiř rank regresyon yöntemi, belirli bir model tarafından kabul edilen deęerli yapısal bilgilerden yararlanılarak boyut küçültme işlemini gerçekleştirir ve önceki modele özgü parametrelerin azaltılmasını mümkün kılmaktadır (Feng ve Li, 2006). Çok deęiřkenli indirgenmiř rank regresyon yöntemi, çok deęiřkenli istatistiksel yöntem için önemli bir araç olup verilerin boyutunu indirgeyerek yorumlamayı sağlamaktadır ve az sayıda deęiřkenler ile analizi gerçekleřtirecek olan modellere izin vermektedir (Heinen ve Rengifo, 2008). Çok deęiřkenli indirgenmiř rank regresyon yöntemi, çok deęiřkenli tahmine yönelik bir yaklařım olup; kanonik korelasyon analizi, artık analizi, kanonik diskriminant analizi ve temel bileřenler analizi gibi birçok standart çok deęiřkenli yöntem analizleriyle yakından baęlantılıdır (Schmidli, 2013). Anderson (1951) tarafından geliştirilen çok deęiřkenli indirgenmiř rank regresyon yöntemi, ekonometri ve zaman serisi analizleri gibi birçok alanlarda kullanılmaktadır (Anderson, 1999). Çok deęiřkenli indirgenmiř rank regresyon yöntemi; Anderson (1951), Izenman (1975) gibi birçok bilim insanı tarafından incelenmiřtir (Tso, 1981). Genel bir parametre azaltma yöntemi olan çok deęiřkenli indirgenmiř rank regresyon yöntemi, istatistiksel analizlerde daha iyi sonuçlar elde eden alternatif bir yol olarak karřımıza çıkmaktadır (Feng ve Li, 2006).

3.2. Çok Değişkenli İndirgenmiş Rank Regresyon Modelinin Parametreleri ve Parametrelerin Elde Ediliş Yöntemleri

Anderson (1951) tarafından geliştirilmiş olan çok değişkenli indirgenmiş rank regresyon yöntemi aşağıdaki gibi gösterilmektedir. Şöyle ki,

Varsayalım ki, $Y = [Y_1 Y_2 Y_3 \dots Y_q]^T$ ve "q" boyutlu bağımlı değişken ve $X = [X_1 X_2 X_3 \dots X_p]^T$ ve "p" boyutlu bağımsız değişken olsun. Bu ifadeler kapsamında çok değişkenli indirgenmiş rank regresyon yöntemi için denklem aşağıdaki gibidir.

$$Y = \alpha + \beta^T X + \varepsilon \quad (3.3)$$

Burada çok sayıda bağımlı değişkenin ve bağımsız değişkenlerin bulunduğu durumlar için şu ifadeler yer almaktadır.

Y : Bağımlı değişken vektörü olup $[Y_1 Y_2 Y_3 \dots Y_q]^T$ şeklinde gösterilir ve (q x n) boyutludur.

X : Bağımsız değişken vektörü olup $[X_1 X_2 X_3 \dots X_p]^T$ şeklinde gösterilir ve (p x n) boyutludur

α : Kesişim parametresi olarak ifade edilmekte olup q boyutludur.

β : Katsayı matrisi olarak ifade edilir ve $\begin{bmatrix} \beta_{11} & \beta_{12} & \dots & \beta_{1q} \\ \beta_{21} & \beta_{22} & \dots & \beta_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{p1} & \beta_{p2} & \dots & \beta_{pq} \end{bmatrix}$ şeklinde

gösterilmektedir. Bu katsayı matrisi ise (p x q) boyutlu olarak ifade edilmiştir.

ε : Hata terimlerin matris vektörü olarak bilinir ve q boyutludur. Aynı zamanda hata vektörü $\varepsilon^T \varepsilon = I$ varsayımını sağlayacak şekilde olmalıdır.

Buradan yola çıkılarak oluşturulacak model denklemi şu şekildedir.

$$[Y_1 Y_2 \dots Y_q]_{q \times n}^T = \alpha_{q \times n} + \beta_{q \times p}^T [X_1 X_2 \dots X_p]_{p \times n}^T + [\varepsilon_1 \varepsilon_2 \dots \varepsilon_q]_{q \times n} \quad (3.4)$$

Velu ve Reinsel (2013) tarafından yazılan Çok Değişkenli İndirgenmiş Rank Regresyon (*Multivariate Reduced Rank Regression*) adlı eserde, yer alan çok değişkenli indirgenmiş rank regresyon yöntemine göre, model için parametre sayısını azaltma gibi bir durum söz konusu olmaktadır (Velu ve Reinsel, 2013). Bu durumda çok değişkenli regresyon yöntemini elde etmek için Model(3.1) 'de kapalı formülünde yer alan "c" katsayı matrisinin rankını düşürülerek aşağıdaki gibi bir varsayımda bulunulmuştur (Velu ve Reinsel, 2013).

$$\text{rank}(C) = r \leq \min(p, q) \quad (3.5)$$

Bu varsayıma göre, “ $p \leq q$ ” ve “ $p > q$ ” gibi her iki durum için rank değeri eşit olacak şekilde uygulanmaktadır. Model (3.5)' de ifade edildiği üzere, bu varsayım altında c katsayı matrisi, tam ranklı iki alt boyutlu bir matris ürünü olarak yazılmaktadır. Öyle ise c katsayı matrisinin rankının r olduğu bilindiğine göre varsayım altında aşağıdaki gibi yazılmaktadır (Velu ve Reinsel, 2013).

$$C = A_{q \times r} B_{r \times p} \quad (3.6)$$

Model (3.1)' de katsayı matrisini yerine koyacak olursak,

$$Y = ABX + \varepsilon \quad (3.7)$$

Y : Bağımlı değişken vektörü olup $[Y_1 \ Y_2 \ Y_3 \ \dots \ Y_q]^T$ şeklinde gösterilir ve $(q \times n)$ boyutludur.

C: Katsayı matrisidir ve $C = AB$ şeklinde tanımlanmakta olup $(q \times p)$ boyutludur.

X : Bağımsız değişken vektörü olup $[X_1 \ X_2 \ X_3 \ \dots \ X_p]^T$ şeklinde gösterilir ve $(p \times n)$ boyutludur

ε : Hata terimlerin matris vektörü olarak bilinir ve $[\varepsilon_1 \ \varepsilon_2 \ \dots \ \varepsilon_q]$ şeklinde ifade edilmekte olup $(q \times n)$ boyutludur.

Buradan yola çıkılarak oluşturulacak model denklemi şu şekildedir.

$$[Y_1 \ Y_2 \ \dots \ Y_q]_{q \times n}^T = A_{q \times r} B_{r \times p} [X_1 \ X_2 \ \dots \ X_p]_{p \times n}^T + [\varepsilon_1 \ \varepsilon_2 \ \dots \ \varepsilon_q]_{q \times n} \quad (3.8)$$

Regresyon yöntemleri için sık karşılaşılan sorunlardan biri de çoklu bağlantı problemidir. Bir veri setinde bu problemin incelenmesi için koşul indeksi bu konuda faydalı olmaktadır. Koşul indeksi, regresyon yöntemlerinde çoklu bağlantı ölçüsü olarak bilinmektedir ve “CI” şeklinde ifade edilmektedir. Koşul indeksi, bağımsız değişkenlerden elde edilmiş olan maksimum özdeğer ile minimum özdeğerin bölümünden oluşan karekök değeri olarak hesaplanmaktadır. Birden büyük bir değer alan koşul indeksi; 15 değerinin altında bir değere sahipse ($CI < 15$) veride çoklu bağlantı probleminin zayıf olduğunu, 15 ile 30 değerleri arasında ise ($15 \leq CI \leq 30$), çoklu bağlantı probleminin orta düzeyde olabileceğini ve 30 değerinin üstünde bir değer hesaplanmışsa ($CI > 30$), bu durumda veride çoklu bağlantı problemlerinin çok kuvvetli olduğunu göstermektedir (Shrestha, 2020:40). Buna göre, aşağıda koşul indeksine ait matematiksel ifade şu şekilde yer almaktadır:

$$CI = \sqrt{\frac{\lambda_{max}}{\lambda_{min}}} ; \begin{cases} \text{Çoklu Bağlantı zayıf düzeyde, } CI < 15 \\ \text{Çoklu Bağlantı orta düzeyde, } 15 \leq CI \leq 30 \\ \text{Çoklu Bağlantı çok kuvvetli, } CI > 30 \end{cases} \quad (3.9)$$

3.3. Çok Değişkenli İndirgenmiş Rank Regresyon Modelinde Kovaryans Matrisinin Önemi

Gözlem sayısı büyük olan bir veri setinin kovaryans matrislerinin tahmini, çok değişkenli istatistikte ve çok değişkenli regresyon yöntemleri için önemli bir temel sorun oluşturmaktadır (Vershynin, 2012). Kovaryans matrisi çok değişkenli indirgenmiş rank regresyon yönteminde ve diğer regresyon yöntemlerinde önemli bir yere sahip olmaktadır. Kovaryans matrisi varyans – kovaryans matrisi olarak adlandırılan ve $X_1, X_2 \dots X_p$ gibi rastgele değişkenler arasındaki varyans ve kovaryansları içeren ($n \times n$) boyutlu kare matris olarak bilinmektedir. Bu matrisin köşegen elemanları varyansları içerirken köşegen haricindeki elemanları ise iki değişken arasındaki kovaryans değerlerini göstermektedir (Montgomery ve Runger, 2010).

Kovaryans matrisi bütün çok değişkenli regresyon analizlerinde ve diğer birçok uygulamalar için gerekli olmaktadır. Kitle kovaryans matrisi hakkında bilgi edinebilmek için örneklem kovaryans matrisine ihtiyaç duyulmaktadır. Örneklem kovaryans matrisi ise örneklem boyutu küçük olan veriler için daha iyi sonuç verirken, büyük boyutlu verilerde çok iyi performans göstermemektedir. Bundan dolayıdır ki, büyük boyutlu veriler için kovaryans matrisini düzeltmenin birkaç yolu bulunmaktadır. Bu yollardan biri daraltılmış kovaryans tahmin edici yöntemleri (*shrinkage covariance estimator methods*) iken diğeri özdeğer yöntemleri (*eigen methods*) olarak bilinmektedir. Bir başka yöntem ise veriler arasında boyut indirgeme yöntemi olup farklı yaklaşımlar söz konusudur (Fan, vd., 2008). Lin ve Perlman (1985), Wong ve arkadaşları (2003) ve birçok bilim insanları tarafından kovaryans matrisinin tahmin edicisini, örneklem kovaryans matrisi için pozitif tanımlı ve yansız olduğunu belirtirken büyük boyutlu kovaryans matrisleri için ise kararsız olduğunu belirtmektedirler (Huang, vd., 2006).

Buna göre klasik bir biçimde olan örneklem kovaryans matrisi aşağıdaki gibi gösterilmektedir.

$$S = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})(X_i - \bar{X})^T ; i = 1, 2, \dots \infty \quad (3.10)$$

$$\bar{X} = \frac{1}{n-1} \sum_{i=1}^n X_i ; i = 1, 2, \dots \infty \quad (3.11)$$

Çok büyük boyutlu veriler için örnek kovaryans matrisinin öz değerleri gerçek öz değerler için zayıf bir tahmin olmaktadır. Yani öz değerlerin en küçük olanı sıfıra yaklaşırken en büyük olan öz değer ise sıfırdan uzaklaşarak daha büyük bir değer alabilmektedir. Bu durum ise kovaryans matrisinin kötü koşullara sahip olduğunu göstermektedir (Fisher ve Sun, 2011). Bazı durumlar vardır ki veri içerisindeki küçük bir değer, kovaryans matrisinin duyarlı olmasına yol açmaktadır. Bu durumda matrisin tersi alınsa bile elde edilen sonuçların tam anlamıyla veriyi yansıtamadığı görülmüştür (Izenman, 2008). Bazı durumlarda ise matrisin tersi alınamamaktadır. Bu gibi durumlarda regresyon modellerinde karşımıza çıkacak olan sorunlardan birkaçı, çoklu bağlantı problemleri ve bağımsız değişkenler arasındaki yüksek korelasyonlardır (Izenman, 2008). Bundan dolayıdır ki, genellikle büyük boyutlu verilerde çok değişkenli regresyon yöntemlerini, makul bir biçimde gözlenen regresyon yöntemlerinden farklı olarak pozitif tanımlı ve simetrik olduğu bilinen bir matrise sahip olması beklenmektedir. Ancak çok büyük boyutlu matrisler verilerden de anlaşılacağı üzere kötü koşullar eğiliminde olmaktadır. Bunun sebebi ise örneklem sayısının artması durumudur. Eğer örneklem sayısı (n) artarsa determinant sıfıra yaklaşır ve kovaryans matrisi için kötü koşullar söz konusu olmaktadır (Riley, 1955).

Burada çalışmakta olduğum tezin asıl amacını oluşturacak yöntem,

“Bağımsız değişkenlerin çoklu bağlantılarından dolayı kötü koşullar altında beta katsayılarının en küçük kareler tahmin yöntemi, aşırı derecede büyük örneklemlemlere sahip kararsız tahmin ediciler üretmesidir (Perez-Melo ve Kibria, 2020)” ifadesiyle “Stats” dergisinde yer alan ve 2020 yılında yayınlamış olan bu makalede belirtilmektedir. Bu amaç doğrultusunda çok değişkenli regresyon yöntemi için kararsız görülen tahmin edici yöntemlerini kararlı tahmin ediciler olarak belirleyerek beta katsayıları üzerindeki kararsızlığı gidermek hedeflenmiştir.

Çoklu bağlantı problemi, değişkenler arasında çok yüksek oranda ilişki olması durumunda ortaya çıkmakta olup korelasyon matrisiyle ilgili bir durumdur. İki değişkenli bir regresyon analizi için çoklu bağlantı problemi daha kolay görülürken, çok değişkenli regresyon analizlerinde bu problemi görebilmek için çok değişkenli istatistik yöntemlerinden yararlanılarak sorunun giderilebilmesi sağlanmaktadır (Tabachnick, vd., 2007). Çok değişkenli regresyon yönteminde bağımsız değişkenlerin birbiriyle daha çok ilişkili olduklarında çoklu bağlantı problemiyle

karşılaşılmaktadır. Çok değişkenli regresyon yönteminde veya diğer regresyon yöntemlerinde de çokça karşılaşılan bir problem konusu olmuştur (Aljandali, 2017). Çoklu bağlantı problemi regresyon katsayılarının yanlış yorumlanmasına ve belirtme katsayısı olan R^2 'nin yüksek çıkması durumunda dahi katsayılar hakkında yapılacak olan çıkarımların güvenilmez olduğunu göstermektedir. Bu sorunun çözümü için birkaç yöntem geliştirilmiştir. Bunlardan biri de iz (*ridge*) regresyon yöntemi olarak bilinmektedir (Aljandali, 2017). İz (*ridge*) regresyon, çoklu bağlantı problemini çözmeye fayda sağlayacak alternatif bir yöntemdir (Perez-Melo ve Kibria, 2020). Çoklu bağlantı problemi regresyon katsayılarının tahminleri üzerinde ciddi bir etkiye sahip olmaktadır. Aynı zamanda bu problem, regresyon katsayılarının kararlı olmalarını etkilediğinden, bu probleme karşı en küçük kareler yöntemine göre daha az duyarlı olan alternatif bir çözüm olarak iz (*ridge*) regresyon yöntemi karşımıza çıkmaktadır (Montgomery ve Runger, 2010). İz (*ridge*) regresyon yöntemi, hata kareler ortalamasına sahip olan kararlı regresyon katsayılarını bulmaya çalışmaktadır. Bu regresyon çoklu bağlantının olduğu durumlar için uygun olmaktadır (Montgomery ve Runger, 2010). İz (*ridge*) regresyon yöntemi üç farklı şekilde nitelendirilmektedir. Buna göre;

- i) Hata kareler ortalamasını en az indiren bir tahmin edici yöntemiyle
- ii) En küçük kareler tahmin edicisini başlangıç noktasına küçülterek daraltılmış tahmin edici (*shrinkage estimator*) yöntemiyle
- iii) Bayes tahmin edici yöntemiyle (Izenman, 2008).

Çoklu bağlantı problemi aynı zamanda bağımsız değişkenlerin yani X 'lerin ortogonal olmadığı durumları da ele alarak, bu problemi çözmek için Hoerl ve Kennard (1970) tarafından geliştirilmiş olan en küçük kareler tahminleri yerine, bir iz (*ridge*) tahmini kullanarak sorunun giderilebileceğini göstermektedir (Zhang ve İbrahim, 2005). Burada " k " iz (*ridge*) parametresi olup β 'lar üzerinde iz (*ridge*) tahmini uygulanmasını sağlamaktadır (Zhang ve İbrahim, 2005). Tahmin de yer alan " k " parametresi 0 (sıfır) değerini alırsa maksimum olabilirlik tahminleri elde edilmektedir (Algama, 2008).

Bu açıklamalar doğrultusunda iz (*ridge*) tahmin edicisinin matematiksel olarak ifadesi aşağıdaki gibidir. Buna göre,

$$\hat{\beta}(k) = (X^T X + kI_p)^{-1} X^T Y \quad (3.12)$$

İz (*ridge*) regresyon üzerinde bazı tahmin yöntemleri kullanılmaktadır. Bu tahmin yöntemlerinden biri daraltılmış kovaryans tahmin edici yöntemi (*shrinkage covariance estimation method*) olup kötü koşullandırılmış örneklem kovaryans matrisi için meydana gelecek olan sorunları çözmeyi ve matrisin tersinin alınamadığı özel durumlarda çoklu bağlantı gibi bir sorunun çözümü için kolaylık sağladığından dolayı gerek duyulmuştur (Fisher ve Sun, 2011). İz (*ridge*) regresyon yöntemi, varyans büyüklüğünü azaltarak modelde yer alan katsayıları sıfıra yakın olacak şekilde indirgeyen bir daraltma yöntemidir ve $X^T X$ matrisine pozitif bir değer eklenerek hesaplanmaktadır (Algama, 2018). Efron (1975) tarafından araştırılmış olan daraltılmış kovaryans tahmin edici yönteminin, kovaryans matrisleri için elde edilen tahminleri iyileştirilmesine yönelik yaygın bir tahmin edici yöntemi olduğu görülmüştür. Çok büyük boyutlu matrislerden elde edilen öz değerler, gerçek öz değerlere oldukça uzak olduğu bilinmektedir. Bu durum ise örnek kovaryans matrisinin kötü koşullandırılmış olduğunu ve tekil olmasına neden olmaktadır (Fisher ve Sun, 2011). Bundan dolayıdır ki, yüksek düzeyde korelasyonların olması çoklu bağlantı sorununa neden olmaktadır. Bu sorunun çözümü için ise çeşitli yöntemler kullanılmaktadır. Bu yöntemlerden biride yukarıda da adı geçtiği gibi daraltılmış kovaryans tahmin edici yöntemi (*shrinkage covariance estimation method*) olarak ifade edilmiştir.

Daraltılmış kovaryans tahmin edici yönteminin bir diğer tanımı ise, regresyon katsayılarında önemsiz olan verilerin sıfıra yakın değerlere sahip olacak şekilde sınırlandırılması olarak tanımlanmıştır (Simila ve Tikka, 2007). Daraltılmış kovaryans tahmin edici yöntemi, maksimum olabilirlik yöntemine eş değer olup sonlu örnekler için sayısal olarak daha kararlı olmaktadır. Stein (1956) tarafından da ifade edilen tahmin edici yöntemi var olan iki tahmin edicinin optimal ağırlık ortalaması olarak tanımlanmaktadır (Bernardini ve Cubadda, 2015).

Kovaryans matrisleri için ortaya çıkmış bir diğer sorun ise yukarıda da bahsedildiği üzere matrisin tersinin alınamaması olarak ifade edilmiştir. Bu sorunun sebeplerinde biri olarak ise örnek kovaryans matrisinin tekil olma durumu veya matris boyutunun (p), gözlem sayısından (n) büyük olması olarak gösterilmektedir (Van Wieringen ve Peeters, 2016). Kovaryans matrisinin tersinin alınabilmesi ve matrisin tekil olmaması için iyi koşullandırılmış olması gerekmektedir. Böylece kovaryans matrisi tam ranklı olup kolayca tersinin alınabildiği görülmektedir (Schäfer ve Strimmer, 2005). Kötü koşullandırılmış bir kovaryans matrisinin tersi

alınmadığından sayısal olarak sonuç vermediği görülmektedir. Bu durumda, bağımsız değişkenler arasında yukarıda da bahsedildiği üzere çoklu bağlantı gibi, regresyon yöntemlerinde karşımıza çıkacak temel sorunlardan bir tanesi görülmektedir. Bu durum değişkenler arasında çok yüksek düzeyde ilişki olduğunu belirtmektedir (Izenman, 2008). Kötü koşullandırılmış bir kovaryans matrisinin çözümü için, Hadamard (1902) tarafından üç koşul sunulmuştur. Buna göre,

- i) Problemin tek çözümü olmalı (*existence*)
- ii) Var olan çözümün tek olma durumu (*uniqueness*)
- iii) Çözüm problemin verilerine bağlı olmamalı (*stability*)

Jaques Hadamard tarafından ortaya konulan bu şartlara göre, bir tanesi sağlanmadığı takdirde kovaryans matrisinin tekil ve kötü koşullandırılmış olduğu belirtilmektedir (Erbaş, 2010:43).

Çok büyük boyutlu verilerde elde edilen kovaryans matrisleri için üzerinde durduğumuz sorunları (kötü koşullandırılmış kovaryans matrisi, matris tersinin alınamama durumu ve pozitif tanımlı olmayan matris gibi...) giderebilmek adına uygulanacak olan daraltılmış kovaryans tahmin edici yöntemleri önerilmiştir (Pamukçu, vd., 2019). Buna göre Pamukçu ve arkadaşları (2019) tarafından da yayınlanmış olan makalelerinde büyük boyutlu kovaryans matrisler için daraltılmış kovaryans tahmin edici yöntemlerinin tanımları şu şekildedir (Pamukçu, vd., 2019).

Maksimum olabilirlik ve Deneysel Bayes Tahmin Edici yöntemi, gerektiği ölçüde küçük bir miktar kaygı oluştuğunda kullanılan bir yöntemdir. SDE ve SRE, Kolmogorov'un son çalışmalarından olup SDE, Şartlandırılmış İz (*ridge*) Diyagonal Kovaryans Tahmin Edici yöntemi olarak bilinir ve değişkenlerdeki ölçek bağımlılığını ortadan kaldıran yöntem olarak tanımlanmaktadır (Pamukçu, vd., 2019). Dışbükey Toplam Kovaryans Tahmin Edicisi olarak bilinen CSE ise öz değerlerin ortalamalarının küçültülmesi olarak tanımlanmıştır. Bozdoğan'ın üzerinde çalıştığı BCSE olarak tanımlanmış olan kovaryans tahmin edicisi, veri kümesindeki farklı türlere veya değişkenlere izin veren genel kareler toplamının her bir bileşeninin boyutuyla ölçülmesini mümkün kılan kareler toplamının sapması olarak tanımlanmıştır (Pamukçu, vd., 2019). Chen ve arkadaşları (2010) tarafından önerilmiş olan ve Gauss modeli altında matris boyutunun gözlem sayısından büyük olduğu durumlarda kullanılan bir diğer kovaryans tahmin edicisi ise Oracle yaklaşım ile daraltılmış tahmin edici (OAS) yöntemidir (Pamukçu, vd., 2019). Bu tanımlamalardan

yola çıkılarak aşağıda elde edilmiş olan daraltılmış kovaryans tahmin edici yöntemlerinin matematiksel modelleri şu şekilde ifade edilmektedir. Buna göre;

i) Maksimum Olabilirlik / Bayes Tahmin Edici Yöntemi (MLE/EB):

$$\hat{\Sigma}_{MLE/EB} = \hat{\Sigma}_{MLE} + \frac{p-1}{n \operatorname{tr}(\hat{\Sigma})} I_p \quad (3.12)$$

I_p : $p \times p$ boyutlu birim matristir.

$\hat{\Sigma}_{MLE}$: Bağımsız değişkenler arasındaki kovaryans matrisi olup,

$$S_{xx(MLE)} = (n-1)S_{MLE} ; S_{MLE} = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})(X_i - \bar{X})^T$$

şeklinde ifade edilmektedir.

ii) Şartlandırılmış Ridge (iz) Tahmin Edici Yöntemi:

$$\hat{\Sigma}_{SRE} = \hat{\Sigma}_{MLE} + \hat{\rho} I_p \quad (3.13)$$

$$\hat{\rho} = p(p-1)[2n \operatorname{tr}(\hat{\Sigma}) - p]^{-1}$$

$\hat{\Sigma}_{MLE}$: Bağımsız değişkenler arasındaki kovaryans matrisi olup,

$$S_{xx(MLE)} = (n-1)S_{MLE} ; S_{MLE} = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})(X_i - \bar{X})^T$$

şeklinde ifade edilmektedir.

