

**ÇUKUROVA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

YÜKSEK LİSANS TEZİ

SEVGİ AKGÜNEŞ KESTİR

**T.C. YÜKSEKÖĞRETİM KURULU
DOKÜMANTASYON MERKEZİ**

REGRESYON TAHMİN EDİCİLERİ VE KARŞILAŞTIRMA ÇALIŞMALARI

112413

MATEMATİK ANABİLİM DALI

ADANA, 2001

ÇUKUROVA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

REGRESYON TAHMİN EDİCİLERİ VE KARŞILAŞTIRMA
ÇALIŞMALARI

SEVGİ AKGÜNEŞ KESTİR
YÜKSEK LİSANS TEZİ
MATEMATİK ANABİLİM DALI

Bu tez 24/09/2001 Tarihinde Aşağıdaki Jüri Üyeleri Tarafından Oybirliği ile
Kabul Edilmiştir.

İmza.....

Doç.Dr.Selâhattin KAÇIRANLAR
DANIŞMAN

İmza.....

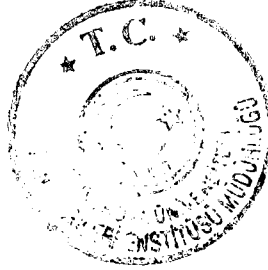
Prof.Dr. Altan ÇABUK
ÜYE

İmza.....

Doç.Dr. Fadullah SAKALLIOĞLU
ÜYE

Bu tez Enstitümüz Matematik Anabilim Dalında hazırlanmıştır.

Kod No: 1903



Prof.Dr. MELİH BORAL
Enstitü Müdürü
İmza ve Mühür

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

İÇİNDEKİLER

Sayfa No

| | |
|----------------------------------------------------------------------------|-----|
| ÖZ..... | I |
| ABSTRACT..... | II |
| TEŞEKKÜR..... | III |
| | |
| 1. GİRİŞ..... | 1 |
| 1.1. Temel Tanımlar, Karşılaştırma Kriterleri..... | 2 |
| 1.2. Modeller..... | 9 |
| 1.2.1. Çoklu Lineer Regresyon Modeli..... | 9 |
| 1.2.2. Standartlaştırılmış Model..... | 9 |
| 1.2.3. Kanonik Model..... | 10 |
| | |
| 2. REGRESYON TAHMİN EDİCİLERİ..... | 11 |
| 2.1. Alışılmış En Küçük Kareler (OLS) Tahmin Edicisi..... | 11 |
| 2.2. Kısıtlı En Küçük Kareler (RLS) Tahmin Edicisi..... | 16 |
| 2.3. Genelleştirilmiş En Küçük Kareler (GLS) Tahmin Edicisi..... | 20 |
| 2.4. Stein Tahmin Edicisi..... | 20 |
| 2.5. Ridge Regresyon (ORR) Tahmin Edicisi..... | 22 |
| 2.6. Kısıtlı Ridge Regresyon (RRR) Tahmin Edicisi..... | 37 |
| 2.7. Genelleştirilmiş Ridge Regresyon (GRR) Tahmin Edicisi..... | 39 |
| 2.8. Hemen Hemen Yansız Tahmin Ediciler Ailesi..... | 41 |
| 2.9. Hemen Hemen Yansız Ridge Regresyon (AURR) Tahmin Edicisi..... | 44 |
| 2.10. Ön Bilgiye Dayalı Ridge Regresyon Tahmin Edicisi..... | 48 |
| 2.11. Küresel Olmayan Hatalı Lineer Modellerde Ridge Tahmin Edicileri..... | 60 |
| 2.12. Shrunken LS Tahmin Edici..... | 62 |
| 2.13. Bayes Tahmin Edicisi..... | 62 |
| 2.14. Preliminary Test Tahmin Edicisi..... | 65 |
| 2.15. Temel Bileşenler Tahmin Edicisi..... | 67 |

| | |
|---------------------------------------------------------------------------------------------------------------------------|-----|
| 2.16. Mixed Tahmin Edicisi..... | 68 |
| 2.17. Minimax Tahmin Edicisi..... | 69 |
| 2.18. Lineer Minimax Tahmin Edicisi..... | 70 |
| 2.19. Liu Tahmin Edicisi..... | 71 |
| 2.20. Hemen Hemen Yansız Genelleştirilmiş Liu Tahmin Edicisi..... | 78 |
| | |
| 3. TAHMİN EDİCİLERİN KARŞILAŞTIRILMASI..... | 80 |
| | |
| 3.1. MSE Altında Regresyon Tahmin Edicilerinin Karşılaştırılması..... | 80 |
| 3.2. OLS ve PC Tahmin edicilerinin Karşılaştırılması..... | 85 |
| 3.3. OLS ve ORR Tahmin Edicilerinin Karşılaştırılması..... | 86 |
| 3.4. ORR Tahmin Edicilerinin Karşılaştırılması..... | 89 |
| 3.5. PC Tahmin Edicilerinin Karşılaştırılması..... | 94 |
| 3.6. ORR ve PC Tahmin Edicilerinin Karşılaştırılması..... | 96 |
| 3.7. OLS ve RLS Tahmin Edicilerinin Karşılaştırılması..... | 99 |
| 3.8. ORR ve OLS Tahmin Edicilerinin Karşılaştırılması..... | 107 |
| 3.9. RLS ve ORR Tahmin Edicilerinin Karşılaştırılması..... | 111 |
| 3.10. AUGRR ve OLS Tahmin Edicilerinin Karşılaştırılması..... | 116 |
| 3.11. RRR Tahmin Edicisinin MSE Kriterine Göre Performansı..... | 121 |
| 3.12. Modified Ridge Regresyon Tahmin Edicisi ile Kısıtlanmış Ridge Regresyon Tahmin Edicisinin Karşılaştırılması..... | 127 |
| 3.13. Gauss-Markoff Modelindeki Minimum Yanlılığa Sahip En İyi Lineer Tahmin Ediciler..... | 133 |
| 3.14. Lineer Tahmin Edicilerin Karşılaştırılması..... | 137 |
| 3.15. OLS ile İterasyon Tahmin Edicilerinin Karşılaştırılması..... | 140 |
| 3.16. ORR ile İterasyon Tahmin Edicilerinin Karşılaştırılması..... | 141 |
| 3.17. Shrunken Tahmin Edici ile İterasyon Tahmin Edicinin Karşılaştırılması..... | 142 |
| 3.18. PC ile İterasyon Tahmin Edicinin Karşılaştırılması..... | 144 |
| 3.19. Homojen Lineer Tahmin Edicilerin Karşılaştırılması..... | 147 |
| 3.20. Minimum Risk Tahmin Edicileri..... | 151 |

| | |
|---------------------------------------------------------------------------------------------------------------|-----|
| 3.21. GLS ile Minimum Risk Tahmin Edicilerinin Karşılaştırılması..... | 152 |
| 3.22. MCMSE ile MARLE nin Karşılaştırılması..... | 154 |
| 3.23. Heterojen Lineer Tahmin Edicilerin Karşılaştırılması..... | 156 |
| 3.24. Lineer Regresyonda En İyi Homojen ve Heterojen Tahmin Edicilerin Karşılaştırılması..... | 169 |
| 3.25. Bir Regresyon Modelinde Stokastik Olarak Kısıtlanmış Lineer Tahmin Edicilerin Karşılaştırılması..... | 176 |
| 3.26. Lineer Modellerde Mixed ve Minimax Tahmin Edicilerinin Karşılaştırılması..... | 181 |
| 3.27. Lineer Regresyonda Admissible ve Minimax Lineer Tahmin Ediciler Arasındaki İlişki..... | 190 |
| 3.28. Mixed Regresyon Tahmin Edicilerinin Karşılaştırılması..... | 196 |
| 3.29. Genelleştirilmiş Liu Tahmin Edicisinin Bilinen Liu Tahmin Edicisine Üstünlüğü..... | 200 |
| KAYNAKLAR..... | 205 |
| ÖZGEÇMİŞ..... | 212 |

ÖZ

YÜKSEK LİSANS TEZİ

**REGRESYON TAHMİN EDİCİLERİ VE KARŞILAŞTIRMA
ÇALIŞMALARI**

SEVGİ AKGÜNEŞ KESTİR

ÇUKUROVA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
MATEMATİK ANABİLİM DALI

Danışman: Doç.Dr. Selahattin KAÇIRANLAR

Yıl: 2001, Sayfa: 212

Jüri : Doç.Dr. Selahattin KAÇIRANLAR

: Prof.Dr. Altan ÇABUK

: Doç.Dr. Sadullah SAKALLIOĞLU

Bu çalışmanın amacı lineer regresyon tahmin edicilerini incelemek, bunlar arasındaki ilişkileri belirlemek ve bunların karşılaştırmalarını inceleyerek hangi durumda hangi tahmin edicinin kullanılacağına karar vermektir.

Anahtar Kelimeler: Çoklu iç ilişki, yanlış tahmin edici, OLS tahmin edici, ridge tahmin edici, hata kareleri ortalaması kriteri.

ABSTRACT

MSc THESIS

REGRESSION ESTIMATORS AND COMPARING STUDIES

SEVGİ AKGÜNEŞ KESTİR

DEPARTMENT OF MATHEMATICS
INSTITUTE OF NATURAL AND APPLIED SCIENCES
UNIVERSITY OF ÇUKUROVA

Supervisor: Doç.Dr. Selahattin KAÇIRANLAR

Year: 2001, Pages: 212

Jury : Doç.Dr. Selahattin KAÇIRANLAR

: Prof.Dr. Altan ÇABUK

: Doç.Dr. Sadullah SAKALLIOĞLU

The aim of this thesis is to investigate linear regression estimators, to determine the relations between this estimators and to decide which estimator will be used under which conditions by examining the comparisons.

Key Words: Multicolinearity, biased estimator, Ordinary Least Squares estimator, Ridge estimator, mean square error criterion.

TEŐEKKÜR

Tez alıőmam boyunca beni ynlendiren , yardımlarını esirgemeyen , sabır gsteren sayın hocam Do. Dr. Selahattin KAIRANLAR' a , Arő. Gr. Orkun COŐKUNTUNCEL' e ve bu uzun alıőmada manevi desteėiyle yanımda olan eőim Fatih'e teőekkr ederim.



1. GİRİŞ

Çoklu lineer regresyon, değişkenler arasında var olan ilişkilerin ortaya çıkarılmasını sağlayan sosyal bilimler, tıp, mühendislik gibi pek çok alanda yaygın bir şekilde kullanılan istatistiksel yöntemlerin en önemlilerinden birisidir. Veri analizi yapan bir araştırmacı tarafından bilim ve teknolojinin hemen hemen her alanında model kurmak için kullanılır. Özellikle son yıllarda bu konuda yapılan çalışmalar hız kazanmıştır. Regresyon katsayılarını tahmin etmek için kullanılan en yaygın yöntem en küçük kareler yöntemidir. Ancak en küçük kareler tahmin edicisinin doğru sonuçlar vermesi bir takım varsayımların sağlanmasına bağlıdır. Veri vektörleri ortogonal olmadığında en küçük kareler tahmini yanıltıcı sonuçlar vermektedir. Ayrıca tahmin etmede kullanılan vektörler arasında çoklu iç ilişki varsa en küçük kareler yöntemi doğru sonuçlara götürmemektedir. Bu nedenle en küçük kareler tahmin edicisine alternatif olarak çok sayıda yanlı tahmin edici tanımlanmıştır. Hoerl ve Kennard (1970,a,b), Stein (1960), Liu (1993), Akdeniz ve Kaçıranlar (1995) ve pek çok araştırmacı yanlı kestiriciler konusunda önemli çalışmalar yapmışlardır.

Çalışmanın birinci bölümünde, tezde kullanılacak olan temel tanım ve teoremler verilmiştir. Ayrıca, çalışma boyunca kullanılacak modeller tanımlanmıştır. İkinci bölümde sırasıyla tahmin ediciler ve bunların istatistiksel özellikleri incelenmiştir. Öncelikle yaygın kullanıma sahip olan en küçük kareler tahmin edicisi incelenmiştir. Sonra kısıtlı en küçük kareler, genelleştirilmiş en küçük kareler, Stein tahmin edicisi incelenip bunların özellikleri verilmiştir. Yanlı tahmin ediciler içerisinde en yaygın kullanıma sahip olan Ridge tahmin edicisi ve son zamanlarda üzerinde fazla çalışılan Liu tahmin edicisi incelenerek bunların istatistiksel özellikleri verilmiştir.

Son olarak üçüncü bölümde yukarıda adı geçen tahmin ediciler ve bunlara ek olarak verdiğimiz birçok tahmin edicinin karşılaştırılması göz önüne alınmıştır. Karşılaştırılan tahmin edicilerin hangi durumlarda daha iyi olduğu teoremlerle ifade edilmiştir.

1.1. Temel Tanımlar

Tanım 1.1.1: A bir simetrik matris ve $Q = x'Ax$ bir karesel form olsun. Sıfırdan farklı her x vektörü için $Q > 0$ ise A ya pozitif tanımlı (p.d.) ; en az bir $x \neq 0$ için $Q = 0$ ve her x için $Q \geq 0$ ise A ya yarı pozitif tanımlı (p.s.d.) matris denir.

Tanım 1.1.2 : A matrisinin negatif olmayan tanımlı (n.n.d.) matris olması için gerek ve yeter koşul A nın p.d. veya p.s.d. matris olmasıdır.

Tanım 1.1.3 : Bir matrisin özdeğerleri ve ortogonal özvektörleri terimleri cinsinden yazımına “ tekil değer ayrışımı” (SVD) denir.

$n \times p$ tipindeki X matrisinin SVD si

$$X = H \Lambda^{\frac{1}{2}} G' \quad (1.1.1)$$

şeklindedir. Burada H $n \times p$ tipinde , $H'H = I$ olacak şekilde standartlaştırılmış , sütunları X in örneklem koordinatlarından oluşan bir matristir. $\Lambda^{\frac{1}{2}}$ ise $p \times p$ tipinde $X'X$ matrisinin $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$ şeklinde sıralı özdeğerlerinin kareköklerinden oluşan köşegen matristir. G matrisi $p \times p$ tipinde $X'X$ in özvektörlerinden oluşan $X'X = G \Lambda G'$ olacak şekildeki matristir.

Tanım 1.1.4: (Tek değişkenli yansızlık) β_i nin b_i tahmin edicisinin beklenen değeri ; β_i bilinmeyen parametresine eşit ise yani $E(b_i) = \beta_i$ ise b_i “ yansız ” olarak adlandırılır. Eğer $E(b_i) \neq \beta_i$ ise b_i yanlıdır ve yanlılığın büyüklüğü $E(b_i) - \beta_i$ ifadesi ile elde edilir.

Tanım 1.1.5: (Tek değişkenli geçerlilik , yeterlilik) Bir yansız tahmin edicinin varyansı diğerinin varyansından daha küçük ise , bu tahmin edici diğerinden daha iyidir (geçerlidir) denir.

b_{i1} ve b_{i2} , β_i nin iki yansız tahmin edicisi olsun .Yani $E(b_{i1}) = E(b_{i2}) = \beta_i$ olsun. Ayrıca

$$Var(b_{i1}) = E(b_{i1} - E(b_{i1}))^2 = E(b_{i1} - \beta_i)^2$$

$$Var(b_{i2}) = E(b_{i2} - \beta_i)^2$$

olsun . Eğer $Var(b_{i1}) < Var(b_{i2})$ ise b_{i1} tahmin edicisi b_{i2} den daha iyidir (geçerlidir) denir. Eğer $Var(b_{i1}) = Var(b_{i2})$ ise her iki tahmin edicinin geçerliliği aynıdır.

Fakat dikkat edilmesi gereken husus varyansların karşılaştırılmasının , sadece yansız tahmin ediciler için kullanılmasıdır. Yanlı tahmin edicilerde sadece varyansın karşılaştırılması anlamsız olur.

Tanım 1.1.6: (Tek değişkenlilerde mse) b_i , β_i nin bir yanlı tahmin edicisi ise hata kareleri ortalaması (mean square error, mse)

$$\begin{aligned} mse(b_i) &= E(b_i - \beta_i)^2 \\ &= E(b_i - E(b_i) + E(b_i) - \beta_i)^2 \\ &= E(b_i - E(b_i))^2 + [E(b_i) - \beta_i]^2 + 2E(b_i - E(b_i))(E(b_i) - \beta_i) \end{aligned}$$

olarak tanımlanır. Üçüncü terim sıfır olduğundan

$$mse(b_i) = Var(b_i) + [Bias(b_i)]^2$$

olarak elde edilir. Burada

$$Bias(b_i) = E(b_i) - \beta_i$$

dir.

Tek Değişkenli MSE Kriteri: b_{i1} ve b_{i2} , β_i nin iki yanlı tahmin edicisi olsun. $MSE(b_{i1}) < MSE(b_{i2})$ ise b_{i1} , b_{i2} ye tercih edilir. Böylece çeşitli tahmin ediciler arasında mse kriterine göre tercih yapılabilir.

Çok Değişkenli Kavramlar

Çok Değişkenlilerde Yansızlık: $p \times 1$ tipindeki β parametre vektörünün b tahmin edici vektörü için $E(b) = \beta$ ise bu b tahmin edicisi “yansız”dır.

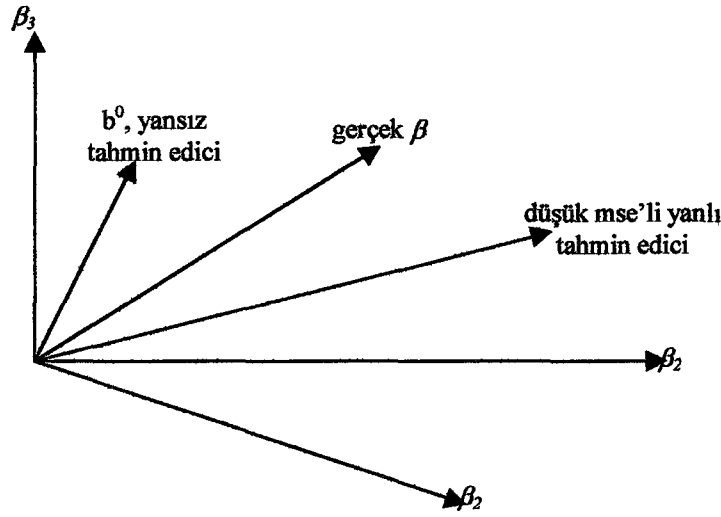
Çok Değişkenlilerde Geçerlilik: b_1 ve b_2 , β nin iki yansız tahmin edicisi ve $Var(b_1)$ ve $Var(b_2)$ bu tahmin edicilerin varyansları olsun. Eğer $Var(b_2) - Var(b_1)$ bir p.s.d matris ise b_1 , b_2 ye göre daha geçerlidir, iyidir.

Çok Değişkenlilerde mse: β nin yanlı tahmin edicisi b nin mse si

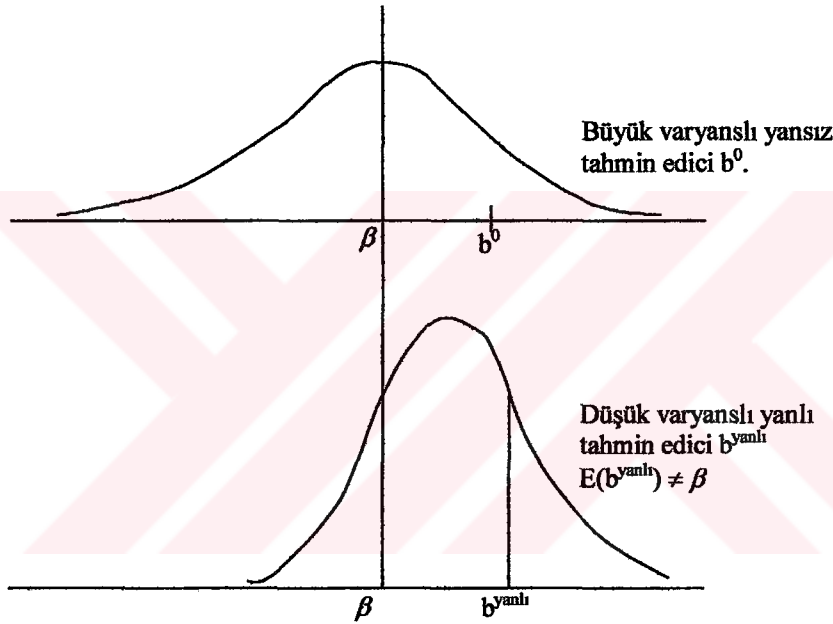
$$\begin{aligned}
 mse(b) &= E[(b - \beta)'(b - \beta)] \\
 &= E[(b - E(b) + E(b) - \beta)'(b - E(b) + E(b) - \beta)] \\
 &= E(b - E(b))'(b - E(b)) + (E(b) - \beta)'(E(b) - \beta) \\
 &= tr[Var(b)] + [Bias(b)]' [Bias(b)]
 \end{aligned}$$

şeklindedir. İstatistikte $E[(b - \beta)'(b - \beta)]$ ifadesine, “kayıp” veya “risk” fonksiyonunun karesinin beklenen değeri denir.

V vektörünün öklid uzunluğunu $\|V\| = (V'V)^{\frac{1}{2}}$ olarak tanımlarsak yukarıdaki ifadedeki $MSE(b)$; b ile β arasındaki öklid uzaklığının karesinin ortalamasıdır. Dolayısıyla düşük mse ye sahip tahmin edici, doğru parametreye daha yakındır.



Şekil 1.1.1. β parametresiyle tahmin edici arasındaki uzaklık.



Şekil 1.1.2.

Şekil (1.1.1) de β ile yanlı tahmin edici arasındaki öklid uzaklığı , β ile b^0 arasındaki öklid uzaklığından daha küçüktür.

Çok Değişkenlilerde Genelleştirilmiş mse: B n.n.d. bir matris olmak üzere b tahmin edicisinin genelleştirilmiş mse'si

$$GMSE(b) = E \left[(b - \beta)' B' (b - \beta) \right]$$

şeklinde tanımlanır.

GMSE Karşılaştırma Kriteri: Gunst ve Mason (1976) , Theobald (1974), Toro-Vizcarrando ve Wallace (1968) ve diğerleri, tahmin ediciler arasında karşılaştırma yapmak için genelleştirilmiş hata kareleri ortalaması (gmse) kriterini tavsiye etmişlerdir. Ancak tahmin edicilerin iyiliğini belirlemek için bu kriter her zaman iyi değildir. Şimdi bu kriteri ve geçerli olduğu koşulları inceleyelim.

$\hat{\beta}_1$ ve $\hat{\beta}_2$ regresyon katsayılarının iki tahmin edicisi olsun . Bu tahmin edicilerin geçerliliklerinin karşılaştırılmasında genellikle kullanılan teknik , performanslarının karesel kayıp fonksiyonları terimleri cinsinden belirlenmesidir.

Theobald (1974) , her karesel kayıp fonksiyonu için $\hat{\beta}_1$ nin $\hat{\beta}_2$ ya tercih edilmesi için gerek ve yeter koşulu, $MSE(\hat{\beta}_i) = E(\hat{\beta}_i - \beta)(\hat{\beta}_i - \beta)'$, $i = 1, 2$ olmak üzere $MSE(\hat{\beta}_2) - MSE(\hat{\beta}_1)$ nin p.s.d. matris olması şeklinde belirtmiştir.

Tahmin edicinin seçilmesinde zor bir yöntem olduğundan bu kriter kullanılırken problemler ortaya çıkabilir. Bu problemlerden biri tüm β ve σ^2 değerleri için $MSE(\hat{\beta}_2) - MSE(\hat{\beta}_1)$ in tanımsız bir matris olmasıdır. Bu durumda $\hat{\beta}_1$ ve $\hat{\beta}_2$ gmse kriteri terimleri cinsinden karşılaştırılmaz . Böyle bir durumda kriter , hangi tahmin edicinin seçilmesi gerektiğini söyleyemez .

Bir başka problem ise kriter tarafından tanımlanan karar kuralı ; tahmin edicilerin ; bilinmediği varsayılan parametrelerin gerçek değerlerindeki özelliklere bağlı olmasıdır. Bu nedenle $MSE(\hat{\beta}_2) - MSE(\hat{\beta}_1)$, β ve σ^2 nin tüm değerleri için p.s.d. olmadıkça , $\hat{\beta}_1$ in $\hat{\beta}_2$ ye bu kriterin terimleri cinsinden tercih edildiği söylenemez . Ayrıca farklı sınıflara ait tahmin ediciler karşılaştırıldığında ; parametre uzayının bir alt kümesi üzerinde bir tahmin edicinin , diğerinden daha iyi olduğu söylenebilir.

$\hat{\beta}_1$ ve $\hat{\beta}_2$ farklı sınıflarda seçilmiş tahmin ediciler olmak üzere $\Omega(\hat{\beta}_1, \hat{\beta}_2)$ tüm β ve σ^2 için $MSE(\hat{\beta}_2) - MSE(\hat{\beta}_1)$ nin bir p.s.d. matris olduğu kümeyi gösterebilir. Böyle bir koşulda iki durum ortaya çıkar . $\Omega(\hat{\beta}_1, \hat{\beta}_2)$ kümesi boş

olmayabilir . Bu durumda Toro-Vizcarrando (1968) in önerdiği gibi bazı istatistiksel testlerle karşılaştırma yapılabilir .

Eğer $\Omega(\hat{\beta}_1, \hat{\beta}_2)$ boş küme ise o zaman parametrelerin hiçbir değeri için $\hat{\beta}_1$, $\hat{\beta}_2$ ya tercih edilemez . $\hat{\beta}_1$ nin , $\hat{\beta}_2$ yerine seçimi bu kritere dayalı olarak yapılamaz. Ancak bu $\hat{\beta}_2$ nin aynı kritere göre $\hat{\beta}_1$ ya tercih edilebileceğini göstermez. Çünkü tahmin ediciler parametre değerlerinin hiçbiri için karşılaştırılmaz.

Predictive mse (Tahmin Edici mse): Tahmin edici mse (predictive mse)

$$\begin{aligned} PMSE(b) &= E(Y - Xb)'(Y - Xb) \\ &= E(b - \beta)' X'X(b - \beta) \end{aligned}$$

şeklinde tanımlanır.

$Y = X\beta + \varepsilon$ modeline göre Y nin ; değişkenlerin r tane yeni değerine karşılık gelen tahmini; $r \times p$ tipindeki X_f matrisi ile gösterilir. Bu matris $X_f\beta$ olarak da gösterilir.

