

**UKUROVA NİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

DOKTORA TEZİ

Tuğba SÖKÜT AÇAR

**GENEL LİNEER İSTATİSTİKSEL MODELDE YANLI
TAHMİN EDİCİLER VE TANILAMA METOTLARI**

İSTATİSTİK ANABİLİM DALI

ADANA-2016

**ÇUKUROVA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

**GENEL LİNEER İSTATİSTİKSEL MODELDE YANLI TAHMİN
EDİCİLER VE TANILAMA METOTLARI**

Tuğba SÖKÜT AÇAR

DOKTORA TEZİ

İSTATİSTİK ANABİLİM DALI

Bu Tez 22/11/2016 Tarihinde Aşağıdaki Jüri Üyeleri Tarafından
Oybirliği/Oyçokluğu ile Kabul Edilmiştir.

.....
Prof. Dr. Mahmude Revan ÖZKALE
DANIŞMAN

.....
Prof. Dr. Selahattin KAÇIRANLAR
ÜYE

.....
Prof. Dr. Aylin ALIN
ÜYE

.....
Prof. Dr. Meral ÇETİN
ÜYE

.....
Doç. Dr. Ebru ÖZGÜR GÜLER
ÜYE

Bu Tez Enstitümüz İstatistik Anabilim Dalında hazırlanmıştır.

Kod No:

**Prof. Dr. Mustafa GÖK
Enstitü Müdürü**

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

ÖZ

DOKTORA TEZİ

**GENEL LİNEER İSTATİSTİKSEL MODELDE YANLI TAHMİN
EDİCİLER VE TANILAMA METOTLARI**

Tuğba SÖKÜT AÇAR

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

Danışman : Prof. Dr. Mahmude Revan ÖZKALE
Yıl: 2016, Sayfa: 148
Jüri : Prof. Dr. M. Revan ÖZKALE
: Prof. Dr. Selahattin KAÇIRANLAR
: Doç. Dr. Ebru ÖZGÜR GÜLER
: Prof. Dr. Meral ÇETİN
: Prof. Dr. Aylin ALIN

Regresyon tanılama metotları, kaldırıcı/etkin gözlemlerin belirlenmesine yönelik ilişkisiz ve sabit varyanslı lineer istatistiksel model altında sıklıkla ele alınmıştır. Ancak, gerçek uygulamalarda modele ait hata terimleri her zaman ilişkisiz ve sabit varyanslı olmayabilir. Bu araştırma otokorelasyon ve çoklu iç ilişki problemine sahip lineer istatistiksel modelde regresyon tanılama metotlarını içermektedir. Hatalar arasındaki ilişki, birinci dereceden otoregresif süreç ile modellenmiştir. İlişkili hataya sahip lineer istatistiksel modelde genelleştirilmiş en küçük kareler tahmin edici ve ridge regresyon tahmin edici kullanarak etki tanılama yöntemleri teorik olarak incelenmiştir. Teorik incelemelerde kaldırıcı gözlem izdüşüm/yarı izdüşüm matrisinin köşegen elemanları ile, regresyon parametresi üzerindeki etkin gözlem DFBETA ile, yanıt değişken üzerindeki etkin gözlem DFFITS ile, regresyon eğrisi üzerindeki etkin gözlem COOK uzaklığı ile, güven elipsoidi üzerindeki etkin gözlem Cook-Weisberg ve Andrews-Predigon ile, regresyon parametresinin varyans-kovaryans matrisi üzerindeki etkin gözlem COVRATIO ile belirlenmiştir. Model geçerliliği OCV, GCV ve C_p istatistikleri ile incelenmiştir. Teorik incelemelere ek olarak ortak bir nümerik örnek ve her bir başlık için Monte Carlo simülasyon çalışması gerçekleştirilmiştir.

Anahtar Kelimeler: Otokorelasyon, Çoklu İç İlişki, Kaldırıcı, Etkinlik, Model Geçerliliği

ABSTRACT

PhD THESIS

BIASED ESTIMATORS AND DIAGNOSTIC METHODS IN GENERAL LINEAR STATISTICAL MODEL

Tuğba SÖKÜT AÇAR

ÇUKUROVA UNIVERSITY
INSTITUTE OF NATURAL AND APPLIED SCIENCES
DEPARTMENT OF STATISTICS

Supervisor : Prof. Dr. Mahmude Revan ÖZKALE
Year: 2016, Pages: 148
Jury : Prof. Dr. M. Revan ÖZKALE
: Prof. Dr. Selahattin KAÇIRANLAR
: Assoc. Prof. Dr. Ebru ÖZGÜR GÜLER
: Prof. Dr. Meral ÇETİN
: Prof. Dr. Aylin ALIN

Regression diagnostics methods have been often studied for determination of leverage/influential observations under the linear statistical model with uncorrelated and homoscedastic errors. However, the errors of the model may not be always uncorrelated and homoscedastic in real applications. This research includes regression diagnostics methods in the linear statistical model with correlated errors and correlated regressors. The correlation between the error terms is modeled by first order autoregressive process. Regression diagnostics methods have been studied theoretically by using generalized least square and ridge regression estimators. Leverage observation is determined by projection/quasi-projection matrix, influential observation on the regression coefficients is determined by DFBETA, on the response vector is determined by DFFITS, on the regression curve is determined by Cook distance, on the confidence ellipsoid is determined by Cook-Weisberg and Andrews-Pregibon, on the variance-covariance matrix of the regression parameters are determined by COVRATIO. Model validation is examined by OCV, GCV and Cp statistics. In addition to the theoretical investigations, a common numerical example and Monte Carlo simulation study for each topic have been performed.

Key Words: Autocorrelation, Multicollinearity, Leverage, Influence, Model Validation

TEŞEKKÜR

Öğrenimim süresince değerli bilgi ve deneyimlerinden yararlandığım, insani ve ahlaki değerleri ile örnek edindiğim, bilgi ve deneyimlerini her zaman benimle cömertçe paylaşan, birlikte çalışmaktan onur duyduğum saygıdeğer hocam Prof. Dr. Mahmude Revan ÖZKALE'ye teşekkürü bir borç bilirim.

Tezin izleme ve değerlendirme sürecinde katkıları ile çalışmamı yönlendiren değerli tez izleme komite üyeleri Prof. Dr. Selahattin KAÇIRANLAR ve Doç. Dr. Ebru ÖZGÜR GÜLER'e sonsuz teşekkürlerimi sunarım.

Doktora programını başka bir ilden sürdürmek oldukça güçtür. Verdikleri destekler ile bu güçlükleri aşmamı sağlayan Çanakkale Onsekiz Mart Üniversitesi İstatistik Bölüm Başkanı Yrd. Doç. Dr. Burcu MESTAV başta olmak üzere bana güvenen Çanakkale Onsekiz Mart Üniversitesi akademik ve idari personeline çok teşekkür ederim.

Yardımları ile sürekli yanımda olan değerli arkadaşım Arş. Gör. Dr. Selma TOKER'e teşekkür ederim.

Son olarak sonsuz teşekkürlerim Adana'daki evim olan, sıcacık yuvalarında her zaman yerim olduğunu bildiğim çok kıymetli ZÜMRÜT ailesine, doktora eğitimime başladığım günden bu yana çalışmamın her aşamasında bana büyük bir özveri ile destek veren sevgili eşim Erhan AÇAR'a, eğitim hayatım boyunca maddi ve manevi destek sağlayarak, aldığım her kararın arkasında olan ve kendileri ile gurur duyduğum sevgili ailemedir.

Sizlerin desteği olmasa bu süreç çok daha zor geçerdi.

İÇİNDEKİLER

SAYFA

ÖZ.....	I
ABSTRACT	II
TEŞEKKÜR	III
İÇİNDEKİLER.....	IV
ÇİZELGELER DİZİNİ.....	VI
SİMGELER VE KISALTMALAR	X
1. GİRİŞ.....	1
2. İLİŞKİSİZ VE SABİT VARYANSLI HATAYA SAHİP LİNEER İSTATİSTİKSEL MODEL	5
2.1. Çoklu İç İlişki Problemi ve Sıradan Ridge Regresyon Tahmin Edici	6
2.2. Rezidülere Dayalı Ölçümler	7
2.3. Aykırı ve Etkin Gözlem.....	9
2.4. Etkinlik Eğrisine Dayalı Ölçümler	10
2.5. Güven Elipsoidine Dayalı Ölçümler.....	12
2.6. Model Geçerliliği.....	15
3. GENEL LİNEER İSTATİSTİKSEL MODEL VE ÇOKLU İÇ İLİŞKİ.....	19
3.1. Otokorelesyon Kaynağı ve Belirlenmesi.....	19
3.2. Otokorelasyon Altında Çoklu İç İlişkinin Belirlenmesi	25
3.3. Yanlı Tahmin Ediciler	26
4. GENEL LİNEER İSTATİSTİKSEL MODELDE TANILAMA METOTLARI	29
4.1. Kaldıraç Gözlem.....	30
4.2. Regresyon Parametrelerine Dayalı Tanılama	31
4.2.1. GEKK için Bazı Özel Durumlar.....	33
4.3. Yanıt Değişkene Dayalı Tanılama.....	34
4.3.1. DFFITS.....	34
4.3.2. Cook Uzaklığı.....	36
4.4. Güven Elipsoidine Dayalı Tanılama.....	36

4.4.1. COVRATIO	37
4.4.2. Cook-Weisberg.....	39
4.4.3. Andrews Pregibon	44
5. GENEL LİNEER İSTATİSTİKSEL MODELDE MODEL GEÇERLİLİĞİ.....	49
5.1. Sıradan Çapraz Geçerlilik, OCV	49
5.2. Genelleştirilmiş Çapraz Geçerlilik, GCV	52
5.3. Cp	54
6. NÜMERİK ÖRNEK: FRANSA EKONOMİ VERİSİ	57
7. MONTE CARLO SİMÜLASYON UYGULAMASI	93
7.1. Regresyon Katsayıları ve Yanıt Değişkene Dayalı Tanılama için Monte Carlo Simülasyon.....	93
7.2. Güven Elipsoidine Dayalı Ölçümler için Monte Carlo Simülasyon.....	110
7.3. Model Geçerliliği için Monte Carlo Simülasyon	124
8. SONUÇ VE ÖNERİLER	133
KAYNAKLAR.....	137
ÖZGEÇMİŞ.....	143
EKLER	144

ÇİZELGELER DİZİNİ

SAYFA

Çizelge 6.1.	Fransa ekonomi verisi	57
Çizelge 6.2.	Fransa ekonomi verisi için $h_{i,i}$ ve $h_{i,i}^R$ değerleri	64
Çizelge 6.3.	Fransa ekonomi verisi için $DFBETA_i$ ve $DFBETA_i^R$ değerleri.....	64
Çizelge 6.4.	Fransa ekonomi verisi için $DFFITS_i$ ve $DFFITS_i^R$ değerleri.....	65
Çizelge 6.5.	Fransa ekonomi verisi için D_i ve D_i^R değerleri.....	66
Çizelge 6.6.	Fransa ekonomi verisi için $COVRATIO_i$ ve $COVRATIO_i^R$ değerleri	66
Çizelge 6.7.	Fransa ekonomi verisi için CW_i ve CW_i^R değerleri	67
Çizelge 6.8.	Fransa ekonomi verisi için AP_i ve AP_i^R değerleri	67
Çizelge 6.9.	Fransa ekonomi verisi rezidüleri	70
Çizelge 6.10.	Fransa ekonomi verisi için $OCV, OCV^R, GCV, GCV^R, C_p$ ve C_p^R değerleri	71
Çizelge 6.11.	Fransa Ekonomi verisi için k_{opt} ile uydurulan modelin geçerliliği.....	74
Çizelge 7.1.	Farklı ρ, γ ve n için 1-inci gözleme ait kaldırmaç değerler.....	95
Çizelge 7.2.	Farklı ρ, γ ve n için 7-inci gözleme ait kaldırmaç değerler.....	96
Çizelge 7.3.	Farklı ρ, γ ve n için 1-inci gözleme ait DFFITS değerleri.....	97
Çizelge 7.4.	Farklı ρ, γ ve n için 7-inci gözleme ait DFFITS değerleri.....	98
Çizelge 7.5.	Farklı ρ, γ ve n için 1-inci gözleme ait Cook Uzaklığı.....	99
Çizelge 7.6.	Farklı ρ, γ ve n için 7-inci gözleme ait Cook Uzaklığı.....	100

Çizelge 7.7. Farklı ρ , γ ve n için $DFBETA_{0,1}$ ve $DFBETA_{0,1}^R$ değerleri.....	101
Çizelge 7.8. Farklı ρ , γ ve n için $DFBETA_{0,7}$ ve $DFBETA_{0,7}^R$ değerleri.....	102
Çizelge 7.9. Farklı ρ , γ ve n için $DFBETA_{1,1}$ ve $DFBETA_{1,1}^R$ değerleri.....	103
Çizelge 7.10. Farklı ρ , γ ve n için $DFBETA_{1,7}$ ve $DFBETA_{1,7}^R$ değerleri.....	104
Çizelge 7.11. Farklı ρ , γ ve n için $DFBETA_{2,1}$ ve $DFBETA_{2,1}^R$ değerleri.....	105
Çizelge 7.12. Farklı ρ , γ ve n için $DFBETA_{2,7}$ ve $DFBETA_{2,7}^R$ değerleri.....	106
Çizelge 7.13. Farklı ρ , γ ve n için $DFBETA_{3,1}$ ve $DFBETA_{3,1}^R$ değerleri.....	107
Çizelge 7.14. Farklı ρ , γ ve n için $DFBETA_{3,7}$ ve $DFBETA_{3,7}^R$ değerleri.....	108
Çizelge 7.15. $\gamma_2 = 0.99$ iken OCV^R, GCV^R, C_p^R değerleri ve ilgili optimum ridge yanlılık parametreleri.....	129
Çizelge 7.16. $\gamma_2 = 0.90$ iken OCV^R, GCV^R, C_p^R değerleri ve ilgili optimum ridge yanlılık parametreleri.....	129
Çizelge 7.17. $\gamma_2 = 0.70$ iken OCV^R, GCV^R, C_p^R değerleri ve ilgili optimum ridge yanlılık parametreleri.....	130
Çizelge 7.18. $\gamma_2 = 0.50$ iken OCV^R, GCV^R, C_p^R değerleri ve ilgili optimum ridge yanlılık parametreleri.....	130

ŞEKİLLER DİZİNİ

SAYFA

Şekil 2.1.	Yanıt ve açıklayıcı değişken arasındaki serpilme grafiği	9
Şekil 3.1.	dw testi için karar şeması	21
Şekil 6.1.	Fransa ekonomi verisi için ACF ve PACF grafikleri.....	58
Şekil 6.2.	Açıklayıcı değişkenler arasındaki ilişkiler.....	61
Şekil 6.3.	Otokorelasyona karşı koşul sayısı.....	62
Şekil 6.4.	Ridge yanlılık parametresine karşı koşul sayısı	62
Şekil 6.5.	i -inci gözlemin silinmesi ile elde edilen koşul sayıları	63
Şekil 6.6.	Kaldıraç, DFFITS, COOK, COVRATIO, CW ve AP değerlerinin kutu grafikleri	69
Şekil 6.7.	ARR tahmin edicisi için yanlılık parametresine karşı OCV^R , Fransa ekonomi verisi	72
Şekil 6.8.	ARR tahmin edicisi için yanlılık parametresine karşı GCV^R , Fransa ekonomi verisi	72
Şekil 6.9.	ARR tahmin edicisi için yanlılık parametresine karşı C_p^R , Fransa ekonomi verisi	73
Şekil 6.10.	Otokorelasyona karşı $h_{i,i}$ ve $h_{i,i}^R$ değerleri	75
Şekil 6.11.	Yanlılık parametresine karşı $h_{i,i}$ ve $h_{i,i}^R$ değerleri	76
Şekil 6.12.	GEKK tahmin edicisi için otokorelasyona karşı $DFBETA_i$ grafiği	77
Şekil 6.13.	ARR tahmin edicisi için otokorelasyona karşı $DFBETA_i^R$ grafiği....	78
Şekil 6.14.	Ridge yanlılık parametresine karşı $DFBETA_i^R$ grafiği.....	79
Şekil 6.15.	Otokorelasyona karşı $DFFITS_i$ ve $DFFITS_i^R$ değerleri.....	80
Şekil 6.16.	Ridge yanlılık parametresine karşı $DFFITS_i^R$ değerleri.....	81
Şekil 6.17.	Otokorelasyona karşı D_i ve D_i^R değerleri.....	82
Şekil 6.18.	Ridge yanlılık parametresine karşı D_i^R değerleri	83
Şekil 6.19.	Otokorelasyona karşı $COVRATIO_i$ ve $COVRATIO_i^R$ değerleri.....	84

Şekil 6.20.	Ridge yanlılık parametresine karşı $COVRATIO_i^R$ değerleri	85
Şekil 6.21.	Otokorelasyona karşı CW_i ve CW_i^R değerleri	86
Şekil 6.22.	Ridge yanlılık parametresine karşı CW_i^R değerleri	87
Şekil 6.23.	Otokorelasyona karşı AP_i ve AP_i^R değerleri	88
Şekil 6.24.	Ridge yanlılık parametresine karşı AP_i^R değerleri	89
Şekil 6.25.	Otokorelasyona karşı OCV ve OCV^R değerleri	90
Şekil 6.26.	Otokorelasyona karşı GCV ve GCV^R değerleri	90
Şekil 6.27.	Otokorelasyona karşı C_p ve C_p^R değerleri	91
Şekil 7.1.	$\gamma^2 = 0.99$ iken farklı ρ ve σ değerlerinde ARR ve GEKK için COVRATIO ve AP sonuçlarının kutu grafikleri	112
Şekil 7.2.	ARR tahmin edicisi için ridge yanlılık parametresine karşı OCV değerleri, Monte Carlo simülasyon	126
Şekil 7.3.	ARR tahmin edicisi için ridge yanlılık parametresine karşı GCV, Monte Carlo simülasyon	127
Şekil 7.4.	ARR tahmin edicisi için ridge yanlılık parametresine karşı C_p , Monte Carlo simülasyon	128

SİMGELER VE KISALTMALAR

ACF	: Otokorelasyon fonksiyonu
AR(1)	: Birinci dereceden otoregresif süreç
ARR	: Otokorelasyonlu ridge regresyon
BLUE	: En iyi lineer yansız tahmin edici
EKK	: En küçük kareler
GEKK	: Genelleştirilmiş en küçük kareler
GLİM	: Genel lineer istatistiksel model
KS	: Koşul sayısı
MSE	: Hata kareler ortalaması
PACF	: Kısmi otokorelasyon fonksiyonu
RR	: Sıradan ridge regresyon
WN	: Ak gürültü serisi
$\hat{b}$	: EKK tahmin edici
$\hat{b}^r$	: RR tahmin edici
$\hat{\beta}$	: GEKK tahmin edici
$\hat{\beta}^R$	: ARR tahmin edici
<i>bias</i>	: Yanlılık
<i>cov</i>	: Kovaryans
<i>det</i>	: Determinant
d_l	: <i>dw</i> testi için alt sınır değer
d_u	: <i>dw</i> testi için üst sınır değer
<i>dw</i>	: Durbin-watson testi
(<i>i</i>)	: <i>i</i> -inci gözlemi veri kümesinden çıkartmak
<i>H</i>	: GEKK tahmin edicisi için izdüşüm matrisi
H^R	: ARR tahmin edicisi için yarı izdüşüm matrisi
<i>k</i>	: Ridge yanlılık parametresi

L	: EKK tahmin edicisi için izdüşüm matrisi
L^r	: RR tahmin edicisi için yarı izdüşüm matrisi
P	: Dönüşüm matrisi
tr	: Trace
var	: Varyans
ρ	: Otokorelasyon katsayısı



1. GİRİŞ

Regresyon analizi, istatistik literatüründe önemli bir yere sahiptir. Birçok istatistiksel analizde olduğu gibi regresyon analizinde de amaç gözlenen veriyi mümkün olduğunca basit, faydalı bir şekilde özetlemektir. Veri kümesinin uydurulan bir model üzerinden yorumlanması regresyon analizinin temel mantığıdır. Eğer modeldeki parametreler birinci dereceden polinom olarak modele dâhil edilmiş ise model, lineer model olarak adlandırılır. Modelin oluşturulmasında gözlenen bir yanıt değişken ve bir veya birden fazla gözlenen açıklayıcı değişken kullanılır. Açıklayıcı değişkenlerin tümünü içeren matris, açıklayıcı değişken matrisi olarak adlandırılır. Yanıt değişken ve açıklayıcı değişken/değişkenler arasındaki fonksiyonel ilişki bilinmeyen regresyon parametrelerine bağlıdır. Bu parametrelerin tahmininde kullanılacak yöntemler regresyon analizinin varsayımlarına dayanır. Lineer regresyon analizinin temel varsayımları;

- i. Açıklayıcı değişken matrisi stokastik olmayan bir matristir,
- ii. Açıklayıcı değişken matrisi tam kolon ranklıdır; açıklayıcı değişkenler lineer bağımsızdır,
- iii. Hata vektörünün elemanları gözlenemeyen rastgele değişkendir,
- iv. Yanıt değişken, açıklayıcı değişkenin lineer fonksiyonudur,
- v. Hata terimi, sıfır ortalamalı ve sabit varyanslı normal dağılıma sahiptir,
- vi. Hata terimleri birbiriyle ilişkisizdir,

sağlandığında modele en küçük kareler (EKK) yöntemi uygulanarak bilinmeyen parametreler tahmin edilip, ilgilenilen analizler yapılır.

EKK ile uydurulan bir lineer istatistiksel model şüphesiz ki literatürde en çok kullanılan istatistiksel yöntemdir (Chatterjee ve Hadi, 1988). EKK yönteminin

en iyi yansız kestirici olması için temel varsayımların sağlanması gerekmektedir. Ancak uygulamalarda bu varsayımların ihlali ile karşılaşilmektedir. Varsayımların sağlanmadığı durumlarda modele EKK yöntemi uygulandığında regresyon katsayılarının EKK tahmin edicileri hala lineer yansız tahmin ediciler olmalarına rağmen varyansları büyüyecektir. Aynı zamanda kovaryans da değişiklik gösterecektir. Bu durumda tahmin edicinin güvenilirliği azalacaktır. Varyans ve kovaryansın büyük olması güven aralıklarının daha geniş olmasına, bunun sonucunda ise istatistiksel hipotez testlerinin geçerliliğinin kaybedilmesine neden olacaktır.

Bu araştırmada, sayfa 1’de verilen (ii) ve (vi) varsayımların ihlalleri üzerinde durulmuştur.

Uygulamalarda (ii) varsayımının ihlali durumunda karşılaşılan problem çoklu iç ilişki olarak adlandırılmaktadır. Çoklu iç ilişkinin pek çok nedeni olabilir. Bunlardan bazıları şöyledir:

- Ø Uygulanan veri toplama metodu,
- Ø Modeldeki veya kitledeki zorunluluklar,
- Ø Modelin aşırı tahmin edilmesi.

Çoklu iç ilişkinin giderilmesi için veriye yeni değişkenler eklenebilir veya çoklu ilişkili değişkenlerden bir veya bir kaç modelden çıkarılabilir. Ancak değişkenlerin modelden çıkarılmasıyla modele gerçekten katkı yapan bir değişken çıkarılmış olabilir ki bu da bilgi kaybına sebep olur. Modeldeki değişkenleri çıkarmadan regresyon katsayılarını tahmin etmek için genellikle çoklu iç ilişkiye en küçük kareler yönteminden daha az duyarlı olan yöntemler kullanılır. Bu yöntemlerin kullanılması yanlış tahmin edicilerin ortaya çıkmasına neden olmuştur. Çoklu iç ilişki sorunu ile baş etmek için kullanılan en popüler yöntem Hoerl ve Kennard (1970) tarafından önerilen sıradan ridge regresyon tahmin edicidir.

Uygulamalarda hataların sabit varyans ve ilişkisiz olması varsayımının ihlali ile de sıklıkla karşılaşılmaktadır. Sabit varyansın sağlanmadığı durumlarda değişen varyans, hatalar arası ilişkisizliğin sağlanmadığı durumlarda ise otokorelasyon problemi ortaya çıkmaktadır. Bu araştırmada sayfa 1’de verilen (vi) varsayımın ihlali üzerinde durulmuştur. Otokorelasyon probleminin meydana gelmesindeki başlıca nedenler şöyledir:

- Ø Zaman serisi verileri ile çalışmak,
- Ø Denklemden bulunması gerektiği halde bulunmayan açıklayıcı değişkenlerin varlığı,
- Ø Modelin yanlış belirlenmesi,
- Ø Yanıt değişkeninde sistematik yani rassal olmayan ölçme hatalarının bulunması.

Otokorelasyon problemini ortadan kaldırmak için ilk olarak hatalar arasındaki ilişkinin modellenmesi gerekmektedir. İlişkili hatalar, otoregresif modeller (AR(p)), hareketli ortalama modeller (MA(q)) ya da otoregresif hareketli ortalama modeller (ARMA(p,q)) ile modellenir. Otokoregresif modeller, bir serinin herhangi bir dönemdeki gözlem değerini ondan önceki belirli sayıda dönemin gözlem değeri ve hata teriminin doğrusal birleşimi olarak ifade eden modellerdir. Otokoregresif modeller içerdikleri geçmiş dönem gözlem değeri sayısına göre isimlendirilirler (1 dönem gecikmeli ilişki ise, 1-inci dereceden otoregresif süreç vb.). Hareketli ortalama modelleri, bir serinin herhangi bir dönemdeki gözlem değerini aynı dönemdeki hata terimi ve ondan önceki belirli sayıda dönemin hata terimine bağlı olarak açıklayan modellerdir. Hareketli ortalama modelleri içerdikleri geçmiş dönem hata terimi sayısına göre 1-inci dereceden, 2-inci dereceden vb. olarak adlandırılırlar. Bir seride hem p -inci dereceden otoregresif hem de q -uncu dereceden hareketli ortalama modeli gözleniyor ise bu model otoregresif hareketli ortalama model olarak adlandırılmaktadır. Hatalar arasındaki

ilişki modellendikten sonra uygun dönüşüm matrisi ile model dönüştürülür ve dönüştürülmüş modelde hatalar sabit varyanslı ve ilişkisiz olarak elde edilir.

Regresyon analizinde seçilen modelin verideki tüm gözlemler için uygun olması oldukça önemlidir. Fakat uygulamada bir veya birden çok gözlem, verinin çoğunluğunun oluşturduğu modelden farklılık gösterebilir (Weisberg, 1985). Bu farklılıklar kaldıraç ve etkin gözlem kavramlarını ortaya çıkarmaktadır. Kaldıraç gözlem regresyon doğrusunun üzerinde ancak verinin merkezine uzakta olan gözlem, etkin gözlem ise hem verinin merkezinden uzakta hem de regresyon doğrusunun eğimini değiştiren gözlem olduğu için regresyon sonuçlarını etkilemektedir. Bu nedenle analizlere başlanmadan önce veri kümesindeki kaldıraç ve etkin gözlemler incelenmelidir. İstatistik literatüründe kaldıraç ve etkin gözlemleri belirlemek için önerilen birçok yöntem regresyon varsayımlarının sağlandığı ve EKK'nın en iyi yansız tahmin edici olduğu durumlarda verilmiştir.

Bu araştırmada, sayfa 1'de verilen (ii) ve (vi) varsayımlarının ihlalleri durumunda uygun tahmin ediciler verilmiş, bu tahmin ediciler için daha önce tanımlanmayan etkinlik ölçümleri önerilmiş ve tahmin edicilerin etkinlik performansları karşılaştırılmıştır.

Araştırmanın genel hattı sırası ile şöyledir: 2. bölümde ilişkisiz ve sabit varyanslı model altında daha önce ele alınan tanılama metotları verilmiştir. 3. bölümde, sayfa 1'de verilen (ii) ve (vi) varsayımlarının ihlalleri durumunda incelemelerde kullanılacak genel lineer istatistiksel model ve bu model altında yanlı ve yansız olmak üzere iki tahmin edici tanıtılmıştır. 4. bölümde birinci derecen otoregresif (first order autoregressive, AR(1)) hatalı lineer modelde bir önceki bölümde tanıtılan tahmin ediciler için etki tanılama metotları verilmiştir. 5. bölümde genel lineer modelin geçerliliğinin testi için istatistikler önerilmiştir. 6. bölümde tanılama metotları bir nümerik örnek üzerinden incelenmiş, 7. bölüm de ise her bir başlık için Monte Carlo simülasyon çalışmasına yer verilmiştir. 8. bölümde tez çalışmasının sonuçlarına değinilmiş ve yeni araştırmacılara yol göstermesi açısından öneriler sunulmuştur.

2. İLİŞKİSİZ VE SABİT VARYANSLI HATAYA SAHİP LİNEER İSTATİSTİKSEL MODEL

Tuğba SÖKÜT AÇAR

2. İLİŞKİSİZ VE SABİT VARYANSLI HATAYA SAHİP LİNEER İSTATİSTİKSEL MODEL

Klasik varsayımların sağlandığı çoklu lineer model

$$y = Z\beta + u, \quad u \sim N(0, \sigma_u^2 I_n) \quad (2.1)$$

ele alınsın. y , yanıt değişkenlerin $(n \times 1)$ tipinde gözlenen vektörü, $\mathbf{1}$, $(n \times 1)$ tipinde 1'lerden oluşan vektör, $X = (x_1, x_2, \dots, x_r)$, açıklayıcı değişkenlerin $(n \times r)$ tipinde bilinen matrisi olmak üzere $Z = [\mathbf{1} \ X]'$ dir. $\beta = [\beta_0 \ \beta_1]'$, β_0 bilinmeyen parametre ve r boyutlu β_1 ile bilinmeyen parametreleri içeren vektördür. $u_i \sim N(0, \sigma_u^2)$, $E(u_i u_j) = 0$ ($i = 1, \dots, n$, $j = 1, \dots, n$, $i \neq j$) ile u hata vektörü, $u \sim N(0, \sigma_u^2 I_n)$ dağılır. Burada I_n , n –boyutlu birim matrisi ifade etmektedir.

(2.1) modelindeki regresyon parametrelerinin tahminleri EKK ile bulunur. Bilinmeyen parametre vektörü β 'nın EKK tahmin edicisi hata kareler toplamı minimum yapılarak elde edilir. $u = y - Z\beta$ olmak üzere hata kareler toplamı;

$$S(\beta) = \sum_{i=1}^n u_i^2 = u'u = (y - Z\beta)'(y - Z\beta) = y'y - 2\beta'Z'y + \beta'Z'Z\beta$$

olarak elde edilir. β 'nın EKK tahmin edicisi,

$$\frac{\partial S(\beta)}{\partial \beta} = -2Z'y + 2Z'Z\beta = 0$$

$$Z'Z\beta = Z'y$$

normal denklemlerin çözümü ile

$$\hat{b} = (Z'Z)^{-1}Z'y \quad (2.2)$$

elde edilir. $\hat{b}$ tahmin edicisi beklenen değeri

$$E(\hat{b}) = E[(Z'Z)^{-1}Z'y] = (Z'Z)^{-1}ZE(Z\beta + u) = \beta$$

ve varyans-kovaryans matrisi

$$\text{cov}(\hat{b}) = (Z'Z)^{-1}Z'\text{cov}(y)Z(Z'Z)^{-1} = \sigma_u^2(Z'Z)^{-1}$$

2. İLİŞKİSİZ VE SABİT VARYANSLI HATAYA SAHİP LİNEER İSTATİSTİKSEL MODEL Tuğba SÖKÜT AÇAR

ile Gauss-Markov teoremi gereğince en iyi lineer yansız tahmin edicidir (best linear unbiased estimator, BLUE).

2.1. Çoklu İç İlişki Problemi ve Sıradan Ridge Regresyon Tahmin Edici

Çoklu lineer istatistiksel modelde, genellikle açıklayıcı değişkenlerin bağımsız olduğu varsayılır, ancak uygulamada bu varsayımın ihlali söz konusu olabilir. Çoklu iç ilişki, bir ya da daha fazla açıklayıcı değişkenin lineer bağımlı olması durumunu ifade eder.

$t_1, t_2, \dots, t_p$ sabitler ve Z_j ($j = 1, 2, \dots, p$) Z matrisinin j -inci sütunu olmak üzere

$$\sum_{j=1}^p t_j Z_{.j} = 0$$

eşitliğini sağlayacak şekilde sıfırdan farklı en az bir t_j ($j = 1, 2, \dots, p$) sabiti var ise tam çoklu iç ilişki durumu ortaya çıkmaktadır. Bu durumda $Z'Z$ matrisi singüler olduğu için EKK tahmin edici kullanılamaz.

$$\sum_{j=1}^p t_j Z_{.j} \approx 0$$

olması yaklaşık çoklu iç ilişki durumunu ifade eder. $Z'Z$ matrisi singülere yakın olmakla birlikte tersi bulunabileceği için EKK yöntemi kullanılabilir, ancak EKK tahminlerinin varyansı büyük olacaktır.

Çoklu iç ilişkinin belirlenmesinde $Z'Z$ matrisinin karakteristik yapısından faydalanılması çalışmalarda sıklıkla kullanılmaktadır. $Z'Z$ matrisinin en büyük özdeğerinin en küçük özdeğere oranı olarak tanımlanan koşul sayısı ($KS = \frac{\lambda_{maks}}{\lambda_{min}}$) belirleyicisi literatürde sıklıkla başvurulmuş çoklu iç ilişki belirleme yöntemlerinden biridir. $KS < 100$ ise çok ciddi boyutta çoklu iç ilişki problemi olmadığı, $100 < KS < 1000$ ise kuvvetli çoklu iç ilişki olduğu, $KS > 1000$ ise veri kümesinde birden fazla çoklu iç ilişki problemi olduğu kabul edilmektedir.

2. İLİŞKİSİZ VE SABİT VARYANSLI HATAYA SAHİP LİNEER İSTATİSTİKSEL MODEL Tuğba SÖKÜT AÇAR

Çoklu iç ilişki problemini ortadan kaldırmak için önerilen tahmin prosedürleri yanlış tahmin edicilerin ortaya çıkmasına neden olmuştur. Hoerl ve Kennard (1970), EKK yönteminde $Z'Z = \begin{pmatrix} \mathbf{1}'\mathbf{1} & \mathbf{1}'X \\ X'\mathbf{1} & X'X \end{pmatrix}$ matrisinin tersini almadan önce $X'X$ matrisinin köşegen elemanlarına bir k ($0 < k \leq \infty$) sabiti ekleyerek sıradan ridge regresyon (RR) tahmin edicisini

$$\hat{b}^r = (Z'Z + kI_p^*)^{-1}Z'y, \quad k \geq 0, \quad I^* = \text{diag}(0,1,\dots,1) \quad (2.3)$$

olarak önermiştir. k , ridge yanlışlık parametresini ifade etmektedir. $k = 0$ olduğunda (2.3) eşitliği ile verilen $\hat{b}^r$ tahmin edicisi (2.2) eşitliği ile verilen $\hat{b}$ tahmin edicisine dönüşmektedir. RR, $\text{bias}(\hat{b}^r) = -k(Z'Z + kI_p^*)^{-1}\beta$ ile yanlış bir tahmin edicidir. RR tahmin edicisinin varyans-kovaryans matrisi ise

$$\text{cov}(\hat{b}^r) = \sigma_u^2(Z'Z + kI_p^*)^{-1}Z'Z(Z'Z + kI_p^*)^{-1}$$

dir.

2.2. Rezidülere Dayalı Ölçümler

Regresyon tanılama yöntemlerinde rezidüler önemli bir yere sahiptir. Rezidüler gözlenen veri ve varsayılan model arasındaki tutarlılığın derecesini belirler (Myers, 1990). Bazı durumlarda ölçeklendirilmiş rezidülerle çalışmak faydalıdır (Montgomery ve ark., 2001). Ölçeklendirilmiş rezidüler aykırı gözlemlerin belirlenmesinde yardımcıdır. Çalışmanın bu bölümünde (2.1) modeli altında EKK ve RR tahmin edicileri için rezidü ve ölçeklendirilmiş rezidüler tanımlanmıştır.

EKK tahmin edicisi için rezidü vektörü

$$\hat{u} = y - \hat{y} = y - Z\hat{b} = y - Z(Z'Z)^{-1}Z'y = (I - L)y$$

ile elde edilir. Burada $L = Z(Z'Z)^{-1}Z'$ izdüşüm matrisidir. $l_{i,j}$, L matrisinin i -inci satır ve j -inci sütununu göstermek üzere, L matrisinin teorik incelemelerde önemli sonuçlara neden olan özellikleri vardır. Bunlar kısaca şöyledir;

2. İLİŞKİSİZ VE SABİT VARYANSLI HATAYA SAHİP LİNEER İSTATİSTİKSEL MODEL Tuğba SÖKÜT AÇAR

- vii. L simetrik bir matristir ($L' = L$),
- viii. L idempotent bir matristir ($L^2 = L$),
- ix. $trace(L) = rank(L) = p$,
- x. $\sum_j l_{i,j}^2 = l_{i,i} = l_{i,i}^2 + \sum_{j \neq i} l_{i,j}^2 \quad i, j = 1, \dots, n$
- xi. $0 \leq l_{i,i} \leq 1, \quad i = 1, \dots, n$.

RR tahmin edicisi için rezidü vektörü,

$$\hat{u}^r = y - \hat{y}^r = y - Z\hat{b}^r = y - Z(Z'Z + kI_p^*)^{-1}Z'y = (I - L^r)y$$

ile verilir. Burada $L^r = Z(Z'Z + kI_p^*)^{-1}Z'$ yarı izdüşüm matrisidir. L^r simetrik ancak idempotent bir matris olmadığından dolayı Tripp (1983) bu matrisi yarı izdüşüm matrisi olarak adlandırmıştır.

Aykırı gözlemi belirlemede rezidüleri yararlı kılan bir ölçüm PRESS rezidülerdir. PRESS rezidü ön tahmin hata kareler ortalamasını ifade etmektedir. PRESS rezidü, ilgili tahmin yöntemi ile uydurulan i -inci yanıt değer i -inci gözleme ait bilinen yanıt değerden ne kadar saptığını gösterir.

EKK tahmin edicisi için Allen (1974) tarafından önerilen PRESS rezidü,

$$\hat{u}_{(i)} = y_i - \hat{y}_{(i)} = \frac{\hat{u}_i}{1 - l_{i,i}}, \quad i = 1, \dots, n$$

ile bulunur. Burada, $\hat{y}_{(i)}$, i -inci gözlem çıkartılıp $(n - 1)$ gözlem üzerinden EKK ile uydurulan i -inci gözlemi göstermekte, $l_{i,i}$, L matrisinin i -inci köşegen elemanını göstermektedir.

RR tahmin edicisi için Montgomery ve Friedman (1993) tarafından verilen PRESS rezidü,

$$\hat{u}_{(i)}^r = y_i - \hat{y}_{(i)}^r = \frac{\hat{u}_i^r}{1 - l_{i,i}^r}, \quad i = 1, \dots, n$$

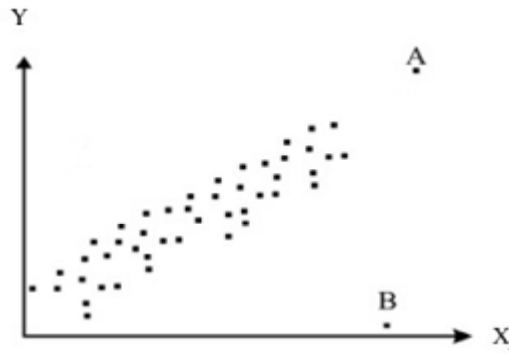
ile bulunur. Burada, $\hat{y}_{(i)}^r$, i -inci gözlem çıkartılıp kalan $(n - 1)$ gözlem üzerinden RR ile uydurulan i -inci gözlemi göstermekte, $l_{i,i}^r$, L^r matrisinin i -inci köşegen elemanını göstermektedir.

2. İLİŞKİSİZ VE SABİT VARYANSLI HATAYA SAHİP LİNEER İSTATİSTİKSEL MODEL Tuğba SÖKÜT AÇAR

PRESS rezidünün küçük olması ilgili tahmin yöntemi ile uydurulan i -inci gözlemin bilinen değerine çok benzer olduğunu gösterir (Montgomery ve ark., 2001). Eğer i -inci gözlem silinip, yanıt değişkeninin tahmini kalan gözlemin i -incisi üzerinden yapılırsa bilinen ve tahmin edilen değerlerin karşılaştırması ön tahmin hatası ile yapılır.

2.3. Aykırı ve Etkin Gözlem

Aykırı ve etkin gözlemlerin belirlenmesine yönelik ölçümlerin birçoğu (2.1) modeli altında sıklıkla ele alınmıştır; Belsley ve ark. (1980), Cook ve Weisberg (1982), Cook (1977), Chatterjee ve Hadi (1986).



Şekil 2.1. Yanıt ve açıklayıcı değişken arasındaki serpilme grafiği

Açıklayıcı değişkenlerin gözlenen matrisinin satırları r -boyutlu R^r uzayında işaretlendiğinde n tane gözlemin serpilme diyagramı $r = 1$ iken Şekil 2.1'dekine benzer olarak verilsin. Bu serpilme diyagramında veri merkezinden uzakta olan A, “kaldıraç gözlem” olarak adlandırılmaktadır. A gözlemi küçük rezidüye sahip olmasına rağmen açıklayıcı değişkenlerin merkezinden oldukça uzaktır. Yani, A gözlemi regresyon doğrusunun eğimini değiştirmeyen ancak veri merkezinin oldukça uzağında bulunan bir gözlemdir. B gözlemi ise, hem regresyon doğrusuna uzakta bir gözlem olduğu için büyük rezidüye sahip hem de çekim gücü oldukça yüksek olduğundan dolayı “etkin gözlem” olarak adlandırılmaktadır. Etkin

2. İLİŞKİSİZ VE SABİT VARYANSLI HATAYA SAHİP LİNEER İSTATİSTİKSEL MODEL Tuğba SÖKÜT AÇAR

gözlemler regresyon doğrusunun eğimini değiştirmektedir. Etkin ya da kaldıraç gözlem/gözlemlerin varlığı regresyon sonuçları üzerinde farklı sonuçlara neden olduğundan dolayı bu tür gözlem ya da gözlemlerin regresyon analizine geçilmeden önce belirlenmesi büyük önem taşımaktadır. Kaldıraç gözlem aynı zamanda etkin gözlem olmayı, etkin gözlem ise aynı zamanda kaldıraç gözlem olmayı gerektirmez.

Kaldıraç olan gözlemlerin belirlenmesi izdüşüm matrisine dayanmaktadır. İzdüşüm matrisinin köşegen elemanları X uzayının uzağına düşen gözlemleri belirler.

2.4. Etkinlik Eğrisine Dayalı Ölçümler

Bu başlık altında üç istatistik ele alınmıştır. Bunlardan ilki, Cook uzaklığıdır. Cook uzaklığı, β parametre uzayında tahmin edilen parametre vektörünün konumunu belirler. İkinci istatistik DFFITS istatistiğidir. DFFITS, y yanıt vektörü üzerinde etkili olan gözlemi belirlemede kullanılır. Son incelenecek ölçüm DFBETA istatistiğidir. DFBETA istatistiği de tahmin edilen parametre vektörüne etki eden gözlemi belirlemede kullanılır.

Cook, DFFITS ve DFBETA ölçümleri EKK tahmin edici için sıklıkla ele alınmıştır (Cook 1977; Chatterjee ve Hadi, 1986, vb.). Steece (1986) sabit varyanslı ve ilişkisiz hataya sahip lineer istatistiksel modelde ridge regresyon üzerinden kaldıraç gözlemlerin belirlenmesini incelemiştir. Walker ve Birch (1988), kaldıraç ve etkinlik kavramlarını ilişkisiz hataya sahip lineer istatistiksel modelde ridge regresyon üzerinden ele almıştır.

Regresyon sonuçlarında i -inci gözlemin etkinlik ölçümünün önemli bir sınıfı Hampel (1974) tarafından tanıtılan etkinlik fonksiyonu fikrine dayanır. Etkinliğin belirlenmesinde sıklıkla kullanılan istatistik Cook uzaklığıdır. Cook (1977)'un önerdiği bu istatistik, etkinliğin ölçümünde yanıt değişken ve açıklayıcı değişken uzayında gözlemin konumunu belirlemektedir. Yani, etkiyi hem y_i kestiricisinin

2. İLİŞKİSİZ VE SABİT VARYANSLI HATAYA SAHİP LİNEER İSTATİSTİKSEL MODEL Tuğba SÖKÜT AÇAR

uydurulan değerine olan uyumunu hem de z_i 'nin geriye kalan gözlemlerden uzaklığı anlamında ele alır. Cook uzaklık ölçümü EKK tahmin edicisi için

$$D_i^{EKK} = \frac{(\hat{b}_{(i)} - \hat{b})' Z' Z (\hat{b}_{(i)} - \hat{b})}{p \hat{\sigma}^2}$$

olarak Cook (1977) tarafından tanımlanmıştır. Burada $\hat{\sigma}^2 = \frac{(y - z\hat{b})'(y - z\hat{b})}{n-p}$, σ^2 'nin EKK tahmin edicisidir. Walker ve Birch (1988), RR tahmin edicisi için Cook uzaklığını

$$D_i^{RR} = \frac{(\hat{b}_{(i)}^r - \hat{b}^r)' Z' Z (\hat{b}_{(i)}^r - \hat{b}^r)}{p \hat{\sigma}^2}$$

olarak önermişlerdir. Cook (1977) ile Montgomery ve Peck (1982), Cook uzaklık değeri 1'den büyük olan gözlemlerin etkin olduğunu belirtmişlerdir.

Welsch ve Kuh (1977) tarafından verilen DFFITS ölçümü, veri kümesinden elde edilen i -inci tahmin edilen yanıt değerden i -inci gözlem çıkartılarak elde edilen i -inci tahmin edilen yanıt değer arasındaki fark alınarak tanımlanmıştır. EKK tahmin edicisi altında Welsch ve Kuh (1977) DFFITS ölçümünü

$$DFFITS_i^{EKK} = \frac{z_i(\hat{b} - \hat{b}_{(i)})}{\hat{\sigma}_{(i)} \sqrt{l_{i,i}}}$$

olarak önerirken, iken Walker ve Birch (1988) RR tahmin edicisi altında DFFITS ölçümünü

$$DFFITS_i^{RR} = \frac{z_i(\hat{b}^r - \hat{b}_{(i)}^r)}{\hat{\sigma}_{(i)} \left\{ \sum_j (l_{i,j}^r)^2 \right\}^{1/2}}$$

olarak vermiştir. Burada $\hat{\sigma}_{(i)} = \sqrt{\frac{(n-p)\hat{\sigma}^2 - [\hat{u}_i^2 / (1-l_{i,i})]}{n-p-1}}$, i -inci gözlem veri kümesinden çıkartıldığında σ 'nin EKK tahmin edicisidir.

DFBETA ölçümü, i -inci gözlem veriden çıkartıldığında hesaplanacak olan yeni regresyon denkleminin parametrelerinde meydana gelecek olan değişimi hesaplamak üzere önerilmiştir.

2. İLİŞKİSİZ VE SABİT VARYANSLI HATAYA SAHİP LİNEER İSTATİSTİKSEL MODEL Tuğba SÖKÜT AÇAR

EKK tahmin edicisi için DFBETA ölçümü Belsley ve arkadaşları (1980) tarafından aşağıdaki gibi önerilmiştir:

$$DFBETA_i^{EKK} = \hat{b} - \hat{b}_{(i)}.$$

RR tahmin edicisi için DFBETA ölçümü ise aşağıdaki gibidir:

$$DFBETA_i^{RR} = \hat{b}^r - \hat{b}_{(i)}^r.$$

2.5. Güven Elipsoidine Dayalı Ölçümler

Etkinliğe dayalı tanılama ölçümlerinden, DFFITS, DFBETA, Cook Uzaklığı ilgili gözlemin yanıt değişken ya da regresyon parametresi üzerine olan etkilerini yorumlamada kullanılmasına rağmen bu ölçümler *i*-inci gözlemin silinmesiyle güven elipsoidinin merkezindeki değişim hakkında bilgi vermez. *i*-inci gözlemin etkisine dayalı ölçümlerin alternatif bir sınıfı *i*-inci gözlemin veri kümesinden çıkartılmasıyla güven elipsoidinin hacmindeki değişikliğe dayanır. Bu ölçümler COVRATIO, Cook-Weisberg (CW) ve Andrews-Pregibon (AP) istatistikleridir (Chatterjee ve Hadi, 1988).

Güven elipsoidine dayalı etki tanılama ölçümleri ile ilgili daha önce yapılan araştırmalar şöyle derlenmiştir: Acarlar ve Gamgam (2010), çoklu doğrusal regresyonda etkin gözlemlerin belirlenmesinde kullanılan yöntemleri karşılaştırırken COVRATIO oranından faydalanmışlardır. İbrahim ve arkadaşları (2013), dairesel regresyon modelinde aykırı gözlemlerin varlığının belirlenmesinde COVRATIO istatistiğini kullanırken, Gapor ve arkadaşları (2014), basit lineer fonksiyonel ilişkili modelde aykırı gözlemi belirlemek için COVRATIO istatistiğini kullanmışlardır. Chatterjee ve Hadi (1986), ilişkisiz ve sabit varyanslı modelde etkinlik ölçümlerinden AP, COVRATIO ve CW istatistiklerini incelemiştir. Draper ve John (1981), regresyonda birden fazla etkili ve aykırı gözlem varlığında AP istatistiğini incelemiştir. Martin (1992), hataların ilişkili olduğu lineer istatistiksel modelde COVRATIO ve AP istatistiklerini en genel hali ile ele almıştır.