I_p : $p \times p$ boyutlu birim matristir.

iii) Şartlandırılmış Ridge (iz) diyagonal Kovaryans Tahmin Edici Yöntemi:

$$\hat{\Sigma}_{SDE} = (1 - \hat{\rho})\hat{\Sigma}_{MLE} + \hat{\rho} \operatorname{Diag}(\hat{\Sigma}_{MLE}) \quad (3.14)$$

$$\hat{\rho} = p(p-1)[2n \operatorname{tr}(\hat{R}) - p]^{-1}, \quad R : \text{korelasyon matrisidir.}$$

$\hat{\Sigma}_{MLE}$: Bağımsız değişkenler arasındaki kovaryans matrisi olup,

$$S_{xx(MLE)} = (n-1)S_{MLE} ; S_{MLE} = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})(X_i - \bar{X})^T$$

şeklinde ifade edilmektedir.

iv) Bozdoğan tarafından önerilmiş olan Yakınsak Toplam Kovaryans Tahmin Edici Yöntemi:

$$\hat{\Sigma}_{BCSE} = \hat{\rho}_{BCSE} \hat{\Sigma} + (1 - \hat{\rho}_{BCSE}) \hat{D} \quad (3.15)$$

$$\hat{\rho}_{BCSE} = \frac{1}{a}, \quad a = \frac{1}{n-1} \sum_{j=1}^p \operatorname{Var}(x_j)$$

$$\hat{D} = \frac{\operatorname{tr}(\hat{\Sigma})}{p} I_p, \quad p \geq 2 \text{ için}$$

v) Oracle Yaklaşım ile Daraltılmış Tahmin Edici Yöntemi:

$$\hat{\Sigma}_{OAS} = (1 - \hat{\rho}_{OAS}) \hat{\Sigma}_{MLE} + \hat{\rho}_{OAS} \hat{D} \quad (3.16)$$

$$\hat{\rho}_{OAS} = \min\left(\frac{1}{(n-1-\frac{2}{p})\hat{\Phi}}, 1\right)$$

$$\hat{\Phi} = \frac{tr(\hat{\Sigma}^2) - \frac{(tr(\hat{\Sigma}))^2}{p}}{(1-\frac{2}{p})tr(\hat{\Sigma}^2) + (tr(\hat{\Sigma}))^2}; \quad tr(\hat{\Sigma}^2) \geq (tr(\hat{\Sigma}))^2$$

$$0 \leq \hat{\Phi} \leq 1$$

vi) Dışbükey Toplam Kovaryans Tahmin Edicisi:

$$\hat{\Sigma}_{CSE} = \hat{\rho}\hat{\Sigma}_{MLE} + (1 - \hat{\rho})\hat{D} \quad (3.17)$$

$\hat{\Sigma}_{MLE}$: Bağımsız değişkenler arasındaki kovaryans matrisi olup,

$$S_{xx(MLE)} = (n-1)S_{MLE}; \quad S_{MLE} = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})(X_i - \bar{X})^T$$

şeklinde ifade edilmektedir.

$$\hat{\rho} = \frac{n}{n+m}; \quad 0 < m < \frac{2[p(1+\beta) - 2]}{p-\beta} \text{ ve } \beta = \frac{(tr(\hat{\Sigma}_{MLE}))^2}{tr(\hat{\Sigma}_{MLE})^2}$$

$$p \geq 2$$

$$\hat{D} = \frac{tr(\hat{\Sigma})}{p} I_p; \quad p \geq 2 \text{ için}$$

şeklinde ifade edilmektedir.

3.4. Çok Değişkenli İndirgenmiş Rank Regresyon Modelinde Kovaryans Matrisi için Model Seçimi

Çok değişkenli indirgenmiş rank regresyon modeli için model seçiminden önce bahsedilmesine gerek duyduğumuz, simülasyon uygulamalarında meydana gelen bir durum söz konusu olmuştur. Buna göre, hata matrisleri için R paketinde yer alan, Hoff ve arkadaşları (2019) tarafından geliştirilmiş olan “Rstiefel Paketi” kullanılmaktadır (Hoff, 2019). Çok değişkenli indirgenmiş rank regresyon modeli için önemli bir yere sahip olan bu paket, hatalar için kare matris olmayan bir matris oluştuğunda kullanılmaktadır. R’ de yer alan “Rstiefel Paketi” çok büyük verileri ortonormal matris olarak hesapladıktan sonra matris tersinin transpozuna eşit olduğu durumu gösterir (Hoff, 2019). Bir matrisin transpozu ile kendisinin çarpımının birim matris olduğu ve matrisin tersinin transpozuna eşit olduğu bilinmektedir. Ancak matris kare olmadığında ve tersinin transpozuna eşit olmadığı durumlar söz konusu olmaktadır. Bu ifadeden yola çıkılarak Hoff ve arkadaşları (2019) tarafından geliştirilmiş olan “Rstiefel Paketi” ne gerek duyulmuştur (Hoff, 2009).

Rstiefel Paketi, matrisi önce ortonormal matrise çevirmekte olup daha sonra matrisin transpozu ile tersinin çarpımını birim matrise dönüştürmektedir. Buna göre “Rstiefel Paketi” içinde ortonormal matrisler için kullanılmakta olan $r \leq q$ varsayımı altında ortonormal sütunlara sahip $q \times r$ boyutlu “Stiefel Manifold’ u” kullanılmıştır (Hoff, 2009). Stiefel Manifold, çeşitli kısıtlamalara maruz kalan gözlem sayısına bağlı çok büyük boyutlu veriler için bir yaklaşım modeli olarak tanımlanmaktadır (Markdahl, vd., 2020). Buna göre Stiefel Manifold $q \times r$ boyutlu olup \mathbb{R}^n ’deki r boyutlu ortonormal vektör uzayı olarak ifade edildiğinde (Markdahl, vd., 2020), aşağıdaki gibi tanımlanmaktadır. Şöyle ki,

$$St(q, r) = \{X \in \mathbb{R}^{q \times r} : X^T X = I_{r \times r}\} \quad (3.18)$$

Rstiefel Paketi için kullanılan “Von Mises Fisher Dağılımı” iki parametrelili üstel bir dağılıma (De Cao ve Aziz, 2020) sahip bir yöntem olup kapalı bir biçime sahip olmaktadır. Von Mises Fisher Dağılımı, Von Mises tarafından tanıtılmıştır (Dhillon ve Sra, 2003). Çok değişkenli normal dağılıma benzeyen bu dağılım, yönlü veriler için ortaya çıkmıştır (Dhillon ve Sra, 2003). Buna göre x değişkeni p boyutlu rastgele bir vektör olsun. Birim noktalar kümesi S^{p-1} olarak gösterilir ve $S^{p-1} = \{x \in \mathbb{R}^p : \|x\| = 1\}$ şeklinde ifade edilmektedir. Buradan p boyutlu Von Mises Fisher Dağılımının olasılık yoğunluk fonksiyonu, p boyutlu rastgele bir vektör olan x üzerinde dağılmaktadır ve $M_p(\mu, \kappa)$ olarak aşağıdaki gibi gösterilmektedir (Dhillon ve Sra, 2003).

$$M_p(\mu, \kappa) = c_p(\kappa) e^{\kappa \mu^T x}; \quad \begin{array}{l} x \in S^{p-1} \\ \|\mu\| = 1 \\ \kappa \geq 0 \end{array} \quad (3.19)$$

$$c_p(\kappa) = \frac{\kappa^{\frac{p}{2}-1}}{(2\pi)^{\frac{p}{2}} I_{p/2-1}(\kappa)}$$

olup, $c_p(\kappa)$ normalleştirme sabiti olarak bilinmektedir.

S^{p-1} : Birim noktalar kümesi olarak ifade edilmekte olup S_p olarak da yazılabilen p boyutlu bir kümedir.

Çok değişkenli regresyon yöntemleri için model seçimi farklı yollarla uygulanmaktadır. Gerçek modellere yakın en iyi model seçimi yapabilmek için en iyi modellerin seçilmesinde yardımcı olacak nesnel veri yöntemlerine ihtiyaç duyulmaktadır (Bedrick ve Tsai, 1994). Doğrusal modellerde modelleri karşılaştırmak için R^2 veya düzeltilmiş R^2 üzerinden bir yaklaşımda bulunmaktadır. Uyum iyiliğinin ölçüsü olan R^2 den en uygun olanı seçilirken burada “model karmaşıklığı”

ihmal edilmektedir (Portet, 2020:121). Bu ifadeler ışığında doğrusal modeller için R^2 ve düzeltilmiş R^2 matematiksel ifade olarak aşağıdaki gibi gösterilmektedir.

$$R^2 = \frac{SSR}{SS} = \frac{\sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2} \quad (3.20)$$

$$R_{adj.}^2 = 1 - (1 - R^2) \frac{n - 1}{n - k} \quad (3.21)$$

k : Modeldeki bağımsız değişken sayısıdır.

Bir modelin en iyisi seçilirken “az sayıda değişkenler ile çalışma” ilkesi gereğince basit bir model kullanılmalıdır. Bununla ilgili JohnsonandOmland (2004) tarafından model seçimi hakkında,

“Az sayıda değişkenler ile işlem yapmak varyans ve yanlılık arasındaki bir değiş tokuştur. Parametrelerin az oluşu tahmin ediciler için yüksek yanlılığa ve tüm faktörleri iyi tanımlamayan bir uyum modeline neden olmaktadır (Portet, 2020:121)” sözüyle ifade etmektedir. Doğrusal bir regresyon modelinde ileri (*forward*), geri (*backward*) ve adımsal (*stepwise*) gibi model seçimleri bulunmaktadır (Rao, vd., 2001).

Bilindiği üzere çok değişkenli regresyon yöntemi, birden çok değişkenin yer aldığı bir model olup değişkenlerin aralarındaki ilişkileri açıklamada kullanılan bir yöntem olarak bilinmektedir (Seghouane, 2006). Örneklem boyutunun küçük olduğu bir durum için çok değişkenli regresyon yöntemini tercih etmek bazı sorunlara neden olmaktadır. Bunun sebebi olarak ise parametre sayısının fazla olmasından kaynaklandığı söylenebilir (Seghouane, 2006).

Regresyon yöntemlerinde ve çok değişkenli regresyon yöntemleri için bağımlı ve bağımsız değişkenler üzerinde en iyi tahminde bulunabilmek adına bilgi kriterlerine ihtiyaç duyulmuştur (Timm, 2002). Model seçim kriterleri, gerçek model ile gerçek modele yakın en uygun model altında elde edilmiş olan tahmin değişkenleri arasındaki Kulback-Leibler uzaklığının minimize edilmiş yöntemi olarak tanımlanmaktadır (Pamukçu, 2017). Bilgi kriterleri (AIC, BIC, KIC, CAIC, HQIC... gibi) her zaman diğerinden daha iyi olmayabilir. Ancak bu kriterler kıyaslandığında belirli bir model yöntemleri için daha iyi bir sonuç vermektedir (Seghouane, 2006). Akaike (1974), Sugiura (1978), Schwartz (1978) ve HannanandQuinn (1979) tarafından çoklu regresyon modeli için elde edilen bilgi kriterlerini, çok değişkenli regresyon yöntemi veya çok değişkenli indirgenmiş rank regresyon yönteminde uygulanması adına daha geniş bir kapsamda hesaplanması için bazı çalışmaları bulunmaktadır (Timm, 2002).

Bilgi kriterleri için bilim insanları tarafından belirtilen bu ifadeler kapsamında matematiksel ifade olarak aşağıda gösterilmektedir (Timm, 2002).

$$AIC = -2\log|\widehat{\Sigma}_q| + 2p \quad (3.22)$$

Çok değişkenli AIC bilgi kriteri ise,

$$AIC_q = n\log|\widehat{\Sigma}_q| + 2qp + p(p + 1) \quad (3.23)$$

Kulback ve Leibler tutarsızlığı göz önünde bulundurularak çoklu regresyon yöntemi için log olabilirlik farkı tahmin edilerek BedrickandTsai (1994) tarafından elde edilmiş olan bir başka kriter ise,

$$AIC_{cq} = (n - q - p - 1)\log|\widehat{\Sigma}_q| + (n + q)p \quad (3.24)$$

BIC ve HQIC kriterleri ise aşağıdaki gibi ifade edilmektedir. Buna göre,

$$BIC_q = n\log|\widehat{\Sigma}_q| + qp\log(n) \quad (3.25)$$

$$HQIC_q = n\log|\widehat{\Sigma}_q| + 2qp\log\log(n) \quad (3.26)$$

n : Gözlem sayısı

$\widehat{\Sigma}_q$: Varyans – kovaryans matrisi olup q boyutludur.

Model kriterlerinden biri olan Kulback bilgi kriteri (KIC), Akaike bilgi kriterlerinden yola çıkılarak geliştirilmiştir. “Kulback’ın simetrik sapması”ndan tahmin edilmektedir (Seghouane, 2006). Bu bilgi kriteri uygun varsayım altında meydana getirilen model ile uygun model (yaklaşık) arasındaki Kulback’ın yansız tahmin edicisi olarak bilinmektedir (Seghouane, 2006). Bu model seçimine alternatif olan bir başka bilgi kriteri ise KIC_c 'dir. Bu bilgi kriteri, KIC bilgi kriterine benzer ve örneklemelerin az olduğu durumlarda güçlü model seçimlerini geliştirmektedir. Bu model çok değişkenli regresyon modelleri için uygun olmamakla birlikte yanlış bir tahmin modeli üretmekte olup parametrelerin yetersiz olduğu durumlarda uyum sağlayamamaktadır (Seghouane, 2006). Sebebi olarak da çok değişkenli regresyon yönteminde bilinmeyen parametrelerin olması olarak görülmektedir. Çok değişkenli regresyon modeli için yanlış bir tahmin yöntemi altında elde edilmiş olan vektör düzeltmesi (vc) anlamına gelen yeni bir kriter olarak belirtilen KIC_{vc} , AIC_c ve KIC_c bilgi kriterlerinden elde edilmiştir. Buna göre çok değişkenli için elde edilmiş olan bu bilgi kriteri aşağıdaki gibi ifade edilmektedir (Seghouane, 2006). Şöyle ki,

$$KIC_{vc} = -2\log|\widehat{\Sigma}_q| + \frac{nq(2p + q + 1)}{n - q - p - 1} + \frac{nq}{n - p - \frac{q - 1}{2}} + \frac{2qp + q^2 - q}{2} \quad (3.27)$$

4. UYGULAMA

Çok değişkenli indirgenmiş rank regresyon yöntemi için öncelikle benzetim çalışmaları ile sonuçları inceleyeceğiz ardından ise gerçek veri setlerinden oluşan iki uygulama göstereceğiz. Üzerinde analiz yaptığımız uygulamalar bağımlı ve bağımsız değişkenlerin birden fazla olan örneklerdir.

Öncelikle $q = 6$ tane bağımlı değişken ve $p = 5$ tane bağımsız değişkenden oluşan bir Monte Carlo benzetim çalışması (Bozdogan, 2003:42) üzerinde durularak gözlem sayısının 100, 500 ve 1000 olarak alındığı durumlar incelenmiştir.

$$Y = a + \beta^T X + \varepsilon \quad (3.28)$$

Çok değişkenli indirgenmiş rank regresyon yönteminde yer alan hatalar, standart normal dağılımlı $N(0,1)$ değerler olarak belirlenmiştir. Benzetim çalışmalarında yer alan σ değeri ise hataların standart sapması olarak belirlenmiştir. Bu değer, hata katsayılarının standart normal dağılımlı olarak belirlenmesi sebebiyle aynı değeri almıştır. Benzetim çalışmalarında elde edilecek olan sonuçlar, aşağıdaki durumlarla ilgili genel bir bilgi sunarak, çok değişkenli indirgenmiş rank regresyon yöntemi için fayda sağlamaktadır. Buna göre,

- i. Gözlem sayısı arttığında, daraltılmış kovaryans tahmin edici yöntemlerinin, β katsayıları üzerinden elde edilen hata kareler ortalaması üzerindeki etkisi,
- ii. Gözlem sayısı sabitken, korelasyon değerlerinin arttığı durumlar göz önüne alınarak daraltılmış kovaryans tahmin edici yöntemlerinin, hata kareler ortalaması üzerindeki etkisi,
- iii. Daraltılmış kovaryans tahmin edici yöntemler kullanılarak elde edilmiş olan bilgi kriterlerinin durumu,
- iv. Modelin performansını ölçmek için kullandığımız eğitim – test modelinde daraltılmış kovaryans tahmin edici yöntemlerin, hata kareler ortalaması üzerindeki etkisi gibi durumlar incelenmiştir.

Diğer bir simülasyon çalışması ise Hibrit kovaryans yapılarının kullanıldığı bir örnek çalışmadır. Burada 6 tane bağımsız değişken ve 5 tane de bağımlı değişken yer almaktadır (Pamukçu, 2019:2991). Bu iki çalışma R programı kullanılarak yapılmıştır. Buna göre aşağıda Bozdogan (2003) tarafından oluşturulmuş benzetim çalışması ile hibrit kovaryans yapılarının kullanıldığı diğer bir benzetim çalışması yer almaktadır.

4.1. Benzetim Çalışması – 1

İlk olarak beş bağımsız değişkenden oluşan birinci benzetim çalışması üzerinde çok değişkenli indirgenmiş rank regresyon yöntemi için analiz yapılacaktır. Burada $\varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4$ ve ε_5 değerleri standart normal dağılıma $N(0, 1)$ sahiptir. Korelasyon parametresi olarak ifade edilen " c " ise tahmini değişkenler arasındaki çoklu bağlantı katsayısını kontrol etmektedir. Aşağıda Benzetim Çalışması – 1 için ifade edilen örnek veri seti yer almaktadır. Buna göre,

$$X_1 = 10 + \varepsilon_1$$

$$X_2 = 10 + 0.3\varepsilon_1 + a\varepsilon_2 \rightarrow \text{Burada } a = \sqrt{1 - c^2}$$

$$X_3 = 10 + 0.3\varepsilon_1 + 0.5604a\varepsilon_2 + 0.8282a\varepsilon_3$$

$$X_4 = -8 + x_1 + 0.5x_2 + 0.3x_3 + 0.5\varepsilon_4$$

$$X_5 = -5 + 0.5x_1 + x_2 + 0.5\varepsilon_5$$

4.2. Benzetim Çalışması – 2

İkinci bir çalışma olarak altı bağımsız değişkenden oluşan bir veri seti kullanılmıştır. Burada Hibrit kovaryans yapılarıyla oluşturulmuş örnek bir veri seti incelenmiştir. Buna göre Z_1, Z_2, Z_3, Z_4, Z_5 ve Z_6 rasgele değişkenleri standart normal dağılımlı $N(0, 1)$ değişkenlerdir. Modelde kullanılmakta olan bağımsız X_1, X_2, X_3, X_4, X_5 ve X_6 değişkenler, standart normal dağılımlı Z rasgele değişkenlerine bağlı şekilde türetilmiştir. Simülasyon denkleminde yer alan " a " ise bağımsız değişkenler arasındaki korelasyonu kontrol eden korelasyon parametresi olup, $a_1 = 0.5$ ve $a_2 = 0.3$ olarak alınmıştır. Aşağıda Benzetim Çalışması – 2 için ifade edilen örnek bir veri seti yer almaktadır. Buna göre,

$$X_1 = \sqrt{1 - a_1^2}Z_1 + a_1Z_6$$

$$X_2 = \sqrt{1 - a_1^2}Z_2 + a_1Z_6$$

$$X_3 = \sqrt{1 - a_1^2}Z_3 + a_1Z_6$$

$$X_4 = \sqrt{1 - a_2^2}Z_4 + a_2Z_6$$

$$X_5 = \sqrt{1 - a_2^2}Z_5 + a_2Z_6$$

$$X_6 = \sqrt{1 - a_2^2}Z_6 + a_2Z_6$$

Hedefimiz çok değişkenli indirgenmiş rank regresyon yöntemi için veri elde ederek çok sayıda bağımlı ve bağımsız değişkenler ile çalışmaktır. Regresyon yönteminde ortaya çıkan problemleri, çok değişkenli regresyon yöntemi üzerinde giderebilmek amaçlanmıştır. Bunun için ise daraltılmış kovaryans tahmin edici

yöntemlerini de kullanarak β katsayıları üzerinden elde edilen hata kareler ortalamalarını elde ederek, gerçek β değerlerini tahmin etmeyi sağlamaktayız. Regresyon katsayıları üzerinden elde edilen sonuçlar ve tahminler, gerçek veriler üzerinde veya gerçek verilere benzeyen durumlar için değerlendirilmiştir. Çok değişkenli indirgenmiş rank regresyon yöntemiyle birlikte kullanılacak olan daraltılmış kovaryans tahmin edici yöntemleri, çoklu bağlantı ve düşük gözlem sayısına sahip veri için meydana gelen birtakım problemleri çözebilmeyi hedeflemektedir.

İlk önce Bozdoğan (2003) tarafından oluşturulan benzetim çalışmasını, çok değişkenli indirgenmiş rank regresyon yöntemi için inceledikten sonra ardından Hibrit kovaryans yapılarının kullanıldığı simülasyon çalışmasını inceleyeceğiz. Daha sonra simülasyon çalışmaları için ise gerekli gördüğümüz daraltılmış kovaryans tahmin edici yöntemlerinin, çok değişkenli indirgenmiş rank regresyon yöntemindeki durumuna bakacağız.

Aşağıda Tablo 4.1. olarak gösterilecek olan kısımda çok değişkenli indirgenmiş rank regresyon yöntemi için gerekli olan daraltılmış kovaryans tahmin edici yöntemlerine yer verilmiştir. Şöyle ki;

Tablo 4.1. Simülasyon çalışmasında yer alacak olan daraltılmış kovaryans tahmin edici yöntemleri

| | Daraltılmış kovaryans tahmin edici yöntemleri | Açıklama |
|---|---|---|
| 1 | MLE / EB | Maksimum Olabilirlik / Bayes Tahmin Edici Yöntemi |
| 2 | SRE | Şartlandırılmış Ridge(iz) Tahmin Edici Yöntemi |
| 3 | SDE | Şartlandırılmış Ridge(iz) diyagonal Kovaryans Tahmin Edici Yöntemi |
| 4 | BCSE | Bozdoğan tarafından önerilmiş Yakınsak Kovaryans Tahmin Edici Yöntemi |
| 5 | CSE | Dışbükey Toplam Kovaryans Tahmin Edicisi |
| 6 | OAS | Oracle Yaklaşım ile Daraltılmış Tahmin Edici Yöntemi |

Tablo 4.2. Gözlem sayılarının 100, 500 ve 1000 değerlerini aldığı bir veri analizi için korelasyon parametrelerinin de 0.3, 0.5, 0.7 ve 0.9 gibi katsayılara sahip olduğu bir durumda $r = 3$ için daraltılmış kovaryans tahmin edici yöntemlerinin etkisiyle β katsayıları üzerinden elde edilmiş olan hata kareler ortalama değerleri

| (n, σ, c) | $MSE_{MLE/EB}$ | MSE_{SRE} | MSE_{SDE} | MSE_{BCSE} | MSE_{CSE} | MSE_{OAS} |
|------------------|----------------|-------------|-------------|--------------|-------------|-------------|
| (100,1,0.3) | 0.1179 | 0.4647 | 3.3953 | 50.5519 | 14.4329 | 18.2232 |
| (100,1,0.5) | 0.1254 | 0.5116 | 2.6293 | 49.9606 | 12.9336 | 15.7639 |
| (100,1,0.7) | 0.1807 | 0.7425 | 2.4999 | 52.5500 | 11.3998 | 13.3405 |
| (100,1,0.9) | 0.3808 | 1.5117 | 2.5520 | 49.1622 | 8.0687 | 8.9795 |
| (500,1,0.3) | 0.0085 | 0.0232 | 0.1117 | 49.5935 | 4.9636 | 5.3516 |
| (500,1,0.5) | 0.0096 | 0.0266 | 0.0916 | 48.9482 | 3.8052 | 4.0873 |
| (500,1,0.7) | 0.0122 | 0.0369 | 0.0830 | 49.0433 | 2.3842 | 2.5265 |
| (500,1,0.9) | 0.0228 | 0.0863 | 0.0871 | 50.6522 | 1.2630 | 1.3128 |
| (1000,1,0.3) | 0.0038 | 0.0072 | 0.0307 | 48.2357 | 2.2972 | 2.4140 |
| (1000,1,0.5) | 0.0042 | 0.0090 | 0.0289 | 51.1932 | 1.8905 | 1.9785 |
| (1000,1,0.7) | 0.0048 | 0.0107 | 0.0225 | 47.591 | 0.8244 | 0.8545 |
| (1000,1,0.9) | 0.0083 | 0.0240 | 0.0193 | 46.2803 | 0.3581 | 0.3659 |

Yukarıdaki Tablo 4.2. de daraltılmış kovaryans tahmin edicilerin etkisiyle β katsayıları üzerinden elde edilmiş olan hata kareler ortalama değerleri verilmiştir. Buna göre Tablo 4.2. de ifade edilmiş olan daraltılmış kovaryans tahmin edici yöntemlerinden MLE / EB kovaryans tahmin edici yöntemi, genel itibariyle diğer daraltılmış kovaryans tahmin edici yöntemlerinden daha iyi olduğu görülmektedir. Gözlem sayısının arttığı durumlar incelendiğinde, gözlem sayısı ne kadar artarsa β katsayıları üzerinden elde edilen hata kareler ortalamalarının sıfıra doğru daha fazla yaklaştığı görülmüştür. Bunun anlamı ise tahmini β katsayılarının, gerçek β katsayılarına yakın değerlere sahip olacağını göstermektedir. Gözlem sayısının sabit ve korelasyon değerlerinin arttığı durumları incelediğimizde ise sıfıra daha yakın değerleri aldığını görmekteyiz. Bu değerlerden gözlem sayısının 1000 değerini aldığı ve korelasyon katsayısının da 0.3 gibi bir değere sahip olduğu durumda β katsayıları üzerinden elde edilmiş olan hata kareler ortalamasının ortalama olarak 0.0038 gibi bir değere sahip olduğu görülmüştür ve sonuç olarak hata kareler ortalamaları üzerinde MLE / EB daraltılmış kovaryans tahmin edici yönteminin daha etkili olduğu tespit

edilmiştir. Böylelikle daraltılmış kovaryans tahmin edici yöntemlerinden en etkili olan MLE / EB kovaryans tahmin edicisinin, çok değişkenli indirgenmiş rank regresyon yöntemi için daha etkili olduğuna karar verilmiştir. Tablo 4.2. ye baktığımızda BCSE daraltılmış kovaryans tahmin edici yönteminin, diğer kovaryans tahmin edici yöntemlerinden daha büyük değerlere sahip olduğu görülmektedir. BCSE daraltılmış kovaryans tahmin edici yönteminin bu kadar yüksek değere sahip olmasına sebep olan durum ceza katsayısının çok fazla artabileceği görüşü savunulabilir. Bu durumu daha net bir şekilde görebilmek adına daraltılmış kovaryans tahmin edici yöntemlerinin determinant değerlerine bakmamız gerekmektedir. Aşağıda daraltılmış kovaryans tahmin edici yöntemlerinin determinant katsayıları yer almaktadır.