Eğer yeni değerler , X değerlerinin bir tekrarı gibi ise , Y nin gerçek tahmini $X_f\beta$ dir. β yerine herhangi bir b tahmin edicisi yazıldığında , herhangi bir X_f için tahmin edici mse

$$\begin{aligned} PMSE(b, X_f) &= E(Y - X_f b)'(Y - X_f b) \\ &= E(b - \beta)' X_f' X_f (b - \beta) \end{aligned}$$

şeklinde tanımlanır.

Matris mse: Matris mse $MSE(b) = E(b - \beta)(b - \beta)'$ şeklindeki $p \times p$ tipindeki bir matrisle tanımlanır. Bu ifade ayrıca

$$\begin{aligned} MSE(b) &= E(b - E(b) + E(b) - \beta)(b - E(b) + E(b) - \beta)' \\ &= E(b - E(b))(b - E(b))' + (E(b) - \beta)(E(b) - \beta)' \end{aligned}$$

$$= \text{Var}(b) + [\text{Bias}(b)][\text{Bias}(b)]'$$

şeklinde yazılabilir.

Eğer her iki tarafın izi (trace) alınırsa

$$\text{tr}(MSE(b)) = \text{mse}(b)$$

$$= E(b - \beta)'(b - \beta)$$

elde edilir.

Theobald (1974) , her p.d. B matrisi için $GMSE(b_1) > GMSE(b_2)$ olması için gerek ve yeter koşulun $MSE(b_1) - MSE(b_2)$ nin p.d. matris olması gerektiğini gösterdi.

Çok Değişkenli mse Kriteri: b_1 ve b_2 tahmin edicilerini ele alalım . Eğer $\Delta = MSE(b_2) - MSE(b_1)$ farkı p.s.d. matris ise o zaman b_1, b_2 ye tercih edilir.

Δ farkının p.s.d. olup olmadığını kontrol etmek yerine ; x sıfırdan farklı herhangi bir sütun vektörü olmak üzere , her x için $x'\Delta x > 0$ olup olmadığını kontrol etmek daha kolaydır.

Eğer her $x \neq 0$ için $x'\Delta x > 0$ ise her x için $MSE(x'b_2) > MSE(x'b_1)$ dir. Öte yandan eğer her $x \neq 0$ için $MSE(x'b_2) > MSE(x'b_1)$ ise o zaman $x'\Delta x > 0$ dir. Çünkü

$$\begin{aligned} x'E(b_1 - \beta)(b_1 - \beta)'x &= \text{tr} \left[E(b_1 - \beta)(b_1 - \beta)'xx' \right] \\ &= E \left[\text{tr}(b_1 - \beta)(b_1 - \beta)'xx' \right] \\ &= E(b_1 - \beta)xx'(b_1 - \beta) \\ &= E(x'b_1 - x'\beta)'(x'b_1 - x'\beta) \\ &= MSE(x'b_1) \end{aligned}$$

ve benzer olarak $x'E(b_2 - \beta)(b_2 - \beta)'x = MSE(x'b_2)$ dir. Dolayısıyla b_1 tahmin edicisinin b_2 tahmin edicisine tercih edilmesi için güçlü bir kriter: $MSE(x'b_2) \geq MSE(x'b_1)$, tüm $x \neq 0$ için şeklinde yazılabilir (Wallace, 1972).

Bu ifade ; $\Delta = MSE(b_2) - MSE(b_1)$ olmak üzere ve “ ≥ 0 ” p.s.d. matrisi belirtmek üzere ; $\Delta \geq 0$ ifadesine eşdeğerdir.

Kesin eşitsizlik için yani $\Delta > 0$ olması için Δ nın p.d. olması gerekmektedir. b_1 in b_2 ye tercih edilmesi için bir zayıf kriter ise $MSE(b_2) \geq MSE(b_1)$ olmasıdır.

1.2. Modeller

1.2.1. Çoklu Lineer Regresyon Modeli

Çoklu lineer regresyon modeli

$$Y = X\beta + \varepsilon, \varepsilon \approx N(0, \sigma^2 I) \quad (1.2.1)$$

şeklinindedir. Burada Y , $n \times 1$ tipinde bağımlı değişkenler üzerinde bir gözlem vektörü; X $n \times p$ tipinde tam kolon ranklı gözlemlerin bir sabit matrisi; β $p \times 1$ tipinde bilinmeyen parametrelerin (regresyon katsayılarının) bir kolon vektörü ve ε $n \times 1$ tipinde hataların vektörüdür.

1.2.2. Standartlaştırılmış Model

Tahmin edicileri hesaplariken $X'X$ matrisinin bazı özellikleri çok önemlidir. Örneğin en küçük kareler tahmin edicisini hesaplamak için $X'X$ matrisinin tersinin hesaplanması gereklidir. $X'X$ matrisinin tersi hesaplanırken;

- 1) $X'X$ sifıra yakın bir determinanta sahip ise
- 2) $X'X$ in elemanları çok farklı büyüklükte ise

ciddi hatalar ortaya çıkar. Yukarıdaki ilk koşul bağımsız değişkenlerin hepsi veya bazıları içilişkiye sahip (bağımlı) iseler ortaya çıkar. Bu sorunu gidermek için bir çözüm, değişkenleri dönüştürmek ve böylece yeniden parametrelendirilmiş regresyon modelini ele almaktır. Bu işlemler sonucunda elde edilen modele standartlaştırılmış formda regresyon modeli denir. Bu modelde artık $X'X$ korelasyon matrisi formunda olup tüm girişleri -1 ile +1 arasındadır ve tersini hesaplamak kolaydır. $X'Y$ ise açıklayıcı değişkenin her biri ile bağımlı değişken arasındaki korelasyon katsayılarının vektörüdür. Böyle bir modelde hesaplanan parametreler de standartlaştırılmış formdadır. Ayrıca standartlaştırılmış modelde elde edilen tahminler ilişkisizdir.

1.2.3. Kanonik Model

$\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_p)$, köşegen elemanları $X'X$ in özdeğerleri olan bir köşegen matris ve T , $T'X'XT = \Lambda$ ve $TT' = T'T = I_p$ koşullarını sağlayan $X'X$ in özvektörlerinden oluşan bir $p \times p$ tipinde matris olsun. Böylece (1.2.1) deki regresyon modeli

$$Y = Z\alpha + \varepsilon, \varepsilon \approx N(0, \sigma^2 I_n) \quad (1.2.3)$$

şeklinde yeniden yazılabilir. Burada $Z = XT$, $\alpha = T'\beta$ dir. Ayrıca $Z'Z = T'X'XT = \Lambda$ olduğu da kolayca görülebilir. Dolayısıyla $\alpha = T'\beta$ bağıntısından faydalanarak istenildiğinde kanonik formda tahmin edilen parametrelerden orijinal parametrelere geçilebilir.

2. REGRESYON TAHMİN EDİCİLERİ

Bu bölümde literatürde yer alan bazı önemli tahmin ediciler özellikleriyle birlikte incelenecektir.

2.1. Alışılmış En Küçük Kareler (OLS) Tahmin Edicisi

$$Y = \beta_0 + \beta_1 X + \varepsilon \quad (2.1.1)$$

şeklindeki basit regresyon modelini ele alalım. Bu modelde X bağımsız, Y bağımlı değişkendir. Açıkça görüldüğü gibi bu bir doğru denklemdir; dolayısıyla Y bir doğrudur. β_0 eksenin kesim noktası; β_1 eğim ve ε ise rastgele hatadır.

Basit regresyonda hataların sıfır ortalamalı ve σ^2 bilinmeyen varyanslı normal dağılıma sahip oldukları ve bağımsız oldukları varsayılır.

X in; Y bağımlı değişkeni bir rastgele değişken iken, veri analizcisi tarafından kontrol edilen ve ihmal edilebilir hata ile ölçülen bağımsız bir değişken olarak düşünülmesi uygundur. Bunun anlamı Y nin; X in mümkün olan tüm değerlerine karşılık bir olasılık dağılımına sahip olduğudur. Bu dağılımın ortalaması

$$E\left(\frac{Y}{X}\right) = \beta_0 + \beta_1 X \quad (2.1.2)$$

ve varyansı

$$Var\left(\frac{Y}{X}\right) = Var(\beta_0 + \beta_1 X + \varepsilon) = \sigma^2 \quad (2.1.3)$$

dir. Görüldüğü gibi Y nin ortalaması X in bir doğrusal fonksiyonudur. Ancak Y nin varyansı X in değerlerine bağlı değildir. Ayrıca hatalar ilişkisiz olduğundan bağımlı değişkenler de ilişkisizdir.

Parametrelerin En Küçük Kareler Tahmini:

En küçük kareler metodu (OLS) basit regresyon modelindeki β_0 ve β_1 in tahmin edilmesinde kullanılır. β_0 ve β_1 ; Y_i gözlem değerleri ile $Y = \beta_0 + \beta_1 X$ doğrusu arasındaki farkın kareleri toplamları minimum olacak şekilde tahmin edilir.

$(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$ şeklindeki n tane veri için modelimiz,

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \quad i = 1, 2, \dots, n \quad (2.1.4)$$

olsun. Bu durumda LS kriterimiz,

$$S(\beta_0, \beta_1) = \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 X_i)^2 \quad (2.1.5)$$

dir. β_0 ve β_1 in OLS tahmin edicileri $\hat{\beta}_0$ ve $\hat{\beta}_1$ olsun. $\hat{\beta}_0$ ve $\hat{\beta}_1$ nm

$$\begin{aligned} \frac{\partial S}{\partial \beta_0}(\hat{\beta}_0, \hat{\beta}_1) &= -2 \sum_{i=1}^n (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i) = 0 \\ \frac{\partial S}{\partial \beta_1}(\hat{\beta}_0, \hat{\beta}_1) &= -2 \sum_{i=1}^n (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i) X_i = 0 \end{aligned} \quad (2.1.6)$$

denklemlerini sağlaması gerekir. Bu denklemlere “OLS normal denklemleri” denir. Normal denklemler çözüldüğünde,

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$$

ve

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n Y_i X_i - \left(\sum_{i=1}^n X_i \right) \left(\sum_{i=1}^n Y_i \right) / n}{\sum_{i=1}^n X_i^2 - \left(\sum_{i=1}^n X_i \right)^2 / n} \quad (2.1.7)$$

elde edilir.

Görüldüğü gibi $\hat{\beta}_0$ ve $\hat{\beta}_1$; Y_i gözlemlerinin lineer kombinasyonudur. $E(\hat{\beta}_0) = \beta_0$ ve $E(\hat{\beta}_1) = \beta_1$ olduğundan OLS yansız tahmin edicidir.

$\hat{\beta}_0$ ve $\hat{\beta}_1$ nin OLS tahmin edicileri ile ilgili önemli bir sonuç , Gauss-Markov teoremi ile verilebilir. Bu teoreme göre $Y = \beta_0 + \beta_1 X + \varepsilon$ modeli için ; $E(\varepsilon) = 0$, $Var(\varepsilon) = \sigma^2$ ve ilişkisiz hatalar varsayımı altında ; OLS tahmin edicileri ; Y_i lerin lineer kombinasyonu olan diğer tahmin edicilerle karşılaştırıldığında yansızdırlar ve minimum varyansa sahiptirler.

Çoklu Lineer Regresyonda En Küçük Kareler Tahmin Edicisi:

Birden fazla bağımsız değişken içeren regresyon modeline çoklu lineer regresyon modeli denir. k tane bağımsız değişken içeren bir çoklu regresyon modeli

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \varepsilon \quad (2.1.8)$$

şeklindedir. Bu model k boyutlu uzayda bir hiper düzlem belirtir.

n ($n > k$) tane gözlemimiz olsun. (2.1.8) modelini ele alalım. Y_i ; i-nci gözlenmiş bağımlı değişkeni ve X_{ij} ise X_j bağımsız değişkenin i-nci gözlenmiş değeri olmak üzere çoklu regresyon modeli

$$\begin{aligned} Y_i &= \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_k X_{ik} + \varepsilon_i \\ &= \beta_0 + \sum_{j=1}^k \beta_j X_{ij} + \varepsilon_i \quad i = 1, 2, \dots, n \end{aligned} \quad (2.1.9)$$

olur. OLS fonksiyonu

$$\begin{aligned} S(\beta_0, \beta_1, \dots, \beta_k) &= \sum_{i=1}^n \varepsilon_i^2 \\ &= \sum_{i=1}^n \left(Y_i - \beta_0 - \sum_{j=1}^k \beta_j X_{ij} \right)^2 \end{aligned} \quad (2.1.10)$$

dir.

S fonksiyonu $\beta_0, \beta_2, \dots, \beta_k$ ya göre minimize edilmelidir. Böylece $p = k + 1$ tane normal denklem oluşur. Bu denklemlerin çözümü $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k$ şeklindeki OLS tahmin edicileri olacaktır.

LS kriteri

$$S(\beta) = \sum_{i=1}^n \varepsilon_i^2 = \varepsilon' \varepsilon = (Y - X\beta)' (Y - X\beta) \quad (2.1.11)$$

dir ve bu ifadenin minimize edilmesi gerekir. Bunun için türevi alınıp sıfıra eşitlenir ve β yerine $\hat{\beta}$ yazılırsa

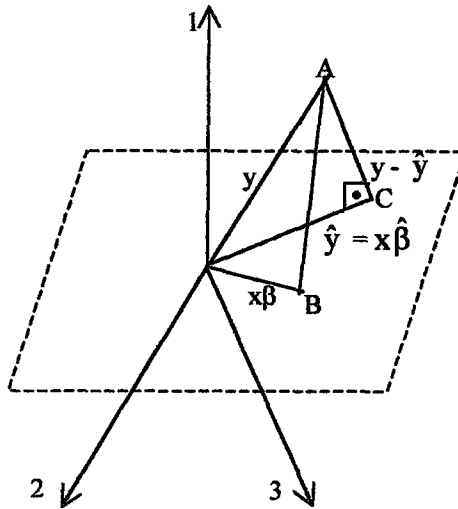
$$X'X\hat{\beta} = X'Y \quad (2.1.12)$$

normal denklemi elde edilir. Bu normal denklemin çözülmesiyle β nun OLS tahmin edicisi

$$\hat{\beta} = (X'X)^{-1} X'Y \quad (2.1.13)$$

şeklinde bulunur.

OLS nin Geometrik Yorumu



Şekil 2.1.1.

$Y = [Y_1, Y_2, \dots, Y_n]$ gözlemlerini , orijinden A noktasına olan bir vektör olarak düşünebiliriz. Y_1, Y_2, \dots, Y_n ler , n boyutlu örneklem uzayının koordinatlarını oluşturmaktadır.

X matrisi p tane $n \times 1$ tipinde sütun vektörlerinden oluşmaktadır. Bu sütunlardan her biri örneklem uzayında orijinden başlayan bir vektör belirtir. Bu p tane vektör, tahmin uzayı adı verilen p boyutlu bir alt uzay oluşturur.

Bu alt uzaydaki herhangi bir noktayı ; $1, X_1, X_2, \dots, X_k$ vektörlerinin bir lineer kombinasyonu olarak ifade edebiliriz. Bu sebepten tahmin uzayındaki her nokta $X\beta$ formundadır.

$X\beta$ vektörü ; B noktasını gösterebilir . B nin A ya olan uzaklığının karesi $S(\beta) = (Y - X\beta)'(Y - X\beta)$ dir. Y gözlem vektörü ile tanımlanan A noktasının tahmin uzayına olan uzaklığının karesinin minimize edilmesi ; tahmin uzayında A ya en yakın noktayı bulmayı gerektirir. Uzaklığın karesi ; tahmin uzayındaki nokta ; A dan tahmin uzayına indirilen dik doğrunun yani normalin , tahmin uzayını kestiği nokta ise minimum olur. Bu nokta c noktasıdır.

Bu nokta $\hat{Y} = X\hat{\beta}$ vektörü ile tanımlanır. Böylece

$$Y - \hat{Y} = Y - X\hat{\beta} \quad (2.1.14)$$

tahmin uzayına dik olduğundan

$$X'(Y - X\hat{\beta}) = 0 \quad (2.1.15)$$

veya

$$X'X\hat{\beta} = X'Y \quad (2.1.16)$$

dir. Böylece LS normal denklemleri elde edilmiş olur.

OLS Özellikleri

(1.2.1) modelini ele alıp OLS tahmin edicisinin yanlılığını inceleyelim.

$$E(\varepsilon) = 0 \text{ ve } (X'X)^{-1} X'X = I \quad (2.1.17)$$

olduğundan

$$\begin{aligned}
 E(\hat{\beta}) &= E[(X'X)^{-1} X'Y] \\
 &= E[(X'X)^{-1} X'(X\beta + \varepsilon)] \\
 &= E[(X'X)^{-1} XX'\beta + (X'X)^{-1} X'\varepsilon] \\
 &= \beta
 \end{aligned} \tag{2.1.18}$$

olur. Bu yüzden $\hat{\beta}$, β nın bir yansız tahmin edicisidir. Şimdi de $\hat{\beta}$ nın varyans-kovaryans matrisini inceleyelim.

$$Cov(\hat{\beta}) = E\left\{ [\hat{\beta} - E(\hat{\beta})][\hat{\beta} - E(\hat{\beta})]'\right\} \tag{2.1.19}$$

şeklindeki varyans-kovaryans matrisi ile varyans özelliği belirtilir. Bu ifadede $Cov(\hat{\beta})$, $p \times p$ tipinde bir simetrik matristir ve j -nci köşegen elemanı $\hat{\beta}_j$ nin varyansını gösterir. Ayrıca ij -nci eleman ise i -nci eleman ile j -nci eleman arasındaki kovaryansı gösterir.

(2.1.19) dan $\hat{\beta}$ nın varyans-kovaryans matrisi

$$Cov(\hat{\beta}) = \sigma^2 (X'X)^{-1} \tag{2.1.20}$$

olur. Eğer $C = (X'X)^{-1}$ olursa $\hat{\beta}_j$ nin varyansı $\sigma^2 c_{jj}$ dir. Ayrıca $\hat{\beta}_i$ ile $\hat{\beta}_j$ arasındaki kovaryans $\sigma^2 c_{ij}$ olur.

β nın OLS tahmin edicisi, β nın “ en iyi lineer yansız tahmin edicisi ” (BLUE) dir. Bundan başka; eğer ε_i hatalarının normal dağılıma sahip olduğunu kabul edersek, $\hat{\beta}$ ayrıca β nın maximum likelihood tahmin edicisi olur. Maximum likelihood tahmin edicisi β nın “ minimum varyanslı yansız tahmin edicisi ” (MVUE) dir

2.2. Kısıtlanmış En Küçük Kareler (RLS) Tahmin Edicisi

Çoklu içilişkililik ile ilgili kaynaklarda en çok üzerinde durulan konu, çoklu içilişkililiğin örneklem varyansını büyütmesi ve anlamlı katsayıların yanlışlıkla modelden çıkarılmasına sebep olmasıdır. Bu problemi çözmek için genellikle iki

metot kullanılır. Bunlar kısıtlanmış en küçük kareler (RLS) metodu ve ridge regresyon metodudur. RLS metodu , tahmin edicinin örneklem varyansının küçültülmesini daha sonra da tahmin edicinin daha küçük hata kareleri ortalamasına (MSE) sahip olmasını sağlar . RLS tahmin metodunda lineer modelin ; modeldeki regresyon katsayıları üzerindeki bazı ön kısıtlamalara göre tahmini yapılır.

Varsayılan kısıtlamalar doğru olduğunda , bu metod OLS tahmin edicisinden daha küçük örneklem varyansına sahip yansız tahmin edici bulunmasını sağlar.

Kısıtlamalar doğru olmadığında örneklem varyansındaki azalma yine sağlanır. Ancak tahmin edici yanlı olur. Belirlenen bağlayıcı kısıtlamalar yanlış olsa bile tahmin edicinin mse sini azaltır.

(1.2.1) deki modeli göz önüne alalım. X deki iki veya daha fazla değişkenin ; çoklu içilişkililik problemi yaratacak kadar ilişkili (bağımlı) olduğunu kabul edelim.

Parametreler üzerindeki kısıtlamalar kümesi

$$R\beta = r \quad (2.2.1)$$

olsun. Burada r , qx1 tipinde bilinen elemanlar vektörü ; R ise qxp tipinde ayrı ayrı parametreler veya parametrelerin bazı lineer kombinasyonlarının yapısı ile ilgili bilgi veren ön bilgi matrisidir.

RLS tahmin edicisinin elde edilmesi için

$$L\left(\beta, \frac{\sigma^2}{Y}, X\right) = L = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left[-\frac{1}{2\sigma^2} (Y - X\beta)' (Y - X\beta)\right] \quad (2.2.2)$$

likelihood fonksiyonunun maksimize edilmesi gerekir. Bu ise hata kareleri toplamının (SSE) yani $(Y - X\beta)' (Y - X\beta)$ nın minimize edilmesiyle olur. Bu işlem

$$\begin{aligned} S^* &= (Y - X\beta)' (Y - X\beta) + \mu'(r - R\beta) \\ &= Y'Y - 2\beta'X'Y + \beta'X'X\beta + \mu'(r - R\beta) \end{aligned} \quad (2.2.3)$$

Lagrange ifadesinde, (2.2.1) kısıtlaması altında yapılır. Bu ifadede μ ; $R\beta = r$ nin m satırındaki m kısıtlama ile ilgili Lagrange çarpanının mx1 tipindeki vektörüdür. (2.2.3) ün β ve μ ye göre türevini alıp bunları sıfıra eşitlediğimizde $\beta = b_R$ için

$$\frac{\partial S^*}{\partial \beta} = 0 = -2X'Y + 2X'Xb_R - R'\mu \quad (2.2.4)$$

$$\frac{\partial S^*}{\partial \mu} = 0 = r - Rb_R \quad (2.2.5)$$

şeklindeki normal denklemleri elde ederiz. b_R ifadesindeki R , β için kısıtlanmış çözümü belirtir ve R matrisi ile ilgisi yoktur.

Normal denklemleri b_R ve μ için çözmek için soldan $(X'X)^{-1}$ ile çarpalım.

$$-2\hat{\beta} + 2b_R - (X'X)^{-1} R'\mu = 0 \quad (2.2.6)$$

olur. Bu ifadede, $\hat{\beta} = (X'X)^{-1} X'Y$ şeklindeki OLS tahmin edicisidir. Bir sonraki adımda (2.2.6) nın soldan R ile çarpılması ve (2.2.5) in kullanılmasıyla ve $R(X'X)^{-1} R'$ nin tersinin olması koşuluyla

$$\mu = 2[R(X'X)^{-1} R']^{-1} (r - Rb_R) \quad (2.2.7)$$

ifadesini elde ederiz. μ nün bu değerini (2.2.6) da yerine yazılması ve b_R ye göre çözülmesiyle

$$b_R = \hat{\beta} - (X'X)^{-1} R'[R(X'X)^{-1} R']^{-1} (R\hat{\beta} - r) \quad (2.2.8)$$

ifadesi elde edilir. Bu ifade kısıtlanmış en küçük kareler (RLS) tahmin edicisidir.

RLS Özellikleri

$R\beta = r$ hipotezi doğru ise yani $R\beta - r = \phi = 0$ ise o zaman kısıtlanmış tahmin edici b_R en iyi lineer yansız tahmin edici olur. Ancak $R\beta = r$ doğru değilse yani $R\beta - r \neq 0$ ise b_R tahmin edicisi yanlı olur.

(2.2.8) ifadesinden

$$E(b_R) = \beta - (X'X)^{-1} R'[R(X'X)^{-1} R']^{-1} \phi \quad (2.2.9)$$

yazabiliriz. Burada $E(\hat{\beta}) = \beta$ olarak yazdık. $(X'X)^{-1} R'[R(X'X)^{-1} R']^{-1} \phi$ terimi sıfır olmadığından b_R yanlıdır. Bazı özel hallerde ; kısıtlamalar kesin olduğunda $R\beta - r = \phi = 0$ dır ve ikinci terim sıfırdır.

Yanlılığa rağmen kısıtlamaların kullanılmasındaki amaç varyansı azaltmaktır. b_R tahmin edicisi bir vektör olduğundan varyans-kovaryans matrisini inceleyelim.

$$\begin{aligned}
 Var(b_R) &= E(b_R - E(b_R))(b_R - E(b_R))' \\
 &= E[\hat{\beta} - \beta - PR(\hat{\beta} - \beta)][\hat{\beta} - \beta - PR(\hat{\beta} - \beta)]' \quad (2.2.10) \\
 &= E(\hat{\beta} - \beta)(\hat{\beta} - \beta)' - E(\hat{\beta} - \beta)(\hat{\beta} - \beta)'R'P' - PRE(\hat{\beta} - \beta)(\hat{\beta} - \beta)' + PRE(\hat{\beta} - \beta)(\hat{\beta} - \beta)'R'P' \\
 &= Var(\hat{\beta}) - \sigma^2 PQP'
 \end{aligned}$$

dir. (2.2.8) ve (2.2.9) dan $b_R - E(b_R) = \hat{\beta} - \beta - PR(\hat{\beta} - \beta)$ olduğunu kullandık.

Ayrıca

$$P = (X'X)^{-1}R'Q^{-1} \text{ ve } Q = R(X'X)^{-1}R' \quad (2.2.11)$$

olarak tanımlanır.

Böylece

$$\begin{aligned}
 Var(\hat{\beta}) &= E(\hat{\beta} - \beta)(\hat{\beta} - \beta)' \\
 &= (X'X)^{-1}X'(E(\varepsilon\varepsilon'))X(X'X)^{-1} \quad (2.2.12) \\
 &= \sigma^2(X'X)^{-1}
 \end{aligned}$$

olur. Bu ifade OLS nin varyans-kovaryans matrisidir.

PQP' nmd olduğundan $Var(\hat{\beta}) - Var(b_R) = a$ p.s.d. olur ve dolayısıyla

$$Var(b_R) \leq Var(\hat{\beta}) \quad (2.2.13)$$

eşitsizliğini elde ederiz. Böylelikle varyans-kovaryans matrisi , kısıtlamaların kullanılmasıyla azaltılmış olur.

2.3. Genelleştirilmiş En Küçük Kareler (GLS) Tahmin Edicisi

$$Y = X\beta + \varepsilon, \varepsilon \approx n(0, \sigma^2 \Omega) \quad (2.3.1)$$

genel lineer modelini ele alalım. Burada Y , $n \times 1$ tipinde bağımlı değişken üzerinde gözlemlerin bir vektörü, X $n \times p$ tipinde tam kolon ranklı gözlemlerin bir sabit matrisi, β $p \times 1$ tipinde bilinmeyen parametrelerin bir kolon vektörü ve ε $n \times 1$ tipinde 0 ortalamalı ve $\sigma^2 \Omega$ varyans-kovaryanslı hataların vektörüdür. Burada Ω simetrik p.d. matristir.