2. İLİŞKİSİZ VE SABİT VARYANSLI HATAYA SAHİP LİNEER İSTATİSTİKSEL MODEL Tuğba SÖKÜT AÇAR

Güven elipsoidine dayalı etki tanılama ölçümleri için incelenecek ilk istatistik, i -inci gözlemin silinmesiyle elde edilen kovaryans matrisinin determinantının, tüm veriden elde edilen kovaryans matrisinin determinantına oranı olan COVRATIO istatistiğidir. (2.2)'de verilen EKK tahmin edicisi kullanıldığında COVRATIO istatistiği Belsley ve arkadaşları (1980) tarafından

$$COVRATIO_i^{EKK} = \frac{\det\{cov(\hat{b}_{(i)})\}}{\det\{cov(\hat{b})\}} = \frac{\det\{\hat{\sigma}_{(i)}^2 [Z'_{(i)} Z_{(i)}]^{-1}\}}{\det\{\hat{\sigma}^2 [Z' Z]^{-1}\}}$$

olarak tanımlanmıştır.

COVRATIO, i -inci gözlemin regresyon katsayıları tahmininin doğruluğu üzerindeki etkisini yansıtır. Belsley ve arkadaşları (1980) ile Montgomery ve arkadaşları (2001), COVRATIO değeri 1'e yakın olan gözlemlerin tahminler üzerindeki hassasiyetinin çok düşük olduğunu vurgularken, COVRATIO değeri 1'den büyük olan gözlemlerin tahminler üzerindeki hassasiyetlerinin çok yüksek olduğunu vurgulamıştır. Tüm gözlemlerin varyans-kovaryans matrisi üzerinde eşit etkiye sahip olduğu ideal durum için COVRATIO değerleri yaklaşık olarak 1'dir.

İncelenecek olan ikinci istatistik CW istatistiğidir. Cook-Weisberg (1980) tarafından regresyon katsayılarının oluşturduğu güven elipsoidi üzerinde i -inci gözlemin etkisini ölçmek için önerilmiştir. Cook-Weisberg (1980), CW değerinin pozitif yönde büyük olması durumunda i -inci gözlemin silinmesinin güven elipsoidinin hacminde önemli bir azalışa neden olacağını, negatif yönde büyümesinin ise i -inci gözlemin silinmesinin güven elipsoidinin hacminde önemli bir artışa neden olacağını belirtmiştir.

EKK tahmin edicisi için CW istatistiği

$$CW_i^{EKK} = -\frac{1}{2} \log \left(COVRATIO(\hat{b}, \beta)_i \right) + \frac{p}{2} \log \left(\frac{F_{\alpha, p, n-p}}{F_{\alpha, p, n-p-1}} \right)$$

olarak verilmiştir. Görüldüğü üzere CW_i^{EKK} istatistiği $COVRATIO_i^{EKK}$ ile bağlantılıdır. CW_i^{EKK} eşitliğinin sağ tarafındaki ikinci terim i 'ye bağlı değildir, bu nedenle $COVRATIO_i^{EKK}$ ve CW_i^{EKK} denktirler (Chatterjee ve Hadi, 1988).

2. İLİŞKİSİZ VE SABİT VARYANSLI HATAYA SAHİP LİNEER İSTATİSTİKSEL MODEL Tuğba SÖKÜT AÇAR

Bu başlık altında incelenecek son ölçüm, güven elipsoidinin hacmini veren AP istatistiğidir. AP istatistiği, güven elipsoidinin hacmini ölçmek için Andrews ve Pregibon (1978) tarafından (2.2) eşitliğinde verilen EKK tahmin edicisi altında aşağıdaki gibi tanımlanmıştır.

$u \sim N(0, \sigma_u^2 I_n)$ varsayımı altında $\hat{b} \sim N(\beta, \sigma^2(Z'Z)^{-1})$ 'dir. $\hat{b}$ merkezli β için $\%(1 - \alpha)$ güven elipsoidi

$$\frac{(\beta - \hat{b})' Z' Z (\beta - \hat{b})}{p \hat{\sigma}^2} \leq F_{n, n-p, 1-\alpha} \quad (2.4)$$

olarak tanımlar. Benzer şekilde, eğer i -inci gözlem veri kümesinden çıkartılırsa $\hat{b}_{(i)}$ merkezli β için $\%(1 - \alpha)$ güven elipsoidi

$$\frac{(\beta - \hat{b}_{(i)})' Z'_{(i)} Z_{(i)} (\beta - \hat{b}_{(i)})}{p \hat{\sigma}_{(i)}^2} \leq F_{n, n-p-1, 1-\alpha} \quad (2.5)$$

olarak yazılır (Rao ve Toutenburg, 1999). (2.4) ve (2.5)'de verilen eşitsizlikler p boyutlu uzayda eliptik bir eğrinin çevresini tanımlar. (2.4)'de verilen β için güven elipsoidinin hacmi $\det(Z'Z)$ 'nin karekökü ile ters orantılıdır. Benzer şekilde eşitlik (2.5)'in hacmi $\det(Z'_{(i)} Z_{(i)})$ 'nin karekökü ile ters orantılıdır. Bu nedenle, etkin bir gözlemin atılmasıyla, $(Z'_{(i)} Z_{(i)})$ 'nin determinanı $(Z'Z)$ 'nin determinantına göre azalacaktır. Diğer yandan, büyük rezidüye sahip bir gözlemin atılmasıyla rezidü kareler toplamı $\hat{\sigma}_{(i)}^2$, $\hat{\sigma}^2$ 'ya göre azalacaktır. Andrews ve Pregibon (1978) bu iki bilgiyi bir ölçümde birleştirmiştir. AP istatistiğinin temel mantığı (2.4) ve (2.5)'de verilen elipsoidlerin hacmini

$$\frac{(n-p-1) \hat{\sigma}_{(i)}^2 \det(Z'_{(i)} Z_{(i)})}{(n-p) \hat{\sigma}^2 \det(Z'Z)} \quad (2.6)$$

eşitliğindeki gibi oranlamaktır (Rao ve Toutenburg, 1999).

$W = (Z: y)$ genişletilmiş matrisinin tanımlanması ile (2.6) eşitliğinde verilen ifade

$$\frac{\det(W'_{(i)} W_{(i)})}{\det(W'W)} \quad (2.7)$$

2. İLİŞKİSİZ VE SABİT VARYANSLI HATAYA SAHİP LİNEER İSTATİSTİKSEL MODEL Tuğba SÖKÜT AÇAR

eşitliğinden elde edilir. (2.7) eşitliği i -inci gözlemin atılmasıyla $\det(W'W)$ 'deki göreceli değişimi ölçmektedir. Veri merkezinin uzağında bulunan bir gözlemin atılması determinantta büyük bir azalmaya neden olacaktır ve böylece hacim de büyük ölçüde artacaktır. Bu yüzden eşitlik (2.6)'nın küçük değerleri özel dikkat gerektirir. Kolaylık sağlaması açısından AP istatistiği

$$AP_i^{EKK} = 1 - \frac{\det(W'_{(i)}W_{(i)})}{\det(W'W)}$$

olarak tanımlanır. AP istatistiği sadece aykırı ya da etkin gözlemlerin belirlenmesinde değil aynı zamanda parametre uzayındaki gözlemlerin uzaklığı hakkında da önemli bilgiler verir (Chatterjee ve Hadi, 1988).

Model performansı için EKK tahmin edicisi altında önerilen bu üç istatistik yanlış tahmin ediciler için daha önce ele alınmamıştır.

2.6. Model Geçerliliği

Sıradan çapraz geçerlilik (ordinary cross-validation, OCV), genelleştirilmiş çapraz geçerlilik (generalized cross-validation, GCV) ve kavramsal ön tahmin (conceptual prediction, C_p) bir modelin ön tahmin performansını belirlemek için kullanılan çapraz geçerlilik yöntemleridir (Picard ve Cook, 1984). Uydurulan bir modelin geçerliliğini gösterir yani, düşünülen modelin sağlıklı ve etkin bir model olup olmadığını ifade eder. Model iyi bir model değilse, uydurulan model üzerinden yapılacak tahminlerde güvenilir sonuç vermez.

Model geçerliliği ile ilgili daha önce yapılan araştırmalar şöyle derlenmiştir: Golub ve arkadaşları (1979), OCV ve GCV ölçümlerini ridge regresyon üzerinden incelemişlerdir. Andrews (1991), değişen varyanslı regresyon modelinde OCV ve GCV'nin asimptotik davranışlarını incelemişlerdir. Montgomery ve Friedman (1993), PRESS ölçümünü ridge için genişletmiş ve ridge yanlılık parametresinin seçimi için minimum PRESS ölçümünü kullanmışlardır. Liu ve arkadaşları (1999), tekrarlı ölçümlü modelde ilişkili hatalar için model seçiminde PRESS ölçümünü kullanmışlardır. Tarpey (2000), kısıtlı en küçük kareler için PRESS ölçümünü

2. İLİŞKİSİZ VE SABİT VARYANSLI HATAYA SAHİP LİNEER İSTATİSTİKSEL MODEL Tuğba SÖKÜT AÇAR

incelerken, Özkale ve Kaçıranlar (2007), Liu parametre seçiminde bir ölçüt olarak PRESS istatistiğini incelemişlerdir. Liu ve Jiang (2012), lineer istatistiksel modelde optimum genelleştirilmiş ridge tahmin edicisini belirlemek için GCV kriterini ele almışlardır. Özkale (2014), kısıtlı iki parametrelili regresyon altında OCV, GCV ve Cp ölçümelerini önermiştir.

OCV istatistiği model seçimi için Allen (1974) tarafından PRESS rezidü karelerin ortalaması olarak önerilmiştir. Bundan dolayı PRESS istatistiği olarak da adlandırılmaktadır. OCV’de, tek gözlem veri kümesinden çıkartılarak kalan veriden tahmin yapılır. Bu işlem her gözlem için tekrarlanır.

Verilen n gözlemli, p değişkenli bir veri kümesi için OCV, (2.2)’de verilen EKK tahmin edicisi altında Myers (1990) tarafından aşağıdaki gibi verilmiştir

$$OCV^{EKK} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_{(i)})^2. \quad (2.8)$$

Montgomery ve Friedman (1993), (2.3)’de verilen sıradan ridge regresyon tahmin edici için OCV’yi

$$OCV^{RR} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_{(i)}^r)^2$$

olarak tanımlamıştır.

Golub ve arkadaşları (1979), eşitlik (2.8)’de verilen istatistiğin ağırlıklandırılmış versiyonu olan GCV ölçümünü EKK tahmin edicisi için

$$GCV^{EKK} = \frac{(y - Z\hat{b})'(y - Z\hat{b})}{(n-p)^2}$$

olarak tanımlamıştır.

(2.3) eşitliğinde verilen $\hat{b}^r$ için GCV istatistiğini Golub ve arkadaşları (1979) aşağıdaki gibi tanımlamıştır

$$GCV^{RR} = \frac{(y - Z\hat{b}^r)'(y - Z\hat{b}^r)}{\{n - tr[Z(Z'Z + kI_p^*)^{-1}Z']\}^2}.$$

Mallows (1973) tarafından sıradan ridge yanlılık parametresi k ’nın seçimi için ele alınan Cp istatistiği aşağıda verilmiştir

$$C_p^r = \frac{SS_{Res}(k)}{\hat{\sigma}^2} - n + 2tr(L^R)$$

2. İLİŞKİSİZ VE SABİT VARYANSLI HATAYA SAHİP LİNEER İSTATİSTİKSEL MODEL Tuğba SÖKÜT AÇAR

burada , $SS_{Res}(k) = (y - Z\hat{b}^r)'(y - Z\hat{b}^r)$ ile elde edilen ARR rezidüleri kareler toplamıdır.



2. İLİŞKİSİZ VE SABİT VARYANSLI HATAYA SAHİP LİNEER
İSTATİSTİKSEL MODEL Tuğba SÖKÜT AÇAR



3. GENEL LİNEER İSTATİSTİKSEL MODEL VE ÇOKLU İÇ İLİŞKİ**3.1. Otokorelesyon Kaynağı ve Belirlenmesi**

(2.1) ile verilen çoklu lineer model tekrar ele alınsın. (2.1) modelinden farklı olarak, ε , $(n \times 1)$ tipinde $E(\varepsilon) = 0$, $E(\varepsilon\varepsilon') = \sigma_\varepsilon^2 V$ varsayımları ile hataların gözlenmeyen vektörü olduğunda

$$y = Z\beta + \varepsilon, \quad \varepsilon \sim N(0, \sigma_\varepsilon^2 V) \quad (3.1)$$

modeli genel lineer regresyon model olarak adlandırılmaktadır (Graybill, 1976, s.149). $E(\varepsilon\varepsilon')$ genel halde

$$E(\varepsilon\varepsilon') = E \begin{bmatrix} \varepsilon_1^2 & \varepsilon_1\varepsilon_2 & \cdots & \varepsilon_1\varepsilon_n \\ \varepsilon_2\varepsilon_1 & \varepsilon_2^2 & \cdots & \varepsilon_2\varepsilon_n \\ \vdots & \vdots & \ddots & \vdots \\ \varepsilon_n\varepsilon_1 & \varepsilon_n\varepsilon_2 & \cdots & \varepsilon_n^2 \end{bmatrix}$$

olarak yazılan hatalara ilişkin varyans-kovaryans matrisini ifade etmektedir (Judge ve ark., 1988). Bu matrisin özel durumları ele alınsın:

$$\text{xii.} \quad E(\varepsilon\varepsilon') = \sigma^2 \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix} = \sigma^2 I_n$$

$$\text{xiii.} \quad E(\varepsilon\varepsilon') = \begin{bmatrix} \sigma_1^2 & 0 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \sigma_n^2 \end{bmatrix} = \Omega = \text{diag}(\sigma_1^2, \dots, \sigma_n^2)$$

$$\text{xiv.} \quad E(\varepsilon\varepsilon') = \begin{bmatrix} \sigma^2 & \rho_1 & \rho_2 & \cdots & \rho_{n-1} \\ \rho_1 & \sigma^2 & \rho_1 & \cdots & \rho_{n-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{n-1} & \rho_{n-2} & \rho_{n-3} & \cdots & \sigma^2 \end{bmatrix} = \sigma^2 V$$

(xii)'de verilen varyans-kovaryans matrisi, hatalara ilişkin varsayımının sağlandığı ilişkisiz ve sabit varyanslı hata yapısını göstermektedir. (xiii) de verilen matris, köşegen elemanları eşit olmayan köşegen matrisi göstermekte olup farklı varyanslı ancak ilişkisiz hata yapısını göstermektedir. Bu durum “değişen varyans” olarak adlandırılmaktadır. (xiv)'de verilen matris ise köşegen dışı elemanları sıfır olmayan

3. GENEL LİNEER İSTATİSTİKSEL MODEL VE ÇOKLU İÇ İLİŞKİ

Tuğba SÖKÜT AÇAR

ancak köşegen üzerindeki elemanları eşit olan varyans-kovaryans yapısını göstermektedir. Bu durum “otokorelasyon” olarak adlandırılmaktadır. Araştırmanın amacı doğrultusunda hata varyans-kovaryans matrisi sayfa 19’da verilen (xiv)’deki gibi alınmıştır.

Hatalar arasındaki ilişkiyi modellemek için kullanılan bir dizi prosedür vardır. Bunlardan en yaygın olan modeller otoregresif modellerdir.

En yaygın olarak kullanılan ve birinci dereceden otoregresif olarak bilinen AR(1) sürecinde hatalar arasındaki ilişki

$$\varepsilon_i = \rho \varepsilon_{i-1} + u_i, \quad |\rho| < 1 \quad (3.2)$$

olarak modellenmektedir. Burada ρ , otokorelasyon katsayısı olarak adlandırılmaktadır. $E(u_i) = 0$, $E(u_i^2) = \sigma_u^2$ ve $E(u_i u_j) = 0, i \neq j$ varsayımları ile ak gürültü hatalarıdır (white noise, WN).

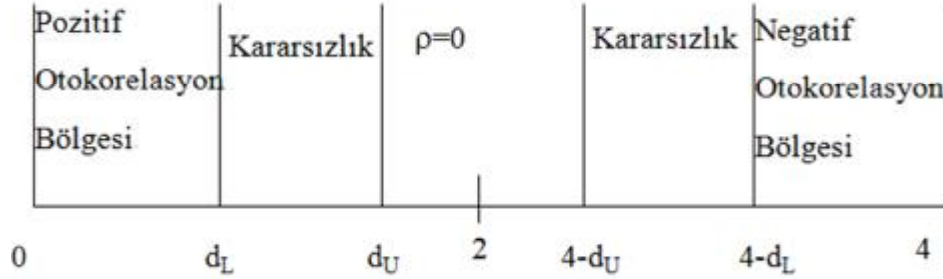
Literatürde birinci dereceden otoregresif hataların belirlenmesinde sıklıkla Durbin-Watson (dw) test istatistiği kullanılmaktadır. Durbin ve Watson (1950) tarafından geliştirilen test istatistiği, yokluk hipotezi $H_0: \rho = 0$ olmak üzere,

$$dw = \frac{\sum_{t=2}^n (\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{t=1}^n \hat{u}_t^2} = \frac{\hat{u}' A \hat{u}}{\hat{u}' \hat{u}}, \quad A = \begin{bmatrix} 1 & -1 & 0 & \cdots & 0 & 0 \\ -1 & 2 & -1 & \cdots & 0 & 0 \\ 0 & -1 & 2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 2 & -1 \\ 0 & 0 & 0 & \cdots & -1 & 1 \end{bmatrix}$$

ile elde edilir. Burada $\hat{u}_t$, EKK ile elde edilen t -inci rezidüdür. dw test istatistiğine ilişkin karar şeması Şekil 3.1. ile verilmiştir.

3. GENEL LİNEER İSTATİSTİKSEL MODEL VE ÇOKLU İÇ İLİŞKİ

Tuğba SÖKÜT AÇAR



Şekil 3.1. dw testi için karar şeması (Gujarati, 2004)

Eğer hesaplanan test istatistiğinin değeri kararsızlık bölgelerine denk geliyorsa Durbin ve Watson (1971)'ın önerdikleri yaklaşım kullanılır. Bu yaklaşımda dw^* ,

$$dw^* = a + bd_U^*$$

formülünden hesaplanır. Burada a ve b ,

$$E(dw) = a + bE(d_U)$$

ve

$$var(dw) = b^2 var(d_U)$$

olacak şekilde seçilir. a ve b 'yi hesaplamak için $E(dw)$, $E(d_U)$, $var(dw)$ ve $var(d_U)$ değerlerine ihtiyaç vardır. Durbin ve Watson (1971) örneklem genişliği n ve açıklayıcı değişken sayısı p 'nin farklı değerleri için $E(d_U)$ ve $var(d_U)$ değerlerini de tablolaştırmışlardır. $E(dw)$ ve $var(dw)$ aşağıdaki şekilde hesaplanır

$$E(dw) = \frac{\Gamma}{n-p}, \Gamma = 2(n-1) - trace[Z^{*'}AZ^*(Z^{*'}Z^*)^{-1}],$$

$$var(dw) = \frac{2}{(n-p)(n-p+2)}[\phi - \Gamma E(dw)].$$

Burada $\Gamma = 2(n-1) - trace[Z^{*'}AZ^*(Z^{*'}Z^*)^{-1}]$ ve $\phi = 2(3n-4) - 2trace[Z^{*'}A^2Z^*(Z^{*'}Z^*)^{-1}] + trace\{Z^{*'}AZ^*(Z^{*'}Z^*)^{-1}\}^2$ dir. dw testinin aşamaları şu şekilde özetlenebilir:

$H_1: \rho > 0$ için test adımları,

3. GENEL LİNEER İSTATİSTİKSEL MODEL VE ÇOKLU İÇ İLİŞKİ

Tuğba SÖKÜT AÇAR

- (1) (2.2) eşitliğinde verilen EKK tahmin edicisi ve ilgili rezidüel ($\hat{u}$) elde edilir.
- (2) İlgili dw istatistik değeri hesaplanır.
- (3) Seçilen önem düzeyi için tablo değerleri olan d_L^* ve d_U^* değerleri elde edilir.
- (4) Eğer $dw < d_L^*$ ise H_0 red edilir aksi durumda red edilemez ve hipotez testi sonuçlanır.
- (5) Eğer $d_L^* < dw < d_U^*$ ise test değeri kararsızlık bölgesine düştüğünden $E(d_u)$, $var(d_u)$ değerleri tablodan belirlenip $E(dw)$ ve $var(dw)$ değerleri ise hesaplanıp a ve b elde edilir.
- (6) $dw^* = a + bd_U^*$ değeri hesaplanır.
- (7) $dw < dw^*$ ise H_0 red edilir aksi durumda red edilemez ve hipotez testi sonuçlanmış olur.

$H_1: \rho < 0$ için test adımları,

- (1) Yukarıda verilen ilk 3 adım tekrarlanır.
- (2) Eğer $dw > 4 - d_L^*$ ise H_0 red edilir aksi durumda red edilemez ve hipotez testi sonuçlanır.
- (3) Eğer $4 - d_U^* < dw < 4 - d_L^*$ ise test değeri kararsızlık bölgesine düştüğünden $E(d_u)$, $var(d_u)$ değerleri tablodan belirlenip $E(dw)$ ve $var(dw)$ değerleri ise hesaplanıp a ve b elde edilir.
- (4) $dw^* = a + b(4 - d_L^*)$ değeri hesaplanır.
- (5) $dw > dw^*$ ise H_0 red edilir aksi durumda red edilemez ve hipotez testi sonuçlanmış olur.

dw istatistiğinin bazı durumlarda dezavantajı vardır. Z matrisi sabit terim içermiyorsa ($Z = X$) ve stokastik olmama varsayımını sağlamıyorsa, modelde bağımsız değişken olarak açıklayıcı değişkenin gecikmeli değerleri yer alıyorsa ve otokorelasyonun derecesi 1'den büyükse dw istatistiğinin gücü azalmaktadır.

3. GENEL LİNEER İSTATİSTİKSEL MODEL VE ÇOKLU İÇ İLİŞKİ

Tuğba SÖKÜT AÇAR

Otokorelasyon varlığı tespit edildikten sonra, problemi ortadan kaldırmak için ilk olarak otokorelasyon katsayısının tahmin edilmesi gerekmektedir. ρ için farklı tahmin yöntemleri Judge ve arkadaşları (1980, s.286-287) tarafından aşağıdaki gibi derlenmiştir.

EKK yöntemi ile rezidüler elde edilir ($\hat{u}$) , (3.2) eşitliğinde verilen modele EKK uygulanması ile otokorelasyon katsayısı

$$\hat{\rho}_1 = \frac{\sum_{t=2}^n \hat{u}_t \hat{u}_{t-1}}{\sum_{t=1}^n \hat{u}_t^2}$$

olarak tahmin edilir. $\hat{\rho}_1$ otokorelasyon katsayısının literatürde sıklıkla kullanılan tahminidir. Theil (1971), ρ_1 'in değiştirilmiş versiyonunu

$$\hat{\rho}_2 = \frac{(n-p) \sum_{t=2}^n \hat{u}_t \hat{u}_{t-1}}{(n-1) \sum_{t=1}^n \hat{u}_t^2}$$

olarak vermişlerdir.

Durbin ve Watson (1950), dw test istatistiğini kullanarak otokorelasyon katsayısının tahminini

$$\hat{\rho}_3 = 1 - \frac{1}{2} dw$$

olarak önermişlerdir.

Theil ve Nagar (1961), dw test istatistiğine dayanarak iyileştirilmiş otokorelasyon katsayısını aşağıdaki gibi vermişlerdir

$$\hat{\rho}_4 = \frac{n^2 \left(1 - \frac{1}{2} dw\right) + p^2}{n^2 - p^2}.$$

V matrisinin bilindiği varsayılın. Bilinmediği durumlar için önce ρ tahmin edilir ve V matrisi biliniyor durumu ile ele alınır. V , simetrik ve pozitif tanımlı bir matris olduğundan dolayı $V^{-1} = P'P$ eşitliğini sağlayan singüler olmayan $(n \times n)$ tipinde P matrisi her zaman vardır. AR(1) sürecinde V , V^{-1} ve P matrislerini Judge ve arkadaşları (1988) aşağıdaki gibi vermişlerdir

3. GENEL LİNEER İSTATİSTİKSEL MODEL VE ÇOKLU İÇ İLİŞKİ

Tuğba SÖKÜT AÇAR

$$V = \frac{1}{1-\rho^2} \begin{bmatrix} 1 & \rho & \dots & \rho^{n-1} \\ \rho & 1 & \dots & \rho^{n-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho^{n-1} & \rho^{n-2} & \dots & 1 \end{bmatrix}$$

$$V^{-1} = \begin{bmatrix} 1 & -\rho & 0 & \dots & 0 & 0 \\ -\rho & 1+\rho^2 & -\rho & \dots & 0 & 0 \\ 0 & -\rho & 1+\rho^2 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1+\rho^2 & -\rho \\ 0 & 0 & 0 & \dots & -\rho & 1 \end{bmatrix}$$

$$P = \begin{bmatrix} \sqrt{1-\rho^2} & 0 & 0 & \dots & 0 & 0 \\ -\rho & 1 & 0 & \dots & 0 & 0 \\ 0 & -\rho & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 & 0 \\ 0 & 0 & 0 & \dots & -\rho & 1 \end{bmatrix}$$

(3.1) eşitliğinde verilen model soldan P ile çarpıldığında,

$$Py = PZ\beta + P\varepsilon$$

ya da

$$y^* = Z^*\beta + \varepsilon^* \quad (3.3)$$

dönüştürülmüş model elde edilir. Burada,

$$y^* = Py = \begin{bmatrix} \sqrt{1-\rho^2}y_1 \\ y_2 - \rho y_1 \\ \vdots \\ y_n - \rho y_{n-1} \end{bmatrix}$$

$$Z^* = PZ = \begin{bmatrix} \sqrt{1-\rho^2} & \sqrt{1-\rho^2}x_{11} & \dots & \sqrt{1-\rho^2}x_{1r} \\ 1-\rho & x_{21} - \rho x_{11} & \dots & x_{2r} - \rho x_{1r} \\ \vdots & \vdots & \ddots & \vdots \\ 1-\rho & x_{n1} - \rho x_{(n-1)1} & \dots & x_{nr} - \rho x_{(n-1)r} \end{bmatrix} \text{ ve}$$

3. GENEL LİNEER İSTATİSTİKSEL MODEL VE ÇOKLU İÇ İLİŞKİ

Tuğba SÖKÜT AÇAR

$$\varepsilon^* = P\varepsilon = \begin{bmatrix} \sqrt{1-\rho^2}\varepsilon_1 \\ \varepsilon_2 - \rho\varepsilon_1 \\ \vdots \\ \varepsilon_n - \rho\varepsilon_{n-1} \end{bmatrix}, \text{dir. (3.3)'de verilen dönüştürülmüş modelde } E(\varepsilon^*) = 0$$

ve $E(\varepsilon^*\varepsilon^{*'}) = \sigma^2 I_n$ varsayımı geçerli olduğundan bu modele uygulanan EKK yöntemi sonucunda elde edilen tahmin edici

$$\hat{\beta} = (Z^{*'}Z^*)^{-1}Z^{*'}y^* = (Z'V^{-1}Z)^{-1}Z'V^{-1}y \quad (3.4)$$

gibi olup genelleştirilmiş en küçük kareler (GEKK) tahmin edici olarak adlandırılmaktadır. GEKK tahmin edici $E(\hat{\beta}) = 0$, $cov(\hat{\beta}) = \sigma^2(Z^{*'}Z^*)^{-1}$ ile (3.3) modeli altında BLUE'dur.

3.2. Otokorelasyon Altında Çoklu İç İlişkinin Belirlenmesi

Trenkler (1984), Bayhan ve Bayhan (1998), Kaçıranlar (2003) ve Özkale (2008) dönüştürülmüş açıklayıcı değişken $Z^* = PZ$ 'nin rankının Z 'nin rankına eşit olduğundan dolayı modeli dönüştürmenin çoklu iç ilişki probleminin şiddetinde önemli farklılığa neden olmadığını belirtmişlerdir.

Çoklu iç ilişkinin belirlenmesine yönelik (2.1) modeli altında kullanılan yöntemler otokorelasyonlu model için genişletilmiştir.

Çoklu iç ilişkiyi belirlemenin en kolay yolu basit korelasyon katsayısını kullanmaktır. Standartlaştırılmış formda $Z^{*'}Z^*$ matrisinin köşegen dışı elemanları (r_{ij}) çoklu iç ilişkinin belirlenmesinde kullanılır. İki açıklayıcı değişken arasında çoklu iç ilişki varsa (r_{ij}) mutlak değerce 1'e yakın olacaktır. Ancak bu yöntemde iki açıklayıcı değişken arasındaki lineer ilişki incelenebilmektedir. Eğer ikiden fazla açıklayıcı değişken arasında bir ilişki varsa, bu ilişkinin basit korelasyon katsayılarına yansıtacağı konusunda kesin bir bilgi yoktur.

Çoklu iç ilişkinin belirlenmesinde kullanılan popüler bir yöntem varyans şişirme faktörüdür (VIF). R_j^2 , z_j^* ($j = 1, 2, \dots, p$) açıklayıcı değişkenin diğer $(p - 1)$ açıklayıcı değişkenler ile regresyonundan elde edilen çoklu belirleyicilik katsayısı olmak üzere, varyans şişirme faktörü

3. GENEL LİNEER İSTATİSTİKSEL MODEL VE ÇOKLU İÇ İLİŞKİ

Tuğba SÖKÜT AÇAR

$$VIF_j = \frac{1}{(1 - R_j^2)}$$

formülü ile bulunur. Eğer z_j^* açıklayıcı değişkeni ile diğer açıklayıcı değişkenler arasında doğrusal bir ilişki yoksa R_j^2 çok küçük olacağından dolayı VIF_j değeri 1'e yaklaşacaktır. Genelde 10'dan büyük VIF_j değerleri, j -inci regresyon katsayısı tahmininin çoklu iç ilişkiden dolayı güvenli bir tahmin olmadığına bir göstergesi olarak kabul edilir.

$Z^{*'}Z^*$ matrisinin özdeğerleri $\lambda_1, \lambda_2, \dots, \lambda_p$ olmak üzere, en büyük özdeğerin en küçük özdeğere oranı olan koşul sayısı (KS) çoklu iç ilişkinin belirlenmesinde sıklıkla başvuru yöntemidir. KS değerinin yorumlanması (2.1) alt başlığındaki ile aynıdır.

3.3. Yanlı Tahmin Ediciler

Otokorelasyonlu hataya sahip lineer istatistiksel modelde açıklayıcı değişkenlerin lineer bağımlı olması durumunda GEKK'ye alternatif tahmin yöntemlerinin bir kısmı Trenkler (1984) tarafından önerilmiştir.

Bu araştırmada Trenkler (1984)'in tanımladığı ridge regresyon tahmin edici Brown (1977)'ye paralel olarak (3.3) modeli altında

$$\hat{\beta}^R = (Z^{*'}Z^* + kI_p^*)^{-1}Z^{*'}y^* = (Z'V^{-1}Z + kI_p^*)^{-1}Z'V^{-1}y, \quad k > 0 \quad (3.5)$$

olarak ele alınmıştır. (3.5)'de verilen tahmin edici otokorelasyonlu ridge regresyon (ARR) tahmin edici olarak adlandırılmıştır. ARR tahmin edici

$$bias(\hat{\beta}^R) = E(\hat{\beta}^R) - \beta = -k(Z^{*'}Z^* + kI_p^*)^{-1}\beta$$

ile yanlı bir tahmin edicidir. Varyans-kovaryans matrisi ise

$$\begin{aligned} cov(\hat{\beta}^R) &= (Z^{*'}Z^* + kI_p^*)^{-1}Z^{*'}var(y^*)Z^*(Z^{*'}Z^* + kI_p^*)^{-1} \\ &= \sigma^2(Z^{*'}Z^* + kI_p^*)^{-1}Z^{*'}Z^*(Z^{*'}Z^* + kI_p^*)^{-1} \end{aligned}$$

ile elde edilir.

3. GENEL LİNEER İSTATİSTİKSEL MODEL VE ÇOKLU İÇ İLİŞKİ

Tuğba SÖKÜT AÇAR

Ridge yanlılık parametresi olan k 'nın seçimi için birçok metot geliştirilmiştir. Firinguetti (1989), varyansdaki azalışın yanlılıktaki artıştan daha büyük olmasını sağlayan yanlılık parametresi için $\hat{\beta}$, GEKK tahmin edicisi, $s^2 = \frac{(y - Z\hat{\beta})'V^{-1}(y - Z\hat{\beta})}{(n-p)}$ ise σ^2 'nin GEKK tahmin edicisi olmak üzere, $k_1 = \frac{ps^2}{\hat{\beta}'\hat{\beta}}$ ve

$k_2 = \frac{ps^2}{\hat{\beta}'Z'V^{-1}Z\hat{\beta}}$ gibi iki yöntem önermişlerdir.

Hataların varyans-kovaryans matrisinin $E(uu') = \sigma_u^2 I_n$ olduğu varsayımı altında Marquardt (1970)'in önerdiği temel bileşenler tahmin edicisini Trenkler (1984) hataların varyans-kovaryans matrisinin $E(\varepsilon\varepsilon') = \sigma_\varepsilon^2 V$ olması durumunda

$$\hat{\beta}_{TBA} = (Z^{*'}Z^*)_p^+ Z^{*'}y^*$$

olarak tanımlamıştır. Burada p , $Z^{*'}Z^*$ matrisinin rankı olmak üzere $(Z^{*'}Z^*)_p^+$, $(Z^{*'}Z^*)$ matrisinin genelleştirilmiş tersini göstermektedir. $Z^{*'}Z^*$ matrisinin spektral ayrışımı kullanılarak, $\left(Z^{*'}Z^* = U\Lambda U' = \begin{pmatrix} U_1 & U_2 \end{pmatrix} \begin{pmatrix} \Lambda_1 & 0 \\ 0 & \Lambda_2 \end{pmatrix} \begin{pmatrix} U_1' \\ U_2' \end{pmatrix} \right)$, $E(\varepsilon\varepsilon') = \sigma_\varepsilon^2 V$ altında temel bileşenler tahmin edici

$$\hat{\beta}_{TBA} = U_1(U_1'Z^{*'}Z^*U_1)^{-1}U_1'Z^{*'}y^*$$

olarak elde edilir.

Hataların varyans-kovaryans matrisinin $E(uu') = \sigma_u^2 I_n$ olduğu varsayımı altında Singh ve arkadaşlarının (1986) önerdiği jackknifed ridge tahmin edicisini Özkale (2008) hataların varyans-kovaryans matrisinin $E(\varepsilon\varepsilon') = \sigma_\varepsilon^2 V$ olması durumunda

$$\hat{\beta}_J = \left[I_p - k^2(Z'V^{-1}Z + kI_p^*)^{-2} \right] \hat{\beta}$$

olarak önermiştir. Burada $\hat{\beta}$, (3.4) eşitliğinde verilen GEKK tahmin edicidir.

Hataların varyans-kovaryans matrisinin $E(uu') = \sigma_u^2 I_n$ olduğu varsayımı altında Mayer ve Wilke (1973)'nin önerdiği shrunk tahmin edicisini Trenkler (1984) hataların varyans-kovaryans matrisinin $E(\varepsilon\varepsilon') = \sigma_\varepsilon^2 V$ olması durumunda

3. GENEL LİNEER İSTATİSTİKSEL MODEL VE ÇOKLU İÇ İLİŞKİ

Tuğba SÖKÜT AÇAR

$$\hat{\beta}_\lambda = \lambda(Z^{*'}Z^*)^{-1}Z^{*'}y^* = \lambda\hat{\beta}$$

olarak vermiştir.

Hataların varyans-kovaryans matrisinin $E(uu') = \sigma_u^2 I_n$ olduğu varsayımı altında Trenkler (1978)'in önerdiği iterasyon tahmin edicisini Trenkler (1984) hataların varyans-kovaryans matrisinin $E(\varepsilon\varepsilon') = \sigma_\varepsilon^2 V$ olması durumunda

$$\hat{\beta}_{m,\alpha} = \alpha \sum_{i=0}^m (I_p - \alpha Z^{*'}Z^*)^i Z^{*'}y^*, m = 0, 1, \dots$$

olarak vermiştir. Burada λ_{maks} , $Z^{*'}Z^*$ matrisinin en büyük özdeğeri olmak üzere

$0 < \alpha < \frac{1}{\lambda_{maks}}$ olarak tanımlanır.

4. GENEL LİNEER İSTATİSTİKSEL MODELDE TANILAMA METOTLARI

Tuğba SÖKÜT AÇAR

4. GENEL LİNEER İSTATİSTİKSEL MODELDE TANILAMA METOTLARI

Hatalara ilişkin sabit varyans ve ilişkisizlik varsayımlarının ihlal edildiği istatistiksel model altında etki tanılama ölçümleri istatistik literatüründe kısmen yer almaktadır. Puterman (1988), AR(1) hatalı ortalama sabit model $y = \beta_0 \mathbf{1} + \varepsilon$, $\mathbf{1}$, $(n \times 1)$ tipindeki $\mathbf{1}$ 'lerin vektörüdür) ve orijine doğru regresyon modelinde ($y = \beta_0 x_1 + \varepsilon$) ilk gözlemin kaldıraç nokta olma durumunu ve GEKK ile tahmin edilen regresyon parametreleri üzerindeki etkinliğini incelemiştir. Stemann ve Trenkler (1993), birden fazla açıklayıcı değişken içeren AR(1) hatalı istatistiksel modelde otokorelasyon katsayısının ilk gözleme etkisini araştırmıştır. Barry ve ark. (1997), Puterman (1988)'in çalışmasını AR(2) hatalı istatistiksel modele uyarlamıştır. Haslett (1999), ilişkili hataya sahip regresyon modellerinde etkin gözlemleri alt kümeler halinde modele dâhil edip DFBETA ve Cook's D istatistiklerini ele almıştır. Roy ve Guria (2004), AR(1) hatalı istatistiksel modelde yanıt değişken üzerindeki etkin gözlemleri incelemiştir. Literatürde AR(1) hatalı istatistiksel modelde yanlış tahmin ediciler için etki tanılama ölçümlerine değinilmemiştir. Bu bölümde ARR yanlış tahmin edici kullanıldığında yanıt değişken ve regresyon katsayısı üzerinde etkin gözlemi belirlemek için ölçümler önerilmiştir. Buna ek olarak, güven elipsoidi üzerindeki etkin gözlemleri belirlemek için kullanılan COVRATIO, Cook-Weisberg ve Andrews-Pregibon istatistikleri genel lineer istatistiksel model altında daha önce ele alınmamıştır. Bu nedenden dolayı güven elipsoidi üzerindeki etkin gözlemi belirlemek için daha önce sabit varyanslı ve ilişkisiz model altında önerilen ölçümler, hem GEKK hem de ARR tahmin ediciler için önerilmiştir.

4.1. Kaldıraç Gözlem

(3.4) eşitliğinde verilen GEKK tahmin edici kullanılarak elde edilen $\hat{y}^*$ değerleri

$$\hat{y}^* = Z^* \hat{\beta} = Hy^*$$

dir. Burada,

$$H = PZ(Z'P'PZ)^{-1}Z'P'$$

olup GEKK tahmin edici için şapka matrisini göstermektedir (Puterman, 1988; Stemmann ve Trenkler, 1993; Barry ve ark., 1997). H matrisi sabit varyanslı ve ilişkisiz model altında elde edilen L matrisi ile benzer özelliklere sahiptir:

- xv. H simetrik bir matristir ($H' = H$)
- xvi. H idempotent bir matristir ($H^2 = H$)
- xvii. $trace(H) = rank(H) = p$
- xviii. $\sum_j h_{i,j}^2 = h_{i,i}$, $i = 1, \dots, n$ $j = 1, \dots, n$
- xix. $0 \leq h_{i,i} \leq 1$, $i = 1, \dots, n$.

(3.5) eşitliğinde verilen ARR tahmin edicisini kullanılarak tahmin edilen yanıt değerler ise

$$\hat{y}^{R*} = Z^* \hat{\beta}^R = H^R y^*$$

olarak elde edilir. Burada,

$$H^R = PZ(Z'P'PZ + kI_p^*)^{-1}Z'P'$$

dir. H ve H^R matrisleri y^* vektörünü sırası ile $\hat{y}^*$ ve $\hat{y}^{R*}$ vektörüne dönüştürmektedir. H matrisi simetrik ve idempotent iken H^R matrisi idempotent bir matris değildir. $\hat{y}^{R*}$ 'nin i -inci değeri H^R 'nin elemanları cinsinden yazıldığında

$$\hat{y}^{R*} = \sum_{j=1}^n h_{i,j}^R y_j^*$$

4. GENEL LİNEER İSTATİSTİKSEL MODELDE TANILAMA METOTLARI

Tuğba SÖKÜT AÇAR

elde edilir. Burada $h_{i,j}^R$, H^R matrisinin i -inci satır, j -nci sütun elemanını göstermektedir. $h_{i,i}$ ve $h_{i,i}^R$ sırası ile H ve H^R matrisinin i -inci köşegen elemanı olmak üzere $h_{i,i}$, $h_{i,i}^R > \frac{2p}{n}$ ise i -inci gözlem kaldıraç olarak adlandırılır (Puterman, 1988).

4.2. Regresyon Parametrelerine Dayalı Tanılama

Genel lineer istatistiksel modelde veri kümesinde yer alan gözlemlerin tahmin edilen regresyon parametreleri üzerine etkisini belirlemek için DFBETA ölçümü kullanılmıştır. DFBETA, silinen bir gözlemin ilgili regresyon katsayısını nasıl etkilediğini gösterir. Lineer istatistiksel modelde DFBETA ölçümü ilk olarak Belsley ve arkadaşları (1980) tarafından önerilmiştir. Bu ölçüm i -inci gözlem veri kümesinde mevcutken ve veri kümesinden çıkarıldığı zaman β vektörünün tahminleri arasındaki farkı ifade etmektedir.

Bu bölümde, birden çok açıklayıcı değişken içeren lineer istatistiksel modelde hataların AR(1) sürecinden geldiği durumlar için DFBETA ölçümü ele alınmıştır. Ayrıca, Puterman (1988)'ın genel lineer modelden indirgediği iki özel model verilmiştir.

Teorem 4.1. AR(1) hatalı GEKK tahmin edicisi kullanıldığında DFBETA ölçümü,

$$DFBETA_i = \begin{cases} \sqrt{1 - \rho^2} (Z^{*'} Z^*)^{-1} \begin{pmatrix} 1 \\ x_{11} \\ x_{12} \\ \vdots \\ x_{1r} \end{pmatrix} e_i^* (1 - h_{i,i})^{-1}, & i = 1 \\ (Z^{*'} Z^*)^{-1} \begin{pmatrix} 1 - \rho \\ x_{i1} - \rho x_{(i-1)1} \\ x_{i2} - \rho x_{(i-1)2} \\ \vdots \\ x_{ir} - \rho x_{(i-1)r} \end{pmatrix} e_i^* (1 - h_{i,i})^{-1}, & i = 2, \dots, n \end{cases} \quad (4.1)$$

olarak önerilmiştir.

İspat 4.1. Genel lineer model altında $DFBETA_i$ ölçümü incelensin. z_i^* , Z^* dönüştürülmüş matrisinin i -inci satır vektörünü gösterebilir. i -inci gözlem

4. GENEL LİNEER İSTATİSTİKSEL MODELDE TANILAMA METOTLARI

Tuğba SÖKÜT AÇAR

çıkartıldığında GEKK tahmin edici $\hat{\beta}_{(i)} = (Z_{(i)}^*{}' Z_{(i)}^*)^{-1} Z_{(i)}^*{}' y_{(i)}^*$ öyle ki $(Z_{(i)}^*{}' Z_{(i)}^*)^{-1} = (Z^{*'} Z^* - z_i^{*'} z_i^*)^{-1}$ ve $Z_{(i)}^*{}' y_{(i)}^* = Z^{*'} y^* - z_i^{*'} y_i^*$ olmak üzere $\hat{\beta} - \hat{\beta}_{(i)}$ farkı elde edilmek istensin.

$$\hat{\beta}_{(i)} = (Z^{*'} Z^* - z_i^{*'} z_i^*)^{-1} (Z^{*'} y^* - z_i^{*'} y_i^*)$$

şeklinde ifade edilir. $Z^{*'} Z^*$, singüler olmayan bir matris olduğundan dolayı Ek-Lemma 1'de verilen SMV teoremi uygulandığında,

$$\begin{aligned} \hat{\beta}_{(i)} &= \left[(Z^{*'} Z^*)^{-1} + \frac{(Z^{*'} Z^*)^{-1} z_i^{*'} z_i^* (Z^{*'} Z^*)^{-1}}{1 - z_i^* (Z^{*'} Z^*)^{-1} z_i^{*'}} \right] (Z^{*'} y^* - z_i^{*'} y_i^*) \\ &= \hat{\beta} - \frac{(Z^{*'} Z^*)^{-1} z_i^{*'} (y_i^* - \hat{y}_i^*)}{1 - h_{i,i}} \\ &= \hat{\beta} - \frac{(Z^{*'} Z^*)^{-1} z_i^{*'} e_i^*}{1 - h_{i,i}}. \end{aligned} \quad (4.2)$$

olarak elde edilir.

$$DFBETA_i = \hat{\beta} - \hat{\beta}_{(i)}$$

olmak üzere, (4.2) eşitliği $DFBETA_i$ 'de yerine yazılırsa

$$DFBETA_i = \frac{(Z^{*'} Z^*)^{-1} z_i^{*'} e_i^*}{1 - h_{i,i}}$$

eşitliği elde edilir. Dönüştürülmüş açıklayıcı değişken matrisi Z^* göz önüne alınılıp ilk gözlem ayrı tutulduğunda Teorem 4.1 ispatlanmış olunur.

Teorem 4.2. AR(1) hatalı ARR tahmin edicisi kullanıldığında DFBETA ölçümü

$$DFBETA_i^R = \frac{(Z^{*'} Z^* + kI_p)^{-1} z_i^{*'} e_i^{R*}}{1 - h_{i,i}^R}$$

olarak önerilmiştir. Burada e_i^{R*} , ARR ile elde edilen i -inci rezidüyü ifade etmektedir.

İspat 4.2. (3.5) eşitliğinde verilen ARR tahmin edici için DFBETA

$$DFBETA_i^R = \hat{\beta}^R - \hat{\beta}_{(i)}^R, \quad i = 1, \dots, n, j = 1, \dots, p.$$

ile elde edilir. i -inci gözlem veri kümesinden çıkartıldığında ARR tahmin edici

$$\begin{aligned}\hat{\beta}_{(i)}^R &= (Z_{(i)}^{*'} Z_{(i)}^* + kI_p^*)^{-1} Z_{(i)}^{*'} y_{(i)}^* \\ &= (Z^{*'} Z^* + kI_p^* - z_i^{*'} z_i^*)^{-1} (Z^{*'} y^* - z_i^{*'} y_i^*)\end{aligned}$$

olup, $(Z^{*'} Z^* + kI_p^* - z_i^{*'} z_i^*)^{-1}$ için Ek-Lemma 1'de verilen SMV teoremi kullanıldığında $\hat{\beta}_{(i)}^R$ aşağıdaki gibi elde edilir

$$\begin{aligned}\hat{\beta}_{(i)}^R &= \left[(Z^{*'} Z^* + kI_p^*)^{-1} + \frac{(Z^{*'} Z^* + kI_p^*)^{-1} z_i^{*'} z_i^* (Z^{*'} Z^* + kI_p^*)^{-1}}{1 - h_{i,i}^R} \right] \\ &\quad \times (Z^{*'} y^* - z_i^{*'} y_i^*) \\ &= \hat{\beta}^R - \frac{(Z^{*'} Z^* + kI_p^*)^{-1} z_i^{*'} e_i^{R*}}{1 - h_{i,i}^R}\end{aligned}\tag{4.3}$$

ve Teorem 4.2 ispatlanır.

4.2.1. GEKK için Bazı Özel Durumlar

Puterman (1988) kaldıraç ve DFBETA ölçümünü, AR(1) hataya sahip ortalama sabit model ve orijine doğru regresyon modeli üzerinden ele almıştır. Araştırmanın bu bölümünde, X matrisinin özel durumlarına inerek Puterman (1988)'in sonuçları ele alınmıştır.

AR(1) hataları ile $y = \beta_0 \mathbf{1} + \varepsilon$ modeli altında $h_{i,i}$ ve (4.1) eşitliğinin özel hali olarak yazılabilen $DFBETA_i$ ölçümü sırası ile aşağıda verilmiştir (Puterman, 1988),

$$h_{i,i} = \begin{cases} \frac{1+\rho}{n-(n-2)\rho} & i = 1 \\ \frac{1-\rho}{n-(n-2)\rho} & i = 2, \dots, n \end{cases}$$

$$DFBETA_i = \begin{cases} \sqrt{\frac{1+\rho}{1-\rho}} \frac{e_i^*}{(n-1)(1-\rho)} & i = 1 \\ \frac{e_i^*}{(n-1)-(n-3)\rho} & i = 2, \dots, n \end{cases}.$$

AR(1) hatalı $y = \beta_0 x_1 + \varepsilon$ modelinde $h_{i,i}$ ve $DFBETA_i$ ölçümü sırası ile aşağıda verilmiştir (Puterman, 1988),

4. GENEL LİNEER İSTATİSTİKSEL MODELDE TANILAMA METOTLARI
Tuğba SÖKÜT AÇAR

$$h_{i,i} = \begin{cases} \frac{(1-\rho)^2 x_{11}^2}{(1-\rho)^2 x_{11}^2 + \sum_{j=2}^n (x_{j1} - \rho x_{(j-1)1})^2} & i = 1 \\ \frac{(x_{i1} - \rho x_{(i-1)1})^2}{(1-\rho)^2 x_{11}^2 + \sum_{j=2}^n (x_{j1} - \rho x_{(j-1)1})^2} & i = 2, \dots, n \end{cases}$$

$$DFBETA_i = \begin{cases} \frac{\sqrt{1-\rho^2} x_{11} e_1^*}{\sum_{j=2}^n (x_{j1} - \rho x_{(j-1)1})^2} & i = 1 \\ \frac{(x_{i1} - \rho x_{(i-1)1})^2 e_i^*}{(1-\rho)^2 x_{11}^2 + \sum_{j=2, j \neq i}^n (x_{j1} - \rho x_{(j-1)1})^2} & i = 2, \dots, n \end{cases}.$$

Puterman (1988), AR(1) hatalı orijine doğru regresyon modelinde $\lim_{\rho \rightarrow 1} h_{1,1} = 0$ olduğunu belirtirken, AR(1) hatalı ortalama sabit modelde $\lim_{\rho \rightarrow 1} h_{1,1} = 1$ olduğunu belirtmiştir. Stemann ve Trenkler (1993), Puterman (1988)'ın sonuçlarını birden fazla açıklayıcı değişken içeren modele genişleterek sabit terimin varlığında ve yokluğunda ilk gözlemin kaldıraçlılığını incelemiş ve sabit terim modelde olduğu zaman $\lim_{\rho \rightarrow 1} h_{1,1} = 1$, sabit terim modelde olmadığı zaman ise $\lim_{\rho \rightarrow 1} h_{1,1} = 0$ olduğu sonucunu elde etmişlerdir.