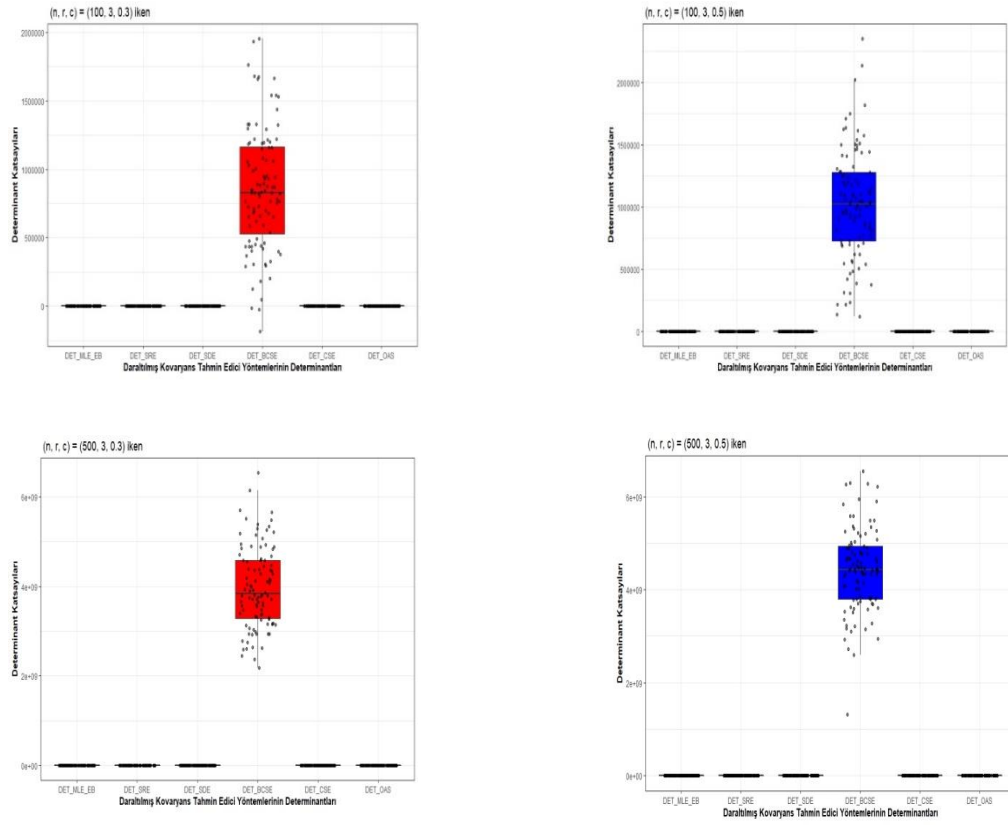
Tablo 4.3. Gözlem sayılarının 100, 500 ve 1000 değerlerini aldığı bir veri seti için korelasyon parametrelerinin de 0.3, 0.5, 0.7 ve 0.9 gibi katsayılara sahip olduğu bir durumda daraltılmış kovaryans tahmin edici yöntemlerinin ortalama yönünden determinant değerleri

| (n, σ, c) | MLE / EB_{DET} | SRE_{DET} | SDE_{DET} | $BCSE_{DET}$ | CSE_{DET} | OAS_{DET} |
|------------------|------------------|-------------|-------------|--------------|-------------|-------------|
| (100,1,0.3) | 0.0354 | 0.0426 | 0.0278 | 8.33E+05 | 0.8426 | 1.4971 |
| (100,1,0.5) | 0.0249 | 0.0304 | 0.0192 | 1.01E+06 | 0.4650 | 0.7722 |
| (100,1,0.7) | 0.0121 | 0.0155 | 0.0089 | 1.43E+06 | 0.1405 | 0.2010 |
| (100,1,0.9) | 0.0019 | 0.0029 | 0.0012 | 3.08E+06 | 0.0123 | 0.0150 |
| (500,1,0.3) | 0.0362 | 0.0376 | 0.0345 | 3.96E+09 | 0.1446 | 0.1569 |
| (500,1,0.5) | 0.0240 | 0.0250 | 0.0228 | 4.66E+09 | 0.0786 | 0.0837 |
| (500,1,0.7) | 0.0115 | 0.0122 | 0.0109 | 6.07E+09 | 0.0276 | 0.0287 |
| (500,1,0.9) | 0.0016 | 0.0018 | 0.0015 | 1.23E+10 | 0.0029 | 0.0030 |
| (1000,1,0.3) | 0.0352 | 0.0359 | 0.0344 | 1.33E+11 | 0.0818 | 0.0842 |
| (1000,1,0.5) | 0.0243 | 0.0248 | 0.0237 | 1.54E+11 | 0.0484 | 0.0494 |
| (1000,1,0.7) | 0.0113 | 0.0116 | 0.0110 | 2.05E+11 | 0.0184 | 0.0186 |
| (1000,1,0.9) | 0.0016 | 0.0016 | 0.0015 | 3.99E+11 | 0.0021 | 0.0021 |

Tablo 4.3. de daraltılmış kovaryans tahmin edici yöntemlerinin ortalama yönünden hesaplanmış determinant değerleri gösterilmiştir. Buna göre Tablo 4.3. de ifade edilmiş olan daraltılmış kovaryans tahmin edici yöntemlerinden BCSE kovaryans tahmin edici yönteminin determinant değeri, diğer kovaryans tahmin edici yöntemlerine göre büyük olduğu görülmektedir. Bu durum BCSE kovaryans tahmin

edicişinin, iz (*ridge*) cezasının fazla olduğunu göstermektedir. Yani çok deęişkenli indirgenmiş rank regresyon yönteminin uygulandıęı bir veri için veyahut dięer regresyon yöntemlerinde uygulanacak bir veride sorunların giderilmesinde BCSE daraltılmış kovaryans tahmin edici yönteminin kullanılmasının uygun olmadığı görölmektedir.

Bir dięer test edebileceğimiz durum ise daraltılmış kovaryans tahmin edici yöntemlerinden elde edilen determinant katsayılarının kutu grafiklerine bakmak olabilir. Aşağıda bu durumu gösteren kutu grafikleri yer almaktadır.



Şekil 4.1. $r = 3$ ve $c = (0.3$ ve $0.5)$ deęerleri için daraltılmış kovaryans tahmin edici yöntemlerinin ortalaması alınarak elde edilmiş olan determinant deęerlerinin kutu grafikleri

Şekil 4.1. de daraltılmış kovaryans tahmin edici yöntemlerinin ortalaması alınarak elde edilmiş olan determinant katsayılarının kutu grafikleri gösterilmiştir. Buna göre Şekil 4.1. de görölmekte olan kutu grafiğinde, daraltılmış kovaryans tahmin edici yöntemlerini incelediğimizde BCSE daraltılmış kovaryans tahmin edici yönteminin determinant deęerleri, dięer daraltılmış kovaryans tahmin edici yöntemlerinin determinant deęerlerinden daha büyük olduğu görölmektedir. Buna göre bu tahmin edici yönteminin uygun olmadığı yukarıdaki kutu grafiğinden de anlaşılmaktadır. Aynı şekilde, CSE ve OAS daraltılmış kovaryans tahmin edici

yöntemlerini incelediğimizde hata kareler ortalama değerleri, Tablo 4.2. de gösterildiği gibi yüksek olsa bile gözlem sayısının artması durumunda sıfıra yaklaştığını görmekteyiz. Bundan dolayıdır ki, bu daraltılmış kovaryans tahmin edici yöntemlerinin determinant değerleri incelendiğinde, çok değişkenli indirgenmiş rank regresyon yöntemi için problemlerin giderilmesinde uygulanabilir olduğu görülmektedir. Ancak yine de kesin bir öngöründe bulunmak doğru olmayabilir.

Tablo 4.4. Daraltılmış Kovaryans Tahmin Edici Yöntemleri kullanılarak $n = 100$, $r = 3$ ve $c = 0.3$ değerleri için bilgi kriterlerinin ortalama yönünden karşılaştırılması

| Daraltılmış kovaryans tahmin edici yöntemleri | AIC_q | AIC_{cq} | BIC_q | $HQIC_q$ | KIC_{vc} |
|---|----------|------------|----------|----------|------------|
| MLE/EB | 454.7612 | 850.9899 | 502.9163 | 456.3920 | 201.0094 |
| SRE | 607.8092 | 985.6721 | 655.9643 | 609.4400 | 197.9485 |
| SDE | 859.6169 | 1207.263 | 907.772 | 861.2477 | 192.9123 |
| BCSE | 2139.671 | 2333.71 | 2187.826 | 2141.302 | 167.3113 |
| CSE | 1567.422 | 1830.131 | 1615.577 | 2043.585 | 178.7562 |
| OAS | 1702.475 | 1948.978 | 1750.63 | 1704.106 | 176.0552 |

Tablo 4.5. Daraltılmış Kovaryans Tahmin Edici Yöntemleri kullanılarak $n = 100$, $r = 3$ ve $c = 0.5$ değerleri için bilgi kriterlerinin ortalama yönünden karşılaştırılması

| Daraltılmış kovaryans tahmin edici yöntemleri | AIC_q | AIC_{cq} | BIC_q | $HQIC_q$ | KIC_{vc} |
|---|----------|------------|----------|----------|------------|
| MLE/EB | 490.5136 | 882.4519 | 538.6687 | 492.1443 | 200.2944 |
| SRE | 650.4291 | 1023.178 | 698.5842 | 652.0599 | 197.0961 |
| SDE | 876.0626 | 1221.735 | 924.2177 | 877.6934 | 192.5834 |
| BCSE | 2158.993 | 2350.714 | 2207.148 | 2160.624 | 166.9248 |
| CSE | 1544.797 | 1810.221 | 1592.952 | 2014.908 | 179.2087 |
| OAS | 1653.768 | 1906.116 | 1701.923 | 1655.399 | 177.0293 |

Tablo 4.4. ve Tablo 4.5. de daraltılmış kovaryans tahmin edici yöntemleri kullanılarak elde edilmiş bilgi kriterlerinin ortalama yönünden karşılaştırılması gösterilmiştir. Buna göre bilgi kriterleri yönünden Tablo 4.4. ve Tablo 4.5. teki

değerlere baktığımızda, $n = 100$ ve korelasyon parametrelerinin $c = 0.3, 0.5$ olduğu durumlar için çok değişkenli indirgenmiş rank regresyon yöntemi için kullandığımız AIC_q bilgi kriteri, Kulback – Leibler tutarsızlığının göz önünde bulundurulduğu BedrickandTsai (1994) tarafından elde edilmiş olan AIC_{cq} bilgi kriteri ve BIC bilgi kriteri ile $HQIC_q$ bilgi kriteri daraltılmış kovaryans tahmin edici yöntemlerinden MLE / EB daraltılmış kovaryans tahmin edici yöntemini, en iyi kovaryans tahmin edicisi olarak gösterdiği tespit edilmiştir. Bu bağlamda bir diğer bilgi kriteri olan Kulback’ın yansız tahmin edicisi olarak bilindiği, çok değişkenli regresyon modelleri için yanlış bir tahmin yöntemi olarak elde edilmiş olan, aynı zamanda “vektör düzeltmesi” anlamına da gelen KIC_{vc} bilgi kriteri, diğer bilgi kriterlerinden ayrılarak OAS daraltılmış kovaryans tahmin edici yöntemini, en iyi kovaryans tahmin edici yöntem olarak görmüştür. Burada KIC_{vc} bilgi kriterinin hesaplama yönünden daraltılmış kovaryans tahmin edici yöntemlerini iyi yansıtamadığı görülmektedir. Bilgi kriterlerinin, daraltılmış kovaryans tahmin edici yöntemlerini tahmin etme yönünden sıralamaya konulduğunda gözlem sayısının 100, rank değerinin 3 ve korelasyon katsayılarının 0.3 ve 0.5 olduğu durumlarda ilk sırada AIC_q bilgi kriteri yer alırken son sırada ise AIC_{cq} bilgi kriterinin yer aldığı görülmektedir.

Tablo 4.6. Daraltılmış Kovaryans Tahmin Edici Yöntemleri kullanılarak $n = 500, r = 3$ ve $c = 0.3$ değerleri için bilgi kriterlerinin ortalama yönünden karşılaştırılması

| Daraltılmış kovaryans tahmin edici yöntemleri | AIC_q | AIC_{cq} | BIC_q | $HQIC_q$ | KIC_{vc} |
|--|---------------------------|------------------------------|---------------------------|----------------------------|------------------------------|
| MLE/EB | 960.5577 | 3379.6643 | 1056.9959 | 980.1719 | 189.0026 |
| SRE | 1383.4103 | 3792.3685 | 1479.8485 | 1403.0245 | 187.3112 |
| SDE | 2208.0635 | 4597.2300 | 2304.5017 | 2227.6776 | 184.0126 |
| BCSE | 10638.3199 | 12825.1603 | 10734.7582 | 10657.9341 | 150.2915 |
| CSE | 5408.6970 | 7721.0480 | 5505.1350 | 8255.8270 | 171.2100 |
| OAS | 5535.1861 | 7844.5017 | 5631.6244 | 5554.8003 | 170.7041 |

Tablo 4.7. Daraltılmış Kovaryans Tahmin Edici Yöntemleri kullanılarak $n = 500$, $r = 3$ ve $c = 0.5$ değerleri için bilgi kriterlerinin ortalama yönünden karşılaştırılması

| Daraltılmış kovaryans tahmin edici yöntemleri | AIC_q | AIC_{cq} | BIC_q | $HQIC_q$ | KIC_{vc} |
|---|------------|------------|------------|------------|------------|
| MLE/EB | 973.0498 | 3391.8566 | 1069.4881 | 992.6640 | 188.9526 |
| SRE | 1452.5642 | 3859.8627 | 1549.0024 | 1472.1784 | 187.0346 |
| SDE | 2296.5294 | 4683.5727 | 2392.9676 | 2316.1435 | 183.6587 |
| BCSE | 10791.1163 | 12974.2896 | 10887.5546 | 10810.7305 | 149.6803 |
| CSE | 5219.9834 | 7536.8638 | 5316.4216 | 8115.0545 | 171.9649 |
| OAS | 5328.9975 | 7643.2616 | 5425.4357 | 5348.6117 | 171.5288 |

Tablo 4.6. ve Tablo 4.7. de gözlem sayısının 500 değerine sahip olduğu durumların incelenmesiyle oluşturulmuş ve daraltılmış kovaryans tahmin edici yöntemlerinin de etkisiyle elde edilmiş bilgi kriterlerinin ortalama yönünden karşılaştırılması gösterilmiştir. Tablo 4.6. ve Tablo 4.7. de görüldüğü üzere, gözlem sayısının 500 olduğu ve korelasyon katsayısının arttığı bir durumda bilgi kriterleri yönünden, MLE / EB daraltılmış kovaryans tahmin edici yönteminin, β katsayıları üzerinden elde edilmiş olan hata kareler ortalaması üzerinde daha etkili olduğu görülmektedir. Bu bağlamda KIC_{vc} bilgi kriteri OAS daraltılmış kovaryans tahmin edici yöntemini daha etkili gösterdiğinden diğer daraltılmış kovaryans tahmin edici yöntemlerinden ayrılmaktadır. Bu bakımdan “vektör düzeltmesi” anlamındaki KIC_{vc} bilgi kriteri β katsayıları üzerindeki hata kareler ortalamasını iyi yansıtamadığı görülmüştür.

Tablo 4.8. Gözlem sayılarının 100, 500 ve 1000 değerlerinin alındığı bir veri için korelasyon parametrelerinin de 0.3, 0.5, 0.7 ve 0.9 olduğu bir durumda eğitim – test modelinde $r = 3$ için β katsayıları üzerinden elde edilen hata kareler ortalama değerleri

| (n, σ, c) | $MSE_{MLE/EB}^*$ | MSE_{SRE}^* | MSE_{SDE}^* | MSE_{BCSE}^* | MSE_{CSE}^* | MSE_{OAS}^* |
|------------------|------------------|---------------|---------------|----------------|---------------|---------------|
| (100,1,0.3) | 34.5198 | 35.8856 | 53.8728 | 155.2126 | 38.7072 | 50.8593 |
| (100,1,0.5) | 32.7538 | 34.0875 | 50.9575 | 147.1144 | 35.5303 | 40.5925 |
| (100,1,0.7) | 34.3150 | 36.1170 | 51.1596 | 155.7021 | 36.2963 | 36.7788 |
| (100,1,0.9) | 34.3320 | 37.3347 | 52.9480 | 152.6841 | 39.2014 | 39.5739 |

| | | | | | | |
|--------------|---------|---------|---------|----------|---------|---------|
| (500,1,0.3) | 29.6000 | 29.8092 | 32.4019 | 145.1783 | 33.8699 | 34.0059 |
| (500,1,0.5) | 32.0092 | 32.2795 | 34.8816 | 156.8340 | 35.7521 | 35.8793 |
| (500,1,0.7) | 33.0158 | 33.4077 | 35.4725 | 161.7422 | 35.8284 | 35.9161 |
| (500,1,0.9) | 32.5493 | 33.1018 | 35.3017 | 159.2844 | 33.8950 | 33.9386 |
| (1000,1,0.3) | 32.7235 | 32.8697 | 34.2344 | 162.0628 | 35.9285 | 35.9983 |
| (1000,1,0.5) | 31.5437 | 31.7031 | 32.8174 | 156.1685 | 34.7023 | 34.7878 |
| (1000,1,0.7) | 31.6930 | 31.8337 | 33.0368 | 156.9041 | 32.5644 | 32.5889 |
| (1000,1,0.9) | 33.9671 | 34.3211 | 35.1640 | 167.8635 | 35.2002 | 35.2190 |

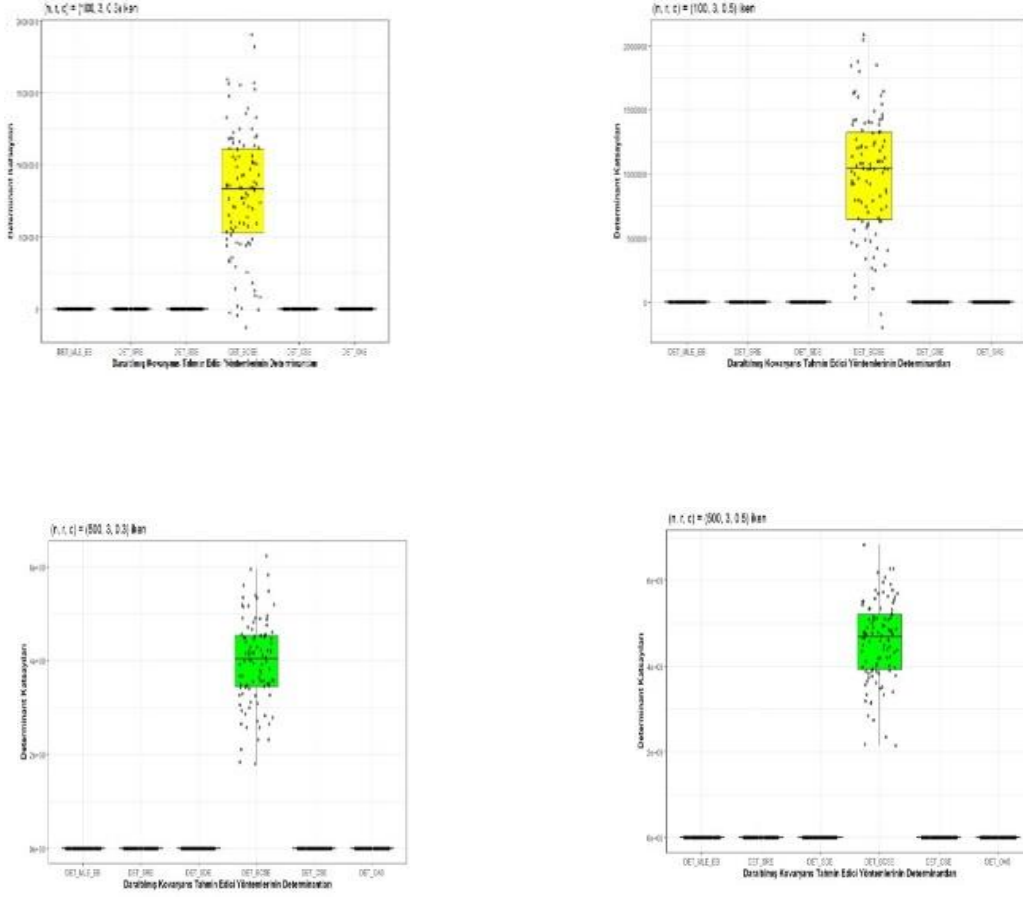
Tablo 4.8. de eğitim – test yöntemiyle elde edilmiş β katsayıları üzerindeki hata kareler ortalama değerleri gösterilmiştir. Burada Tablo 4.8. de belirtildiği gibi çok değişkenli indirgenmiş rank regresyon yöntemi için oluşturulmuş olan eğitim – test modelinin tahmin edilen β katsayıları üzerindeki hata kareler ortalamasıdır. Tablodaki değerler karekökü alınarak hesaplanmıştır. Buna göre rank değerinin $r = 3$ olarak belirlendiği çok değişkenli indirgenmiş rank regresyon yöntemi için $c = 0.3$ değerini aldığı ve gözlem sayısının da $n = 100, 500$ ve 1000 olarak belirlendiği durumlar için, daraltılmış kovaryans tahmin edici yöntemlerinden MLE / EB tahmin edicinin hata kareler ortalaması üzerinde daha etkili olduğu görülmektedir. Bu bağlamda şunu da belirtmek gerekir ki, $n = 500$ ve $n = 1000$ gözlem sayıları için MLE / EB daraltılmış kovaryans tahmin edicisinden sonra SRE daraltılmış kovaryans tahmin edici yönteminin, hata kareler ortalaması üzerinde etkili olduğu görülmüştür. Diğer durumları da incelediğimizde yine MLE / EB daraltılmış kovaryans tahmin edici yönteminin ve sonrasında ise yine SRE kovaryans tahmin edici yönteminin de hata kareler ortalaması üzerinde daha etkili olduğu görülmektedir.