β nm genelleştirilmiş en küçük kareler tahmin edicisi (GLS)

$$\beta^* = (X' \Omega^{-1} X)^{-1} X' \Omega^{-1} Y \quad (2.3.2)$$

şeklindedir. GLS tahmin edicisinin yanlılığı $Bias(\beta^*) = 0$ ve varyansı $Var(\beta^*) = \sigma^2 (X' \Omega^{-1} X)^{-1}$ olur. GLS tahmin edicisinin mse si $MSE(\beta^*) = \sigma^2 (X' \Omega^{-1} X)^{-1}$ olur.

2.4. Stein Tahmin Edicisi

(1.2.1) deki lineer regresyon modelini ele alalım.(1.1.5) deki X in (Tekil Değer Ayrışımı) SVD si kullanılarak lineer regresyon modelini

$$Y = H \Lambda^{\frac{1}{2}} G' \beta + \varepsilon \quad (2.4.1)$$

şeklinde yazabiliriz. Bu modeli ise $X^* = H \Lambda^{\frac{1}{2}}$ ve $\gamma = G' \beta$ olacak şekilde

$$Y = X^* \gamma + \varepsilon \quad (2.4.2)$$

olarak yazabiliriz. (2.4.1) ve (2.4.2) den, $\theta = \Lambda^{\frac{1}{2}} G' \beta$ olmak üzere

$$Y = H\theta + \varepsilon \quad (2.4.3)$$

şeklinde yazılabilir. Bu ifadeyi soldan H' ile çarptığımızda $W = H'Y$ ve $V = H'U$ sıfır ortalamalı ve $\sigma^2 I$ varyansa sahip olmak üzere

$$W = \theta + V \quad (2.4.4)$$

eşitliğini elde ederiz. $\beta = G\Lambda^{-\frac{1}{2}}\theta$ olduğundan β nın Stein-rule tahmin edicisi ; $c > 0$ herhangi bir sabit olmak üzere

$$\begin{aligned} \hat{\beta}_s &= G\Lambda^{-\frac{1}{2}}\hat{\theta}_s \\ &= G\Lambda^{-\frac{1}{2}}\left[1 - \frac{c\sigma^2}{W'W}\right]W \\ &= \left[1 - \frac{c\sigma^2}{W'W}\right]G\Lambda^{-\frac{1}{2}}W \\ &= \left[1 - \frac{c\sigma^2}{b'X'Xb}\right]b \end{aligned} \quad (2.4.5)$$

şeklindedir.

$$\begin{aligned} b &= G\Lambda^{-\frac{1}{2}}W \\ b &= (X'X)^{-1}X'Y \end{aligned} \quad (2.4.6)$$

β nın OLS tahmin edicisidir. Ayrıca

$$W = \Lambda^{\frac{1}{2}}G'b \quad (2.4.7)$$

$$W'W = b'G\Lambda G'b = b'X'Xb \quad (2.4.8)$$

dir.

β nın James-Stein tahmin edicisi ; c bir sabit olmak üzere

$$\begin{aligned} \hat{\beta}_{JS} &= G\Lambda^{-\frac{1}{2}}\hat{\theta}_{JS} \\ &= \left[1 - \left(\frac{c}{t}\right) \frac{\hat{\varepsilon}'\hat{\varepsilon}}{b'X'Xb}\right]b, \quad t = n - p \end{aligned} \quad (2.4.9)$$

şeklindedir. Ayrıca σ^2 yerine , yansız tahmin edicisi

$$s^2 = \frac{\hat{\varepsilon}'\hat{\varepsilon}}{t} = \frac{(Y - Xb)'(Y - Xb)}{t} \quad (2.4.10)$$

yazılır.

Stein-Rule Tahmin Edicisinin Özellikleri

Stein-rule tahmin edicisinin yanlılığı

$$\text{Bias}(\hat{\beta}_{SR}) = E(\hat{\beta}_{SR} - \beta) = \frac{tc\sigma^2}{\beta'X'X\beta} \beta \quad (2.4.11)$$

olur.MSE matrisi

$$\begin{aligned} \text{MSE}(\hat{\beta}_{SR}) &= E(\hat{\beta}_{SR} - \beta)(\hat{\beta}_{SR} - \beta)' \\ &= \sigma^2(X'X)^{-1} + \frac{tc\sigma^2}{(\beta'X'X\beta)^2} [\beta\beta'\{4I + c(t+2)\} - 2(\beta'X'X\beta)(X'X)^{-1}] \end{aligned} \quad (2.4.12)$$

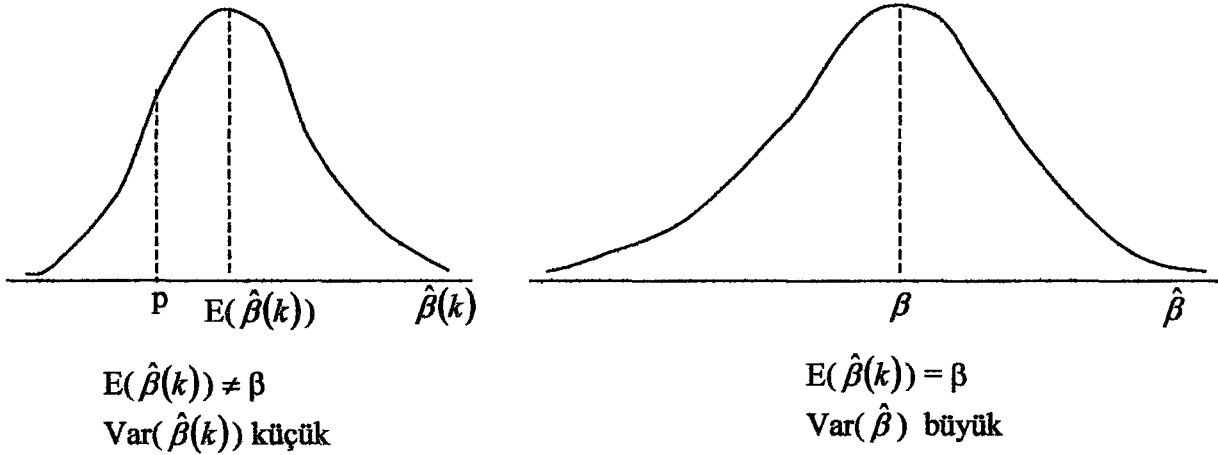
olur. Skaler mse ise

$$\begin{aligned} \text{mse}(\hat{\beta}_{SR}) &= E(\hat{\beta}_{SR} - \beta)'(\hat{\beta}_{SR} - \beta) \\ &= \text{trMSE}(\hat{\beta}_{SR}) \\ &= \sigma^2 \text{tr}(X'X)^{-1} + \frac{tc\sigma^4}{(\beta'X'X\beta)^2} [\beta'\beta[4 + c(t+2)]] - 2(\beta'X'X\beta)\text{tr}(X'X)^{-1} \end{aligned} \quad (2.4.13)$$

olarak bulunur.

2.5. Ridge Regresyon (ORR) Tahmin Edicisi

Ortogonal olmayan verilere LS metodu uygulandığında regresyon katsayılarının uygun tahminleri elde edilmez. Yani çoklu içilişkinin olduğu verilerde OLS tahmin edicisi kullanıldığında büyük örneklem varyansları ortaya çıkar. Bunun



Şekil 2.5.1

Şekil (2.5.1) de, yanlı tahmin edicinin varyansı, yansız tahmin edicinin varyansından küçüktür. Yanlı tahmin edici kullanıldığında β üzerindeki güven aralıkları daha dardır. Yanlı tahmin edicinin varyansının daha küçük olması, ayrıca $\hat{\beta}(k)$ in ; β nin, $\hat{\beta}$ dan daha iyi bir tahmin edicisi olduğunu gösterir.

Hoerl ; OLS tahminlerindeki büyüme ve genel kararsızlık durumlarında

$$\hat{\beta}(k) = [X'X + kI]^{-1} X'Y \quad k \geq 0 \quad (2.5.2)$$

$$= WX'Y \quad (2.5.3)$$

tahmin edicisinin kullanılabilirliğini belirtmiştir.

(2.5.2) ye dayalı olarak yapılan tahmin ve analize “ridge regresyon” adı verilmiştir. Bir ridge tahmini ile bilinen bir tahmin arasındaki ilişki

$$\hat{\beta}(k) = [I_p + k(X'X)^{-1}]^{-1} \hat{\beta} \quad (2.5.4)$$

$$= Z\hat{\beta} \quad (2.5.5)$$

alternatif formu ile verilmiştir. $\hat{\beta}(k)$, W ve Z nin bazı özellikleri aşağıdaki şekildedir.

i) $\varepsilon_i(W)$ ve $\varepsilon_i(Z)$; W ve Z nin özdeğerleri olsun. O zaman λ_i , $X'X$ in özdeğerleri olmak üzere

$$\varepsilon_i(W) = \frac{1}{(\lambda_i + k)} \quad (2.5.6)$$

$$\varepsilon_i(Z) = \frac{\lambda_i}{(\lambda_i + k)} \quad (2.5.7)$$

şeklinde yazılır.

Bu sonuçları W ve Z nin (2.5.3) ve (2.5.5) deki tanımlarından ve $|W - \varepsilon I| = 0$ ve $|Z - \varepsilon I| = 0$ karakteristik denklemlerin çözülmesinden elde edilir.

$$\begin{aligned} \text{ii) } Z &= I - k(X'X + kI)^{-1} & (2.5.8) \\ &= I - kW \end{aligned}$$

Bu bağıntı ise Z nin $Z = (X'X + kI)^{-1} X'X = WX'X$ alternatif formunda yazılıp (2.5.8) deki ifadenin soldan W^{-1} ile çarpılmasıyla elde edilir.

iii) $k \neq 0$ için $\hat{\beta}(k)$; $\hat{\beta}$ dan daha kısadır yani

$$(\hat{\beta}(k))' (\hat{\beta}(k)) < \hat{\beta}' \hat{\beta} \quad (2.5.9)$$

dir. Tanımdan dolayı $\hat{\beta}(k) = Z\hat{\beta}$ dır. Z nin tanımından ve $X'X$ üzerindeki varsayımlardan dolayı Z p.d. simetrik bir matristir. Böylece

$$(\hat{\beta}(k))' (\hat{\beta}(k)) \leq \varepsilon_{\max}^2(Z) \hat{\beta}' \hat{\beta} \quad (2.5.10)$$

bağıntısı elde edilir. λ_1 , $X'X$ in en büyük özdeğeri olmak üzere $\varepsilon_{\max}(Z) = \frac{\lambda_1}{(\lambda_1 + k)}$ şeklinde yazılır ve (2.5.9) daki eşitsizlik oluşturulur. (2.5.7) ve (2.5.8) deki ifadelerden $Z(0) = I$ ve $k \rightarrow \infty$ iken $Z \rightarrow 0$ olduğunu söyleyebiliriz.

$\hat{\beta}(k)$ tahmini için rezidü kareler toplamı

$$\phi^*(k) = (Y - X\hat{\beta}(k))' (Y - X\hat{\beta}(k)) \quad (2.5.11)$$

şeklindedir ve bu ifade

$$\phi^*(k) = Y'Y - (\hat{\beta}(k))' X'Y - k(\hat{\beta}(k))' (\hat{\beta}(k)) \quad (2.5.12)$$

formunda yazılabilir.

Bu ifade gösterir ki ; $\hat{\beta}(k)$ 'nin uzunluğunun karesine dayalı bir değişiklik yapıldığında $\hat{\beta}(k)$ 'nin tüm kareler toplamı $\phi^*(k)$; $\hat{\beta}(k)$ 'nin regresyon kareler toplamından küçük olur.

Ridge İzi

a) Ridge izinin tanımı :

Ridge izi ; $\hat{\beta}(k)$ 'nin elemanlarının k değerlerine karşı oluşturulan izdüşümdür. k genellikle 0-1 aralığında seçilir

Eğer çoklu içilişkililik fazla ise; regresyon katsayılarındaki kararsızlık ridge izinden kolaylıkla görülebilir. k arttığında ridge tahminlerinden bazıları büyük oranlarda değişir. Bazı k değerlerinde $\hat{\beta}(k)$ ridge tahminleri kararlı olur. Objektif olan ; $\hat{\beta}(k)$ ridge tahminlerinin kararlı olduğu küçük k değerlerini seçmektir.

$\hat{\beta}_i(k)$ ve $\hat{\beta}_j(k)$ bağımsız değişkenleri için oluşturulan izdüşüm aşağıdaki gibi olsun

sonucu olarak parametreler üzerindeki güven aralıkları geniş olur; anlamlı regresyon katsayıları modelden çıkarılır ve nokta tahmin kötü olur.

Bu problemin çözümünde genellikle iki tahmin prosedürü kullanılır. Bunlardan biri daha önce verdiğimiz RLS metodu ve diğeri de ridge regresyon (RR) metodudur. RLS metodu daha küçük varyanslı tahmin edici bulunmasını sağlar. RR metodu ise daha küçük MSE ye sahip tahmin edici bulunmasını sağlar.

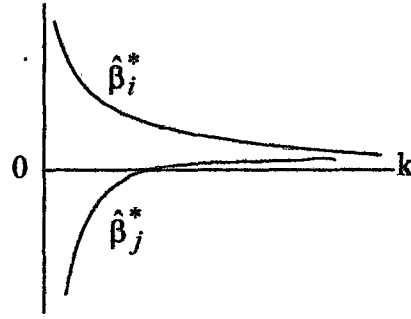
LS metodundaki problem OLS tahmin edicisi $\hat{\beta}$ nın; β nın yansız tahmin edicisi olmasını gerektirmesidir. OLS tahmin edicisi yansız tahmin ediciler arasında minimum varyansa sahip olanıdır. Ancak bazen bu varyans küçük olmayabilir.

Bu problemi çözmek için β nın tahmin edicisinin yansız olması gerekliliğinden vazgeçelim. β nın yansız olan OLS tahmin edicisi $\hat{\beta}$ dan daha küçük varyansa sahip $\hat{\beta}(k)$ gibi bir yanlı tahmin edici bulabileceğimizi varsayalım. $\hat{\beta}(k)$ tahmin edicisinin mse σ ,

$$\begin{aligned} MSE(\hat{\beta}(k)) &= E(\hat{\beta}(k) - \beta)^2 \\ &= tr[Var(\hat{\beta}(k))] + [E(\hat{\beta}(k)) - \beta]^2 \\ &= tr[Var(\hat{\beta}(k))] + [Bias(\hat{\beta}(k))]^2 \end{aligned} \quad (2.5.1)$$

şeklinde tanımlanır.

mse , $\hat{\beta}(k)$ nın β dan olan uzaklığının karesinin beklenen değeridir. $\hat{\beta}(k)$ da küçük bir miktar yanlılığa müsaade ederek $\hat{\beta}(k)$ ın varyansı daha küçük yapılabilir ; öyle ki $\hat{\beta}(k)$ nın mse si, yansız tahmin edici $\hat{\beta}$ MSE'sinden küçüktür.



Şekil 2.5.2

Bu şekilde $\hat{\beta}_i(k)$ ve $\hat{\beta}_j(k)$ zıt yönde hareket etmektedirler. Bunun anlamı bu iki değişken arasında negatif korelasyon vardır ve korelasyon katsayısı büyüktür. Dolayısıyla ridge izindeki kararsızlık; çoklu iç ilişkinin neden olduğu korelasyondan kaynaklanmaktadır. Ridge çözümlerinin ; ridge izinin kararsız olduğu bölgelerde çoklu iç ilişkiden etkilendiği açıktır.

Hoerl ve Kennard ; yanlılık parametresi k nin uygun bir değerinin, ridge izinin incelenmesiyle elde edilebileceğini belirtmişlerdir.

b) Ridge İzinin Karakterizasyonu :

B , β vektörünün herhangi bir tahmini olsun. O zaman rezidü kareler toplamı

$$\begin{aligned}\phi &= (Y - XB)'(Y - XB) \\ &= (Y - X\hat{\beta})'(Y - X\hat{\beta}) + (B - \hat{\beta})'X'X(B - \hat{\beta}) \\ &= \phi_{\min} + \phi(B)\end{aligned}\tag{2.5.13}$$

şeklinde yazılabilir.

ϕ sabitinin oluşturduğu şekil; $\hat{\beta}$ merkezli hiperelipsoidlerin (yani β nin OLS tahmininin) yüzeyidir.

ϕ nin değeri ; ϕ_{\min} ifadesine $(B - \hat{\beta})$ daki karesel formun değerinin eklenmesiyle elde edilir.

$\phi_0 > 0$ belirli bir artış olmak üzere B_0 in $\phi = \phi_{\min} + \phi_0$ bağıntısını sağlayan değerleri vardır. Bununla beraber $X'X$ in küçük bir özdeğeri varsa $\hat{\beta}$ nin β ya olan uzaklığı artacaktır. Ayrıca $X'X$ in içlişkililiği arttıkça, $\hat{\beta}$ nin uzunluğu büyüyecektir. Öte yandan iç ilişkililik arttıkça; rezidü kareler toplamında belirgin bir artış olmadan, $\hat{\beta}$ dan hareket edilebilir

Kareler toplamının minimum yapılması düşüncesinden hareket edilirse, amaç regresyon vektörünün uzunluğunun küçültülmesi yönünde olmalıdır.

Ridge izi belirli bir ϕ değeri için seçilen B minimum olacak şekilde kareler toplamı yüzeyine doğru izlenecek yol olarak tanımlanabilir.

Bu ise şu şekilde yazılabilir : $B'B$ yi

$$(B - \hat{\beta})' X'X (B - \hat{\beta}) = \phi_0 \quad (2.5.14)$$

kısıtlamasına göre minimize edelim. Bir Lagrange problemi olarak ; $\left(\frac{1}{k}\right)$ Lagrange çarpanı olmak üzere

$$F = B'B + \left(\frac{1}{k}\right) \left[(B - \hat{\beta})' X'X (B - \hat{\beta}) - \phi_0 \right] \quad (2.5.15)$$

eşitliğinin minimize edilmesi gerekir. Bu durumda

$$\frac{\partial F}{\partial B} = 2B + \frac{1}{k} \left[2(X'X)B - 2(X'X)\hat{\beta} \right] = 0 \quad (2.5.16)$$

olur. Bu ifade, k (2.5.14) deki kısıtlamayı sağlamak üzere

$$B = \hat{\beta}(k) = (X'X + kI)^{-1} X'Y \quad (2.5.17)$$

bağıntısına indirgenir. Pratikte $k \geq 0$ seçilmesi ve böylece ϕ_0 in hesaplanması daha kolaydır. $\hat{\beta}(k)$ terimleri cinsinden rezidü kareler toplamı.

$$\begin{aligned}\phi^*(k) &= (Y - X\hat{\beta}(k))' (Y - X\hat{\beta}(k)) \\ &= \phi_{\min} + k^2 \hat{\beta}(k)' (X'X)^{-1} \hat{\beta}(k)\end{aligned}\quad (2.5.18)$$

şeklinde bulunur.

Bu yolun tamamen eşdeğeri olan durum şu şekildedir :

Eğer R^2 de regresyon vektörü B nin uzunluğu belirli ise o zaman $\hat{\beta}(k)$; B nin minimum kareler toplamını veren değeridir. Yani $\hat{\beta}(k)$; B nin

$$F_1 = (Y - XB)' (Y - XB) + \left(\frac{1}{k}\right) (B'B - R^2) \quad (2.5.19)$$

fonksiyonunu minimize eden değerdir.

c) Ridge İzinin Likelihood Karakterizasyonu

Hata vektörünün 0 ortalamalı ve $\sigma^2 I$ varyanslı normal dağılıma sahip olduğu varsayımı kullanıldığında likelihood fonksiyonu

$$(2\pi\sigma^2)^{-n/2} \exp\left[-\frac{1}{2\sigma^2} (Y - X\beta)' (Y - X\beta)\right] \quad (2.5.20)$$

olur. Bu fonksiyonun esası; kuvvettteki

$$(Y - X\beta)' (Y - X\beta) = (Y - X\hat{\beta})' (Y - X\hat{\beta}) + (\beta - \hat{\beta})' X'X (\beta - \hat{\beta}) \quad (2.5.21)$$

şeklinde yazılabilen karesel formdur.

(2.5.13) ve (2.5.21) deki ifadeler gösterir ki ; rezidü kareler toplamındaki artış, likelihood fonksiyonunun değerindeki bir azalmaya denktir.

Eşit likelihood'un oluşturduğu şekil ; $\hat{\beta}$ merkezli hiperelipsoidin yüzeyi üzerindedir. Bu sebeple ridge izi, likelihood uzayına doğru olan (izlenen) yoldur.

Bu durumda neden bu özel durumun önemli olduğu (incelendiği) sorusu akla gelebilir. Kareler toplamı için geçerli olan sebep bunun için de geçerlidir.

Her ne kadar uzun vektörler, kısa vektörler ile aynı likelihood değerlerini verse de her zaman aynı fiziksel anlamda değildirler.

Belirtilen kısıtlama $\hat{\beta}$ nın olası değerleri üzerindedir ve genel lineer modelin formulasyonunda kesin (belirli) değildir.

Ridge Regresyonun Mse Özellikleri

a) Ridge Tahmin Edicisinin Varyansı ve Yanlılığı :

$\hat{\beta}(k)$ 1 ; mse açısından incelemek için öncelikle $E[L_1^2(k)]$ için bir ifade elde etmemiz gerekir.

$$E[L_1^2(k)] = E\left[(\hat{\beta}(k) - \beta)'(\hat{\beta}(k) - \beta)\right] \quad (2.5.22)$$

$$= E\left[(\hat{\beta} - \beta)' Z' Z (\hat{\beta} - \beta)\right] + (Z\beta - \beta)' (Z\beta - \beta) \quad (2.5.23)$$

$$= \sigma^2 \text{tr}(X'X)^{-1} Z'Z + \beta'(Z - I)' (Z - I)\beta \quad (2.5.24)$$

$$= \sigma^2 \left[\text{tr}(X'X + kI)^{-1} + k \text{tr}(X'X + kI)^{-2} \right] + k^2 \beta'(X'X + kI)^{-2} \beta \quad (2.5.25)$$

$$= \sigma^2 \sum_{i=1}^p \frac{\lambda_i}{(\lambda_i + k)^2} + k^2 \beta'(X'X + kI)^{-2} \beta \quad (2.5.26)$$

$$= \gamma_1(k) + \gamma_2(k) \quad (2.5.27)$$

elde edilir.

$\gamma_2(k)$; $Z\hat{\beta}$ nın β ya olan uzaklığının karesidir. $Z=I$ olduğunda $k=0$ olduğundan bu uzaklık sıfır olacaktır. Bu sebepten $\gamma_2(k)$; $\hat{\beta}$ yerine $\hat{\beta}(k)$ kullanıldığı durumdaki yanlışlığın karesi olarak düşünülebilir.

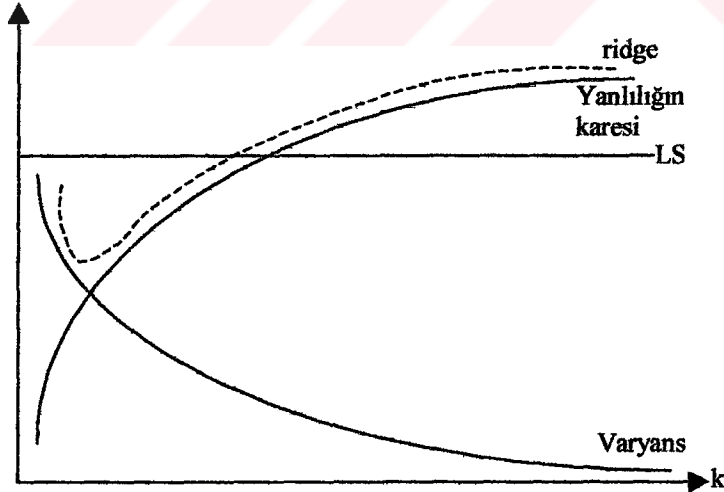
$\gamma_1(k)$ ise, parametre tahminlerinin varyansları toplamıdır. Y rasgele değişkeni terimleri cinsinden

$$\hat{\beta}(k) = Z\hat{\beta} = Z(X'X)^{-1}X'Y \quad (2.5.28)$$

olur. O zaman

$$\begin{aligned} \text{Var}(\hat{\beta}(k)) &= Z(X'X)^{-1}X'\text{Var}(Y)X(X'X)^{-1}Z' \\ &= \sigma^2 Z(X'X)^{-1}Z' \\ &= \sigma^2 (X'X + kI)^{-1}X'X(X'X + kI)^{-1} \end{aligned} \quad (2.5.29)$$

olur. Tüm $\hat{\beta}(k_i)$ lerin varyansları toplamı (2.5.29) daki matrisin köşegen elemanlarının toplamıdır.



Şekil 2.5.3

Şekil (2.5.3) ; varyanslar, yanlılığın karesi ve k parametresi arasındaki ilişkiyi gösterir. k arttıkça toplam varyans azalır. Ayrıca k arttıkça yanlılığın karesi de artar. Noktalı grafik ise $\gamma_1(k)$ ve $\gamma_2(k)$ nin toplamını yani $E[L_1^2(k)]$ yı gösterir. Grafikten ; $\hat{\beta}(k)$ nin mse si , bilinen çözüm $\hat{\beta}$ nin mse sinden daha küçük olacak şekilde k değerlerinin bulunma olasılığının olduğunu söyleyebiliriz. Bu olasılık $\gamma_1(k)$ ve $\gamma_2(k)$ nin matematiksel özellikleri yardımıyla da söylenebilir.

$\gamma_2(k)$ monoton artarken, $\gamma_1(k)$ bir monoton azalan fonksiyondur. Bununla beraber en önemli özellik ise; her bir fonksiyonun orijin komşuluğundaki türevinin değeridir. Bu türevler

$$\lim_{k \rightarrow 0^+} \left(\frac{d\gamma_1}{dk} \right) = -2\sigma^2 \sum_{i=1}^p \left(\frac{1}{\lambda_i^2} \right) \quad (2.5.30)$$

$$\lim_{k \rightarrow 0^+} \left(\frac{d\gamma_2}{dk} \right) = 0 \quad (2.5.31)$$

Böylelikle ortogonal bir $X'X$ için $k \rightarrow 0^+$ iken $\gamma_1(k)$ nin türevi negatiftir ve $-2p\sigma^2$ ye yaklaşır. Ayrıca $X'X$ de içi ilişki olduğunda ve $\lambda_p \rightarrow 0$ iken türev $-\infty$ a yaklaşır.