4.3. Yanıt Değişkene Dayalı Tanılama

4.3.1. DFFITS

Tahmin edilen yanıt değer üzerinde etkin olan gözlemi belirlemek amacı ile (3.3) eşitliğinde verilen dönüştürülmüş model altında DFFITS ölçümü ele alınmıştır. DFFITS tek gözlemin etkisini inceleyen popüler bir ölçüdür. Tüm gözlemleri içeren modelden elde edilen i -inci gözlemin tahmin edilen yanıt değeri ile i -inci gözlem veri kümesinden çıkarıldıktan sonra elde edilen i -inci gözlemin uydurulan değeri arasındaki farkı ölçer.

(3.4) eşitliğinde verilen GEKK tahmin edicisi kullanıldığında elde edilen DFFITS istatistiğini Roy ve Gruia (2004),

$$DFFITS_i = \frac{\hat{y}_i^* - \hat{y}_{(i)}^*}{SE(z_i^* \hat{\beta})}$$

şeklinde vermiştir.

Teorem 4.3. AR(1) hatalı ARR tahmin edicisi kullanıldığında DFFITS istatistiği

4. GENEL LİNEER İSTATİSTİKSEL MODELDE TANILAMA METOTLARI

Tuğba SÖKÜT AÇAR

$$DFFITS_i^R = \left(\frac{h_{i,i}^R}{1 - h_{i,i}^R} \right) \frac{e_i^{R*}}{s_{(i)} \left[\sum_{j=1}^n (h_{i,j}^R)^2 \right]^{\frac{1}{2}}}$$

olarak önerilmiştir. Burada, $s_{(i)}$, i -inci gözlem silindiğinde σ 'nın GEKK tahmin edicisidir.

İspat 4.3. (3.5) eşitliğinde verilen ARR tahmin edicisi kullanıldığında

$$DFFITS_i^R = \frac{\hat{y}_i^{R*} - \hat{y}_{(i)}^R}{SE(z_i^* \hat{\beta}^R)} = \frac{z_i^* [\hat{\beta}^R - \hat{\beta}_{(i)}^R]}{SE(z_i^* \hat{\beta}^R)}, i = 1, \dots, n, j = 1, \dots, p$$

olarak yazılır. Payda kısmı $var(\hat{y}_i^{R*}) = \sigma^2 \sum_{j=1}^n h_{i,j}^2$ olduğundan dolayı standartlaştırmayı göstermektedir. ARR ile uydurulan i -inci yanıt değerinin standart sapması

$$\begin{aligned} SE(z_i^* \hat{\beta}^R) &= s \left[z_i^* (Z^{*'} Z^* + kI_p^*)^{-1} Z^{*'} Z^* (Z^{*'} Z^* + kI_p^*)^{-1} z_i^{*'} \right]^{1/2} \\ &= \sigma \left[\sum_{j=1}^n h_{i,j}^2 \right]^{1/2} \end{aligned}$$

olmak üzere i -inci gözlemin silindiği durum ile

$$DFFITS_i^R = \frac{z_i^* [\hat{\beta}^R - \hat{\beta}_{(i)}^R]}{\sigma_{(i)} \left[\sum_j h_{i,j}^2 \right]^{1/2}}$$

elde edilir. Hata kareler ortalaması yanıt değerler ve tahmin edilen yanıt değerlerin bir fonksiyonudur. Tahmin edilen yanıt değerler ile gerçek yanıt değerler $Z^{*'} Z^*$ 'ın özdeğerlerine bağlı değildir yani çoklu iç ilişkiyi etkilemez. Bu nedenle, standartlaştırmada σ 'nın GEKK tahmin edicisi kullanılacaktır (Walker ve Birch,

1988). i -inci gözlem silindiğinde σ 'nın GEKK tahmin edicisi $s_{(i)} = \sqrt{\frac{(n-p)s^2 - \frac{e_i^{*2}}{1-h_{i,i}}}{n-p-1}}$

(EK-Teorem 1) olarak elde edilmektedir. Eşitlik (4.3)'ün kullanılması ve gerekli düzenlemelerin yapılması ile Teorem 4.3'ün ispatı tamamlanır.

4.3.2. Cook Uzaklığı

Etkinliğin belirlenmesinde sıklıkla kullanılan diğer bir ölçüm Cook uzaklığıdır. Cook (1977)'un önerdiği bu ölçüm, y ve X -uzayında gözlemin konumunu belirlemektedir. (3.3) eşitliğinde verilen dönüştürülmüş model altında GEKK tahmin edicisi için Cook uzaklığı Haslett (1999) tarafından aşağıdaki şekilde verilmiştir,

$$D_i = \frac{(\hat{\beta} - \hat{\beta}_{(i)})' Z' P' P Z (\hat{\beta} - \hat{\beta}_{(i)})}{ps^2}.$$

Teorem 4.4. AR(1) hatalı ARR tahmin edicisi kullanıldığında Cook uzaklık ölçümü

$$D_i^R = \frac{\sum_j (h_{i,j}^R)^2 (e_i^{R*})^2}{(1 - h_{i,i}^R)^2} \frac{1}{ps^2}$$

olarak önerilmiştir.

İspat 4.4. (3.5) eşitliğinde verilen ARR tahmin edicisi kullanıldığında

$$D_i^R = \frac{(\hat{\beta}^R - \hat{\beta}_{(i)}^R)' Z' P' P Z (\hat{\beta}^R - \hat{\beta}_{(i)}^R)}{ps^2}, \quad i = 1, \dots, n, j = 1, \dots, p$$

olarak yazılır. (4.3) eşitliğini kullanarak $\hat{\beta}^R - \hat{\beta}_{(i)}^R = \frac{(Z^{*'} Z^* + kI_p^*)^{-1} Z_i^{*'} e_i^{R*}}{1 - h_{i,i}^R}$ farkı D_i^R 'de

yerine yazılırsa

$$D_i^R = \frac{e_i^{R*'} Z_i^* (Z^{*'} Z^* + kI_p^*)^{-1} Z^{*'} Z^* (Z^{*'} Z^* + kI_p^*)^{-1} Z_i^{*'} e_i^{R*}}{(1 - h_{i,i}^R)^2 ps^2}$$

elde edilir ve gerekli düzenlemelerin yapılması ile Teorem 4.4'ün ispatı tamamlanır.

4.4. Güven Elipsoidine Dayalı Tanılama

Tahmin edilen regresyon katsayıları üzerinde i -inci gözlemin etkisini belirlemenin bir ölçümü i -inci gözlem veri kümesinde iken ve i -inci gözlem veri kümesinden çıkartılmasıyla güven elipsoidindeki değişime dayandırılabilir

4. GENEL LİNEER İSTATİSTİKSEL MODELDE TANILAMA METOTLARI

Tuğba SÖKÜT AÇAR

(Chatterjee ve Hadi, 1986). Güven elipsoidine dayalı etkinlik ölçümleri COVRATIO, Cook-Weisberg ve Andrews-Pregibon olmak üzere üç başlık altında önerilmiştir.

4.4.1. COVRATIO

Teorem 4.5. AR(1) hatalı GEKK tahmin edicisi için COVRATIO istatistiği

$$COVRATIO_i = \left\{ \frac{n-p}{n-p-1+r_i^{*2}} \right\}^p \frac{1}{1-h_{i,i}}$$

olarak önerilmiştir. Burada r_i^* , GEKK ile elde edilen i –inci harici student türü rezidüdür (externally studentized residual).

İspat 4.5. (3.4) eşitliğinde verilen GEKK tahmin edicisi için COVRATIO istatistiği,

$$COVRATIO_i = \frac{\det[cov(\hat{\beta}_{(i)})]}{\det[cov(\hat{\beta})]}$$

olarak yazılabilir. $cov(\hat{\beta}) = s^2(Z^{*'}Z^*)^{-1}$ olmak üzere i -nci gözlem veri kümesinden çıkartıldığında $cov(\hat{\beta}_{(i)}) = s_{(i)}^2(Z_{(i)}^{*'}Z_{(i)}^*)^{-1}$ elde edilir. Bu durumda

$$COVRATIO_i = \frac{\det[(Z_{(i)}^{*'}Z_{(i)}^*)^{-1}s_{(i)}^2]}{\det[(Z^{*'}Z^*)^{-1}s^2]}$$

şeklinde yazılır. Determinant özelliklerinden Ek-Lemma 6'nın kullanılması ile

$$COVRATIO_i = \frac{(s_{(i)}^2)^p \left[\frac{1}{\det(Z_{(i)}^{*'}Z_{(i)}^*)} \right]}{(s^2)^p \left[\frac{1}{\det(Z^{*'}Z^*)} \right]} = \left(\frac{s_{(i)}^2}{s^2} \right)^p \left(\frac{\det(Z^{*'}Z^*)}{\det(Z_{(i)}^{*'}Z_{(i)}^*)} \right) \quad (4.4)$$

yazılır. (4.4) eşitliğindeki $\left(\frac{s_{(i)}^2}{s^2} \right)^p$ oranı ele alınsın. GEKK için harici student türü

rezidü $r_i^* = \frac{e_i^*}{s_{(i)}\sqrt{1-h_{i,i}}}$ olmak üzere $[n-p-1+r_i^{*2}]s_{(i)}^2 = (n-p)s^2$ dir. Bu durumda,

$$\left(\frac{s_{(i)}^2}{s^2} \right)^p = \left\{ \frac{(n-p)}{n-p-1+r_i^{*2}} \right\}^p \quad (4.5)$$

4. GENEL LİNEER İSTATİSTİKSEL MODELDE TANILAMA METOTLARI

Tuğba SÖKÜT AÇAR

olarak elde edilir. Eşitlik (4.4)'ün ikinci parçası olan $\left(\frac{\det(Z^{*'}Z^*)}{\det(Z_{(i)}^{*'}Z_{(i)}^*)} \right)$ oranı ele alınsın. Ek-Lemma 2'nin uygulanması ile

$$\begin{aligned}\det(Z^{*'}Z^* - z_i'z_i) &= \det(Z^{*'}Z^*)(1 - z_i(Z^{*'}Z^*)^{-1}z_i') \\ &= \det(Z^{*'}Z^*)(1 - h_{i,i})\end{aligned}$$

bulunur. Bu durumda eşitlik (4.4)'deki ikinci parça,

$$\left(\frac{\det(Z^{*'}Z^*)}{\det(Z_{(i)}^{*'}Z_{(i)}^*)} \right) = \frac{1}{(1-h_{i,i})} \quad (4.6)$$

olarak elde edilir. (4.5) ve (4.6) eşitlikleri (4.4)'de yerine yazılarak Teorem 4.5'in ispatı tamamlanır.

Teorem 4.6. AR(1) hatalı ARR tahmin edicisi için COVRATIO istatistiği

$$COVRATIO_i^R = \left\{ \frac{n-p}{n-p-1+r_i^{*2}} \right\}^p \frac{1-h_{i,i}}{(1-h_{i,i}^R)^2}$$

olarak önerilmiştir. Burada r_i^* , GEKK ile elde edilen i -inci harici student türü rezidüdür.

İspat 4.6. (3.5) eşitliğinde verilen ARR için i -inci gözlem veri kümesinde iken ve çıkartıldığında varyans-kovaryans matrisinin determinantları oranı

$$COVRATIO_i^R = \frac{\det[cov(\hat{\beta}_{(i)}^R)]}{\det[cov(\hat{\beta}^R)]}$$

olarak yazılabilir. Eşitlik (3.5) ve (4.3)'te verilen tahmin edicilerin varyans-kovaryans matrisleri sırası ile aşağıdaki gibi elde edilir:

$$\begin{aligned}cov(\hat{\beta}^R) &= s^2(Z^{*'}Z^* + kI_p^*)^{-1}Z^{*'}Z^*(Z^{*'}Z^* + kI_p^*)^{-1} \\ cov(\hat{\beta}_{(i)}^R) &= (Z_{(i)}^{*'}Z_{(i)}^* + kI_p^*)^{-1}Z_{(i)}^{*'}var(y_{(i)}^*)Z_{(i)}^*(Z_{(i)}^{*'}Z_{(i)}^* + kI_p^*)^{-1} \\ &= s_{(i)}^2(Z_{(i)}^{*'}Z_{(i)}^* + kI_p^*)^{-1}Z_{(i)}^{*'}Z_{(i)}^*(Z_{(i)}^{*'}Z_{(i)}^* + kI_p^*)^{-1}.\end{aligned}$$

Bu durumda

$$\begin{aligned}
 COVRATIO_i^R &= \frac{\det \left[(Z^{*'}_{(i)} Z^*_{(i)} + kI_p^*)^{-1} Z^{*'}_{(i)} Z^*_{(i)} (Z^{*'}_{(i)} Z^*_{(i)} + kI_p^*)^{-1} s_{(i)}^2 \right]}{\det \left[(Z^{*'} Z^* + kI_p^*)^{-1} Z^{*'} Z^* (Z^{*'} Z^* + kI_p^*)^{-1} s^2 \right]} \\
 &= \frac{\det \left[(Z^{*'}_{(i)} Z^*_{(i)} + kI_p^*)^{-1} \right] \det(Z^{*'}_{(i)} Z^*_{(i)}) \det \left[(Z^{*'}_{(i)} Z^*_{(i)} + kI_p^*)^{-1} \right]}{\det \left[(Z^{*'} Z^* + kI_p^*)^{-1} \right] \det(Z^{*'} Z^*) \det \left[(Z^{*'} Z^* + kI_p^*)^{-1} \right]} \\
 &\quad \times \left(\frac{s_{(i)}^2}{s^2} \right)^p \tag{4.7}
 \end{aligned}$$

olarak yazılır.

(4.7) eşitliğindeki $\det \left[(Z^{*'}_{(i)} Z^*_{(i)} + kI_p^*)^{-1} \right]$ inceleyelim. $(Z^{*'}_{(i)} Z^*_{(i)} + kI_p^*)^{-1} = (Z^{*'} Z^* - z'_i z_i + kI_p^*)^{-1}$ olduğundan

$$\det \left[(Z^{*'}_{(i)} Z^*_{(i)} + kI_p^*)^{-1} \right] = \det \left[(Z^{*'} Z^* - z'_i z_i + kI_p^*)^{-1} \right]$$

dir. $A = Z^{*'} Z^* + kI_p^*$, $u = z'_i$ ve $v' = z_i$ olmak üzere Ek-Lemma 2 uygulanırsa

$$\begin{aligned}
 \det(Z^{*'} Z^* + kI_p^* - z'_i z_i) &= \det(Z^{*'} Z^* + kI_p^*) \left(1 - z_i (Z^{*'} Z^* + kI_p^*)^{-1} z'_i \right) \\
 &= \det(Z^{*'} Z^* + kI_p^*) (1 - h_{i,i}^R) \tag{4.8}
 \end{aligned}$$

bulunur. (4.5), (4.6) ve (4.8) eşitliği (4.7)'de yerine yazılırsa Teorem 4.6.'nın ispatı tamamlanır.

4.4.2. Cook-Weisberg

Teorem 4.7. AR(1) hatalı GEKK tahmin edicisi için Cook-Weisberg istatistiği

$$CW_i = -\frac{1}{2} \log(COVRATIO_i) + \frac{p}{2} \log \left(\frac{F_{\alpha, p, n-p}}{F_{\alpha, p, n-p-1}} \right)$$

olarak önerilmiştir. Burada $F_{\alpha, p, n-p}$ ve $F_{\alpha, p, n-p-1}$ sırasıyla $(p, n-p)$ ve $(p, n-p-1)$ serbestlik dereceleri ile merkezi F dağılımının α üst noktasıdır.

İspat 4.7. Klasik F testini göz önüne alalım. $H_0: A\beta = c$ 'yi test etmek istediğimizde bu test için doğal bir istatistik $A\hat{\beta} - c$ 'yi test etmektir. Eğer $A\hat{\beta}$, c 'den yeterince farklı ise H_0 red edilir. Ancak, $A\hat{\beta}$ 'daki her eleman farklı

4. GENEL LİNEER İSTATİSTİKSEL MODELDE TANILAMA METOTLARI

Tuğba SÖKÜT AÇAR

hassasiyetlere sahip olduğundan dolayı c 'nin her elemanı ile eşleştirilmemelidir.

$\hat{\beta}$ 'nin her bir elemanın hassasiyetini birleştirmenin bir yolu $(A\hat{\beta} - c)'(Var[A\hat{\beta}])^{-1}(A\hat{\beta} - c)$ karesel formunu kullanmaktır. Eğer σ^2 'nin yansız tahmin edicisi $s^2 = RSS/(n - p)$ kullanılırsa,

$(A\hat{\beta} - c)'[A(Z^{*'}Z^*)^{-1}A']^{-1}(A\hat{\beta} - c)/s^2$ 'yi elde ederiz (Seber ve Lee, 2003).

(3.4) eşitliğinde verilen GEKK tahmin edicisi ele alındığında normallik varsayımı altında

$$\hat{\beta} - \beta \sim N \left[0, \sigma^2(Z^{*'}Z^*)^{-1} \right]$$

elde edilir.

$$\hat{\beta} - \beta = A\hat{\beta} - \beta$$

eşitliği yazılsın. $A = I_p$ olmak üzere $(A\hat{\beta} - c)'(Var[A\hat{\beta}])^{-1}(A\hat{\beta} - c)$ karesel formu,

$$(\hat{\beta} - \beta)'[\sigma^2(Z^{*'}Z^*)^{-1}]^{-1}(\hat{\beta} - \beta)$$

olarak elde edilir. Ek-Lemma 3 gereğince

$$(\hat{\beta} - \beta)'[\sigma^2(Z^{*'}Z^*)^{-1}]^{-1}(\hat{\beta} - \beta)$$

yazılır. Bu karesel form, $A = I_p$ ve bu da idempotent olduğundan dolayı merkezi χ^2 dağılır. Ayrıca, Rencher ve Schaalje (2007, s. 118)'nin verdiği ikinci sonuç teoremi gereğince $[I_p - Z^*(Z^{*'}Z^*)^{-1}Z^{*'}]$ matrisi idempotent olduğundan dolayı

$$\frac{RSS}{\sigma^2} = \frac{(n-p)s^2}{\sigma^2} = \frac{y^*[I_p - Z^*(Z^{*'}Z^*)^{-1}Z^{*'}]y^{*'}}{\sigma^2} \sim \chi_{n-p}^2$$

olur. H_0 doğru ise,

$$F = \frac{\frac{(RSS_H - RSS)}{p}}{\frac{RSS}{n-p}} = \frac{\frac{(\hat{\beta} - \beta)'[\sigma^2(Z^{*'}Z^*)^{-1}]^{-1}(\hat{\beta} - \beta)}{p}}{\frac{\frac{(n-p)s^2}{\sigma^2}}{n-p}}$$

4. GENEL LİNEER İSTATİSTİKSEL MODELDE TANILAMA METOTLARI

Tuğba SÖKÜT AÇAR

eşitliği pay ve paydaları merkezi ki-kare dağıldığından dolayı $\hat{\beta}$ merkezli β 'nin güven elipsoidi

$$E = \frac{(\hat{\beta} - \beta)' Z^{*'} Z^* (\hat{\beta} - \beta)}{p s^2} \leq F_{\alpha, p, n-p}$$

elde edilir. Benzer işlemler yardımı ile i -nci gözlem veri kümesinden çıkartıldığında

$$E_{(i)} = \frac{(\hat{\beta}_{(i)} - \beta)' Z_{(i)}^{*'} Z_{(i)}^* (\hat{\beta}_{(i)} - \beta)}{p s_{(i)}^2} \leq F_{\alpha, p, n-p-1}$$

elde edilir.

Cook ve Weisberg (1980)'in tanımladığı CW istatistiği, tüm gözlemler ile tahmin edilen parametre vektörü için güven elipsoidinin, i -nci gözlemin çıkartılması ile elde edilen parametre vektörü için güven elipsoidine oranlanmasına dayanır. Yani,

$$CW_i = \log \frac{Volume(E)}{Volume(E_{(i)})}$$

dir. Bir elipsoidin hacmi, karesel formda ilgili matrisin determinantının karekökünün tersidir (Chatterjee ve Hadi, 1988). Bundan dolayı, CW_i ,

$$CW_i = \log \left\{ \left[\frac{\det((Z^{*'} Z^*)^{-1})}{\det((Z_{(i)}^{*'} Z_{(i)}^*)^{-1})} \right]^{\frac{1}{2}} \times \left(\frac{s^2}{s_{(i)}^2} \right)^{p/2} \left(\frac{F_{\alpha, p, n-p}}{F_{\alpha, p, n-p-1}} \right)^{p/2} \right\}$$

Eşitlik (4.5) ve (4.6)'nın kullanılması ile

$$CW_i = \log \left\{ \left[(1 - h_{i,i}) \left(\frac{n-p-1+r_i^{*2}}{n-p} \right)^p \right]^{\frac{1}{2}} \left(\frac{F_{\alpha, p, n-p}}{F_{\alpha, p, n-p-1}} \right)^{p/2} \right\} \quad (4.9)$$

olarak yazılır. (4.9) eşitliğinin ilk parçası $COVRATIO_i$ 'nin tersidir. Böylece Teorem 4.7 ispatlanmıştır.

4. GENEL LİNEER İSTATİSTİKSEL MODELDE TANILAMA METOTLARI

Tuğba SÖKÜT AÇAR

Teorem 4.8. AR(1) hatalı ARR tahmin edicisi altında Cook-Weisberg istatistiği

$$CW_i^R = -\frac{1}{2} \log(COVRATIO_i^R) + \frac{p}{2} \log \left(\frac{F_{\alpha,p,n-p,\xi}}{F_{\alpha,p,n-p-1,\xi}} \right)$$

olarak önerilmiştir. Burada $F_{\alpha,p,n-p,\xi}$ ve $F_{\alpha,p,n-p-1,\xi}$ sırasıyla $(p, n-p)$ ve $(p, n-p-1)$ serbestlik dereceleri ve $\xi = \frac{1}{2} \frac{k^2}{\sigma^2} \beta' (Z^{*'} Z^*)^{-1} \beta$ merkezi olmama parametresi ile merkezi olmayan F dağılımının α üst noktasıdır.

İspat 4.8. (3.5) eşitliğinde verilen ARR tahmin edicisi ele alındığında normallik varsayımı altında

$$\hat{\beta}^R - \beta \sim N \left[-k(Z^{*'} Z^* + kI_p^*)^{-1} \beta, \sigma^2 (Z^{*'} Z^* + kI_p^*)^{-1} Z^{*'} Z^* (Z^{*'} Z^* + kI_p^*)^{-1} \right]$$

elde edilir.

$$\hat{\beta}^R - \beta = (Z^{*'} Z^* + kI_p^*)^{-1} Z^{*'} Z^* \hat{\beta} - \beta = A \hat{\beta} - \beta$$

eşitliği yazılsın. $A = (Z^{*'} Z^* + kI_p^*)^{-1} Z^{*'} Z^*$ olmak üzere $(A \hat{\beta} - c)' (Var[A \hat{\beta}])^{-1} (A \hat{\beta} - c)$ karesel formu,

$$(\hat{\beta}^R - \beta)' \left[\sigma^2 (Z^{*'} Z^* + kI_p^*)^{-1} Z^{*'} Z^* (Z^{*'} Z^* + kI_p^*)^{-1} \right]^{-1} (\hat{\beta}^R - \beta)$$

olarak elde edilir. Ek-Lemma 4 gereğince

$$(\hat{\beta}^R - \beta)' \left[\sigma^2 (Z^{*'} Z^* + kI_p^*)^{-1} Z^{*'} Z^* (Z^{*'} Z^* + kI_p^*)^{-1} \right]^{-1} (\hat{\beta}^R - \beta)$$

karesel formu $A \Sigma = I_p$ olduğundan dolayı merkezi olmayan χ^2 dağılır. Öyleki merkezi olmama parametresi $\xi = \frac{1}{2} \frac{k^2}{\sigma^2} \beta' (Z^{*'} Z^*)^{-1} \beta$ 'dir. Ayrıca, Rencher ve Schaalje (2007, s. 118)'nin verdiği ikinci sonuç teoremi gereğince

$$\frac{RSS}{\sigma^2} = \frac{(n-p)s^2}{\sigma^2} = \frac{y^* [I_p - Z^* (Z^{*'} Z^*)^{-1} Z^{*'}] y^{*'}}{\sigma^2} \sim \chi_{n-p}^2$$

olur. H_0 doğru ise,

4. GENEL LİNEER İSTATİSTİKSEL MODELDE TANILAMA METOTLARI

Tuğba SÖKÜT AÇAR

$$F = \frac{\frac{(RSS_H - RSS)}{p}}{\frac{RSS}{n-p}} = \frac{\frac{(\hat{\beta}^R - \beta)' \left[\sigma^2 (Z^{*'} Z^* + kI_p^*)^{-1} Z^{*'} Z^* (Z^{*'} Z^* + kI_p^*)^{-1} \right]^{-1} (\hat{\beta}^R - \beta)}{p}}{\frac{\left[\frac{(n-p)s^2}{\sigma^2} \right]}{n-p}}$$

eşitliği pay kısmı merkezi olmayan ki-kare, payda kısmı ise merkezi ki-kare dağıldığından dolayı $\hat{\beta}^R$ merkezli β 'nin güven elipsoidi

$$E^R = \frac{(\hat{\beta}^R - \beta)' \left[(Z^{*'} Z^* + kI_p^*)^{-1} Z^{*'} Z^* (Z^{*'} Z^* + kI_p^*)^{-1} \right]^{-1} (\hat{\beta}^R - \beta)}{ps^2} \leq F_{\alpha, p, n-p, \xi}$$

elde edilir. Benzer işlemler yardımı ile i -inci gözlem veri kümesinden çıkartıldığında

$$E_{(i)}^R = \frac{(\hat{\beta}_{(i)}^R - \beta)' \left[(Z_{(i)}^{*'} Z_{(i)}^* + kI_p^*)^{-1} Z_{(i)}^{*'} Z_{(i)}^* (Z_{(i)}^{*'} Z_{(i)}^* + kI_p^*)^{-1} \right]^{-1} (\hat{\beta}_{(i)}^R - \beta)}{ps_{(i)}^2} \leq F_{\alpha, p, n-p-1, \xi}$$

elde edilir.

Cook ve Weisberg (1980)'in tanımladığı CW istatistiği, tüm gözlemler ile tahmin edilen parametre vektörü için güven elipsoidinin, i -nci gözlemin çıkartılması ile elde edilen parametre vektörü için güven elipsoid oranlarına dayanır. Yani,

$$CW_i^R = \log \frac{Volume(E^R)}{Volume(E_{(i)}^R)}$$

eşitliğindeki gibidir. Bir elipsoidin hacmi, karesel formda ilgili matrisin determinantının karekökünün tersidir (Chatterjee ve Hadi, 1988). Bundan dolayı, CW_i^R ,

$$CW_i^R = \log \left\{ \left[\frac{\det \left((Z^{*'} Z^* + kI_p)^{-1} Z^{*'} Z^* (Z^{*'} Z^* + kI_p)^{-1} \right)}{\det \left((Z_{(i)}^{*'} Z_{(i)}^* + kI_p)^{-1} Z_{(i)}^{*'} Z_{(i)}^* (Z_{(i)}^{*'} Z_{(i)}^* + kI_p)^{-1} \right)} \right]^{\frac{1}{2}} \times \left(\frac{S^2}{S_{(i)}^2} \right)^{p/2} \left(\frac{F_{\alpha, p, n-p, \xi}}{F_{\alpha, p, n-p-1, \xi}} \right)^{p/2} \right\}$$

Eşitlik (4.5), (4.6) ve (4.8)'in kullanılması ile

$$CW_i^R = \log \left\{ \left[\frac{(1-h_{i,i}^R)^2}{1-h_{i,i}} \left(\frac{n-p-1+r_i^{*2}}{n-p} \right)^p \right]^{\frac{1}{2}} \left(\frac{F_{\alpha, p, n-p, \xi}}{F_{\alpha, p, n-p-1, \xi}} \right)^{p/2} \right\} \quad (4.10)$$

olarak yazılır. (4.10) eşitliğinin ilk parçası $COVRATIO_i^R$ 'nin tersidir. Böylece Teorem 4.8 ispatlanmıştır.

4.4.3. Andrews Pregibon

Teorem 4.9. AR(1) hatalı GEKK tahmin edicisi için Andrews-Pregibon istatistiği

$$AP_i = h_{i,i} + (1 - h_{i,i}) \frac{r_i^{*2}}{n-p}$$

olarak önerilmiştir. Burada, r_i^{**} , i -nci dâhili student türü rezidüdür. (internally studendized residual).

İspat 4.9. (3.4) eşitliğinde verilen GEKK tahmin edicisi kullanıldığında güven elipsoidinin hacmi için $W = (Z^* : y^*)$ birleştirilmiş matrisi ele alalım. $W'W$ matrisi

$$W'W = \begin{pmatrix} Z^{*'} Z^* & Z^{*'} y^* \\ y^{*'} Z^* & y^{*'} y^* \end{pmatrix}$$

olarak yazılır.

Teorik incelemelerde $Z^{*'} Z^*$ matrisinin singüler olmayan bir matris olduğu varsayılın. Parçalanmış matrisin determinantını bulmak için Ek-Lemma 5 uygulandığında

$$\begin{aligned} \det(W'W) &= \det(Z^{*'} Z^*) \det(y^{*'} y^* - y^{*'} Z^* (Z^{*'} Z^*)^{-1} Z^{*'} y^*) \\ &= \det(Z^{*'} Z^*) y^{*'} (I_n - H)' y^* = \det(Z^{*'} Z^*) SSE(\hat{\beta}, \beta) \end{aligned}$$

4. GENEL LİNEER İSTATİSTİKSEL MODELDE TANILAMA METOTLARI

Tuğba SÖKÜT AÇAR

elde edilir. Burada $SSE(\hat{\beta}, \beta) = (y^* - Z^*\hat{\beta})'(y^* - Z^*\hat{\beta})$ olup GEKK için rezidü hata kareler toplamıdır. GEKK için σ^2 'nin yansız tahmin edicisi $s^2 = \frac{SSE(\hat{\beta}, \beta)}{n-p}$ olduğundan dolayı

$$\det(W'W) = (n-p)s^2 \det(Z^{*'}Z^*)$$

olarak yazılır. Benzer tanımdan yola çıkarak i -inci gözlem çıkartıldığında birleştirilmiş matris $W_{(i)} = (Z_{(i)}^* : y_{(i)}^*)$ olmak üzere $W_{(i)}'W_{(i)}$ parçalanmış matrisin determinantını Lemma 4.3'den yararlanarak

$$\det(W_{(i)}'W_{(i)}) = (n-p-1)s_{(i)}^2 \det(Z_{(i)}^{*'}Z_{(i)}^*)$$

yazılabilir. Böylece determinantların oranı

$$\frac{\det(W_{(i)}'W_{(i)})}{\det(W'W)} = \frac{(n-p-1)s_{(i)}^2 \det(Z_{(i)}^{*'}Z_{(i)}^*)}{(n-p)s^2 \det(Z^{*'}Z^*)} \quad (4.11)$$

olarak bulunur.

Veri merkezinin uzağında olan bir gözlemin çıkartılması determinantta büyük bir azalmaya neden olurken hacimde büyük bir artışa neden olur. Bu nedenle (4.11) eşitliğinin küçük değerler alması büyük bir dikkat gerektirir. İşiyi kolaylaştırmak için

$$AP_i = 1 - \frac{\det(W_{(i)}'W_{(i)})}{\det(W'W)}$$

olarak tanımlansın. $W_{(i)}'W_{(i)} = W'W - w_i'w_i$, için $w_i = (z_i^* : y_i^*)$, W matrisinin i -inci satır elemanı olmak üzere Ek-Lemma 2'nin kullanılması ile

$$\det(W_{(i)}'W_{(i)}) = \det(W'W - w_i'w_i) = \det(W'W)(1 - w_i(W'W)^{-1}w_i')$$

yazılır. Burada $w_i(W'W)^{-1}w_i'$, $W(W'W)^{-1}W'$ izdüşüm matrisinin i -inci köşegen elemanıdır. Bu eşitlik AP_i 'de yerine yazılırsa

$$AP_i = w_i(W'W)^{-1}w_i'$$

elde edilir. Parçalanmış matrisin tersi

$$(W'W)^{-1} = \begin{pmatrix} Z^{*'}Z^* & Z^{*'}y^* \\ y^{*'}Z^* & y^{*'}y^* \end{pmatrix}^{-1} = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$$

4. GENEL LİNEER İSTATİSTİKSEL MODELDE TANILAMA METOTLARI

Tuğba SÖKÜT AÇAR

olarak verelim. Graybill (1983, s. 183)'de verilen parçalanmış matrisin tersi yöntemini kullanarak

$$A_{11} = (Z^{*'}Z^*)^{-1} + \frac{(Z^{*'}Z^*)^{-1}Z^{*'}y^*y^{*'}Z^*(Z^{*'}Z^*)^{-1}}{SSE(\hat{\beta}, \beta)}$$

$$A_{12} = -\frac{(Z^{*'}Z^*)^{-1}Z^{*'}y^*}{SSE(\hat{\beta}, \beta)}$$

$$A_{21} = -\frac{y^{*'}Z^*(Z^{*'}Z^*)^{-1}}{SSE(\hat{\beta}, \beta)}$$

$$A_{22} = [y^{*'}y^* - y^{*'}Z^*(Z^{*'}Z^*)^{-1}Z^{*'}y^*]^{-1} = \frac{1}{SSE(\hat{\beta}, \beta)}$$

elde edilir. Böylece $W(W'W)^{-1}W'$ matrisi aşağıdaki gibi yazılır

$$W(W'W)^{-1}W' = Z^*(Z^{*'}Z^*)^{-1}Z^{*'} + \frac{\begin{bmatrix} Z^*(Z^{*'}Z^*)^{-1}Z^{*'}y^*y^{*'}Z^*(Z^{*'}Z^*)^{-1}Z^{*'} - y^{*'}y^*Z^*(Z^{*'}Z^*)^{-1}Z^{*'} \\ -Z^*(Z^{*'}Z^*)^{-1}Z^{*'}y^*y^{*'} + y^{*'}y^* \end{bmatrix}}{SSE(\hat{\beta}, \beta)}.$$

Gerekli düzenlemeler yapılarak,

$$W(W'W)^{-1}W' = H + \frac{(I_n - H)y^*y^{*'}(I_n - H)}{SSE(\hat{\beta}, \beta)} = H + \frac{e^*e^{*'}}{SSE(\hat{\beta}, \beta)}$$

elde edilir. Bu durumda $W(W'W)^{-1}W'$ matrisinin i -inci elemanı ve dolayısıyla i -inci gözleme ait AP istatistiği aşağıdaki gibi elde edilir

$$AP_i = w_i(W'W)^{-1}w_i' = h_{i,i} + \frac{e_i^{*2}}{SSE(\hat{\beta}, \beta)}.$$

Velleman ve Welsh (1981) bazı paket programların ve birkaç yazarın rezidü yerine dâhili student türü rezidüyü kullandıklarının altını çizmiştir. GEKK tahmin edicisi için dâhili student türü rezidü, $r_i^{**} = \frac{e_i^*}{s\sqrt{1-h_{i,i}}}$, σ 'ya bağlı olmadığından dolayı bu rezidüyü kullanmanın uygun olduğunu belirtmişlerdir. Atkinson (1981), σ^2 'nin tahmini s^2 ise $\frac{e_i^*}{s}$ oranının σ parametresinden bağımsız olduğunu

4. GENEL LİNEER İSTATİSTİKSEL MODELDE TANILAMA METOTLARI

Tuğba SÖKÜT AÇAR

belirtmiştir. Bu bilgiler altında GEKK tahmin edicisi için dâhili student türü rezidü kullanıldığında $r_i^{**2} s^2 (1 - h_{i,i}) = e_i^{*2}$ olarak yazılır. Yeni rezidüler AP_i^R 'de yerine yazılırsa Teoremi 4.9 ispatlanır.

Teorem 4.10. AR(1) hatalı ARR tahmin edicisi için Andrews-Pregibon istatistiği

$$AP_i^R = h_{i,i}^R + \frac{(e_i^{R*})^2}{\sum_{i=1}^n y_i^{*2} - \sum_{i=1}^n y_i^* \sum_{j=1}^n h_{i,i}^R y_j^*}$$

olarak önerilmiştir.

İspat 4.10. (3.5) eşitliğinde verilen ARR tahmin edicisi altında güven elipsoidinin hacmini bulmak için, $W = (Z^* : y^*)$ birleştirilmiş matrisi göz önüne alınsın. Çoklu iç ilişki problemi altında $Z^{*'}Z^*$ matrisi kötü koşullu olduğu için $(W'W + kI^A)$ matrisini incelenmesi gerekir. Öyle ki $I^A = \begin{pmatrix} I_p^* & 0_{p \times 1} \\ 0_{1 \times p} & 0_{1 \times 1} \end{pmatrix}$, $(p+1) \times (p+1)$ tipinde bir matris ve $W'W + kI^A = \begin{pmatrix} Z^{*'}Z^* + kI_p^* & Z^{*'}y^* \\ y^{*'}Z^* & y^{*'}y^* \end{pmatrix}$ dir.

ARR için AP istatistiği

$$AP_i^R = 1 - \frac{\det(W_{(i)'}W_{(i)} + kI^A)}{\det(W'W + kI^A)}$$

olarak yazılır, burada $W_{(i)'}W_{(i)} + kI^A = W'W - w_i'w_i + kI^A$ dir. Ek-Lemma 2'nin uygulanması ile elde edilen

$$\det(W_{(i)'}W_{(i)} + kI^A) = \det(W'W + kI^A) (1 - w_i(W'W + kI^A)^{-1}w_i')$$

ifadesi AP_i^R eşitliğinde yerine yazılırsa

$$AP_i^R = w_i(W'W + kI^A)^{-1}w_i'$$

bulunur. Yani i -inci gözleme ait AP değeri $W(W'W + kI^A)^{-1}W'$ yarı izdüşüm matrisinin (quasi-projection matrix) i -inci köşegen elemanıdır. AP_i^R hesaplamak için ilk olarak $(W'W + kI^A)$ parçalanmış matrisin tersi elde edilsin.

$$(W'W + kI^A)^{-1} = \begin{pmatrix} Z^{*'}Z^* + kI_p^* & Z^{*'}y^* \\ y^{*'}Z^* & y^{*'}y^* \end{pmatrix}^{-1} = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix},$$

$$B_{11} = (Z^{*'}Z^* + kI_p^*)^{-1} + (Z^{*'}Z^* + kI_p^*)^{-1}Z^{*'}y^*B_{22}y^{*'}Z^*(Z^{*'}Z^* + kI_p^*)^{-1}$$

4. GENEL LİNEER İSTATİSTİKSEL MODELDE TANILAMA METOTLARI
Tuğba SÖKÜT AÇAR

$$B_{12} = -(Z^{*'}Z^* + kI_p^*)^{-1}Z^{*'}y^*B_{22}$$

$$B_{21} = -y^{*'}Z^*(Z^{*'}Z^* + kI_p^*)^{-1}B_{22}$$

$$B_{22} = \left[y^{*'}y^* - y^{*'}Z^*(Z^{*'}Z^* + kI_p^*)^{-1}Z^{*'}y^* \right]^{-1} = (y^{*'}(I_n - L^*)y^*)^{-1}$$

olmak üzere $W(W'W + kI^A)^{-1}W'$ matrisi

$$\begin{aligned} W(W'W + kI^A)^{-1}W' &= Z^*(Z^{*'}Z^* + kI_p^*)^{-1}Z^{*'} \\ &+ \frac{Z^*(Z^{*'}Z^* + kI_p^*)^{-1}Z^{*'}y^*y^{*'}Z^*(Z^{*'}Z^*)^{-1}Z^{*'}}{y^{*'}(I_n - H^R)y^*} \\ &- \frac{y^*y^{*'}Z^*(Z^{*'}Z^* + kI_p^*)^{-1}Z^{*'}}{y^{*'}(I_n - H^R)y^*} \\ &- \frac{Z^*(Z^{*'}Z^*)^{-1}Z^{*'}y^*y^{*'}}{y^{*'}(I_n - H^R)y^*} + \frac{y^*y^{*'}}{y^{*'}(I_n - H^R)y^*} \end{aligned}$$

olarak elde edilir. Gerekli düzenlemeler yapıldıktan sonra

$$\begin{aligned} W(W'W + kI^A)^{-1}W' &= H^R + \frac{(I_n - H^R)y^*y^{*'}(I_n - H^R)}{y^{*'}(I_n - H^R)y^*} \\ &= H^R + \frac{(e^{R*})(e^{R*})'}{y^{*'}(I_n - H^R)y^*} \end{aligned}$$

bulunur. Böylece, $W(W'W + kI^A)^{-1}W'$ matrisinin i -inci köşegen elemanı olan AP istatistiğinin ARR tahmin edicisi altında ispatı tamamlanır.

5. GENEL LİNEER İSTATİSTİKSEL MODELDE MODEL GEÇERLİLİĞİ

Tuğba SÖKÜT AÇAR

5. GENEL LİNEER İSTATİSTİKSEL MODELDE MODEL GEÇERLİLİĞİ

Ridge regresyonda k 'nın seçimi için kullanılan birçok yöntem, regresyon katsayılarının kararlılığını iyileştirmeye yönelik önerilmiştir. Eğer istatistiksel model, yanıt değişken y 'nin öngörüsü için kullanılacaksa katsayıların kararlılığını sağlamak modelin öngörü (predictive) performansını iyileştirebilir ancak bu modelin kestirim performansına dolaylı bir saldırıdır (Montgomery ve Friedman, 1993).

Montgomery ve Friedman (1993) öngörü performansı yüksek bir model elde etmek için k 'nın seçimini PRESS istatistiğini minimum yaparak elde etmişlerdir. Golub ve arkadaşları (1979), iyi bir genelleştirilmiş ridge tahmini için GCV istatistiğini minimum yaparak k 'yı elde etmeyi önermişlerdir. Bu bölümde GEKK ile uydurulan modelin öngörü performans ölçümleri ile ARR ile uydurulan modelin öngörü performans ölçümleri elde edilmiştir. Ayrıca, incelenen ölçümün otokorelasyon katsayısına göre limit incelemeleri yapılmıştır.

5.1. Sıradan Çapraz Geçerlilik, OCV

OCV, literatürde sıklıkla kullanılan geçerlilik ölçümüdür. OCV istatistiği, tek gözlemin atılması ya da jackknife tekniğine dayanır. i -inci gözlem olmadan model uydurma ve yeni veri kümesi ile uydurulan modelde i -inci gözleme ait yanıt değişkeni tahmin etmek için kullanma mantığına dayalıdır. Bu işlem tüm gözlemler için tekrarlanır.

OCV istatistiği uydurulan bir istatistiksel modelin tahmin performansını ölçmek için kullanıldığında uygun ridge yanlılık parametresinin seçiminin OCV'yi minimum yapacak şekilde olması gerekmektedir. Ancak OCV, k 'nın konveks bir fonksiyonu olmadığından, OCV'nin yerel minimum noktasındaki k değerini seçmek gerekir. Montgomery ve Friedman (1993), k 'ya karşı OCV istatistiğinin grafiğinin çizilmesinin bu konuda yardımcı olacağını belirtmişlerdir.

5. GENEL LİNEER İSTATİSTİKSEL MODELDE MODEL GEÇERLİLİĞİ

Tuğba SÖKÜT AÇAR

Otokorelasyon ve çoklu iç ilişki problemleri altında GEKK ve ARR tahmin edicileri için OCV istatistiğini ele alalım.

Teorem 5.1. AR(1) hatalı GEKK tahmin edicisi kullanıldığında OCV istatistiği

$$OCV = \frac{1}{n} \sum_{i=1}^n \left(\frac{e_i^*}{1 - h_{i,i}} \right)^2$$

olarak önerilmiştir.

İspat 5.1. (3.4)'de verilen GEKK tahmin edicisi kullanıldığında sıradan çapraz geçerlilik ölçümünü elde etmeye GEKK için PRESS rezidüleri elde edilerek başlansın. GEKK için PRESS rezidü, i -inci gözlem silindiğinde elde edilen GEKK rezidüyü ifade eder yani

$$e_{(i)}^* = y_i^* - \hat{y}_{(i)}^* = y_i^* - z_i^* \hat{\beta}_{(i)}$$

(4.2) eşitliğinde verilen $\hat{\beta}_{(i)}$ kullanılarak

$$\begin{aligned} e_{(i)}^* &= y_i^* - z_i^* \left[\hat{\beta} - \frac{(Z^{*'}Z^*)^{-1}z_i^{*'}(y_i^* - \hat{y}_i^*)}{1 - h_{i,i}} \right] \\ &= y_i^* - \frac{z_i^* \hat{\beta} (1 - h_{i,i}) + h_{i,i} e_i^*}{1 - h_{i,i}} \\ &= \frac{y_i^* (1 - h_{i,i}) - \hat{y}_i^* (1 - h_{i,i}) + h_{i,i} e_i^*}{1 - h_{i,i}} \\ &= \frac{e_i^*}{1 - h_{i,i}} \end{aligned}$$

olarak elde edilir. Sıradan çapraz geçerlilik PRESS rezidülerin kareleri toplamından elde edildiğinden dolayı GEKK için OCV istatistiği

$$OCV = \frac{1}{n} \sum_{i=1}^n \left(\frac{e_i^*}{1 - h_{i,i}} \right)^2 \quad (5.1)$$

olarak elde edilmiş ve Teorem 5.1'in ispatı yapılmıştır.

(5.1) eşitliğinin özel durumlarını inceleyelim.

$\lim_{\rho \rightarrow 0} OCV = \lim_{\rho \rightarrow 0} \frac{1}{n} \sum_{i=1}^n \left(\frac{e_i^*}{1 - h_{i,i}} \right)^2$ limiti incelensin. $\lim_{\rho \rightarrow 0} P'P = I_n$ olmak üzere

5. GENEL LİNEER İSTATİSTİKSEL MODELDE MODEL GEÇERLİLİĞİ

Tuğba SÖKÜT AÇAR

$\lim_{\rho \rightarrow 0} Z^*(Z^{*'}Z^*)^{-1}Z^{*'} = \lim_{\rho \rightarrow 0} PZ(Z'P'PZ)^{-1}Z'P' = Z(Z'Z)^{-1}Z'$ olduğundan dolayı

$$\lim_{\rho \rightarrow 0} OCV = \lim_{\rho \rightarrow 0} \frac{1}{n} \sum_{i=1}^n \left(\frac{e_i^*}{1 - h_{i,i}} \right)^2 = \frac{1}{n} \sum_{i=1}^n \left(\frac{\hat{u}_i}{1 - l_{i,i}} \right)^2$$

olarak elde edilir. Yani $\rho \rightarrow 0$ iken ilişkisiz ve sabit varyanslı modelde EKK için tanımlanan OCV istatistiğine yakınsar.

Teorem 5.2. AR(1) hatalı ARR tahmin edicisi kullanıldığından OCV istatistiği

$$OCV^R = \frac{1}{n} \sum_{i=1}^n \left(\frac{e_i^{R*}}{1 - h_{i,i}^R} \right)^2$$

olarak önerilmiştir.

İspat 5.2. (3.5) eşitliğinde verilen ARR tahmin edicisi için PRESS rezidü, i -inci gözlem silindiğinde elde edilen ARR rezidü olmak üzere,

$$e_{(i)}^{R*} = y_i^* - \hat{y}_{(i)}^{R*} = y_i^* - z_i^* \hat{\beta}_{(i)}^R$$

dir. (4.3) eşitliği kullanılarak

$$\begin{aligned} e_{(i)}^{R*} &= y_i^* - z_i^* \left[\hat{\beta}^R - \frac{(Z^{*'}Z^* + kI_p)^{-1} z_i^{*'} e_i^{R*}}{1 - h_{i,i}^R} \right] \\ &= y_i^* - \frac{z_i^* \hat{\beta}^R (1 - h_{i,i}^R) + h_{i,i}^R e_i^{R*}}{1 - h_{i,i}^R} \\ &= \frac{y_i^* (1 - h_{i,i}^R) - \hat{y}_{(i)}^{R*} (1 - h_{i,i}^R) + h_{i,i}^R e_i^{R*}}{1 - h_{i,i}^R} \\ &= \frac{e_i^{R*}}{1 - h_{i,i}^R} \end{aligned}$$

elde edilir. OCV istatistiği PRESS rezidülerin kareleri toplamından elde edildiğinden dolayı ARR için OCV istatistiği

$$OCV^R = \frac{1}{n} \sum_{i=1}^n \left(\frac{e_i^{R*}}{1 - h_{i,i}^R} \right)^2 \quad (5.2)$$

olarak elde edilmiş ve Teorem 5.2'nin ispatı tamamlanmıştır.