Tablo 4.9. Gözlem sayılarının 100, 500 ve 1000 değerlerini aldığı bir veri seti için korelasyon parametrelerinin de 0.3, 0.5, 0.7 ve 0.9 gibi katsayılara sahip olduğu bir durumda $r = 3$ için eğitim – test modeli üzerinde uygulanmış olan daraltılmış kovaryans tahmin edici yöntemlerinin ortalama yönünden determinant değerleri

| (n, σ, c) | MLE / EB_{DET}^* | SRE_{DET}^* | SDE_{DET}^* | $BCSE_{DET}^*$ | CSE_{DET}^* | OAS_{DET}^* |
|------------------|--------------------|---------------|---------------|----------------|---------------|---------------|
| (100,1,0.3) | 0.0362 | 0.0435 | 0.0285 | 7.36E+05 | 0.8826 | 1.6025 |
| (100,1,0.5) | 0.0232 | 0.0285 | 0.0178 | 1.01E+06 | 0.4413 | 0.7391 |
| (100,1,0.7) | 0.0114 | 0.0147 | 0.0083 | 1.44E+06 | 0.1366 | 0.1992 |

| | | | | | | |
|--------------|--------|--------|--------|----------|--------|--------|
| (100,1,0.9) | 0.0018 | 0.0028 | 0.0012 | 3.19E+06 | 0.0137 | 0.0167 |
| (500,1,0.3) | 0.0348 | 0.0362 | 0.0332 | 3.86E+09 | 0.1411 | 0.1534 |
| (500,1,0.5) | 0.0232 | 0.0242 | 0.022 | 4.51E+09 | 0.0769 | 0.0819 |
| (500,1,0.7) | 0.0114 | 0.012 | 0.0107 | 6.20E+09 | 0.0269 | 0.0278 |
| (500,1,0.9) | 0.0017 | 0.0018 | 0.0015 | 1.22E+10 | 0.0029 | 0.0030 |
| (1000,1,0.3) | 0.0359 | 0.0366 | 0.035 | 1.28E+11 | 0.0836 | 0.0861 |
| (1000,1,0.5) | 0.0238 | 0.0243 | 0.0232 | 1.48E+11 | 0.0476 | 0.0486 |
| (1000,1,0.7) | 0.0113 | 0.0116 | 0.011 | 2.05E+11 | 0.0184 | 0.0186 |
| (1000,1,0.9) | 0.0016 | 0.0017 | 0.0015 | 3.98E+11 | 0.0022 | 0.0022 |

Tablo 4.9. da eğitim – test yöntemiyle elde edilmiş olan daraltılmış kovaryans tahmin edici yöntemlerinin ortalama yönünden determinant değerleri gösterilmiştir. Buna göre Tablo 4.9 da eğitim – test modeli için daraltılmış kovaryans tahmin edici yöntemlerinden elde edilmiş olan determinant değerlerinin ortalamasıdır. Tabloya bakıldığında daraltılmış kovaryans tahmin edici yöntemlerinden biri olan BCSE kovaryans tahmin edicisinin ortalama yönünden determinant değerlerinin çok yüksek olduğu görülmektedir. Bu bağlamda BCSE daraltılmış kovaryans tahmin edicisinin meydana getirmiş olduğu iz (*ridge*) cezasının yüksek olduğu anlaşılabilmektedir. Çok değişkenli indirgenmiş rank regresyon yöntemi veya diğer regresyon yöntemlerinde problemlerin giderilebilmesi için hem eğitim – test yönteminde hem de eğitim – test yöntemi olmadan yapılan analizler sonucunda görülmüştür ki, rank değerinin bilindiği durumlarda ($r = 3$ gibi) BCSE daraltılmış kovaryans tahmin edici yönteminin, çok değişkenli indirgenmiş rank regresyon yöntemi için uygun olmadığı kararla varılmıştır. Alternatif olarak daha kapsamlı bir şekilde görebilmek için eğitim – test modelinde kullanılan daraltılmış kovaryans tahmin edici yöntemlerinin determinant katsayılarının elde edildiği kutu grafiklerine baktığımızda,



Şekil 4.2. $r = 3$, $c = (0.3 \text{ ve } 0.5)$ değerleri için eğitim – test modeli üzerinden daraltılmış kovaryans tahmin edici yöntemlerinin ortalaması alınarak elde edilmiş olan determinant değerlerinin kutu grafikleri

Şekil 4.2.'deki eğitim – test yöntemi için daraltılmış kovaryans tahmin edici yöntemlerinin ortalaması alınarak elde edilmiş determinant değerlerinin kutu grafikleri gösterilmiştir. Buna göre Şekil 4.2. de eğitim – test yöntem için elde edilen kutu grafiğini incelediğimizde, çok değişkenli indirgenmiş rank regresyon yöntemi için daraltılmış kovaryans tahmin edici yöntemi olan BCSE tahmin edicisinin, determinant katsayısının aşırı fazla olduğu görülmektedir. Burada da aynı sonucu veren daraltılmış kovaryans tahmin edicisinin çok değişkenli indirgenmiş rank regresyon yöntemlerinde ve diğer regresyon yöntemleri için uygun olmadığı görülmektedir. Tahmini β katsayılarının sonucunda elde edilmiş olan hata kareler ortalamasında BCSE daraltılmış kovaryans tahmin edicisinin, büyük değer almasını hem eğitim – test yöntemi kullanılarak hem de eğitim – test olmadan yapılan analizler sonucunda görülmüştür. Bu bakımdan daraltılmış kovaryans tahmin edici yöntemlerinin determinant değerleri de incelendiğinde iz (*ridge*) cezasının diğer tahmin edici

yöntemlerinin hesaplanan değerlerinden çok fazla yüksek çıkmasından dolayı uygun olmadığı görülmüştür.

Böylelikle Bozdoğan (2003) tarafından oluşturulmuş Uygulama 1 simülasyon çalışması rankın 3 olduğu durumlar için tamamlanmıştır. Bundan sonra ise rank değerinin 4 olduğu durumlar için incelenecektir.

Tablo 4.10. Gözlem sayılarının 100, 500 ve 1000 değerlerini aldığı bir veri analizi için korelasyon parametrelerinin de 0.3, 0.5, 0.7 ve 0.9 gibi katsayılara sahip olduğu bir durumda $r = 4$ için daraltılmış kovaryans tahmin edici yöntemlerinin etkisiyle β katsayıları üzerinden elde edilmiş olan hata kareler ortalama değerleri

| (n, σ, c) | $MSE_{MLE / EB}$ | MSE_{SRE} | MSE_{SDE} | MSE_{BCSE} | MSE_{CSE} | MSE_{OAS} |
|------------------|------------------|-------------|-------------|--------------|-------------|-------------|
| (100,1,0.3) | 0.1451 | 0.5715 | 3.8232 | 66.0569 | 18.1476 | 22.9075 |
| (100,1,0.5) | 0.1632 | 0.6345 | 3.8000 | 69.2733 | 16.1552 | 19.8317 |
| (100,1,0.7) | 0.2457 | 0.9900 | 3.3523 | 66.4585 | 14.2141 | 16.7133 |
| (100,1,0.9) | 0.4759 | 1.9224 | 3.6145 | 65.1925 | 10.5854 | 11.7884 |
| (500,1,0.3) | 0.0119 | 0.0307 | 0.1614 | 65.8004 | 6.0531 | 6.5379 |
| (500,1,0.5) | 0.0128 | 0.0355 | 0.1333 | 65.0987 | 4.9158 | 5.2712 |
| (500,1,0.7) | 0.0157 | 0.0498 | 0.1078 | 69.8786 | 3.1756 | 3.3731 |
| (500,1,0.9) | 0.0296 | 0.1113 | 0.1293 | 65.3503 | 1.5767 | 1.6385 |
| (1000,1,0.3) | 0.0050 | 0.0101 | 0.0375 | 66.8401 | 2.9399 | 3.0920 |
| (1000,1,0.5) | 0.0051 | 0.0108 | 0.0341 | 64.3797 | 2.3554 | 2.4637 |
| (1000,1,0.7) | 0.0064 | 0.0163 | 0.0305 | 69.9956 | 1.4009 | 1.4522 |
| (1000,1,0.9) | 0.0104 | 0.0314 | 0.0295 | 65.3850 | 0.5056 | 0.5167 |

Yukarıdaki Tablo 4.10. da $r = 4$ için daraltılmış kovaryans tahmin edicilerin etkisiyle β katsayıları üzerinden elde edilmiş olan hata kareler ortalama değerleri verilmiştir. Buna göre Tablo 4.10'da gösterilen tablo incelendiğimizde rank değerinin arttığı bir durumda yani ($r = 4$) için daraltılmış kovaryans tahmin edici yöntemlerinden olan MLE / EB tahmin edicisinin β katsayıları üzerinden elde edilmiş hata kareler ortalamaları üzerinde daha etkili olduğu görülmektedir. Tablo değerlerine bakıldığında, sadece gözlem sayısının arttığı bir durumda dahi MLE / EB kovaryans tahmin edici yönteminin hata kareler ortalamaları üzerinde daha etkili olduğu görülmüştür. Gözlem sayısının artması durumunda, MLE / EB daraltılmış kovaryans

tahmin edici yönteminin sifira daha çok yaklaştığı görülmektedir. Bu durum ise tahmin edilen β katsayılarının gerçek β katsayılarına yaklaşık bir değerle tahminde bulunduğunu açıklamaktadır.

Gözlem sayısının sabit olduğunu varsaydığımızda, korelasyon katsayılarının arttığı bir durumu inceleyelim. Buna göre eğer gözlem sayısı sabit ve korelasyon değerleri artarsa aynı şekilde MLE / EB daraltılmış kovaryans tahmin edici yönteminin hata kareler ortalaması üzerinde daha etkili olduğunu görebilmekteyiz. Bir kıyaslama yapıldığında ise verinin, $n = 1000$ olduğu durum incelendiğinde, MLE / EB daraltılmış kovaryans tahmin edici yönteminin sifira daha yakın olduğu görülmektedir. Çok değişkenli indirgenmiş rank regresyon yöntemi için en uygun analiz ise, tablo 4.10. a göre gözlem sayısının 1000 değerinin aldığı, rankın 4 ve korelasyon katsayısının da 0.3 gibi bir değere sahip olduğu bir durum olarak gözükmektedir. Bu tablo değerine göre, MLE / EB daraltılmış kovaryans tahmin edici yönteminin β katsayıları üzerinden elde edilen hata kareler ortalaması için daha etkili olduğu görülmektedir.

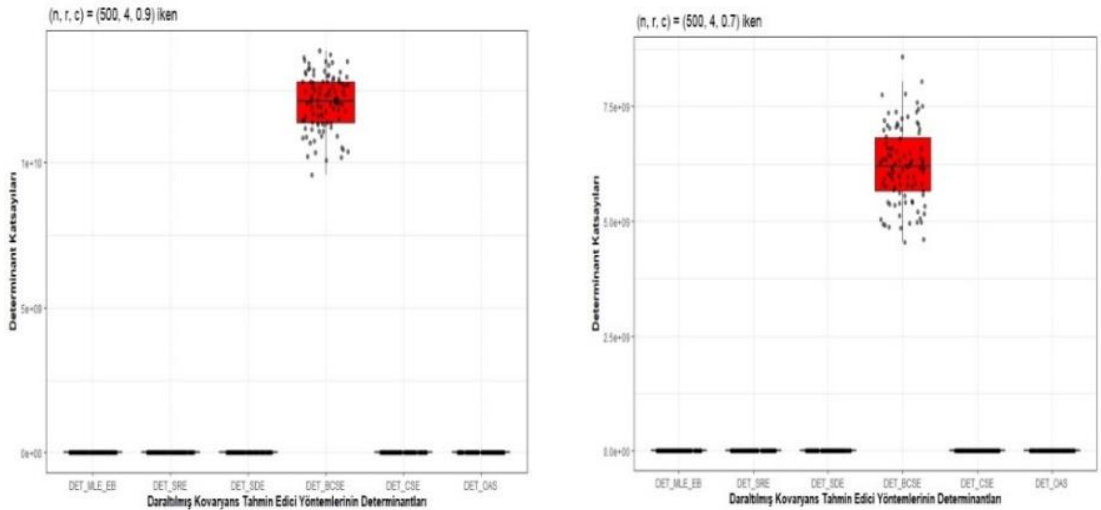
Tablo 4.10. da daraltılmış kovaryans tahmin edici yöntemleri için bir sıralamada bulunduğumuz da sırasıyla, MLE / EB kovaryans tahmin edicisinin, SRE kovaryans tahmin edicisinin ve SDE kovaryans tahmin edicisinin β katsayıları üzerinden elde edilen hata kareler ortalaması için daha iyi ve daha etkili olduğunu söyleyebiliriz. Bu bağlamda tabloda da görüldüğü üzere, daraltılmış kovaryans tahmin edici yöntemlerinden biri olan BCSE kovaryans tahmin edicisinin, hata kareler ortalaması değerleri, rankın arttığı bir durumda da diğer kovaryans tahmin edici yöntemlerinden yüksek olduğu görülmektedir. BCSE daraltılmış kovaryans tahmin edici yönteminin diğer kovaryans tahmin edici yöntemlerinden bu derece farklı olmasının nedenlerinden biri ise aynı şekilde iz (*ridge*) cezasının fazla olmasından kaynaklanmaktadır. Daha net bir şekilde görebilmek adına daraltılmış kovaryans tahmin edici yöntemlerinin determinant değerlerini incelediğimizde,

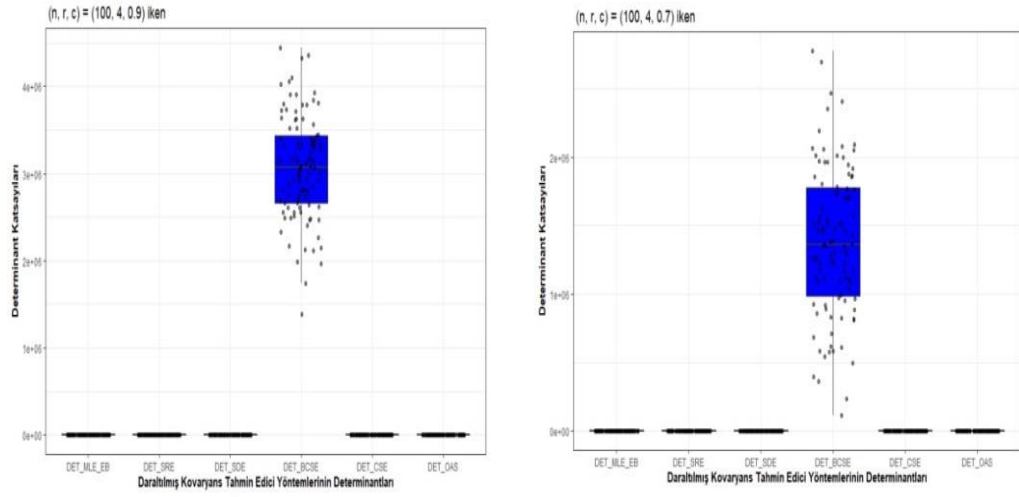
Tablo 4.11. Gözlem sayısının 100, 500 ve 1000 değerlerini alarak oluşturulmuş bir veri için korelasyon parametrelerinin de 0.3, 0.5, 0.7 ve 0.9 değerlerini aldığı bir durumda daraltılmış kovaryans tahmin edici yöntemlerinin $r=4$ için ortalama yönünden determinant değerleri

| (n, σ, c) | MLE / EB_{DET} | SRE_{DET} | SDE_{DET} | $BCSE_{DET}$ | CSE_{DET} | OAS_{DET} |
|------------------|------------------|-------------|-------------|--------------|-------------|-------------|
| (100,1,0.3) | 0.0363 | 0.0436 | 0.0285 | 7.82E+05 | 0.8777 | 1.5920 |
| (100,1,0.5) | 0.0251 | 0.0308 | 0.0194 | 9.17E+05 | 0.4728 | 0.7960 |

| | | | | | | |
|--------------|----------|----------|----------|----------|----------|----------|
| (100,1,0.7) | 0.0115 | 0.0148 | 0.0084 | 1.41E+06 | 0.1347 | 0.1941 |
| (100,1,0.9) | 0.0019 | 0.0029 | 0.0012 | 3.08E+06 | 0.0127 | 0.0153 |
| (500,1,0.3) | 0.0350 | 0.0364 | 0.0334 | 3.93E+09 | 0.1420 | 0.1545 |
| (500,1,0.5) | 0.0239 | 0.0250 | 0.0227 | 4.73E+09 | 0.0780 | 0.0830 |
| (500,1,0.7) | 1.15E+04 | 1.21E+04 | 1.08E+04 | 6.09E+15 | 2.70E+04 | 2.79E+04 |
| (500,1,0.9) | 0.0016 | 0.0018 | 0.0015 | 1.21E+10 | 0.0029 | 0.0029 |
| (1000,1,0.3) | 0.0349 | 0.0355 | 0.0341 | 1.34E+11 | 0.0809 | 0.0832 |
| (1000,1,0.5) | 0.0248 | 0.0253 | 0.0242 | 1.52E+11 | 0.0497 | 0.0507 |
| (1000,1,0.7) | 0.0113 | 0.0116 | 0.0109 | 2.00E+11 | 0.0182 | 0.0184 |
| (1000,1,0.9) | 0.0016 | 0.0016 | 0.0015 | 4.05E+11 | 0.0021 | 0.0021 |

Tablo 4.11. de $r = 4$ için daraltılmış kovaryans tahmin edici yöntemlerinin ortalama yönünden hesaplanmış determinant değerleri gösterilmiştir. Buna göre Tablo 4.11. de ifade edilen tahmin ediciler incelendiğinde BCSE daraltılmış kovaryans tahmin edici yönteminin determinant değerlerinin diğer kovaryans tahmin edici yöntemlerinden çok yüksek olduğu görülmektedir. Bu durum ise BCSE tahmin edici yönteminin, çok değişkenli indirgenmiş rank regresyon yöntemi için uygun olmadığına karar verilmiştir. Tablo 4.10. da hesaplanan BCSE daraltılmış kovaryans tahmin edici yönteminin bu kadar yüksek çıkmasına sebep iz (*ridge*) cezasının fazla olmasından kaynaklanmaktadır. Bunun sebeplerini görebilmek adına daraltılmış kovaryans tahmin edici yöntemlerinin ortalama bakımından determinant değerlerine bakılmıştır.





Şekil 4.3. $r = 4$, $c = (0.7 \text{ ve } 0.9)$ değerleri için daraltılmış kovaryans tahmin edici yöntemlerinin ortalaması alınarak elde edilmiş olan determinant değerlerinin kutu grafikleri

Şekil 4.3. de $r = 4$ için daraltılmış kovaryans tahmin edici yöntemlerin ortalaması alınarak elde edilmiş olan determinant değerlerinin kutu grafikleri gösterilmiştir. Yukarıdaki kutu grafiği incelendiğinde BCSE daraltılmış kovaryans tahmin edici yönteminin iz (*ridge*) cezasının fazla olduğu görülmektedir. Yani çok değişkenli indirgenmiş rank regresyon yöntemlerinin sorunlarının giderilmesinde BCSE daraltılmış kovaryans tahmin edici yönteminin hem $r = 3$ hem de $r = 4$ durumlarında uygun olmadığı görülmektedir. Aynı şekilde gözlem sayısının artmasında veya gözlem sayısı sabitken korelasyon değerleri için incelendiğinde, BCSE daraltılmış kovaryans tahmin edici yönteminin her iki durum için de uygun olmadığına karar verilmiştir.

Tablo 4.12. Daraltılmış Kovaryans Tahmin Edici Yöntemleri kullanılarak $n = 1000$, $r = 4$ ve $c = 0.5$ değerleri için bilgi kriterlerinin ortalama yönünden karşılaştırılması

| Daraltılmış kovaryans tahmin edici yöntemleri | AIC_q | AIC_{cq} | BIC_q | $HQIC_q$ | KIC_{vc} |
|---|-----------|------------|-----------|------------|------------|
| MLE/EB | 1490.0252 | 6413.2249 | 1607.2578 | 1515.9838 | 187.9214 |
| SRE | 2102.8652 | 7018.7109 | 2220.0979 | 2128.8239 | 186.6957 |
| SDE | 3521.2548 | 8420.0797 | 3638.4874 | 3547.2135 | 183.8589 |
| BCSE | 25549.707 | 30184.191 | 25666.94 | 25575.666 | 139.802 |
| CSE | 8971.0445 | 13804.4719 | 9088.2771 | 14916.9372 | 172.9594 |

| | | | | | |
|-----|-----------|------------|-----------|-----------|----------|
| OAS | 9086.3945 | 13918.4377 | 9203.6271 | 9112.3532 | 172.7287 |
|-----|-----------|------------|-----------|-----------|----------|

Tablo 4.13. Daraltılmış Kovaryans Tahmin Edici Yöntemleri kullanılarak $n = 1000, r = 4$ ve $c = 0.9$ değerleri için bilgi kriterlerinin ortalama yönünden karşılaştırılması

| Daraltılmış kovaryans tahmin edici yöntemleri | AIC_q | AIC_{cq} | BIC_q | $HQIC_q$ | KIC_{vc} |
|---|----------|------------|----------|----------|------------|
| MLE/EB | 2009.244 | 6926.213 | 2126.477 | 2035.203 | 186.883 |
| SRE | 2862.373 | 7769.104 | 2979.606 | 2888.332 | 185.1767 |
| SDE | 3719.099 | 8615.55 | 3836.332 | 3745.058 | 183.4633 |
| BCSE | 23262.99 | 27924.91 | 23380.22 | 23288.94 | 144.3755 |
| CSE | 5872.877 | 10743.48 | 5990.109 | 12037.38 | 179.1557 |
| OAS | 5905.305 | 10775.52 | 6022.538 | 5931.264 | 179.0908 |

Tablo 4.12 ve Tablo 4.13 de gözlem değerlerinin 1000 değerinin olduğu ve $r = 4$ için daraltılmış kovaryans tahmin ediciler kullanılarak elde edilmiş olan bilgi kriterleri gösterilmiştir. Buna göre Tablo 4.12. ve Tablo 4.13.'de görüldüğü gibi bilgi kriterleri, daraltılmış kovaryans tahmin edici yöntemlerinden MLE / EB daraltılmış kovaryans tahmin edici yöntemini, β katsayıları üzerinden elde edilen hata kareler ortalamaları için daha etkili olduğunu matematiksel olarak ifade etmektedir. Ancak burada şunu belirtmek gerekir ki, bilgi kriterlerinden yalnızca bir tanesi yani Kulback – Leibler'in ortaya koymuş olduğu bilgi kriteri yani (KIC_{vc}) diğer bilgi kriterlerinin aksine OAS bilgi kriterinin daha etkili olduğunu göstermektedir. Bu durum ise KIC_{vc} bilgi kriterinin rankın 4, korelasyon katsayılarının da 0.5 ve 0.9 olduğu bir durum da dahi kötü bir sonuca vardığını göstermiş yani bilgi kriterinin iyi bir bilgi kriteri olmadığına analizler sonucunda karar verilmiştir. Bu sonuçlara göre, gözlem sayısının değerleri fark etmeksizin veyahut korelasyon değerlerinin artması yada azalması durumunda görülmüştür ki, hem β katsayıları üzerinden elde edilen hata kareler ortalama değerleri hem de bilgi kriterlerinin ortaya koymuş olduğu değerler için daraltılmış kovaryans tahmin edici yöntemlerinden olan Maksimum Olabilirlik / Bayes Tahmin Edicisi yani kısaca (MLE / EB) yöntemi, çok değişkenli indirgenmiş rank regresyon yöntemi için en uygun tahmin edici olarak görülmektedir. Yapılan analizler

sonucunda, MLE / EB daraltılmış kovaryans tahmin edici yöntemi, tahmin edilen β katsayılarının gerçek β değerlerini yaklaşık olarak daha iyi tahmin ettiğine karar verilmiştir.