Öte yandan $\lim_{k \rightarrow 0^+} \left(\frac{d\gamma_2}{dk} \right) = 0$ ifadesi gösterir ki $\gamma_2(k)$ orijinde sıfırdır. Bu özelliklerden şu sonucu varılabilir :

$k > 0$ olarak alınır ve küçük bir yanlılığa müsaade edilirse varyans azaltılabilir; dolayısıyla da tahminin mse si düzeltilebilir.

b) MS Fonksiyonu ile İlgili Teoremler :

Teorem 2.5.1 : $\gamma_1(k)$ toplam varyansı ; k nm bir sürekli ve monoton azalan fonksiyonudur.

Sonuç 2.5.1 : $\gamma_1(k)$ toplam varyansının k ya göre birinci türevi $\gamma_1'(k)$; $k \rightarrow 0^+$ ve $\lambda_p \rightarrow 0$ iken $-\infty$ a yaklaşır.

Teorem 2.5.2 : Yanlılığın karesi $\gamma_2(k)$; k nın bir sürekli ve monoton artan fonksiyonudur.

İspat : (2.5.26) dan $\gamma_2(k) = k^2 \beta'(X'X + kI)^{-2} \beta$ dır. Eğer Λ ; $X'X$ in özdeğerlerinin matrisi ve P ise $X'X = P'\Lambda P$ olacak şekildeki bir ortogonal dönüşüm ise o zaman

$$\alpha = P\beta \quad (2.5.32)$$

olmak üzere

$$\gamma_2(k) = k^2 \sum_{i=1}^p \frac{\alpha_i^2}{(\lambda_i + k)^2} \quad (2.5.33)$$

dir. Tüm i ler için $\lambda_i > 0$ ve $k \geq 0$ olduğundan her bir $\lambda_i + k$ elemanı pozitiftir, ve toplamda tekillik yoktur. Yani toplam sıfırdan büyüktür.

$\gamma_2(0) = 0$ olduğu açıktır. Bu durumda $k \geq 0$ için $\gamma_2(k)$ sürekli bir fonksiyondur. $k > 0$ için (2.5.32) eşitliği

$$\gamma_2(k) = \sum_{i=1}^p \frac{\alpha_i^2}{\left(1 + \frac{\lambda_i}{k}\right)^2} \quad (2.5.34)$$

şeklinde yazılabilir.

Tüm i ler için $\lambda_i > 0$ olduğundan; artan k değerleri için $\frac{\lambda_i}{k}$ nın monoton azalan fonksiyon olduğu açıktır ve $\gamma_2(k)$ nın herbir terimi monoton artandır. Dolayısıyla $\gamma_2(k)$ monoton artandır.

Sonuç 2.5.2 : Yanlılığın karesi $\gamma_2(k)$ bir üst limit olarak $\beta'\beta$ ya yaklaşır.

İspat : (2.5.34) den $\lim_{k \rightarrow \infty} \gamma_2(k) = \sum_{i=1}^p \alpha_i^2 = \alpha'\alpha = \beta'p'p\beta = \beta'\beta$

Sonuç 2.5.3 : $k \rightarrow 0^+$ iken $\gamma_2'(k)$ türevi sifra yaklaşır.

İspat : (2.5.33) den

$$\frac{d\gamma_2(k)}{dk} = 2k \sum_{i=1}^p \frac{\lambda_i \alpha_i^2}{(\lambda_i + k)^3} \quad (2.5.35)$$

eşitliği yazılabilir.

Her terimdeki $\frac{2k\lambda_i\alpha_i^2}{(\lambda_i + k)^3}$ sürekli bir fonksiyondur. $k \rightarrow 0^+$ iken her bir terimin limiti sıfırdır.

Teorem 2.5.3 : $E[L_1^2(k) < E[L_1^2(0)]] = \sigma^2 \sum_{i=1}^p \left(\frac{1}{\lambda_i} \right)$ olacak şekilde her zaman bir $k > 0$ bulunur.

İspat : (2.5.26) , (2.5.33) ve (2.5.35) eşitliklerinden

$$\begin{aligned} \frac{dE[L_1^2(k)]}{dk} &= \frac{d\gamma_1(k)}{dk} + \frac{d\gamma_2(k)}{dk} \\ &= -2\sigma^2 \sum_{i=1}^p \frac{\lambda_i}{(\lambda_i + k)^3} + 2k \sum_{i=1}^p \frac{\lambda_i \alpha_i^2}{(\lambda_i + k)^3} \end{aligned} \quad (2.5.36)$$

$$\gamma_1(0) = \sigma^2 \sum_{i=1}^p \left(\frac{1}{\lambda_i} \right) \text{ ve } \gamma_2(0) = 0$$

olduğu görülür.

Teorem (2.5.1.) ve Teorem (2.5.2) den ; $\gamma_1(k)$ ve $\gamma_2(k)$ nin karşılıklı olarak monoton azaldığı veya karşılıklı olarak monoton arttığını görmüştük. Bunların birinci türevleri her zaman sıfırdan büyük veya eşit ya da sıfırdan küçük veya eşittir. Bu yüzden teoremin ispatı için $\frac{dE[L_1^2(k)]}{dk} < 0$ olacak şekilde her zaman bir $k > 0$ olduğunun gösterilmesi yeterli olacaktır. Bunun için gerekli koşulu ise (2.5.36) dan

$$k < \frac{\sigma^2}{\alpha_{\max}^2} \quad (2.5.37)$$

şeklinde elde ederiz.

c) MSE Fonksiyonu Üzerinde Yorumlar

$E[L_1^2(k)] = \gamma_1(k) + \gamma_2(k)$ nin özellikleri bu ifadenin bir minimumu olacağını gösterir. Ayrıca $k \rightarrow \infty$ iken $\gamma_2(k)$, $\beta'\beta$ limit değerine yaklaştığından $\beta'\beta$ nin değeri büyürken minimum, $k=0$ değerini aldığımda elde edilir.

$\beta'\beta$ bilinmeyen regresyon vektörünün uzunluğunun karesi olduğundan $k \neq 0$ için bir değer seçilmesi ve dolayısıyla $\beta'\beta$ ya bir üst sınır belirlemeden daha küçük mse yi elde etmek imkansızmış gibi görünebilir.

Öte yandan pratikte $\beta'\beta$ nin sonsuz olamayacağı açıktır ve $\hat{\beta}(k)$; β ya $\hat{\beta}$ nin olduğundan daha yakın olacak şekilde bir k değeri veya k değerleri bulunabilir. Bir başka deyişle matematiksel sınırsızlık ile pratikteki sınırsızlık farklı şeylerdir.

Ridge Regresyonunun Bir Genel Formu

$X'X$ bir köşegen matris olacak şekilde genel lineer regresyon problemi her zaman (1.2.3) deki kanonik forma indirgenebilir. Özel olarak $A = (\delta_i, \lambda_i)$ $X'X$ in özdeğerlerinin matrisi olmak üzere $X'X = TAT'$ olacak şekilde bir ortogonal dönüşümü vardır.

$$X = ZT' \quad (2.5.38)$$

ve

$$\alpha = T'\beta, \quad Z'Z = A \quad \text{ve} \quad \alpha'\alpha = \beta'\beta \quad (2.5.39)$$

olmak üzere

$$Y = Z\alpha + \varepsilon \quad (2.5.40)$$

olsun. O zaman $K = (\delta_i, k_i)$ $k_i \geq 0$ olmak üzere genel ridge tahmin prosedürü

$$\alpha^* = (Z'Z + K)^{-1}(Z'Y) \quad (2.5.41)$$

eşitliğinden tanımlanır. Genel formda her kononik form için k_i değerleri Z ile tanımlanmıştır.

$$(L_1^*)^2 = (\hat{\alpha}^* - \alpha)'(\hat{\alpha}^* - \alpha) \quad (2.5.42)$$

şeklinde tanımlanarak k_i leri için optimal değerleri $k_i = \frac{\sigma^2}{\alpha_i^2}$ olduğu görülebilir. Ridge

izinin grafiksel eşdeğeri yoktur. Fakat $k_i = \frac{\hat{\sigma}^2}{\hat{\alpha}_i^2}$ alınarak bir iteratif prosedür

kullanılabilir.

Ridge regresyon tahmin edicisinin bir başka tanımı ise aşağıdaki gibidir.

(1.2.1) deki lineer regresyon modelini ele alalım. $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_p$; $X'X$ in özdeğerleri ve $V'V = VV' = I$ olmak üzere $V'X'XV = \Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_p)$ olsun. Tanımdan dolayı V nin i -nci sütunu , $X'X$ in λ_i özdeğerine karşılık gelen normalleştirilmiş özvektörüdür.

Hoerl ve Kennard'ın ORR tahmin edicisi; D , $p \times p$ tipinde negatif olmayan elemanlardan oluşan bir köşegen matris olmak üzere

$$\hat{\beta}(k) = (X'X + XDV')^{-1} X'Y \quad (2.5.43)$$

şeklinde tanımlanır. Ayrıca D nin enaz bir tane pozitif elemanı olduğunu kabul ediyoruz. Aksi halde $\hat{\beta}(k)$, OLS tahmin edicisine eşit olur. Bu tahmin edici ise yanlıdır.

Bu tahmin edicinin ortalaması

$$(X'X + VDV')^{-1} X'X\beta \quad (2.5.44)$$

ve varyans-kovaryans matrisi

$$\sigma^2 (X'X + VDV')^{-1} X'X (X'X + VDV')^{-1} \quad (2.5.45)$$

şeklindedir.

2.6. Kısıtlanmış Ridge Regresyon (RRR) Tahmin Edicisi

RLS ve ORR tahmin edicilerinin elde edilmesinde kullanılan iki yaklaşımın birleştirilmesiyle yeni bir tahmin edici olan kısıtlanmış ridge regresyon (RRR) tahmin edicisi elde edilir. Bu yaklaşımın kullanılmasındaki asıl neden şudur : parametreler üzerindeki kısıtlamalar doğru olduğunda, orijinal parametrelerin yeniden düzenlenmiş (reparametrize edilmiş) halleri elde edilir. Yeniden düzenlenmiş (reparametrize edilmiş) modelde hala çoklu iç ilişkililik problemi varsa RLS tahmin edicisinin ORR tahmin

edicisi doğrultusunda düzenlenmesi gerekir. Bir başka deyişle ORR düşüncesinin, RLS tahmin edicisine uygulanmasıyla yeni bir tahmin edici olan RRR tahmin edicisi oluşturulur.

(1.2.1.) deki lineer regresyon modelini ele alalım. İki veya daha fazla değişkenin; modelde çoklu iç ilişkililik problemi yaratacak şekilde ilişkili olduğunu kabul edelim.

$\hat{\beta}_{RLS}$; β nın (2.2.8) deki RLS tahmin edicisi ve $S = X'X$ ve $W = (I_p + kS^{-1})^{-1}$ olmak üzere β nın RRR tahmin edicisi

$$\hat{\beta}_{RRR}(k) = W\hat{\beta}_{RLS}, \quad k \geq 0 \quad (2.6.1)$$

şeklinde tanımlanır. (2.6.1) deki RRR tahmin edicisi; belirli bir φ değeri için $\beta'\beta$ nın

$$(\beta - \hat{\beta}_{RLS})' X'X (\beta - \hat{\beta}_{RLS}) = \varphi \quad (2.6.2)$$

ye göre minimize edilmesi ve $\hat{\beta}$ yerine $\hat{\beta}_{RLS}$ yazılmasıyla RR tahmin edicisinde olduğu gibi elde edilir.

$R\beta = r$ kısıtlaması kullanılarak ; $\delta = r - R\beta$ olmak üzere

$$E(\hat{\beta}_{RRR}(k)) = W\beta + WS^{-1}R'(RS^{-1}R')^{-1}\delta \quad (2.6.3)$$

olduğu belirlenebilir.

Böylece $\underline{0}$; $qx1$ tipindeki sıfır vektörü olmak üzere ; $k=0$ ve $\delta = \underline{0}$ olmadıkça β nın RRR tahmin edicisi her zaman bir yanlı tahmin edicidir.

$$S(k) = X'X + kI_p \quad (2.6.4)$$

ve

$$\eta = S^{-1}R'(RS^{-1}R')^{-1}\delta \quad (2.6.5)$$

ve

$$A = S^{-1}R'(RS^{-1}R')^{-1}RS^{-1} \quad (2.6.6)$$

olmak üzere RRR tahmin edicisinin yanlılığı

$$Bias(\hat{\beta}_{RRR}) = -W\eta - kS^{-1}(k)\beta \quad (2.6.7)$$

varyans- kovaryans matrisi

$$Var(\hat{\beta}_{RRR}) = \sigma^2(W'S^{-1}W') - \sigma^2(WAW') \quad (2.6.8)$$

ve mse si

$$MSE(\hat{\beta}_{RRR}) = \sigma^2 tr(W'S^{-1}W') - \sigma^2 tr(WAW') + 2k\eta'W'S^{-1}(k)\beta + \eta'W'W\eta + k^2\beta'S^{-2}(k)\beta \quad (2.6.9)$$

şeklindedir.

2.7. Genelleştirilmiş Ridge Regresyon (GRR) Tahmin Edicisi

Hoerl ve Kennard (1970a) ridge regresyon prosedürünün, her bir bağımsız değişken için ayrı ayrı yanlılık parametresi kullanımını sağlayan bir genişlemesini kullanmışlardır. Genelleştirilmiş ridge regresyon (GRR) ; verinin, ortogonal değişken uzayına dönüşümü ile ilgilenir.

Lineer regresyon modelinin (1.2.3) deki kanonik formunu ele alalım.

α nın OLS tahmin edicisi

$$\begin{aligned} \hat{\alpha}_{OLS} &= (Z'Z)^{-1}Z'Y \\ &= A^{-1}Z'Y \end{aligned} \quad (2.7.1)$$

şeklindedir. α nın GRR tahmin edicisi;

$$\hat{\alpha}_{GRR} = (A + K)^{-1} Z'Y \quad (2.7.2)$$

$$= (A + K)^{-1} A \hat{\alpha}_{OLS} \quad (2.7.3)$$

$$= [I - (A + K)^{-1} K] \hat{\alpha}_{OLS} \quad (2.7.4)$$

olarak bulunur. Burada $i = 1, 2, \dots, p$ için k_i değerleri yanlışlık parametreleri olmak üzere $K = \text{diag}(k_1, k_2, \dots, k_p)$ dir.

$\alpha = T'\beta$ olduğundan β nın GRR tahmin edicisi

$$\hat{\beta}_{GRR} = T' \hat{\alpha}_{GRR} \quad (2.7.5)$$

$$= [(X'X) + TKT']^{-1} X'Y \quad (2.7.6)$$

dir. $\hat{\alpha}_{GRR}$ nin yanlışlığı ise

$$\text{Bias}(\hat{\alpha}_{GRR}) = -(A + K)^{-1} K \alpha \quad (2.7.7)$$

şeklindedir. (2.7.4) den $i = 1, 2, \dots, p$ için α tahmin edicisinin i -nci bileşeni α_i nin mse si

$$\text{mse}(\alpha_i) = \frac{(\alpha_i^2 k_i^2 + \sigma^2 \lambda_i)}{(\lambda_i + k_i)^2} \quad (2.7.8)$$

şeklinde bulunur.

2.8. Hemen Hemen Yansız Regresyon Tahmin Edicileri

OLS tahmin edicisi; basit lineer regresyon modelinde en iyi yansız tahmin edici olmasına rağmen, veride çoklu içilişkililik olduğunda istenmeyen sonuçlar ortaya çıkarabilir. Çoklu iç ilişkililik problemini çözmek için kullanılan yöntemlerden bazıları ridge regresyon tahmin edicisi ve Stein tipi tahmin edicinin kullanılmasıdır.

Yansız Tahmin Ediciler Ailesi

$$Y = X\beta + \varepsilon \quad (2.8.1)$$

modelini ele alalım.

$$E(\varepsilon) = 0, \quad E(\varepsilon\varepsilon') = \sigma^2 I \quad (2.8.2)$$

olsun. β nın OLS tahmin edicisi

$$b = (X'X)^{-1} X'Y \quad (2.8.3)$$

şeklindedir ve b en iyi lineer yansız tahmin edicidir.

Vinod ve Ullah (1981) yanlı tahmin edicilerin genel bir sınıfını ; D p x p tipinde bilinen bir matris ; c bir pozitif sabit ve B verilen bir p.d. matris ve

$$a = \frac{cs}{b' B b'} \quad \text{ve} \quad s = (Y - Xb)' (Y - Xb) \quad (2.8.4)$$

olmak üzere

$$\hat{\beta} = (I + aD)^{-1} b \quad (2.8.5)$$

şeklinde tanımlamıştır.

D nin ve B nin farklı değerlerinin alınmasıyla birçok bilinen tahmin edici (2.8.4) ve (2.8.5) den elde edilebilir. $D=I$ için $\hat{\beta}$ Stein tipi tahmin edici olur.

$D=I$ ve $B = X'X$ olarak alınırsa ; $\hat{\beta}$ Ullah ve Ullah (1978) in bir çift k-sınıfı tahmin edicisi olur. Eğer $D = (X'X)^{-1}$ alınırsa $\hat{\beta} = (X'X + aI)^{-1} X'Y$, Ullah ve ark. (1981) in işlemsel ORR tahmin edicisi ailesi olur. $\hat{\beta}$ tahmin edicisi yanlıdır.

$$n = T - p \quad (2.8.6)$$

olmak üzere

$$E(\hat{\beta} - \beta) = - \left(cn\sigma^2 / \beta'B\beta \right) D\beta \quad (2.8.7)$$

şeklinde verilmiştir.

Şimdi yansız $\tilde{\beta}$ tahmin edicilerin ailesini

$$\tilde{\beta} = \hat{\beta} + \left(cn\sigma^2 / \beta'B\beta \right) D\beta \quad (2.8.8)$$

olarak oluşturalım.

(2.8.7) nin kullanılmasıyla

$$E(\tilde{\beta}) = \beta \quad (2.8.9)$$

olduğu görülür.

δ nin tahmin edicisi olan $\hat{\delta}$ nin skaler mse si Wallace (1972) tarafından

$$trE(\hat{\delta} - \delta)(\hat{\delta} - \delta)' = E(\hat{\delta} - \delta)'(\hat{\delta} - \delta) \quad (2.8.10)$$

şeklinde tanımlanmıştır.

Teorem 2.8.1 : Skaler mse kriterine göre tüm β ve σ^2 değerleri için $MSE(\tilde{\beta}) - MSE(\hat{\beta}) \leq 0$ dir. Vinod ve Ullah (1981), belirli koşullar altında $MSE(\hat{\beta}) \leq MSE(b)$ olduğunu göstermişlerdir. Böylelikle aynı koşullar altında $MSE(\tilde{\beta}) \leq MSE(b)$ olur.

İspat : (2.8.8) in kullanılmasıyla

$$MSE(\tilde{\beta}) = V(\tilde{\beta})$$

$$= E \left[\left(\hat{\beta} - \beta \right) + \left(\frac{cn\sigma^2}{\beta'B\beta} \right) D\beta \right] X \left[\left(\hat{\beta} - \beta \right) + \left(\frac{cn\sigma^2}{\beta'B\beta} \right) D\beta \right]' \quad (2.8.11)$$

yazılabilir. (2.8.7) nin kullanılmasıyla (2.8.11) ifadesi

$$MSE(\tilde{\beta}) = MSE(\hat{\beta}) - \left(\frac{cn\sigma^2}{\beta'B\beta} \right)^2 D\beta\beta'D' \quad (2.8.12)$$

olarak yazılabilir.

(2.8.12) deki eşitliğin her iki tarafının trace'nin alınmasıyla istenen sonuç elde edilir.

Hatırlatmalar

(2.8.4) deki eşitlikle $D=I$ alındığında $\hat{\beta}$ bir Stein tipi tahmin edici olur. Teoremden skaler mse kriterine göre Stein tipi tahmin ediciden daha iyi olan, yanlılığı düzeltilmiş bir Stein tipi tahmin edici elde edilir. Ullah ve ark .(1981), belirli koşullar altında işlemsel ORR tahmin edicisinin OLS tahmin edicisinden daha iyi olduğunu göstermiştir. Yukarıdaki teorem; işlemsel ORR tahmin edicisinden daha iyi bir yanlılığı düzeltilmiş işlemsel ORR tahmin edicisi elde edilebileceğini gösterir.

2.9. Hemen Hemen Yansız Ridge (AURR) Regresyon Tahmin Edicisi

Çoklu regresyon modelinde, bağımsız değişkenler arasında ilişki olduğunda parametrelerin bilinen en küçük kareler, tahmin edicilerinin hesaplanması problemi ortaya çıkar. Kullanılan yansız tahmin edicilerin varyansı çok büyük olur. Bu durumlarda varyansı küçük fakat yanlı tahmin ediciler kullanılır. Bunlardan bir Hoerl ve Kennard in ridge tahmin edicisidir. Fakat bazı durumlarda yanlılık çok fazla olur.

Ridge tahmin edicisinin yanlılığının büyük olduğu durumlarda Jack-Knife prosedürü kullanılarak yanlılık azaltılabilir. Sonuçta elde edilecek tahmin edici ridge tahmin edicisi ile benzer formdadır. Dolayısıyla aynı asimptotik özelliklere sahip olur. Ayrıca Jack-Knife prosedürü; regresyon parametrelerinin güven aralıklarının bulunması için kolay bir metod ortaya koyar. Yalnız dikkat edilmesi gereken husus, Jack-Knife prosedürü, yansız tahmin edicinin uygun olmadığı ve yerine yanlı tahmin edicinin kullanılabildiği durumlarda uygulanır.

Çoklu ilişkinin olduğu durumlarda en küçük kareler tahmin edicisi (aynı zamanda BLUE) kullanılabilir. Fakat iç ilişki probleminden dolayı hesaplamalarda uygun değildir. Ridge tahmin edicisi hesaplamalarda büyük avantaj sağlar.

Ridge tahmin edicisinde Jack-Knife prosedürünü uygulamak için önce geliştirilmiş ridge tahmin edicisini ele alalım. Daha sonra prosedürü uygulayıp yeni tahmin ediciyi elde edelim.

Ridge Tahmin Edicisine Jack-Knife Prosedürünün Uygulanması

Y , $n \times 1$ tipinde gözlemlerden oluşan bağımlı değişken vektörü; X $n \times p$ tipinde

$\lim_{n \rightarrow \infty} \frac{X'X}{n}$ sonlu olan stokhostik olmayan regressörler matrisi; β $p \times 1$ tipinde bilinmeyen

regresyon katsayıları vektörü ve ε $n \times 1$ tipinde bilinen varsayımlara sahip, yani ($E(\varepsilon) = 0$) sıfır ortalamalı ve ($E(\varepsilon\varepsilon') = \sigma^2 I$) $\sigma^2 I$ varyans-kovaryans matrisine sahip olan hata vektörü olmak üzere

$$Y = X\beta + \varepsilon \quad (1.2.1)$$

şeklindeki lineer regresyon modelini ele alalım.

Sütunları, $X'X$ in normalleştirilmiş özvektörleri olan $p \times p$ tipindeki bir T matrisini alalım. O zaman, (1.2.1) lineer regresyon modelini $Z = XT$ ve $\alpha = T'\beta$ olmak üzere

$$Y = Z\alpha + \varepsilon \quad (2.9.1)$$

şeklinde yazabiliriz. λ_i , $X'X$ in i -nci özdeğeri olmak üzere $Z'Z = T'X'XT = \Lambda = \text{diag}(\lambda_1, \lambda_2, \lambda_3, \dots, \lambda_p)$ dir. Λ nın i -nci elemanına (köşegen elemanına) k_i pozitif sabitini eklersek, α nın genelleştirilmiş ridge tahmin edicisini; $K = \text{diag}(k_1, k_2, k_3, \dots, k_p)$

ve $A = \Lambda + K$ olmak üzere

$$\hat{\alpha}_{GRR} = (\Lambda + K)^{-1} Z'Y = A^{-1} Z'Y \quad (2.9.2)$$

şeklinde yazabiliriz.

$\alpha = T'\beta$ ve $TT' = I$ olduğundan β nın genelleştirilmiş ridge tahmin edicisi; $Z'Z = X'X + K_*$ ve $K_* = TKT'$ olmak üzere,

$$\hat{\beta} = G\hat{\alpha}_{GRR} = (Z'Z)^{-1} X'Y \quad (2.9.3)$$

şeklindedir. Bundan başka α nın OLS tahmin edicisi,

$$g = A^{-1} Z'Y \quad (2.9.4)$$

olarak verildiğinden, α nın genelleştirilmiş ridge tahmin edicisini

$$\hat{\alpha}_{GRR} = A^{-1} \Lambda g = (I - A^{-1}K)g \quad (2.9.5)$$

şeklinde yazabiliriz.

$(X'X)^{-1}$ ve dolayısıyla $(Z'Z)^{-1}$ in $\frac{1}{n}$ inci dereceden olması durumunda

$$\text{Bias}(\hat{\alpha}_{GRR}) = -A^{-1}K\alpha \quad (2.9.6)$$

nın derecesi $\frac{1}{n}$ dir. Ayrıca

$$\text{Bias}(\hat{\beta}) = T \text{Bias}(\hat{\alpha}_{GRR}) = -TA^{-1}K\alpha = -TA^{-1}KT'\beta = -T[-(A+K)^{-1}KT'\beta] \quad (2.9.7)$$

de $\frac{1}{n}$ inci derecedendir.

Theobald (1974) ; b , β nın OLS tahmin edicisi olmak üzere $K = kI$ ve $0 < k < \frac{2\sigma^2}{(\beta'\beta)}$ olduğunda $MSE(b) - MSE(\hat{\beta})$ nın pozitif tanımlı olduğunu göstermiştir.

Bu sebepten ridge tahmin edicisi tercih edilir. Fakat ridge tahmin edicisi yanlıdır ve yanlılık yüksek olduğu zaman nokta tahminler güvenilir olmazlar.

Y_{-i} , i-nci koordinatı silinmiş Y vektörü ve Z_{-i} , i-nci sütunu silinmiş Z matrisini gösterebilir. Ayrıca α_{-i} ise (2.9.2) deki Z ve Y nin Z_{-i} ve Y_{-i} ile yer değiştirilmesi ile elde edilmiş olsun.

Böylece

$$(\hat{\alpha}_{GRR})_{-i} = (Z'_{-i}Z_{-i} + K)^{-1}Z'_{-i}Y_{-i} \quad (2.9.8)$$

Z_i , Z nin i-nci sütun vektörü ve Y_i ise Y nin i-nci koordinatı olsun. O zaman

$$(\hat{\alpha}_{GRR})_{-i} = (Z'Z - Z_iZ'_i + K)^{-1}(Z'Y - Z_iY_i) \quad (2.9.9)$$

yazabiliriz. Bu ifadeyi $U_i = Y_i - Z_i \hat{\alpha}_{GRR}$ ve $W_i = Z_i (Z'Z)^{-1} Z_i$ olmak üzere

$$(\hat{\alpha}_{GRR})_{-i} = \hat{\alpha}_{GRR} - \frac{A^{-1} Z_i U_i}{1 - W_i} \quad (2.9.10)$$

şeklinde basite indirgeyebiliriz.