5. GENEL LİNEER İSTATİSTİKSEL MODELDE MODEL GEÇERLİLİĞİ

Tuğba SÖKÜT AÇAR

(5.2) eşitliğinin özel durumlarını inceleyelim. İlk olarak

$$\lim_{\rho \rightarrow 0} OCV^R = \lim_{\rho \rightarrow 0} \frac{1}{n} \sum_{i=1}^n \left(\frac{e_i^{R*}}{1-h_{i,i}^R} \right)^2 \text{ limiti incelensin:}$$

$\lim_{\rho \rightarrow 0} P'P = I_n$ ve $k > 0$ olan bir sabit olduğundan dolayı $\lim_{\rho \rightarrow 0} (Z^{*'}Z^* + kI^*)^{-1} = \lim_{\rho \rightarrow 0} (Z'Z + kI^*)^{-1}$ bulunur. Böylece,

$$\lim_{\rho \rightarrow 0} \frac{1}{n} \sum_{i=1}^n \left(\frac{e_i^{R*}}{1-h_{i,i}^R} \right)^2 = \frac{1}{n} \sum_{i=1}^n \left(\frac{\hat{u}_i^r}{1-l_{i,i}^R} \right)^2$$

olarak elde edilir. Yani $\rho \rightarrow 0$ iken OCV istatistiği EKK tahmin edici kullanıldığında Montgomery ve Friedman (1993) tarafından verilen OCV istatistiğine yakınsar.

$$\lim_{k \rightarrow 0} OCV^R \text{ limiti ise } \lim_{k \rightarrow 0} \frac{1}{n} \sum_{i=1}^n \left(\frac{e_i^{R*}}{1-h_{i,i}^R} \right)^2 = \frac{1}{n} \sum_{i=1}^n \left(\frac{e_i^*}{1-h_{i,i}} \right)^2 \text{ olmak}$$

üzere GEKK tahmin edicisi için tanımlanan OCV istatistiğine dönüşür.

5.2. Genelleştirilmiş Çapraz Geçerlilik, GCV

Genelleştirilmiş çapraz geçerlilik, sıradan çapraz geçerliliğin dönüşümler altında sabit versiyonudur. Golub ve arkadaşları (1979), GCV istatistiğinin model seçiminde kullanılabileceği gibi modeller arasında karşılaştırma yapabilmek için ve ridge yanlılık parametresinin seçimi için de kullanılabileceğini belirtmişlerdir. Ayrıca Golub ve arkadaşları (1979), GCV kriterinin OCV kriterinden daha üstün olduğunu belirtmişlerdir. Liu ve Jiang (2012) optimum genelleştirilmiş ridge regresyon tahmin edicisini GCV ile belirlemişlerdir. Montgomery ve arkadaşları (2001), Myers (1990) ridge regresyon için GCV istatistiğini vermişlerdir. Bu çalışmada biz GCV istatistiğini, GEKK ve ARR tahmin edicileri için tanımlayıp uygulama aşamasında uygun yanlılık parametresinin seçiminde kullanacağız.

Teorem 5.3. AR(1) hatalı GEKK tahmin edicisi kullanıldığında GCV istatistiği

$$GCV = \frac{e^{*'}e^*}{(n-p)^2}$$

olarak elde edilmektedir.

İspat 5.3. (3.4) eşitliğinde verilen GEKK tahmin edicisi kullanıldığında genelleştirilmiş çapraz geçerlilik ölçümü

$$GCV = \frac{(y^* - Z^* \hat{\beta})'(y^* - Z^* \hat{\beta})}{\left[\text{tr} \{ I_n - Z^* (Z^{*'} Z^*)^{-1} Z^{*'} \}^2 \right]} \quad (5.3)$$

olarak önerilmiştir. (5.3) eşitliğinin pay kısmı GEKK için hata kareler toplamı olan $e^{*'} e^*$ ifade etmekte olup payda kısmında yer alan $\text{trace} \left(I_n - Z^* (Z^{*'} Z^*)^{-1} Z^{*'} \right)$

4. bölüm sayfa 30'da verilen (xvii) özelliğinden dolayı $(n - p)$ 'dir. Böylece, Teorem 5.3'ün ispatı tamamlanmıştır.

(5.3) eşitliğinin özel durumlarını inceleyelim.

$$\lim_{\rho \rightarrow 0} GCV = \lim_{\rho \rightarrow 0} \frac{e^{*'} e^*}{(n-p)^2} = \frac{\hat{u}' \hat{u}}{(n-p)^2} \text{ olup ilişkisiz ve sabit varyanslı modelde}$$

EKK tahmin edicisi için verilen GCV istatistiğine yakınsar.

Teorem 5.4. AR(1) hatalı ARR tahmin edicisi kullanıldığında GCV istatistiği

$$GCV^R = \frac{e^{R*'} e^{R*}}{(n - \text{tr}(H^R))^2}$$

olarak önerilmiştir.

İspat 5.4. (3.5) eşitliğinde verilen ARR tahmin edicisi kullanıldığında genelleştirilmiş çapraz geçerlilik ölçümü

$$GCV^R = \frac{(y^* - Z^* \hat{\beta}^R)'(y^* - Z^* \hat{\beta}^R)}{\left[\text{tr} \{ I_n - Z^* (Z^{*'} Z^* + k I_p^*)^{-1} Z^{*'} \} \right]^2} \quad (5.4)$$

olarak yazılır. (5.4) eşitliğinin pay kısmı ARR için hata kareler toplamı olan $e^{R*'} e^{R*}$ ifade etmekte olup payda kısmında yer alan $\text{tr} \{ I_n - Z^* (Z^{*'} Z^* + k I_p^*)^{-1} Z^{*'} \}$ gösterimi ise $(n - \text{tr}(H^R))$ 'dir. Böylece, Teorem 5.4'ün ispatı tamamlanmıştır.

(5.4) eşitliğinin özel durumlarını inceleyelim.

$$\lim_{\rho \rightarrow 0} GCV^R = \lim_{\rho \rightarrow 0} \frac{e^{R*'} e^{R*}}{(n - \text{tr}(H^R))^2} = \frac{(y - Z \hat{b}^R)'(y - Z \hat{b}^R)}{(n - \text{tr}(L^R))^2}$$

5. GENEL LİNEER İSTATİSTİKSEL MODELDE MODEL GEÇERLİLİĞİ

Tuğba SÖKÜT AÇAR

olup ilişkisiz ve sabit varyanslı model altında sıradan ridge regresyon tahmin edici için çapraz geçerlilik ölçümüne yakınsar.

$$\lim_{k \rightarrow 0} GCV^R = \lim_{\rho \rightarrow 0} \frac{e^{R*'} e^{R*}}{(n - tr(H^R))^2} = \frac{e^{*'} e^*}{(n - p)^2}$$

olup GEKK tahmin edici için tanımlanan çapraz geçerlilik ölçümüne yakınsar.

5.3. Cp

Mallows'un 1973 yılında model seçimi için önerdiği Cp istatistiği, birçok yazar tarafından sıradan ridge regresyon tahmin edici altında ele alınmış ve iyi bir ridge yanlılık parametre tahmin için Cp istatistiği incelenmiştir. Cp istatistiği herhangi bir tahmin yöntemi ile uydurulan tahmin değerinin standartlaştırılmış hata kareler ortalamasını ifade etmektedir (Hocking, 1976).

Teorem 5.5. Tüm açıklayıcı değişkenleri içeren model için AR(1) hatalı GEKK tahmin edicisi kullanıldığında Cp istatistiği

$$C_p = p$$

olarak önerilmiştir.

İspat 5.5. (3.4) eşitliğinde verilen GEKK tahmin edicisi için Cp ölçümü incelenir. Bu çalışmada değişken seçimi ile ilgilenilmediğinden dolayı tüm açıklayıcı değişkenleri içeren model için Cp ölçümü incelenmiştir.

i -nci gözleme ait hata kareler ortalaması $MSE(\hat{y}_i^*) = var(\hat{y}_i^*) + [bias(\hat{y}_i^*)^2]$ ile elde edilir. Standartlaştırılmış hata kareler ortalaması,

$$\sum_{i=1}^n \frac{MSE(\hat{y}_i^*)}{\sigma^2} = \sum_{i=1}^n \frac{var(\hat{y}_i^*)}{\sigma^2} + \sum_{i=1}^n \frac{[bias(\hat{y}_i^*)^2]}{\sigma^2} \quad (5.5)$$

olarak elde edilir. (5.5)'de verilen eşitliğin sağ tarafında yer alan ikinci parça

$$\begin{aligned} \sum_{i=1}^n [bias(\hat{y}_i^*)]^2 &= E(Z^* \beta - Z^* \hat{\beta})' E(Z^* \beta - Z^* \hat{\beta}) \\ &= E\{[Z^* \beta - HZ^* \beta]' [Z^* \beta - HZ^* \beta]\} \\ &= E\{(Z^* \beta)' (I - H)^2 (Z^* \beta)\}. \end{aligned}$$

Graybil (1976)'nın verdiği karesel formun beklenen değeri yardımı ile

5. GENEL LİNEER İSTATİSTİKSEL MODELDE MODEL GEÇERLİLİĞİ

Tuğba SÖKÜT AÇAR

$$\sum_{i=1}^n [bias(\hat{y}_i^*)]^2 = SSres(\hat{\beta}, \beta) - \sigma^2 tr(I - H) = 0 \quad (5.6)$$

olarak elde edilir.

(5.5)'de verilen eşitliğin sağ tarafında yer alan birinci parça, 4. bölüm sayfa 30'da verilen (xvi) ve (xvii) özelliklerinden dolayı

$$\begin{aligned} \sum_{i=1}^n \frac{var(\hat{y}_i^*)}{\sigma^2} &= \sum_{i=1}^n \frac{var(z_i^* \hat{\beta})}{\sigma^2} = \sum_{i=1}^n z_i^* (Z^{*'} Z^*)^{-1} z_i^{*'} \\ &= tr[H^2] = p \end{aligned} \quad (5.7)$$

olarak elde edilir.

(5.6) ve (5.7)'de verilen eşitlikler (5.5)'de yerine yazılırsa

$$C_p = tr(H) = p$$

bulunur ve böylece Teorem 5.5'in ispatı tamamlanır.

Teorem 5.6. AR(1) hatalı ARR tahmin edicisi kullanıldığında C_p istatistiği

$$C_p^R = \frac{SSres(\hat{\beta}^R, \beta)}{s^2} - n + 2tr(H^R) \quad (5.8)$$

olarak önerilmiştir.

İspat 5.6. (3.5) eşitliğinde verilen ARR tahmin edicisi için C_p ölçümü incelensin. i -nci gözleme ait hata kareler ortalaması

$$MSE(\hat{y}_i^{R*}) = var(\hat{y}_i^{R*}) + [bias(\hat{y}_i^{R*})^2]$$

ile elde edilir. Standartlaştırılmış hata kareler ortalaması,

$$\sum_{i=1}^n \frac{MSE(\hat{y}_i^{R*})}{\sigma^2} = \sum_{i=1}^n \frac{var(\hat{y}_i^{R*})}{\sigma^2} + \sum_{i=1}^n \frac{[bias(\hat{y}_i^{R*})^2]}{\sigma^2} \quad (5.9)$$

olarak elde edilir. (5.9)'de verilen eşitliğin ilk parçasını inceleyelim

$$\begin{aligned} \sum_{i=1}^n \frac{var(\hat{y}_i^{R*})}{\sigma^2} &= \sum_{i=1}^n \frac{var(z_i^* \hat{\beta}^R)}{\sigma^2} \\ &= \sum_{i=1}^n z_i^* (Z^{*'} Z^* + kI^*)^{-1} Z^{*'} Z^* (Z^{*'} Z^* + kI^*)^{-1} z_i^{*'} \\ &= tr(H^R)^2. \end{aligned} \quad (5.10)$$

5. GENEL LİNEER İSTATİSTİKSEL MODELDE MODEL GEÇERLİLİĞİ

Tuğba SÖKÜT AÇAR

(5.9)'da verilen eşitliğin ikinci parçasını inceleyelim. Ridge regresyon kullanıldığında katsayılar yanlıdır ve uydurulan değerdeki yanlılık rezidülerin hata kareler toplamını şişirir (Myers, 1990). Hata kareler toplamı,

$$SSres(\hat{\beta}^R, \beta) = (y^* - Z^* \hat{\beta}^R)'(y^* - Z^* \hat{\beta}^R) \quad (5.11)$$

olarak yazılır. Graybill (1976)'in verdiği karesel formun beklenen değerini kullanarak

$$E[(y^* - Z^* \hat{\beta}^R)'(y^* - Z^* \hat{\beta}^R)] = \sigma^2 tr(I - H^R)^2 + \beta' Z^{*'} tr(I - H^R)^2 Z^* \beta \quad (5.12)$$

olarak yazılabilir. (5.11) ve (5.12) eşitliklerinin kullanılması ile

$$\begin{aligned} \sum_{i=1}^n [bias(\hat{y}_i^{R*})]^2 &= E(Z^* \beta - Z^* \hat{\beta}^R)' E(Z^* \beta - Z^* \hat{\beta}^R) \\ &= [Z^* \beta - H^R Z^* \beta]' [Z^* \beta - H^R Z^* \beta] = (Z^* \beta)' (I - H^R)^2 (Z^* \beta). \\ &= SSres(\hat{\beta}^R, \beta) - \sigma^2 tr(I - H^R)^2. \end{aligned} \quad (5.13)$$

(5.10) ve (5.13)'de verilen eşitlikler (5.9)'da yerine yazılırsa

$$\begin{aligned} Cp^R &= tr(H^R)^2 + \frac{SSres(\hat{\beta}^R, \beta) - \sigma^2 tr(I - H^R)^2}{\sigma^2} \\ &= \frac{SSres(\hat{\beta}^R, \beta)}{\sigma^2} - n + 2tr(H^R) \end{aligned}$$

bulunup. σ^2 yerine yansız tahmin edicisi olan s^2 kullanıldığında Teorem 5.6'nın ispatı tamamlanır.

6. NÜMERİK ÖRNEK: FRANSA EKONOMİ VERİSİ

Bu bölümde, Malinvaud (1966) tarafından kullanılan ve Fransa ekonomisindeki ithalat faaliyetleri ile ilgili verilere dayanan Fransa ekonomi verisi analiz edilmiştir. Değişkenler y , ithalatları, x_1 , yerli üretimi, x_2 , stok oluşumunu ve x_3 , yerli tüketimi olmak üzere, tüm ölçümler 1949-1966 yılları arasında Fransız frangı cinsinden elde edilmiştir. Chatterjee ve arkadaşları (2000), aynı veri kümesinin 1949-1959 yıllarını kapsayan verileri ile çoklu iç ilişkinin öngörü üzerindeki etkisini belirlemek için kullanmış ancak verideki olası otokorelasyon problemi üzerinde durmamışlardır. Bu çalışmada Fransa ekonomi verisinin 1953-1966 yılları arasındaki değerleri kullanılmıştır. 3 açıklayıcı değişken kullanarak

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon \quad (6.1)$$

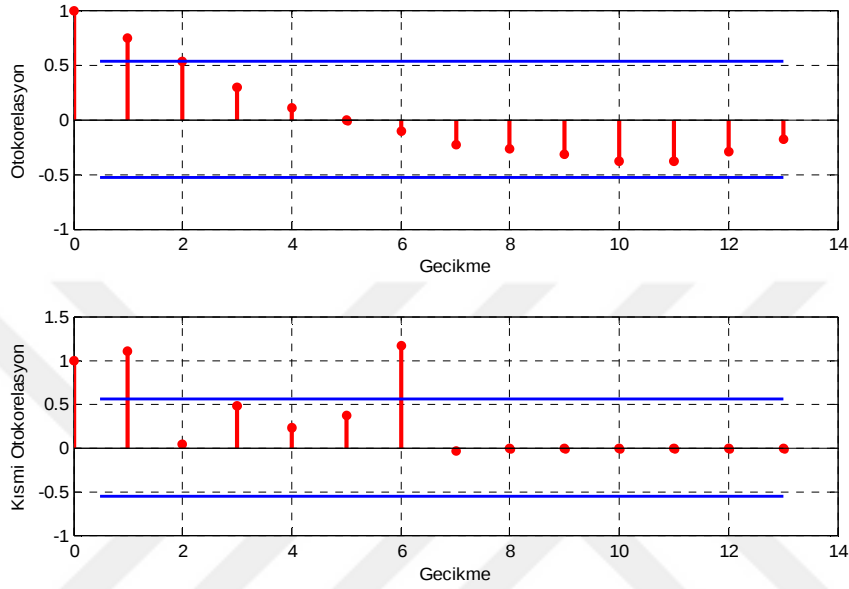
modeli oluşturulmuştur. 14 gözlem içeren veri kümesi Çizelge 6.1'de verilmiştir. Veri kümesinde X matrisi ve y vektörü merkezileştirilmiş ve standartlaştırılmış formda analize dâhil edilmiştir.

Çizelge 6.1. Fransa ekonomi verisi

Gözlem	y	x_1	x_2	x_3
1	18.8	180.8	1.1	132.1
2	20.4	190.7	2.2	137.7
3	22.7	202.1	2.1	146.0
4	26.5	212.4	5.6	154.1
5	28.1	226.1	5.0	162.3
6	27.6	231.9	5.1	164.3
7	26.3	239.0	0.7	167.6
8	31.1	258.0	5.6	176.8
9	33.3	269.8	3.9	186.6
10	37.0	288.4	3.1	199.7
11	43.3	304.5	4.6	213.9
12	49.0	323.4	7.0	223.8
13	50.3	336.8	1.2	232.0
14	56.6	353.9	4.5	242.9

6. NÜMERİK ÖRNEK: FRANSA EKONOMİ VERİSİ Tuğba SÖKÜT AÇAR

Veri kümesinde otokorelasyonun varlığını ve yapısını belirleyebilmek için otokorelasyon fonksiyonu (ACF) ve kısmi otokorelasyon fonksiyonuna (PACF) ait katsayıların gecikmelere göre grafiksel gösterimi Şekil 6.1’de verilmiştir.



Şekil 6.1. Fransa ekonomi verisi için ACF ve PACF grafikleri

Şekil 6.1’de verilen ACF grafiği incelendiğinde gecikmelere bağlı olarak otokorelasyonların sifıra yaklaştığı, PACF grafiğinden ise birinci gecikmenin önemli olduğu görülmüştür. Ancak, otoregresif sürecin derecesini belirlemede kesin bir yöntem olmamıştır. Bundan dolayı derece belirlemede diğer yöntemler göz önüne alınmıştır. Otokorelasyonun AR(1) sürecinden olup olmadığı Durbin-Watson test istatistiği kullanılarak test edilmiştir. DW istatistiğinin aşamaları aşağıda verilmiştir.

Hipotez:

$$H_0: \rho = 0 \text{ (Otokorelasyon yoktur.)}$$

$$H_1: \rho \neq 0 \text{ (Otokorelasyon vardır.)}$$

Test İstatistiği:

$$dw = \frac{\sum_{i=2}^n (\hat{u}_i - \hat{u}_{i-1})^2}{\sum_{i=1}^n \hat{u}_i^2} = \frac{\hat{u}' A \hat{u}}{\hat{u}' \hat{u}} = 0.8569$$

Burada $\hat{u}_i$, EKK ile elde edilen i -inci rezidü ve

$$A = \begin{bmatrix} 1 & -1 & 0 & \cdots & 0 & 0 \\ -1 & 2 & -1 & \cdots & 0 & 0 \\ 0 & -1 & 2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 2 & -1 \\ 0 & 0 & 0 & \cdots & -1 & 1 \end{bmatrix}_{14 \times 14} \text{ 'dır.}$$

Karar: 0.05 önem düzeyinde dw testi için alt ve üst sınır değerleri $n = 14$, $p = 4$ için $d_L^* = 0.767$ ve $d_U^* = 1.779$ olarak elde edilmiştir. $d_L^* < dw < d_U^*$ olduğundan test kararsızlık bölgesine düşmüş ve bundan dolayı Durbin ve Watson (1971)'in önerdikleri yaklaşım kullanılmıştır.

$$E(d) = \frac{\Gamma}{n-p} = 2.246, \Gamma = 2(n-1) - \text{tr}[Z'AZ(Z'Z)^{-1}],$$

$$\text{Var}(d) = \frac{2}{(n-p)(n-p+2)} (\phi - \Gamma E(d)) = 0.217,$$

$$\phi = 2(3n-4) - 2\text{tr}[Z'A^2Z(Z'Z)^{-1}] + \text{tr}[\{Z'AZ(Z'Z)^{-1}\}^2]$$

olarak hesaplanmış ve Judge ve arkadaşlarının (1988, s. 995) çalışmasında verilen dw üst sınırının ortalama ve varyansı $E(d_U) = 2.532$, $\text{Var}(d_U) = 0.195$ olarak elde edilmiştir. $E(dw) = a + bE(d_U)$ ve $\text{Var}(dw) = b^2\text{Var}(d_U)$ olmak üzere a ve b değerleri sırası ile -0.142 ve 1.056 olarak hesaplanmıştır. Son olarak $dw^* = a + bd_U^* = 1.736$ olmak üzere $dw < dw^*$ olduğundan H_0 hipotezi red edilmiş ve veri kümesinde otokorelasyon olduğu saptanmıştır. Şekil 6.1'den kısmi otokorelasyon incelendiğinde 1-inci ve 6-ıncı gözleme ait kısmi otokorelasyonların yaklaşık 2 standart sapma sınırının (0.6030 ve -0.6030) dışında olduğu görülmüştür. Veri kümesinin AR(2) olabileceğinden şüphelendiğimizden dolayı, yüksek dereceden otoregresif süreci test etmek için Godfrey (1978)'in önerdiği test prosedürü uygulanmıştır.

AR(2) sürecinde hatalar arasındaki ilişki

$$\varepsilon_i = \rho_1 \varepsilon_{i-1} + \rho_2 \varepsilon_{i-2} + u_i, \quad u_i \sim N(0, \sigma_u^2)$$

6. NÜMERİK ÖRNEK: FRANSA EKONOMİ VERİSİ Tuğba SÖKÜT AÇAR

ile modellenir. Fransa ekonomi verisinde

$H_0: \varepsilon_t$, AR(1) sürecinden gelmektedir.

$H_1: \varepsilon_t$, AR(2) sürecinden gelmektedir.

hipotezinin testi için Godfrey (1978)'in yöntemi uygulanmış ve test istatistiği $\phi = 0.5369$ olarak bulunmuştur. 0.05 önem düzeyinde $\chi_1^2 = 3.84$ ile karşılaştırıldığında $\phi < \chi_1^2$ olduğundan dolayı H_0 hipotezi red edilememiştir. Yani hataların AR(1) sürecinden geldiği sonucuna varılmıştır.

Hataların AR(1) sürecinden ilişkili olduğu belirlendikten sonra otokorelasyon katsayısı

$\hat{\rho}_1 = 0.5475$ (Griffiths ve ark., 1993, s. 525),

$\hat{\rho}_2 = 0.4212$ (Theil, 1971),

$\hat{\rho}_3 = 0.5716$ (Durbin ve Watson, 1950),

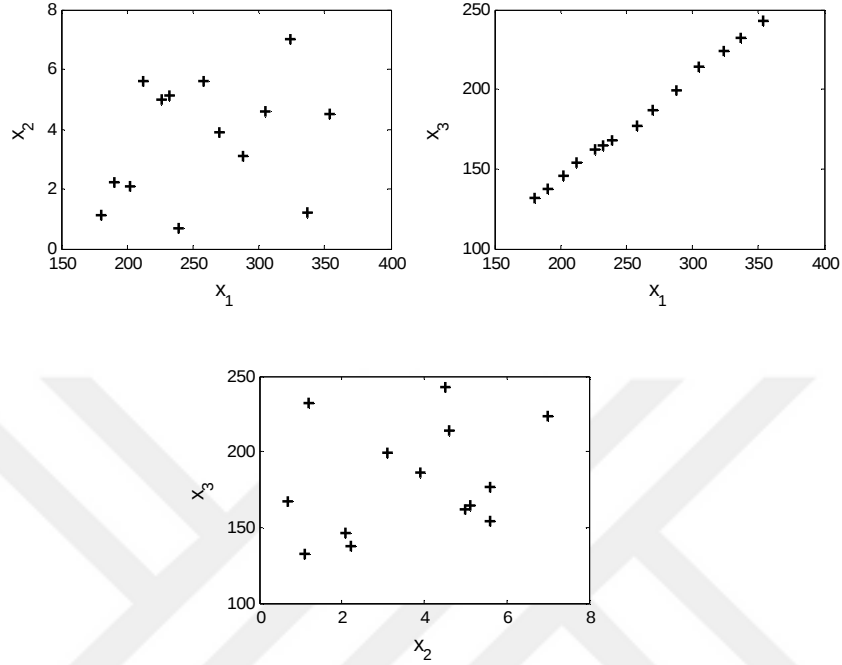
$\hat{\rho}_4 = 0.7113$ (Theil ve Nagar, 1961)

olarak tahmin edilmiştir. $\hat{\rho}_1 = 0.5475$ tahmini kullanılarak AR(1) süreci için P matrisi türetilmiş ve GEKK tahmin edicisi,

$$\hat{\beta} = \begin{pmatrix} 0.0082 \\ -0.5198 \\ 0.0690 \\ 1.5056 \end{pmatrix}$$

olarak elde edilmiştir.

Çoklu iç ilişkiyi belirlemek için koşul sayısı kullanılmıştır. $Z^{*'}Z^*$ matrisinin özdeğerleri $\lambda_1 = 0.0010$, $\lambda_2 = 0.6108$, $\lambda_3 = 1.5201$ ve $\lambda_4 = 3.3644$ olmak üzere $KS = \frac{\lambda_{max}}{\lambda_{min}} = 3364.4$ olarak belirlenmiş olup veri kümesinde güçlü çoklu iç ilişki probleminin olduğu saptanmıştır. Şekil 6.2'de orijinal açıklayıcı değişkenler arasındaki ilişkiler grafiklendirilmiştir.



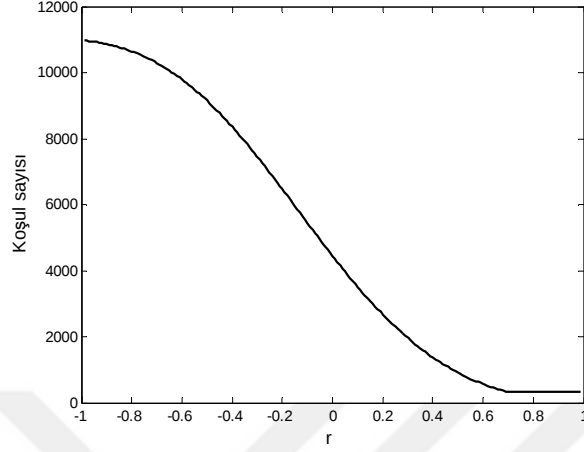
Şekil 6.2. Açıklayıcı değişkenler arasındaki ilişkiler

X matrisinin merkezileştirilme ve standartlaştırma işleminden sonra $X'X$, korelasyon matrisi formunda elde edilmiştir. Buna göre, yerli üretim (x_1) ile stok oluşumu (x_2) arasındaki korelasyon 0.2828, yerli üretim (x_1) ile yerli tüketim (x_3) arasındaki 0.9987, stok oluşumu (x_2) ile yerli tüketim (x_3) arasındaki ise 0.2875 olarak hesaplanmıştır. Böylece, yerli üretim ile yerli tüketim arasındaki ilişkiler veri kümesinde çoklu iç ilişki problemini ortaya çıkartmıştır. Firinguetti (1989)'nin önerdiği $k_1 = \frac{ps^2}{\hat{\beta}'\hat{\beta}} = 0.0012$ ile ARR tahmin edicisi

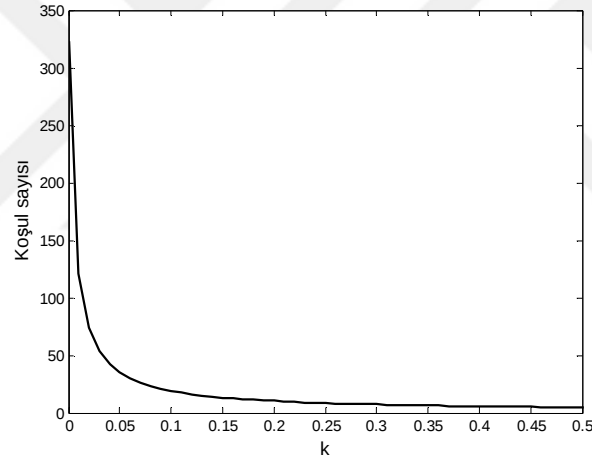
$$\hat{\beta}^R = \begin{pmatrix} 0.0084 \\ 0.0436 \\ 0.0677 \\ 0.9416 \end{pmatrix}$$

elde edilmiştir.

Koşul sayısının otokorelasyon ve ridge yanlılık parametresine karşı grafikleri Şekil 6.3 ve Şekil 6.4'de verilmiştir.



Şekil 6.3. Otokorelasyona karşı koşul sayısı

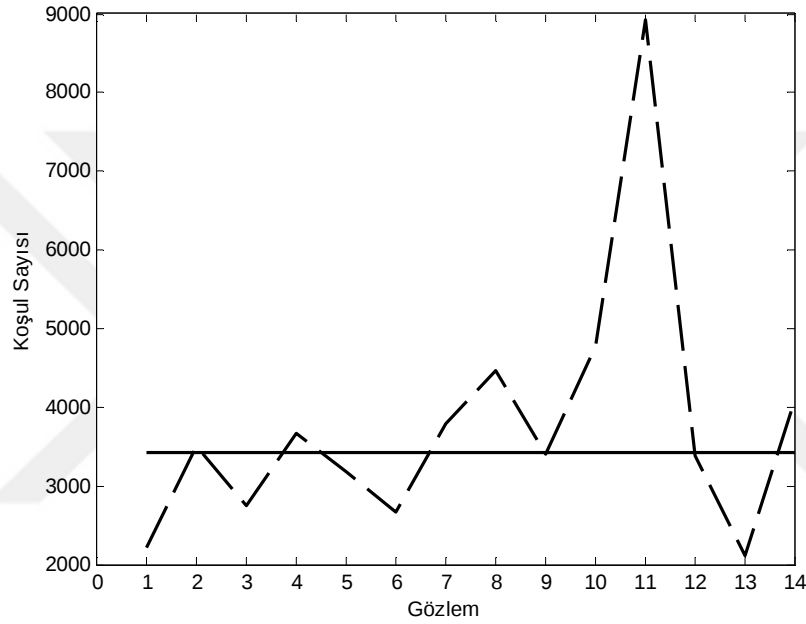


Şekil 6.4. Ridge ynlılık parametresine karşı koşul sayısı

Şekil 6.3'den görüldüğü üzere, otokorelasyon katsayısı pozitif yönde arttıkça yani $\rho \rightarrow 1$ iken koşul sayısı azalırken, otokorelasyonun negatif yönde artması ile yani $\rho \rightarrow -1$ iken koşul sayısının arttığı gözlenmiştir. Ancak, bu artış ve azalışlar çoklu iç ilişki problemini ortadan kaldıracak kadar etkili değildir. Şekil 6.4, k arttıkça koşul sayısının azaldığını göstermektedir. Bu azalışlar çoklu iç ilişki problemini ortadan kaldıracak kadar önemlidir.

6. NÜMERİK ÖRNEK: FRANSA EKONOMİ VERİSİ Tuğba SÖKÜT AÇAR

Koşul sayısına ilişkin önemli bir grafik Şekil 6.5’de verilmiştir. Şekil 6.5’de i -inci gözlemin silinmesi ile koşul sayısındaki değişim gösterilmiştir. 1-inci gözlemi silmek koşul sayısında önemli bir düşüşe neden olurken, 8-inci ve 11-inci gözlemi silmenin koşul sayısında önemli bir artışa neden olduğu görülmektedir. Bu azalış ve artışlar çoklu iç ilişki problemini ortadan kaldırmamaktadır.



Şekil 6.5. i -inci gözlemin silinmesi ile elde edilen koşul sayıları (-----, i -inci gözlem olmadan, _____, tüm gözlemler modelde iken)

Hem otokorelasyon hem de çoklu iç ilişki problemi olduğu belirlenen Fransa ekonomisi verisinde etkin gözlemler teoride ele alınan yöntemler ile incelenmiştir. İlk olarak kaldırılacak gözlemler GEKK ve ARR tahmin edicileri yardımı ile tespit edilmiştir.

6. NÜMERİK ÖRNEK: FRANSA EKONOMİ VERİSİ Tuğba SÖKÜT AÇAR

Çizelge 6.2. Fransa ekonomi verisi için $h_{i,i}$ ve $h_{i,i}^R$ değerleri ($\hat{\rho}_1 = 0.5475$, $k_1 = 0.0012$)

Gözlem	GEKK	ARR	Gözlem	GEKK	ARR
1	0.5490	0.5458	8	0.6035	0.3994
2	0.1220	0.1176	9	0.0840	0.0835
3	0.1124	0.1002	10	0.1161	0.1158
4	0.3253	0.2603	11	0.4982	0.3011
5	0.0730	0.0700	12	0.2662	0.2531
6	0.1082	0.0927	13	0.5377	0.5343
7	0.3145	0.2769	14	0.2899	0.2896

Çizelge 6.2’den görüldüğü üzere hem GEKK hem de ARR tahmin edicisi kullanıldığında 1-inci 8-inci, 11-inci ve 13-üncü gözlemler diğer gözlemlere oranla daha büyük kaldıraç değerlere sahiptir. GEKK ve ARR için kaldıraç değerler karşılaştırıldığında tüm gözlemler için ARR ile elde edilen kaldıraç değerlerin GEKK ile elde edilen değerlerden daha küçük olduğu görülmektedir. Bununla birlikte ARR tahmin edicisi kullanıldığında 8-inci gözlemin kaldıraç değeri GEKK ile elde edilen değere göre önemli ölçüde küçülmüştür.

Çizelge 6.3. Fransa ekonomi verisi için $DFBETA_i$ ve $DFBETA_i^R$ değerleri ($\hat{\rho}_1 = 0.5475$, $k_1 = 0.0012$)

Göz.	GEKK				ARR			
	β_0	β_1	β_2	β_3	β_0	β_1	β_2	β_3
1	0.0207	-0.1670	-0.0081	0.0896	0.0213	-0.0976	-0.0085	0.0187
2	0.0011	0.0143	0.0003	-0.0175	0.0007	0.0035	0.0002	-0.0056
3	0.0002	-0.0063	-0.0001	0.0058	0.0008	-0.0092	-0.0002	0.0076
4	-0.0009	0.0532	-0.0022	-0.0507	0.0007	-0.0183	0.0016	0.0165
5	-0.0021	0.0263	-0.0007	-0.0242	-0.0018	0.0107	-0.0006	-0.0089
6	-0.0040	-0.1069	-0.0024	0.1123	-0.0046	-0.0524	-0.0029	0.0588
7	-0.0042	-0.1903	0.0122	0.1878	-0.0051	-0.1040	0.0146	0.1012
8	-0.0043	-0.4017	-0.0093	0.4024	-0.0064	-0.2690	-0.0149	0.2709
9	-0.0044	0.0201	0.0032	-0.0271	-0.0043	0.0067	0.0031	-0.0136
10	-0.0048	0.0080	0.0037	-0.0225	-0.0047	-0.0005	0.0036	-0.0137
11	-0.0047	0.4790	-0.0041	-0.4938	-0.0006	0.0250	-0.0004	-0.0269
12	0.0038	0.1013	0.0060	-0.0880	0.0031	0.0394	0.0050	-0.0287
13	0.0058	0.0963	-0.0241	-0.0548	0.0055	0.0507	-0.0225	-0.0120
14	0.0075	0.0120	0.0081	0.0288	0.0077	0.0169	0.0082	0.0244

Regresyon katsayıları üzerinde etkin gözlemi belirlemek adına GEKK ve ARR tahmin edicileri için DFBETA değerleri Çizelge 6.3’te verilmiş ve DFBETA

6. NÜMERİK ÖRNEK: FRANSA EKONOMİ VERİSİ Tuğba SÖKÜT AÇAR

değerlerine göre diğer gözlemlere oranla aykırı duran gözlemler incelenmiştir. Böylece, β_0 ve β_2 parametreleri üzerinde etkin gözlem gözlenmez iken β_1 ve β_3 parametreleri üzerinde etkin gözlemler belirlenmiştir. 8-inci gözlem hem GEKK hem de ARR tahmin edicisi kullanıldığında β_1 ve β_3 parametreleri üzerinde etkin iken, GEKK kullanıldığında β_1 ve β_3 parametreleri üzerinde etkin olan 11-inci gözlemin etkinliği, ARR tahmin edicisi kullanıldığında ortadan kalkmıştır. GEKK ve ARR tahmin edicileri DFBETA istatistiği üzerinde farklı tutumlar sergilemişlerdir.

Çizelge 6.4. Fransa ekonomi verisi için $DFFITS_i$ ve $DFFITS_i^R$ değerleri ($\hat{\rho}_1 = 0.5475, k_1 = 0.0012$)

Gözlem	GEKK	ARR	Gözlem	GEKK	ARR
1	2.6314	2.6903	8	-0.7923	-1.1107
2	0.0952	0.0613	9	-0.3549	-0.3438
3	0.0208	0.0630	10	-0.4618	-0.4516
4	-0.1308	0.0936	11	-0.9055	-0.1007
5	-0.1433	-0.1241	12	0.5000	0.4011
6	-0.3417	-0.3762	13	1.3008	1.2181
7	-0.6441	-0.7636	14	1.2372	1.2552

Yanıt değişken üzerinde etkin olan gözlemleri belirlemek için $DFFITS_i$ ve $DFFITS_i^R$ ($i = 1, \dots, 14$) değerleri incelenmiş ve sonuçlar Çizelge 6.4'te verilmiştir. 1-inci, 11-inci, 13-üncü ve 14-üncü gözlemlerin GEKK tahmin edicisi ile uydurulan yanıt değişken üzerinde etkin olduğu saptanırken, 1-inci, 8-inci, 13-üncü ve 14-üncü gözlemlerin ARR tahmin edicisi ile uydurulan yanıt değişken üzerinde etkin olduğu saptanmıştır. GEKK ve ARR tahmin edicileri DFFITS istatistiği açısından karşılaştırıldığında, 8-inci ve 11-inci gözlemler hariç diğer tüm gözlemler için DFFITS değerlerinin yakın olduğu görülmektedir.

6. NÜMERİK ÖRNEK: FRANSA EKONOMİ VERİSİ Tuğba SÖKÜT AÇAR

Çizelge 6.5. Fransa ekonomi verisi için D_i ve D_i^R değerleri ($\hat{\rho}_1 = 0.5475$, $k_1 = 0.0012$)

Gözlem	GEKK	ARR	Gözlem	GEKK	ARR
1	1.1786	1.2242	8	0.1667	0.1960
2	0.0025	0.0010	9	0.0304	0.0283
3	0.0001	0.0010	10	0.0502	0.0478
4	0.0047	0.0019	11	0.2086	0.0013
5	0.0055	0.0040	12	0.0645	0.0396
6	0.0293	0.0304	13	0.4046	0.3522
7	0.1047	0.1299	14	0.3002	0.3082

Çizelge 6.5’den görüldüğü üzere GEKK ve ARR tahmin edicileri kullanıldığında 1-inci gözlem başta olmak üzere 8-inci, 13-üncü ve 14-üncü gözlemlere ait Cook uzaklığı değerleri diğer gözlemlere oranla oldukça yüksektir. Ayrıca GEKK tahmin edicisi kullanıldığında 11-inci gözleme ait Cook uzaklık değeri ARR tahmin edicisi ile elde edilen Cook uzaklık değerine göre oldukça büyüktür.

Çizelge 6.6. Fransa ekonomi verisi için $COVRATIO_i$ ve $COVRATIO_i^R$ değerleri ($\hat{\rho}_1 = 0.5475$, $k_1 = 0.0012$)

Gözlem	GEKK	ARR	Gözlem	GEKK	ARR
1	0.4766	0.4698	8	3.2133	1.4004
2	1.6866	1.6698	9	0.9430	0.9419
3	1.7145	1.6685	10	0.8880	0.8874
4	2.2237	1.8503	11	2.1383	1.1022
5	1.4666	1.4573	12	1.5461	1.4924
6	1.1382	1.0997	13	1.8103	1.7840
7	1.5161	1.3625	14	0.5331	0.5326

Çizelge 6.6’da GEKK ve ARR tahmin edicileri ile elde edilen parametre vektörüne ait varyans-kovaryans matrisi üzerindeki etkili gözlemi belirlemek için COVRATIO değerleri verilmiştir. 1-inci ve 14-üncü gözlemlere ait COVRATIO değerleri hem GEKK hemde ARR tahmin edicileri kullanıldığında diğer gözlemlere oranla aykırıdır. GEKK tahmin edicisi kullanıldığında varyans-kovaryans matrisi üzerinde aykırı olan 8-inci gözlem ARR tahmin edicisi ile aykırı

6. NÜMERİK ÖRNEK: FRANSA EKONOMİ VERİSİ Tuğba SÖKÜT AÇAR

değildir. Yani ARR tahmin edicisi kullanıldığında 8-inci gözlemin varyans-kovaryans matrisi üzerindeki etkisi kaybolmuştur.

GEKK ve ARR tahmin edicileri için COVRATIO değerleri birbirine yakın değerler olsa da ARR tahmin edicisi ile her zaman daha küçük COVRATIO değerleri elde edilmiştir.

Çizelge 6.7. Fransa ekonomi verisi için CW_i ve CW_i^R değerleri ($\hat{\rho}_1 = 0.5475, k_1 = 0.0012$)

Gözlem	GEKK	ARR	Gözlem	GEKK	ARR
1	0.2833	0.2879	8	-0.6709	-0.2582
2	-0.3486	-0.3461	9	-0.0579	-0.0598
3	-0.3568	-0.3457	10	-0.0278	-0.0301
4	-0.4868	-0.3974	11	-0.4672	-0.1384
5	-0.2787	-0.2781	12	-0.3051	-0.2900
6	-0.1520	-0.1373	13	-0.3840	-0.3792
7	-0.2953	-0.2444	14	0.2273	0.2253

Çizelge 6.7’de Fransa ekonomi verisi için Cook-Weisberg değerleri verilmiştir. ARR tahmin edicisi kullanıldığında 8-inci ve 11-inci gözlemlerin CW değerleri GEKK ile elde edilen değerlere göre önemli ölçüde küçülmüştür. GEKK tahmin edici kullanıldığında 8-inci gözleme ait CW değeri diğer gözlemlere oranla daha büyük iken ARR tahmin edicisi kullanıldığında diğer gözlemlerden aykırı duran bir gözlem tespit edilmemiştir.

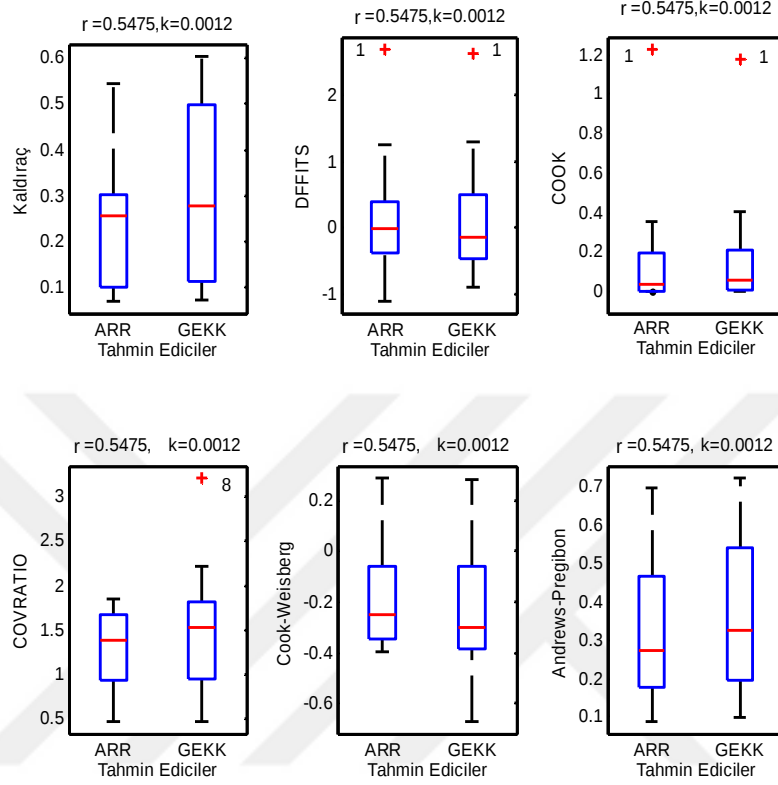
Çizelge 6.8. Fransa ekonomi verisi için AP_i ve AP_i^R değerleri ($\hat{\rho}_1 = 0.5475, k_1 = 0.0012$)

Gözlem	GEKK	ARR	Gözlem	GEKK	ARR
1	0.7237	0.6980	8	0.6209	0.4744
2	0.1283	0.1198	9	0.2053	0.1771
3	0.1127	0.1030	10	0.2512	0.2218
4	0.3279	0.2618	11	0.5404	0.3021
5	0.0991	0.0865	12	0.3184	0.2824
6	0.1943	0.1884	13	0.6020	0.5813
7	0.3771	0.3625	14	0.4987	0.4659

6. NÜMERİK ÖRNEK: FRANSA EKONOMİ VERİSİ Tuğba SÖKÜT AÇAR

Çizelge 6.8’de gözlemlerin güven elipsoidinin hacmi üzerindeki etkilerini gösteren Andrews-Pregibon değerleri verilmiştir. Hem GEKK hem de ARR tahmin edicileri kullanıldığında 1-inci ve 13-üncü gözlemlere ait AP değerleri diğer gözlemlere oranla daha yüksektir. GEKK ile elde edilen 8-inci gözleme ait AP değeri ARR tahmin edici kullanıldığında önemli ölçüde azalmıştır. ARR tahmin edicisi ile elde edilen AP değerleri her zaman GEKK tahmin edicisi ile elde edilen ölçüm değerlerinden daha küçüktür.

Çizelge 6.2-Çizelge 6.8’de tüm gözlemler için ilgili ölçüm değerleri tek tek verilmiştir. İlgili ölçüme göre aykırı duran gözlemi kolayca belirlemenin bir yolu grafiksel gösterimden faydalanmaktır. Kutu grafikleri, veri kümesinin %75’inden büyük ya da %25’inden küçük değer alan ve aykırı olan gözlemleri kolayca belirlemek için literatürde sıklıkla başvurulmuş bir yöntemdir. Fransa ekonomi verisinde GEKK ve ARR tahmin edicileri kullanıldığında elde edilen kaldıraç, DFFITS, COOK, COVRATIO, CW ve AP değerlerine ait kutu grafikleri Şekil 6.6 ile verilmiştir.



Şekil 6.6. Kaldıraç, DFFITS, COOK, COVRATIO, CW ve AP değerlerinin kutu grafikleri

Şekil 6.6'dan görüldüğü üzere, yanıt değişken üzerinde etkin gözlemleri belirlemede DFFITS ve COOK değerlerine ait kutu grafikleri incelenmiş ve 1-inci gözlemin DFFITS ve COOK değerleri açısından aykırı olduğu saptanmıştır. Ayrıca tahmin edilen otokorelasyon katsayısı ve ridge yanlılık parametresi altında 8-inci gözlem COVRATIO üzerinde aykırı olarak belirlenmiştir. Kutu grafiği üzerinden Kaldıraç, CW ve AP değerlerine göre aykırı duran gözlem gözlenmemiştir.

1-inci gözlem AR(1) yapısından kaynaklı sistematik olarak etkin çıkarken, 8-inci ve 11-inci gözlemler bazı istatistikler üzerinde etkin bulunmuştur. Ancak bu gözlemleri silmek otokorelasyon ve çoklu iç ilişkinin gücünde önemli değişikliklere sebep olduğundan veri kaybetmek yerine gözlemlerin etkinliğini

6. NÜMERİK ÖRNEK: FRANSA EKONOMİ VERİSİ Tuğba SÖKÜT AÇAR

azaltan ARR tahmin edicisini kullanmak daha uygun bulunmuştur. Bu kapsamda birde otokorelasyon ve çoklu iç ilişki problemleri altında GEKK ve ARR ile uydurulan modelin geçerlilikleri karşılaştırılmıştır. Firinguetti (1989)'nin önerdiği k_1 ve k_2 ridge yanlılık parametresi Judge ve ark (1980)'nin verdiği dört farklı otokorelasyon tahmini için elde edilmiştir. Ancak unutulmamalıdır ki, Firinguetti (1989)'nin önerdiği yanlılık parametresi hata kareler ortalamasını minimum yapmayı amaçlamaktadır. Yani, k_1 ve k_2 kullanılarak elde edilen OCV^R , GCV^R ve C_p^R değerleri $k = 0$ 'da elde edilen değerlerden büyük çıkabilir. Bizim amacımız modelin öngörü performansını yani geçerliliği iyi kılacak bir yanlılık parametresi seçmektir. Bu bağlamda ARR'yi GEKK'dan üstün kılacak ridge yanlılık parametresinin seçimi en küçük OCV^R , GCV^R ve C_p^R değerlerini veren k olarak grafiksel gösterim yardımı ile belirlenmiştir.

İlk olarak GEKK ve ARR tahmin edicileri için rezidüler ve PRESS rezidüler incelenmiştir.

Çizelge 6.9. Fransa ekonomi verisi rezidüleri ($\hat{\rho}_1 = 0.5475, k_1 = 0.0012$)

Gözlem	Rezidü		PRESS Rezidü	
	GEKK	ARR	GEKK	ARR
1	0.0370	0.0382	0.0821	0.0840
2	0.0070	0.0046	0.0080	0.0052
3	0.0016	0.0051	0.0018	0.0057
4	-0.0046	0.0038	-0.0068	0.0051
5	-0.0143	-0.0126	-0.0154	-0.0135
6	-0.0260	-0.0303	-0.0292	-0.0334
7	-0.0222	-0.0286	-0.0323	-0.0396
8	-0.0117	-0.0268	-0.0294	-0.0446
9	-0.0308	-0.0299	-0.0337	-0.0327
10	-0.0326	-0.0319	-0.0368	-0.0360
11	-0.0182	-0.0031	-0.0362	-0.0044
12	0.0202	0.0167	0.0276	0.0224
13	0.0225	0.0212	0.0486	0.0456
14	0.0405	0.0411	0.0570	0.0578

Çizelge 6.9'da $\hat{\rho}_1 = 0.5475, k_1 = 0.0012$ 'da ki GEKK ve ARR tahmin edicilerine ait rezidüler ve PRESS rezidüler verilmiştir. Montgomery ve arkadaşları

6. NÜMERİK ÖRNEK: FRANSA EKONOMİ VERİSİ Tuğba SÖKÜT AÇAR

(2001, s. 138), PRESS rezidülerin ilgili sıradan rezidülerinden oldukça büyük olduğu gözlemlerin etkin olabileceğini ve etkin gözlem içeren model üzerinden tahmin yapmanın da sağlıklı olmadığını belirtmiştir. Görüldüğü üzere, bazı gözlemlere (1, 8, 11, 13 ve 14) ait PRESS rezidüler ilgili sıradan rezidülerden diğer gözlemlere oranla oldukça büyüktür. Bu durum, gözlemlerin etkin olma potansiyellerinin bir göstergesi olarak yorumlanabilmektedir. Unutulmamalı ki, tanılama yöntemlerinde de bu gözlemlerin etkin olma potansiyellerinin yüksek olduğu belirlenmiştir.