Tablo 4.14. Gözlem sayılarının 100, 500 ve 1000 değerlerinin alındığı bir veri için korelasyon parametrelerinin de 0.3, 0.5, 0.7 ve 0.9 olduğu bir durumda eğitim – test modelinde $r = 4$ için β katsayıları üzerinden elde edilen hata kareler ortalama değerleri

| (n, σ, c) | $MSE_{MLE/EB}^*$ | MSE_{SRE}^* | MSE_{SDE}^* | MSE_{BCSE}^* | MSE_{CSE}^* | MSE_{OAS}^* |
|------------------|------------------|---------------|---------------|----------------|---------------|---------------|
| (100,1,0.3) | 38.9154 | 40.2559 | 60.7347 | 174.8305 | 42.9527 | 64.7815 |
| (100,1,0.5) | 38.5661 | 40.4202 | 57.6645 | 173.1380 | 43.3012 | 51.5421 |
| (100,1,0.7) | 38.2805 | 40.3415 | 56.0614 | 186.2858 | 42.1578 | 43.5712 |
| (100,1,0.9) | 39.6396 | 43.0351 | 60.0229 | 177.0732 | 44.7777 | 45.2944 |
| (500,1,0.3) | 37.9808 | 38.2828 | 41.6516 | 185.6595 | 43.0070 | 43.1097 |
| (500,1,0.5) | 37.9867 | 38.3227 | 41.4173 | 185.7704 | 42.2009 | 42.3208 |
| (500,1,0.7) | 38.0128 | 38.4223 | 41.1333 | 185.9782 | 40.8563 | 40.9586 |
| (500,1,0.9) | 38.1650 | 38.8221 | 41.6055 | 186.3208 | 39.8148 | 39.8692 |
| (1000,1,0.3) | 36.8494 | 36.9811 | 38.5646 | 182.4803 | 39.7489 | 39.8304 |
| (1000,1,0.5) | 36.8404 | 36.9898 | 38.3987 | 182.5014 | 39.1018 | 39.1661 |
| (1000,1,0.7) | 36.8377 | 37.0284 | 38.1732 | 182.5328 | 38.2489 | 38.2852 |
| (1000,1,0.9) | 36.9165 | 37.2762 | 38.2144 | 182.5667 | 38.0133 | 38.0312 |

Tablo 4.14. de $r = 4$ için eğitim – test yöntemi kullanılarak β katsayıları üzerinden elde edilmiş olan hata kareler ortalama değerleri ifade edilmiştir. Buna göre eğitim – test yöntemi için oluşturulmuş olan daraltılmış kovaryans tahmin edici yöntemlerinin etkisiyle, yukarıdaki Tablo 4.14. de görüldüğü üzere, β katsayıları üzerinden elde edilmiş olan hata kareler ortalama değerleri için MLE / EB daraltılmış kovaryans tahmin edici yönteminin daha etkili olduğu görülmektedir. Bu durum ise MLE / EB daraltılmış kovaryans tahmin edici yönteminin, çok değişkenli indirgenmiş rank regresyon yönteminde ve eğitim – test modeli için oluşturulmuş olan çok değişkenli indirgenmiş rank regresyon yöntemi için β katsayıları üzerinden elde edilen hata kareler ortalama değerlerinin daha etkili olduğunu göstermektedir. Yani; tahmin edilen β değerlerinin, gerçek β değerlerini yaklaşık olarak sifıra yakın olacak şekilde

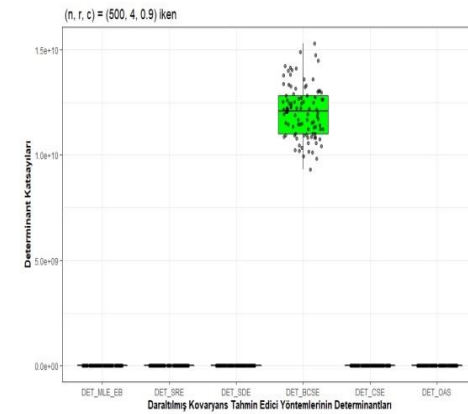
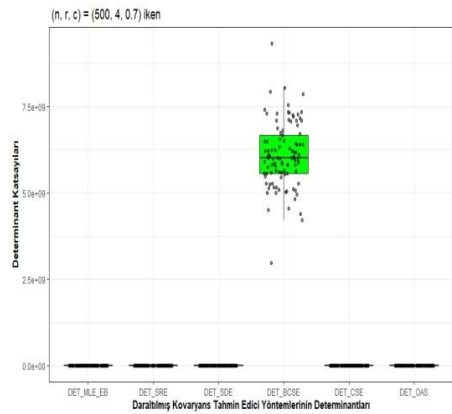
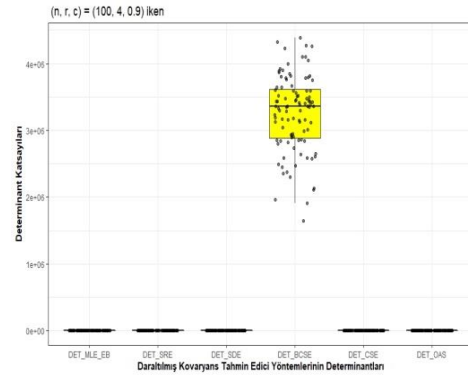
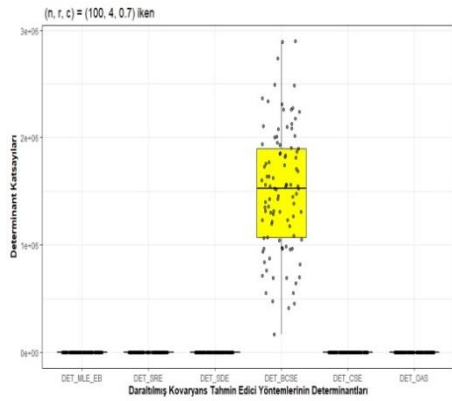
daha iyi tahminde bulunduğunu göstermektedir. Tabloda da görüldüğü üzere tıpkı rankın 3 olduğu durumda olduğu gibi, rankın 4 olduğu durumda ve gözlem sayılarının aynen korunduğu veri setinde daraltılmış kovaryans tahmin edici yöntemlerinden biri olan BCSE kovaryans tahmin edici yönteminin, diğer kovaryans tahmin edici yöntemlerinden daha yüksek değerlere sahip olduğu görülmüştür. Bunun sebeplerini incelemek için daraltılmış kovaryans tahmin edici yöntemlerinin determinant değerlerine bakılabilir. Buna göre,

Tablo 4.15. Gözlem sayılarının 100, 500 ve 1000 değerlerini aldığı bir veri seti için korelasyon parametrelerinin de 0.3, 0.5, 0.7 ve 0.9 gibi katsayılara sahip olduğu bir durumda $r = 4$ için eğitim – test modeli üzerinde uygulanmış olan daraltılmış kovaryans tahmin edici yöntemlerinin ortalama yönünden determinant değerleri

| (n, σ, c) | MLE/EB_{DET}^* | SRE_{DET}^* | SDE_{DET}^* | $BCSE_{DET}^*$ | CSE_{DET}^* | OAS_{DET}^* |
|------------------|------------------|---------------|---------------|----------------|---------------|---------------|
| (100,1,0.3) | 0.0362 | 0.0435 | 0.0285 | 7.47E+05 | 0.9505 | 1.7793 |
| (100,1,0.5) | 0.0255 | 0.0312 | 0.0197 | 9.24E+05 | 0.4902 | 0.8294 |
| (100,1,0.7) | 0.0114 | 0.0148 | 0.0084 | 1.36E+06 | 0.1352 | 0.1956 |
| (100,1,0.9) | 0.0019 | 0.0029 | 0.0012 | 3.06E+06 | 0.0130 | 0.0158 |
| (500,1,0.3) | 0.0346 | 0.0360 | 0.0330 | 3.90E+09 | 0.1396 | 0.1518 |
| (500,1,0.5) | 0.0236 | 0.0246 | 0.0224 | 4.45E+09 | 0.0768 | 0.0818 |
| (500,1,0.7) | 0.0110 | 0.0116 | 0.0103 | 6.13E+09 | 0.0259 | 0.0269 |
| (500,1,0.9) | 0.0016 | 0.0017 | 0.0014 | 1.21E+10 | 0.0028 | 0.0028 |
| (1000,1,0.3) | 0.0353 | 0.0360 | 0.0345 | 1.30E+11 | 0.0822 | 0.0846 |
| (1000,1,0.5) | 0.0240 | 0.0245 | 0.0234 | 1.48E+11 | 0.0482 | 0.0492 |
| (1000,1,0.7) | 0.0111 | 0.0115 | 0.0108 | 2.02E+11 | 0.0181 | 0.0183 |
| (1000,1,0.9) | 0.0016 | 0.0016 | 0.0015 | 3.96E+11 | 0.0021 | 0.0021 |

Tablo 4.15. de $r = 4$ için eğitim – test yöntemiyle elde edilmiş olan daraltılmış kovaryans tahmin edici yöntemlerinin ortalama yönünden determinant değerleri gösterilmiştir. Yukarıdaki tabloda da görüldüğü üzere BCSE daraltılmış kovaryans tahmin edici yönteminin determinant değeri, diğer kovaryans tahmin edici yöntemlerinden fazladır. Bu durum ise BCSE daraltılmış kovaryans tahmin edici yönteminin iz (*ridge*) cezasının fazla olmasından kaynaklanmaktadır. Tablodaki verileri incelediğimizde çok değişkenli indirgenmiş rank regresyon yöntemi için

veyahut diğer regresyon yöntemleri için karşılaşılan problemlerin çözümlerinde BCSE daraltılmış kovaryans tahmin edici yönteminin önerilemeyeceği görülmüştür. Tablo 4.15. de gözlem sayılarının sabit bir değer aldığı durumlar incelendiğinde korelasyon değerlerinin artması durumunda determinant katsayılarının sıfıra daha yakın olduğu görülmektedir. Bu durum ise daraltılmış kovaryans tahmin edici yöntemlerinin β katsayıları üzerinden elde edilen hata kareler ortalama değerleri için daha etkili olduğunu göstermektedir. Örneğin; SRE daraltılmış kovaryans tahmin edici yönteminin determinant değerleri, korelasyon değerlerinin artması halinde sıfıra yakın değerler aldığı Tablo 4.15.'te görülmektedir. Bu daraltılmış kovaryans tahmin edici yönteminin hata kareler ortalamasına Tablo 4.14. te verilen değerler kapsamında incelediğimizde ise görülmektedir ki, daraltılmış kovaryans tahmin edici yöntemlerinden bir olan MLE / EB daraltılmış kovaryans tahmin edici yönteminden sonra uygun olan ikinci bir tahmin edici yöntemi olarak görülmektedir. Yani; MLE / EB tahmin edici yönteminden sonra SRE daraltılmış kovaryans tahmin edici yöntemi, β katsayıları üzerinden elde edilen hata kareler ortalaması için en etkili ikinci yöntem olarak karşımıza çıkmaktadır.



Şekil 4.4. $r = 4$, $c = (0.7 \text{ ve } 0.9)$ değerleri için eğitim – test modeli üzerinden daraltılmış kovaryans tahmin edici yöntemlerinin ortalaması alınarak elde edilmiş olan determinant değerlerinin kutu grafikleri

Şekil 4.4. de $r = 4$ için eğitim – test yöntemi üzerinden daraltılmış kovaryans tahmin edici yöntemlerinin ortalaması alınarak elde edilmiş olan determinant değerlerinin kutu grafikleri gösterilmiştir. Yukarıdaki şekilde gösterildiği üzere eğitim – test modeli için elde edilen kutu grafiğinde, aynen Tablo 4.15. de belirtildiği üzere BCSE daraltılmış kovaryans matrisinin iz (*ridge*) cezasının fazla olduğu görülmektedir. Bu durum ise BCSE daraltılmış kovaryans tahmin edici yönteminin, çok değişkenli indirgenmiş rank regresyon yöntemi için karşılaştığımız problemlerin çözümünde uygun olmayacağını göstermektedir. Burada, BCSE daraltılmış kovaryans tahmin edici yöntemi için gözlem sayısının arttığı bir durum incelendiğinde, iz (*ridge*) cezasının arttığı görülmektedir. Yani; gözlem sayısının artması daraltılmış kovaryans tahmin edici yöntemini sıfırdan uzaklaştırdığı gibi ayrıca regresyon yönteminde sorunların büyümesine yol açmaktadır. Bu yüzden bu daraltılmış kovaryans tahmin edici yönteminin çok değişkenli indirgenmiş rank regresyon yöntemi için önerilmemesine bu analiz sonucunda da karar verilmiştir.

Uygulama 1 (Bozdogan, 2003) simülasyon çalışması kapsamında, çok değişkenli indirgenmiş rank regresyon yöntemi için belirlenmiş olan rank değerinin matris boyutlarından küçük olması şartı gereğince, rankın 3 ve 4 olduğu durumlar incelenmiş bulunmaktadır. Bu bağlamda her iki durumda incelenerek hem çok değişkenli indirgenmiş rank regresyon yönteminde hem de çok değişkenli indirgenmiş rank regresyon yöntemi üzerinden elde edilen eğitim – test verisi kullanılarak oluşturulmuş veri seti için, MLE / EB daraltılmış kovaryans tahmin edici yönteminin, β katsayıları üzerinden elde edilen hata kareler ortalama değerleri için en uygun tahmin edici yöntemi olarak görülmüştür.

Benzetim Çalışması – 1 burada tamamlanmaktadır. Bundan sonraki analizler ise Benzetim Çalışması – 2 için elde edilmiştir.

Tablo 4.16. Daraltılmış kovaryans tahmin edici yöntemleriyle beraber kullanılarak elde edilmiş olan β değerlerinin, gözlem sayısının $n = (100, 500 \text{ ve } 1000)$ olduğu bir veride $r = (3 \text{ ve } 4)$ için meydana gelmiş olduğu hata kareler ortalama değerleri

| (n, r) | $MSE_{MLE/EB}$ | MSE_{SRE} | MSE_{SDE} | MSE_{BCSE} | MSE_{CSE} | MSE_{OAS} |
|----------|----------------|-------------|-------------|--------------|-------------|-------------|
| (100,3) | 0.0283 | 0.1331 | 0.3801 | 49.0710 | 8.2406 | 29.1864 |
| (500,3) | 0.0033 | 0.0072 | 0.0178 | 46.8121 | 1.4718 | 1.8216 |

| | | | | | | |
|----------|--------|--------|--------|---------|---------|---------|
| (1000,3) | 0.0016 | 0.0027 | 0.0050 | 52.6475 | 0.6594 | 0.7454 |
| (100,4) | 0.0376 | 0.1809 | 0.7258 | 76.4753 | 11.8091 | 42.4368 |
| (500,4) | 0.0044 | 0.0103 | 0.0214 | 65.6777 | 2.1704 | 2.6838 |
| (1000,4) | 0.0020 | 0.0033 | 0.0067 | 69.3698 | 0.8014 | 0.9054 |

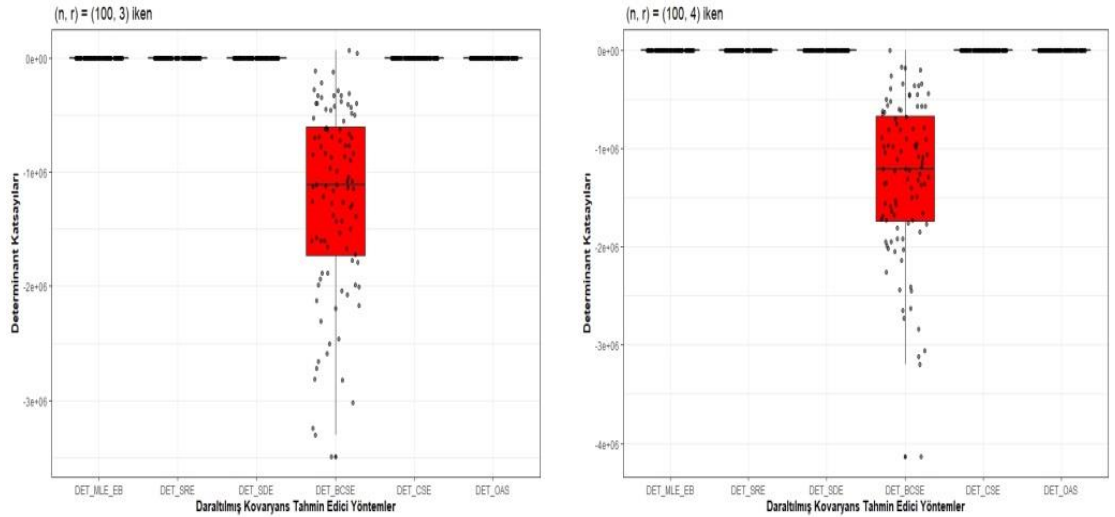
Benzetim Çalışması – 2 için Tablo 4.16. da β katsayıları üzerinden hesaplanan hata kareler ortalama değerleri elde edilmiştir. Burada rank değerinin artması veya azalması durumunda elde edilen verilerin nasıl değiştiği incelenmektedir. Bundan sonraki verilecek olan tablolar için rank değerlerinin durumuna göre yorumlar yapılmıştır. Buna göre ilk tablo için yapılacak olan değerlendirme ise rankın 3 veya 4 olduğu bir durumda daraltılmış kovaryans tahmin edici yöntemlerinden MLE / EB kovaryans tahmin edicisi, hibrit kovaryans yapılarının (Bozdogan, 2004) kullanıldığı veri setinde β katsayıları üzerinden elde edilen hata kareler ortalama değerleri için daha etkili olduğu görülmektedir. Gözlem sayısının sabit olduğu bir durumda rank değerinin artması durumunda dahi MLE / EB daraltılmış kovaryans tahmin edici yönteminin daha etkili olduğu görülmüştür. Ancak rank değerinin azalması durumundaki durum hata kareler ortalaması üzerinde daha etkili olmaktadır. Gözlem sayısı arttığında rank değerleri sabitse bu durum gözlem sayısının artmasının daha faydalı olacağını Benzetim Çalışması – 2 için görebilmekteyiz. Şöyle ki, eğer rank değerleri sabit ve gözlem değerleri artıyorsa, Tablo 4.16. da görüldüğü üzere gözlem sayısı arttıkça daraltılmış kovaryans tahmin edici yöntemlerinden MLE / EB kovaryans tahmin edicisinin sifıra daha yakın bir değer aldığı görülmektedir. Yani bu durum gözlem sayısının 1000 değerini aldığı ve aynı zamanda rank değeri sabit olduğunda, tahmin edilen β katsayılarının gerçek β katsayılarını, sifıra yaklaştıkça yaklaşık olarak daha iyi tahminde bulunmaktadır.

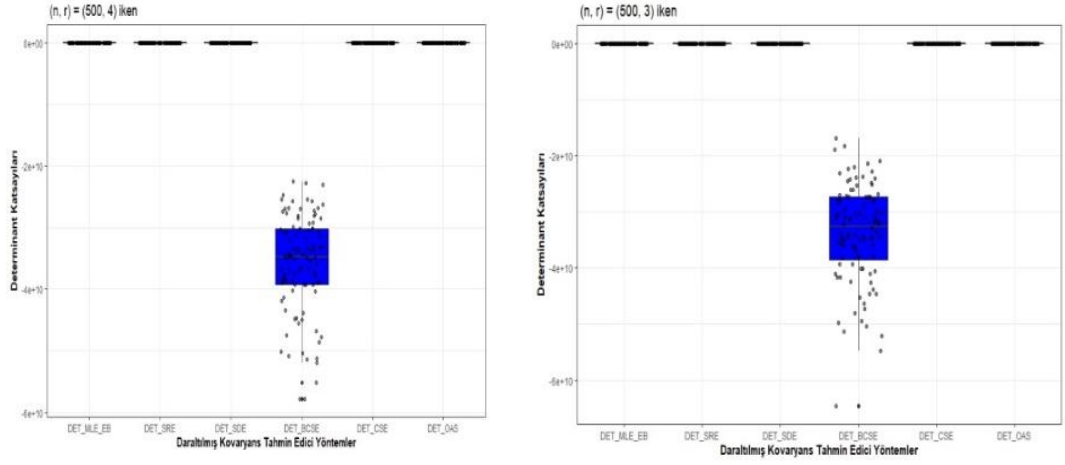
Tablo 4.17. Gözlem sayısının $n = (100, 500 \text{ ve } 1000)$ ve rank değerlerinin de 3 ve 4 olduğu bir veride daraltılmış kovaryans tahmin edici yöntemlerinin determinant değerlerinin ortalama yönünden karşılaştırılması

| (n, r) | MLE / EB_{DET} | SRE_{DET} | SDE_{DET} | $BCSE_{DET}$ | CSE_{DET} | OAS_{DET} |
|----------|------------------|-------------|-------------|--------------|-------------|-------------|
| (100,3) | 0.1086 | 0.1283 | 0.0888 | -1.24E+06 | 0.8641 | 1.7253 |
| (500,3) | 0.1125 | 0.1163 | 0.1082 | -3.53E+10 | 0.2524 | 0.2811 |
| (1000,3) | 0.1146 | 0.1165 | 0.1124 | -2.39E+12 | 0.1817 | 0.1877 |

| | | | | | | |
|----------|--------|--------|--------|-----------|--------|--------|
| (100,4) | 0.1090 | 0.1287 | 0.0891 | -1.19E+06 | 0.8703 | 1.7458 |
| (500,4) | 0.1144 | 0.1183 | 0.1100 | -3.62E+10 | 0.2579 | 0.2875 |
| (1000,4) | 0.1143 | 0.1162 | 0.1121 | -2.40E+12 | 0.1815 | 0.1876 |

Tablo 4.17. de daraltılmış kovaryans tahmin edicilerinin determinant değerleri gösterilmiştir. Buna göre Tablo 4.16. ve Tablo 4.17. de ki verileri incelediğimizde BCSE daraltılmış kovaryans tahmin edici yöntemi hibrit kovaryans yapısının kullanıldığı Benzetim Çalışması – 2’de çok değişkenli indirgenmiş rank regresyon yöntemi için uygun olmadığı görülmektedir. Benzetim Çalışması – 1’de ki sonuçlara göre ortalama yönünden pozitif yönde uzaklaşan BCSE daraltılmış kovaryans tahmin edici yönteminin determinant değerleri, Benzetim Çalışması – 2’de negatif yönde uzaklaştığı görülmektedir. Her iki örnek çalışma için BCSE daraltılmış kovaryans tahmin edici yönteminin uygun olmadığı görülmüştür. Buna göre BCSE daraltılmış kovaryans tahmin edici yöntemi dışındaki diğer daraltılmış kovaryans tahmin edici yöntemleri arasında bir sıralama yapıldığı zaman sırasıyla, Maksimum Olabilirlik/ Bayes tahmin edici (MLE / EB) yöntemi, Şartlandırılmış Ridge (iz) Tahmin Edici (SRE) yöntemi ve üçüncü olarak da Şartlandırılmış Ridge (iz) Diyagonal Tahmin Edici (SDE) yöntemi olarak çok değişkenli indirgenmiş rank regresyon yöntemlerinde ve diğer regresyon yöntemlerinde uygulanabilir olduğu görülmüştür.





Şekil 4.5. Gözlem sayılarının 100, 500 ve 1000 değerlerine sahip olduğu bir veri seti için $r = (3 \text{ ve } 4)$ değerlerinin kullanılmasıyla elde edilmiş olan daraltılmış kovaryans tahmin edici yöntemlerinin kutu grafikleri

Şekil 4.5. de Benzetim Çalışması – 2 için ifade edilen daraltılmış kovaryans tahmin edicilerinin kutu grafikleri gösterilmiştir. Hibrit kovaryans yapısının kullanıldığı Benzetim Çalışması – 2 için oluşturulan Şekil 4.5. teki kutu grafiğinde de görüldüğü gibi rankın 3 ve 4 olduğu durumlar için BCSE daraltılmış kovaryans tahmin edici yönteminin negatif yönde sıfırdan uzaklaştığı görülmektedir. Bu bağlamda BCSE daraltılmış kovaryans tahmin edici yönteminin çok değişkenli indirgenmiş rank regresyon yöntemi için uygun olmadığı kutu grafiğinde de görülmüştür. Bu sonuçlara göre BCSE daraltılmış kovaryans matrisinin iz (*ridge*) cezası negatif yönde artarak devam ettiği gözlemlenmiştir.

Tablo 4.18. $n = 100$ ve $r = 3$ olarak alındığı bir durum için bilgi kriterlerinin daraltılmış kovaryans tahmin edici yöntemlerine etkisi

| Daraltılmış kovaryans tahmin edici yöntemleri | AIC_q | AIC_{cq} | BIC_q | $HQIC_q$ | KIC_{vc} |
|---|-----------|------------|-----------|-----------|------------|
| MLE/EB | 71.2850 | 513.5310 | 119.4400 | 72.9160 | 208.6790 |
| SRE | 159.3580 | 591.0350 | 207.5130 | 160.9890 | 206.9180 |
| SDE | 212.5740 | 637.8650 | 260.7290 | 214.2050 | 205.8530 |
| BCSE | 1697.9060 | 1944.9570 | 1746.0610 | 176.1470 | 1699.5370 |
| CSE | 1121.6410 | 1437.8440 | 1169.7960 | 1537.417 | 187.6720 |
| OAS | 1401.0030 | 1683.6830 | 1449.1580 | 1402.6340 | 182.0850 |

Tablo 4.19. $n = 100$ ve $r = 4$ olarak alındığı bir durum için bilgi kriterlerinin daraltılmış kovaryans tahmin edici yöntemlerine etkisi

| Daraltılmış kovaryans tahmin edici yöntemleri | AIC_q | AIC_{cq} | BIC_q | $HQIC_q$ | KIC_{vc} |
|--|-----------|------------|-----------|-----------|------------|
| MLE/EB | 66.4640 | 509.2880 | 114.6190 | 68.0950 | 208.7750 |
| SRE | 183.8880 | 612.6220 | 232.0440 | 185.5190 | 206.4270 |
| SDE | 277.9810 | 695.4230 | 326.1360 | 279.6120 | 204.5450 |
| BCSE | 2203.3300 | 2389.730 | 2251.4850 | 166.0380 | 2204.9610 |
| CSE | 1400.0270 | 1682.8230 | 1448.1820 | 1769.9050 | 182.1040 |
| OAS | 1725.1870 | 1968.9650 | 1773.3420 | 1726.8180 | 175.6010 |

Tablo 4.18 ve Tablo 4.19. da Benzetim Çalışması – 2 için daraltılmış kovaryans tahmin edici yöntemlerinin etkisiyle hesaplanmış olan bilgi kriterleri gösterilmiştir. Hibrit kovaryans yapılarının kullanıldığı Benzetim Çalışması – 2 için elde edilen bilgi kriterlerinin, daraltılmış kovaryans tahmin edici yöntemlerine etkisiyle gözlem sayısının 100 ve rankın 3 ve 4 olduğu Tablo 4.18. ve Tablo 4.19. da gösterilmektedir. Yukarıdaki verilen tablolara göre bilgi kriterleri, daraltılmış kovaryans tahmin edici yöntemlerinden MLE / EB daraltılmış kovaryans tahmin edici yöntemini β katsayıları üzerinden elde edilen hata kareler ortalaması için daha etkili bulmuşlardır. Ancak KIC_{vc} bilgi kriteri, diğer dört bilgi kriterinden farklı bir sonuç elde ederek OAS bilgi kriterini daha etkili olduğunu göstermektedir. Bu durum ise KIC_{vc} bilgi kriterinin çok değişkenli indirgenmiş rank regresyon yöntemi için kötü bir sonuç elde ettiğini ve güvenilir olmadığını göstermektedir.