Şimdi Miller (1974) u takip edelim ve

$$P_i = n \hat{\alpha}_{GRR} - (n-1) (\hat{\alpha}_{GRR})_{-i} \quad (2.9.11)$$

değerlerini tanımlayalım. Bu değerlerden Jack-Knife tahmin edicisi

$$\bar{P} = \frac{1}{n} \sum P_i = \hat{\alpha}_{GRR} + \frac{n-1}{n} A^{-1} \sum_{i=1}^n \frac{Z_i U_i}{(1 - W_i)} \quad (2.9.12)$$

şeklinde elde edilir.

U_i uzaklıklarında denge eksikliği neticesinde regresyon modelleri dengesiz olduğunda (2.9.11) deki değerler gözlemlere göre simetrik olarak tanımlanır. Bundan başka $[\hat{\alpha}_{GRR} - (\hat{\alpha}_{GRR})_{-i}]$ nin varyansı, W_i nin bir artan fonksiyonu olduğundan

$$Q_i = \hat{\alpha}_{GRR} + n(1 - W_i) [\hat{\alpha}_{GRR} - (\hat{\alpha}_{GRR})_{-i}] \quad (2.9.13)$$

değerleri tanımlanır.

Bu durumda uygun Jack-Knife uygulanmış tahmin edici

$$\hat{\alpha}_{AURR} = \bar{Q} = \frac{1}{n} \sum Q_i = \hat{\alpha}_{GRR} + A^{-1} \sum Z_i U_i \quad (2.9.14)$$

olarak verilir. Bu ifade

$$\hat{\alpha}_{AURR} = (I + A^{-1}K) \hat{\alpha}_{GRR} = \left[I - (A^{-1}K)^2 \right] \beta \quad (2.9.15)$$

şeklinde de yazılabilir.

$$\text{Bias}(\hat{\alpha}_{AURR}) = -(A^{-1}K)^2 \alpha \quad (2.9.16)$$

nin $\frac{1}{n^2}$ inci dereceden olduğu açıktır.

$|\dots|_i$, i-nci bileşenin mutlak değeri olmak üzere $|Bias(\hat{\alpha}_{GRR})_i|$ ve $|Bias(\hat{\alpha}_{AURR})_i|$ yi karşılaştırdığımızda yanlılıktaki azalmayı açıkça görebiliriz.

$$|Bias(\hat{\alpha}_{GRR})_i| - |Bias(\hat{\alpha}_{AURR})_i| = \frac{\lambda_i k_i}{(\lambda_i + k_i)^2} |\alpha_i| \quad (2.9.17)$$

pozitifdir.

Yanlılıktaki bu azalma $Bias(\tilde{\beta})$, $Bias(\hat{\beta})$ dan daha küçük dereceye sahip olsa bile $|Bias(\tilde{\beta})|$ ile $|Bias(\hat{\beta})|$ nin her bir bileşeninde olmayabilir. Fakat $\Delta = \sum \{ |Bias(\hat{\beta})_i| - |Bias(\tilde{\beta})_i| \}$ pozitiftir. Çünkü,

$$\Delta = \beta' \left\{ T \left[(A^{-1}K)^2 - (A^{-1}K)^4 \right] T' \right\} \beta \quad (2.9.18)$$

ifadesinde $\left\{ T \left[(A^{-1}K)^2 - (A^{-1}K)^4 \right] T' \right\}$ pozitiftir ve dolayısıyla $\Delta > 0$ dir.

2.10. Ön Bilgiye Dayalı Ridge Regresyon Tahmin Edicisi

$Y = X\beta + \varepsilon$ lineer modelindeki X matrisinin sütunları arasında lineer bağıntı varsa yani çoklu iç ilişkililik varsa OLS tahmin edicisini yani en iyi lineer yansız tahmin ediciyi (BLUE) içeren yansız tahmin ediciler güvenilir olmazlar. Bu durumda kullanılan yanlı tahmin ediciler arasında en çok kullanılan Hoerl ve Kennard (1970) in RR tahmin edicisidir. Bu tahmin edicilerin ailesinde $k=0$ olduğunda BLUE elde edilir. Ayrıca X matrisi ve Y ne olursa olsun k artarken $\hat{\beta}(k)$, 0 a yaklaşır. Hoerl ve Kennard (1970a)

i) $\hat{\beta}(k)$ nın; aynı rezidü kareler toplamına sahip tahmin edicilerin denklik sınıfındaki en kısa tahmin edici olduğunu

ii) $\hat{\beta}(k)$ nın; aynı uzunluğa sahip tahmin edicilerin denklik sınıfındaki minimum rezidü kareler toplamına sahip olduğunu

iii) OLS tahmin edicisi $\hat{\beta}$ nın beklenen uzunluğunun, β nın uzunluğundan fazla olduğunu göstermişlerdir.

Ancak tüm bunlar $k > 0$ için $\hat{\beta}(k)$ nın β nın bir tahmin edicisi olarak seçilmesinde zorlayıcı sebep değildir. $k > 0$ artarken $\hat{\beta}(k)$ nın izi, parametre uzayında $\hat{\beta}$ dan 0 a doğru bir eğri oluşturur. Bu eğri tek şekilde elde edilir, öyle ki : rezidü kareler toplamı mümkün olduğu kadar yavaş artarken , $\hat{\beta}(k)$ dan 0 a olan uzaklık mümkün olduğu kadar azalır.

Hoerl ve Kennard, k artarken ilk olarak $\hat{\beta}(k)$ ile β arasındaki uzaklığın beklenen değerinin azalması gerektiğini göstermişlerdir. Yani

$$E \left(\hat{\beta}(k) - \beta \right)' \left(\hat{\beta}(k) - \beta \right) < E \left(\hat{\beta} - \beta \right)' \left(\hat{\beta} - \beta \right) \quad (2.10.1)$$

olacak şekilde, her zaman k nın pozitif değerlerinin olduğunu göstermişlerdir.

Swindel ve Chapman (1973); Theobald (1974) daha genel olarak

$$E \left(\underline{L}' \hat{\beta}(k) - \underline{L}' \beta \right)^2 < E \left(\underline{L}' \hat{\beta} - \underline{L}' \beta \right)^2 \quad \forall \underline{L} \neq 0 \quad (2.10.2)$$

olması için gerek ve yeter koşulun; k nın hiçbir zaman boş olmayan

$$(0, +\infty) ; \left[(X'X)^{-1} - \sigma^{-2} \beta \beta' \right]; \text{ nnd ise}$$

$$K = K(X, \beta, \sigma^2) = \left(0, \frac{-2}{\min \text{ kök} \left[(X'X)^{-1} - \sigma^{-2} \beta \beta' \right]} \right); \text{ aksi halde} \quad (2.10.3)$$

açık aralıkta bulunması olduğunu göstermişlerdir.

Böylece, $\hat{\beta}(k)$ nin, parametrelerin sıfırdan farklı tüm lineer kombinasyonları için, $\hat{\beta}$ dan daha küçük mse ye sahip tahmin ediciler ortaya çıkarması için gerek ve yeter koşul k nin K da olmasıdır.

Böylece β nin BLUE olmayan tahmin edicileri için $\hat{\beta}(k)$ ailesinin göz önüne alınması için zorlayıcı sebeplerdir.

Son olarak Hoerl ve Kennard; k nin bir optimum seçimi; dolayısıyla $\hat{\beta}(k)$ nin bir optimum seçimi., modelin bilinmeyen parametrelerine dayansa bile bir şeylerin yapılabileceğini göstermişlerdir. Yani β nin tahmin edicisi $\hat{\beta}(k)$, ridge izinin incelenmesi ile seçilebilir. Bir başka deyişle küçük k değerleri için $\hat{\beta}(k)$ nin p koordinatının grafiği ve $(Y - X\hat{\beta}(k))'(Y - X\hat{\beta}(k))$ şeklindeki rezidü kareler toplamının incelenmesi ile seçilebilir.

Parametre uzayında herhangi bir b^* noktası başlangıç noktası olarak alınsın. Bu durumda

$$b(k, b^*) = (X'X + kI)^{-1}(X'Y + kb^*) \quad (2.10.4)$$

şeklindeki tahmin edici ailesi, $b^*=0$ olduğunda Hoerl ve Kennard'ın ridge tahmin edicisini; $k=0$ olduğunda ise BLUE yi içerir.

k sıfırdan farklı olarak arttığında, $b(k, b^*)$ nin izi, parametre uzayında $\hat{\beta}$ dan b^* a bir eğri oluşturur. Bu eğri tek şekilde belirlenir ; öyle ki : rezidü kareler toplamı mümkün olduğu kadar yavaş artarken, $b(k, b^*)$ nin b^* a olan uzaklığı mümkün olduğu kadar hızlı azalır. b^* a doğru eğri oluşturan $b(k, b^*)$ ve b^* üzerindeki hipotezleri veya ön bilgiyi mümkün olduğu kadar iyi yansıtacak şekilde seçilebilir.

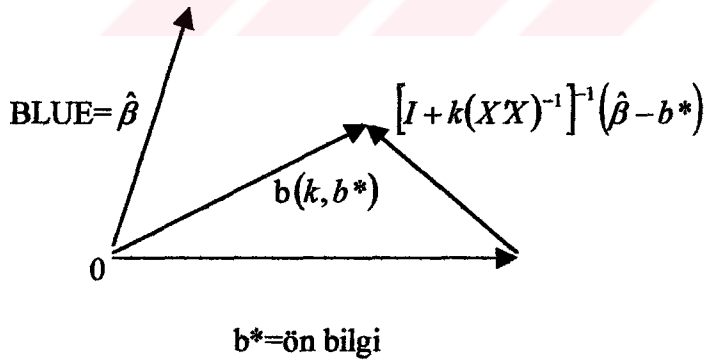
O zaman, b^* , β üzerindeki ön bilgi ve $b(k, b^*)$ ($k \geq 0$) ise β nin b^* ön bilgisine dayalı tahmin edicileri ailesi olarak adlandırılabilir. b^* nin bu şekilde elde edilmesiyle, özel bir k nin seçilmesi ve dolayısıyla $b(k, b^*)$ nin seçilmesi, $\hat{\beta}$ ve $b(k, b^*) = b^*$ ($k \rightarrow \infty$) arasında bir bağlantı olarak görülebilir.

$\hat{\beta}$ nin b^* a sezgisel olarak küçültülmesi; aslında $\hat{\beta}$ nin 0 a küçültülmesinden daha güçlüdür.

$Y = X\beta + \varepsilon$ modelindeki eğim parametreleri için parametre uzayı Ω , p boyutlu öklid uzayıdır. Orjin sabit ; b^* ve β bilinmeyen diğer sabit noktalarıdır.

$$b(k, b^*) = (X'X + kI)^{-1} (X'Y + kb^*) \quad k \geq 0 \quad (2.10.5)$$

şeklinde tanımlanan $b(k, b^*)$; b^* , $\hat{\beta} - b^*$ ile aşağıdaki şekildeki gibi ilişkili bir rasgele değişkendir.



Şekil (2.10.1)

Yani

$$\begin{aligned}
 b(k, b^*) &= (X'X + kI)^{-1} (X'Y + kb^*) \\
 &= (X'X + kI)^{-1} X'X (X'X)^{-1} (X'Y + kb^*) \\
 &= (I + k(X'X)^{-1})^{-1} (\hat{\beta} - b^* + b^* + k(X'X)^{-1} b^*) \\
 &= (I + k(X'X)^{-1})^{-1} (\hat{\beta} - b^*) + b^*
 \end{aligned} \tag{2.10.6}$$

olur. $\hat{\beta} = b^*$ olduğunda $b(k, b^*) = b^*$ olduğu açıktır. $\hat{\beta} \neq b^*$ olduğunda

i) $b(0, b^*) = \hat{\beta}$ ((2.10.5) den)

ii) $b(k, b^*) (k \rightarrow \infty) = \lim_{k \rightarrow \infty} b(k, b^*)$

$$\begin{aligned}
 &= b^* + \lim_{k \rightarrow \infty} (I + k(X'X)^{-1})^{-1} (\hat{\beta} - b^*) \\
 &= b^*
 \end{aligned}$$

olur. Tüm negatif olmayan k lar için, tüm kökler sıfıra yaklaşırken, k sınırsız olarak artarken $[I + k(X'X)^{-1}]^{-1}$ p.d. bir matristir. Yani $[I + k(X'X)^{-1}]^{-1}$; $k \rightarrow \infty$ iken sıfır matrisidir.

iii) $b(k, b^*)$ ile b^* arasındaki $\|b(k, b^*) - b^*\|$ öklid uzaklığı ; yani $\|(I + k(X'X)^{-1})^{-1} - (\hat{\beta} - b^*)\|$, k artarken monoton azalır.

İki simetrik p.d. matrisin toplamının ve çarpımının yine p.d. bir matris olduğunun kullanılmasıyla

$$0 \leq k \text{ ve } 0 < \Delta k$$

$$\begin{aligned}
&\Rightarrow \Delta k(\mathbf{X}'\mathbf{X})^{-1} > 0 \\
&\Rightarrow \mathbf{I} + [\mathbf{I} + k(\mathbf{X}'\mathbf{X})^{-1}]^{-1} \Delta k(\mathbf{X}'\mathbf{X})^{-1} - \mathbf{I} > 0 \\
&\Rightarrow (\mathbf{I} + k(\mathbf{X}'\mathbf{X})^{-1})^{-1} - [\mathbf{I} + (k + \Delta k)(\mathbf{X}'\mathbf{X})^{-1}]^{-1} > 0 \\
&\Rightarrow \left\{ [\mathbf{I} + k(\mathbf{X}'\mathbf{X})^{-1}]^{-1} + [\mathbf{I} + (k + \Delta k)(\mathbf{X}'\mathbf{X})^{-1}]^{-1} \right\} \\
&\quad \mathbf{X} \left\{ [\mathbf{I} + k(\mathbf{X}'\mathbf{X})^{-1}]^{-1} - [\mathbf{I} + (k + \Delta k)(\mathbf{X}'\mathbf{X})^{-1}]^{-1} \right\} > 0 \\
&\Rightarrow [\mathbf{I} + k(\mathbf{X}'\mathbf{X})^{-1}]^{-2} - [\mathbf{I} + (k + \Delta k)(\mathbf{X}'\mathbf{X})^{-1}]^{-2} > 0 \\
&\Rightarrow (\hat{\beta} - b^*)' [\mathbf{I} + k(\mathbf{X}'\mathbf{X})^{-1}]^{-2} (\hat{\beta} - b^*) - (\hat{\beta} - b^*)' [\mathbf{I} + (k + \Delta k)(\mathbf{X}'\mathbf{X})^{-1}]^{-2} (\hat{\beta} - b^*) > 0
\end{aligned}$$

şeklinde ispat yapılabilir.

iv) $b(k, b^*)$; $(\hat{\beta} - b^*)'(\hat{\beta} - b^*)$ ifadesini $(\mathbf{Y} - \mathbf{X}\hat{\beta})'(\mathbf{Y} - \mathbf{X}\hat{\beta}) = c$ olacak şekilde minimize eder.

$\frac{1}{k}$ Lagrange çarpanı olmak üzere

$$\varphi = (\hat{\beta} - b^*)'(\hat{\beta} - b^*) + \frac{1}{k} \left[(\mathbf{Y} - \mathbf{X}\hat{\beta})'(\mathbf{Y} - \mathbf{X}\hat{\beta}) - c \right]$$

fonksiyonunu ele alalım. $\varphi, \hat{\beta}$ nın bir konveks fonksiyonudur. φ nun $\hat{\beta}$ ya göre türevinin alınıp 0 a eşitlenmesiyle iddia sağlanır.

v) $b(k, b^*)$; $(\mathbf{Y} - \mathbf{X}\hat{\beta})'(\mathbf{Y} - \mathbf{X}\hat{\beta})$ ya $(\hat{\beta} - b^*)'(\hat{\beta} - b^*) = c$ olacak şekilde minimize eder ve bunun sağlanması ise

$$\varphi = (\mathbf{Y} - \mathbf{X}\hat{\beta})'(\mathbf{Y} - \mathbf{X}\hat{\beta}) + k \left[(\hat{\beta} - b^*)'(\hat{\beta} - b^*) - c \right]$$

nin yazılmasıyla (iv) deki gibidir.

Eğer $b(k, b^*)$; β parametrelerinin tüm sıfırdan farklı lineer kombinasyonları için Gauss-Markov'un BLUE Tahmin edicisinden daha küçük mse ye sahip bir tahmin edici oluşturuyorsa yani eğer

$$E \left(L' b(k, b^*) - L' \beta \right)^2 < E \left(L' \hat{\beta} - L' \beta \right)^2 \quad \forall L \neq 0 \quad (2.10.7)$$

ise $b(k, b^*)$; β nın b^* ön bilgisine dayalı iyi bir ridge tahmin edicisi olarak adlandırılır.

(2.10.7) nin gerçekleşmesi için gerek ve yeter koşul, k nın

$$K = K(b^*, X, \beta, \sigma^2) = \left\{ \begin{array}{l} (0, +\infty); \left[(X'X)^{-1} - (b^* - \beta)(b^* - \beta)' \sigma^{-2} \right] \text{ n.n.d. ise} \\ 0, \frac{-2}{\min k \delta k \left[(X'X)^{-1} - (b^* - \beta)(b^* - \beta)' \sigma^{-2} \right]}; \text{ aksi halde} \end{array} \right\} \quad (2.10.8)$$

açık aralığında olmasıdır.

Eğer b^* ön bilgisi β ya yeterli derecede yakın ise o zaman (2.10.5) den dolayı BLUE ile ön bilginin her karşılaştırılmasına; mse açısından bakıldığında ön bilgi; BLUE ya tercih edilebilir.

Özel olarak $(X'X)^{-1} - (b^* - \beta)(b^* - \beta)' / \sigma^2$ nin n.n.d. olması ve dolayısıyla her $b(k, b^*)$, $k \neq 0$ in (2.10.7) e göre iyi bir tahmin edici olması için gerek ve yeter koşul.

$$(b^* - \beta)' (b^* - \beta) \leq \sigma^2 \text{ minkök } (X'X)^{-1} \quad (2.10.9)$$

dur.

$K(b^*, X, \beta, \sigma^2)$ aralığı ; ön bilgi, X matrisi ve model parametreleri ne olursa olsun boş değildir. Yani mse ye göre karşılaştırıldığında her zaman BLUE ya tercih edilebilecek bir ön bilgi vardır.

Swindel; oluşturduğu tahmin edici ile ORR tahmin edicisini yani $MSE(\hat{\beta}(k))$ ve $MSE(b(k, b^*))$ nin karşılaştırmasını yapmamıştır.

$\hat{\beta}(k)$ ve $b(k, b^*)$ nin mse matrislerinin karşılaştırılması zordur. Çünkü iki tahmin edici arasındaki farkın iki nedeni olabilir :

- 1) OLS nin küçültüldüğü nokta yani 0 veya b^*
- 2) Ridge parametresi k tarafından belirlenen küçültme derecesi

İyi bir ön bilgi şu şekilde tanımlanır.

Tanım 2.10.1 : k nin tüm pozitif değerleri için aynı k değeri kullanılarak $\hat{\beta}(k)$ ve $b(k, b^*)$ hesaplandığında eğer $MSE(\hat{\beta}(k)) - MSE(b(k, b^*))$ p.s.d. ise , b^* ön bilgisi iyi olarak adlandırılır.

Hatırlatma 2.10.1 : $k > 0$ kısıtlaması yapılmıştır. Çünkü $k=0$ olursa o zaman tüm b^* için $\hat{\beta} = \hat{\beta}(k) = b(k, b^*)$ olur. Bu ise üç tahmin edicinin aynı riske sahip olduğunu belirtir.

Aşağıdaki teorem ; b^* nin iyi olması için gerek ve yeter koşulu verir.

Teorem (2.10.1) : $B = \beta\beta' - (\beta - b^*)(\beta - b^*)'$ olsun. b^* ön bilgisine bağlı olan tahmin edicinin iyi olması için gerek ve yeter koşul B nin p.s.d. olmasıdır.

İspat : $S(k) = (X'X + kI)^{-1}$ olsun. Daha önceki varsayımları ve k nin nonstokastik olduğunu kabul edelim. $\hat{\beta}(k)$ ve $b(k, b^*)$ nin mse matrisleri

$$MSE(\hat{\beta}(k)) = \sigma^2 S(k) X' X S(k) + k^2 S(k) \beta \beta' S(k) \quad (2.10.10)$$

$$MSE(b(k, b^*)) = \sigma^2 S(k) X' X S(k) + k^2 S(k) (\beta - b^*) (\beta - b^*)' S(k) \quad (2.10.11)$$

şeklindedir. $\hat{\beta}(k)$ ve $b(k, b^*)$ nin hesaplanmasında aynı k değeri kullanılırsa

$$MSE(\hat{\beta}(k)) - MSE(b(k, b^*)) = k S(k) B S(k) k \quad (2.10.12)$$

olur.

$k > 0$ iken $k S(k)$ p.d. olur. Dolayısıyla $MSE(\hat{\beta}(k)) - MSE(b(k, b^*))$ nin p.s.d olması için gerek ve yeter koşul B nin p.s.d olmasıdır.

Teorem (2.10.1) e göre sezgisel olarak ; $\hat{\beta}(k)$ ve $b(k, b^*)$ aynı k değeri için hesaplandığında bu iki tahmin edicinin varyans–kovaryans matrislerinin aynı olduğu söylenebilir. Bu sebepten (2.10.12) eşitliği sadece iki tahmin edicinin yanlılık vektörleri arasındaki farkı yansıtır. Dolayısıyla b^* nin uygun olarak seçilmesiyle $b(k, b^*)$ nin riski $\hat{\beta}(k)$ nin riskinden daha küçük olur.

Hatırlatma 2.10.2 : Teorem (2.10.1) ; iyi ön bilgi kümesinin; B nin p.s.d. olması koşulu tarafından karakterize edildiğini belirtir. $b^* = \beta \in B$ (yani $\beta \beta'$ p.s.d.) olduğundan bu kümenin hiçbir zaman boş olmadığı açıktır.

Hatırlatma 2.10.3 : $\hat{\beta}(k)$ ve $b(k, b^*)$ tahmin edicileri aynı k değeri için hesaplandığında $MSE(b(k, b^*)) - MSE(\hat{\beta}(k))$ nin p.s.d. olması için gerek ve yeter koşulun B nin p.s.d. olması olduğunu belirten bir teoremin ispatlanması yeterlidir.

Hatırlatma 2.10.4 : İyi ön bilgi tanımı; tüm pozitif k değerleri için $MSE(\hat{\beta}(k)) - MSE(b(k, b^*))$ nin p.s.d. olması kısıtlamasını içerir. İspattan, aynı gerek ve yeter koşulun iyi ön bilginin bir başka tanımı için de geçerli olduğu açıktır. Bu tanım : “ k nin özel değerlerinin seçilmesi ile mse matrisleri farklı psd matrisler olur” şeklindedir.

Tanım 2.10.2 : Bir ridge parametresi eğer $E(\underline{L}'\hat{\beta}(k)-\underline{L}'\beta)^2$ veya $E(\underline{L}'b(k,b^*)-\underline{L}'\beta)^2$ ni özel bir $\underline{L} \neq \underline{0}$ için minimize ediyorsa bu parametre optimaldir.

Hatırlatma 2.10.5 : Bu tanım Lee ve Campbell (1985) in tanımına benzerdir. Ancak Lee ve Campbell (1985) in tanımında minimize edilen fonksiyon $E(\hat{\beta}(k)-\beta)'(\hat{\beta}(k)-\beta)$ dir. Lee ve Campbell 'in optimal ridge parametresi bu fonksiyonun bir yerel minimumunu verir.

Teorem 2.10.2 : b^* m B psd olacak şekilde olduğunu kabul edelim. O zaman k_0 ve k_* , k nın $E(\underline{L}'\hat{\beta}(k)-\underline{L}'\beta)^2$ ve $E(\underline{L}'b(k,b^*)-\underline{L}'\beta)^2$ yi minimize eden değerleri olmak üzere $E(\underline{L}'\hat{\beta}(k_0)-\underline{L}'\beta)^2 \geq E(\underline{L}'b(k_*,b^*)-\underline{L}'\beta)^2$ olur.

İspat : b^* ; B p.s.d. olacak şekilde olsun. O zaman Teorem (2.10.1) den

$$E(\underline{L}'\hat{\beta}(k_0)-\underline{L}'\beta)^2 \geq E(\underline{L}'b(k_*,b^*)-\underline{L}'\beta)^2$$

olur. k_* m tanımından

$$E(\underline{L}'b(k_0,b^*)-\underline{L}'\beta)^2 \geq E(\underline{L}'b(k_*,b^*)-\underline{L}'\beta)^2$$

elde edilir. Dolayısıyla

$$E(\underline{L}'\hat{\beta}(k_0)-\underline{L}'\beta)^2 \geq E(\underline{L}'b(k_*,b^*)-\underline{L}'\beta)^2$$

olur.

Hatırlatma 2.10.6 : k_0 ve k_* , \underline{L} nin fonksiyonlarıdır ve tek değildirler. Bununla beraber k_0 ve k_* m tek olmamaları bir problem yaratmaz. Örneğin eğer $E(\underline{L}'\hat{\beta}(k)-\underline{L}'\beta)^2$ k nın birden fazla değeri için minimize ediliyorsa bu k değerlerinden herhangi biri k_0 olarak seçilebilir.

Hatırlatma (2.10.7) : B nin negatif yarı tanımlı (n.s.d.) olması

$$E(L' b(k, L) - L' \beta)^2 \geq E(L' b(k_0, L) - L' \beta)^2$$

olması için yeterli koşuldur.

Aşağıdaki örnek ; B nin p.s.d. olmasının ; özel bir L için tahmin edicilerin hesaplanmasında aynı k değeri kullanıldığında $L' b(k, b^*)$ nin mse sinin en fazla $L' \hat{\beta}(k)$ nin mse si kadar olması için gerekli olmadığını gösterir.

$d_1 = 9$ ve $d_2 = 1$ olmak üzere $X'X = \text{diag}(d_1, d_2)$ olsun. Ayrıca $L' = (20, 10)$ ve $\beta' = (4, 0)$ olsun. $b^{*'} = (3, 1)$ şeklindeki seçimini ele alalım. b^* in bu şekildeki seçimi B nin tanımsız bir matris olduğunu belirtir.