Model geçerliliğini en iyi kılan ridge yanlılık parametresini belirlemeden önce hata kareler ortalamasını minimum yapan ridge yanlılık parametresi değerlerine göre OCV, GCV ve C_p istatistikleri elde edilmiştir. Bunun için, OCV, GCV ve C_p istatistikleri Firinguetti (1989)'nın önerdiği iki farklı ridge yanlılık parametresi (k_1 ve k_2) tahmini dikkate alınarak uygulanmıştır.

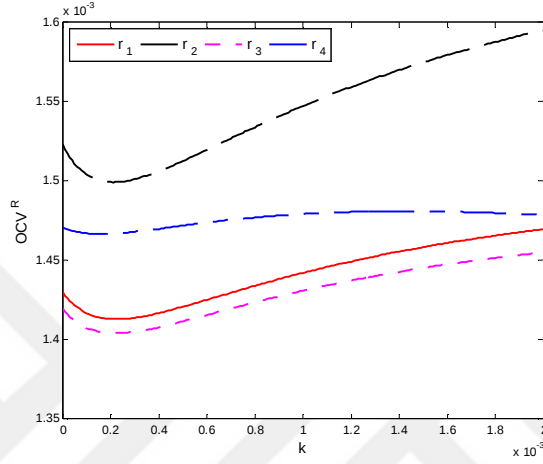
Çizelge 6.10. Fransa ekonomi verisi için OCV , OCV^R , GCV , GCV^R , C_p ve C_p^R değerleri

	$\hat{\rho} = 0.5475$ $k_1 = 0.0012$	$\hat{\rho} = 0.5475$ $k_2 = 0.0089$
OCV	0.001430210861775	0.001430210861775
OCV^R	0.001450522366980	0.001496248280296
GCV	0.784457930693725e-4	0.784457930693725e-4
GCV^R	0.759579050282469e-4	0.799543781316027e-4
C_p	4.000000000000000	4.000000000000000
C_p^R	3.677685898456254	4.314197198262147

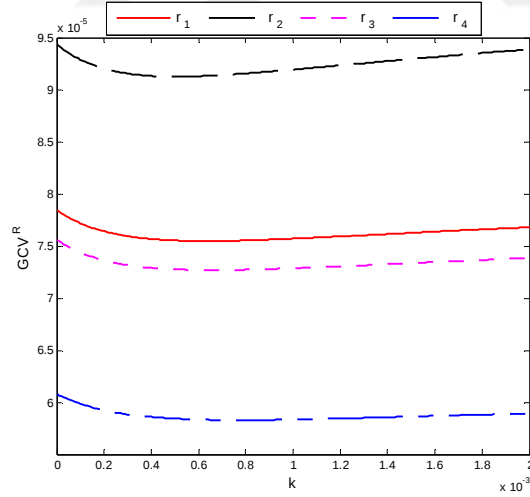
Uygulanan çapraz geçerlilik ölçüleri sonucunda elde edilen değerlerin küçük olması istenir. Çizelge 6.10'da verilen sonuçlar hata kareler ortalamasını minimum yapmak için önerilen k_1 ve k_2 gibi belirli ridge yanlılık parametrelerine göre sabit sonuçlardır. Çizelge 6.10 incelendiğinde k_2 ridge yanlılık parametresi ile elde edilen ARR tahmin edicisinin OCV^R , GCV^R ve C_p^R değerlerinin k_1 ile elde edilen değerlerden daha büyük olduğu görülmüştür. Ayrıca, $OCV < OCV^R$ iken $GCV > GCV^R$ ile $C_p > C_p^R$ olduğu gözlenmiştir. Bu durum OCV, GCV ve C_p

istatistiklerini minimum yapan k 'nın hata kareler ortalamasını minimum yapan k tahminlerinden farklı olduğunu göstermektedir.

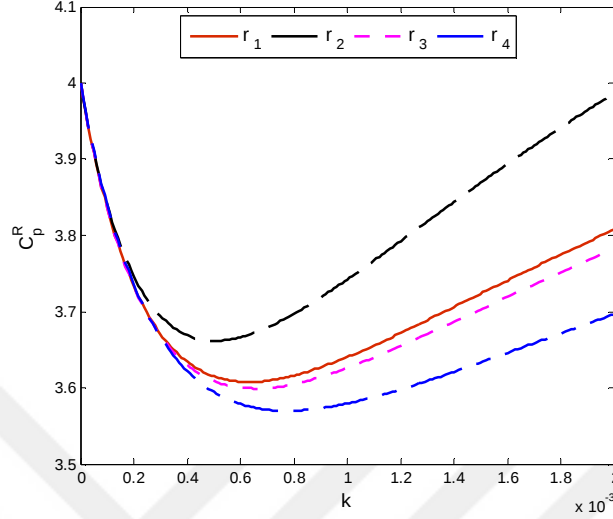
OCV, GCV ve Cp istatistiklerinin farklı k değerlerine göre davranışları Şekil 6.7-Şekil 6.9 ile verilmiştir.



Şekil 6.7. ARR tahmin edicisi için yanlılık parametresine karşı OCV^R , Fransa ekonomi verisi



Şekil 6.8. ARR tahmin edicisi için yanlılık parametresine karşı GCV^R , Fransa ekonomi verisi



Şekil 6.9. ARR tahmin edicisi için yanlılık parametresine karşı C_p^R , Fransa ekonomi verisi

Şekil 6.7-Şekil 6.9 incelendiğinde tahmin edilen dört farklı ρ değeri için OCV^R , GCV^R ve C_p^R istatistiklerini minimum yapan tek birer k değerlerinin olduğu görülmektedir. Detaylı incelemeler yapıldığında,

OCV^R 'yi minimum yapan k değerinin $k_{opt} \cong 0.0002$, GCV^R 'yi minimum yapan k değerinin ise tüm ρ değerlerinde $k_{opt} \cong 0.0007$ olduğu tespit edilmiştir. Ayrıca, Şekil 6.7'den görüldüğü üzere $\hat{\rho}_3$ katsayısında OCV^R değerleri $\hat{\rho}_1$, $\hat{\rho}_2$ ve $\hat{\rho}_4$ katsayısı ile elde edilen OCV^R değerlerinden daha küçüktür. GCV^R istatistiğinin $\hat{\rho}_4$ katsayısı ise tahmin edildiğinde $\hat{\rho}_1$, $\hat{\rho}_2$ ve $\hat{\rho}_3$ 'e göre daha küçük değerler aldığı Şekil 6.8'den görülmektedir.

Şekil 6.9'dan görüldüğü üzere, belirli bir k değerine kadar C_p^R değerleri azalmış ve bu noktadan sonra hızlı bir artış göstermiştir. Otokorelasyon katsayısı $\hat{\rho}_1$, $\hat{\rho}_3$ ve $\hat{\rho}_4$ olarak tahmin edildiğinde C_p^R istatistiğini minimum yapan ridge yanlılık parametresi $k \cong 0.0007$ iken otokorelasyon katsayısı $\hat{\rho}_2$ ile tahmin edildiğinde ise C_p istatistiğini minimum yapan ridge yanlılık parametresi $k \cong$

6. NÜMERİK ÖRNEK: FRANSA EKONOMİ VERİSİ Tuğba SÖKÜT AÇAR

0.0005 olarak tespit edilmiştir. Ayrıca, C_p^R istatistiği $\hat{\rho}_4$ 'da minimum değerini alırken sırası ile, $\hat{\rho}_3$, $\hat{\rho}_1$ ve $\hat{\rho}_2$ 'de daha büyük değerler almıştır.

Grafikler yardımı ile dört farklı otokorelasyon katsayısında OCV^R , GCV^R ve C_p^R 'yi minimum yapan k_{opt} değerleri belirlendikten sonra aşağıdaki sonuç elde edilmiştir.

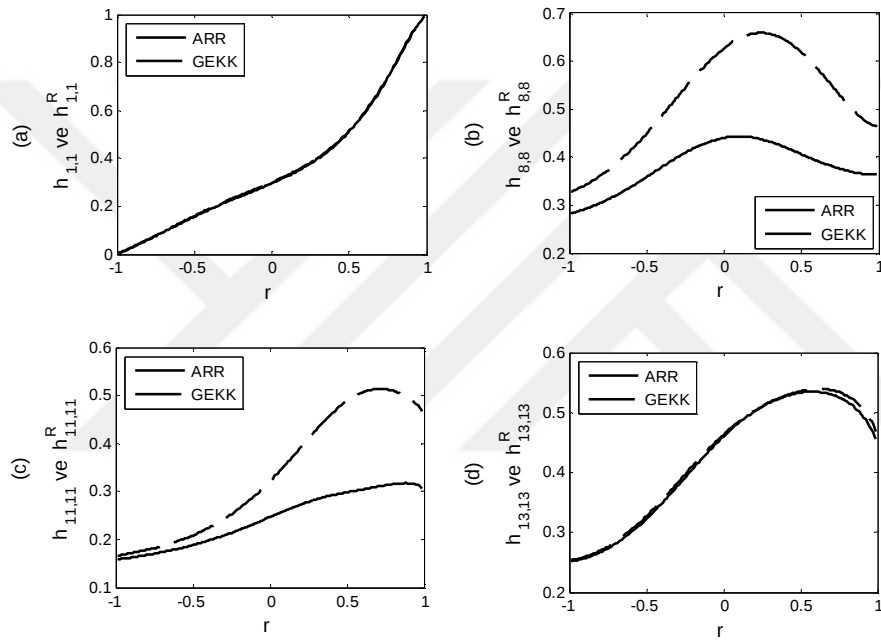
Çizelge 6.11. Fransa Ekonomi verisi için k_{opt} ile uydurulan modelin geçerliliği

	OCV	OCV^R	
$\hat{\rho}_1 = 0.5475$ $k_{opt} = 0.0002$	0.001430210861775	0.001413119541939	
$\hat{\rho}_2 = 0.4212$ $k_{opt} = 0.0002$	0.001522943761705	0.001499443522314	
$\hat{\rho}_3 = 0.5716$ $k_{opt} = 0.0002$	0.001419611653615	0.001404356814974	
$\hat{\rho}_4 = 0.7113$ $k_{opt} = 0.0002$	0.001470982157826	0.001466810938919	
	GCV	GCV^R	C_p^R
$\hat{\rho}_1 = 0.5475$ $k_{opt} = 0.0007$	0.78445793069e-4	0.75495351406e-4	3.609258673784382
$\hat{\rho}_2 = 0.4212$ $k_{opt} = 0.0005$	0.94349114671e-4	0.91260797745e-4	3.661594623514298
$\hat{\rho}_3 = 0.5716$ $k_{opt} = 0.0007$	0.75597049190e-4	0.72687387407e-4	3.599728745370028
$\hat{\rho}_4 = 0.7113$ $k_{opt} = 0.0007$	0.60816709038e-4	0.5831730136e-4	3.571461103583944

Çizelge 6.10 ile Çizelge 6.11 karşılaştırıldığında görülmüştür ki Firinguetti (1989)'nin en düşük hata kareler ortalamasını elde etmek için önerdiği k_1 ve k_2 parametreleri ile elde edilen OCV, GCV ve C_p değerleri minimum değildir. Bu durumda hata kareler ortalamasını minimum yapan k ile ARR tahmin edicisi kullanılarak uydurulan modelin geçerliliği en iyi değildir. Çizelge 6.10'da OCV değerleri incelendiğinde ARR ile elde edilen OCV değerleri GEKK ile elde edilenden daha büyük iken, k_{opt} kullanıldığında ARR ile elde edilen OCV değerleri GEKK ile elde edilenden daha küçüktür.

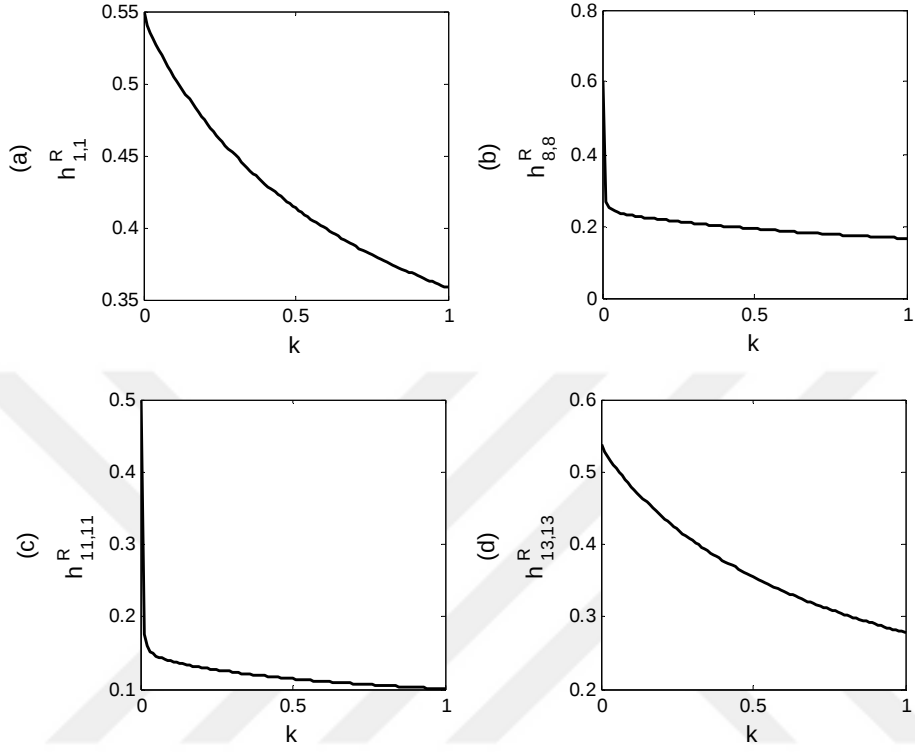
6. NÜMERİK ÖRNEK: FRANSA EKONOMİ VERİSİ Tuğba SÖKÜT AÇAR

Şu ana kadar ki değerlendirmeler tahmin edilen otokorelasyon katsayısı ve yanlılık parametresi üzerinden incelenmiştir. $-1 < \rho < 1$ ve $0 < k < \infty$ aralıklarında kaldırıcı değerler, DFBETA, DFFITS, COOK, COVRATIO, CW, AP, OCV, GCV ve Cp ölçümlerinin tutumları Şekil 6.10 - Şekil 6.27 ile incelenmiştir. Grafikselsel incelemelerde etkin olma potansiyelleri yüksek olan 1-nci, 8-inci, 11-inci ve 13-üncü gözlemler kullanılmıştır.



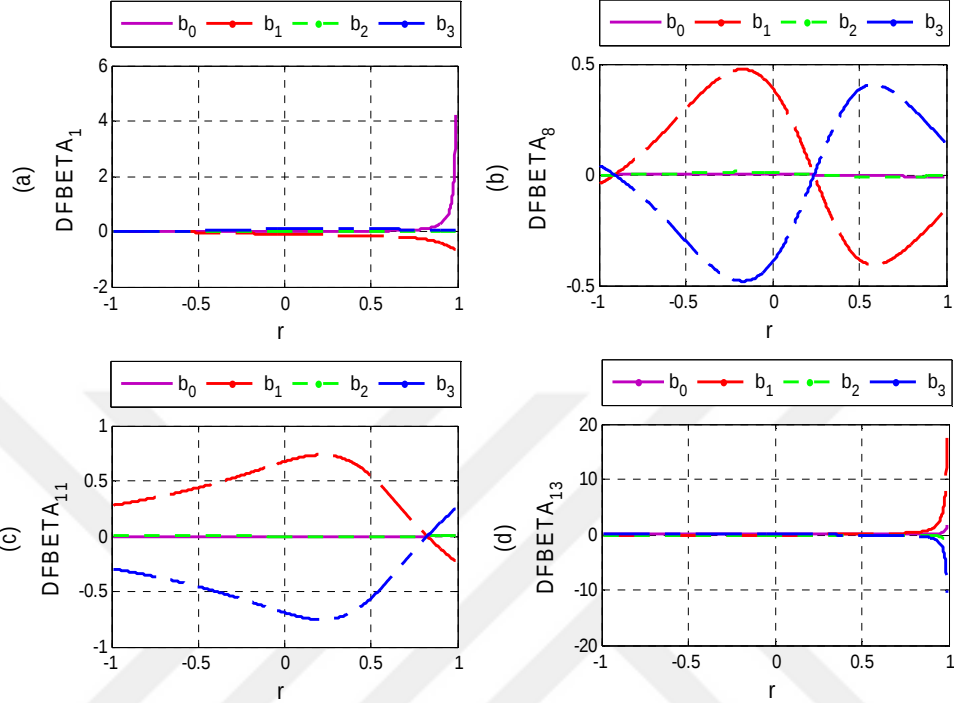
Şekil 6.10. Otokorelasyona karşı $h_{i,i}$ ve $h_{i,i}^R$ değerleri ($i = 1, 8, 11, 13$)

Şekil 6.10'dan görüldüğü üzere 1-inci ve 13-üncü gözlemde GEKK ve ARR tahmin edicileri ile elde edilen kaldırıcı değerler farklı otokorelasyonlarda birbirine çok yakın değerler almıştır (Şekil 6.10 a ve d). Ancak, bu durum 8-inci ve 11-inci gözlemlerin kaldırıcı değerleri için geçerli değildir. Açıkça görülmektedir ki, ARR tahmin edicisi kullanıldığında 8-inci ve 11-inci gözlemler için elde edilen kaldırıcı değerler GEKK ile elde edilen değerlerden daha küçüktür. Genel olarak $|\rho| \rightarrow 1$ iken $h_{i,i}$, $h_{i,i}^R$ ($i \neq 1$) değerleri azalmakta ve bu azalış $\rho \rightarrow -1$ 'de daha belirgindir.



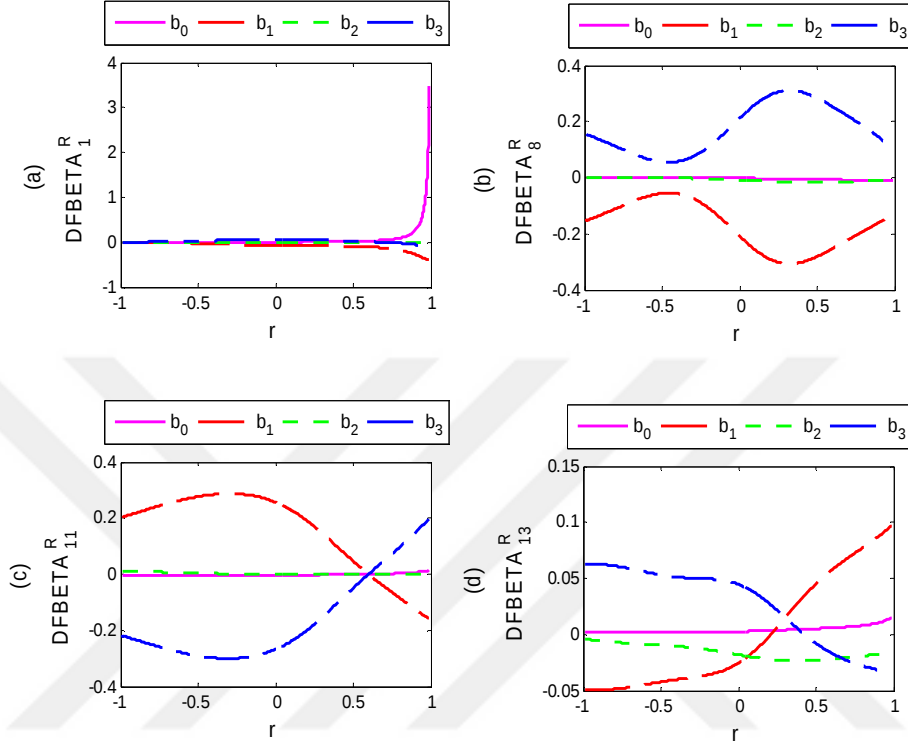
Şekil 6.11. Yanlılık parametresine karşı $h_{i,i}$ ve $h_{i,i}^R$ değerleri ($i = 1, 8, 11, 13$)

Şekil 6.11'den görüldüğü üzere, tüm gözlemler için kaldıraç değerler k arttıkça azalmaktadır. Bu durumda, ARR tahmin edicisi ile elde edilen kaldıraç değerler her zaman GEKK ile elde edilen kaldıraç değerlerden daha küçüktür.



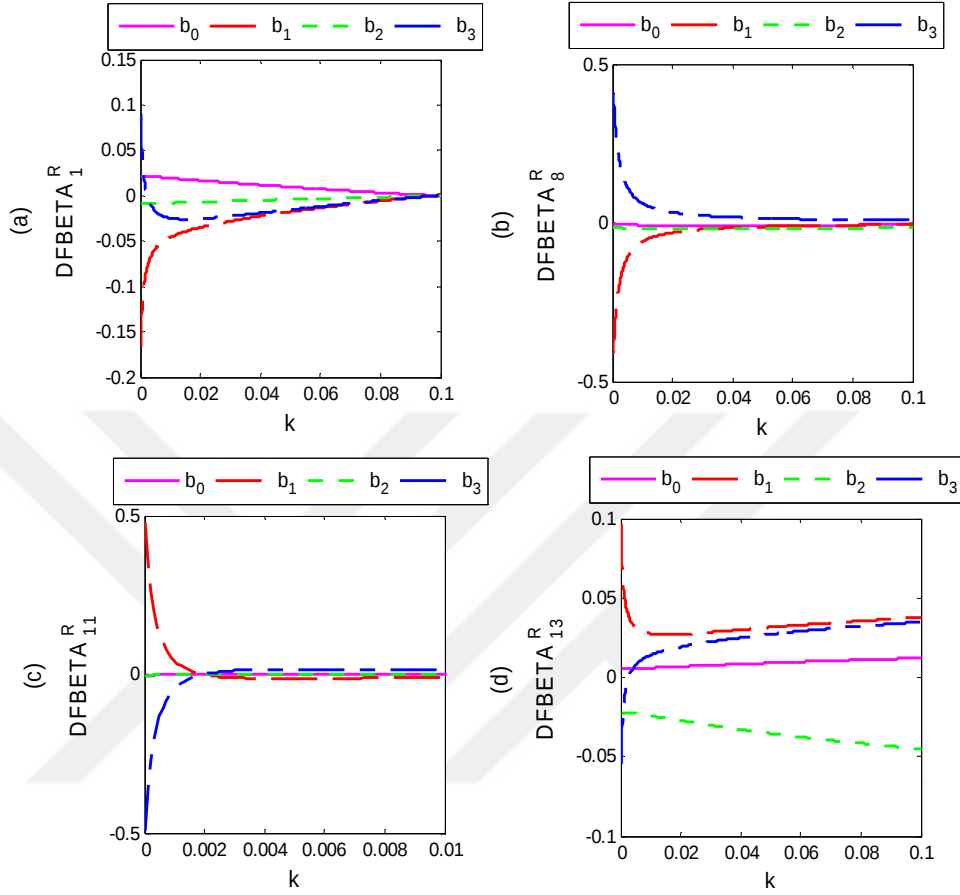
Şekil 6.12. GEKK tahmin edicisi için otokorelasyona karşı $DFBETA_i$ grafiği ($i = 1, 8, 11, 13$)

Şekil 6.12’de GEKK tahmin edicisi kullanıldığında farklı ρ değerleri için gözlemlerin regresyon katsayıları üzerindeki etkisi incelenmiştir. $\rho \rightarrow 1$ iken β_0 ve β_1 üzerinde 1-inci gözlemin etkisi artmaktadır. Farklı ρ değerlerinde β_2 ve β_3 üzerinde birinci gözlemin etkisi β_0 ve β_1 ’e göre belirgin değildir. Çizelge 6.2 ve Çizelge 6.3’e göre etkin olma potansiyeli yüksek olan 8-inci, 11-inci ve 13-üncü gözlemlerin farklı ρ değerleri için β_1 ve β_3 regresyon katsayısı üzerindeki etkisinin değiştiği görülmüştür.



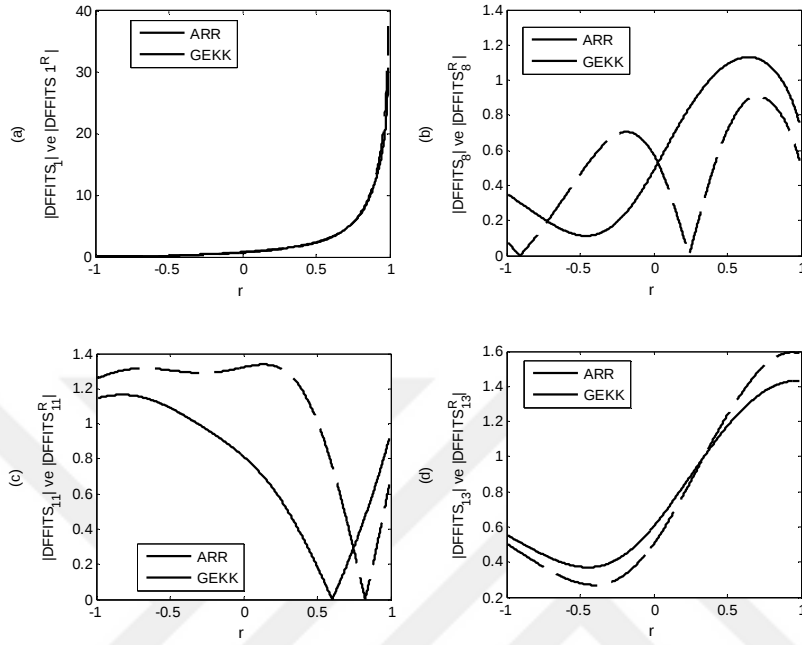
Şekil 6.13. ARR tahmin edicisi için otokorelasyona karşı $DFBETA_i^R$ grafiği ($i = 1, 8, 11, 13$)

Şekil 6.13’de ARR tahmin edicisi kullanıldığında farklı ρ değerleri için gözlemlerin regresyon katsayıları üzerindeki etkisi verilmiştir. $\rho \rightarrow 1$ iken ARR ile elde edilen β_0 ve β_1 regresyon katsayıları üzerinde birinci gözlemin etkisi artmaktadır. Farklı ρ değerlerinde ARR tahmin edicisi ile elde edilen β_2 ve β_3 katsayıları üzerinde birinci gözlemin etkisi yoktur. 8-inci, 11-inci ve 13-üncü gözlemlerin farklı ρ değerleri için β_1 ve β_3 regresyon katsayısı üzerindeki etkisinin değiştiği görülmüştür.



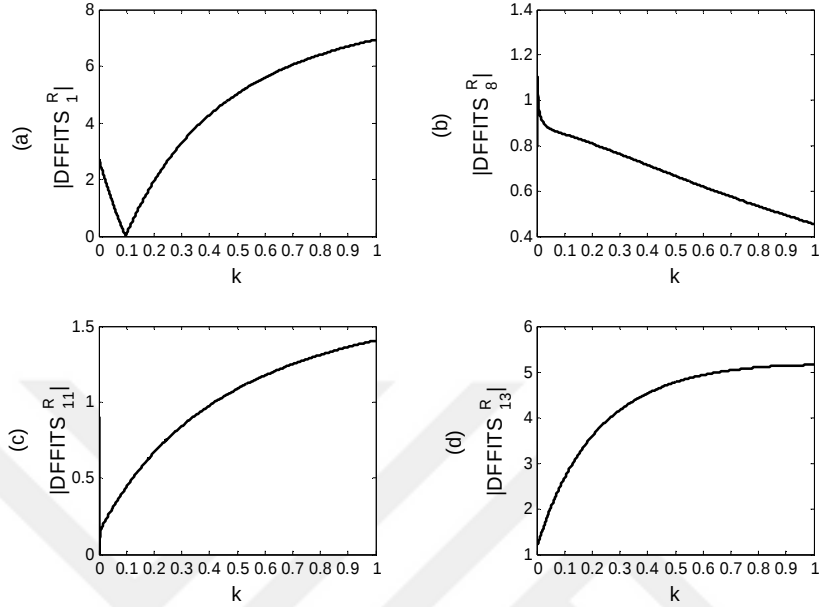
Şekil 6.14. Ridge yanlılık parametresine karşı $DFBETA_i^R$ grafiği ($i = 1, 8, 11, 13$)

Şekil 6.14’de farklı ridge yanlılık parametre değerlerine karşılık gözlemlerin ARR ile elde edilen regresyon katsayıları üzerindeki etkisi verilmiştir. 13-üncü gözlemin β_2 regresyon katsayısı üzerindeki etkisi hariç diğer tüm gözlemlerin parametre vektörü üzerindeki etkisi k arttıkça azalmaktadır.



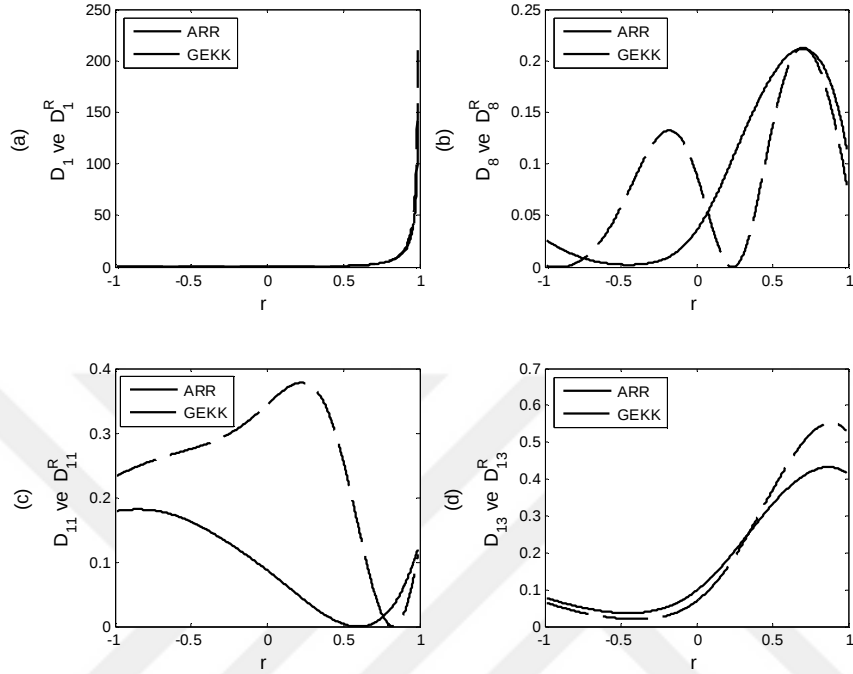
Şekil 6.15. Otokorelasyona karşı $DFFITS_i$ ve $DFFITS_i^R$ değerleri ($i = 1, 8, 11, 13$)

Şekil 6.15, farklı otokorelasyon değerlerinde GEKK ve ARR tahmin edicileri için mutlak değerce DFFITS değerlerini göstermektedir. $\rho \rightarrow 1$ 'de $|DFFITS_1|$ ve $|DFFITS_1^R|$ değerleri artarken $\rho \rightarrow -1$ 'de azalmaktadır (Şekil 6.15.a). Yanıt değişken üzerinde ilk dönüştürülmüş gözlemin etkisi GEKK ve ARR tahmin edicileri kullanıldığında hemen hemen aynıdır. $\hat{\rho}_1 = 0.5475$ iken etkin olma potansiyeli yüksek olan 13-üncü gözlemin farklı ρ için davranışı izlendiğinde, ρ 'nun pozitif yönde artması ile $|DFFITS_{13}|$ ve $|DFFITS_{13}^R|$ değerlerinde artış gözlenmiştir. Yani otokorelasyon pozitif yönde arttıkça 13-üncü gözlemin yanıt değişken üzerindeki etkisi artmaktadır (Şekil 6.15.d). 8-inci ve 11-inci gözlemler için otokorelasyon katsayısına göre DFFITS değerleri belirgin bir yapı sergilememiştir.



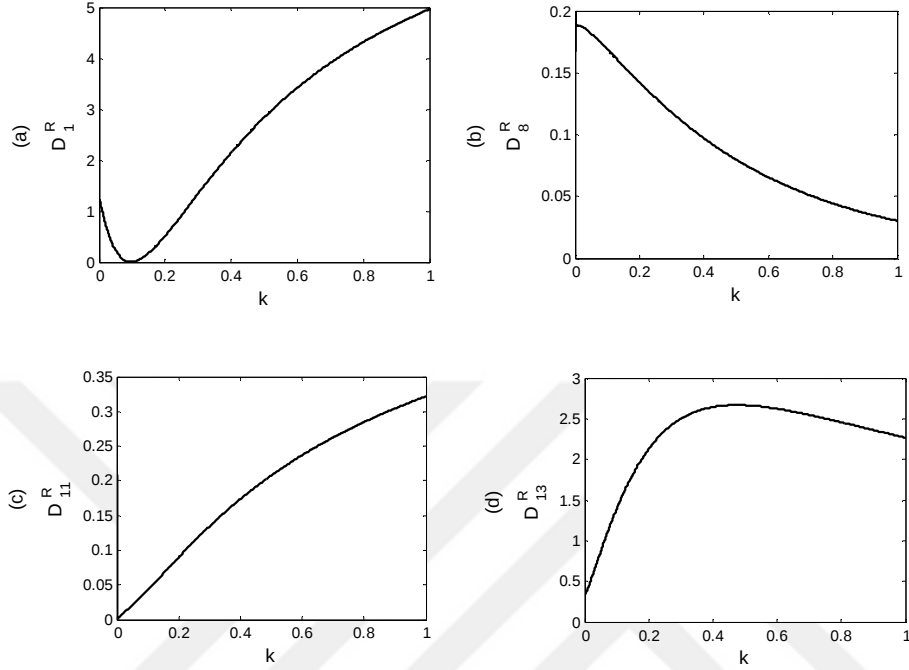
Şekil 6.16. Ridge yanalılık parametresine karşı $DFFITS_i^R$ değerleri
($i = 1, 8, 11, 13$)

Şekil 6.16'dan söyleyebiliriz ki, $k \cong 0.1$ ' e kadar ARR tahmin edicisi ile uydurulan yanıt değişken üzerinde 1-inci gözlemin etkisi azalırken, $k \cong 0.1$ ' den sonra etki artmıştır. $|DFFITS_i^R|$ ($i = 11, 13$) değerleri k 'ya bağlı olarak artarken, $|DFFITS_8^R|$ değerleri k 'ya bağlı olarak azalmaktadır. Walker ve Birch (1988, s. 224) ilişkisiz ve sabit varyanslı hataya sahip lineer istatistiksel modelde k 'ya karşı DFFITS grafiğini elde etmiş ve bazı gözlemlerin k 'ya bağlı olarak arttığını belirtirken, bazılarının k 'ya bağlı olarak azaldığını belirtmiştir. İlişkili hataya sahip lineer istatistiksel modelde Walker ve Birch (1988)'e paralel olarak bazı gözlemlerin DFFITS değerleri k 'ya bağlı olarak artarken bazıları k 'ya bağlı olarak azalmaktadır.



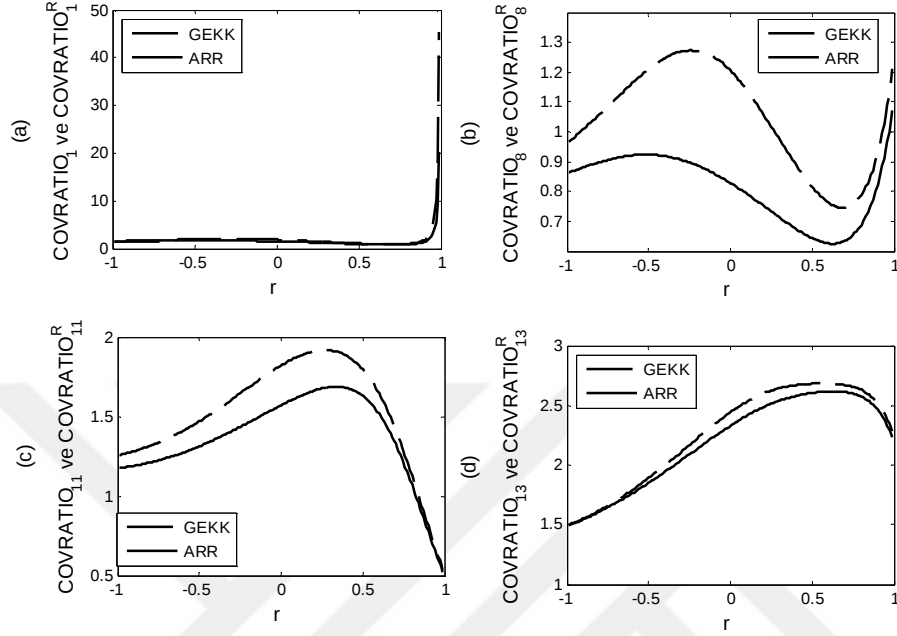
Şekil 6.17. Otokorelasyona karşı D_i ve D_i^R değerleri ($i = 1, 8, 11, 13$)

Şekil 6.17'den görüldüğü üzere 1-inci ve 13-üncü gözlemler için otokorelasyon katsayısının farklı değerlerinde Cook grafiği ilgili gözlemlerin DFFITS grafiği ile benzer yapı göstermiştir. $\rho \rightarrow 1$ iken D_1 ve D_1^R değerleri hızlı bir şekilde artmakta, $\rho \rightarrow -1$ iken D_1 ve D_1^R değerleri sıfıra yakınsamaktadır. Otokorelasyon katsayısına göre Cook uzaklıklarının grafiği analiz edildiğinde GEKK ve ARR tahmin edicisi ile ele elde edilen birinci dönüştürülmüş gözlemin Cook değerleri arasında önemli bir farklılık olmadığı görülmüştür.



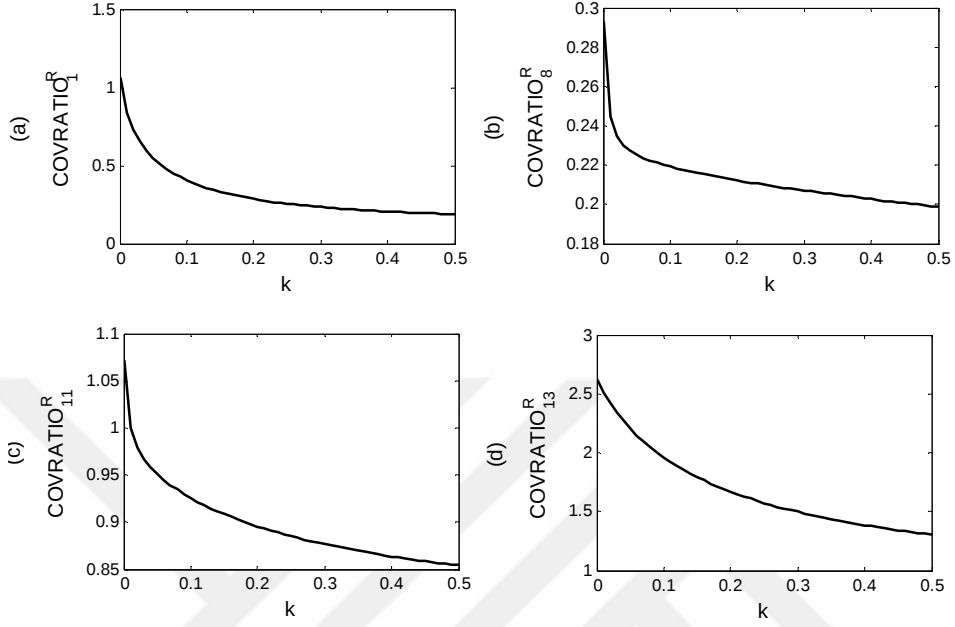
Şekil 6.18. Ridge yanlılık parametresine karşı D_i^R değerleri ($i = 1, 8, 11, 13$)

Şekil 6.18’de farklı k değerleri için $D_i^R, i = 1, 8, 11, 13$ değerlerinin grafiği verilmiştir. $D_1^R, k \cong 0.1$ ’e kadar azalırken, $k \cong 0.1$ ’den sonra hızlı bir şekilde artmıştır. D_8^R değeri k arttıkça azalırken D_{11}^R değeri k arttıkça artmaktadır. D_{13}^R değeri $k \cong 0.3$ ’e kadar hızlı bir şekilde artarken, bu değerden sonra yavaşça azalmıştır. Şekil 6.18 genel olarak değerlendirildiğinde farklı gözlemlerin k ’ya göre D_i^R davranışlarının değiştiği görülmektedir.



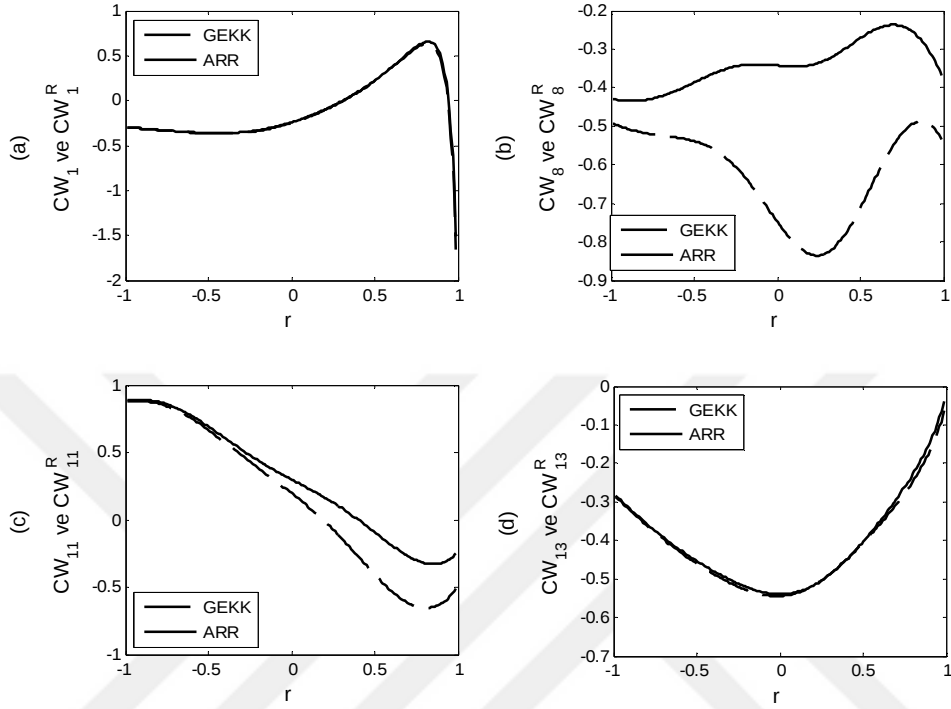
Şekil 6.19. Otokorelasyona karşı $COVRATIO_i$ ve $COVRATIO_i^R$ değerleri ($i = 1, 8, 11, 13$)

Şekil 6.19’da farklı ρ değerleri için GEKK ve ARR tahmin edicileri ile elde edilen COVRATIO ölçümlerinin grafiği verilmiştir. İlk dönüştürülmüş gözlemin parametre vektörüne ait varyans-kovaryans matrisi üzerindeki etkisi hem GEKK hem de ARR tahmin edicileri kullanıldığında benzer yapı sergilemekle birlikte $\rho \rightarrow 1$ iken hızlı bir artış göstermektedir. Diğer gözlemlerin varyans-kovaryans matrisi üzerindeki etkisi ρ ’ya göre belirgin bir yapı sergilememektedir. Ancak Şekil 6.19 üzerinden görülmektedir ki, farklı ρ değerlerinde GEKK tahmin edicisi ile elde edilen COVRATIO ölçüm değerleri ARR ile elde edilenlerden daha büyüktür.



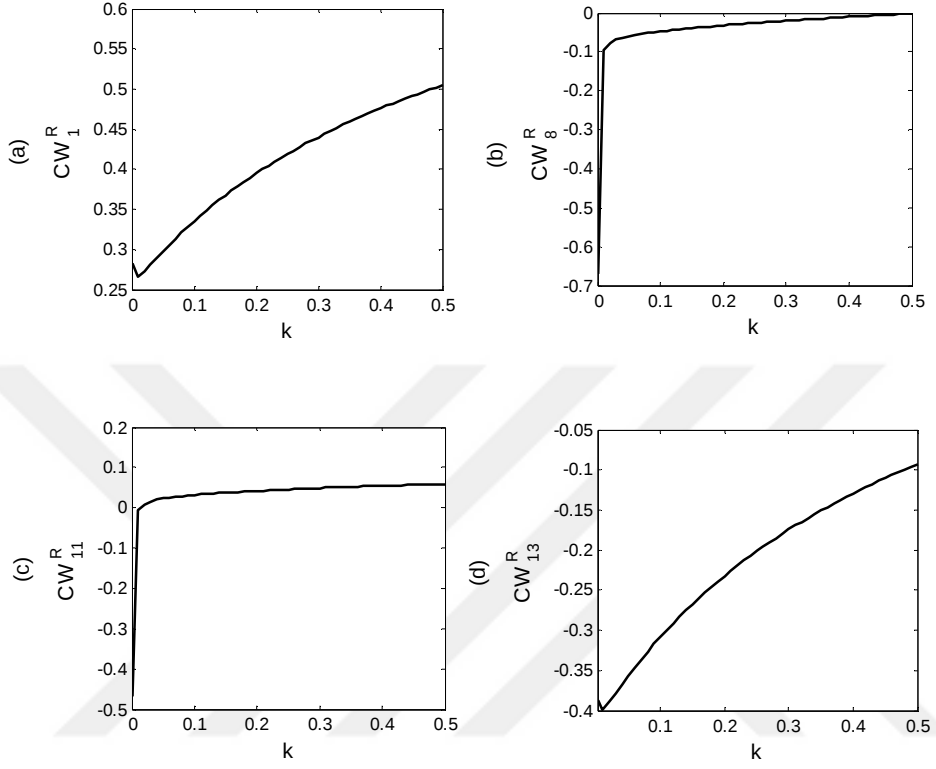
Şekil 6.20. Ridge yanlılık parametresine karşı $COVRATIO_i^R$ değerleri ($i = 1, 8, 11, 13$)

Şekil 6.20’de farklı k değerleri için COVRATIO ölçümlerinin grafiği verilmiştir. Net olarak görülmektedir ki, k arttıkça COVRATIO değerleri azalmaktadır.



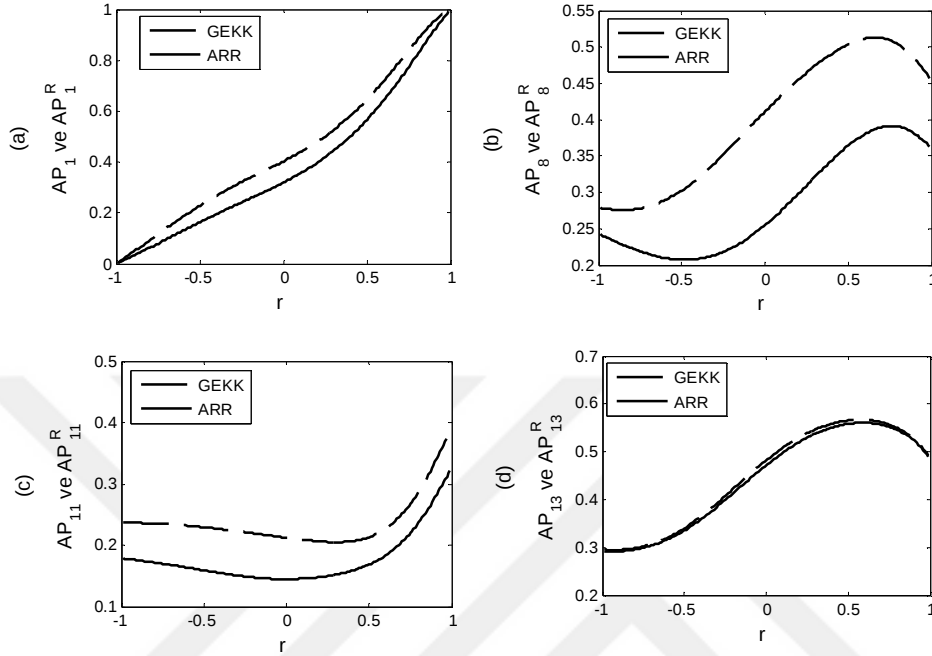
Şekil 6.21. Otokorelasyona karşı CW_i ve CW_i^R değerleri ($i = 1, 8, 11, 13$)

Şekil 6.21’de farklı ρ değerlerinde GEKK ve ARR tahmin edicileri ile elde edilen CW ölçümlerinin grafiği verilmiştir. İlk dönüştürülmüş gözleme ait CW değerleri hem GEKK hem de ARR tahmin edicileri kullanıldığında benzer yapı sergilemekle birlikte, $\rho \rightarrow 1$ iken CW_1 ve CW_1^R değerleri negatif yönde artmaktadır. Diğer gözlemlerin CW değerleri ρ ’ya göre belirgin bir yapı sergilememekte ancak tüm gözlemler için otokorelasyona karşı tutum COVRATIO tutumlarının tersidir.