Tablo 4.20. $n = 1000$ ve $r = 3$ olarak alındığı bir durum için bilgi kriterlerinin daraltılmış kovaryans tahmin edici yöntemlerine etkisi

| Daraltılmış kovaryans tahmin edici yöntemleri | AIC_q | AIC_{cq} | BIC_q | $HQIC_q$ | KIC_{vc} |
|--|---------|------------|----------|----------|------------|
| MLE/EB | 51.9820 | 4992.4390 | 169.2150 | 77.9410 | 190.7970 |
| SRE | 65.1120 | 5005.4110 | 182.3450 | 91.0710 | 190.7710 |
| SDE | 82.6610 | 5022.7490 | 199.8940 | 108.6200 | 190.7360 |

| | | | | | |
|------|------------|------------|------------|-----------|------------|
| BCSE | 16724.7370 | 21465.1200 | 16841.9700 | 157.4520 | 16750.6960 |
| CSE | 3399.6290 | 8299.9130 | 3516.8620 | 9214.1410 | 184.1020 |
| OAS | 3642.7030 | 8540.0710 | 3759.9360 | 3668.6620 | 183.6160 |

Tablo 4.21. $n = 1000$ ve $r = 4$ olarak alındığı bir durum için bilgi kriterlerinin daraltılmış kovaryans tahmin edici yöntemlerine etkisi

| Daraltılmış kovaryans tahmin edici yöntemleri | AIC_q | AIC_{cq} | BIC_q | $HQIC_q$ | KIC_{vc} |
|--|------------|------------|------------|-----------|------------|
| MLE/EB | 51.8710 | 4992.3290 | 169.1040 | 77.8300 | 190.7980 |
| SRE | 68.3960 | 5008.6550 | 185.6290 | 94.3550 | 190.7650 |
| SDE | 95.1820 | 5035.1200 | 212.4150 | 121.1410 | 190.7110 |
| BCSE | 21206.2830 | 25892.8880 | 21323.5160 | 148.4890 | 21232.2420 |
| CSE | 4283.5900 | 9173.2670 | 4400.8220 | 9944.8920 | 182.3340 |
| OAS | 4585.6540 | 9471.7060 | 4702.8870 | 4611.6130 | 181.7300 |

Tablo 4.20. ve Tablo 4.21. de gözlem sayısının 1000 değerini aldığı durum incelenmiş olup Benzetim Çalışması – 2 için daraltılmış kovaryans tahmin edici yöntemlerinin etkisiyle hesaplanmış olan bilgi kriterleri gösterilmiştir. Yukarıdaki Tablo 4.20. de elde edilen analizler incelendiğinde, dört bilgi kriterlerinin sonuçlarına göre Maksimum Olabilirlik / Bayes Tahmin Edici (MLE / EB) yöntemi β katsayıları üzerinden elde edilen hata kareler ortalamaları için daha etkili bulunmaktadır. Yine burada KIC_{vc} bilgi kriterinin diğer bilgi kriterlerinden daraltılmış kovaryans tahmin edici yöntemlerini farklı bir şekilde hesapladığı görülmektedir. Yani bu bilgi kriteri çok değişkenli indirgenmiş rank regresyon yöntemi için güvenilir bir sonuç vermemektedir. Aynı şekilde aşağıda yer alacak olan Tablo 4.21. de rankın 4 olduğu durumlar için de görülmektedir. Buna göre hem $r = 3$ hem de $r = 4$ gibi durumlar için bilgi kriterleri, Maksimum Olabilirlik / Bayes Tahmin edici yöntemini çok değişkenli indirgenmiş rank regresyon yöntemi için en iyi kovaryans tahmin edici olarak göstermektedirler. Yalnız bu bilgi kriterlerinden KIC_{vc} , diğer bilgi kriterlerine göre daraltılmış kovaryans tahmin edici yöntemlerini daha iyi yansıtamadığı görülmüştür.

Tablo 4.22. Eğitim – Test modeli kullanılarak oluşturulmuş olan bir veri için daraltılmış kovaryans tahmin edici yöntemleriyle elde edilmiş olan β değerlerinin, gözlem sayısının $n = (100, 500$ ve $1000)$ olduğu bir veride $r = (3$ ve $4)$ için meydana gelmiş olduğu hata kareler ortalama değerleri

| (n, r) | $MSE_{MLE/EB}^*$ | MSE_{SRE}^* | MSE_{SDE}^* | MSE_{BCSE}^* | MSE_{CSE}^* | MSE_{OAS}^* |
|----------|------------------|---------------|---------------|----------------|---------------|---------------|
| (100,3) | 4.1600 | 4.3540 | 4.1580 | 18.7250 | 6.4510 | 17.5070 |
| (500,3) | 3.8100 | 3.8450 | 3.7960 | 18.0940 | 3.9610 | 4.0170 |
| (1000,3) | 4.0440 | 4.0610 | 4.0370 | 19.4530 | 4.0150 | 4.0180 |
| (100,4) | 4.7000 | 4.9380 | 4.7130 | 21.4440 | 7.0140 | 16.9830 |
| (500,4) | 4.4630 | 4.5060 | 4.4460 | 21.4460 | 4.6600 | 4.7300 |
| (1000,4) | 4.4070 | 4.4280 | 4.3970 | 21.3170 | 4.4990 | 4.5120 |

Tablo 4.22. de eğitim – test yöntemi için hesaplanmış daraltılmış kovaryans tahmin edici yöntemlerinin etkisiyle meydana gelen β katsayılarının hata kareler ortalama değerleri gösterilmiştir. Buna göre Tablo 4.22. deki değerleri incelediğimizde rankın 3 ve 4 olduğu durumlar için daraltılmış kovaryans tahmin edici yöntemlerinden MLE / EB kovaryans tahmin edici yöntemi β katsayıları üzerinden elde edilen hata kareler ortalaması üzerinde daha etkili olduğu görülmektedir. Tablo 4.22. incelendiğinde aynı şekilde görülmektedir ki, BCSE daraltılmış kovaryans tahmin edici yönteminin diğer kovaryans tahmin edici yöntemlerinden daha fazla sıfırdan uzaklaştığı görülmektedir. Nasıl ki MLE / EB daraltılmış kovaryans tahmin edici yöntemi, tahmin edilen β katsayılarının gerçek β katsayılarını yaklaşık olarak sıfıra yaklaştıkça daha iyi tahminde bulunuyorsa aynı şekilde BCSE daraltılmış kovaryans tahmin edici yöntemi de β katsayılarını sıfırdan uzaklaşarak tahminde bulunmayı zorlaştırmaktadır.

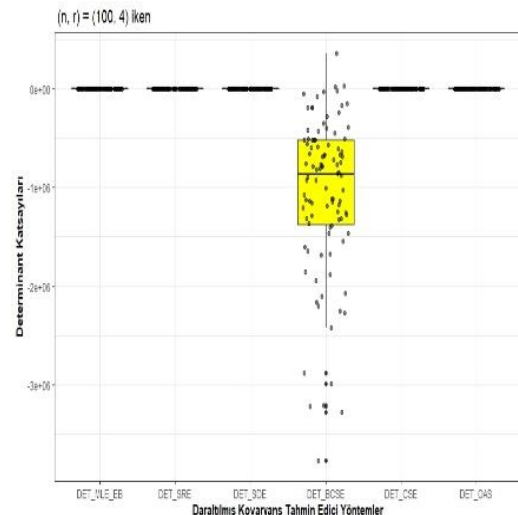
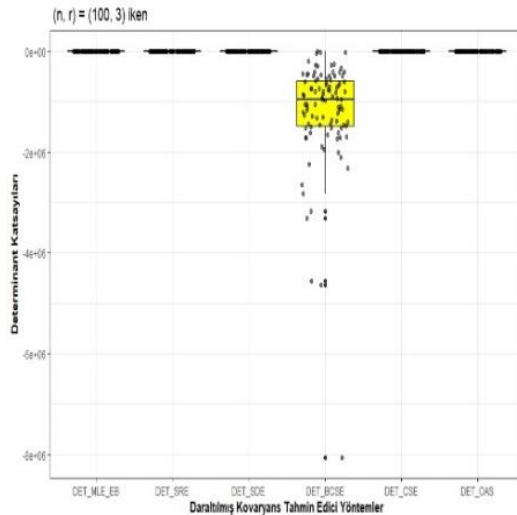
Tablo 4.16. ve Tablo 4.22. de ki çok değişkenli indirgenmiş rank regresyon yöntemi için daraltılmış kovaryans tahmin edici yöntemleri üzerinden elde edilen hata kareler ortalama değerleri incelendiğinde, BCSE daraltılmış kovaryans tahmin edici yönteminin diğer daraltılmış kovaryans tahmin edici yöntemlerine göre büyük değerlere sahip olduğunu göstermektedir. Bu durum ise çok değişkenli indirgenmiş rank regresyon yöntemi için uygun olmayan daraltılmış kovaryans tahmin edici yöntemi olduğunu göstermektedir. Burada BCSE daraltılmış kovaryans tahmin edici yönteminin iz (*ridge*) cezasının fazla olduğunu söyleyebiliriz. Daraltılmış kovaryans tahmin edici yöntemlerinden BCSE tahmin edici yönteminin bu durumunu daha net

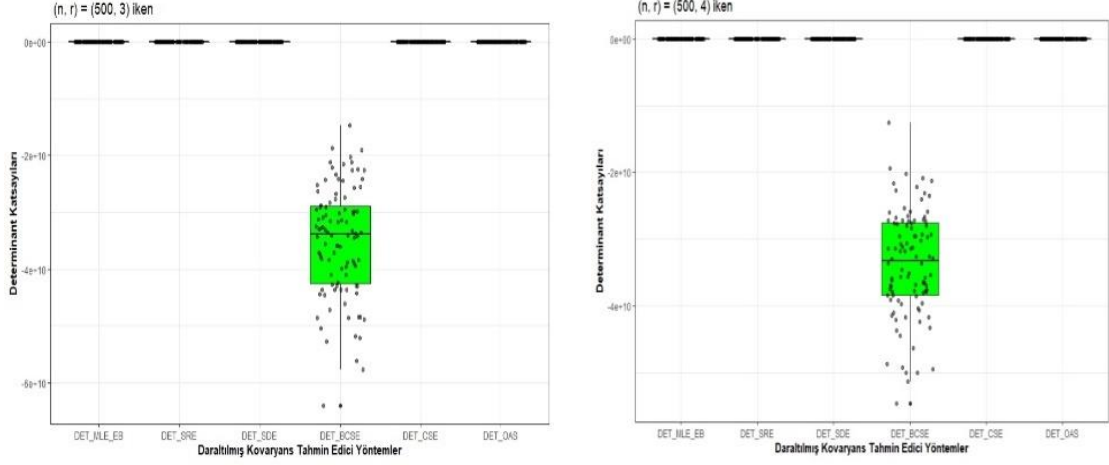
ifade edebilmek için tahmin edici yöntemlerinin determinant değerlerine bakmak yeterli olabilmektedir.

Tablo 4.23. Gözlem sayısının $n = (100, 500 \text{ ve } 1000)$ ve rank değerlerinin de 3 ve 4 olduğu bir veride eğitim – test modeli kullanılarak elde edilmiş olan daraltılmış kovaryans tahmin edici yöntemlerinin determinant değerlerinin ortalama yönünden karşılaştırılması

| (n, r) | MLE / EB_{DET}^* | SRE_{DET}^* | SDE_{DET}^* | $BCSE_{DET}^*$ | CSE_{DET}^* | OAS_{DET}^* |
|----------|--------------------|---------------|---------------|----------------|---------------|---------------|
| (100,3) | 0.1025 | 0.1213 | 0.0836 | -1.27E+06 | 0.8905 | 1.7950 |
| (500,3) | 0.1135 | 0.1174 | 0.1092 | -3.49E+10 | 0.2526 | 0.2812 |
| (1000,3) | 0.1167 | 0.1186 | 0.1144 | -2.30E+12 | 0.1864 | 0.1929 |
| (100,4) | 0.1002 | 0.1190 | 0.0818 | -1.21E+06 | 0.7903 | 1.5920 |
| (500,4) | 0.1129 | 0.1168 | 0.1086 | -3.48E+10 | 0.2528 | 0.2820 |
| (1000,4) | 0.1176 | 0.1195 | 0.1153 | -2.35E+18 | 0.1881 | 0.1947 |

Tablo 4.23. de eğitim – test yöntemi kullanılarak elde edilen daraltılmış kovaryans tahmin edicilerin determinant değerleri gösterilmiştir. Buna göre Tablo 4.23. de ki değerler incelendiğinde görülmüştür ki, BCSE daraltılmış kovaryans tahmin edici yönteminin determinant değerleri negatif yönde artmakta ve sıfırdan giderek uzaklaşmaktadır. Bu durum ise yukarıda da belirtildiği üzere iz (*ridge*) cezasının fazla olmasından kaynaklanmaktadır. Buna göre BCSE daraltılmış kovaryans tahmin edici yöntemi çok değişkenli indirgenmiş rank regresyon yöntemi için uygun olmadığı görülmüştür. Aşağıda kutu grafikleri bakımından görsel olarak gösterilmiştir.





Şekil 4.6. Gözlem sayılarının 100, 500 ve 1000 değerlerine sahip olduğu bir veri seti için $r = (3 \text{ ve } 4)$ değerlerinin kullanılmasıyla elde edilmiş olan daraltılmış kovaryans tahmin edici yöntemlerinin kutu grafikleri

Şekil 4.6. da Benzetim Çalışması – 2 için hesaplanmış olan daraltılmış kovaryans tahmin edici yöntemlerinin kutu grafikleri gösterilmiştir. Buna göre, Şekil 4.6. da gösterildiği üzere daraltılmış kovaryans tahmin edici yöntemlerinin determinant değerlerinin elde edilmesiyle oluşturulmuş kutu grafikleri verilmiştir. Yukarıdaki şekilde de gösterildiği gibi BCSE daraltılmış kovaryans tahmin edici yönteminin diğer kovaryans tahmin edici yöntemlerinden farklı bir şekilde negatif yönde sıfırdan uzaklaşmaktadır. Bu durum ise BCSE kovaryans tahmin edici yönteminin iz (*ridge*) cezasının negatif yönde fazla olduğunu göstermektedir. Sonuç olarak; BCSE daraltılmış kovaryans tahmin edici yöntemi çok değişkenli indirgenmiş rank regresyon yöntemi için uygun olmadığı kutu grafiği ile de görülmüştür.

4.3. Gerçek Veri Seti Uygulamaları

Bu bölümde ise gerçek veri setleri ile çalışılarak, çok değişkenli indirgenmiş rank regresyon yöntemi için daraltılmış kovaryans tahmin edici yönteminin, gerçek veri setleri üzerindeki durumu incelenecektir. Bu aşamada farklı rank değerleri için bağımsız değişkenlere ait koşul indeksleri, kovaryans matrislerine ait determinant değerleri ve test-eğitim verileri için hata kareler ortalamaları verilmiştir.

Veri setleri R programı içerisinde yer alan “*Robustbase* ve *ICSNP* paketlerinden” elde edilmiştir. Bu uygulamanın ilki *Robustbase* paketindeki Epilepsi veri setidir. Epilepsi veri seti 59 gözlemden ve 11 değişkenden meydana gelmektedir. Burada, 59 hastadan deneysel çalışmalar sonucunda elde edilmiş veriler bulunmaktadır (Maronna ve Suggest, 2009). Aynı zamanda veri içinde, kısmi nöbetler

geçiren hastalar için 4 farklı şekilde elde edilmiş epilepsi atak nöbet değerleri yer almaktadır. Bu değerlerden ilk ikisi, ilk 2 haftadaki epilepsi atak nöbet değerlerini gösterirken, son iki değer ise son 2 haftanın epilepsi atak nöbet değerlerini göstermektedir (Thall ve Vaill, 1990).

Epilepsi verisi için aşağıda yer alacak olan bilgiler ile bu veri seti hakkında temel bir bilgi sahibi olmamızı bizlere göstermektedir. Şöyle ki,

Tablo 4.24. Epilepsi veri setine ait değişkenlerin sözel ifadeleri

| Değişkenler | Açıklama |
|-------------------|--|
| Y_1 | Hasta takibi sonrası ilk atak nöbet değerleri |
| Y_2 | Hasta takibi sonrası ikinci atak nöbet değerleri |
| Y_3 | Hasta takibi sonrası üçüncü atak nöbet değerleri |
| Y_4 | Hasta takibi sonrası dördüncü atak nöbet değerleri |
| Y_{sum} | Dört defa takip edilen bir hastanın toplam epilepsi atak nöbet sayı değerleri |
| Base | Hastalara ilaç verilmeden önce 8 haftalık dönem içerisinde kaydedilmiş olan epilepsi atak nöbet sayı değerleri |
| Base ₄ | Epilepsi hastalarının geçirmiş oldukları atak nöbet değerlerinin dörde bölünmüş olan değerlerdir. |
| Age | Hastaların yaşı |
| Age ₁₀ | Hastaların yaşlarının 10'a bölünmüş olan değerleridir. |
| ID | Hastaların kimlik numaraları |
| TRT | Epilepsi hastaları üzerinde kullanılan anti – epilepsi ilacı olan progabide ilacının uygulanıp uygulanmadığını belirten ve plasebo progabide seviyelerini de gösteren bir faktördür. |

Tablo 4.24. de Epilepsi veri setine ait değişkenler ve açıklamaları gösterilmiştir. Buna göre yukarıda verilen bilgiler doğrultusunda, çok değişkenli indirgenmiş rank regresyon yöntemi için Epilepsi veri setinden ID, TRT, Y_{sum} , Base₄ ve Age₄ faktörleri çıkarılarak analiz edilmiştir. Burada Y_1, Y_2, Y_3, Y_4 değişkenleri bağımlı değişkenler olarak; Base ve Age değişkenleri ise bağımsız değişkenler olarak alınmıştır. Epilepsi veri seti böylelikle çok değişkenli indirgenmiş rank regresyon yöntemi için uygun hale getirilmiştir.

İkinci bir uygulama yöntemi ise “*LASER*” veri seti ile yapılan bir çalışmadır. Bu veri seti “ICSNP paketinden elde edilmiştir. Genç Finlilerdeki kalp ve damarlarına ilişkin risk çalışması anlamına gelen bu veri seti için yaş aralığının 26 – 42 arasında olduğu 223 hasta arasında ölçüm yapılmıştır. Veri seti içinde 32 değişken

bulunmaktadır (Nordhausen, vd., 2018). Bu veri setinin çok deęişkenli indirgenmiř rank regresyon yöntemine uygun hale gelmesi için, veri setinin içindeki “cinsiyet, ortalama kalp atıř farkları, ortalama kalp debi farkları, ortalama kan damarlarında oluřan direnç farkları ve ortalama nabız dalga hızı farkları” çıkarılmıřtır.

Böylelikle “LASERI” veri seti, 7 tane bağımsız deęişkenden ve 16 tane bağımlı deęişkenden oluřan bir veri seti olup, çok deęişkenli regresyon yöntemi için uygun hale getirilmiřtir. Bu açıklamalardan sonra veri seti içinde yer almıř olan deęişkenlerin tanımları ařağıdaki gibi belirtilmektedir. řöyle ki,

Tablo 4.25. LASERI veri setine ait deęişkenlerin sözel ifadeleri

| Deęişkenler | Açıklama |
|--------------------|--|
| Cinsiyet | Erkek ve Kadın (Faktör Grubu) |
| Age | Yař |
| Height | Boy (cm) |
| Weight | Kilo (kg) |
| Waist | Bel çevresi (cm) |
| BMI | Vücut Kitle İndeksi |
| WHR | Bel – Kalça oranı |
| HRT1 | Ortalama kalp atıř hızı (Dinlenme ařamasında iken) |
| HRT 2 | Ortalama kalp atıř hızı (Tilt testi – 2. Dakika) |
| HRT 3 | Ortalama kalp atıř hızı (Tilt testi – 5. Dakika) |
| HRT 4 | Ortalama kalp atıř hızı (Tilt testi sonrası – 5. Dakika) |
| PWVT1 | Ortalama nabız dalga hızı (Dinlenme ařamasının 10. Dakikasında iken) |
| PWVT2 | Ortalama nabız dalga hızı (Tilt testi – 2. Dakika) |
| PWVT3 | Ortalama nabız dalga hızı (Tilt testi – 5. Dakika) |
| PWVT4 | Ortalama nabız dalga hızı (Tilt testi sonrası – 5. Dakika) |
| COT1 | Ortalama kalp debisi (Dinlenme ařamasının 10. Dakikasında iken) |
| COT2 | Ortalama kalp debisi (Tilt testi – 2. Dakika) |
| COT3 | Ortalama kalp debisi (Tilt testi – 5. Dakika) |
| COT4 | Ortalama kalp debisi (Tilt testi sonrası – 5. Dakika) |
| SVRIT1 | Ortalama sistemik vasküler direnç indeksi (Dinlenme ařamasının 10. Dakikasında iken) |
| SVRIT2 | Ortalama sistemik vasküler direnç indeksi (Tilt testi – 2. Dakika) |
| SVRIT3 | Ortalama sistemik vasküler direnç indeksi (Tilt testi – 5. Dakika) |
| SVRIT4 | Ortalama sistemik vasküler direnç indeksi (Tilt testi sonrası – 5. Dakika) |

Tablo 4.25. de LASERI veri setine ait deęişkenler ve açıklamaları ifade edilmiřtir. Buna göre LASERI veri seti için bağımlı deęişkenler grubu olarak ortalama kalp atıř hızları, ortalama nabız dalga hızları, ortalama kalp debileri ve ortalama

sistemik vasküler direnç indeksi yer alırken bağımsız değişkenler grubunda ise yaş, boy, kilo değerleri, vücut kitle indeksi, bel çevresi ve bel – kalça oranları yer almaktadır.

Yukarıdaki bilgiler ifade edildikten sonra, bu iki uygulama sırasıyla analiz edilmiştir. Öncelikle Epilepsi veri seti üzerinden elde edilmiş olan sonuçlar, daha sonra ise LASERI veri seti üzerinden elde edilmiş olan sonuçlar Tablo 4.26’da gösterilmektedir.

Tablo 4.26. Epilepsi veri seti üzerinden elde edilen, rank değerlerinin 1 ve 2 değerlerini aldığı, daraltılmış kovaryans tahmin edici yöntemlerinin etkisiyle oluşturulmuş bilgi kriterleri değerleri

| Rank değerleri | Daraltılmış kovaryans tahmin edici yöntemleri | AIC _q | AIC _{cq} | BIC _q | HQIC _q | KIC _{vc} |
|----------------|---|------------------|-------------------|------------------|-------------------|-------------------|
| 1 | MLE/EB | 285.8695 | 358.5629 | 296.4898 | 286.3574 | 68.3075 |
| | SRE | 285.8695 | 358.5629 | 296.4898 | 286.3574 | 68.3075 |
| | SDE | 286.7159 | 359.3089 | 297.3362 | 287.2037 | 68.2788 |
| | BCSE | 431.2715 | 486.7139 | 441.8918 | 431.7594 | 63.3786 |
| | CSE | 288.0741 | 360.5060 | 298.6944 | 279.1166 | 68.2328 |
| | OAS | 288.1738 | 360.5939 | 298.7941 | 288.6617 | 68.2294 |
| 2 | MLE/EB | 286.5948 | 359.2022 | 297.2151 | 287.0827 | 68.2829 |
| | SRE | 286.5948 | 359.2022 | 297.2151 | 287.0827 | 68.2829 |
| | SDE | 287.4363 | 359.9439 | 298.0566 | 287.9242 | 68.2544 |
| | BCSE | 431.4830 | 486.9002 | 442.1033 | 431.9709 | 63.3715 |
| | CSE | 288.7859 | 361.1333 | 299.4062 | 279.6880 | 68.2087 |
| | OAS | 288.8851 | 361.2207 | 299.5054 | 289.3730 | 68.2053 |

Tablo 4.26. da Epilepsi veri setine ait veriler, daraltılmış kovaryans tahmin edicilerin etkisiyle hesaplanmış olan bilgi kriterleri değerleri gösterilmiştir. Buna göre Yukarıdaki Epilepsi verisi üzerinden elde edilmiş olan bilgi kriterleri değerleri, veri üzerinde stabilizasyon yapılarak elde edilmiştir. Çok değişkenli indirgenmiş rank regresyon yönteminin incelenmesi için rank değerlerinin durumu, bu regresyon yönteminde çok önemli bir yere sahip olmaktadır. Epilepsi veri setine baktığımızda

bağımlı değişken sayısının 4 tane olduğunu, bağımsız değişken sayısının ise 2 tane olduğunu görebilmekteyiz. Buna göre, epilepsi veri seti üzerinde, çok değişkenli indirgenmiş rank regresyon yöntemini uygulayabilmemiz için regresyon yönteminin şartlarından biri olan bağımlı ve bağımsız değişken sayılarını kapsayacak şekilde rank değerinin küçük veya küçük eşit olması gerekmektedir. Aynı şekilde diğer uygulama örneği için de bu durum geçerli olmaktadır.