Eğer $k = 1$ ise (2.10.12) eşitliği

$$E(L' \hat{\beta}(k) - L' \beta)^2 - E(L' b(k, b^*) - L' \beta)^2 = +55$$

olur. Dolayısıyla sadece özel bir L ile çalışıldığında b^* in olası seçimlerini; B nin psd olacak şekildeki seçimlerine kısıtlamak gereksizdir.

Bir Rastgele Değişken Olarak Ön Bilgi

Uygulamalarda (1.2.1.) modelinde veri analizi ile karşılaşıldığında; β üzerindeki b^* ön bilgisinin belirli olarak ele alınması ve β nin b^* a bağlı tahmin edicilerinin bulunması mantıklıdır. Bu yüzden β nin b^* ön bilgisine dayalı iyi bir $b(k, b^*)$ ridge tahmin edicisine sahip olması için gerek ve yeter koşulun genel hali en kullanışlısıdır.

b^* in; p boyutlu bir rasgele değişkenin karşılaşılan değeri B olarak alınmasıyla formülün genelleştirilmesi zor değildir.

$$b(k, B) = (X'X + kI)^{-1}(X'Y + kB) \quad k \geq 0 \quad (2.10.13)$$

Y ye ve B ön bilgisine dayalı, β nin ridge tahmin edicileri ailesi ve $\underline{L}'b(k, B)$; $0 \leq k$ ise; $\underline{L} \neq 0$ rasgele seçilmiş olmak üzere $\underline{L}'\beta$ nin uygun tahmin edicileri ailesidir.

B nin sonlu bir $E(B)$ beklenen değerine sahip olduğunu kabul edelim. $\underline{L}'b(k, B)$ nin beklenen değeri

$$E(\underline{L}'b(k, B)) = \underline{L}'[I + k(X'X)^{-1}]^{-1}\beta + k\underline{L}'(X'X + kI)^{-1}E(B) \quad (2.10.14)$$

ve $\underline{L}'b(k, B)$ nin yanlışlığı

$$\text{Bias}(\underline{L}'b(k, B)) = k\underline{L}'(X'X + kI)^{-1}[E(B) - \beta] \quad (2.10.15)$$

dir.

Böylece B, β nin bir yansız tahmin edicisi ise $\underline{L}'b(k, B)$; $\underline{L}'\beta$ nin bir yansız tahmin edicisidir.

Bundan başka, B nin Y den bağımsız olduğunu ve V(B) sonlu varyans-kovaryans matrisine sahip olduğunu kabul edersek, $\underline{L}'b(k, B)$ nin varyans – kovaryans matrisi

$$V(\underline{L}'b(k, B)) = \underline{L}'(X'X + kI)^{-1}[\sigma^2 X'X + k^2 V(B)](X'X + kI)^{-1} \underline{L} \quad (2.10.16)$$

olur. $\underline{L}'b(k, B)$ nin mse si

$$\begin{aligned} \text{MSE}(\underline{L}'b(k, B)) &= V(\underline{L}'b(k, B)) + B^2(\underline{L}'b(k, B)) \\ &= \underline{L}'(X'X + kI)^{-1}[\sigma^2 X'X + k^2 \text{MSE}(B)](X'X + kI)^{-1} \underline{L} \quad (2.10.17) \end{aligned}$$

olur. $\underline{L}'\beta$ nin BLUE sinin mse si ise,

$$\text{MSE}(\underline{L}'\hat{\beta}) = \sigma^2 \underline{L}'(X'X)^{-1} \underline{L} \quad (2.10.18)$$

olur. Eğer

$$MSE(L' b(k, B)) < MSE(L' \hat{\beta}) \quad \forall L \neq 0 \quad (2.10.19)$$

ise $b(k, B)$, β nın B ön bilgisine dayalı iyi bir ridge tahmin edicisi olarak tanımlanır.

(2.10.19) için gerek ve yeter koşul ; k nın,

$$K = K(MSE(B), X, \sigma^2) = \left\{ \begin{array}{l} (0, +\infty) ; [(X'X)^{-1} - \sigma^{-2} MSE(B)] \text{ nmd ise} \\ \left(0, \frac{-2}{\min kök[(X'X)^{-1} - \sigma^{-2} MSE(B)]} \right) ; \text{ aksi halde} \end{array} \right\} \quad (2.10.20)$$

aralığında olmalıdır.

2.11. Küresel Olmayan Hatalı Lineer Modellerde

Ridge Regresyon Tahmin Edicileri

Çoklu lineer regresyon modelini ele alalım. $\Omega = \Omega(0)$ nın elemanları; q boyutlu öklid uzayının açık bir alt kümesinde bulunan q x 1 tipindeki bilinmeyen parametre vektörü θ nın bir fonksiyonudur.

β nın GLS tahmin edicisi

$$\beta^* = (X'\Omega X)^{-1} X'\Omega Y \quad (2.11.1)$$

ve $k \geq 0$ olmak üzere β nın ORR tahmin edicisi

$$\beta^*(k) = (X'\Omega X + kI_p)^{-1} X'\Omega Y \quad (2.11.2)$$

şeklinde tanımlanır.

Trenkler (1984) β^* ve $\beta^*(k)$ nin mse matrislerini karşılaştırmış ve $\beta^*(k)$ nin β^* dan daha iyi olması için gerek ve yeter koşulun $\beta' \left[\frac{2}{k} I_p + (X' \Omega X)^{-1} \right] \beta < \sigma^2$ olduğunu göstermiştir.

Eğer θ bilinmiyorsa; θ yerine onun tutarlı tahmin edicisi $\hat{\theta}$ nin yazılmasıyla (2.11.1) ifadesinden “kullanılabilir GLS tahmin edicisi”,

$$\hat{\beta} = (X' \hat{\Omega} X)^{-1} X' \hat{\Omega} Y \quad (2.11.3)$$

şeklinde bulunur. Burada $\hat{\Omega} = \Omega(\hat{\theta})$, Ω da θ yerine $\hat{\theta}$ nin yazılmasıyla elde edilmiştir.

Rothenberg (1984), kovaryans parametre tahminler üzerindeki belirli regülerlik koşulları altında, $\hat{\beta}$ nin bir lineer fonksiyonunun asimptotik dağılımını elde etmiştir.

(2.11.2) de θ yerine $\hat{\theta}$ ve k yerine

$$\hat{k} = \frac{as^2}{\hat{\beta}' c \hat{\beta}} \quad (2.11.4)$$

yazılarak ridge regresyon tahmin edicileri ailesi

$$\hat{\beta}_R = \left[X' \hat{\Omega} X + \frac{as^2}{\hat{\beta}' c \hat{\beta}} I_p \right]^{-1} X' \hat{\Omega} Y \quad (2.11.5)$$

şeklinde elde edilir.

Burada $a \geq 0$ bir skaler c p x p tipinde bir p.d. bir matris ve $s^2 = (n-p)^{-1} (Y - X \hat{\beta})' \hat{\Omega} (Y - X \hat{\beta})$ dir.

2.12. Shrunken Ls Tahmin Edicisi

$Y = X\beta + \varepsilon$ lineer regresyon modelini ele alalım. β nin OLS tahmin edicisi

$$\hat{\beta}_{OLS} = (X'X)^{-1} X'Y \quad (2.12.1)$$

dir.

α ; $0 < \alpha < 1$ şeklinde bir sabit olmak üzere shrunken LS tahmin edicisi

$$\hat{\beta}_{SLS} = \alpha \hat{\beta}_{OLS} \quad (2.12.2)$$

şeklindedir. Yanlı olan bu tahmin edicinin ortalaması

$$\alpha\beta \quad (2.12.3)$$

ve varyans-kovaryansı matrisi

$$\alpha^2 \sigma^2 (X'X)^{-1} \quad (2.12.4)$$

şeklindedir.

2.13. Bayes Tahmin Edicisi

β ,

$$\beta \quad E(\beta) = \theta \quad \text{ve} \quad V(\beta) = F \quad (2.13.1)$$

şeklinde bilinen ön ortalamaya ve bilinen varyansa sahip bir rasgele değişken olsun.

$$E\left(\frac{\varepsilon}{\beta}\right) = 0 \quad \text{ve} \quad V\left(\frac{\varepsilon}{\beta}\right) = \sigma^2 I \quad (2.13.2)$$

olduğunu kabul edelim. Varyans

$$E(p'\beta - a - L'Y) = 0 \quad (2.13.3.a)$$

ya göre minimize edildiğinde

$$\begin{aligned} V &= \text{Var}(p'\beta - a - L'Y) \\ &= \text{Var}(p'\beta) + \text{Var}(L'Y) - 2\text{Cov}(p'\beta, L'Y) \\ &= p'Fp + L'(XFX' + \sigma^2 I)L - 2p'FX'L \end{aligned} \quad (2.13.3.b)$$

elde edilir. (2.13.3.b) deki eşitlik

$$a = (p' - L'X)\theta \quad (2.13.4)$$

olduğunu belirtir.

$$(XFX' + \sigma^2 I)L = XFp \quad (2.13.5)$$

elde etmek için türev alalım. Bu durumda

$$L' = FX'(XFX' + \sigma^2 I)^{-1} \quad (2.13.6)$$

olur. $XFX' + \sigma^2 I$ pozitif tanımlıdır. Tam ranklı ve tam ranklı olmayan durumlar için optimum tahmin edici (2.13.4), (2.13.6) eşitliklerinden

$$p'\hat{\beta}_b = p'\theta + p'FX'(XFX' + \sigma^2 I)^{-1}(Y - X\theta) \quad (2.13.7)$$

olarak bulunur. Bu tahmin edici (2.13.1) ve (2.13.2) varsayımlarına göre minimum varyanslı yansız tahmin edicidir.

θ bilinmediğinde ve $p' = L'X$ olduğunda yani $p'\beta$ tahmin edilebilir olduğunda (2.13.3) ifadesi

$$V = L'L\sigma^2 \quad (2.13.8)$$

eşitliğine indirgenir. Bu eşitlikteki V yi $p' = L'X$ ile minimize edelim. Sonuçta Gauss-Markov teoreminden OLS tahmin edicisi elde edilir.

(2.13.3) deki optimizasyon probleminin düzenlenmiş bir hali; a sabit ve c bir vektör olmak üzere

$$V = \text{Var}(p'\beta - a - c'b) \quad (2.13.9)$$

ifadesinin

$$E(p'\beta - a - c'b) = 0 \quad (2.13.10)$$

eşitliğine göre minimize edilmesidir. Böylece

$$V = \text{Var}(p'\beta) + \text{Var}(c'b) - 2\text{Cov}(p'\beta, c'b) \quad (2.13.11)$$

$$= p'Fp + c'[UU'FUU' + \sigma^2(X'X)^+]c - 2p'FUU'c$$

olarak bulunur. $E(p'\beta - a - c'b) = 0$ eşitliği

$$p'\theta - cUU'\theta = a \quad (2.13.12)$$

olduğunu gösterir. (2.13.11) deki eşitliğin türevini aldığımızda

$$[UU'FUU' + \sigma^2(X'X)^+]c = p'FUU' \quad (2.13.13)$$

eşitliği elde edilir. $UU'FUU' + \sigma^2(X'X)^+$ nnd dir. (2.13.12) ve (2.13.13) eşitliklerinden

$$p'\hat{\beta} = p'\theta + p'FUU'[UU'FUU' + \sigma^2(X'X)^+]^+(b - V) \quad (2.13.14)$$

elde edilir.

θ bilinmediğinde (2.13.12), $p' = c'UU'$ ve $p'\beta$ tahmin edilebilir koşulu haline gelir. Problem ise

$$V = c'(X'X)^+ c\sigma^2 \quad (2.13.15)$$

eşitliğinin

$$p' - c'UU' = 0 \quad (2.13.16)$$

ya göre minimize edilmesi haline gelir.

λ , Lagrange çarpanlarının bir vektörü olsun.

$$T = c'(X'X)^+ c - (p' - c'UU')\lambda \quad (2.13.17)$$

olsun. T nin c'ye göre türevini alıp sonucu sifıra eşitlediğimizde

$$(X'X)^+ c + UU'\lambda = 0 \quad (2.13.18)$$

elde ederiz. Elde ettiğimiz bu ifadeyi $X'X$ ile çarparsak

$(X'X)^+ UU' = U\Lambda^{-1}U'UU' = (X'X)^+$ olduğundan

$$UU'c + X'X\lambda = 0 \quad (2.13.19)$$

ve

$$\lambda = -(X'X)^+ UU'c = -(X'X)^+ p \quad (2.13.20)$$

olarak buluruz.

$c=p$ eşitliğini elde etmek için (2.13.20) eşitliğini (2.13.18) deki yerine yazalım.

Böylece b çözüm olarak elde edilir.

2.14. Preliminary Test (PT) Tahmin Edicisi

$Y = X\beta + \varepsilon$, $\varepsilon \sim N(0, \sigma^2 I)$ lineer regresyon modelini ele alalım. Regresyonda birçok araştırmacı analize $H_0 : R\beta = r$ ya da $\phi = 0$ hipotezini $H_1 : R\beta \neq r$ veya $\phi \neq 0$ hipotezine karşı test ederek başlar. Bu hipotezi test etmek için

$$w = \frac{(Rb-r)' [R(X'X)^{-1} R']^{-1} (Rb-r) / m}{(Y-Xb)' (Y-Xb) / T-p} \sim F(m, T-p)$$

eşitliği ile verilen F testi için w istatistiği kullanılır.

c ; $\int_c^{\infty} dF(w) = \alpha$ şeklinde elde edilen kritik değer olmak üzere $w > c$ ise

H_0 reddedilir. Eğer H_0 kabul edilirse $Y = X\beta + \varepsilon$ modelindeki β için kısıtlı en küçük kareler (RLS) tahmin edicisi kullanılır. Aksi halde OLS tahmin edicisi kullanılır.

Bu prosedürde tahmin kısmı; önemliliğin bir ön testine bağlıdır. Bu yolla preliminary- test (PT) tahmin edicisi elde edilir. PT tahmin edicisi

$$\begin{aligned} b_{PT} &= I(w)b_R + (1-I(w))b \\ &= b - I(w)(X'X)^{-1} R' [R(X'X)^{-1} R']^{-1} (Rb-r) \end{aligned} \quad (2.14.1)$$

şeklinde tanımlanır. Burada b_R kısıtlanmış tahmin edici ; $I(w)$ ise

$$I(w) = 1 ; w \leq c \text{ ise}$$

$$I(w) = 0 ; w > c \text{ ise}$$

şeklindeki indikatör fonksiyondur.

Eğer $w > c$ ise $R\beta = r$ hipotezi reddedilir ve $I(w) = 0$ ve $b_{PT} = b$ olur.

Eğer kısıtlamalar stokastik ise yani $r = R\beta + V$ ise o zaman PT tahmin edicisinin stokastik hali $b_{PT} = I(w)b_M + (1-I(w))b$ şeklinde yazılabilir. Burada b_M ; “mixed tahmin edici” dir

$\Delta = MSE(b) - MSE(b_{PT})$ p.s.d. olduğunda b_{PT} tahmin edicisi OLS tahmin edicisine tercih edilir.

w merkezi olmayan F dağılımına sahiptir. Bu dağılımın merkezi olmama parametresi $\theta = \frac{\phi'Q^{-1}\phi}{2\sigma^2}$ olmak üzere $\theta \leq \frac{1}{4}$ olduğunda tüm $\eta \neq 0$ için

$$MSE(\eta'b) \geq MSE(\eta'b_{PT})$$

ise b_{PT} , OLS ye tercih edilir.

2.15. Temel Bileşenler (PC) Tahmin Edicisi

$Y = X\beta + \varepsilon$ lineer regresyon modelini ele alalım. Bu model için OLS tahmin edicisi $\hat{\beta}_{OLS} = (X'X)^{-1}X'Y$ dir.

$V'V = VV' = I$ olmak üzere $V'X'XV = \Lambda = \text{diag}(\lambda_1, \lambda_2, \lambda_3, \dots, \lambda_k)$ ve $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_k$ olsun . Bu tanıma göre V nin i-nci sütunu $X'X$ in λ_i özdeğerine göre normalleştirilmiş özvektörüdür.

Temel Bileşenler (PC) tahmin edicisi Kendal (1965) tarafından ortaya konulmuştur. Bu tahmin edici çoklu içilişkililik ile ilgili problemler olduğunda yaygın olarak kullanılmıştır. Johnson,Reimer ve Rothrock (1972) PC tahmin edicisinin $X'X$ in r tane özdeğeri ile ilgili bileşenlerin silinmesiyle elde edildiğini ve V_0 ; V nin ilk r sütunundan oluşan $k \times r$ tipinde matris olmak üzere PC tahmin edicisinin

$$\hat{\beta}_{PC} = \hat{\beta}_{OLS} - (X'X)^{-1}V_0(V_0'(X'X)^{-1}V_0)V_0'\hat{\beta}_{OLS} \quad (2.15.1)$$

şeklinde olduğunu göstermişlerdir.

PC tahmin edicisinin ortalaması

$$\beta - (X'X)^{-1}V_0(V_0(X'X)^{-1}V_0)^{-1}V_0'\beta \quad (2.15.2)$$

ve varyans-kovaryans matrisi

$$\sigma^2 \left[(X'X)^{-1} - (X'X)^{-1}V_0(V_0(X'X)^{-1}V_0)^{-1}V_0'(X'X)^{-1} \right] \quad (2.15.3)$$

şeklindedir.

2.16. Mixed Tahmin Edicisi

Klasik anlamda , regresyon parametreleri ile ilgili önbilginin lineer modellere dahil edilmesinde önemli iki yol vardır. Model parametreleri tahmin edildiğinde önbilgi; örneklem gözlemleri ile aynı durumda olan uydurulmuş gözlemlerle formüle edilebilir.

Durbin (1953) tarafından ortaya konulan ve Theil ve Goldberger (1961) tarafından geliştirilen bu yaklaşımın sonucunda elde edilen tahmin edici mixed tahmin edicidir.

$$Y = X\beta + \varepsilon \quad \varepsilon \sim N(0, \sigma^2 I) \quad (2.16.1)$$

lineer regresyon modelini ele alalım.

$$\beta \in B = \left\{ \beta : \beta'R'R\beta \leq \frac{\sigma^2}{c} \right\} \quad c > 0 \quad (2.16.2)$$

olduğunu kabul edelim.

ϕ , mx1 tipinde gözlenemeyen stokastik hata vektörü ; r , mx1 tipinde her zaman sıfır olarak gözlenen uydurulmuş stokastik vektör olmak üzere lineer önbilginin

$$r = R\beta + \phi \quad , \quad \phi \sim N\left(0, \begin{pmatrix} \sigma^2 \\ k \end{pmatrix} I\right) \quad (2.16.3)$$

şeklinde olduğunu kabul edelim. Ayrıca $\text{rank}(R) = m$ olduğunu kabul edelim. Bu durum $R\beta$ nın sifıra yakın olduğu ön düşüncesi stokastik lineer formda verildiğinde ortaya çıkar. (2.16.2) ifadesine göre $R\beta \neq 0$ olması olasılığı 1 dir.

Bu durumda mixed tahmin edici

$$b_R(k) = (X'X + kR'R)^{-1}(X'Y + kR'r) \quad (2.16.4)$$

olur. (2.16.1) ve (2.16.3) e bağlı olarak mixed tahmin edici yanlıdır. Çünkü (2.16.3)

bağıntısı yanlı bir bilgidir ve gerçekte $\phi \sim N\left(-R\beta; \left(\frac{\sigma^2}{k}\right)I\right)$ s şeklindedir.

2.17. Minimax Tahmin Edicisi

Regresyon parametreleri ile ilgili ön bilginin lineer modellere dahil edilmesinde kullanılan metodlar vardır. Model parametreleri tahmin edildiğinde ön bilgi ; örneklem gözlemleri ile aynı durumda olan uydurulmuş gözlemlerle formüle edilebilir. Lineer modellerde ayrıca lineer olmayan önbilgi de kullanılabilir. Eğer , regresyon katsayıları vektörü tüm katsayı uzayı içinde bilinen bir elipsoid tarafından içeriliyorsa veya izdüşümü bir alt uzayda bilinen bir elipsoidde ise bu önbilgi tahminde kullanılabilir. Kayıp matrisinin rankı 1 olduğunda ; karesel riskin maksimumunun ön kısıtlamalara göre minimize edilmesiyle minimax tahmin edici elde edilir.

$Y = X\beta + \varepsilon$, $\varepsilon \sim N(0, \sigma^2 I)$ lineer regresyon modelini ele alalım.

$$\beta \in B = \left\{ \beta : \beta'R'R\beta \leq \frac{\sigma^2}{c} \right\} \quad c > 0 \quad (2.17.1)$$

olduğunu kabul edelim.

Eğer ön bilgi (2.17.1) ise ; a , $m \times 1$ tipinde bir vektör ; $A = aa'$ kayıp matrisi olmak üzere karesel risk altında minimax tahmin edici

$$b_R^*(c) = (X'X + cR'R)^{-1} X'Y \quad (2.17.2)$$

şeklindedir.

2.18. Lineer Minimax Tahmin Edicisi

β , m boyutlu parametre vektörü ve θ m boyutlu bir vektör olsun. Parametre uzayı

$$\Omega = \left\{ \beta : (\beta - \theta)' G (\beta - \theta) \leq 1 \right\} \quad (2.18.1)$$

olsun. G nin pozitif tanımlı olduğunu kabul edelim.

$$p' \hat{\beta} = p' \theta + L'(Y - X\theta) \quad (2.18.2)$$

formundaki lineer tahmin ediciyi ele alalım.

Risk veya mse,

$$\begin{aligned} p'Rp &= p'E(\hat{\beta} - \beta)(\hat{\beta} - \beta)p \\ &= E[p'(\theta - \beta) + L'(Y - X\theta)] \left[(V - \beta)' p + (Y' - \theta'X')L \right] \\ &= (L'X - p')(\beta - \theta)(\beta - \theta)'(XL - p) + \sigma^2 L'L \end{aligned} \quad (2.18.3)$$

olarak bulunur. (2.18.1) deki elipsoidde ya da eşdeğer olarak,

$$(\beta - \theta)(\beta - \theta)' \leq G^{-1} = F \quad (2.18.4)$$

ifadesinde (2.18.3) ün maksimum değeri

$$p'Rp = (L'X - p')F(X'L - p) + \sigma^2 L'L \quad (2.18.5)$$

olur. L ye göre türev alalım.

$$2XFX'L - 2p'FX' + L\sigma^2 = 0 \quad (2.18.6)$$

bulunur. Böylece XFX' nmd olduğundan (2.18.6) eşitliği

$$L' = p'FX'[XFX' + \sigma^2 I]^{-1} \quad (2.18.7)$$

olduğunda minimize edilir. Bu durumda minimax tahmin edici,

$$p'\hat{\beta} = p'\theta + p'FX'(XFX' + \sigma^2 I)^{-1}(Y - X\theta) \quad (2.18.8)$$

olarak bulunur.

G nin tam ranklı olmadığı durumda elipsoid $(\beta - \theta)(\beta - \theta)' \leq F$ bölgesi ile yer değiştirdiğinde yukarıdaki sonuçlar geçerli olur.

2.19. Liu Tahmin Edicisi

$Y = X\beta + \varepsilon$, $E(\varepsilon) = 0$, $Cov(\varepsilon) = \sigma^2 I$ (1.2.1) lineer regresyon modelini ele alalım. Çoklu iç ilişkililik ortaya çıktığında OLS tahmin edicisi güvenilir olmaz. Çoklu iç ilişkililik problemini çözmek için kullanılan metodlardan ikisi ridge tahmini ve $\hat{\beta}_s = c\hat{\beta}$ $0 < c < 1$ şeklindeki Stein tahminidir.

Stein ve ridge tahminlerinin avantajları ve dezavantajları vardır. Ridge tahmini pratikte etkilidir ancak bu tahmin edici k nın karmaşık bir fonksiyonudur ve k yı seçerken karmaşık eşitliklerle karşılaşılabilir. Stein tahmininin avantajı , c nin lineer bir fonksiyon olmasıdır. Ancak Stein tahmin edicinin her bir elemanının küçülmesi ayırdır. Dolayısıyla pratikte çok kullanışlı değildir.

Ridge tahmin edicisi $\hat{\beta}(k)$ ve Stein tahmin edicisi $\hat{\beta}_s$ nin avantajlarını birleştiren, $0 < d < 1$ olmak üzere $\hat{\beta}_d = (X'X + I)^{-1}(X'Y + d\hat{\beta})$ şeklindeki yanlı tahmin ediciyi ele alalım. $\hat{\beta}_d$ nin genelleştirilmiş formu $\hat{\beta}(k)$ ninki ile aynıdır ve $\hat{\beta}_d$ nin $\hat{\beta}(k)$ ye göre avantajı , d nin lineer bir fonksiyonu olmasıdır. (1.2.1) lineer regresyon modelinin $0 = k^2\beta + c'$ ile genişletilip OLS tahmininin kullanılmasıyla $\hat{\beta}(k)$ elde edilir. Eğer lineer regresyon modeli $d\hat{\beta} = \beta + c'$ ile genişletilip OLS tahmini kullanılırsa $\hat{\beta}_d$ elde edilir. $\hat{\beta}_{Gd}$ tahmini, $\hat{\beta}_{Gd} = (X'X + I)^{-1}(X'X + \text{diag}(d_1, d_2, \dots, d_{p+1}))$ ifadesine genişletilebilir. Burada $i = 1, 2, 3, \dots, p+1$ olmak üzere $0 < d_i < 1$ parametrelerdir.

d nin seçimi : (1.2.1) deki model

$$Y = \beta_0 I + \bar{X}\beta_1 + \varepsilon \quad (2.19.1)$$

şeklinde yeniden yazılsın. Burada \bar{X} , merkezleştirilmiş ve $I = (1, \dots, 1)'$ olsun. Bu durumda \bar{Y} , β_0 ı tahmin etmekte kullanılabilir. Dolayısıyla sadece β_1 in tahmininin ele alınması gerekir.