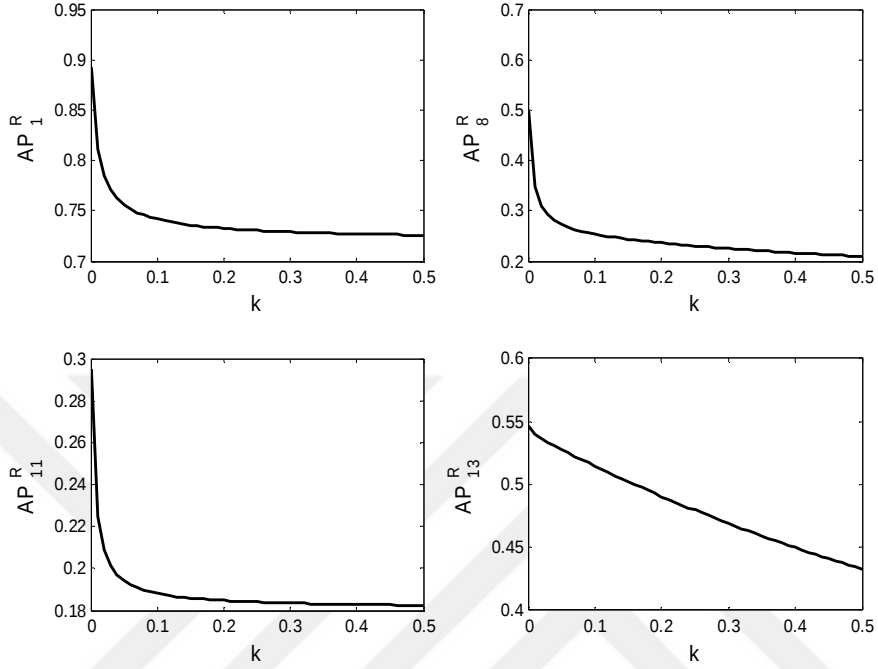
Şekil 6.22. Ridge yanalılık parametresine karşı CW_i^R değerleri ($i = 1, 8, 11, 13$)

Şekil 6.22’de farklı k değerleri için CW_i^R ölçümlerinin grafiği verilmiştir. CW_1^R , CW_8^R ve CW_{11}^R değerleri $k \cong 0.01$ ’e kadar azalırken, bu değerden sonra artış göstermiştir. CW_{13}^R değerleri ise $k \cong 0.01$ ’e kadar çok az artış gösterirken bu değerden sonra hızlı bir azalış göstermiştir.



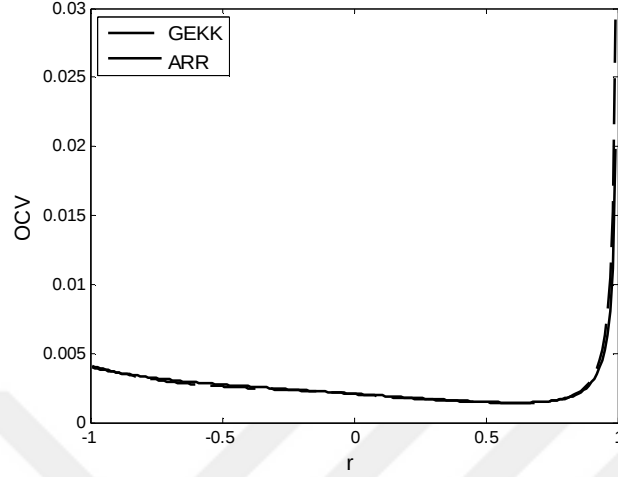
Şekil 6.23. Otokorelasyona karşı AP_i ve AP_i^R değerleri ($i = 1, 8, 11, 13$)

Şekil 6.23’de farklı ρ değerleri için AP ölçümlerinin grafiği verilmiştir. İlk dönüştürülmüş gözlemin güven elipsoidinin hacmi üzerindeki etkisi hem GEKK hem de ARR tahmin edicileri kullanıldığında benzer yapı sergilemekle birlikte $\rho \rightarrow 1$ iken AP_1 ve AP_1^R değerleri artmakta, $\rho \rightarrow -1$ iken ise azalmaktadır. Diğer gözlemlerin güven elipsoidinin hacmi üzerindeki etkisi ρ ’ya göre belirgin bir yapı sergilememektedir. Ancak Şekil 6.23 üzerinden görülmektedir ki, farklı ρ değerlerinde ARR tahmin edicisi ile elde edilen AP değerleri GEKK ile elde edilenlerden daha küçüktür.

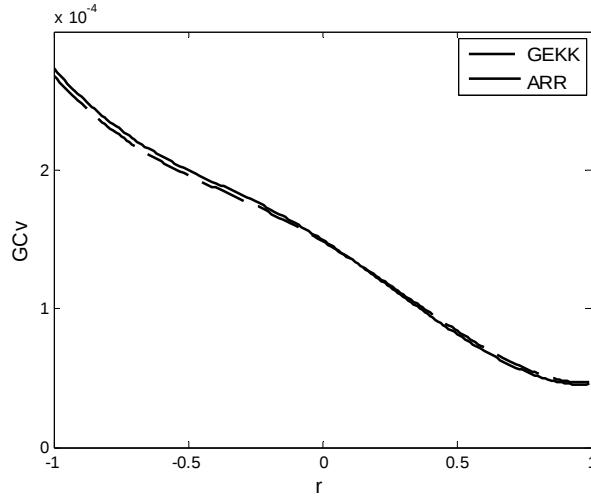


Şekil 6.24. Ridge yanlılık parametresine karşı AP_i^R değerleri ($i = 1, 8, 11, 13$)

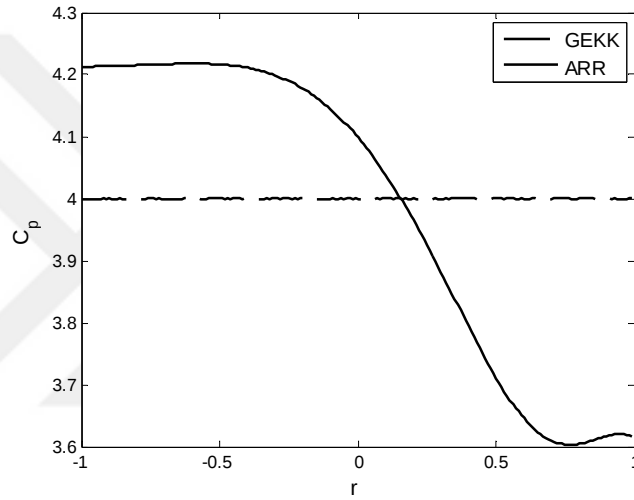
k 'ya karşı $AP_i^R, i = 1, 8, 11, 13$ grafiklerinin verildiği Şekil 6.24, k arttıkça AP_i^R değerlerinin azaldığını göstermektedir.

Şekil 6.25. Otokorelasyona karşı OCV ve OCV^R değerleri

Şekil 6.25’den görüldüğü üzere, $\rho \rightarrow -1$ iken OCV ve OCV^R değerleri artış göstermektedir. OCV ve OCV^R değerlerinin ρ ’nun yaklaşık 0.7 değerine kadar azalmakta ve bu değerden sonra hızlı bir artış göstermektedir. Sabit k değerinde farklı ρ ’ya göre OCV ve OCV^R değerleri birbirine yakın değerler almıştır.

Şekil 6.26. Otokorelasyona karşı GCV ve GCV^R değerleri

Şekil 6.26'da GCV ölçümlerinin otokorelasyon katsayısına karşı grafiği verilmiştir. Buna göre, hem GEKK hem de ARR tahmin edicileri için $\rho \rightarrow -1$ iken GCV değerlerinin arttığı, $\rho \rightarrow 1$ ise azaldıkları gözlenmiştir. GEKK ve ARR tahmin edicilerinin model geçerliliği üzerindeki etkileri karşılaştırıldığında, negatif otokorelasyon bölgesinde GCV^R , GCV 'den daha yüksek iken, pozitif otokorelasyon bölgesinde GCV , GCV^R 'dan daha yüksek değerler almıştır.



Şekil 6.27. Otokorelasyona karşı C_p ve C_p^R değerleri

Şekil 6.27'de, C_p ve C_p^R değerlerinin otokorelasyon katsayısına karşı grafiği verilmiştir. Buna göre, $\rho \rightarrow -1$ iken C_p^R istatistiği artmaktadır. Otokorelasyonun pozitif yönde artması ile C_p^R istatistiği $\rho \cong 0.8$ 'e kadar hızlı bir şekilde azalmakta, $\rho \cong 0.8$ 'den sonra yavaş bir artış gösterip tekrar azalmaktadır. σ yansız tahmin edildiğinden GEKK tahmin edici için C_p istatistiği otokorelasyondan bağımsızdır. Böylece her otokorelasyon noktasında GEKK için C_p sabit terim içeren açıklayıcı değişken sayısı olan $p=4$ 'e eşittir. Bu durumda Şekil 6.27, negatif otokorelasyon bölgesinde $C_p^R > C_p$ iken, $\rho \cong 0.2$ değerinden sonra $C_p > C_p^R$ olduğunu göstermektedir.

6. NÜMERİK ÖRNEK: FRANSA EKONOMİ VERİSİ Tuğba SÖKÜT AÇAR

Nümerik örnek ile 14 gözlemlı bir veri kümesinde tüm gözlemlerin parametre vektörü, yanıt vektörü, güven elipsoidi üzerindeki etkinlikleri değerlendirilmiştir. Tüm gözlemler için ARR ile elde edilen kaldıraç, COVRATIO ve AP değerleri GEKK ile elde edilen değerlere göre daha küçük elde edilmiştir. Bununla birlikte 1-inci, 8-inci, 11-inci ve 13-üncü gözlemlerin etkinliklerinin kullanılan tahmin ediciye göre önemli farklılıklar gösterdiği gözlenmiştir. GEKK tahmin edicisi kullanıldığında regresyon katsayısı üzerinde etkin olan gözlemlerin etkisi ARR tahmin edicisi kullanıldığında azalmaktadır. Özellikle, ARR tahmin edicisi kullanıldığında 11-inci gözlemin etkinlik performansının iyileştiği görülmüştür. Otokorelasyon katsayısının 1'e giderken 1-inci gözlemin kaldıraçlığı, regresyon parametreleri üzerindeki etkinliği, yanıt değişken üzerindeki etkisi, varyans-kovaryans matrisi üzerindeki etkisi, güven elipsoidinin hacmi üzerindeki etkisi artmaktadır. Diğer gözlemlerin etkinliği otokorelasyon ve ridge yanlılık parametresine göre değişmektedir. Ridge yanlılık parametresi arttıkça ARR tahmin edicisi ile elde edilen kaldıraç değerler küçülmektedir. İlgili regresyon parametresi üzerinden etkin olan gözlemlerin etkinliği ridge yanlılık parametresi arttıkça azalmaktadır. ARR tahmin edicisi ile uydurulan yanıt vektörü üzerinde bazı gözlemlerin etkinliği ridge yanlılık parametresine bağlı olarak artarken bazılarının etkinliği ridge yanlılık parametresine bağlı olarak azalmaktadır. Parametre vektörüne ait varyans-kovaryans matrisi ve parametre vektörüne ait güven elipsoidinin hacmi üzerinde etkin olan gözlemlerin etkinliği ridge yanlılık parametresi arttıkça azalmaktadır.

Otokorelasyon ve çoklu iç ilişki problemlerinin var olduğu veri kümesinde ARR tahmin edicisi kullanmak gözlemlerin etkinliği açısından GEKK kullanmaktan daha iyidir. İki tahmin edici kullanılarak uydurulan modelin ön görüşü performansı değerlendirilmiş ve ARR tahmin edicisi kullanıldığında modelin öngörü performansını GEKK'dan daha iyi kılan bir ridge yanlılık parametresinin olduğu görülmüştür. En iyi ridge yanlılık parametresi otokorelasyon katsayısından etkilenmektedir.

7. MONTE CARLO SİMÜLASYON UYGULAMASI

Otokorelasyonlu hataya sahip lineer istatistiksel modelde simülasyon çalışması ilk olarak Firinguetti (1989) tarafından uygulanmıştır. İlişkili hataya ek olarak çoklu iç ilişkiye sahip açıklayıcı değişkenleri üretmek için McDonald ve Galarneau (1975) tarafından verilen aşağıdaki eşitlik kullanılmıştır

$$x_{ij} = (1 - \gamma^2)^{1/2} d_{ij} + \gamma d_{i(r+1)}, i = 1, \dots, n, j = 1, \dots, r. \quad (7.1)$$

Burada d_{ij} , normal rastgele sayıları, γ^2 ise herhangi iki açıklayıcı değişken arasındaki ilişkiyi göstermektedir.

Bu bölümde açıklayıcı değişkenler ilişkili olacak şekilde üretildikten sonra teoride incelenen her bir başlık için Monte Carlo simülasyon çalışması gerçekleştirilmiş ve aşamaları alt bölümler halinde verilmiştir.

7.1. Regresyon Katsayıları ve Yanıt Değişkene Dayalı Tanılama için Monte Carlo Simülasyon

Otokorelasyon katsayısına ve çoklu iç ilişkinin şiddetine bağlı olarak kaldıraç ve etkin gözlemlerin davranışlarını belirlemek adına Monte Carlo simülasyon çalışması uygulanmıştır.

- Ø Örneklem genişliği $n = 20, 100$ ve 500 olarak belirlenmiştir.
- Ø Açıklayıcı değişken sayısı için $r = 3, p = 4$ olarak seçilmiştir.
- Ø Çoklu iç ilişkinin şiddeti $\gamma^2 = 0.30, 0.50, 0.70$ ve 0.90 olarak keyfi seçilmiş ve açıklayıcı değişkenler (7.1) eşitliğinden üretilmiştir. X değişkenleri üretildikten sonra açıklayıcı değişkenler matrisi $Z = [1 \ X]$ olarak alınmıştır.
- Ø Brown (1977)'e paralel olarak X matrisi, merkezleştirilmiş ve standartlaştırılmış forma dönüştürülürken, yanıt vektörü y orjinal formunu korumuştur.

- Ø $\beta=(2, 2, 2, 2)'$ ve $\sigma_u=1$ alınmıştır.
- Ø Otokorelasyon katsayısı $|\rho| = 0.50, 0.70, 0.90, 0.99$ olarak seçilerek hatalar (3.2)'de ki modelden üretilmiştir.
- Ø Yanıt değişkene ait değerler aşağıdaki modelden elde edilmiştir

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \varepsilon_i \quad (7.2)$$
 burada $\varepsilon \sim N(0, \sigma_u^2 V)$ dir.
- Ø Her veri kümesi için 7-inci gözlem etkin gözlem olarak değiştirilmiştir. 7-inci gözlemi etkin olarak belirlemek için ilk olarak her bir veri kümesinde y vektöründeki maksimum değer belirlenmiş ve daha sonra bu maksimum değere 5 eklenmiş ve bu değer 7-inci gözleme atanmıştır. Açıklayıcı değişken üzerinde herhangi bir değişiklik yapılmamıştır. Hata terimleri üretildikten sonra y üzerindeki 7-inci gözlem etkin yapıldığından dolayı, y vektörünün varyans-kovaryans matrisi değişmiştir. Ancak her bir denemede dw testi hataların AR(1) olduğunu göstermiştir. Açıklayıcı değişken matrisi ilişkili olarak üretildikten sonra 7-inci gözlem etkin olarak değiştirildiğinden dolayı, 7-inci gözlem çoklu bağlantı-etkin (collinearity-influent) gözlem değildir. Diğer bir deyişle, bu gözlem veri kümesinde olsun ya da olmasın çoklu iç ilişki problemi vardır. 7-inci gözlem etkin olarak değiştirilmesine ek olarak bu çalışmada 1-inci gözlem de ele alınmıştır. Bunun nedeni literatürde AR(1) hatalı modelde 1-inci gözlemin sadece kaldırıcı özelliğinin incelenmesidir.
- Ø Ridge yanlılık parametresi, k , Firinguetti (1989)'nin önerdiği k_1 'deki gibi tahmin edilmiştir.
- Ø Son olarak her defasında yeni bir hata terimi üreterek simülasyon 1000 deneme üzerinden gerçekleştirilmiş ve ortalama istatistikler elde edilmiştir.

7. MONTE CARLO SİMÜLASYON UYGULAMASI Tuğba SÖKÜT AÇAR

Farklı otokorelasyon katsayısı, çoklu iç ilişki gücü ve örneklem genişliğinde etkinlik ölçümlerinin ve kaldırıcı gözlemlerin davranışı Çizelge 7.1-Çizelge 7.14 arasında gösterilmiştir.

Çizelge 7.1. Farklı ρ , γ ve n için 1-inci gözleme ait kaldırıcı değerler

ρ	n	$\gamma=0.90$		$\gamma=0.70$		$\gamma=0.50$		$\gamma=0.30$	
		GEKK	ARR	GEKK	ARR	GEKK	ARR	GEKK	ARR
-0.99	20	0.0013	0.0011	0.0015	0.0013	0.0018	0.0016	0.0022	0.0019
	100	0.0005	0.0004	0.0004	0.0003	0.0004	0.0002	0.0003	0.0002
	500	0.0002	0.0001	0.0002	0.0001	0.0001	0.0001	0.0001	0.0001
-0.90	20	0.0136	0.0109	0.0159	0.0135	0.0187	0.0163	0.0228	0.0199
	100	0.0057	0.0039	0.0046	0.0033	0.0037	0.0027	0.0029	0.0023
	500	0.0019	0.0014	0.0017	0.0014	0.0015	0.0012	0.0012	0.0010
-0.70	20	0.0434	0.0340	0.0507	0.0422	0.0595	0.0511	0.0719	0.0624
	100	0.0184	0.0128	0.0149	0.0110	0.0121	0.0092	0.0096	0.0075
	500	0.0061	0.0046	0.0054	0.0044	0.0048	0.0040	0.0040	0.0034
-0.50	20	0.0750	0.0561	0.0868	0.0701	0.1012	0.0849	0.1217	0.1034
	100	0.0317	0.0225	0.0259	0.0192	0.0211	0.0161	0.0170	0.0134
	500	0.0107	0.0080	0.0095	0.0076	0.0084	0.0069	0.0072	0.0059
0.50	20	0.1850	0.1506	0.1918	0.1637	0.2021	0.1762	0.2198	0.1913
	100	0.0528	0.0466	0.0487	0.0438	0.0456	0.0417	0.0434	0.0402
	500	0.0162	0.0135	0.0149	0.0130	0.0138	0.0122	0.0124	0.0112
0.70	20	0.2558	0.2369	0.2591	0.2436	0.2644	0.2501	0.2738	0.2581
	100	0.0670	0.0637	0.0647	0.0622	0.0631	0.0610	0.0620	0.0603
	500	0.0170	0.0156	0.0163	0.0153	0.0157	0.0149	0.0149	0.0143
0.90	20	0.5066	0.5016	0.5072	0.5031	0.5083	0.5046	0.5104	0.5065
	100	0.1644	0.1636	0.1638	0.1632	0.1634	0.1629	0.1631	0.1626
	500	0.0384	0.0380	0.0382	0.0379	0.0380	0.0378	0.0378	0.0376
0.99	20	0.9134	0.9130	0.9134	0.9130	0.9134	0.9131	0.9135	0.9133
	100	0.6679	0.6679	0.6679	0.6679	0.6679	0.6679	0.6679	0.6679
	500	0.2852	0.2852	0.2852	0.2852	0.2852	0.2852	0.2852	0.2852

7. MONTE CARLO SİMÜLASYON UYGULAMASI Tuğba SÖKÜT AÇAR

Çizelge 7.2. Farklı ρ , γ ve n için 7-inci gözleme ait kaldırıcı değerler

ρ	n	$\gamma=0.90$		$\gamma=0.70$		$\gamma=0.50$		$\gamma=0.30$	
		GEKK	ARR	GEKK	ARR	GEKK	ARR	GEKK	ARR
-0.99	20	0.2050	0.1532	0.2135	0.1685	0.2212	0.1815	0.2291	0.1931
	100	0.0231	0.0158	0.0238	0.0177	0.0245	0.0191	0.0252	0.0203
	500	0.0083	0.0052	0.0078	0.0056	0.0074	0.0057	0.0069	0.0055
-0.90	20	0.2058	0.1525	0.2146	0.1687	0.2226	0.1823	0.2306	0.1943
	100	0.0241	0.0167	0.0249	0.0193	0.0257	0.0210	0.0265	0.0224
	500	0.0084	0.0056	0.0079	0.0061	0.0075	0.0061	0.0070	0.0059
-0.70	20	0.2034	0.1468	0.2127	0.1642	0.2210	0.1785	0.229	0.1908
	100	0.0267	0.0181	0.0278	0.0213	0.0288	0.0235	0.0298	0.0251
	500	0.0086	0.0057	0.0081	0.0062	0.0076	0.0062	0.0071	0.0060
-0.50	20	0.1933	0.1343	0.2027	0.1520	0.2110	0.1663	0.2189	0.1784
	100	0.0296	0.0195	0.0310	0.0233	0.0322	0.0258	0.0335	0.0279
	500	0.0085	0.0056	0.0080	0.0060	0.0075	0.0060	0.0071	0.0059
0.50	20	0.0896	0.0728	0.0951	0.0779	0.0998	0.0818	0.1042	0.0845
	100	0.0305	0.0208	0.0317	0.0248	0.0329	0.0275	0.0342	0.0296
	500	0.0038	0.0030	0.0036	0.0031	0.0035	0.0031	0.0034	0.0031
0.70	20	0.0755	0.0646	0.0799	0.0678	0.0835	0.0703	0.0869	0.0717
	100	0.0283	0.0196	0.0292	0.0231	0.0302	0.0255	0.0314	0.0275
	500	0.0030	0.0025	0.0029	0.0026	0.0029	0.0026	0.0028	0.0026
0.90	20	0.0552	0.0472	0.0584	0.0493	0.0611	0.0508	0.0633	0.0515
	100	0.0254	0.0176	0.0262	0.0206	0.0270	0.0228	0.0280	0.0245
	500	0.0025	0.0022	0.0024	0.0023	0.0024	0.0023	0.0024	0.0023
0.99	20	0.0313	0.0240	0.0339	0.0257	0.0359	0.0269	0.0374	0.0272
	100	0.0196	0.0121	0.0204	0.0149	0.0211	0.0169	0.0221	0.0185
	500	0.0018	0.0016	0.0018	0.0017	0.0018	0.0017	0.0018	0.0017

7. MONTE CARLO SİMÜLASYON UYGULAMASI Tuğba SÖKÜT AÇAR

Çizelge 7.3. Farklı ρ , γ ve n için 1-inci gözleme ait DFFITS değerleri

ρ	n	$\gamma=0.90$		$\gamma=0.70$		$\gamma=0.50$		$\gamma=0.30$	
		GEKK	ARR	GEKK	ARR	GEKK	ARR	GEKK	ARR
-0.99	20	0.0005	0.0001	0.0010	0.0006	0.0017	0.0010	0.0025	0.0016
	100	-0.0002	0.0000	-0.0002	0.0000	-0.0002	-0.0000	-0.0002	-0.0001
	500	-0.0000	-0.0000	-0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
-0.90	20	0.0043	0.0007	0.0101	0.0051	0.0166	0.0098	0.0250	0.0153
	100	-0.0015	0.0004	-0.0016	0.0002	-0.0016	-0.0002	-0.0015	-0.0006
	500	-0.0003	-0.0002	-0.0001	0.0001	0.0000	0.0003	0.0002	0.0004
-0.70	20	0.0084	-0.0029	0.0252	0.0090	0.0443	0.0217	0.0694	0.0367
	100	-0.0061	-0.0005	-0.0064	-0.0009	-0.0062	-0.0019	-0.0055	-0.0030
	500	-0.0014	-0.0007	-0.0008	0.0002	-0.0003	0.0006	0.0000	0.0008
-0.50	20	0.0056	-0.0123	0.0305	0.0030	0.0591	0.0192	0.0978	0.0382
	100	-0.0121	-0.0025	-0.0119	-0.0027	-0.0111	-0.0040	-0.0097	-0.0054
	500	-0.0023	-0.0009	-0.0014	0.0005	-0.0007	0.0011	-0.0001	0.0013
0.50	20	-0.0089	-0.0465	0.0013	-0.0489	0.0151	-0.0520	0.0390	-0.0567
	100	-0.0307	-0.0229	-0.0265	-0.0183	-0.0216	-0.0150	-0.0151	-0.0115
	500	-0.0033	-0.0011	-0.0026	0.0001	-0.0020	0.0006	-0.0014	0.0006
0.70	20	0.0015	-0.0419	0.0093	-0.0426	0.0207	-0.0443	0.0405	-0.0471
	100	-0.0283	-0.0226	-0.0248	-0.0183	-0.0202	-0.0149	-0.0138	-0.0110
	500	-0.0029	-0.0012	-0.0024	-0.0002	-0.0019	0.0002	-0.0013	0.0003
0.90	20	0.0217	-0.0274	0.0285	-0.0266	0.0390	-0.0272	0.0577	-0.0287
	100	-0.0270	-0.0221	-0.0239	-0.0180	-0.0194	-0.0145	-0.0129	-0.0101
	500	-0.0022	-0.0009	-0.0017	0.0001	-0.0012	0.0005	-0.0007	0.0008
0.99	20	0.1067	0.0206	0.1163	0.0248	0.1329	0.0243	0.1644	0.0218
	100	-0.0351	-0.0267	-0.0324	-0.0237	-0.0283	-0.0208	-0.0220	-0.0169
	500	-0.0018	-0.0007	-0.0012	0.0005	-0.0006	0.0013	0.0002	0.0018

7. MONTE CARLO SİMÜLASYON UYGULAMASI Tuğba SÖKÜT AÇAR

Çizelge 7.4. Farklı ρ , γ ve n için 7-inci gözleme ait DFFITS değerleri

ρ	n	$\gamma=0.90$		$\gamma=0.70$		$\gamma=0.50$		$\gamma=0.30$	
		GEKK	ARR	GEKK	ARR	GEKK	ARR	GEKK	ARR
-0.99	20	14.993	15.801	15.061	16.059	15.114	16.230	15.163	16.380
	100	12.131	10.884	12.296	11.603	12.448	12.036	12.603	12.356
	500	11.496	0.91	12.145	11.593	11.792	11.494	12.384	12.179
-0.90	20	16.301	16.978	16.373	17.293	16.437	17.503	16.505	17.688
	100	11.308	10.281	11.453	11.002	11.594	11.378	11.743	11.647
	500	0.9494	0.8793	0.9258	0.9042	0.9022	0.8921	0.8749	0.8687
-0.70	20	19.678	19.930	19.752	20.423	19.840	20.754	19.954	21.045
	100	11.708	10.623	11.867	11.426	12.025	11.829	12.193	12.120
	500	0.8120	0.7516	0.7916	0.7726	0.7714	0.7625	0.7486	0.7432
-0.50	20	23.160	22.645	23.201	23.417	23.290	23.919	23.448	24.350
	100	12.540	11.302	12.716	12.214	12.899	12.677	13.098	13.018
	500	0.7633	0.7020	0.7436	0.7229	0.7244	0.7144	0.7036	0.6972
0.50	20	18.154	17.165	18.332	17.918	18.500	18.407	18.683	18.760
	100	12.568	11.622	12.674	12.350	12.812	12.703	12.990	12.986
	500	0.5062	0.4791	0.4972	0.4863	0.4902	0.4843	0.4851	0.4813
0.70	20	14.127	13.634	14.331	14.171	14.493	14.513	14.637	14.743
	100	11.829	11.018	11.920	11.649	12.045	11.957	12.214	12.216
	500	0.4777	0.4596	0.4718	0.4647	0.4679	0.4641	0.4663	0.4639
0.90	20	0.9973	0.9677	10.157	10.086	10.292	10.337	10.391	10.481
	100	11.210	10.470	11.309	11.049	11.434	11.345	11.600	11.595
	500	0.5055	0.4931	0.5022	0.4972	0.5009	0.4982	0.5020	0.5003
0.99	20	0.6718	0.6402	0.6921	0.6856	0.7063	0.7125	0.7153	0.7277
	100	0.9992	0.9294	10.120	0.9870	10.268	10.181	10.452	10.447
	500	0.5786	0.5636	0.5747	0.5674	0.5735	0.5692	0.5757	0.5728

7. MONTE CARLO SİMÜLASYON UYGULAMASI Tuğba SÖKÜT AÇAR

Çizelge 7.5. Farklı ρ , γ ve n için 1-inci gözleme ait Cook Uzaklığı

ρ	n	$\gamma=0.90$		$\gamma=0.70$		$\gamma=0.50$		$\gamma=0.30$	
		GEKK	ARR	GEKK	ARR	GEKK	ARR	GEKK	ARR
-0.99	20	0.0000	0.0000	0.0001	0.0000	0.0001	0.0000	0.0001	0.0001
	100	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	500	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
-0.90	20	0.0036	0.0025	0.0041	0.0028	0.0054	0.0035	0.0081	0.0046
	100	0.0009	0.0005	0.0007	0.0004	0.0006	0.0003	0.0004	0.0003
	500	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
-0.70	20	0.0232	0.0145	0.0274	0.0177	0.0350	0.0221	0.0496	0.0286
	100	0.0053	0.0029	0.0042	0.0024	0.0034	0.0020	0.0027	0.0017
	500	0.0009	0.0006	0.0008	0.0006	0.0007	0.0005	0.0006	0.0004
-0.50	20	0.0533	0.0305	0.0614	0.0378	0.0762	0.0466	0.1047	0.0587
	100	0.0115	0.0065	0.0093	0.0053	0.0075	0.0044	0.0060	0.0038
	500	0.0023	0.0014	0.0020	0.0013	0.0018	0.0012	0.0015	0.0010
0.50	20	0.1757	0.1353	0.1772	0.1398	0.1843	0.1478	0.2072	0.1594
	100	0.0193	0.0156	0.0174	0.0142	0.0159	0.0133	0.0148	0.0127
	500	0.0034	0.0026	0.0032	0.0025	0.0029	0.0023	0.0026	0.0022
0.70	20	0.1999	0.1867	0.1942	0.1857	0.1957	0.1888	0.2105	0.1962
	100	0.0192	0.0174	0.0180	0.0164	0.0171	0.0158	0.0164	0.0154
	500	0.0026	0.0022	0.0025	0.0022	0.0024	0.0022	0.0023	0.0021
0.90	20	0.2314	0.2249	0.2246	0.2231	0.2266	0.2263	0.2345	0.2327
	100	0.0230	0.0240	0.0240	0.0230	0.0231	0.0222	0.0221	0.0216
	500	0.0024	0.0025	0.0025	0.0025	0.0025	0.0025	0.0025	0.0025
0.99	20	10.041	0.9284	11.053	0.8822	11.156	0.8616	12.062	0.8618
	100	0.0287	0.0281	0.0294	0.0268	0.0280	0.0259	0.0264	0.0250
	500	0.0032	0.0031	0.0031	0.0031	0.0031	0.0031	0.0031	0.0031

7. MONTE CARLO SİMÜLASYON UYGULAMASI Tuğba SÖKÜT AÇAR

Çizelge 7.6. Farklı ρ , γ ve n için 7-inci gözleme ait Cook Uzaklığı

ρ	n	$\gamma=0.90$		$\gamma=0.70$		$\gamma=0.50$		$\gamma=0.30$	
		GEKK	ARR	GEKK	ARR	GEKK	ARR	GEKK	ARR
-0.99	20	304.27	211.04	284.91	212.06	274.40	216.17	273.06	225.09
	100	91.136	51.948	92.436	55.335	93.917	58.452	95.630	61.285
	500	26.628	12.692	25.359	13.710	24.101	14.191	22.677	14.060
-0.90	20	159.36	106.81	146.83	106.85	139.69	108.46	137.33	112.17
	100	19.853	11.506	20.084	12.907	20.412	14.065	20.838	15.025
	500	0.4399	0.2300	0.4197	0.2615	0.3997	0.2736	0.3771	0.2724
-0.70	20	91.360	56.296	83.530	56.864	79.155	58.232	77.718	60.618
	100	10.109	0.5654	10.252	0.6548	10.439	0.7249	10.663	0.7797
	500	0.2289	0.1190	0.2180	0.1350	0.2074	0.1413	0.1958	0.1412
-0.50	20	75.250	42.568	69.364	43.818	66.156	45.518	65.066	47.677
	100	0.8290	0.4420	0.8430	0.5209	0.8611	0.5836	0.8829	0.6333
	500	0.1794	0.0915	0.1705	0.1025	0.1621	0.1073	0.1531	0.1076
0.50	20	62.097	46.689	58.623	43.578	56.875	41.885	56.432	40.641
	100	0.7992	0.4370	0.8015	0.5225	0.8115	0.5870	0.8286	0.6386
	500	0.0788	0.0540	0.0761	0.0572	0.0741	0.0591	0.0727	0.0605
0.70	20	68.776	54.888	64.821	50.438	62.629	47.711	61.924	45.676
	100	0.9539	0.5349	0.9522	0.6342	0.9613	0.7088	0.9809	0.7697
	500	0.0788	0.0607	0.0770	0.0636	0.0759	0.0655	0.0755	0.0671
0.90	20	68.496	53.504	64.693	48.534	62.737	45.648	62.031	43.356
	100	15.780	0.8734	15.742	10.284	15.887	11.501	16.212	12.530
	500	0.1210	0.1025	0.1197	0.1061	0.1194	0.1088	0.1202	0.1116
0.99	20	40.046	25.394	38.399	22.674	37.501	21.181	37.065	19.687
	100	22.263	0.9718	22.386	12.311	22.744	14.431	23.369	16.271
	500	0.3502	0.2949	0.3465	0.3027	0.3461	0.3102	0.3500	0.3191

7. MONTE CARLO SİMÜLASYON UYGULAMASI Tuğba SÖKÜT AÇAR

Çizelge 7.7. Farklı ρ , γ ve n için $DFBETA_{0,1}$ ve $DFBETA_{0,1}^R$ değerleri

ρ	n	$\gamma=0.90$		$\gamma=0.70$		$\gamma=0.50$		$\gamma=0.30$	
		GEKK	ARR	GEKK	ARR	GEKK	ARR	GEKK	ARR
-0.99	20	0.0003	0.0001	0.0007	0.0004	0.0011	0.0007	0.0015	0.0010
	100	-0.0001	0.0000	-0.0001	-0.0000	-0.0001	-0.0000	-0.0001	-0.0000
	500	-0.0000	-0.0000	-0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
-0.90	20	0.0026	0.0004	0.0063	0.0033	0.0104	0.0063	0.0151	0.0095
	100	-0.0005	0.0001	-0.0007	0.0000	-0.0008	-0.0001	-0.0008	-0.0004
	500	-0.0001	-0.0001	-0.0000	0.0000	0.0000	0.0001	0.0001	0.0001
-0.70	20	0.0050	-0.0020	0.0158	0.0059	0.0277	0.0142	0.0420	0.0229
	100	-0.0022	-0.0001	-0.0026	-0.0004	-0.0029	-0.0010	-0.0029	-0.0017
	500	-0.0003	-0.0002	-0.0002	0.0001	-0.0001	0.0002	0.0000	0.0003
-0.50	20	0.0034	-0.0088	0.0196	0.0018	0.0379	0.0131	0.0604	0.0248
	100	-0.0044	-0.0010	-0.0050	-0.0012	-0.0054	-0.0020	-0.0053	-0.0032
	500	-0.0006	-0.0003	-0.0004	0.0001	-0.0002	0.0003	-0.0000	0.0004
0.50	20	-0.0119	-0.1620	0.0021	-0.1539	0.0268	-0.1509	0.0699	-0.1513
	100	-0.0282	-0.0265	-0.0290	-0.0238	-0.0262	-0.0209	-0.0197	-0.0168
	500	-0.0022	-0.0009	-0.0020	0.0000	-0.0017	0.0005	-0.0012	0.0005
0.70	20	0.0028	-0.2538	0.0231	-0.2393	0.0581	-0.2330	0.1196	-0.2325
	100	-0.0364	-0.0390	-0.0394	-0.0363	-0.0361	-0.0317	-0.0267	-0.0245
	500	-0.0028	-0.0015	-0.0026	-0.0004	-0.0023	0.0002	-0.0017	0.0004
0.90	20	0.0776	-0.4836	0.1523	-0.4467	0.2488	-0.4281	0.4008	-0.4230
	100	-0.0633	-0.0847	-0.0786	-0.0862	-0.0757	-0.0760	-0.0557	-0.0570
	500	-0.0036	-0.0026	-0.0036	-0.0005	-0.0029	0.0009	-0.0017	0.0017
0.99	20	13.196	0.8251	23.465	10.329	32.910	0.9362	44.947	0.7157
	100	-0.3513	-0.5272	-0.5284	-0.6217	-0.5736	-0.6059	-0.5069	-0.5228
	500	-0.0125	-0.0186	-0.0133	-0.0049	-0.0078	0.0093	0.0028	0.0208

7. MONTE CARLO SİMÜLASYON UYGULAMASI Tuğba SÖKÜT AÇAR

Çizelge 7.8. Farklı ρ , γ ve n için $DFBETA_{0,7}$ ve $DFBETA_{0,7}^R$ değerleri

ρ	n	$\gamma=0.90$		$\gamma=0.70$		$\gamma=0.50$		$\gamma=0.30$	
		GEKK	ARR	GEKK	ARR	GEKK	ARR	GEKK	ARR
-0.99	20	0.7793	0.9488	0.8367	10.203	0.8599	10.389	0.8626	10.348
	100	0.7649	0.8167	0.8074	0.9434	0.8322	0.9496	0.8462	0.9489
	500	0.6890	0.7281	0.7060	0.7358	0.7167	0.7416	0.7241	0.7467
-0.90	20	0.8517	10.303	0.9145	11.062	0.9407	11.259	0.9452	11.226
	100	0.7896	0.9034	0.8267	0.9178	0.8483	0.9237	0.8605	0.9254
	500	0.4812	0.5010	0.4940	0.5085	0.5023	0.5142	0.5084	0.5188
-0.70	20	10.591	12.694	11.391	13.635	11.750	13.893	11.860	13.898
	100	0.7779	0.8805	0.8134	0.8940	0.8338	0.9004	0.8450	0.9026
	500	0.4076	0.4245	0.4188	0.4314	0.4262	0.4366	0.4318	0.4408
-0.50	20	13.248	15.962	14.301	17.140	14.805	17.463	15.019	17.509
	100	0.7935	0.8982	0.8297	0.9128	0.8507	0.9198	0.8626	0.9228
	500	0.3848	0.4025	0.3959	0.4094	0.4032	0.4144	0.4088	0.4185
0.50	20	20.900	45.382	25.759	43.813	28.533	42.611	30.088	41.589
	100	0.8833	11.088	0.9948	11.770	10.633	12.132	11.102	12.380
	500	0.4089	0.4581	0.4397	0.4777	0.4595	0.4909	0.4740	0.5011
0.70	20	21.011	55.824	27.947	54.041	31.973	52.387	34.244	50.867
	100	0.9840	13.427	11.839	14.748	13.075	15.455	13.933	15.948
	500	0.4665	0.5514	0.5238	0.5887	0.5603	0.6133	0.5864	0.6320
0.90	20	25.254	88.861	37.637	86.030	44.903	82.909	49.008	79.823
	100	15.207	25.073	21.225	29.545	24.966	31.814	27.559	33.331
	500	0.7450	10.381	0.9559	11.833	10.889	12.759	11.823	13.436
0.99	20	32.185	136.182	52.300	131.060	64.110	125.661	70.781	120.404
	100	40.914	81.536	66.360	102.432	82.206	112.293	93.175	118.565
	500	36.045	66.940	57.611	84.417	71.132	94.152	80.514	100.965

7. MONTE CARLO SİMÜLASYON UYGULAMASI Tuğba SÖKÜT AÇAR

Çizelge 7.9. Farklı ρ , γ ve n için $DFBETA_{1,1}$ ve $DFBETA_{1,1}^R$ değerleri

ρ	n	$\gamma=0.90$		$\gamma=0.70$		$\gamma=0.50$		$\gamma=0.30$	
		GEKK	ARR	GEKK	ARR	GEKK	ARR	GEKK	ARR
-0.99	20	-0.0071	-0.0016	-0.0110	-0.0054	-0.0160	-0.0090	-0.0232	-0.0136
	100	0.0039	-0.0002	0.0025	-0.0002	0.0020	0.0000	0.0019	0.0004
	500	0.0015	-0.0002	0.0002	-0.0005	-0.0002	-0.0006	-0.0005	-0.0007
-0.90	20	-0.0654	-0.0098	-0.1061	-0.0483	-0.1560	-0.0843	-0.2288	-0.1289
	100	0.0330	-0.0038	0.0213	-0.0029	0.0177	-0.0000	0.0166	0.0040
	500	0.0135	-0.0020	0.0029	-0.0041	-0.0005	-0.0051	-0.0028	-0.0058
-0.70	20	-0.1248	0.0297	-0.2590	-0.0836	-0.4070	-0.1818	-0.6206	-0.2998
	100	0.1335	0.0101	0.0812	0.0076	0.0658	0.0137	0.0607	0.0251
	500	0.0528	0.0056	0.0171	-0.0046	0.0059	-0.0084	-0.0006	-0.0105
-0.50	20	-0.0793	0.1260	-0.2983	-0.0295	-0.5162	-0.1519	-0.8318	-0.2947
	100	0.2516	0.0310	0.1470	0.0212	0.1160	0.0285	0.1041	0.0453
	500	0.0841	0.0104	0.0289	-0.0072	0.0117	-0.0127	0.0020	-0.0151
0.50	20	0.0325	0.1034	-0.0037	0.1142	-0.0419	0.1281	-0.1156	0.1503
	100	0.1824	0.0654	0.1056	0.0463	0.0780	0.0401	0.0566	0.0352
	500	0.0339	0.0026	0.0152	-0.0011	0.0093	-0.0022	0.0062	-0.0024
0.70	20	-0.0025	0.0458	-0.0131	0.0517	-0.0293	0.0585	-0.0639	0.0691
	100	0.0721	0.0301	0.0444	0.0220	0.0335	0.0189	0.0239	0.0158
	500	0.0128	0.0015	0.0060	0.0001	0.0038	-0.0005	0.0026	-0.0007
0.90	20	-0.0084	0.0079	-0.0103	0.0090	-0.0149	0.0104	-0.0255	0.0125
	100	0.0120	0.0060	0.0085	0.0045	0.0067	0.0039	0.0048	0.0032
	500	0.0015	0.0002	0.0008	-0.0001	0.0004	-0.0002	0.0002	-0.0003
0.99	20	-0.0055	-0.0010	-0.0060	-0.0013	-0.0074	-0.0014	-0.0106	-0.0014
	100	0.0009	0.0005	0.0008	0.0004	0.0007	0.0004	0.0006	0.0004
	500	0.0001	0.0000	0.0000	-0.0000	0.0000	-0.0000	-0.0000	-0.0000

7. MONTE CARLO SİMÜLASYON UYGULAMASI Tuğba SÖKÜT AÇAR

Çizelge 7.10. Farklı ρ , γ ve n için $DFBETA_{1,7}$ ve $DFBETA_{1,7}^R$ değerleri

ρ	n	$\gamma=0.90$		$\gamma=0.70$		$\gamma=0.50$		$\gamma=0.30$	
		GEKK	ARR	GEKK	ARR	GEKK	ARR	GEKK	ARR
-0.99	20	156.35	124.96	111.25	97.38	101.40	92.82	100.17	93.80
	100	68.34	57.10	54.00	53.04	54.82	54.16	60.67	57.39
	500	796.28	357.06	493.53	309.77	396.08	279.87	334.69	248.01
-0.90	20	168.71	132.32	120.32	104.15	109.88	99.75	108.73	101.10
	100	51.71	51.09	43.45	47.52	45.96	48.76	52.52	52.52
	500	559.69	289.84	347.25	244.52	278.87	217.31	235.73	191.67
-0.70	20	195.35	145.78	140.32	117.65	128.84	114.07	128.17	116.50
	100	25.05	41.52	28.64	39.87	35.26	42.33	44.46	47.37
	500	471.06	244.85	292.06	204.85	234.21	181.75	197.43	160.02
-0.50	20	209.47	146.19	152.22	122.73	140.98	121.12	141.49	125.07
	100	-0.26	30.42	13.38	31.54	24.27	35.45	36.13	41.70
	500	422.60	215.13	261.92	179.41	209.67	159.32	176.15	140.06
0.50	20	12.96	24.75	19.65	27.00	26.31	29.91	34.60	33.77
	100	-58.79	-21.57	-31.43	-14.24	-20.32	-0.96	-12.11	-0.517
	500	68.75	39.92	43.84	32.29	35.04	27.96	28.61	23.62
0.70	20	-0.01	12.97	0.66	14.66	12.15	16.74	18.44	19.47
	100	-43.51	-20.05	-25.99	-14.77	-18.55	-11.39	-12.87	-0.804
	500	31.08	19.73	20.44	16.00	16.45	13.72	13.24	11.33
0.90	20	-0.38	0.58	0.07	0.70	0.45	0.83	0.87	10.14
	100	-25.69	-15.40	-18.32	-12.49	-14.48	-10.37	-11.09	-0.81
	500	10.18	0.77	0.74	0.64	0.62	0.55	0.49	0.43
0.99	20	-0.40	0.32	-0.08	0.41	0.20	0.51	0.52	0.65
	100	-17.46	-12.8	-14.62	-11.02	-12.34	-0.94	-0.99	-0.77
	500	0.51	0.49	0.45	0.43	0.39	0.37	0.31	0.28

7. MONTE CARLO SİMÜLASYON UYGULAMASI Tuğba SÖKÜT AÇAR

Çizelge 7.11. Farklı ρ , γ ve n için $DFBETA_{2,1}$ ve $DFBETA_{2,1}^R$ değerleri

ρ	n	$\gamma=0.90$		$\gamma=0.70$		$\gamma=0.50$		$\gamma=0.30$	
		GEKK	ARR	GEKK	ARR	GEKK	ARR	GEKK	ARR
-0.99	20	0.0020	0.0005	0.0015	0.0009	-0.0001	0.0002	-0.0035	-0.0017
	100	-0.0024	0.0002	-0.0020	0.0001	-0.0017	-0.0002	-0.0012	-0.0004
	500	-0.0021	-0.0003	-0.0004	0.0007	0.0003	0.0012	0.0008	0.0013
-0.90	20	0.0184	0.0030	0.0144	0.0079	-0.0005	0.0017	-0.0338	-0.0161
	100	-0.0203	0.0040	-0.0170	0.0020	-0.0149	-0.0012	-0.0109	-0.0036
	500	-0.0186	-0.0010	-0.0049	0.0064	0.0010	0.0097	0.0050	0.0107
-0.70	20	0.0374	-0.0110	0.0397	0.0151	0.0052	0.0064	-0.0837	-0.0334
	100	-0.0825	-0.0068	-0.0647	-0.0085	-0.0549	-0.0151	-0.0398	-0.0193
	500	-0.0730	-0.0168	-0.0290	0.0075	-0.0110	0.0175	0.0012	0.0201
-0.50	20	0.0267	-0.0480	0.0564	0.0061	0.0228	0.0094	-0.0914	-0.0258
	100	-0.1568	-0.0280	-0.1169	-0.0240	-0.0956	-0.0305	-0.0671	-0.0339
	500	-0.1177	-0.0253	-0.0497	0.0139	-0.0222	0.0280	-0.0037	0.0299
0.50	20	-0.0236	-0.0457	0.0021	-0.0370	0.0161	-0.0269	0.0178	-0.0115
	100	-0.1203	-0.0716	-0.0791	-0.0495	-0.0550	-0.0356	-0.0300	-0.0214
	500	-0.0539	-0.0086	-0.0300	0.0017	-0.0200	0.0051	-0.0124	0.0051
0.70	20	0.0021	-0.0216	0.0087	-0.0182	0.0136	-0.0141	0.0138	-0.0076
	100	-0.0472	-0.0322	-0.0327	-0.0230	-0.0230	-0.0164	-0.0124	-0.0095
	500	-0.0207	-0.0045	-0.0121	-0.0006	-0.0083	0.0009	-0.0052	0.0013
0.90	20	0.0088	-0.0047	0.0092	-0.0040	0.0100	-0.0034	0.0097	-0.0026
	100	-0.0075	-0.0063	-0.0060	-0.0048	-0.0044	-0.0035	-0.0024	-0.0019
	500	-0.0024	-0.0006	-0.0015	0.0001	-0.0010	0.0005	-0.0005	0.0006
0.99	20	0.0089	0.0013	0.0087	0.0014	0.0087	0.0013	0.0084	0.0010
	100	-0.0003	-0.0004	-0.0004	-0.0003	-0.0003	-0.0003	-0.0002	-0.0002
	500	-0.0001	-0.0000	-0.0001	0.0000	-0.0000	0.0001	0.0000	0.0001

7. MONTE CARLO SİMÜLASYON UYGULAMASI Tuğba SÖKÜT AÇAR

Çizelge 7.12. Farklı ρ , γ ve n için $DFBETA_{2,7}$ ve $DFBETA_{2,7}^R$ değerleri

ρ	n	$\gamma=0.90$		$\gamma=0.70$		$\gamma=0.50$		$\gamma=0.30$	
		GEKK	ARR	GEKK	ARR	GEKK	ARR	GEKK	ARR
-0.99	20	140.18	100.34	99.12	80.08	89.63	77.63	87.80	76.44
	100	-288.09	-97.42	-159.12	-71.66	-114.24	-60.52	-85.48	-51.57
	500	164.43	90.47	120.27	83.43	102.28	75.92	83.39	62.09
-0.90	20	153.17	107.49	108.47	86.68	98.23	84.46	96.37	76.67
	100	-269.48	-103.40	-148.06	-78.19	-105.71	-64.97	-78.50	-53.62
	500	118.62	69.99	86.29	64.47	73.18	58.60	59.53	48.34
-0.70	20	184.65	122.34	131.28	101.14	119.27	99.75	117.40	93.13
	100	-276.13	-110.16	-150.41	-82.27	-106.22	-67.08	-77.56	-53.93
	500	106.78	62.20	76.49	56.49	64.27	50.97	51.85	41.77
-0.50	20	211.75	128.68	151.24	110.53	137.85	110.90	136.17	110.81
	100	-282.70	-112.97	-152.91	-82.81	-106.94	-66.56	-76.82	-52.47
	500	104.17	58.77	73.48	52.65	61.16	47.20	48.95	38.39
0.50	20	71.25	44.24	59.06	46.46	58.05	49.08	50.08	51.97
	100	-75.63	-39.62	-42.89	-27.49	-29.65	-20.43	-19.59	-13.83
	500	35.22	19.23	23.37	16.23	18.63	13.98	14.34	10.99
0.70	20	38.19	29.90	35.46	31.29	35.15	32.83	27.09	34.49
	100	-39.30	-23.20	-23.60	-16.27	-16.52	-11.88	-10.59	-0.756
	500	19.81	11.56	13.30	0.971	10.59	0.828	0.808	0.642
0.90	20	17.99	19.38	18.39	20.14	17.70	20.91	11.27	21.72
	100	-16.77	-12.02	-11.53	-0.869	-0.835	-0.621	-0.514	-0.356
	500	0.863	0.594	0.639	0.513	0.526	0.440	0.403	0.337
0.99	20	14.38	15.21	13.76	15.68	12.08	16.14	13.39	16.62
	100	-0.973	-0.829	-0.768	-0.612	-0.574	-0.429	-0.343	-0.222
	500	0.508	0.449	0.460	0.411	0.405	0.357	0.318	0.273