Öncelikle bu veri seti için, bilgi kriterleri değerleri incelendiğinde rankın 1 veya 2 olduğu durumlarda, MLE / EB ve SRE daraltılmış kovaryans tahmin edici yöntemlerinin, epilepsi veri seti için en iyi yöntem olduğu görülmektedir. Burada, bilgi kriterlerinden AIC_q , AIC_{cq} ve BIC_q kriterleri yukarıda belirtilen daraltılmış kovaryans tahmin edici yöntemlerini belirlerken, $HQIC_q$ kriteri CSE daraltılmış kovaryans tahmin edici yöntemini, KIC_{vc} bilgi kriteri ise BCSE daraltılmış kovaryans tahmin edici yöntemini en iyi yöntem olarak belirlemiştir.

Bu değerlendirmeden sonra şu sonuç elde edilmiştir. MLE / EB ve SRE bilgi kriterleri, epilepsi veri seti gibi başka veri setleri için de çok değişkenli indirgenmiş rank regresyon yönteminin uygulandığı durumlarda, regresyon yönteminde karşımıza çıkacak olan çoklu bağlantı, kovaryans matrisinin tersinin alınamaması veyahut hataların bağımsızlığı gibi bir sorunda uygulanacak çözüm yollarından biri olduklarını göstermektedirler. Bu iki tahmin edici yöntemi alternatif bir çözüm yolu olarak gözükmektedir.

Tablo 4.27. Epilepsi veri seti için rank değerlerinin 1 ve 2 değerlerini aldığı durumlarda daraltılmış kovaryans tahmin edici yöntemleri kullanılarak elde edilen hata kareler ortalama değerleri

| Rank değerleri | Daraltılmış kovaryans tahmin edici yöntemleri | Eğitim verilerine ait hata kareler ortalama değerleri | Test verilerine ait hata kareler ortalama değerleri |
|----------------|---|---|---|
| 1 | MLE/EB | 8.4419 | 9.8864 |
| | SRE | 8.4419 | 9.8864 |
| | SDE | 8.4722 | 9.8900 |
| | BCSE | 15.6319 | 12.7955 |
| | CSE | 8.5211 | 9.8775 |
| | OAS | 8.5247 | 9.8769 |
| | MLE/EB | 8.4678 | 9.2074 |

| | | | |
|---|------|---------|---------|
| | SRE | 8.4678 | 9.2074 |
| 2 | SDE | 8.4981 | 9.2121 |
| | BCSE | 15.6459 | 12.4529 |
| | CSE | 8.5468 | 9.2114 |
| | OAS | 8.4678 | 9.2074 |

Tablo 4.27. de Epilepsi veri setine ait veriler hem eğitim verileri üzerinden hem de test verileri üzerinden daraltılmış kovaryans tahmin edicilerin etkisiyle hesaplanmış olan hata kareler ortalama değerlerini göstermektedir. Buna göre Tablo 4.27’de hem eğitim verilerine ait hem de test verilerine ait hata kareler ortalama değerleri bulunmuştur. Buna göre $r = 1$ için eğitim verilerine bakıldığında MLE / EB ve SRE daraltılmış kovaryans tahmin edicilerinin hata kareler ortalama değerleri diğer tahmin edici yöntemlerine göre küçük değerlere sahip olduğu görülmektedir. Test verisi için baktığımızda ise, $r = 1$ için CSE ve OAS değerlerinin diğer daraltılmış kovaryans değerlerinin hata kareler ortalamalarına göre küçük değerler aldığı görülmektedir. Bu durum eğitim veri seti için MLE / EB ve SRE daraltılmış kovaryans tahmin edici yöntemlerinin, test veri seti içinde CSE ve OAS daraltılmış kovaryans tahmin edici yöntemlerini model üzerinde tahmin edilen değerlerin, gerçek değerlere daha yakın tahminlerde bulunacağını göstermektedir. Başka bir ifadeyle, yukarıda söz edilen bu daraltılmış kovaryans tahmin edici yöntemleri, diğer daraltılmış kovaryans tahmin edici yöntemlerden daha az oranda hata katsayılarına sahiptirler. Diğer bir sonuç incelendiğinde $r = 2$ için MLE / EB ve SRE daraltılmış kovaryans tahmin edici yöntemleri eğitim veri seti için en iyi kovaryans tahmin edicisi olarak belirlenmişken, test verisi için incelediğimizde MLE / EB, SRE ve OAS daraltılmış kovaryans tahmin edici değerlerinin model üzerinde daha etkili olduğunu görmekteyiz. Bu durumda $r = 2$ için tahmin edilen β değerleri, gerçek β değerlerine daha iyi tahmin ettiği görülmüştür. Yukarıdaki tabloda göze çarpan başka bir durum daha vardır ki, bu da BCSE daraltılmış kovaryans tahmin edici yöntemi için elde edilen sonuçtur. Bu durum, daraltılmış kovaryans tahmin edici yöntemlerinden biri olan BCSE tahmin edicinin sahip olduğu iz (*ridge*) cezasının fazla olmasından kaynaklanmaktadır. Yani, BCSE daraltılmış kovaryans tahmin edici yöntemiyle elde edilmiş olan çok değişkenli regresyon yöntemi için kovaryans matrisinin kötü koşullandırılmış olabileceğini söyleyebilmekteyiz. Bu durumu daha net bir şekilde görebilmek için daraltılmış

kovaryans tahmin edici yöntemlerinin determinant değerlerine bakmak yeterli olacaktır.

Tablo 4.28. Epilepsi veri seti için rank değerlerinin 1 ve 2 değerlerini aldığı durumlarda, daraltılmış kovaryans tahmin edici yöntemleri kullanılarak elde edilmiş olan determinant değerleri

| Rank Değerleri | Daraltılmış kovaryans tahmin edici yöntemleri | Eğitim verilerine ait determinant katsayıları | Test verilerine ait determinant katsayıları |
|----------------|---|---|---|
| 1 | MLE/EB | 27661.85 | 27580.69 |
| | SRE | 27661.85 | 27580.69 |
| | SDE | 27883.87 | 27809.78 |
| | BCSE | 144515.41 | 148359.47 |
| | CSE | 36996.60 | 37171.72 |
| | OAS | 37382.44 | 37566.52 |
| 2 | MLE/EB | 27661.85 | 28746.43 |
| | SRE | 27661.85 | 28746.43 |
| | SDE | 27883.87 | 28988.71 |
| | BCSE | 144515.41 | 158227.35 |
| | CSE | 36996.60 | 38977.10 |
| | OAS | 37382.44 | 39395.77 |

Tablo 4.28. de Epilepsi verisine ait daraltılmış kovaryans tahmin edici yöntemleri kullanılarak elde edilmiş olan determinant değerleri gösterilmiştir. Buna göre Tablo 4.28. de gösterilen daraltılmış kovaryans tahmin edici yöntemlerinin determinant değerleri incelendiğinde, hem rank değerinin 1 olduğu durumlarda hem de 2 olduğu durumlarda görülmektedir ki, BCSE daraltılmış kovaryans tahmin edici yönteminin determinanı yüksek çıkmıştır. Aynı şekilde, Epilepsi veri seti üzerinden elde edilen eğitim – test modeli için de söz konusu olmaktadır. Bu durum ise BCSE yönteminin iz (*ridge*) cezasının fazla olmasından kaynaklanmaktadır. Sonuç olarak; BCSE daraltılmış kovaryans tahmin edici yönteminin etkisiyle elde edilmiş olan çok değişkenli indirgenmiş rank regresyon yönteminin kovaryans matrisi, kötü koşullandırılmış bir matris olduğuna karar verilmiştir. Buna göre, daraltılmış

kovaryans tahmin edici yöntemlerinden BCSE tahmin edici yönteminin, epilepsi veri seti üzerinde uygun olmadığı kanısına varılmıştır.

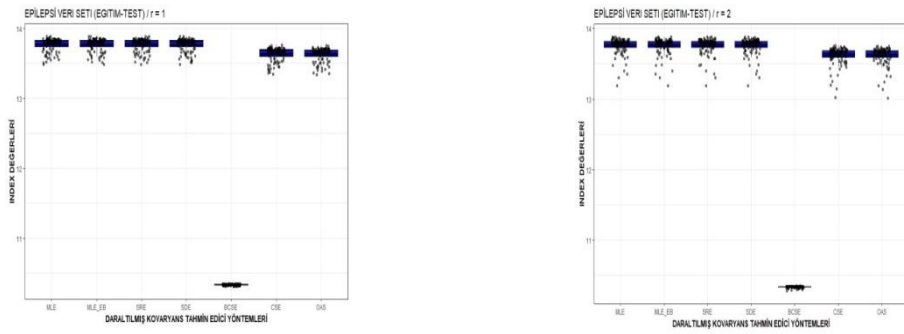
Gerçek veri setleri ile örnek çalışma yaptığımız bu iki veri seti kovaryans matrislerinin tersinin alınabildiği durumlara örnektir. Çok değişkenli indirgenmiş rank regresyon yönteminde ve diğer regresyon yöntemlerinde karşımıza çıkacak olan çoklu bağlantı gibi sorunlar ortaya çıkmaktadır. Bu iki uygulama için çoklu bağlantı sorununu ölçen koşul indeksi değerleri elde edilmiştir. Koşul indeks değerleri kovaryans matrisleri üzerinden elde edilmiştir. Bu değerler elde edilirken daraltılmış kovaryans tahmin edici yöntemler stabilize edilerek oluşturulmuştur. Stabilizasyon işlemi, birbirinden oldukça uzaklaşan öz değerlerin en yakın değerler alacak şekilde genişletilmesi olarak tanımlanmaktadır (Pamukçu, vd., 2019). Çoklu bağlantı probleminin çözümü için alternatif bir yöntem olarak ifade edilen koşul indeks sonuçları öncelikle epilepsi veri seti için daha sonra ise LASERI veri seti için aşağıdaki gibi gösterilmektedir. Buna göre,

Tablo 4.29. Epilepsi veri seti için rank değerlerinin 1 ve 2 olduğu durumlarda daraltılmış kovaryans tahmin edici yöntemlerinin etkisiyle oluşturulmuş olan koşul indeks değerleri

| Daraltılmış Kovaryans Tahmin Edici Yöntemleri | $r = 1$ | $r = 2$ |
|--|---------------------------|---------------------------|
| MLE | 13.7634 | 13.7559 |
| MLE/EB | 13.7634 | 13.7559 |
| SRE | 13.7634 | 13.7559 |
| SDE | 13.7616 | 13.7543 |
| BCSE | 10.3347 | 10.3340 |
| CSE | 13.6290 | 13.6212 |
| OAS | 13.6232 | 13.6153 |

Tablo 4.29. da Epilepsi veri setine ait daraltılmış kovaryans tahmin edici yöntemlerinin etkisiyle oluşturulmuş koşul indeks değerleri gösterilmiştir. Buna göre Tablo 4.29. da elde edilen değerler için daraltılmış kovaryans tahmin edici yöntemlerinden BCSE kovaryans tahmin edici yönteminin koşul indeks değerinin en küçük olduğu görülmektedir. Bu tablodaki değerlerden anlaşılmaktadır ki, veri setinde çoklu bağlantı problemi çok az miktarda bulunmaktadır. Kovaryans matrisinin tersinin alınabildiği durumlar için incelenen bu veri setinde, BCSE daraltılmış kovaryans

tahmin edici yönteminin çoklu bağlantı problemini diğer kovaryans tahmin edici yöntemlerine göre düşürdüğü görülmüştür. Ayrıca bu veri seti incelendiğinde hem rankın 1 olduğu durumlarda hem de rankın 2 olduğu durumlar için aynı durum söz konusu olmaktadır. Bu veri setine benzer durumlar için ayrıca şöyle bir sonuca varılmıştır. Klasik kovaryans matrisinin koşul indeksi, bazı daraltılmış kovaryans tahmin edici yöntemlerinden daha yüksek olduğu için, daraltılmış kovaryans tahmin edici yöntemlerinden en uygun olan kullanılarak, çoklu bağlantı problemi giderilmeye çalışılmıştır. Bu durumu daha net bir şekilde anlayabilmek için koşul indeks değerlerinin de gösterildiği kutu grafiğinin incelenmesi, veri seti hakkında bilgi sahibi olmada faydalı olmaktadır. Buna göre,



Şekil 4.7. Epilepsi veri seti için rank değerlerinin 1 ve 2 olduğu durumlarda daraltılmış kovaryans tahmin edici yöntemlerinin etkisiyle oluşturulmuş olan koşul indeks değerlerinin kutu grafikleri

Şekil 4.7. de Epilepsi veri setine ait koşul indeks değerlerinin kutu grafikleri gösterilmiştir. Yukarıdaki Şekil 4.7. de görüldüğü üzere, daraltılmış kovaryans tahmin edici yöntemlerinden BCSE kovaryans tahmin edicisi, diğer tahmin edici yöntemler göre, veri seti üzerindeki çoklu bağlantı problemini düşürmektedir. Bu veri setindeki analiz sonuçlarına göre, kovaryans matrisinin tersinin hesaplanabildiği ya da çoklu bağlantı probleminin ortaya çıktığı durumlarda, daraltılmış kovaryans tahmin edici yöntemlerinden BCSE tahmin edici yönteminin faydalı olacağı görülmüştür. Epilepsi veri seti için elde edilmiş olan analiz sonuçları burada tamamlanmış bulunmaktadır. Diğer bir uygulama örneği ise LASERI veri setidir. Bu veri seti için elde edilmiş olan sonuçlar aşağıdaki gibi gösterilmiştir.

Tablo 4.30. LASERI veri seti üzerinden elde edilmiş olan, rankın 1, 2, 3, 4, 5, 6 ve 7 değerlerini aldığı, daraltılmış kovaryans tahmin edici yöntemlerinin etkisiyle oluşturulmuş bilgi kriterleri değerleri

| Rank değerleri | Daraltılmış kovaryans tahmin edici yöntemleri | AIC_q | AIC_{cq} | BIC_q | $HQIC_q$ | KIC_{vc} |
|----------------|---|-----------|------------|-----------|-----------|------------|
| 1 | MLE/EB | 5521.1340 | 6350.0660 | 5846.7370 | 5619.1840 | 1044.8000 |
| | SRE | 5521.1340 | 6350.0650 | 5846.7370 | 5619.1840 | 1044.8000 |
| | SDE | 5521.2450 | 6350.1650 | 5846.8480 | 5619.2960 | 1044.7990 |
| | BCSE | 5321.3000 | 6171.7380 | 5646.9030 | 5419.3510 | 1046.5930 |
| | CSE | 5513.6330 | 6343.3720 | 5839.2370 | 5233.6330 | 1044.8680 |
| | OAS | 5513.3330 | 6343.1050 | 5838.9360 | 5611.3840 | 1044.8700 |
| 2 | MLE/EB | 5601.4160 | 6421.7080 | 5927.0190 | 5699.4670 | 1044.0800 |
| | SRE | 5601.4160 | 6421.7070 | 5927.0190 | 5699.4660 | 1044.0800 |
| | SDE | 5600.3900 | 6420.7920 | 5925.9930 | 5698.4410 | 1044.0900 |
| | BCSE | 5489.8370 | 6322.1370 | 5815.4400 | 5587.8880 | 1045.0810 |
| | CSE | 5597.1400 | 6417.8920 | 5922.7430 | 5317.1400 | 1044.1190 |
| | OAS | 5596.9690 | 6417.7390 | 5922.5720 | 5695.0190 | 1044.1200 |
| 3 | MLE/EB | 5602.3210 | 6422.5160 | 5927.9250 | 5700.3720 | 1044.0720 |
| | SRE | 5602.3210 | 6422.5150 | 5927.9240 | 5700.3720 | 1044.0720 |
| | SDE | 5601.3230 | 6421.6250 | 5926.9260 | 5699.3740 | 1044.0810 |
| | BCSE | 5491.0300 | 6323.2010 | 5816.6330 | 5589.0800 | 1045.0700 |
| | CSE | 5598.0560 | 6418.7090 | 5923.6590 | 5318.0560 | 1044.1110 |
| | OAS | 5597.8850 | 6418.5570 | 5923.4880 | 5695.9360 | 1044.1120 |
| 4 | MLE/EB | 5601.8270 | 6422.0750 | 5927.4310 | 5699.8780 | 1044.0770 |
| | SRE | 5601.8270 | 6422.0740 | 5927.4300 | 5699.8780 | 1044.0770 |
| | SDE | 5600.8510 | 6421.2040 | 5926.4550 | 5698.9020 | 1044.0860 |
| | BCSE | 5490.3990 | 6322.6380 | 5816.0020 | 5588.4490 | 1045.0760 |
| | CSE | 5597.5570 | 6418.2640 | 5923.1600 | 5317.5570 | 1044.1150 |
| | OAS | 5597.3860 | 6418.1110 | 5922.9890 | 5695.4370 | 1044.1170 |
| | MLE/EB | 5601.8300 | 6422.0770 | 5927.4330 | 5699.8800 | 1044.0770 |

| | | | | | | |
|---|--------|-----------|-----------|-----------|-----------|-----------|
| | SRE | 5601.8300 | 6422.0770 | 5927.4330 | 5699.8800 | 1044.0770 |
| 5 | SDE | 5600.8540 | 6421.2060 | 5926.4570 | 5698.9050 | 1044.0860 |
| | BCSE | 5490.4020 | 6322.6410 | 5816.0050 | 5588.4520 | 1045.0760 |
| | CSE | 5597.5590 | 6418.2660 | 5923.1630 | 5317.5590 | 1044.1150 |
| | OAS | 5597.3890 | 6418.1140 | 5922.9920 | 5695.4390 | 1044.1170 |
| | MLE/EB | 5601.8290 | 6422.0760 | 5927.4330 | 5699.8800 | 1044.0770 |
| | SRE | 5601.8290 | 6422.0760 | 5927.4320 | 5699.8800 | 1044.0770 |
| 6 | SDE | 5600.8540 | 6421.2060 | 5926.4570 | 5698.9040 | 1044.0860 |
| | BCSE | 5490.4010 | 6322.6400 | 5816.0040 | 5588.4520 | 1045.0760 |
| | CSE | 5597.5590 | 6418.2660 | 5923.1620 | 5317.5590 | 1044.1150 |
| | OAS | 5597.3880 | 6418.1130 | 5922.9910 | 5695.4390 | 1044.1170 |
| | MLE/EB | 5601.8290 | 6422.0760 | 5927.4330 | 5699.8800 | 1044.0770 |
| | SRE | 5601.8290 | 6422.0760 | 5927.4320 | 5699.8800 | 1044.0770 |
| 7 | SDE | 5600.8540 | 6421.2060 | 5926.4570 | 5698.9040 | 1044.0860 |
| | BCSE | 5490.4010 | 6322.6400 | 5816.0040 | 5588.4520 | 1045.0760 |
| | CSE | 5597.5590 | 6418.2660 | 5923.1620 | 5317.5590 | 1044.1150 |
| | OAS | 5597.3880 | 6418.1130 | 5922.9910 | 5695.4390 | 1044.1170 |

Tablo 4.30. da LASERI veri setine ait olan daraltılmış kovaryans tahmin edici yönteminin etkisiyle oluşturulmuş olan bilgi kriterleri gösterilmektedir. Buna göre Tablo 4.30. da gösterilen analiz sonuçlarında görülmektedir ki, mevcut olan üç bilgi kriterinin BCSE daraltılmış kovaryans tahmin edici yöntemini, çok değişkenli indirgenmiş rank regresyon yönteminde en iyi kovaryans tahmin edici olarak göstermektedir. Rank değerlerinin değiştiği durumlarda dahi bu sonuç değişmemiştir. Ancak, rank değerlerinin değiştiği durumlarda bile $HQIC_q$ bilgi kriterinin, CSE daraltılmış kovaryans tahmin edici yöntemini, regresyon yöntemi için en iyi kovaryans tahmin edici olarak görürken, KIC_{vc} bilgi kriteri ise, rank değerinin 1 olduğu durumda SDE daraltılmış kovaryans tahmin edici yöntemini; rank değerinin diğer tüm durumlarında ise MLE / EB ve SRE daraltılmış kovaryans tahmin edici yöntemlerini en iyi kovaryans tahmin edici yöntemi olarak görmektedir. Bu sonuçlar doğrultusunda

daraltılmış kovaryans tahmin edici yöntemlerinin hata kareler ortalama değerleri aşağıdaki gibi elde edilmiştir.

Tablo 4.31. LASERI veri seti için rankın **1, 2, 3, 4, 5, 6 ve 7** değerlerini aldığı durumlarda daraltılmış kovaryans tahmin edici yöntemleri kullanılarak elde edilen hata kareler ortalama değerleri

| Rank Değerleri | Daraltılmış kovaryans tahmin edici yöntemleri | Eğitim verilerine ait hata kareler ortalamaları | Test verilerine ait hata kareler ortalamaları |
|-----------------------|--|--|--|
| 1 | MLE/EB | 1375.235 | 1419.768 |
| | SRE | 1375.235 | 1419.767 |
| | SDE | 1375.407 | 1420.070 |
| | BCSE | 1099.214 | 1189.363 |
| | CSE | 1363.720 | 1410.731 |
| | OAS | 1363.261 | 1410.369 |
| 2 | MLE/EB | 1504.751 | 1501.941 |
| | SRE | 1504.750 | 1501.941 |
| | SDE | 1503.020 | 1500.952 |
| | BCSE | 1327.820 | 1355.365 |
| | CSE | 1497.554 | 1496.208 |
| | OAS | 1497.267 | 1495.981 |
| 3 | MLE/EB | 1506.279 | 1496.176 |
| | SRE | 1506.278 | 1496.176 |
| | SDE | 1504.594 | 1495.273 |
| | BCSE | 1329.597 | 1354.614 |
| | CSE | 1499.093 | 1490.570 |
| | OAS | 1498.806 | 1490.345 |
| 4 | MLE/EB | 1505.445 | 1486.927 |
| | SRE | 1505.444 | 1486.926 |
| | SDE | 1503.798 | 1486.465 |
| | BCSE | 1328.656 | 1333.559 |
| | CSE | 1498.255 | 1480.876 |
| | OAS | 1497.968 | 1480.634 |

| | | | |
|---|--------|----------|----------|
| | MLE/EB | 1505.449 | 1485.488 |
| | SRE | 1505.448 | 1485.488 |
| 5 | SDE | 1503.803 | 1484.944 |
| | BCSE | 1328.661 | 1344.800 |
| | CSE | 1498.259 | 1479.852 |
| | OAS | 1497.972 | 1479.626 |
| | MLE/EB | 1505.448 | 1484.045 |
| | SRE | 1505.447 | 1484.045 |
| 6 | SDE | 1503.802 | 1482.748 |
| | BCSE | 1328.660 | 1330.484 |
| | CSE | 1498.258 | 1478.004 |
| | OAS | 1497.971 | 1477.763 |
| | MLE/EB | 1505.448 | 1477.377 |
| | SRE | 1505.447 | 1477.377 |
| 7 | SDE | 1503.802 | 1476.544 |
| | BCSE | 1328.660 | 1323.417 |
| | CSE | 1498.258 | 1471.307 |
| | OAS | 1497.971 | 1471.064 |

Tablo 4.31 de LASERI veri seti için daraltılmış kovaryans tahmin edici yöntemleri kullanılarak oluşturulan hata kareler ortalama değerleri gösterilmiştir. Buna göre LASERI veri seti için oluşturulan hata kareler ortalama değerlerini incelediğimizde, $r = 1, 2, 3$ ve 4 değerleri için eğitim verileri üzerinden elde edilen sonuçlara göre BCSE ve OAS daraltılmış kovaryans tahmin ediciler diğer daraltılmış kovaryans tahmin edicilere göre model için en uygun olarak belirlenmiştir. Test verileri üzerinden elde edilmiş sonuçlar ise $r = 1, 3$ ve 4 değerlerinde BCSE ve OAS daraltılmış kovaryans tahmin edici yöntemini model üzerinde en uygun tahmin edici yöntemi olarak belirlerken $r = 2$ değerinde test verisi üzerinden elde edilen hata kareler ortalama değeri BCSE ve CSE daraltılmış kovaryans tahmin edici yöntemini model üzerinde en uygun tahmin edici olarak göstermektedir. Rank değerinin ilk dört değeri için incelenen sonuçlar bu şekilde olup rankın diğer değerleri için BCSE ve

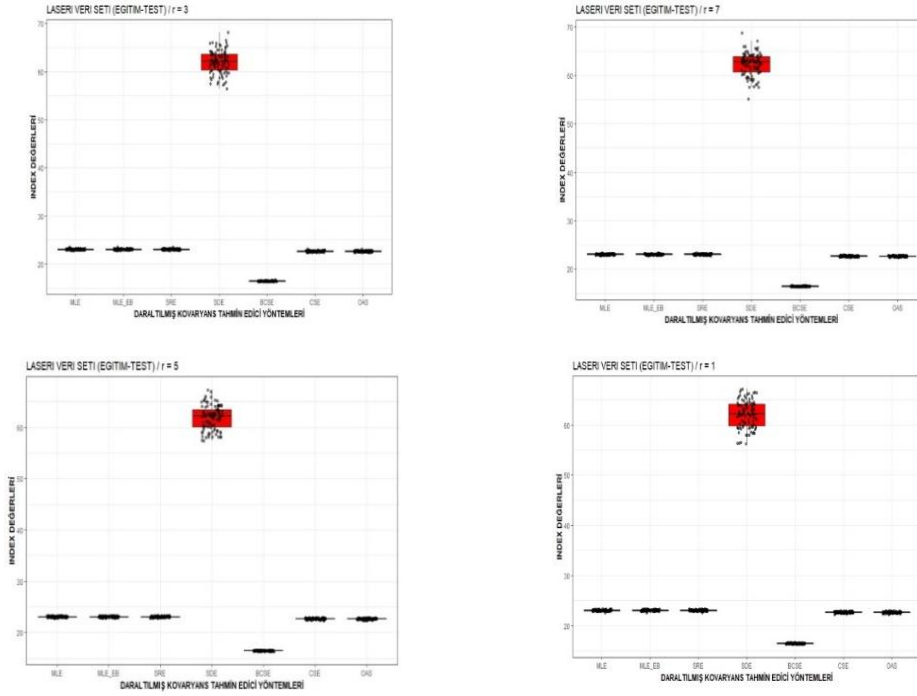
OAS daraltılmış kovaryans tahmin edici yöntemleri en uygun yöntem olarak belirlenmiştir.