$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$, $X'X$ in özdeğerleri ve q_1, q_2, \dots, q_p ise karşılık

gelen öz vektörler olsun. $Q = [q_1, q_2, \dots, q_p]$ ve $\Lambda = \begin{bmatrix} \lambda_1 & . & . & . \\ . & . & . & . \\ . & . & . & . \\ . & . & . & \lambda_p \end{bmatrix}$ olmak üzere

(2.19.1) ifadesi

$$Y = \beta_0 I + Z\alpha + \varepsilon \quad (2.19.2)$$

şeklinde yeniden yazılabilir. Burada $Z = \bar{X}Q$ ve $\alpha = Q'\beta_1$ dir. O zaman

$Z'Z = Q' \begin{pmatrix} \bar{X}' & \bar{X} \end{pmatrix} Q = \Lambda$ olur. Dolayısıyla $\hat{\alpha} = \Lambda^{-1}Z'Y$, $\hat{\alpha}(k) = (\Lambda + kI)^{-1}Z'Y$,

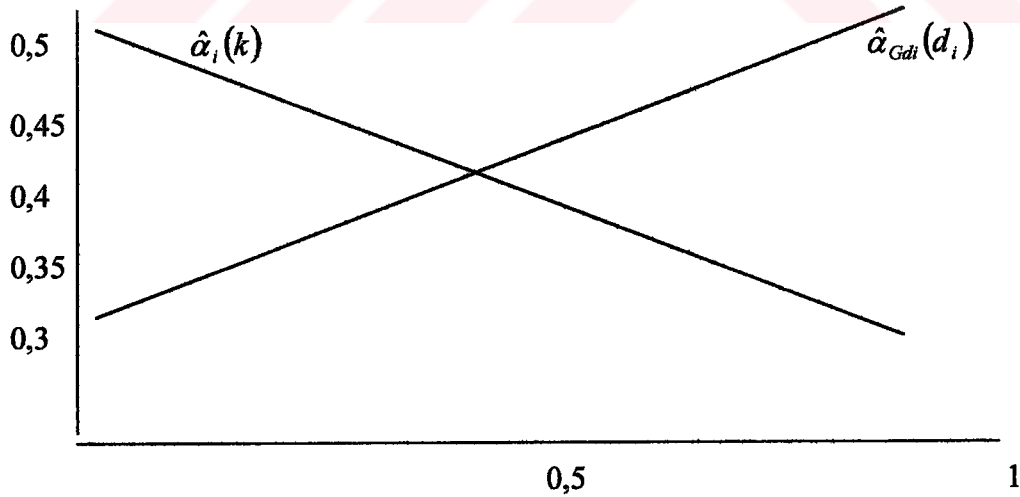
$$\hat{\alpha}_d = (\Lambda + I)^{-1}(Z'Y + d\hat{\alpha}) \text{ ve } \hat{\alpha}_{Gd} = (\Lambda + I)^{-1} \left[Z'Y + \begin{bmatrix} d_1 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & d_p \end{bmatrix} \hat{\alpha} \right] \text{ olur.}$$

Karşılıklı olarak $\hat{\beta}_1 = Q\hat{\alpha}$, $\hat{\beta}_1(k) = Q\hat{\alpha}(k)$, $\hat{\beta}_{1d} = Q\hat{\alpha}_d$ ve $\hat{\beta}_{1Gd} = Q\hat{\alpha}_{Gd}$,

$\hat{\alpha}_i(k) = \frac{\lambda_i}{\lambda_i + k} \hat{\alpha}_i$ ve $\hat{\alpha}_{Gdi}(d_i) = \frac{\lambda_i + d_i}{\lambda_i + 1} \hat{\alpha}_i$ ($i = 1, 2, 3, \dots, p$) olduğuna dikkat edilmesi

gerekir. Dolayısıyla $\hat{\alpha}_i(0) = \hat{\alpha}_i$, $\hat{\alpha}_i(1) = \frac{\lambda_i}{\lambda_i + 1} \hat{\alpha}_i$ ve $\hat{\alpha}_{Gdi}(0) = \frac{\lambda_i}{\lambda_i + 1} \hat{\alpha}_i$, $\hat{\alpha}_{Gdi}(1) = \hat{\alpha}_i$

olur. Bu sebepten $\hat{\alpha}_i(k)$ ve $\hat{\alpha}_{Gdi}(d_i)$ nin grafiği şekil (2.19.1) de olduğu gibi oluşturulabilir.



Şekil 2.19.1.

Şekil (2.19.1) den ; her $0 < k < 1$ için ; $\hat{\alpha}_i(k) = \hat{\alpha}_{Gdi}(d_i)$ olacak şekilde bir $0 < d_i < 1$ bulunabileceği söylenebilir. Pratikte optimal k , genellikle sıfıra yakındır. Dolayısıyla $\hat{\alpha}_{Gd}$ ile $\hat{\alpha}(k)$ nin aynı olabileceği sonucuna varılabilir.

Aşağıdaki teorem , mse açısından incelendiğinde $\hat{\beta}_{1d}$, $\hat{\beta}_1$ den daha iyi olacak şekilde bir d nin her zaman bulunabileceğini gösterir.

Teorem 2.19.1 : $mse(\hat{\beta}_{1d}) < mse(\hat{\beta}_1)$ olacak şekilde bir $0 < d < 1$ her zaman vardır.

İspat : Sadece $mse(\hat{\alpha}_d) < mse(\hat{\alpha})$ olduğunun gösterilmesi yeterlidir. $Cov(\hat{\alpha}_d) = \sigma^2 (\Lambda + I)^{-1} (\Lambda + dI) \Lambda^{-1} (\Lambda + dI) (\Lambda + I)^{-1}$ ve $E(\hat{\alpha}_d) = (\Lambda + I)^{-1} (\Lambda + dI) \alpha$ dir. Böylece,

$$\begin{aligned} mse(\hat{\alpha}_d) &= trCov(\hat{\alpha}_d) + \|E(\hat{\alpha}_d) - \alpha\|^2 \\ &= \sigma^2 \sum_{i=1}^p \frac{(\lambda_i + d)^2}{\lambda_i (\lambda_i + 1)} + (d-1)^2 \sum_{i=1}^p \frac{\alpha_i^2}{(\lambda_i + 1)^2} \\ &= g(d) \end{aligned}$$

olur. O zaman

$$g'(d) = 2\sigma^2 \sum_{i=1}^p \frac{(\lambda_i + d)}{\lambda_i (\lambda_i + 1)^2} + 2(d-1) \sum_{i=1}^p \frac{\alpha_i^2}{(\lambda_i + 1)^2}$$

olur. Bu sebepten

$$g'(1) = 2\sigma^2 \sum_{i=1}^p \frac{1}{\lambda_i (\lambda_i + 1)} > 0$$

dir. Böylelikle $g(d) < g(1)$ olacak şekilde veya eşdeğer olarak $mse(\hat{\alpha}_d) < mse(\hat{\alpha})$ olacak şekilde $0 < d < 1$ aralığında bir d vardır.

Teorem (2.19.1) in ispatından ; $mse(\hat{\alpha}_d)$ nin

$$d = \frac{\sum_{i=1}^p \frac{\alpha_i^2 - \sigma^2}{(\lambda_i + 1)^2}}{\sum_{i=1}^p \frac{\sigma^2 + \lambda_i \alpha_i^2}{\lambda_i (\lambda_i + 1)^2}}$$

de minimize edildiği görülebilir.

α_i^2 ve σ^2 yerine bunların yansız tahminleri olan $\hat{\alpha}_i^2 - \frac{\sigma^2}{\lambda_i}$ ve $\hat{\sigma}^2$ yazılırsa d nin

tahmini

$$\hat{d}_{mn} = 1 - \hat{\sigma}^2 \left[\frac{\sum_{i=1}^p \frac{1}{\lambda_i (\lambda_i + 1)}}{\sum_{i=1}^p \frac{\hat{\alpha}_i^2}{(\lambda_i + 1)^2}} \right] \quad (2.19.3)$$

şeklinde elde edilir. Bu ifade minimum mse tahmini olarak adlandırılır. Ayrıca bu eşitlik $h > 0$ olmak üzere

$$\hat{d}_{mnh} = 1 - h \hat{\sigma}^2 \left[\frac{\sum_{i=1}^p \frac{1}{\lambda_i (\lambda_i + 1)}}{\sum_{i=1}^p \frac{\hat{\alpha}_i^2}{(\lambda_i + 1)^2}} \right] \quad (2.19.4)$$

ifadesine genişletilebilir.

Bundan sonra ; ridge tahminindeki k nun tahminine benzer şekilde yapılan , d nin bazı tahminleri üzerinde durulacaktır. $\hat{\beta}_{1d}$, d nin bir lineer fonksiyonu olduğundan bu durumda hesaplamalar daha kolay olacaktır.

İlk olarak C_L kriterini göz önüne alalım. $H_d = \bar{X}(\bar{X}'\bar{X} + I)^{-1}(\bar{X}'\bar{X} + dI)\bar{X}'$ ve SS_{Resd} ise $\hat{\beta}_{1d}$ kullanılarak elde edilen rezidü kareler toplamı olmak üzere C_L

$$C_L = \frac{SS_{Resd}}{\hat{\sigma}^2} + 2tr(H_d) - (n-2)$$

şeklinde tanımlanır. d , C_L minimize edilecek şekilde seçilebilir. C_L nin, d nin bir karesel fonksiyonu olduğu kolayca görülebilir. Dolayısıyla C_L nin minimumu

$$\hat{d}_{C_L} = 1 - \hat{\sigma}^2 \left[\frac{\sum_{i=1}^p \frac{1}{(\lambda_i + 1)}}{\sum_{i=1}^p \frac{\lambda_i \hat{\alpha}_i^2}{(\lambda_i + 1)^2}} \right] \quad (2.19.5)$$

şeklinde elde edilebilir.

GCV kriteri, d nin seçiminde kullanılan bir başka yöntemdir. GCV ise

$$GCV = \frac{SS_{Resd}}{(n-1-tr(H_d))}$$

şeklinde tanımlanır. d , GCV nin minimize edilmesiyle seçilebilir. Bazı karmaşık işlemlerden sonra, GCV nin minimumunun \hat{d}_{C_L} olduğu gösterilebilir.

McDonald-Galarneau metodu ile

$$\|\hat{\beta}_{1d}\|^2 = \|\hat{\beta}\|^2 - \hat{\sigma}^2 \sum_{i=1}^p \left(\frac{1}{\lambda_i} \right)$$

olacak şekilde d seçilmesi için kullanılır. $\|\hat{\beta}_{1d}\|^2$, d nin bir karesel fonksiyonu olduğundan

$$\hat{d}_{MG} = 1 - \frac{\sum_{i=1}^p \frac{\hat{\alpha}_i^2}{\lambda_i + 1} - \left[\sum_{i=1}^p \frac{\hat{\alpha}_i^2}{\lambda_i + 1} \right]^2 - \hat{\sigma}^2 \sum_{i=1}^p \frac{\hat{\alpha}_i^2}{(\lambda_i + 1)^2} \sum_{i=1}^p \frac{1}{\lambda_i}}{\sum_{i=1}^p \frac{\hat{\alpha}_i^2}{(\lambda_i + 1)^2}} \quad (2.19.6)$$

hesaplanabilir.

Hoerl ve Kennard (1976) nın oluşturduğu iteratif metod (2.19.3) , (2.19.5) ve (2.19.6) ya uygulanabilir.

Örneğin , (2.19.3) için d nin iteratif tahminleri

$$\hat{\alpha}_{d_0} \quad d_0 = 1 - \hat{\sigma}^2 \left[\frac{\sum_{i=1}^p \frac{1}{\lambda_i (\lambda_i + 1)}}{\sum_{i=1}^p \frac{\hat{\alpha}_i^2}{(\lambda_i + 1)^2}} \right]$$

$$\hat{\alpha}_{d(d_0)} \quad d_1 = 1 - \hat{\sigma}^2 \left[\frac{\sum_{i=1}^p \frac{1}{\lambda_i (\lambda_i + 1)}}{\sum_{i=1}^p \frac{\hat{\alpha}_{d_i}^2(d_0)}{(\lambda_i + 1)^2}} \right]$$

$$\hat{\alpha}_{d(d_1)} \quad d_2 = 1 - \hat{\sigma}^2 \left[\frac{\sum_{i=1}^p \frac{1}{\lambda_i (\lambda_i + 1)}}{\sum_{i=1}^p \frac{\hat{\alpha}_{d_i}^2(d_1)}{(\lambda_i + 1)^2}} \right]$$

$T = \frac{tr(\overline{X'X})^{-1}}{p}$ olmak üzere , bu prosedür $\frac{(d_j - d_{j+1})}{d_j} < 20 \cdot T^{-1,3}$ ise işlem

durdurulur. O zaman α nın tahmin edilmesinde $\hat{\alpha}_d(d_j)$ kullanılabilir.

$\hat{\alpha}_{Gd}$ için mse

$$mse(\hat{\alpha}_d) = \sigma^2 \sum_{i=1}^p \frac{(\lambda_i + d_i)^2}{\lambda_i (\lambda_i + 1)^2} + \sum_{i=1}^p \frac{(d_i - 1)^2}{(\lambda_i + 1)^2}$$

dir. Dolayısıyla $i = 1, 2, \dots, p$ için $mse(\hat{\alpha}_{Gd})$, $d_i = \frac{\alpha_i^2 - \sigma^2}{\alpha_i^2 + \left(\frac{\sigma^2}{\lambda_i}\right)}$ de minimize edilir. α_i^2

ve σ^2 yerine bunların yansız tahminleri olan $\hat{\alpha}_i^2 - \frac{\hat{\sigma}^2}{\lambda_i}$ ve $\hat{\sigma}^2$ yazılırsa $i = 1, 2, \dots, p$ için

d_i nin $\hat{d}_i = 1 - \hat{\sigma}^2 \frac{\lambda_i + 1}{\lambda_i \hat{\alpha}_i^2}$ şeklindeki tahminleri elde edilir.

2.20. Hemen Hemen Yansız Genelleştirilmiş Liu Tahmin Edicisi

$Y = X\beta + \varepsilon$ $\varepsilon \sim N(0, \sigma^2 I_n)$ lineer regresyon modelinin

$X'X = \Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_p)$ ve λ_i , $X'X$ in i -nci öz değeri olacak şekilde kanonik formda olduğunu varsayalım. β nin OLS tahmin edicisi

$$\hat{\beta} = (X'X)^{-1} X' Y \quad (2.20.1)$$

dir. β nin genelleştirilmiş Liu tahmin edicisi

$$\begin{aligned} \hat{\beta}_{Gd} &= (\Lambda + I)^{-1} (X'Y + D\hat{\beta}) \\ &= (\Lambda + I)^{-1} (\Lambda + D)\hat{\beta} \end{aligned} \quad (2.20.2)$$

şeklinde verilebilir. $\hat{\beta}_{Gd}$,

$$\hat{\beta}_{Gd} = [I - (\Lambda + I)^{-1} (I - D)]\hat{\beta} \quad (2.20.3)$$

şeklinde yazılabilir. Bundan faydalanarak

$$B(\hat{\beta}_{Gd}) = -(\Lambda + I)^{-1} (I - D)\beta \quad (2.20.4)$$

olduğunu görmek kolaydır.

Kadiyala (1984) nın yanlışlık düzeltilmiş tahmin edici elde etmek için kullandığı yöntem burada uygulanırsa yanlışlığı düzeltilmiş Liu tahmin edicisi

$$\beta_{Gd}^w = \hat{\beta}_{Gd} + (\Lambda + I)^{-1}(I - D)\beta \quad (2.20.5)$$

olarak elde edilir. (2.20.5) ifadesinde Ohtani (1986) ve Nomura (1988) i izleyerek , β yerine $\hat{\beta}_{Gd}$ yazılmasıyla hemen hemen yansız genelleştirilmiş Liu (AUGL) tahmin edicisi

$$\tilde{\beta}_{Gd}^w = (I + (\Lambda + I)^{-1}(I - D))\hat{\beta}_{Gd} \quad (2.20.6)$$

veya

$$\tilde{\beta}_{Gd}^w = (I - (\Lambda + I)^{-2}(I - D)^2)\hat{\beta}_{Gd} \quad (2.20.7)$$

biçiminde elde edilir (Akdeniz ve Kaçıranlar (1994)). Burada

$$B(\tilde{\beta}_{Gd}^w) = -[(\Lambda + I)^{-2}(I - D)^2]\beta \quad (2.20.8)$$

ya da

$$= (\Lambda + I)^{-1}(I - D)B(\hat{\beta}_{Gd}) \text{ ve } Var(\tilde{\beta}_{Gd}^w) = [I + (\Lambda + I)^{-1}(I - D)]^2 Var(\hat{\beta}_{Gd}) \quad (2.20.9)$$

dır. Böylece (2.20.4) ve (2.20.8) deki matrisler köşegen matrisler olduklarından $B(\tilde{\beta}_{Gd}^w) < B(\hat{\beta}_{Gd})$ bulunur. Burada yanlışlık indirgeme çarpanı $(\Lambda + I)^{-1}(I - D)$ dir. Varyans-kovaryans matrisleri incelendiğinde (2.20.9) dan görülebileceği gibi $[I + (\Lambda + I)^{-1}(I - D)]^2$ çarpanıyla $\tilde{\beta}_{Gd}^w$ nin varyansının arttığı görülür. D nin elemanları sıfıra yaklaştırılırsa ve $\lambda_i \geq 0$ olması nedeniyle $[I + (\Lambda + I)^{-1}(I - D)]$ nin köşegen elemanları 2 den küçük kalacaktır. Dolayısıyla varyans-kovaryans matrisleri arasında da aşağıdaki eşitsizlik verilebilir.

$$Var(\hat{\beta}_{Gd}) \leq Var(\tilde{\beta}_{Gd}^w) \leq 4 \cdot Var(\hat{\beta}_{Gd}) \quad (2.20.10)$$

3. TAHMİN EDİCİLERİNİN KARŞILAŞTIRILMASI

3.1. Mse Kriteri Altında Regresyon Tahmin Edicelerinin Karşılaştırılması

Lineer regresyon modelinde katsayıların tahmininde bir çok metod kullanılmıştır. Bunlardan başlıcaları OLS, PC, ORR ve SLS tahmin edicileridir.

Şeçilecek olan tahmin edicinin tipi; performansının ölçmek için kullanılacak kritere bağlıdır.

Gunst ve Mason (1976), Theobald (1974), Toro-Vizcarronda ve Wallace (1968) ve diğerleri; regresyon katsayılarının kesin bir tahminini elde etmek amacıyla tahmin ediciler arasında karşılaştırma yapmak için bir standart olarak MSE kriterinin kullanılmasını önermişlerdir. Ancak tahmin edicilerin performansını ölçmek için bu kriter her zaman geçerli değildir. Örneğin Guilkey ve Price (1981) ; bu kriterin ; ancak tahmin ediciler belirli koşulları sağladığında, alternatif RLS tahmin edicilerinin karşılaştırılmasında kullanılabileceğini göstermişlerdir.

$Y = X\beta + \varepsilon$, $\varepsilon \sim N(0, \sigma^2 I)$ (1.2.1) lineer regresyon modelini ele alalım. β nin OLS tahmin edicisi $\hat{\beta}_{OLS} = (X'X)^{-1} X'Y$ dir. $\hat{\beta}_{OLS}$ yansızdır ve varyans kovaryans matrisi $Cov(\hat{\beta}_{OLS}) = \sigma^2 (X'X)^{-1}$ şeklindedir. $VV = VV' = I$ ve $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_k$ olmak üzere $V'X'XV = \Lambda = diag(\lambda_1, \lambda_2, \dots, \lambda_k)$ olsun.

Tanımdan dolayı ; V nin i-nci sütunu ; $X'X$ in λ_i özdeğerine karşılık gelen normalleştirilmiş özvektörüdür.

PC tahmin edicisi Kendall (1965) tarafından sunulmuştur ve çoklu iç ilişkililik problemi ortaya çıktığında kullanılmıştır. Johnson, Reimer ve Rothrock (1972) ; PC tahmin edicisinin, $X'X$ in r tane en küçük özdeğerine karşılık gelen bileşenlerin

silinmesiyle oluşturulduğunu göstermişlerdir. PC tahmin edicisi ; V_0 , V nin ilk r sütunu ile oluşturulan $k \times r$ tipinde bir matris olmak üzere

$$\hat{\beta}_{PC} = \hat{\beta}_{OLS} - (X'X)^{-1} V_0 \left(V_0' (X'X)^{-1} V_0 \right)^{-1} V_0' \hat{\beta}_{OLS}$$

şeklindedir. Bu $V_0' \hat{\beta}_{PC} = 0$ kısıtlamasına göre rezidü kareler toplamını minimize eden RLS tahmin edicisine eşdeğerdir.

PC tahmin edicisinin ortalaması $\beta - (X'X)^{-1} V_0 \left(V_0' (X'X)^{-1} V_0 \right)^{-1} V_0' \beta$ ve varyans-kovaryans matrisi $\sigma^2 \left\{ (X'X)^{-1} - (X'X)^{-1} V_0 \left(V_0' (X'X)^{-1} V_0 \right)^{-1} V_0' (X'X)^{-1} \right\}$ şeklindedir.

Hoerl ve Kennard (1970) ; D $k \times k$ tipinde negatif olmayan elemanlardan oluşan bir matris olmak üzere ; ORR tahmin edicisini $\hat{\beta}(k) = (X'X + VDV')^{-1} X'Y$ şeklinde oluşturmuşlardır.

D nin en az bir tane pozitif elemanı olduğunu kabul edelim. Aksi halde ORR tahmin edicisi OLS tahmin edicisine eşit olur.

ORR tahmin edicisi yanlıdır ve ortalaması $(X'X + VDV')^{-1} X'X\beta$ ve varyans kovaryans matrisi $\sigma^2 (X'X + VDV')^{-1} X'X (X'X + VDV')^{-1}$ şeklindedir.

SLS tahmin edicisi Mayer ve Willke (1973) tarafından $0 < \alpha < 1$ aralığında bir sabit olmak üzere $\hat{\beta}_{SLS} = \alpha \hat{\beta}_{OLS}$ şeklinde tanımlamıştır. Bu tahmin edici yanlıdır ve ortalaması $\alpha\beta$ ve varyans kovaryans matrisi $\alpha^2 \sigma^2 (X'X)^{-1}$ şeklindedir.

Tahmin Edicilerin Karşılaştırılmaları

Görüldüğü gibi regresyon katsayılarının tahmin edilmesinde bir çok metod kullanılmaktadır. Şimdi bu tahmin edicilerin performanslarının belirlenmesi için kullanılan bir kriter üzerinde durulacak ve bu kriterin hangi durumlarda bu tahmin edicilerin karşılaştırılmaları için uygun olduğu belirlenecektir.

$\hat{\beta}_1$ ve $\hat{\beta}_2$ regresyon katsayılarının iki tahmin edicisi olsun. Tahmin edicilerin performanslarının karşılaştırılmasında sıklıkla kullanılan teknik ; performansların karesel kayıp fonksiyonları terimleri cinsinden belirlenmesidir.

Theobald (1974) ; $i = 1,2$ için $MSE(\hat{\beta}_i) = E(\hat{\beta}_i - \beta)(\hat{\beta}_i - \beta)'$ olmak üzere her karesel kayıp fonksiyonu için $\hat{\beta}_1$ nin $\hat{\beta}_2$ ye tercih edilmesi için gerek ve yeter koşulun $MSE(\hat{\beta}_2) - MSE(\hat{\beta}_1)$ nin bir p.s.d. matris olması olduğunu göstermiştir. Buna göre tahmin edicilerin karşılaştırılmasında aşağıdaki kriter oluşturulabilir.

Gerçek parametre değerlerine göre hesaplama yapıldığında; $\hat{\beta}_1$ nin $\hat{\beta}_2$ ya MSE kriteri altında tercih edilebilmesi için gerek ve yeter koşul $MSE(\hat{\beta}_2) - MSE(\hat{\beta}_1)$ in p.s.d olmasıdır.

Bir tahmin edicinin seçilmesinde zor bir gereklilik olduğundan bu koşulun sağlanmasında problemler ortaya çıkabilir. Bu problemlerden biri ; tüm β ve σ^2 değerleri için eğer $MSE(\hat{\beta}_2) - MSE(\hat{\beta}_1)$ matrisi tanımsız ise o zaman $\hat{\beta}_1$ ve $\hat{\beta}_2$ MSE kriterine göre karşılaştırılmaz. Bu durumda MSE kriterine göre hangi tahmin edicinin seçilmesi gerektiği söylenemez.

İkinci problem ; kriter tarafından tanımlanan karar kuralının; bilinmediği kabul edilen parametrelerin, gerçek değerlerindeki tahmin edicilerin özelliklerine bağlı olmasıdır. Dolayısıyla tüm β ve σ^2 değerleri için $MSE(\hat{\beta}_2) - MSE(\hat{\beta}_1)$ p.s.d.

olmadıkça , bu kritere göre $\hat{\beta}_1$ nın $\hat{\beta}_2$ ya tercih edilebileceğinden emin olunamaz. Ancak SLS, OLS, ORR ve PC tahmin edicilerinden herhangi ikisi karşılaştırıldığında; birinin, ancak parametre uzayının uygun bir alt kümesinde diğerinden daha iyi olduğu görülür.

$\hat{\beta}_1$ ve $\hat{\beta}_2$; SLS ,OLS, PC ve ORR tahmin edicileri sınıfından seçilen iki tahmin edicileri sınıfından seçilen iki tahmin edici olmak üzere ; $\Omega(\hat{\beta}_1, \hat{\beta}_2)$; tüm β ve σ^2 için $MSE(\hat{\beta}_2) - MSE(\hat{\beta}_1)$ nın p.s.d. olduğu kümeyi gösterebilir. Bu durumda iki sonuç ortaya çıkar. Eğer $\Omega(\hat{\beta}_1, \hat{\beta}_2)$ kümesi boş küme değilse o zaman karşılaştırma yapılabilmesi için Toro-Vizcarrando ve Wallace (1968) in önerdiği gibi bazı istatistiksel testlerin uygulanması gerekir.

Eğer $\Omega(\hat{\beta}_1, \hat{\beta}_2)$ kümesi boş ise parametrelerin hiçbir değeri için $\hat{\beta}_1$, $\hat{\beta}_2$ ya tercih edilemez. Ayrıca bu ifade $\hat{\beta}_2$ nın aynı kritere göre $\hat{\beta}_1$ yas tercih edilebileceği anlamına gelmez. Çünkü bu iki tahmin edici parametrelerin hiçbir değeri için karşılaştırılmaz.

Şimdi; bazı parametre değerleri için ; $MSE(\hat{\beta}_2) - MSE(\hat{\beta}_1)$ nın p.s.d. olması için $\hat{\beta}_1$ ve $\hat{\beta}_2$ nın sağlaması gereken koşulları inceleyelim. Bu bize tahmin edicilerin karşılaştırılmasında ne zaman istatistiksel prosedürlerin uygulanması gerektiği ve ne zaman tahmin edicilerin MSE kriterine göre karşılaştırılmayacağını belirtir.

$\hat{\beta}_1$ veya $\hat{\beta}_2$ nın OLS tahmin edicisi olduğu ve $\hat{\beta}_1$ ve $\hat{\beta}_2$ nın her ikisinin de SLS tahmin edicisi olduğu durumda $\Omega(\hat{\beta}_1, \hat{\beta}_2)$ kümesini tanımlayan özellikler incelenecektir.