7. MONTE CARLO SİMÜLASYON UYGULAMASI Tuğba SÖKÜT AÇAR

Çizelge 7.13. Farklı ρ , γ ve n için $DFBETA_{3,1}$ ve $DFBETA_{3,1}^R$ değerleri

ρ	n	$\gamma=0.90$		$\gamma=0.70$		$\gamma=0.50$		$\gamma=0.30$	
		GEKK	ARR	GEKK	ARR	GEKK	ARR	GEKK	ARR
-0.99	20	0.0039	0.0008	0.0048	0.0020	0.0048	0.0022	0.0035	0.0015
	100	-0.0036	0.0002	-0.0028	0.0002	-0.0025	-0.0001	-0.0020	-0.0004
	500	-0.0003	-0.0002	-0.0001	0.0002	0.0001	0.0004	0.0003	0.0005
-0.90	20	0.0356	0.0051	0.0454	0.0176	0.0464	0.0205	0.0335	0.0138
	100	-0.0306	0.0043	-0.0245	0.0029	-0.0217	-0.0007	-0.0174	-0.0041
	500	-0.0031	-0.0014	-0.0013	0.0015	0.0003	0.0031	0.0017	0.0037
-0.70	20	0.0659	-0.0115	0.1067	0.0294	0.1148	0.0409	0.0816	0.0270
	100	-0.1235	-0.0083	-0.0931	-0.0100	-0.0797	-0.0183	-0.0629	-0.0257
	500	-0.0117	-0.0055	-0.0075	0.0019	-0.0035	0.0058	0.0004	0.0069
-0.50	20	0.0394	-0.0492	0.1139	0.0109	0.1302	0.0295	0.0847	0.0173
	100	-0.2297	-0.0313	-0.1651	-0.0276	-0.1367	-0.0367	-0.1045	-0.0451
	500	-0.0176	-0.0074	-0.0123	0.0038	-0.0067	0.0093	-0.0012	0.0101
0.50	20	-0.0060	-0.0172	0.0003	-0.0061	-0.0013	0.0065	-0.0208	0.0249
	100	-0.1412	-0.0696	-0.0926	-0.0506	-0.0669	-0.0387	-0.0405	-0.0260
	500	-0.0019	-0.0026	-0.0037	0.0000	-0.0034	0.0010	-0.0023	0.0011
0.70	20	0.0003	-0.0056	0.0003	-0.0005	-0.0026	0.0053	-0.0148	0.0139
	100	-0.0553	-0.0316	-0.0383	-0.0236	-0.0281	-0.0179	-0.0168	-0.0115
	500	-0.0006	-0.0013	-0.0014	-0.0003	-0.0013	0.0001	-0.0009	0.0002
0.90	20	-0.0008	-0.0000	-0.0017	0.0010	-0.0038	0.0020	-0.0092	0.0036
	100	-0.0095	-0.0066	-0.0076	-0.0053	-0.0058	-0.0040	-0.0035	-0.0024
	500	-0.0001	-0.0002	-0.0002	-0.0001	-0.0002	0.0000	-0.0001	0.0001
0.99	20	-0.0033	-0.0003	-0.0037	-0.0005	-0.0048	-0.0007	-0.0072	-0.0009
	100	-0.0010	-0.0008	-0.0010	-0.0007	-0.0008	-0.0006	-0.0006	-0.0005
	500	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	0.0000	0.0000	0.0000

7. MONTE CARLO SİMÜLASYON UYGULAMASI Tuğba SÖKÜT AÇAR

Çizelge 7.14. Farklı ρ , γ ve n için $DFBETA_{3,7}$ ve $DFBETA_{3,7}^R$ değerleri

ρ	n	$\gamma=0.90$		$\gamma=0.70$		$\gamma=0.50$		$\gamma=0.30$	
		GEKK	ARR	GEKK	ARR	GEKK	ARR	GEKK	ARR
-0.99	20	-263.32	-192.55	-161.16	-128.89	-124.07	-105.18	-98.46	-87.64
	100	265.66	102.99	179.95	94.72	153.94	95.43	143.12	97.81
	500	-772.13	-257.10	-427.93	-215.67	-321.91	-194.47	-271.40	-182.93
-0.90	20	-284.34	-202.82	-173.60	-136.52	-133.39	-111.65	-105.66	-93.08
	100	271.83	111.54	176.85	103.98	150.94	103.84	140.09	105.22
	500	-546.32	-226.48	-303.67	-182.98	-229.13	-160.13	-193.85	-147.76
-0.70	20	-331.18	-219.76	-200.98	-149.52	-153.58	-122.85	-120.90	-102.53
	100	310.50	132.31	201.11	122.33	171.03	121.15	158.21	121.87
	500	-467.50	-195.70	-260.82	-156.95	-197.58	-137.89	-167.93	-128.00
-0.50	20	-360.62	-213.40	-217.13	-148.01	-164.48	-122.57	-128.02	-102.51
	100	350.00	151.34	225.87	138.26	191.56	136.29	176.81	136.72
	500	-427.89	-174.33	-239.71	-139.89	-182.33	-124.11	-155.64	-116.34
0.50	20	-48.59	-0.256	-26.08	-0.153	-14.34	0.065	-0.330	0.405
	100	154.38	88.28	104.24	77.82	90.04	74.86	83.82	73.83
	500	-87.63	-42.22	-52.56	-34.39	-42.02	-31.38	-37.28	-30.17
0.70	20	-16.43	0.55	-0.71	0.60	-0.10	0.73	0.52	0.93
	100	93.56	59.82	67.18	53.97	59.74	52.35	56.56	51.90
	500	-43.91	-23.68	-27.76	-19.78	-22.93	-18.30	-20.86	-17.79
0.90	20	-0.24	0.84	0.11	0.86	0.41	0.93	0.75	10.35
	100	47.33	36.82	39.29	35.21	37.11	34.96	36.30	35.13
	500	-16.80	-11.13	-12.29	-0.996	-10.96	-0.959	-10.49	-0.96
0.99	20	0.060	0.869	0.292	0.879	0.508	0.919	0.745	0.987
	100	30.20	28.25	29.12	28.43	28.93	28.75	29.02	29.16
	500	-0.92	-0.79	-0.84	-0.77	-0.82	-0.78	-0.83	-0.80

Çizelge 7.1'den görüldüğü üzere $\lim_{\rho \rightarrow 1} h_{1,1}^R = 1$ ve $\lim_{\rho \rightarrow -1} h_{1,1}^R = 0$ dir. Küçük örneklem için 7-inci gözleme ait kaldırıcı değer $\rho \rightarrow -1$ iken artmakta, $\rho \rightarrow 1$ iken azalmaktadır. GEKK ve ARR tahmin edicileri için örneklem genişliğinin artması ile gözlemlerin kaldırıcı değerleri azalmaktadır (Çizelge 7.2). Tüm ρ , γ ve n değerleri için ARR tahmin edicisine ait kaldırıcı değerler ilgili

gözlemin GEKK kaldırıcı değerinden daha küçüktür. Düşük, orta ya da şiddetli çoklu iç ilişkinin kaldırıcı değerleri çok etkilememesine rağmen küçük örneklerde çoklu iç ilişkinin şiddeti azaldıkça 1-inci ve 7-inci gözlem için kaldırıcı değerler artmaktadır.

Çizelge 7.3'den elde edilen sonuçlar şöyledir: İlk gözleme ait DFFITS değerleri $\rho \cong 1$ 'de belirgin bir artış göstermiştir. Aynı zamanda, $\rho \cong 1$ 'de $DFFITS_1$ değerleri, $DFFITS_1^R$ değerlerinden önemli ölçüde büyüktür. Çizelge 7.4'den görüldüğü üzere, otokorelasyonun pozitif olduğu noktalarda her zaman $|DFFITS_7| > |DFFITS_7^R|$ 'dir. Çoklu iç ilişkinin şiddeti azaldığında $|DFFITS_i^R|$ ve $|DFFITS_i|$ ($i = 1$ ve 7) değerleri önemli ölçüde artmaktadır. Örneklem genişliği $|DFFITS_i^R|$ ve $|DFFITS_i|$ değerleri üzerinde önemli bir etki yaratmamıştır.

Çizelge 7.5 ve Çizelge 7.6'dan görüldüğü gibi, $i = 1$ ve $i = 7$ için $D_i^R \leq D_i$ tüm ρ , γ ve n için geçerlidir. GEKK ve ARR tahmin edicileri için ilk dönüştürülmüş gözlemin Cook uzaklığı $\rho \rightarrow 1$ iken artmakta, $\rho \rightarrow -1$ iken azalmaktadır (Çizelge 7.5). D_7 ve D_7^R değerleri $\rho \rightarrow -1$ iken artmakta, $\rho \rightarrow 1$ iken azalmaktadır (Çizelge 7.6). D_i ve D_i^R değerleri artan örneklem genişliği ile azalmaktadır.

Negatif otokorelasyon değerlerinde $|DFBETA_{0,1}^R|$ genellikle $|DFBETA_{0,1}|$ 'den daha küçüktür (Çizelge 7.7). Çizelge 7.8'de GEKK ve ARR tahmin edicileri kullanıldığında 7-inci gözlemin sabit terim üzerindeki etkisi verilmiştir. Tüm γ ve n değerleri için $\rho \rightarrow 1$ iken $|DFBETA_{0,7}^R|$ ve $|DFBETA_{0,7}|$ değerleri artmaktadır. Değerlendirilen tüm ρ , γ ve n değerleri için $|DFBETA_{0,7}^R| > |DFBETA_{0,7}|$ geçerlidir. 1-inci ve 7-inci gözlemlerin GEKK ve ARR yöntemi ile tahmin edilen sabit terim üzerindeki etkileri genellikle artan örneklem genişliği ile azalmaktadır. Bu azalış 7-inci gözleme göre 1-inci gözlemde daha net gözlenmektedir.

Çizelge 7.10'dan görüldüğü üzere, negatif otokorelasyon değerlerinde $|DFBETAS_{1,7}|$ ve $|DFBETAS_{1,7}^R|$ oldukça yüksek iken pozitif otokorelasyonda daha küçük değerler almışlardır. GEKK ve ARR ile tahmin edilen β_1 üzerinde otokorelasyon, çoklu iç ilişki ve örneklem genişliği belirli bir etki göstermemiştir. Çizelge 7.12, Çizelge 7.13 ve Çizelge 7.14 göstermiştir ki, $|DFBETA_{2,7}|$, $|DFBETA_{3,7}|$, $|DFBETA_{2,7}^R|$ ve $|DFBETA_{3,7}^R|$ değerleri negatif otokorelasyon bölgesinde çoklu iç ilişki şiddeti azaldıkça azalmaktadır ve pozitif otokorelasyon bölgesinde belirgin bir yapı yoktur. Ayrıca, $|DFBETA_{2,7}|$, $|DFBETA_{3,7}|$, $|DFBETA_{2,7}^R|$ ve $|DFBETA_{3,7}^R|$ değerleri $|\rho \rightarrow 1|$ iken azalmaktadırlar.

7.2. Güven Elipsoidine Dayalı Ölçümler için Monte Carlo Simülasyon

Farklı otokorelasyon katsayısı, ridge yanlılık parametresi ve σ_u altında COVRATIO ve AP ölçümlerinin davranışlarını belirlemek adına Monte Carlo simülasyon çalışması uygulanmıştır. Güven elipsoidine dayalı ölçümlerden CW istatistiğine COVRATIO ile benzer sonuçlar verdiğinden dolayı simülasyon çalışmasında yer verilmemiştir. Simülasyon adımları aşağıda verilmiştir.

- Ø Monte Carlo simülasyon döngüsüne başlamadan önce, $n = 50$, $r = 3$, $p = 4$ olarak seçilmiştir.
- Ø İlişkili açıklayıcı değişkenleri üretmek için McDonald ve Galarneau (1975) tarafından verilen

$$x_{ij} = \begin{cases} 1, & j = 1; i = 1, 2, \dots, n \\ (1 - \gamma^2)^{1/2} d_{ij} + \gamma d_{i(r+1)}, & j = 2, \dots, r; i = 1, \dots, n \end{cases} \quad (7.3)$$

eşitliği kullanılmıştır. Burada d_{ij} , normal rastgele sayıları, γ^2 herhangi iki açıklayıcı değişken arasındaki ilişkiyi göstermektedir. $\gamma^2 = 0.99$ olarak seçilmiştir. x_{ij} 'ler üretildikten sonra X matrisinin son satırı

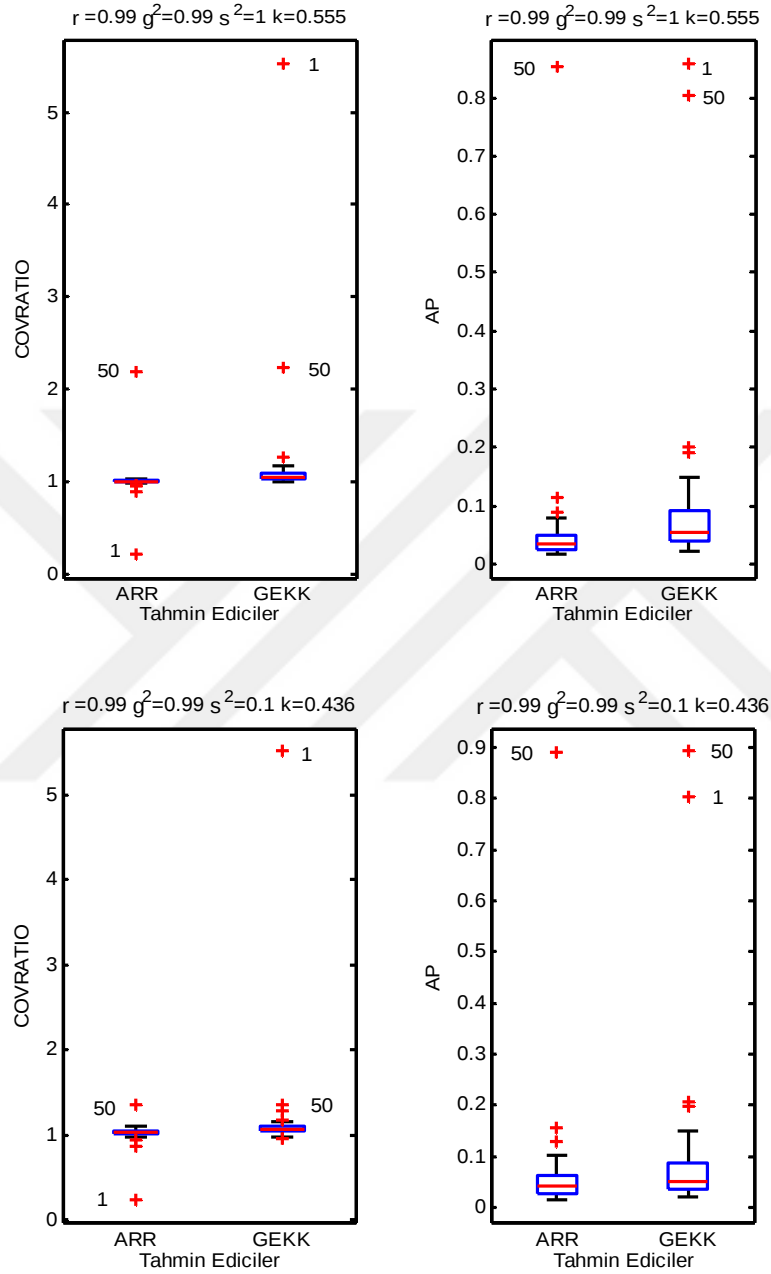
$x_{n,j} = 20 - 0.05(j - 1), j = 1, \dots, r$ olarak değiştirilmiştir. Böylece yüksek kaldıraç gözlem elde edilmiştir (Hadi ve Simonoff, 1993).

- Ø β parametre vektör $\beta = \frac{1}{\sqrt{p}} \mathbf{1}_{p \times 1}$ olarak alınmıştır. Burada $\mathbf{1}_{p \times 1}$; 1'lerden oluşan vektördür (Holgersson, 2015).
- Ø Simülasyon döngüsü normal dağılımdan ak gürültü hataların üretilmesi ile başlamıştır, öyle ki $u \sim N(0, \sigma_u^2)$, $\sigma_u^2 = 0.01, 0.1$ ve 1 olarak alınmıştır.
- Ø Çalışmada otokorelasyon katsayısı $|\rho| = 0.99, 0.90, 0.70, 0.50$ olmak üzere genel lineer modelin hata terimleri (3.2) eşitliğindeki modelden üretilmiştir.
- Ø Yanıt değişkene ait değerler aşağıdaki modelden elde edilmiştir

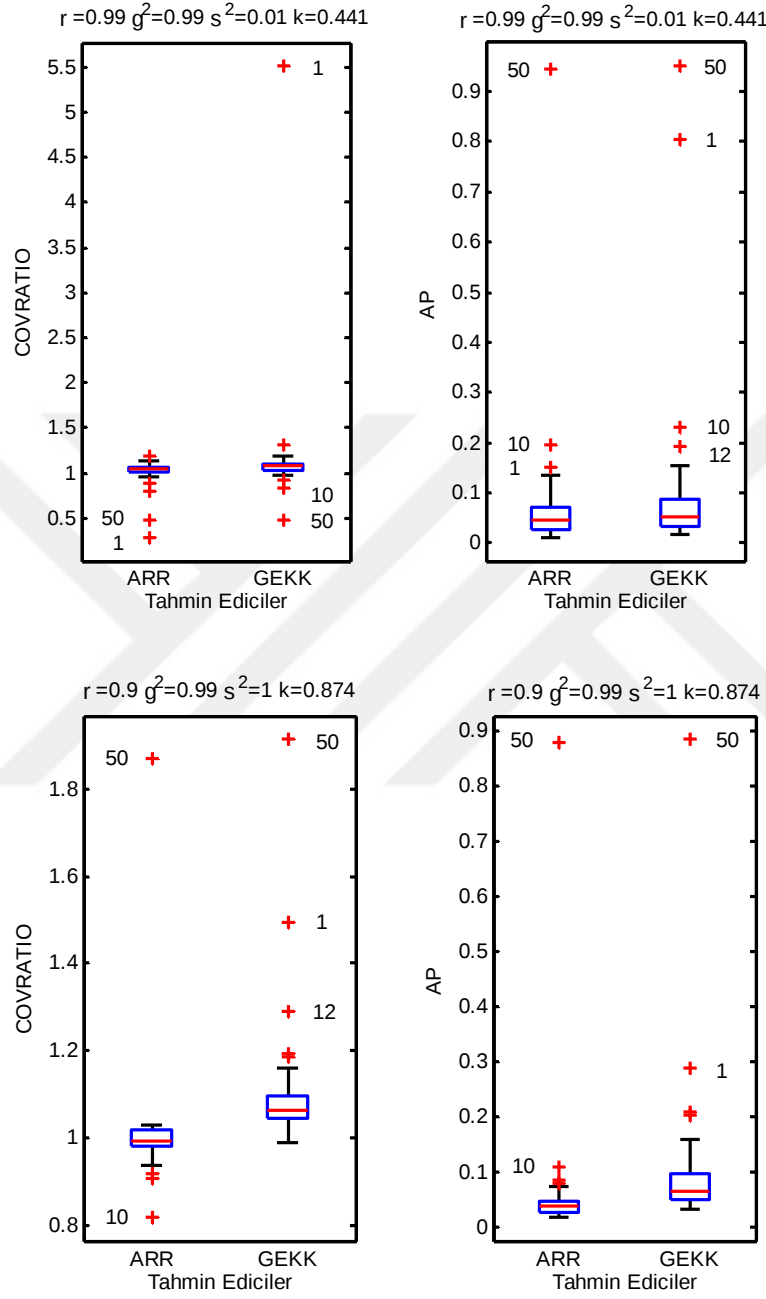
$$y_i = \beta_0 + \sum_{j=1}^r \beta_j x_{ij} + \varepsilon_i.$$
- Ø Son gözlem ait yanıt değeri

$$y_n = \beta_0 + \sum_{j=1}^r \beta_j x_{ij} + 10\varepsilon_n$$

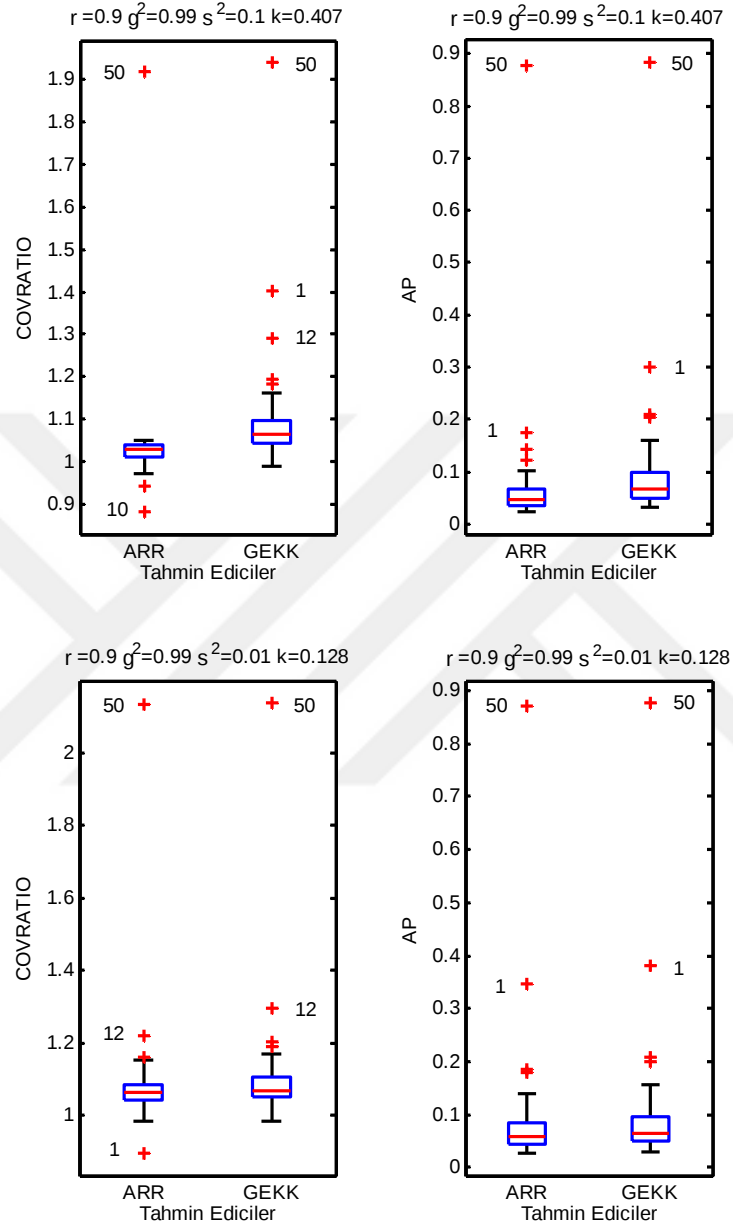
olarak değiştirilmiş ve böylece etkin bir gözlem elde edilmiştir.
- Ø Dönüşüm matrisi olarak adlandırılan P , her bir otokorelasyon değeri için elde edilerek, dönüştürülmüş yanıt vektörü, dönüştürülmüş açıklayıcı değişken matrisi elde edilmiştir.
- Ø Ridge yanlılık parametresi k , $k_1 = \frac{ps^2}{\hat{\beta}'\hat{\beta}}$ ile hesaplanmıştır.
- Ø Deneme yeni hata terimleri oluşturularak 1000 kez tekrarlanmıştır. Her bir deneme için COVRATIO ve AP değerleri elde edilmiş, toplam deneme sayısı olan MCN'ye göre ortalaması alınmış ve kutu grafikleri çizdirilmiştir.



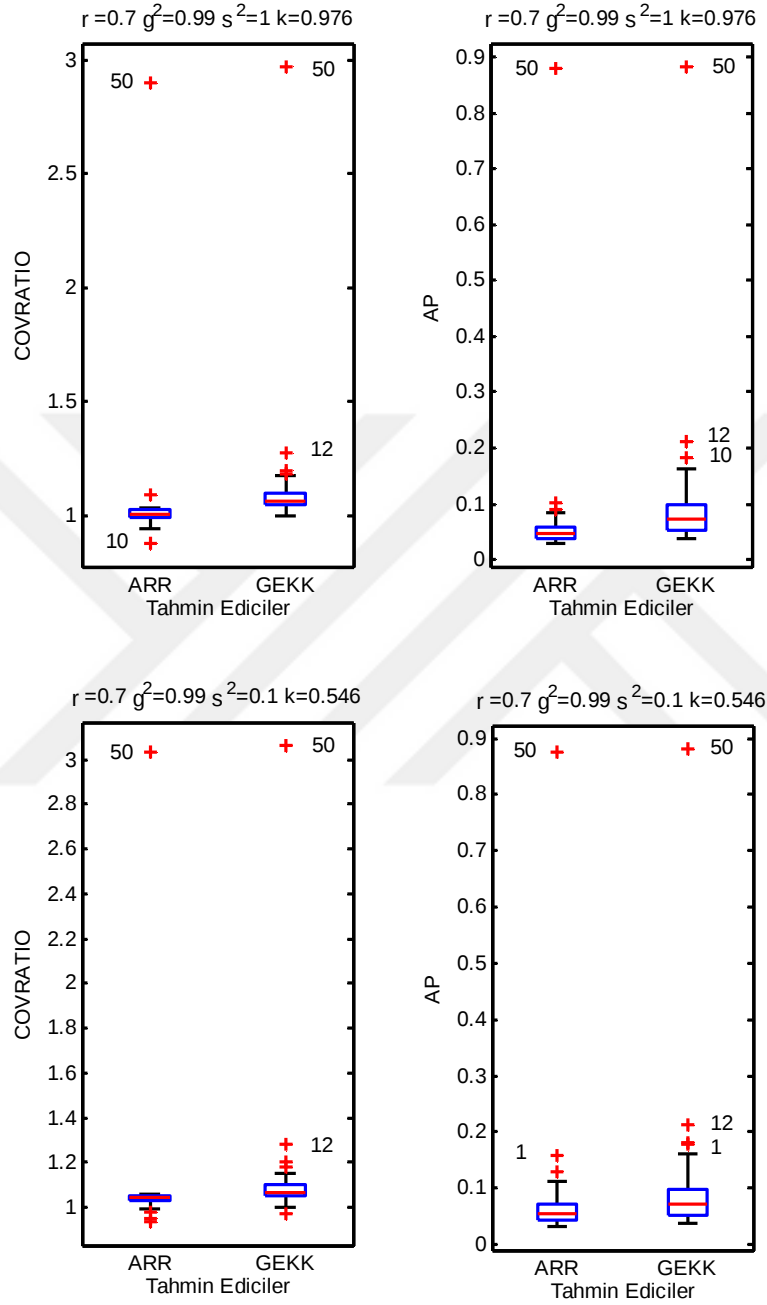
Şekil 7.1. $\gamma^2 = 0.99$ iken farklı ρ ve σ değerlerinde ARR ve GEKK için COVRATIO ve AP sonuçlarının kutu grafikleri



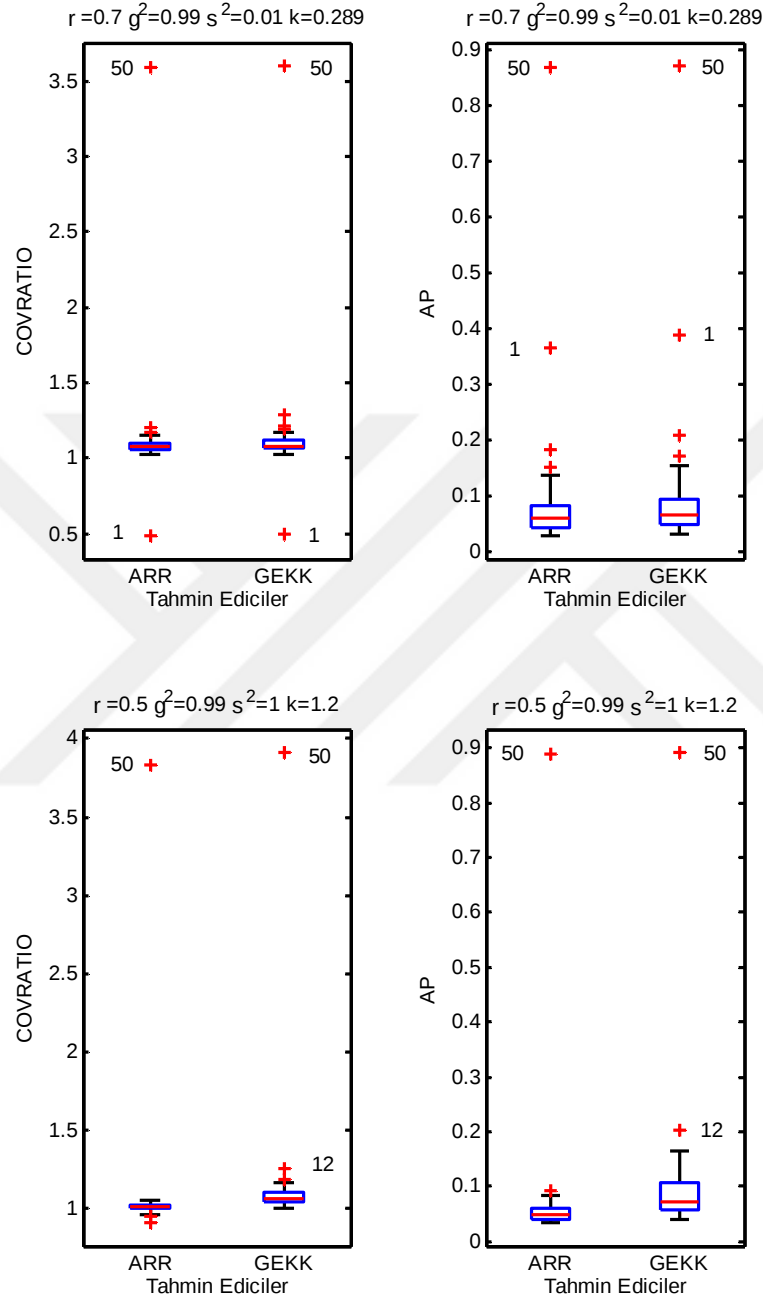
Şekil 7.1. $\gamma^2 = 0.99$ iken farklı ρ ve σ değerlerinde ARR ve GEKK için COVRATIO ve AP sonuçlarının kutu grafikleri (Devam)



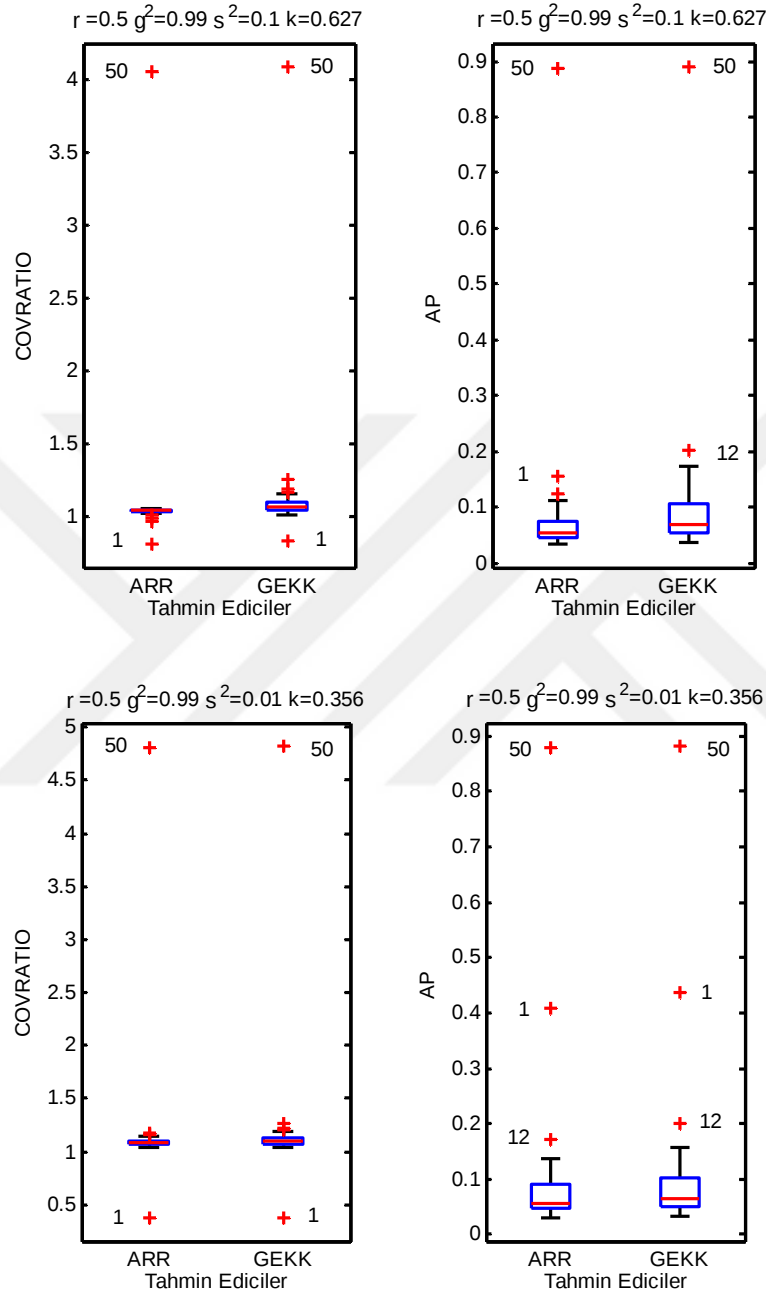
Şekil 7.1. $\gamma^2 = 0.99$ iken farklı ρ ve σ değerlerinde ARR ve GEKK için COVRATIO ve AP sonuçlarının kutu grafikleri (Devam)



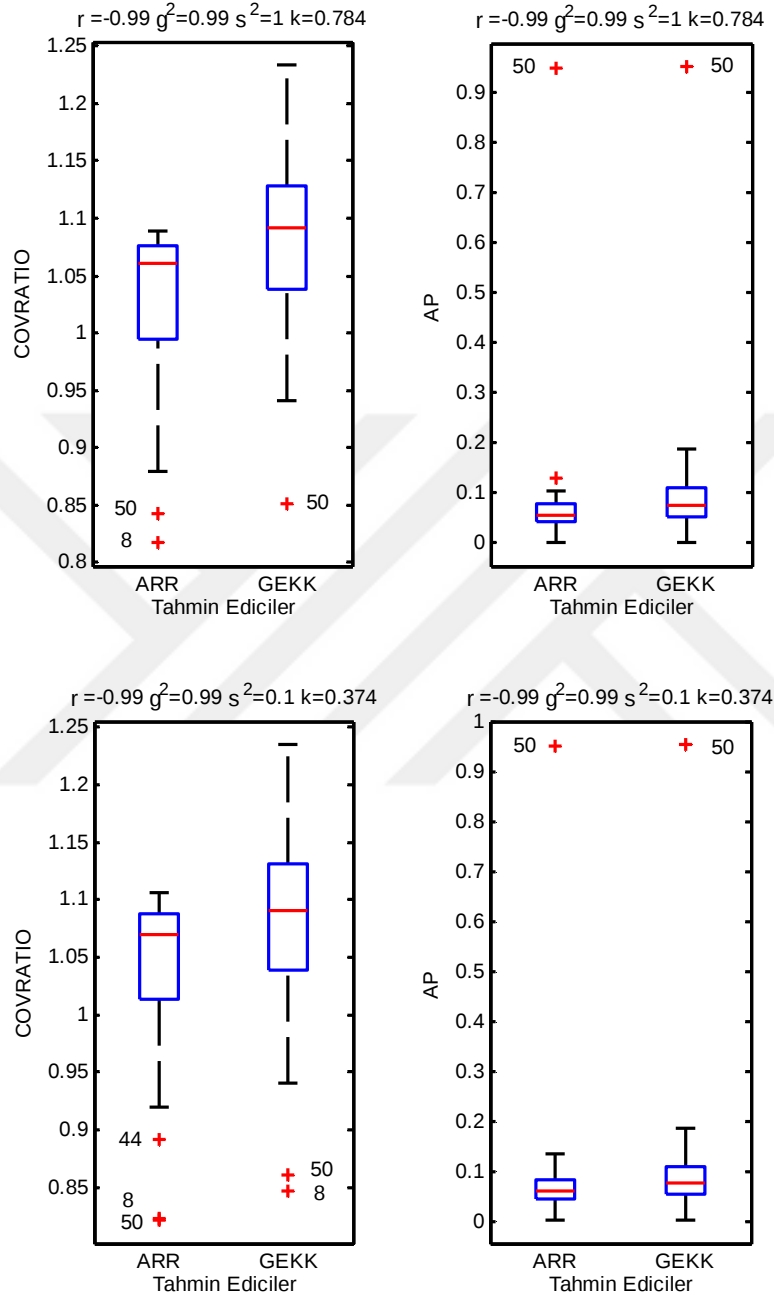
Şekil 7.1. $\gamma^2 = 0.99$ iken farklı ρ ve σ değerlerinde ARR ve GEKK için COVRATIO ve AP sonuçlarının kutu grafikleri (Devam)



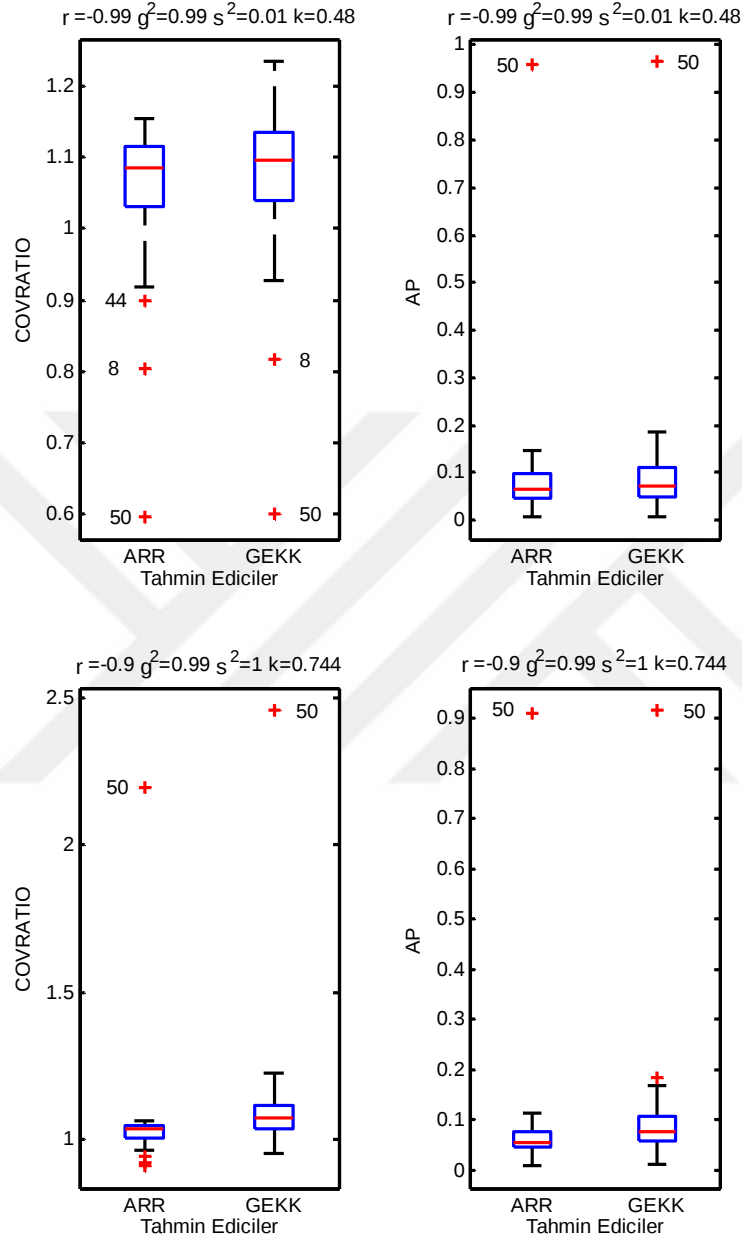
Şekil 7.1. $\gamma^2 = 0.99$ iken farklı ρ ve σ değerlerinde ARR ve GEKK için COVRATIO ve AP sonuçlarının kutu grafikleri (Devam)



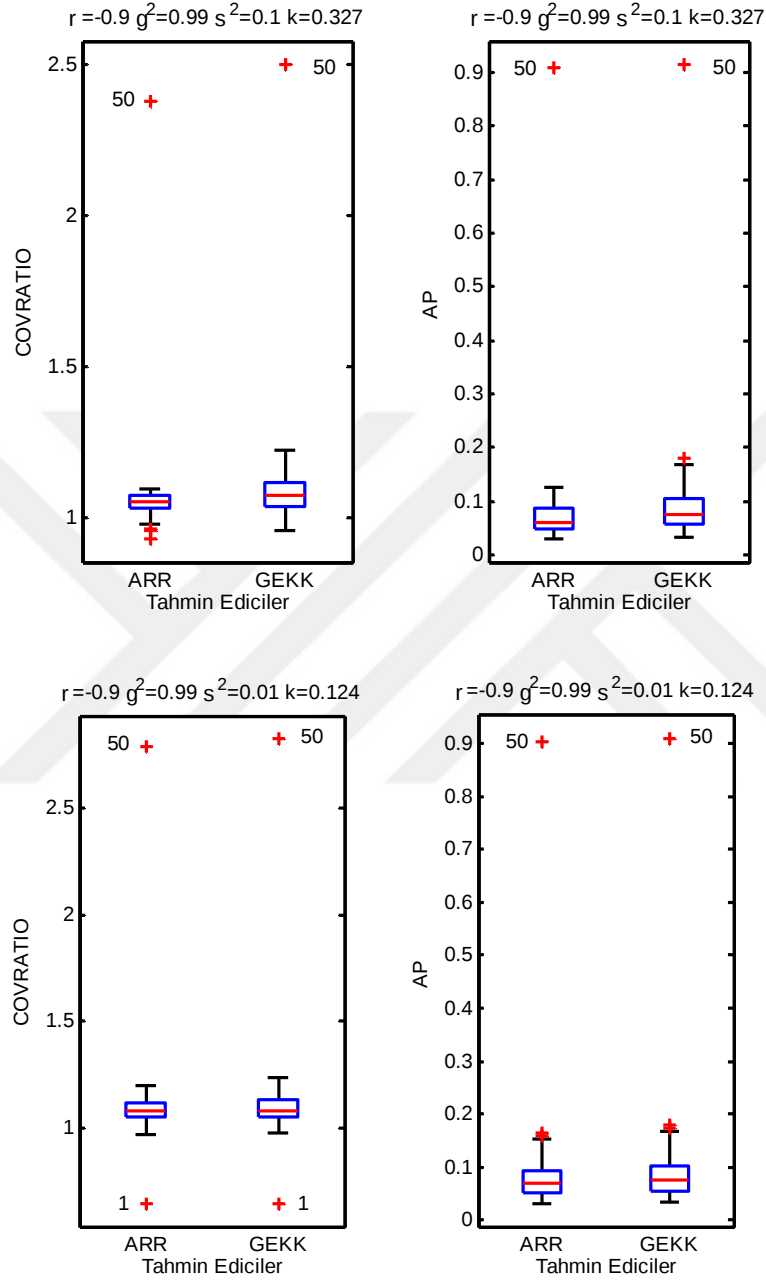
Şekil 7.1. $\gamma^2 = 0.99$ iken farklı ρ ve σ değerlerinde ARR ve GEKK için COVRATIO ve AP sonuçlarının kutu grafikleri (Devam)



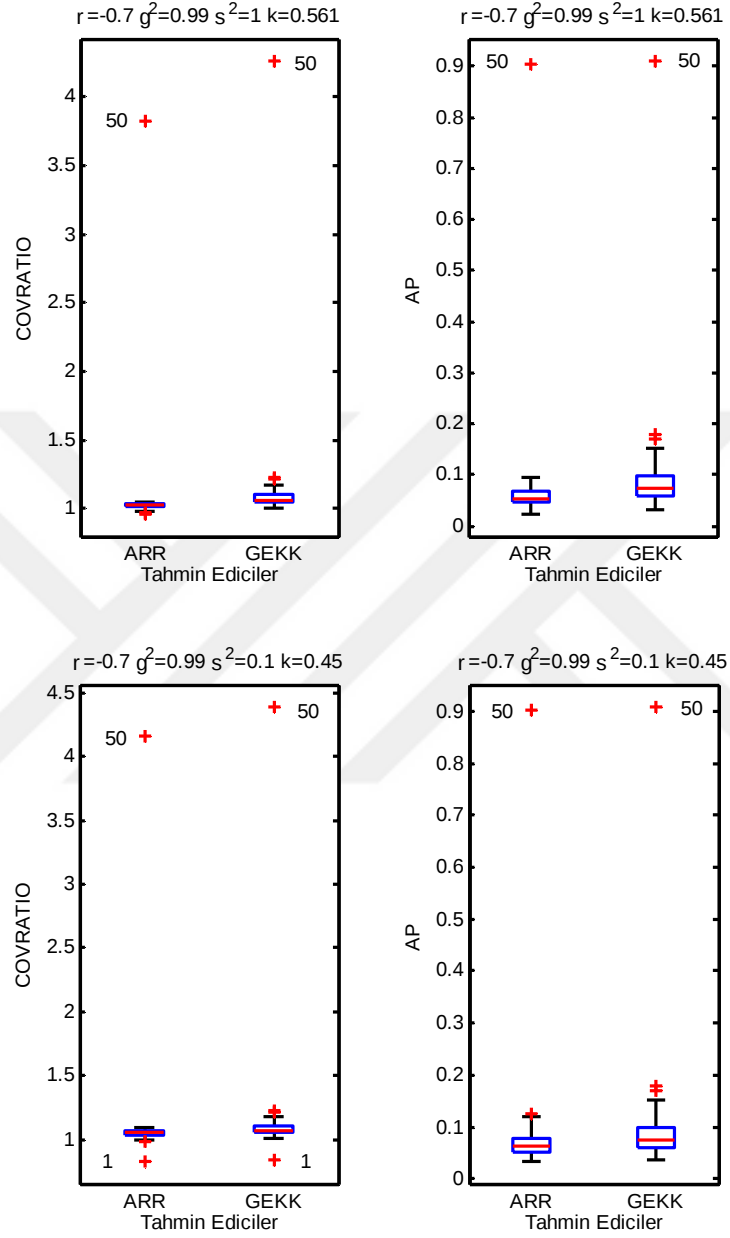
Şekil 7.1. $\gamma^2 = 0.99$ iken farklı ρ ve σ değerlerinde ARR ve GEKK için COVRATIO ve AP sonuçlarının kutu grafikleri (Devam)



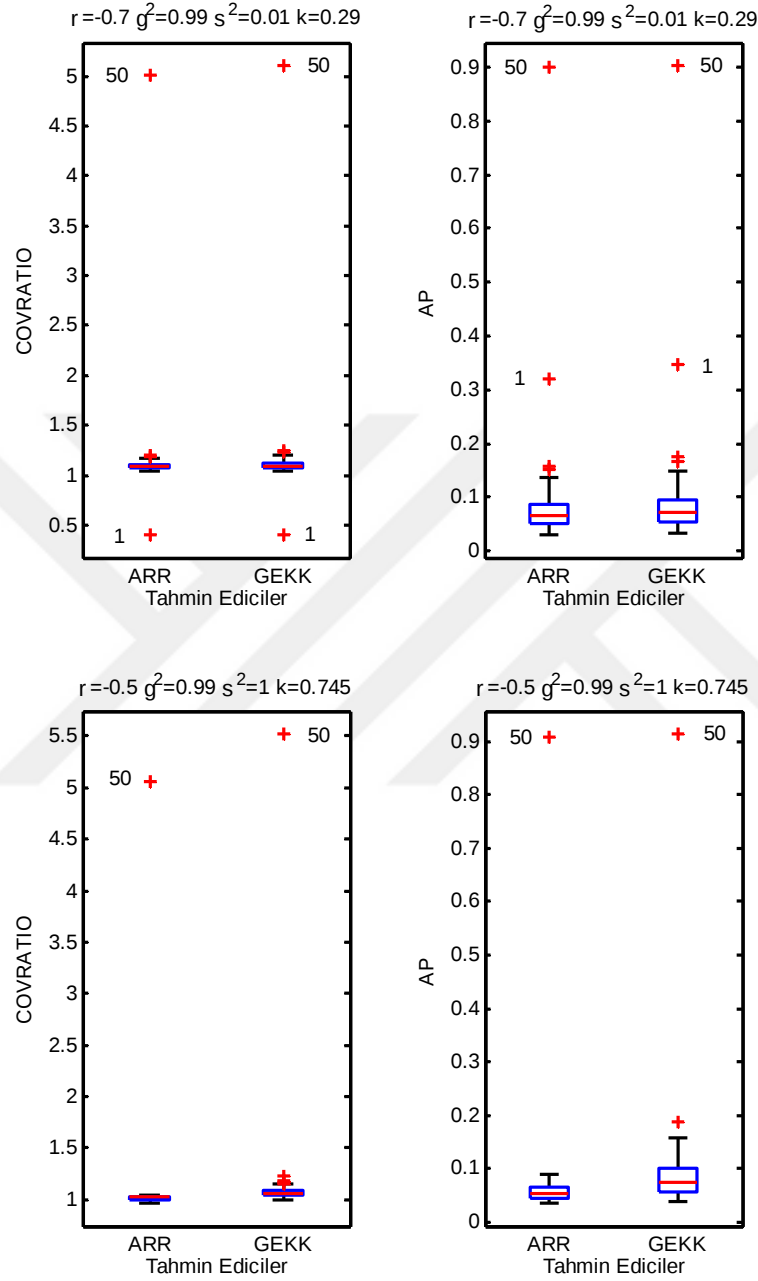
Şekil 7.1. $\gamma^2 = 0.99$ iken farklı ρ ve σ değerlerinde ARR ve GEKK için COVRATIO ve AP sonuçlarının kutu grafikleri (Devam)



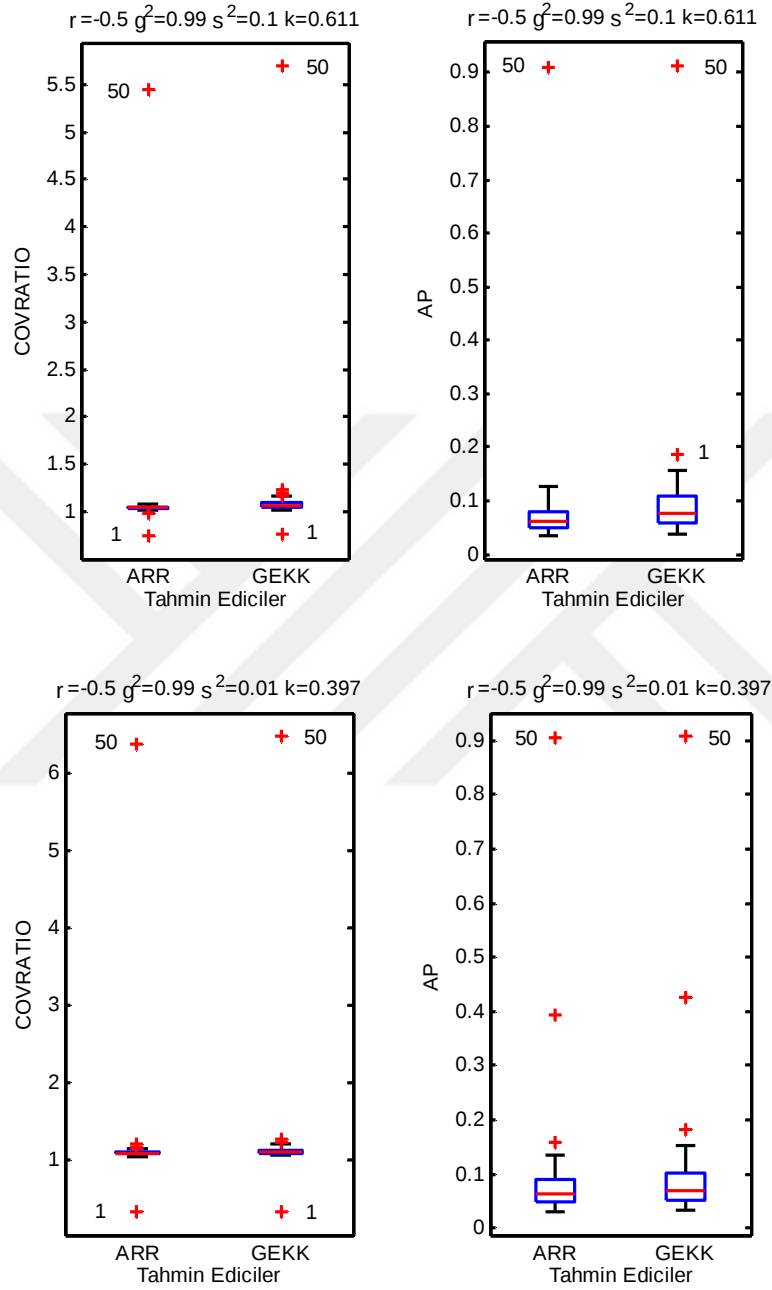
Şekil 7.1. $\gamma^2 = 0.99$ iken farklı ρ ve σ değerlerinde ARR ve GEKK için COVRATIO ve AP sonuçlarının kutu grafikleri (Devam)



Şekil 7.1. $\gamma^2 = 0.99$ iken farklı ρ ve σ değerlerinde ARR ve GEKK için COVRATIO ve AP sonuçlarının kutu grafikleri (Devam)



Şekil 7.1. $\gamma^2 = 0.99$ iken farklı ρ ve σ değerlerinde ARR ve GEKK için COVRATIO ve AP sonuçlarının kutu grafikleri (Devam)



Şekil 7.1. $\gamma^2 = 0.99$ iken farklı ρ ve σ değerlerinde ARR ve GEKK için COVRATIO ve AP sonuçlarının kutu grafikleri (Devam)

Şekil 7.1'den elde edilen sonuçlar aşağıda verilmiştir.