Epilepsi veri seti için uygulanan koşul indeks değerleri, LASERI veri seti için de uygulanmıştır. Buna göre,

Tablo 4.32. LASERI veri seti için rank değerlerinin **1, 2, 3, 4, 5, 6 ve 7** olduğu durumlarda daraltılmış kovaryans tahmin edici yöntemlerinin etkisiyle oluşturulmuş olan koşul indeks değerleri

| Daraltılmış kovaryans tahmin edici yöntemleri | r = 1 | r = 2 | r = 3 | r = 4 | r = 5 | r = 6 | r = 7 |
|--|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
| MLE | 23.0386 | 23.0636 | 23.0681 | 23.0452 | 23.0264 | 23.0341 | 23.0684 |
| MLE/EB | 23.0386 | 23.0636 | 23.0680 | 23.0452 | 23.0264 | 23.0341 | 23.0684 |
| SRE | 23.0386 | 23.0636 | 23.0680 | 23.0452 | 23.0264 | 23.0341 | 23.0684 |
| SDE | 62.1807 | 61.7469 | 62.0114 | 62.3904 | 62.5066 | 62.2831 | 61.9104 |
| BCSE | 16.4391 | 16.4529 | 16.4554 | 16.4427 | 16.4323 | 16.4366 | 16.4556 |
| CSE | 22.6579 | 22.6837 | 22.6889 | 22.6650 | 22.6450 | 22.6519 | 22.6891 |
| OAS | 22.6430 | 22.6689 | 22.6742 | 22.6502 | 22.6300 | 22.6369 | 22.6744 |

Tablo 4.32. de LASERI veri seti için daraltılmış kovaryans tahmin edicinin etkisiyle oluşturulan koşul indeks değerleri gösterilmektedir. Burada Çok değişkenli indirgenmiş rank regresyon yönteminde de karşılaştığımız çoklu bağlantı problemi için LASERI veri seti incelenmiştir. Tablo 4.32. de LASERI veri setine ait koşul indeks değerleri yer almaktadır. Buradaki değerler incelendiğinde, daraltılmış kovaryans tahmin edici yöntemlerinden olan BCSE tahmin edici yönteminin koşul indeksi, diğer tahmin edici yöntemlerinin koşul indeks değerlerinden küçük olduğu görülmüştür. Yukarıdaki değerlere göre klasik kovaryans tahmin edici yönteminin koşul indeksinden daha küçük olan BCSE tahmin edicisinin koşul indeksi çoklu bağlantı problemini azalttığı görülmüştür. Böylelikle bu veri seti için orta düzeyli çoklu bağlantı sorununda, bazı daraltılmış kovaryans tahmin edici yöntemlerinin kullanılmasıyla azaldığı görülmektedir. Rank değerlerinin arttığı veya değiştiği durumlar incelendiğinde, rankın 5 olduğu bir durumda çoklu bağlantı sorununun diğer rank değerlerinde göre daha küçük olduğu görülmüştür. Tablo 4.32. de ki durumu daha net bir şekilde görebilmek adına aşağıda koşul indeks değerlerinin de yer aldığı kutu grafikleri yer almaktadır. Buna göre,



Şekil 4.8. LASERI veri seti için rank değerlerinin 1, 3, 5 ve 7 olduğu durumlarda daraltılmış kovaryans tahmin edici yöntemlerinin etkisiyle oluşturulmuş koşul indeks değerlerinin kutu grafikleri

Şekil 4.8. de LASERI veri setine ait koşul indeks değerlerini kutu grafikleri gösterilmiştir. Yukarıdaki Şekil 4.8 de gösterilen kutu grafikleri incelendiğinde, BCSE kovaryans tahmin edici yönteminin koşul indeks değerinin en küçük olduğu görülmüştür. Burada rank değerlerinin 1, 3, 5 ve 7 değerlerinin aldığı durumlar incelenmiştir. Orta düzeyli çoklu bağlantı sorununda, daraltılmış kovaryans tahmin edici yöntemlerinden BCSE tahmin edici yöntemi kullanıldığında çoklu bağlantı sorununu düşürdüğü kutu grafiğinde de görülmüştür.

Kovaryans matrisinin tersinin alınabildiği bu iki uygulama incelendiğinde şu sonuca varılmıştır. Daraltılmış kovaryans tahmin edici yöntemleri, çok değişkenli regresyon yöntemlerinde kullanıldığında var olan problemlerin giderilmesinde faydalı olmaktadır. Bu tahmin edici yöntemleri aynı zamanda kovaryans matrisinin tersinin alınamadığı durumlar için de kullanılabilir. Ancak güvenilir değildir.

5. SONUÇLAR ve ÖNERİLER

Tez çalışması boyunca elde edilen bulgular analiz edilerek konu ile ilgili bilgi verilmiştir. Öncelikle burada çok değişkenli indirgenmiş rank regresyon yöntemi için uygulamalarda gerekli olan birkaç matematiksel ifadeler yer almaktadır. Daraltılmış kovaryans tahmin edici yöntemlerinden, bilgi kriterlerinden ve çok değişkenli indirgenmiş rank regresyon yöntemi için önemli bir yere sahip olan kovaryans matrisinden bahsedilmiştir. Amacımız kovaryans matrisinin büyük boyutlu verilerde nasıl bir yol izleyeceğini göstermek olmuştur. Ayrıca daraltılmış kovaryans tahmin edici yöntemlerinin kullanılmasıyla, beta (β) katsayıları üzerinden elde edilen hata kareler ortalamasının incelenmesi de gerçekleştirdiğimiz hedeflerimizden biri olmaktadır. Burada öncelikle daraltılmış kovaryans tahmin edici yöntemlerinin kullanılmasıyla beta (β) katsayıları üzerinden elde edilen hata kareler ortalaması incelenerek, hangi daraltılmış kovaryans tahmin edici yönteminin hata kareler ortalaması üzerinde etkili olduğu incelenmiş ve içlerinden Maksimum Olabilirlik / Bayes tahmin edici yöntemi yani, MLE / EB' nin hata kareler ortalaması üzerinde daha etkili olduğuna karar verilmiştir. Daha sonra ise bilgi kriterleri açısından analiz sonuçları elde edilmiş ve daraltılmış kovaryans matrisler için bilgi kriterleri değerlerine bakılmıştır. Elde edilen sonuçlar doğrultusunda, genel bir değerlendirme sonucu olarak, Maksimum Olabilirlik / Bayes tahmin edici yönteminin çok değişkenli indirgenmiş rank regresyon yöntemi için en iyi kovaryans tahmin edici yöntemi olduğuna karar verilmiştir. Çok değişkenli indirgenmiş rank regresyon yöntemi için yapılan uygulamalarda bazı bulgular elde edilmiştir. Buna göre beta(β) katsayıları üzerinden elde edilen hata kareler ortalamaları incelenirken, daraltılmış kovaryans tahmin edici yöntemlerinden BCSE tahmin edici yönteminin, diğer daraltılmış kovaryans tahmin edici yöntemlerinin değerlerinden çok farklı olduğu gözlemlenmiştir. Bu husus ise kovaryans tahmin edici yöntemleri için belirlenmiş olan ceza katsayılarının incelenmesini ve bir değerlendirmede bulunmamızı gerektirmiştir. Ceza katsayısı adını verdiğimiz iz (*ridge*) ceza yöntemi, kovaryans matrislerinin kötü koşullandırılmış olmasından kaynaklanmaktadır. Yani kovaryans matrislerinin iz (*ridge*) cezası, büyük boyutlu veri setleri içinde, tersinin hesaplanmadığı veyahut determinantının alınmadığı durumlarda ortaya çıkmaktadır. Burada determinant değerlerinin hesaplanarak oluşturulduğu daraltılmış kovaryans tahmin edici yöntemleri için iz (*ridge*) cezasının olup olmasına bakılmıştır. Bu bağlamda ifade edilmelidir ki, BCSE daraltılmış kovaryans tahmin edici yönteminin, diğer kovaryans

tahmin edici yöntemlerinden daha fazla iz (*ridge*) cezasına sahip olduğu görülmektedir. Buna göre BCSE daraltılmış kovaryans tahmin edici yönteminin, çok değişkenli indirgenmiş rank regresyon yöntemi için uygun olmadığı görülmüştür. Genel anlamda Maksimum Olabilirlik / Bayes tahmin edici yönteminin hem beta(β) katsayıları üzerinden elde edilen hata kareler ortalamaları yönünden hem de bilgi kriterleri açısından çok değişkenli indirgenmiş rank regresyon yöntemi için daha etkili, daha uyumlu olduğuna karar verilmiştir. Bu durum ise tahmin edilen beta katsayıları için, gerçek beta katsayılarını, Maksimum Olabilirlik / Bayes tahmin edici yöntemi ile sıfıra yakın bir değer alacak şekilde daha iyi tahminde bulunabileceğini göstermiştir. Daraltılmış kovaryans tahmin edici yöntemlerinin determinant değerleri incelenirken, bazı tahmin edicilerin determinant değerlerinin normal olmayan durumlarından dolayı, ceza katsayılarının fazla olmasından şüphelenerek, görsel olarak daha net anlayabilmek için daraltılmış kovaryans tahmin edici yöntemlerinin kutu grafiklerine de bakılmıştır. Kutu grafikleri incelendiğinde, BCSE daraltılmış kovaryans tahmin edici yönteminin diğer kovaryans tahmin edici yöntemlerinden farklı bir şekilde ceza katsayısının çok fazla arttığı veya çok fazla azaldığı görülmüştür. Uygulanan veri setleri için bir de eğitim-test modeli geliştirilmiştir. Burada da incelemeler yapılmış ve aynı şekilde beta katsayıları üzerinden elde edilen hata kareler ortalama değerlerine bakılmıştır. Ayrıca burada bilgi kriterleri ile daraltılmış kovaryans tahmin edici yöntemlerin durumları da incelenmiştir. Beta katsayıları üzerinden elde edilen hata kareler ortalaması için Maksimum Olabilirlik / Bayes Tahmin edici yönteminin, hata kareler ortalaması için daha etkili olduğuna karar verilmiştir ve bilgi kriterleri yönünden incelendiğinde yine Maksimum Olabilirlik / Bayes tahmin edici yöntemi en iyi kovaryans tahmin edici yöntemi olarak gösterilmiştir. Daraltılmış kovaryans tahmin edici yönteminin etkisiyle beta katsayıları üzerinden elde edilen hata kareler ortalaması incelendiğinde görülmüştür ki, BCSE daraltılmış kovaryans tahmin edici yöntemi, aynı şekilde eğitim-test olmadan yapılan analizlerdeki gibi ceza katsayısının fazla olduğunu göstermiştir. Burada BCSE daraltılmış kovaryans tahmin edici yönteminin eğitim-test için iz (*ridge*) cezasının, diğer kovaryans tahmin edici yöntemlerine göre farklı bir şekilde çok fazla arttığı veya azaldığı görülmektedir. Bu durum ise eğitim-test için de BCSE daraltılmış kovaryans tahmin edici yönteminin uygun olmadığı görülmektedir.

Buraya kadar ki anlatılanlar ilk iki simülasyon çalışması üzerinden elde edilen analiz sonuçlarının çıkarımlarıdır. Üçüncü bir uygulama çalışması ile çok değişkenli

indirgenmiş rank regresyon yöntemi için gerçek veri setleri kullanılarak çalışılmıştır. Epilepsi ve LASERI veri setleri ile regresyon yöntemi için analiz sonuçları elde edilmiş ve bazı değerlendirmelerde bulunulmuştur. Yapılan çalışma sonucunda bu iki veri seti için BCSE daraltılmış kovaryans tahmin edici yönteminin veri seti üzerinde daha etkili olduğuna karar verilmiştir. Bilindiği üzere, regresyon yöntemlerinde bazı sorunlar ortaya çıkmaktadır. Bu sorunlardan birkaçı, çoklu bağlantı ve kovaryans matrisi üzerinde yaşanan sorunlardır. Bu sorunları düzeltebilmek veya azaltabilmek için daraltılmış kovaryans tahmin edici yöntemlerine ihtiyaç duyulmuştur. Bu iki veri setinde çoklu bağlantı problemi bulunmaktadır. Aynı zamanda veri setleri kovaryans matrislerinin tersinin alınabildiği durumlar için örnek oluşturmaktadır. Böyle bir durumda daraltılmış kovaryans tahmin edici yöntemlerinden yararlanılarak bazı değerlendirmelerde bulunulmuştur. Öncelikle veri setleri üzerinde çoklu bağlantı sorununu daha net bir şekilde görebilmek için koşul indeks değerleri hesaplanmıştır. Daraltılmış kovaryans tahmin edici yöntemlerinin etkisiyle hesaplanan koşul indeks değerleriyle anlaşılmaktadır ki, klasik kovaryans matrisinden daha uygun veya daha etkili olan daraltılmış kovaryans tahmin edici yöntemleri bulunmaktadır. Buradan çoklu bağlantının olduğu veri setlerinde, daraltılmış kovaryans tahmin edici yöntemlerinin sorunu azalttığı görülmüştür. Bu değerlendirmelerden sonra, kovaryans matrisinin tersinin alınamadığı durumlarda da bu yöntemin uygulanabileceği önerilmektedir. Ancak güvenilir olup olmadığı konusunda temkinli olmak gerekmektedir. Ayrıca çok değişkenli indirgenmiş rank regresyon yöntemi için daraltılmış kovaryans tahmin edici yöntemleri kullanılarak rank değerlerinin tahmin edilmesi de önerilmektedir. Son olarak; iki simülasyon çalışmasında daraltılmış kovaryans tahmin edici yöntemlerinden MLE / EB kovaryans tahmin edici yönteminin çok değişkenli indirgenmiş rank regresyon yöntemi için daha etkili veya daha uygun bir model olduğuna karar verilmişken, bu durum gerçek bir veri seti uygulamasında farklılık göstermiştir. BCSE tahmin edici yönteminin, Epilepsi ve LASERI veri setleri için daha uygun olduğu görülmüştür.

KAYNAKÇA

- Algamal, Z. Y. (2018). Shrinkage estimators for gamma regression model. *Electronic Journal of Applied Statistical Analysis*, 11(1), 253-268.
- Aljandali, A. (2017). *Multivariate methods and forecasting with IBM® SPSS® statistics*. New York, NY: Springer.
- Anderson, T. W. (1999). Asymptotic distribution of the reduced rank regression estimator under general conditions. *The Annals of Statistics*, 27(4), 1141-1154.
- Bedrick, E. J., and Tsai, C. L. (1994). Model selection for multivariate regression in small samples. *Biometrics*, 226-231.
- Bernardini, E., and Cubadda, G. (2015). Macroeconomic forecasting and structural analysis through regularized reduced-rank regression. *International Journal of Forecasting*, 31(3), 682-691.
- Bozdogan, H. (2003). Intelligent Statistical Data Mining with Information Complexity and Genetic Algorithms Hamparsum Bozdogan University of Tennessee, Knoxville, USA. In *Statistical data mining and knowledge discovery* (pp. 47-88). Chapman and Hall/CRC.
- Braak, C. J. T., and Looman, C. W. (1994). Biplots in reduced-rank regression. *Biometrical journal*, 36(8), 983-1003.
- Bura, E., and Cook, R. D. (2003). Rank estimation in reduced-rank regression. *Journal of Multivariate Analysis*, 87(1), 159-176.
- Burnham, A. J., Viveros, R., and MacGregor, J. F. (1996). Frameworks for latent variable multivariate regression. *Journal of chemometrics*, 10(1), 31-45.
- Chen, Y., Wiesel, A., Eldar, Y. C., and Hero, A. O. (2010). Shrinkage algorithms for MMSE covariance estimation. *IEEE Transactions on Signal Processing*, 58(10), 5016-5029.
- Cook, R. D., Forzani, L., and Zhang, X. (2015). Envelopes and reduced-rank regression. *Biometrika*, 102(2), 439-456.
- De Cao, N., and Aziz, W. (2020). The power spherical distribution. *arXiv preprint arXiv:2006.04437*.
- Dhillon, I. S., and Sra, S. (2003). *Modeling data using directional distributions* (pp. 03-06). Technical Report TR-03-06, Department of Computer Sciences, The University of Texas at Austin. URL <ftp://ftp.cs.utexas.edu/pub/techreports/tr03-06.ps.gz>.
- Erbas, Ü. (2010). *Entropi İlkelerinin Boyut İndirgeme Uygulamaları* (Doctoral dissertation, Doktora Tezi).
- Fan, J., Fan, Y., and Lv, J. (2008). High dimensional covariance matrix estimation using a factor model. *Journal of Econometrics*, 147(1), 186-197.
- Feng, Z., and Li, P. (2006, November). Performance-oriented statistical parameter reduction of parameterized systems via reduced rank regression. In *2006 IEEE/ACM International Conference on Computer Aided Design* (pp. 868-875). IEEE.
- Fisher, T. J., and Sun, X. (2011). Improved Stein-type shrinkage estimators for the high-dimensional multivariate normal covariance matrix. *Computational Statistics and Data Analysis*, 55(5), 1909-1918.
- Heinen, A., and Rengifo, E. (2008). Multivariate reduced rank regression in non-Gaussian contexts, using copulas. *Computational statistics and data analysis*, 52(6), 2931-2944.
- Hoff, P. D. (2009). Simulation of the matrix Bingham–von Mises–Fisher distribution, with applications to multivariate and relational data. *Journal of Computational and Graphical Statistics*, 18(2), 438-456.

- Hoff, P. D. (2019). Matrix modeling with rstiefel.
- Huang, J. Z., Liu, N., Pourahmadi, M., and Liu, L. (2006). Covariance matrix selection and estimation via penalised normal likelihood. *Biometrika*, 93(1), 85-98.
- Izenman, A. J. (2008). Modern multivariate statistical techniques. *Regression, classification and manifold learning*, 10, 978-0.
- Koc, E. K., and Bozdogan, H. (2015). Model selection in multivariate adaptive regression splines (MARS) using information complexity as the fitness function. *Machine Learning*, 101(1), 35-58.
- Markdahl, J., Thunberg, J., and Goncalves, J. (2020). High-dimensional Kuramoto models on Stiefel manifolds synchronize complex networks almost globally. *Automatica*, 113, 108736.
- Maronna, M., and Suggests, M. A. S. S. (2009). Package ‘robustbase’.
- Miyashiro, R., and Takano, Y. (2015). Subset selection by Mallows’ Cp: A mixed integer programming approach. *Expert Systems with Applications*, 42(1), 325-331
- Mohebbi, S., Pamukcu, E., and Bozdogan, H. (2019). A new data adaptive elastic net predictive model using hybridized smoothed covariance estimators with information complexity. *Journal of Statistical Computation and Simulation*, 89(6), 1060-1089.
- Montgomery, D. C., and Runger, G. C. (2010). *Applied statistics and probability for engineers*. John Wiley and Sons.
- Mukherjee, A. (2013). *Topics on Reduced Rank Methods for Multivariate Regression* (Doctoral dissertation).
- Nordhausen, K., Sirkia, S., Oja, H., Tyler, D. E., and Nordhausen, M. K. (2018). Package ‘ICSNP’. *R 2165 package version 1.1*, 1, 2166.
- Pamukçu, E. (2017). A new hybrid regression model for undersized sample problem. *Celal Bayar University Journal of Science*, 13(3), 803-813.
- Pamukçu, E. (2019). Choosing the optimal hybrid covariance estimators in adaptive elastic net regression models using information complexity. *Journal of Statistical Computation and Simulation*, 89(16), 2983-2996.
- Perez-Melo, S., and Kibria, B. M. (2020). On some test statistics for testing the regression coefficients in presence of multicollinearity: a simulation study. *Stats*, 3(1), 40-55.
- Portet, S. (2020). A primer on model selection using the Akaike Information Criterion. *Infectious Disease Modelling*, 5, 111-128.
- Rao, C. R., Wu, Y., Konishi, S., and Mukerjee, R. (2001). On model selection. *Lecture Notes-Monograph Series*, 1-64.
- Riley, J. D. (1955). Solving systems of linear equations with a positive definite, symmetric, but possibly ill-conditioned matrix. *Mathematical tables and other aids to computation*, 96-101.
- Schäfer, J., and Strimmer, K. (2005). A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. *Statistical applications in genetics and molecular biology*, 4(1).
- Schmidli, H. (2013). *Reduced Rank Regression: With applications to quantitative structure-activity relationships*. Springer Science and Business Media.
- Shrestha, N. (2020). Detecting multicollinearity in regression analysis. *American Journal of Applied Mathematics and Statistics*, 8(2), 39-42.
- Seghouane, A. K. (2006). Multivariate regression model selection from small samples using Kullback's symmetric divergence. *Signal processing*, 86(8), 2074-2084.

- Similä, T., and Tikka, J. (2007). Input selection and shrinkage in multiresponse linear regression. *Computational Statistics and Data Analysis*, 52(1), 406-422.
- Tabachnick, B. G., Fidell, L. S., and Ullman, J. B. (2007). *Using multivariate statistics* (Vol. 5, pp. 481-498). Boston, MA: Pearson.
- Thall, P. F., and Vail, S. C. (1990). Some covariance models for longitudinal count data with overdispersion. *Biometrics*, 657-671.
- Timm, N. H. (2002). *Applied multivariate analysis*. Springer.
- Tso, M. S. (1981). Reduced-rank regression and canonical analysis. *Journal of the Royal Statistical Society: Series B (Methodological)*, 43(2), 183-189.
- Van Wieringen, W. N., and Peeters, C. F. (2016). Ridge(iz) estimation of inverse covariance matrices from high-dimensional data. *Computational Statistics and Data Analysis*, 103, 284-303.
- Velu, R. P., and Reinsel, G. C. (1987). Reduced rank regression with autoregressive errors. *Journal of econometrics*, 35(2-3), 317-335.
- Velu, R., and Reinsel, G. C. (2013). *Multivariate reduced-rank regression: theory and applications* (Vol. 136). Springer Science and Business Media.
- Vershynin, R. (2012). How close is the sample covariance matrix to the actual covariance matrix? *Journal of Theoretical Probability*, 25(3), 655-686.
- Yee, T. W., and Hastie, T. J. (2003). Reduced-rank vector generalized linear models. *Statistical modelling*, 3(1), 15-41.
- Zhang, J., and Ibrahim, M. (2005). A simulation study on SPSS ridge(iz) regression and ordinary least squares regression procedures for multicollinearity data. *Journal of Applied Statistics*, 32(6), 571-588.

ÖZGEÇMİŞ

Adı ve Soyadı : Salih ARPAGUŞ
Yabancı Dili : İngilizce
Eğitim Durumu
Lise : 19 Mayıs Lisesi
Lisans : 19 Mayıs Üniversitesi Fen Edebiyat Fakültesi
İstatistik Bölümü
Yüksek Lisans :19 Mayıs Üniversitesi Lisansüstü Eğitim
Enstitüsü İstatistik Ana Bilim Dalı
İletişim Bilgileri
ORCID ID : 0000 – 0002 – 5936 – 0655