Simülasyon çalışmasının başlangıcında etkin gözlem olarak değiştirilen son gözlem tüm durumlarda (farklı ρ ve σ_u^2) COVRATIO ve AP ölçümlerine göre aykırı gözlem olarak belirlenmiştir.

σ_u^2 azaldıkça ARR tahmin edicisi ile elde edilen COVRATIO ve AP değerleri GEKK ile elde edilen değerlere yakınsamaktadır. Bu yüzden, GEKK ve ARR tüm durumlarda aynı gözlemi aykırı olarak belirlemesine rağmen, iki tahmin edici için elde edilen COVRATIO ve AP değerleri arasındaki farklar σ_u^2 'nin yüksek değerlerinde artmaktadır.

İlk dönüştürülmüş gözleme ait COVRATIO ve AP değerleri incelendiğinde $\rho \cong 1$ 'de ilk dönüştürülmüş gözlemin veri merkezinden oldukça uzakta olduğu görülmektedir.

Negatif ve pozitif otokorelasyon noktaları karşılaştırıldığında, $\rho = -0.99$ hariç tüm ρ değerlerinde birini, ikinci ve üçüncü kartillerin birbirine oldukça yakın olduğu gözlenmiştir.

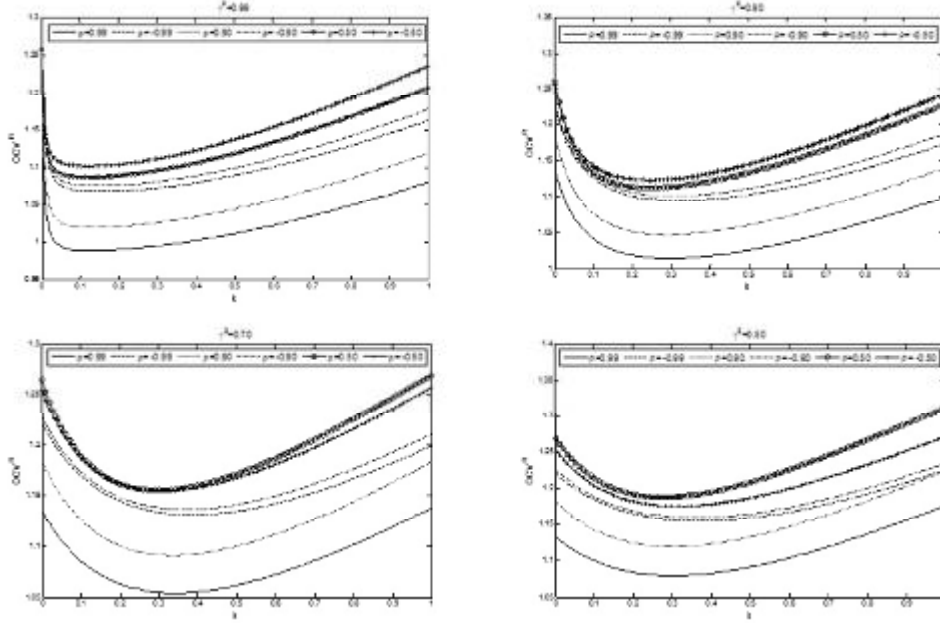
7.3. Model Geçerliliği için Monte Carlo Simülasyon

Otokorelasyon katsayısı ve ridge yanlılık parametresine bağlı olarak OCV, GCV ve C_p ölçümlerinin davranışlarını belirlemek adına Monte Carlo simülasyon çalışması uygulanmıştır.

- Ø Örneklem genişliği $n = 20$, değişken sayısı $r = 3$, $p = 4$ olarak belirlenmiştir.
- Ø Veride çoklu iç ilişkinin gücü $\gamma^2 = 0.50, 0.70, 0.90, 0.99$ olarak seçilmiş ve ilişkili açıklayıcı değişkenleri üretmek için McDonald ve Galarneau (1975) tarafından verilen (7.1) eşitliği kullanılmıştır.

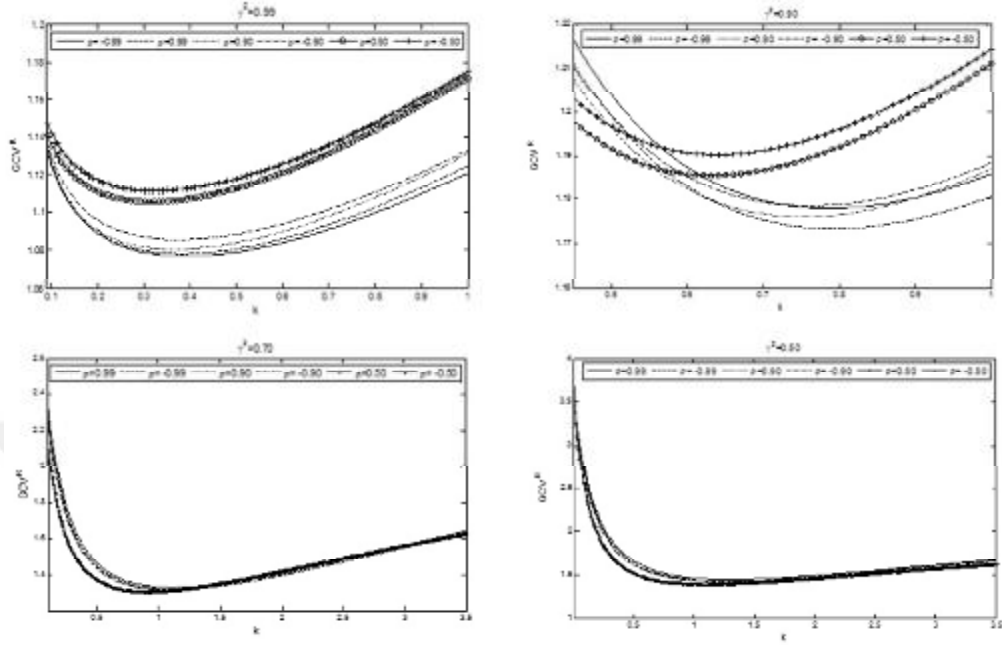
- Ø Açıklayıcı değişkenler (X) üretildikten sonra $Z = [1 \ X]$ olarak alınmıştır. Burada X matrisi, merkezileştirilmiş ve standartlaştırılmış formdadır.
- Ø Parametre vektörü $\beta = (2, 2, 2, 2)'$ olarak alınmıştır.
- Ø Otokorelasyon katsayısı keyfi olarak $|\rho| = 0.50, 0.70, 0.90, 0.99$ olarak seçilmiş ve hatalar arasındaki ilişki eşitlik (3.2)'de ki modellenmiştir.
- Ø Yanıt değişkene ait değerler (7.2) eşitliğindeki modelden elde edilmiştir. Yanıt vektörü y orjinal formda kalmıştır (standartlaştırma yapılamamıştır).
- Ø Otokorelasyon katsayısı ve ridge yanlılık parametresine karşı OCV^R , GCV^R ve C_p^R değerlerinin grafiği çizdirilirken $\sigma_u = 1$ olarak sabitlenmiştir.
- Ø En küçük OCV^R , GCV^R ve C_p^R değerlerini veren k_{opt} değerleri çizelgeler ile sunulmuş ve çizelgelerde farklı ρ ve γ^2 'ye ek olarak farklı σ_u değerleri de dikkate alınmıştır.
- Ø Son olarak her defasında yeni bir hata terimi üreterek simülasyon 1000 deneme üzerinden gerçekleştirilmiş ve ilgili istatistiklerin MCN'ye göre ortalamaları alınmıştır.

Optimum ridge yanlılık parametresi OCV, GCV ve C_p istatistiklerini minimum yapacak şekilde belirlenmiş ve sonuçlar aşağıda sunulmuştur.



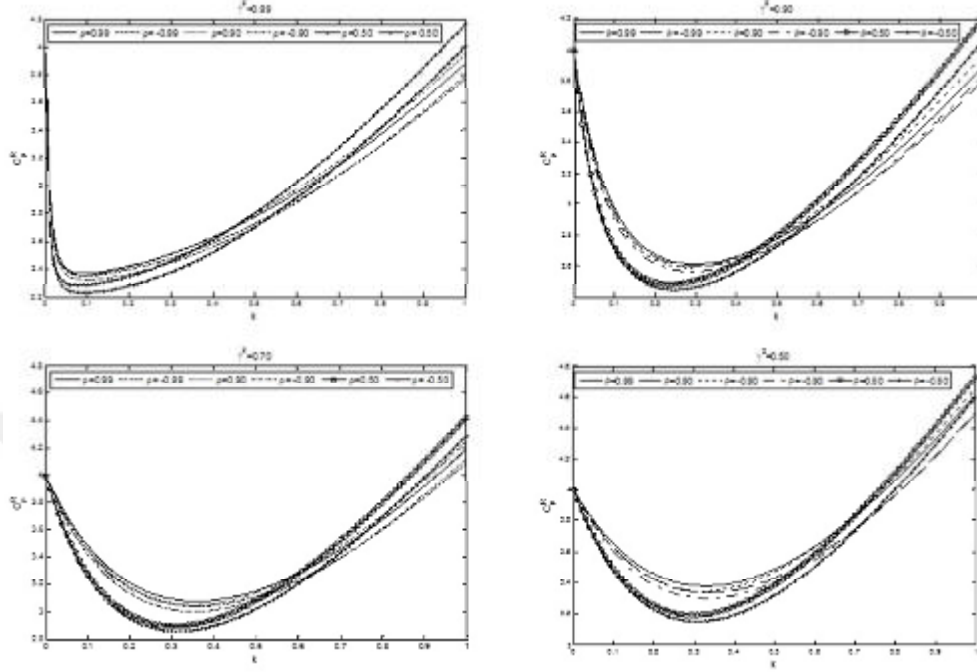
Şekil 7.2. ARR tahmin edicisi için ridge yanlılık parametresine karşı OCV^R değerleri, Monte Carlo simülasyon

Şekil 7.2’de farklı otokorelasyon katsayısı ve çoklu iç ilişki gücünde OCV^R değerlerinin ridge yanlılık parametresine göre grafiği verilmiştir. Çoklu iç ilişki şiddetli iken yani $\gamma^2 = 0.99$ ve $\gamma^2 = 0.90$ ’da k arttıkça OCV^R değerleri hızlı bir şekilde azalmaktadır. γ^2 azaldıkça OCV^R yerel minimum noktasına doğru yavaşça azalmaktadır. Bu durum OCV^R istatistiğini minimum yapan k_{opt} değerini etkiler. Yani γ^2 arttıkça OCV^R ’yi minimum yapan k_{opt} değeri küçülmektedir. Otokorelasyon katsayısının küçük olduğu durumlarda daha büyük OCV^R değerleri elde edilirken, sabit otokorelasyon değerinin işaretinin değişmesi OCV^R değerlerinde önemli bir artış ya da azalışa neden olmamıştır.



Şekil 7.3. ARR tahmin edicisi için ridge yanlılık parametresine karşı GCV, Monte Carlo simülasyon

Şekil 7.3’de farklı otokorelasyon katsayısı ve çoklu iç ilişki gücünde GCV^R değerlerinin ridge yanlılık parametresine göre grafiği verilmiştir. Grafiklerden görüldüğü üzere, $\gamma^2 = 0.99$ iken $|\rho| = 0.50$ ’deki GCV^R değerleri, $|\rho| = 0.99, 0.90$ ’daki GCV^R değerlerinden tüm k değerleri için büyüktür. Ancak, azalan γ^2 ile $|\rho| = 0.50$ ’de elde edilen GCV^R değerleri $|\rho| = 0.99, 0.90$ ’de elde edilen GCV^R değerlerine yakınsamaktadır. γ^2 ’nin tüm değerleri için $|\rho| = 0.50$ ’deki k_{opt} değeri $|\rho| = 0.99, 0.90$ ’dekilerden daha küçüktür. γ^2 ’nin yüksek olmadığı durumlarda GCV^R istatistiğini minimum yapan k_{opt} değerine kadar olan azalış, k_{opt} değerinden sonraki artıştan daha hızlıdır.



Şekil 7.4. ARR tahmin edicisi için ridge yanlılık parametresine karşı C_p , Monte Carlo simülasyon

Şekil 7.4’de farklı otokorelasyon katsayısı ve çoklu iç ilişki gücünde C_p^R değerlerinin ridge yanlılık parametresine göre grafiği verilmiştir. γ^2 artarken minimum C_p^R değerini veren k_{opt} azalmaktadır. $|\rho| = 0.50$ ’de elde edilen k_{opt} değeri $|\rho| = 0.90, 0.99$ ’de elde edilen k_{opt} değerinden daha küçüktür. Şekil 7.4’den görüldüğü üzere, k_{opt} değerinden daha büyük olan bir eşik değeri vardır. Bu k ’nın eşik değerinden önce $|\rho| = 0.50$ ’de elde edilen C_p^R değeri $|\rho| = 0.90, 0.99$ ’da elde edilen C_p değerinden küçük iken, k ’nın eşik değerinden sonra $|\rho| = 0.50$ ’de elde edilen C_p^R değeri $|\rho| = 0.90, 0.99$ ’da elde edilen C_p^R değerinden daha büyüktür.

Çizelge 7.15. $\gamma^2 = 0.99$ iken OCV^R , GCV^R , C_p^R değerleri ve ilgili optimum ridge yanlılık parametreleri

$\gamma^2 = 0.99$	OCV^R/k_{opt}	GCV^R/k_{opt}	C_p^R/k_{opt}
$\rho = 0.99, \sigma_u = 0.1$	0.0102/0.026	0.0123/0.066	2.7017/0.026
$\rho = 0.90, \sigma_u = 0.1$	0.0105/0.026	0.0123/0.062	2.6832/0.025
$\rho = 0.70, \sigma_u = 0.1$	0.0110/0.023	0.0124/0.055	2.6359/0.023
$\rho = 0.50, \sigma_u = 0.1$	0.0112/0.021	0.0124/0.050	2.5950/0.021
$\rho = 0.99, \sigma_u = 0.5$	0.2485/0.076	0.2762/0.208	2.4065/0.066
$\rho = 0.90, \sigma_u = 0.5$	0.2564/0.076	0.2767/0.198	2.3949/0.062
$\rho = 0.70, \sigma_u = 0.5$	0.2680/0.071	0.2799/0.179	2.3559/0.057
$\rho = 0.50, \sigma_u = 0.5$	0.2732/0.065	0.2828/0.163	2.3229/0.053
$\rho = 0.99, \sigma_u = 1$	0.9894/0.129	1.0780/0.380	2.3682/0.089
$\rho = 0.90, \sigma_u = 1$	1.0204/0.125	1.0804/0.365	2.3572/0.086
$\rho = 0.70, \sigma_u = 1$	1.0666/0.124	1.0937/0.337	2.3173/0.082
$\rho = 0.50, \sigma_u = 1$	1.0876/0.118	1.1053/0.316	2.2833/0.079

Çizelge 7.16. $\gamma^2 = 0.90$ iken OCV^R , GCV^R , C_p^R değerleri ve ilgili optimum ridge yanlılık parametreleri

$\gamma^2 = 0.90$	OCV^R/k_{opt}	GCV^R/k_{opt}	C_p^R/k_{opt}
$\rho = 0.99, \sigma_u = 0.1$	0.0111/0.024	0.0195/0.115	3.7480/0.027
$\rho = 0.90, \sigma_u = 0.1$	0.0116/0.024	0.0194/0.110	3.7320/0.026
$\rho = 0.70, \sigma_u = 0.1$	0.0121/0.023	0.0191/0.100	3.6941/0.025
$\rho = 0.50, \sigma_u = 0.1$	0.0123/0.023	0.0189/0.092	3.6575/0.024
$\rho = 0.99, \sigma_u = 0.5$	0.2603/0.158	0.3262/0.442	2.9293/0.165
$\rho = 0.90, \sigma_u = 0.5$	0.2692/0.155	0.3250/0.420	2.9039/0.158
$\rho = 0.70, \sigma_u = 0.5$	0.2808/0.143	0.3251/0.379	2.8424/0.145
$\rho = 0.50, \sigma_u = 0.5$	0.2852/0.133	0.3251/0.346	2.7881/0.134
$\rho = 0.99, \sigma_u = 1$	1.0140/0.293	1.1779/0.786	2.6123/0.294
$\rho = 0.90, \sigma_u = 1$	1.0472/0.288	1.1760/0.751	2.5902/0.280
$\rho = 0.70, \sigma_u = 1$	1.0932/0.267	1.1810/0.684	2.5327/0.257
$\rho = 0.50, \sigma_u = 1$	1.1121/0.247	1.1588/0.631	2.4823/0.238

Çizelge 7.17. $\gamma^2 = 0.70$ iken OCV^R, GCV^R, C_p^R değerleri ve ilgili optimum ridge yanlılık parametreleri

$\gamma^2 = 0.70$	OCV^R/k_{opt}	GCV^R/k_{opt}	C_p^R/k_{opt}
$\rho = 0.99, \sigma_u = 0.1$	0.0113/0.009	0.0281/0.101	3.9630/0.011
$\rho = 0.90, \sigma_u = 0.1$	0.0118/0.009	0.0278/0.099	3.9598/0.011
$\rho = 0.70, \sigma_u = 0.1$	0.0124/0.009	0.0275/0.093	3.9516/0.011
$\rho = 0.50, \sigma_u = 0.1$	0.0126/0.009	0.0271/0.088	3.9431/0.011
$\rho = 0.99, \sigma_u = 0.5$	0.2732/0.140	0.4027/0.578	3.5158/0.156
$\rho = 0.90, \sigma_u = 0.5$	0.2836/0.138	0.3988/0.553	3.4904/0.152
$\rho = 0.70, \sigma_u = 0.5$	0.2964/0.132	0.3936/0.507	3.4298/0.144
$\rho = 0.50, \sigma_u = 0.5$	0.3007/0.127	0.3891/0.468	3.3729/0.138
$\rho = 0.99, \sigma_u = 1$	1.0543/0.339	1.3265/1.120	3.0767/0.360
$\rho = 0.90, \sigma_u = 1$	1.0912/0.330	1.3176/1.070	3.0444/0.350
$\rho = 0.70, \sigma_u = 1$	1.1385/0.310	1.3090/0.980	2.9676/0.330
$\rho = 0.50, \sigma_u = 1$	1.1555/0.300	1.3018/0.910	2.8977/0.310

Çizelge 7.18. $\gamma^2 = 0.50$ iken OCV^R, GCV^R, C_p^R değerleri ve ilgili optimum ridge yanlılık parametreleri

$\gamma^2 = 0.70$	OCV^R/k_{opt}	GCV^R/k_{opt}	C_p^R/k_{opt}
$\rho = 0.99, \sigma_u = 0.1$	0.0113/0.004	0.0324/0.073	3.9875/0.006
$\rho = 0.90, \sigma_u = 0.1$	0.0118/0.004	0.0322/0.072	3.9863/0.006
$\rho = 0.70, \sigma_u = 0.1$	0.0124/0.004	0.0320/0.070	3.9834/0.006
$\rho = 0.50, \sigma_u = 0.1$	0.0127/0.004	0.0318/0.068	3.9802/0.006
$\rho = 0.99, \sigma_u = 0.5$	0.2782/0.099	0.4643/0.606	3.7572/0.114
$\rho = 0.90, \sigma_u = 0.5$	0.2895/0.098	0.4582/0.584	3.7395/0.113
$\rho = 0.70, \sigma_u = 0.5$	0.3037/0.094	0.4486/0.543	3.6961/0.111
$\rho = 0.50, \sigma_u = 0.5$	0.3089/0.092	0.4401/0.510	3.6528/0.109
$\rho = 0.99, \sigma_u = 1$	1.0797/0.302	1.4389/1.293	3.3776/0.329
$\rho = 0.90, \sigma_u = 1$	1.1196/0.300	1.4238/1.247	3.3445/0.321
$\rho = 0.70, \sigma_u = 1$	1.1702/0.286	1.4027/1.163	3.2652/0.309
$\rho = 0.50, \sigma_u = 1$	1.1882/0.276	1.3849/1.094	3.1905/0.298

γ^2, ρ ve σ_u 'nın OCV^R, GCV^R, C_p^R ve ilgili k_{opt} değerleri üzerindeki etkisini görmek için Çizelge 7.15-7.18 hazırlanmıştır. Çizelge 7.15- 7.18'den görüldüğü üzere, γ^2 ve σ_u sabit iken otokorelasyon katsayısı ρ azaldıkça k_{opt} ile elde edilen GCV^R ve C_p^R değerleri azalırken OCV^R değeri artmaktadır. Diğer yandan γ^2 ve ρ sabit tutulduğunda σ_u arttıkça k_{opt} ile elde edilen C_p^R değeri

azalırken, OCV^R ve GCV^R değerleri artmaktadır. ρ ve σ_u sabit tutulduğunda γ^2 azaldıkça k_{opt} ile elde edilen OCV^R, GCV^R, C_p^R değerleri yavaşça azalmaktadır.





8. SONUÇ VE ÖNERİLER

Bu araştırma ile daha önce literatürde var olan sıradan en küçük kareler uygulamak için gerekli olan varsayımların sağlandığı lineer istatistiksel modelde kaldıraç, DFBETA, DFFITS, Cook uzaklığı, COVRATIO, Cook-Weisberg, Andrews-Pregibon, sıradan çapraz geçerlilik, genelleştirilmiş çapraz geçerlilik ve kavramsal ön tahmin kavramlarına değinilmiştir.

Hatalara ait ilişkisizlik varsayımının ihlali olan otokorelasyon problemi altında genelleştirilmiş en küçük kareler tahmin edici verilmiştir. Hatalar arasındaki ilişki birinci dereceden otoregresif süreç olarak ele alınmıştır.

Açıklayıcı değişkenlerin lineer bağımsızlığı varsayımının ihlali olan çoklu iç ilişki problemi otokorelasyon varlığında incelenmiş ve çoklu iç ilişki problemi altında önerilen otokorelasyonlu ridge regresyon tahmin edici verilmiştir.

Otokorelasyon ve çoklu iç ilişki problemlerini aynı anda barındıran veri kümelerinde etki tanılama yöntemleri incelenmiş ve aykırı gözlemlerin varlığının tespitine yönelik ölçümler önerilmiştir.

Otokorelasyonlu hataya sahip lineer istatistiksel modelde genelleştirilmiş en küçük kareler tahmin edici için daha önce bazı yazarlar tarafından tanımlanan kaldıraç, rezidü, DFFITS, Cook Uzaklığı, DFBETA, OCV, GCV ve C_p ölçümlerine ek olarak bu çalışmada genelleştirilmiş en küçük kareler tahmin edici için COVRATIO, Cook-Weisberg ve Andrews-Pregibon istatistikleri tanımlanmıştır.

Otokorelasyonlu ridge regresyon tahmin edici için kaldıraç, rezidü, DFBETA, DFFITS, Cook uzaklığı, COVRATIO, Cook-Weisberg, Andrews Pregibon, OCV, GCV ve C_p istatistikleri tanımlanmıştır.

Önerilen ölçümlerin etkisi AR(1) hatalı veri kümesi kullanarak nümerik analiz ve Monte Carlo simülasyon çalışması ile değerlendirilmiştir. Değerlendirmelerde genelleştirilmiş en küçük kareler ve otokorelasyonlu ridge tahmin edici için ölçüm değerleri karşılaştırmalı olarak verilmiştir.

Gerek teorik gerekse örnek ve simülasyon çalışmalarından aşağıdaki sonuçlar elde edilmiştir.

Ridge yanlılık parametresi gözlemlerin etkinliğinde önemli rol oynamaktadır. Öyle ki, ridge yanlılık parametresi arttıkça ARR tahmin edici ile elde edilen kaldırıcı değerler küçülmektedir. İlgili regresyon parametresi üzerinden etkin olan gözlemlerin etkinliği ridge yanlılık parametresi arttıkça azalmaktadır. ARR ile uydurulan yanıt vektörü üzerinde bazı gözlemlerin etkinliği ridge yanlılık parametresine bağlı olarak artarken bazılarının etkinliği ridge yanlılık parametresine bağlı olarak azalmaktadır. Parametre vektörüne ait varyans-kovaryans matrisi ve parametre vektörüne ait güven elipsoidinin hacmi üzerinde etkin olan gözlemlerin etkinliği ridge yanlılık parametresi arttıkça azalmaktadır.

Regresyon parametresinin güven elipsoidine dayalı etkinlik ölçümleri için COVRATIO, Cook-Weisberg ve Andrews-Pregibon istatistikleri tanımlanmıştır. ARR tahmin edici ile elde edilen COVRATIO ve AP istatistikleri GEKK'ye göre daha küçüktür. Ancak, COVRATIO ve AP değerlerinin küçük elde edilmesi ARR tahmin edicisinin model performansı üzerinde her zaman daha iyi olduğu anlamına gelmemektedir. Bu durumu göstermek için farklı otokorelasyon, varyans değerlerinde sabit yanlılık parametresi ve çoklu iç ilişki gücünde kutu grafikleri çizdirilmiştir. Kutu grafikleri ilk aşamada farklı yanlılık parametresi ve çoklu iç ilişki gücünde çizdirilmiş ancak aykırı gözlemi belirlemede çok etkin olmadığı için bu grafiklere araştırmada yer verilmemiştir. Bu yüzden k ve γ^2 sabit alınmıştır.

Literatürde AR(1) hatalı lineer istatistiksel modelde birinci dönüştürülmüş gözlemin otokorelasyon arttıkça kaldırıcı değer olduğu bilgisi yer almaktadır. Bu çalışma, birinci dereceden otoregresif bir süreçte, model eğer sabit terim içeriyorsa birinci gözlemin parametre vektörüne ait varyans-kovaryans matrisi üzerindeki ve güven elipsoidinin hacmi üzerindeki etkisinin oldukça yüksek olduğunu göstermiştir.

Otokorelasyon ve çoklu ilişki problemi olan veri kümesinde ARR tahmin edicisi kullanıldığında genel olarak gözlemlerin kaldırıcı ve etkin olma

potansiyelleri GEKK'ye göre azalmaktadır. Bu çalışmada ayrıca kaldırma ve etkin gözlemin varlığında model üzerinden yapılacak öngörülerin performansı değerlendirilmiştir. Modelin öngörü performansı için OCV, GCV ve Cp istatistikleri ele alınmıştır. ARR tahmin edicisi kullanıldığında modelin öngörü performansını GEKK'dan daha iyi kılan bir ridge yanlılık parametresi her zaman vardır. En iyi ridge yanlılık parametresi otokorelasyon katsayısından, hataların varyansından, çoklu iç ilişkinin gücünden etkilenmektedir. Bu çalışma ile model geçerliliği için önerilen istatistikleri minimum yapan ridge yanlılık parametresinin Firinguetti (1989)'nın önerdiği iki farklı ridge yanlılık parametre tahminlerinden daha uygun olduğu görülmüştür.

Çalışmadan elde edilen sonuçlar doğrultusunda benzer çalışmalar diğer yanlı tahmin ediciler için ele alınabilir. Hataların AR(1) ile modellendiği bu çalışmadan yola çıkarak farklı dereceden otoregresif süreçler ya da hareketli ortalama modelleri üzerinden regresyonda tanılama yöntemleri incelenebilir.



KAYNAKLAR

- Acarlar, I., ve Gamgam, H., 2010. Çoklu Doğrusal Regresyonda Etkili Gözlem Gruplarının Saptanması İçin Kullanılan Tanı Yöntemlerinin Karşılaştırılması. TÜİK, İstatistik Araştırma Dergisi, 7(1): 83-99.
- Allen, D.M., 1974. The Relationship Between Variable Selection and Data Augmentation and a Method for Prediction. *Technometrics*, 16(1): 125–127.
- Andrews, D.F., and Pregibon, D., 1978. Finding Outliers that Matter. *Journal of the Royal Statistical Society, Series B*, 40: 85-93.
- Andrews, D.W.K., 1991. Asymptotic Optimality of Generalized C_p , Cross-Validation, and Generalized Cross-Validation in Regression with Heteroskedastic Errors. *Journal of Econometrics*, 47: 359-377.
- Atkinson, A.C., 1981. Two Graphical Displays for Outlying and Influential Observations in Regression. *Biometrika*, 68: 13-20.
- Barry, A.M., Burney, S.M.A., and Bhatti, M.I., 1997. Optimum Influence of Initial Observations in Regression Models with AR(2) Errors. *Applied Mathematics and Computations*, 82: 57-65.
- Bayhan, G. M., and Bayhan, M., 1998. Forecasting Using Autocorrelated Errors and Multicollinear Predictor Variables. *Computers and Industrial Engineering*, 34: 413-421.
- Belsley, D.A., Kuh, E., and Welsch, R.E., 1980. *Regression Diagnostics. Identifying Influential Data and Sources of Collinearity*. John Wiley & Sons, New Jersey, 292 p.
- Brown, P.J., 1977. Centering and Scaling in Ridge Regression, *Technometrics*, 19(1): 35-36.
- Chatterjee, S., and Hadi, A.S., 1986. Influential Observations, High Leverage Points, and Outliers in Linear Regression. *Statistical Science*, 1(3): 379-416.

- Chatterjee, S., and Hadi, A.S., 1988. *Sensitivity Analysis in Linear Regression*. John Wiley and Sons Inc., New York, 315 p.
- Chatterjee, S., Hadi, A.S., and Price, B., 2000. *Regression Analysis by Example*, 3rd ed.. John Wiley and Sons Inc., New York, 359 p.
- Cook, R.D., 1977. Detection of Influential Observations in Linear Regression. *Technometrics*, 19: 15-18.
- Cook, R.D., and Weisberg, S., 1980. Characterizations of an Empirical Influence Function for Detecting Influential Cases in Regression. *Technometrics*, 22: 495-508.
- Cook, R.D., and Weisberg, S., 1982. *Residuals and Influence in Regression*. Chapman and Hall, New York, 230 p.
- Draper, N.R., and John, J.A., 1981. Influential Observations and Outliers in Regression. *Technometrics*, 23(1): 21-26.
- Durbin, J. and Watson, G.S., 1950. Testing for Serial Correlation in Least Squares Regression I. *Biometrika*, 37: 409-428.
- Durbin, J. and Watson, G.S., 1971. Testing for Serial Correlation in Least Squares Regression III. *Biometrika*, 58: 1-42.
- Firinguetti, L., 1989. A Simulation Study of Ridge Regression Estimators with Autocorrelated Errors. *Commun. Statist.- Sim. and Comp.*, 18(2): 673-702.
- Ghapor, A.A., Zubairi, Y.Z., Mamun, A.S.M.A., and Imon, A.H.M.R., 2014. On Detecting Outlier in Simple Linear Functional Relationship Model Using Covratio Statistic. *Pakistan Journal of Statistics*, 30(1): 129-142.
- Godfrey, L.G., 1978. Testing for Higher Order Serial Correlation in Regression Equations when the Regressors Include Lagged Dependent Variables. *Econometrica*, 46(6), 1303-1310.
- Golub, G.H., Heath, M., and Wahba, G., 1979. Generalized Cross-Validation as a Method for Choosing a Good Ridge Parameter. *Technometrics*, 21(2): 215-223.

- Graybill, F. A., 1976. Theory and Application of the Linear Model. Duxbery Press, Canada, 704 p.
- Graybill, F. A., 1983. Matrices with Applications in Statistics, 2nd ed. Duxbury Press, California, 461 p.
- Griffiths, W.E., Hill, R.C., and Judge, G.G, 1993. Learning and Practicing Econometrics. John Wiley & Sons, New York, 866 p.
- Gujarati, D.N., 2004. Basic Econometrics, 4th ed. McGraw-Hill, New Jersey, 1002 p.
- Hadi, A.S., and Simonoff, J.S., 1993. Procedures for the Identification of Multiple Outliers in Linear Models. Journal of American Statistical Association, 88: 1264-1272.
- Hampel, F.R., 1974. The Influence Curve and Its Role in Robust Estimation. Journal of the American Statistical Association, 69: 383–393.
- Haslett, J., 1999. A Simple Derivation of Deletion Diagnostic Results for the General Linear Model with Correlated Errors. Journal of the Royal Statistical Society, 61(3): 603-609.
- Hocking, R.R., 1976. The Analysis and Selection of Variables in the Linear Regression. Biometrics, 32: 1-49.
- Hoerl, A. E., and Kennard, R. W., 1970. Ridge Regression: Biased Estimation for Nonorthogonal Problems. Technometrics, 12(1): 55-67.
- Holgersson, H. E. T., 2015. A Note on a Commonly Used Ridge Regression Monte Carlo Design. Commun. Statist.- Theor. Meth., 44: 2176-2179.
- Ibrahim, S., Rambli, A., Hussin, A.G., and Mohamed, I., 2013. Outlier Detection in a Circular Regression Model Using COVRATIO Statistic. Commun. Statist.- Sim. and Comp., 42: 2272-2280.
- Judge, G.G., Griffiths, W.E., Hill, R.C., Lutkepohl, H., and Lee, T.C., 1980. The Theory and Practice of Econometrics, 2nd ed. John Wiley, New York, 1019 p.

- Judge, G.G., Hill, R.C., Griffiths, W.E., Lutkepohl, H., and Lee, T.C., 1988. Introduction to the Theory and Practice of Econometrics, 2nd ed. John Wiley, New York, 1024 p.
- Kaçıranlar, S., 2003. Liu Estimator in the General Linear Regression Model. *Journal of Applied Statistical Science*, 12(3): 229-234.
- Liu, H., Weiss, R.E., Jennrich, R.I., and Wenger, N.S., 1999. PRESS Model Selection in Repeated Measure Data. *Computational Statistics and Data Analysis*, 30(2): 169-184.
- Liu, X.Q., and Jiang, H.Y., 2012. Optimal Generalized Ridge Estimator Under the Generalized Cross-Validation Criterion in Linear Regression. *Linear Algebra Appl.*, 436: 1238–1245.
- Malinvaud, E., 1966. *Statistical Methods of Econometrics*. Rand McNally, Chicago, 631 p.
- Mallows, C.L., 1973. Some Comments on Cp. *Technometrics*, 15(4): 661-675.
- Marquardt, D. W., 1970. Generalized Inverses, Ridge Regression, Biased Linear Estimation, and Nonlinear Estimation, *Technometrics*. 12: 591-612.
- Martin R.J., 1992. Leverage, Influence and Residuals in Regression Models when Observations are Correlated. *Communications in Statistics-Theory and Methods*, 21(5): 1183-1212.
- Mayer, L. S., and Willke, T. A., 1973. On Biased Estimations in Linear Models, *Technometrics* 15, 497-508
- McDonald, G.C., and Galarneau, D.I., 1975. A Monte Carlo Evaluation of Some Ridge Type Estimators. *Journal of the American Statistical Association*, 70: 407-416.
- Montgomery, D.C., and Friedman, D.J., 1993. Prediction Using Regression Models with Multicollinear Predictor Variables. *IIE Trans* 25(3): 73–85.
- Montgomery, D., Peck, E.A., and Vining, G.G., 2001. *Introduction to Linear Regression Analysis*. John Wiley and Sons Inc., New York, 641 p.

- Myers, R.H., 1990. Classical and Modern Regression with Applications. Duxbury Press, California, 488 p.
- Özkale, M.R., and Kaçıranlar, S., 2007. A Prediction-Oriented Criterion for Choosing the Biasing Parameter in Liu Estimation. *Commun. Stat.* 36(10): 1889–1903.
- Özkale M.R., 2008. A Jackknifed Ridge Estimator in the Linear Regression Model with Heteroscedastic or Correlated Errors. *Statistics and Probability Letters*, 78 (18): 3159-3169.
- Özkale M.R., 2014. Predictive Performance of Linear Regression Models. *Statistical Papers*, DOI: 10.1007/s00362-014-0596-4.
- Picard, R.R., and Cook, R. D., 1984. Cross-Validation of Regression Models. *Journal of American Statistical Association*, 79(387): 575-583.
- Puterman, M.L., 1988. Leverage and Influence in Autocorrelated Regression Models. *Journal of the Royal Statistical Society*, 37(1): 76-86.
- Rao, C.R., and Toutenburg, H., 1999. *Linear Models: Least Squares and Alternatives*, 2nd ed. Springer, New York, 427 p.
- Rencher, A.C., and Schaalje, G.B., 2007. *Linear Models in Statistics*, 2nd ed, John Wiley and Sons, Inc, Hoboken, New Jersey, 672 p.
- Roy, S.S., and Guria, S., 2004. Regression Diagnostics in an Autocorrelated Model. *Brazilian Journal of Probability and Statistics*, 18: 103-112.
- Seber, G.A.F., and Lee, A.J., 2003. *Linear Regression Analysis*, 2nd ed, John Wiley and Sons Inc, Hoboken, New Jersey, 557 p.
- Singh, B., Chaubey, U.P., and Dwivedi, T. D., 1986. An Almost Unbiased Ridge Estimator, *Sankhya Ser.B* 48, 342-346.
- Steece, B. M., 1986. Regressor Space Outliers in Ridge Regression. *Commun. Statist.- Theor. Meth.* 15(12): 3599-3605.
- Stemann, D., and Trenkler, G., 1993. Leverage and Cochrane-Orcutt Estimation in Linear Regression. *Commun. Statist-Theor. Meth.* 22(5): 1315-1333.

- Tarpey, T., 2000. A Note on the Prediction Sum of Squares Statistic for Restricted Least Squares. *The American Statistician*, 54(2): 116-118.
- Theil, H., 1971. *Principles of Econometrics*. Wiley, New York, 768 p.
- Theil, H., and Nagar, A.L., 1961. Testing the Independence of Regression Disturbances. *Journal of the American Statistical Association*, 56: 793-806.
- Trenkler, G., 1978. An Iteration Estimators for The Linear Model. In: *Compstat*, Physica-Verlag, 125- 131.
- Trenkler, G., 1984. On the Performance of Biased Estimators in the Linear Regression Model with Correlated or Heteroscedastic Errors. *Journal of Econometrics*, 25: 179-190.
- Tripp, R.E., 1983. Nonstochastic Ridge Regression and Effective Rank of the Regressors Matrix. (unpublished Ph.D. dissertation). Virginia Polytechnic Institute and State University, Dept. of Statistics.
- Velleman, P.F., and Welsch, R.E., 1981. Efficient Computing of Regression Diagnostics. *The American Statistician*, 35: 234-242.
- Walker, E., and Birch, J.B., 1988. Influence Measures in Ridge Regression. *Technometrics*, 30(2): 221-227.
- Weisberg, S., 1985. *Applied Linear Regression*, 2nd ed. Wiley, New York, 310 p.
- Welsch, R.E., and Kuh, E., 1977. *Linear Regression Diagnostics*. Technical Report 923-77, Sloan School of Management, Cambridge, MA.

ÖZGEÇMİŞ

31.07.1984'de Amasya'da doğdu. İlk, orta ve lise öğrenimini Amasya'da tamamladı. 2004 yılında Ondokuz Mayıs Üniversitesi Sinop Fen Edebiyat Fakültesi İstatistik bölümünü kazandı. 2007 yılında Çukurova Üniversitesi Fen Edebiyat Fakültesi İstatistik bölümüne yatay geçiş yaptı ve 2008 yılında bu bölümden mezun oldu. 2008-2011 yılları arasında Çukurova Üniversitesi Fen Bilimleri Enstitüsü İstatistik Anabilim Dalı yüksek lisans programını bitirdi. 2011-2012 güz yarıyılında Çukurova Üniversitesi Fen Bilimleri Enstitüsü İstatistik Anabilim Dalında doktora programına kayıt yaptırdı. Aralık 2011'de Çanakkale Onsekiz Mart Üniversitesi Fen Bilimleri Enstitüsünde araştırma görevlisi olmaya hak kazandı. Halen Çanakkale Onsekiz Mart Üniversitesi Fen-Edebiyat Fakültesi İstatistik bölümünde araştırma görevlisi olarak görev yapmaktadır.



EKLER



Lemma 1 (Sherman- Morrison- Woodbury, SMV): A matrisi $p \times p$ boyutunda singüler olmayan bir matris olsun. a , A matrisinin p boyutlu bir satır vektörünü göstere. $(A - aa')^{-1}$ matrisi aşağıdaki eşitlik yardımı ile hesaplanmaktadır (Myers, 1990)

$$(A - aa')^{-1} = A^{-1} + \frac{A^{-1}aa'A^{-1}}{1 - a'A^{-1}a}.$$

Lemma 2 (Rao ve Toutenburg, 1999): A ($m \times m$) tipinde singüler olmayan pozitif tanımlı bir matris, B ve C ($m \times n$) tipinde matrisler olmak üzere $\det(A - BC') = \det(A)(I - CA^{-1}B)$ eşitliği vardır.

Lemma 3 (Rencher ve Schaalje, 2007, s. 117): $y \sim N_p(0, \sigma^2 I_n)$ olsun. A , $\text{rank}(A) = r$ ile sabitlerin simetrik matrisi olmak üzere

$$y'Ay \sim \chi_r^2,$$

dağılıması için gerek ve yeter koşul A 'nın idempotent bir matris olmasıdır.

Lemma 4 (Rencher ve Schaalje, 2007, s. 117): $y \sim N_p(\mu, \Sigma)$ olsun. A , $\text{rank}(A) = r$ ile sabitlerin simetrik matrisi olmak üzere

$$y'Ay \sim \chi_{r,\xi}^2, \quad \xi = \frac{1}{2} \mu'A\mu$$

dağılıması için gerek ve yeter koşul $A\Sigma'$ 'nin karesel matris olmasıdır.

Lemma 5 (Graybill, 1983): $S, \Omega = \begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{pmatrix}$ şeklinde parçalanmış ($n \times n$) tipinde bir matris olsun. Burada $\Omega_{ij}; n_i \times n_j, i, j = 1, 2, n_1 + n_2 = n$ 'dir.

Eğer Ω singüler olmayan bir matris ise $\det(\Omega) = \det(\Omega_{11}) \det(\Omega_{22} - \Omega_{21}\Omega_{11}^{-1}\Omega_{12})$ 'dir.

Lemma 6 A ($n \times n$) tipinde matris, a bir sabit olmak üzere $\det(aA) = a^n \det(A)$ dır.

Bir kare matrisin çarpmaya göre tersinin determinanı, determinantının tersine eşittir

$$\det(A^{-1}) = \frac{1}{\det(A)}, \det(A) \neq 0.$$

Teorem-1. s ; σ 'nin GEKK tahmin edicisi olmak üzere, i -nci gözlem silindiğinde rezidü standart sapmasının tahmini

$$s_{(i)} = \sqrt{\frac{(n-p)s^2 - \frac{e_i^2}{1-h_{i,i}}}{n-p-1}}$$

olarak elde edilmektedir.

İspat Teorem-1

n gözlemin tümü veri kümesinde iken rezidü varyansı

$$s^2 = \sum_{i=1}^n \frac{(y_i - \hat{y}_i^*)^2}{n-p}$$

iken i -nci gözlem veri kümesinden çıkartıldığında kalan $(n-1)$ gözlem üzerinden elde edilen rezidü varyansı

$$s_{(i)}^2 = \sum_{\substack{j=1 \\ j \neq i}}^n \frac{(y_j - z_j^* \hat{\beta}_{(i)})^2}{n-p} \quad (*)$$

olarak yazılır. (4.2)'de verilen $\hat{\beta}_{(i)}$ (*) eşitliğinde yerine yazılırsa,

$$\begin{aligned} (n-p)s_{(i)}^2 &= \sum_{\substack{j=1 \\ j \neq i}}^n \left(y_j - z_j^* \left[\hat{\beta} - \frac{(Z^{*'}Z^*)^{-1}z_i^{*'}e_i^*}{1-h_{i,i}} \right] \right)^2 \\ &= \sum_{j=1}^n \left[y_j - z_j^* \hat{\beta} + \frac{z_j^*(Z^{*'}Z^*)^{-1}z_i^{*'}e_i^*}{1-h_{i,i}} \right]^2 - \left(y_i - z_i^* \hat{\beta} + \frac{h_{i,i}e_i^*}{1-h_{i,i}} \right)^2 \\ &= \sum_{j=1}^n \left(e_j^* + \frac{h_{i,j}e_i^*}{1-h_{i,i}} \right)^2 - \left(e_i^* + \frac{h_{i,i}e_i^*}{1-h_{i,i}} \right)^2 \\ &= \sum_{j=1}^n \left(e_j^* + \frac{h_{i,j}e_i^*}{1-h_{i,i}} \right)^2 - \left(\frac{e_i^*}{1-h_{i,i}} \right)^2 \\ &= \sum_{j=1}^n e_j^{*2} + \frac{2e_i^*}{1-h_{i,i}} \sum_{j=1}^n e_j^* h_{i,j} + \frac{e_i^{*2}}{(1-h_{i,i})^2} \sum_{j=1}^n h_{i,j}^2 - \frac{e_i^{*2}}{(1-h_{i,i})^2} \end{aligned} \quad (**)$$

elde edilir. $Hy = H\hat{y}^*$ olduğundan dolayı $\sum_{j=1}^n e_j^* h_{i,j} = 0$ yazılır. Buna ek olarak

$H^2 = H$ ve $\sum_{j=1}^n h_{i,j}^2 = h_{i,i}$ olduğundan dolayı (**) eşitliği

$$\begin{aligned} (n-p)s_{(i)}^2 &= \sum_{j=1}^n e_j^{*2} - \frac{e_i^{*2}}{1-h_{i,i}} \\ &= (n-p)s^2 - \frac{e_i^{*2}}{1-h_{i,i}} \end{aligned}$$

yazılır ve böylece

$$s_{(i)}^2 = \frac{(n-p)s^2 - \frac{e_i^{*2}}{1-h_{i,i}}}{(n-p)}$$

elde edilir.