Bu sonuçlar üzerinde durmadan önce aşağıdaki lemmaları inceleyelim.

Lemma 3.1.1 : A sxs tipinde herhangi bir negatif tanımlı (n.d.) matris ve z , sxl tipinde herhangi bir vektör olsun. Eğer $s \geq 2$ ise o zaman $A + zz'$ p.s.d. olamaz.

İspat : zz' matrisinin rankı en fazla 1'e eşittir. Eğer $s \geq 2$ ise o zaman $w'zz'w = 0$ olacak şekilde sıfırdan farklı bir $s \times 1$ tipinde w vektörü vardır. A n.d. olduğundan $w'(A + zz')w = w'Aw < 0$ olur. Bu da $A + zz'$ nün p.s.d olamayacağını gösterir.

Lemma 3.1.2 : A $s \times s$ tipinde herhangi bir p.d matris ve z , $s \times 1$ tipinde herhangi bir vektör olsun . O zaman $A - zz'$ nün p.s.d. olması için gerek ve yeter koşul $z'A^{-1}z \leq 1$ olmasıdır.

İspat : A p.d. olduğundan ; $P'P = A$ olacak şekilde $s \times s$ tipinde tekil olmayan bir P matrisi vardır. Bu ise $A - zz'$ nün p.s.d. olması için gerek ve yeter koşulun $I - (P')^{-1}zz'P^{-1}$ in p.s.d. olması gerektiğini belirtir. Bu matris $(s-1)$ katlı 1 özdeğerine sahiptir ve $I - z'P^{-1}(P')^{-1}z = I - z'A^{-1}z$ özdeğeri 1 katlıdır. Bir matrisin p.s.d olması için gerek ve yeter koşul tüm özdeğerlerinin negatif olmayan olmasıdır. Dolayısıyla $A - zz'$ nün p.s.d. olması için gerek ve yeter koşul $z'A^{-1}z \leq 1$ olmasıdır.

Lemma 3.1.3 : Her SLS tahmin edicisi bir ORR tahmin edicisi olarak yazılabilir.

İspat : Her bir SLS tahmin edicisi ; α , $0 < \alpha < 1$ aralığında bir sabit olmak üzere $\alpha\hat{\beta}_{OLS}$ formundadır. $D = \left(\frac{1-\alpha}{\alpha}\right)\Lambda$ tanımlayalım. O zaman ORR tahmin edicisi

$$\begin{aligned} (X'X + VDV')^{-1}X'X\hat{\beta}_{OLS} &= V(\Lambda + D)^{-1}\Lambda V'\hat{\beta}_{OLS} \\ &= V\left(\Lambda + \left(\frac{1-\alpha}{\alpha}\right)\Lambda\right)^{-1}\Lambda V'\hat{\beta}_{OLS} = \alpha\hat{\beta}_{OLS} \end{aligned}$$

olur. Dolayısıyla; SLS tahmin edicisi ORR tahmin edicisinin bir özel halidir.

3.2. OLS ve PC Tahmin Edicilerinin Karşılaştırılması

Teorem 3.2.1 : V_0 ; V nin ilk r sütunundan oluşturulan $k \times r$ tipinde bir matris

olmak üzere $\hat{\beta}_{PC} = \hat{\beta}_{OLS} - (X'X)^{-1} V_0 (V_0' (X'X)^{-1} V_0)^{-1} V_0' \hat{\beta}_{OLS}$ şeklinde tanımlansın,

i) Eğer $rank(V_0) \geq 2$ ise o zaman $\Omega(\hat{\beta}_{OLS}, \hat{\beta}_{PC})$ boş kümedir.

ii) Eğer $rank(V_0) = 1$ ise o zaman $MSE(\hat{\beta}_{PC}) - MSE(\hat{\beta}_{OLS})$ nin n.n.d. olması için gerek ve yeter koşul

$$\frac{(V_0' \beta)' (V_0' (X'X)^{-1} V_0)^{-1} (V_0' \beta)}{\sigma^2} \geq 1$$

olmasıdır.

İspat : PC tahmin edicisi ; $V_0' \hat{\beta}_{PC} = 0$ kısıtlamasını sağlayan RLS tahmin edicisine eşdeğer olduğundan Guilkey ve Price (1981) in sonuçlarının uygulanması ile ispat yapılır.

Teorem 3.2.2: $MSE(\hat{\beta}_{OLS}) - MSE(\hat{\beta}_{PC})$ nin ps.d olması için gerek ve yeter koşul

$$\frac{(V_0' \beta)' (V_0' (X'X)^{-1} V_0)^{-1} (V_0' \beta)}{\sigma^2} \leq 1$$

olmasıdır.

3.3. OLS ve ORR Tahmin Edicilerinin Karşılaştırılması

Teorem 3.3.1 : $D = \text{diag}(d_1, d_2, \dots, d_k)$ olmak üzere ORR tahmin edicisi $\hat{\beta}(k) = (X'X + VDV')^{-1} X'Y$ şeklinde tanımlansın. O zaman,

i) Eğer d_i ; iki veya daha fazla i için sıfırdan farklı ise o zaman β ve σ^2 nin hiçbir değeri için $MSE(\hat{\beta}(k)) - MSE(\hat{\beta}_{OLS})$ p.s.d. olamaz.

ii) Eğer d_i ; sadece bir i için sıfırdan farklı ise o zaman $MSE(\hat{\beta}(k)) - MSE(\hat{\beta}_{OLS})$ nin p.s.d. olması için gerek ve yeter koşul $Z = (\Lambda + D)^{-1} DV'\beta$ ve $A = \Lambda^{-1} - (\Lambda + D)^{-1} \Lambda (\Lambda + D)^{-1}$ olmak üzere $\frac{Z'A^+Z}{\sigma^2} \geq 1$ olmasıdır. Burada A^+ , A matrisinin genelleştirilmiş tersidir.

İspat : $MSE(\hat{\beta}(k)) - MSE(\hat{\beta}_{OLS})$ nin p.s.d. olması için gerek ve yeter koşul

$$V' \{MSE(\hat{\beta}(k)) - MSE(\hat{\beta}_{OLS})\} V =$$

$$\sigma^2 \{(\Lambda + D)^{-1} \Lambda (\Lambda + D)^{-1} - \Lambda^{-1}\} + \{(\Lambda + D)^{-1} DV'\beta\} \{(\Lambda + D)^{-1} DV'\beta\}' \quad (3.3.1)$$

nin ps.d. olmasıdır. Varsayımdan dolayı iki veya daha fazla i için $d_i > 0$ dir. Genelliği kaybetmeden $i = 1, 2$ için d_i nin pozitif olduğunu kabul edelim. $D_0 = \text{diag}(d_1, d_2)$ ve $\Lambda_0 = \text{diag}(\lambda_1, \lambda_2)$ tanımlayalım. V_0 ; V nin ilk iki sütunu ile oluşturulan $k \times 2$ tipinde bir matris olur. O zaman (3.3.1) ifadesindeki matris

$$\sigma^2 \{(\Lambda_0 + D_0)^{-1} \Lambda_0 (\Lambda_0 + D_0)^{-1} - \Lambda_0^{-1}\} + \{(\Lambda_0 + D_0)^{-1} D_0 V_0' \beta\} \{(\Lambda_0 + D_0)^{-1} D_0 V_0' \beta\}' \quad (3.3.2)$$

olur. Bu ifadedeki ilk matris n.d. olduğundan ve $(\Lambda_0 + D_0)^{-1} V_0' \beta$ 2×1 tipinde bir vektör olduğundan ; (3.3.2) ifadesinde verilen matris Lemma 3.1.1 den dolayı hiçbir

zaman p.s.d. olamaz. Bu ise $MSE(\hat{\beta}(k)) - MSE(\hat{\beta}_{OLS})$ nin hiçbir parametre değeri için p.s.d. olamayacağını belirtir.

iii) $d_j > 0$ olsun ve tüm $i \neq j$ için $d_i = 0$ olduğunu kabul edelim. (3.3.1) eşitliğinden r için $s \neq j$ olmak üzere $(V\{MSE(\hat{\beta}_{RR}) - MSE(\hat{\beta}_{OLS})\}V)_{rs} = 0$ olduğu görülür. Dolayısıyla $MSE(\hat{\beta}(k)) - MSE(\hat{\beta}_{OLS})$ nin p.s.d. olması için gerek ve yeter koşul (3.3.1) deki matrisin (j, j) nci elemanının negatif olmayan olmasıdır. Bu koşul ise $\frac{Z'A + Z}{\sigma^2} > 1$ koşuluna denktir.

Sonuç 3.3.1 : $\hat{\beta}_{SLS} = \alpha \hat{\beta}_{OLS}$ olarak tanımlansın.

i) Eğer $k > 0$ ise o zaman $\Omega(\hat{\beta}_{OLS}, \hat{\beta}_{SLS})$ boş kümedir.

ii) Eğer $k=1$ ise $MSE(\hat{\beta}_{SLS}) - MSE(\hat{\beta}_{OLS})$ nin p.s.d. olması için gerek ve yeter koşul $\frac{(1-\alpha)^2(1-\alpha^2)^{-1} \beta'X'X\beta}{\sigma^2} \geq 1$ olmasıdır.

Bu sonucun bir genelleştirmesi olan aşağıdaki teorem Theobald (1974) tarafından oluşturulmuştur.

Teorem 3.3.2 : $D = \text{diag}(d_1, d_2, \dots, d_k)$ olmak üzere

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

$$Z = (\Lambda + D)^{-1} DV'\beta$$

$$A = \Lambda^{-1} - (\Lambda + D)^{-1} \Lambda (\Lambda + D)^{-1}$$

şeklinde tanımlansın. O zaman $MSE(\hat{\beta}_{OLS}) - MSE(\hat{\beta}(k))$ nin p.s.d. olması için gerek ve yeter koşul

$$\frac{Z'A'Z}{\sigma^2} < 1$$

olmasıdır.

İspat : (3.3.1) eşitliğinden $MSE(\hat{\beta}_{OLS}) - MSE(\hat{\beta}_{RR})$ nin p.s.d. olması için gerek ve yeter koşul

$$-\sigma^2 \left\{ (\Lambda + D)^{-1} \Lambda (\Lambda + D)^{-1} - \Lambda^{-1} \right\} - \left\{ (\Lambda + D)^{-1} DV' \beta \right\} \left\{ (\Lambda + D)^{-1} DV' \beta \right\}' = \sigma^2 A - ZZ'$$

nın ps.d. olmasıdır.

a) Tüm i ler için $d_i > 0$ olduğunu kabul edelim. O zaman $\sigma^2 A$ p.d. olur ve Lemma (3.1.2) nin uygulanmasıyla ispat yapılır.

b) Bazı i ler için $d_i = 0$ olduğunu kabul edelim. Genelliği kaybetmeden tüm $i \leq j$ için $d_i > 0$ ve tüm $i > j$ için $d_i = 0$ olduğunu kabul edelim.

$$D_0 = \text{diag}(d_1, d_2, \dots, d_j)$$

$$\Lambda_0 = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_j)$$

olsun. V_0 ise ; V nin ilk j sütunundan elde edilen $k \times j$ tipindeki matris olsun. O zaman

$$Z_0 = (\Lambda_0 + D_0)^{-1} D_0 V_0' \beta \quad \text{ve} \quad A_0 = \Lambda_0^{-1} - (\Lambda_0 - D_0)^{-1} \Lambda_0 (\Lambda_0 + D_0)^{-1} \quad \text{olmak üzere}$$

$$\sigma^2 A - ZZ' = \begin{bmatrix} \sigma^2 A_0 - Z_0 Z_0' & 0 \\ 0 & 0 \end{bmatrix}$$

olur. Dolayısıyla $MSE(\hat{\beta}_{OLS}) - MSE(\hat{\beta}(k))$ nin p.s.d. olması için gerek ve yeter koşul

$\sigma^2 A_0 - Z_0 Z_0'$ nün p.s.d. olmasıdır. Ayrıca $\sigma^2 A_0$ in p.d. olduğu biliniyor. Bu sebepten

Lemma 3.1.2 den $\sigma^2 A_0 - Z_0 Z_0'$ nün p.s.d. olması için gerek ve yeter koşul

$$\frac{Z_0' A^{-1} Z_0}{\sigma^2} \leq 1$$

olmasıdır ki bu durum teoremdesine denktir.

Sonuç 3.3.2 : Eğer $\hat{\beta}_{SLS} = \alpha \hat{\beta}_{OLS}$ ise $MSE(\hat{\beta}_{OLS}) - MSE(\hat{\beta}_{SLS})$ nin p.s.d. olması için gerek ve yeter koşul $\frac{(1-\alpha)^2 (1-\alpha^2)^{-1} \beta' X' X \beta}{\sigma^2} \leq 1$ olmasıdır.

3.4. ORR Tahmin Edicilerinin Karşılaştırılmaları

Teorem 3.4.1 : $D = \text{diag}(d_1, d_2, \dots, d_k)$ ve $H = \text{diag}(h_1, h_2, \dots, h_k)$ olmak üzere

$$\hat{\beta}_1(k) = (X'X + VDV')^{-1} X'Y \text{ ve } \hat{\beta}_2(k) = (X'X + VHV')^{-1} X'Y$$

olarak tanımlansın. Bu durumda

i) Tüm i ler için $d_i \geq h_i$ ise o zaman $MSE(\hat{\beta}_2(k)) - MSE(\hat{\beta}_1(k))$ ifadesi bazı parametre değerleri için p.s.d. olur

ii) Eğer iki veya daha fazla i için $d_i < h_i$ ise o zaman $\Omega(\hat{\beta}_1(k), \hat{\beta}_2(k))$ boş kümedir.

iii) i lerden sadece biri için $d_i < h_i$ ise o zaman $MSE(\hat{\beta}_2(k)) - MSE(\hat{\beta}_1(k))$ p.s.d. olacak şekilde β ve σ^2 değerleri vardır.

İspat (i) : $MSE(\hat{\beta}_2(k)) - MSE(\hat{\beta}_1(k))$ smatrisinin p.s.d. olması için gerek ve yeter koşul

$$\begin{aligned}
V' \{MSE(\hat{\beta}_2(k)) - MSE(\hat{\beta}_1(k))\} V = \\
\sigma^2 \{(\Lambda + H)^{-1} \Lambda (\Lambda + H)^{-1} - (\Lambda + D)^{-1} \Lambda (\Lambda + D)^{-1}\} + \\
+ \{(\Lambda + H)^{-1} H V' \beta\} \{(\Lambda + H)^{-1} H V' \beta\}' \quad (3.4.1) \\
- \{(\Lambda + D)^{-1} D V' \beta\} \{(\Lambda + D)^{-1} D V' \beta\}'
\end{aligned}$$

ifadesinin p.s.d. olmasıdır.

Eğer $\beta = 0$ ise o zaman (3.4.1) ifadesi

$$\sigma^2 \{(\Lambda + H)^{-1} \Lambda (\Lambda + H)^{-1} - (\Lambda + D)^{-1} \Lambda (\Lambda + D)^{-1}\}$$

ifadesine indirgenir. Bu ifade $i=1,2,\dots,k$ için (i, i) inci elemanı

$$\sigma^2 \{\lambda_i (\lambda_i + h_i)^{-1} - \lambda_i (\lambda_i + d_i)^{-1}\}$$

şeklinde olan köşegen matris belirtir. Tüm i ler için $d_i \geq h_i$ varsayımından dolayı bu matris p.s.d. olmalıdır.

İspat (ii) : İki veya daha fazla i değeri için $d_i < h_i$ olduğunu kabul edelim.

Genelliği kaybetmeden $i=1,2$ için $d_i < h_i$ olduğunu kabul edelim.

$$\Lambda_0 = \text{diag}(\lambda_1, \lambda_2)$$

$$D_0 = \text{diag}(d_1, d_2)$$

$$H_0 = \text{diag}(h_1, h_2)$$

ve V_0 ; V nin ilk iki sütunu ile oluşturulan $k \times 2$ tipinde matris olsun. (i) nin ispatından elde edilen sonuçtan ; $MSE(\hat{\beta}_2(k)) - MSE(\hat{\beta}_1(k))$ nin p.s.d. olması için gerek ve yeter koşul

$$\sigma^2 \left\{ (\Lambda + H)^{-1} \Lambda (\Lambda + H)^{-1} - (\Lambda + D)^{-1} \Lambda (\Lambda + D)^{-1} \right\} + \left\{ (\Lambda + H)^{-1} H V' \beta \right\} \left\{ (\Lambda + H)^{-1} H V' \beta \right\}' \quad (3.4.2)$$

ün p.s.d. olmasıdır. 2x2 tipindeki matris

$$\begin{aligned} & \sigma^2 \left\{ (\Lambda_0 + H_0)^{-1} \Lambda_0 (\Lambda_0 + H_0)^{-1} - (\Lambda_0 + D_0)^{-1} \Lambda_0 (\Lambda_0 + D_0)^{-1} \right\} + \\ & + \left\{ (\Lambda_0 + H_0)^{-1} H_0 V_0' \beta \right\} \left\{ (\Lambda_0 + H_0)^{-1} H_0 V_0' \beta \right\}' \end{aligned} \quad (3.4.3)$$

şeklindedir.

$i = 1, 2$ için $d_i < h_i$ olduğundan bu ifadedeki ilk matris n.d. dir. Bundan başka $\left\{ (\Lambda_0 + H_0)^{-1} H_0 V_0' \beta \right\}$ ise 2x1 tipinde bir vektördür. Dolayısıyla Lemma 3.1.1 den; (3.4.3) deki matris hiçbir zaman p.s.d. olamaz. Bu ise $\Omega(\hat{\beta}_1(k), \hat{\beta}_2(k))$ kümesinin boş olması gerektiğini gösterir.

İspat (iii) : i -lerden sadece bir tanesi için $d_i < h_i$ olsun. Genelliği kaybetmeden $d_1 < h_1$ olduğunu kabul edelim. İki durum ele alalım.

a) $k=1$ olduğunu kabul edelim. O zaman $MSE(\hat{\beta}_2(k)) - MSE(\hat{\beta}_1(k))$ nin p.s.d olması için gerek ve yeter koşul

$$\sigma^2 \left\{ \lambda_1 (\lambda_1 + h_1)^{-2} - \lambda_1 (\lambda_1 + d_1)^{-2} \right\} + \left\{ (\lambda_1 + h_1)^{-2} h_1^2 - (\lambda_1 + d_1)^{-2} d_1^2 \right\} \beta^2$$

nin negatif olmayan olmasıdır. Bu koşulun sağlandığı parametre değerleri vardır.

b) $k \geq 2$ olsun

$$\Lambda_2 = \text{diag}(\lambda_2, \lambda_3, \dots, \lambda_k)$$

$$D_2 = \text{diag}(d_2, d_3, \dots, d_k)$$

$$H_2 = \text{diag}(h_2, h_3, \dots, h_k)$$

ve V_1 $k \times 1$ tipinde ve V_2 $k \times (k-1)$ tipinde olmak üzere $V = [V_1, V_2]$ olsun.

Eğer β ; V_1 tarafından gerilen alt uzayda ise o zaman (3.4.1) deki matris

$$\begin{bmatrix} Z & 0 \\ 0 & \bar{Z} \end{bmatrix} \text{ formundadır. Bu ifadede}$$

$$Z = \sigma^2 \left\{ \lambda_1 (\lambda_1 + h_1)^{-2} - \lambda_1 (\lambda_1 + d_1)^{-2} \right\} + \left\{ (\lambda_1 + h_1)^{-2} h_1^2 - (\lambda_1 + d_1)^{-2} d_1^2 \right\} \cdot V_1' \beta \beta' V_1$$

ve

$$\bar{Z} = \sigma^2 \left\{ (\Lambda_2 + H_2)^{-1} \Lambda_2 (\Lambda_2 + H_2)^{-1} - (\Lambda_2 + D_2)^{-1} \Lambda_2 (\Lambda_2 + D_2)^{-1} \right\}$$

dir. Bu matrisin p.s.d. olması için gerek ve yeter koşul Z nin negatif olmayan ve \bar{Z} nin p.s.d. olmasıdır. \bar{Z} matrisi tüm $\sigma^2 > 0$ değerleri için p.s.d. dir ve eğer β sıfırdan farklı ise Z nin negatif olmayan olduğu σ^2 değerleri vardır. Dolayısıyla $MSE(\hat{\beta}_{RR}^2) - MSE(\hat{\beta}_{RR}^1)$; bazı parametre değerleri için p.s.d. dir.

Sonuç 3.4.1 : $\hat{\beta}_{SLS} = \alpha \hat{\beta}_{OLS}$ ve $D = \text{diag}(d_1, d_2, \dots, d_k)$ olmak üzere $\hat{\beta}(k) = (X'X + VDV')X'Y$ olarak tanımlansın. O zaman

i) $MSE(\hat{\beta}_{SLS}) - MSE(\hat{\beta}(k))$ nin p.s.d. olduğu parametre değerlerinin olması için gerek ve yeter koşul i lerden en çok biri için $\left(\frac{1-\alpha}{\alpha} \right) d_i > d_i$ olmasıdır.

ii) $MSE(\hat{\beta}(k)) - MSE(\hat{\beta}_{OLS})$ nin p.s.d. olduğu β ve σ^2 değerlerinin olması için gerek ve yeter koşul i lerden en çok biri için $d_i > \left(\frac{1-\alpha}{\alpha} \right) d_i$ olmasıdır.

Teorem 3.4.2 : $0 \leq \alpha_1 < \alpha_2 < 1$ olsun ve $i=1,2$ için $\hat{\beta}^i_{SLS} = \alpha_i \hat{\beta}_{OLS}$ olarak tanımlansın. O zaman

i) $MSE(\hat{\beta}_{SLS}^2) - MSE(\hat{\beta}_{SLS}^1)$ nin p.s.d. olması için gerek ve yeter koşul

$$\frac{(\alpha_2^2 - \alpha_1^2)^{-1} \{(1 - \alpha_1)^2 - (1 - \alpha_2)^2\} \beta' X' X \beta}{\sigma^2} \leq 1$$

olmasıdır.

ii) Eğer $k \geq 2$ ise $\Omega(\hat{\beta}_{SLS}^2, \hat{\beta}_{SLS}^1)$ boş kümedir

iii) Eğer $k=1$ ise $MSE(\hat{\beta}_{SLS}^1) - MSE(\hat{\beta}_{SLS}^2)$ nin p.s.d. olması için gerek ve yeter

koşul

$$\frac{(\alpha_2^2 - \alpha_1^2)^{-1} \{(1 - \alpha_1)^2 - (1 - \alpha_2)^2\} \beta' X' X \beta}{\sigma^2} \geq 1$$

olmasıdır.

İspat (i) : $MSE(\hat{\beta}_{SLS}^2) - MSE(\hat{\beta}_{SLS}^1)$ matrisi

$$\begin{aligned} & \{\alpha_2^2 \sigma^2 (X'X)^{-1} + (1 - \alpha_2)^2 \beta' \beta\} - \{\alpha_1^2 \sigma^2 (X'X)^{-1} + (1 - \alpha_1)^2 \beta' \beta\} = \\ & = \sigma^2 (\alpha_2^2 - \alpha_1^2)^2 (X'X)^{-1} - \{(1 - \alpha_1)^2 - (1 - \alpha_2)^2\} \beta' \beta \end{aligned}$$

şeklindedir.

$\sigma^2 (\alpha_2^2 - \alpha_1^2)^2 (X'X)^{-1}$ p.d. olduğundan , $MSE(\hat{\beta}_{SLS}^2) - MSE(\hat{\beta}_{SLS}^1)$ nin p.s.d.

olması için gerek ve yeter koşul Lemma 3.1.2 den

$$\frac{(\alpha_2^2 - \alpha_1^2)^{-1} \{(1 - \alpha_1)^2 - (1 - \alpha_2)^2\} \beta' X' X \beta}{\sigma^2} \leq 1$$

olmasıdır.

ii) Bu sonuç Lemma 3.1.3 ve Teorem 3.4.1 in (ii) bölümünden elde edilir.

iii) (i) de elde edilen sonuçtan $MSE(\hat{\beta}_{SLS}^1) - MSE(\hat{\beta}_{SLS}^2)$ ifadesi

$$-\sigma^2 (\alpha_2^2 \alpha_1^2) (X'X)^{-1} + \{(1 - \alpha_1)^2 - (1 - \alpha_2)^2\} \beta' \beta'$$

şeklinde bulunur. $k=1$ olduğundan bu ifade bir skalerdir. Bu ise $MSE(\hat{\beta}_{SLS}^1) - MSE(\hat{\beta}_{SLS}^2)$ nin p.s.d. olması için gerek ve yeter koşulun (ii) deki eşitsizliğin sağlanması olduğunu belirtir.

Teorem 3.4.2 ; $\hat{\beta}_1(k)$ ve $\hat{\beta}_2(k)$ nin her ikisi de SLS tahmin edicisi olduğunda, tüm parametre değerleri için $MSE(\hat{\beta}_2(k)) - MSE(\hat{\beta}_1(k))$ nin tanımsız bir matris olamayacağını belirtir. Bu da; herhangi iki SLS tahmin edicisi karşılaştırıldığında ; bazı β ve σ^2 değerleri için, MSE kriterine göre birinin daima diğerinden iyi olması gerektiğini gösterir. Teorem 3.4.1 ve Sonuç 3.3.2 nin her ikisi de SLS tahmin edicisi olmayan iki RR tahmin edicisinin karşılaştırılmasında bunun doğru olmadığını gösterir.

Özel olarak $\hat{\beta}_1(k)$ ve $\hat{\beta}_2(k)$; Teorem 3.4.1 deki gibi tanımlansın. Eğer iki veya daha fazla i için $d_i < h_i$ ve iki veya daha fazla i için $h_i < d_i$ ise o zaman β ve σ^2 nin tüm olası değerleri için $MSE(\hat{\beta}_2(k)) - MSE(\hat{\beta}_1(k))$ tanımsız bir matris olmalıdır. Bu durumda $\hat{\beta}_1(k)$ ve $\hat{\beta}_2(k)$ MSE kriterine göre karşılaştırılmaz. Bu tahmin edicilerin karşılaştırmaları için başka bir kriter kullanılmalıdır.

3.5. PC Tahmin Edicilerinin Karşılaştırılması

Teorem 3.5.1 : $j=1,2,3$ için V_j , V nin r_j sütunlarını içermek ve $rank[V_1, V_2, V_3] = r_1 + r_2 + r_3$ olmak üzere V_1 , V_2 , V_3 ; V nin alt matrisleri olsun. $\hat{\beta}_{PC}^i$ ise; $\begin{bmatrix} V_1' \\ V_2' \\ V_3' \end{bmatrix} \hat{\beta}_{PC}^i = 0$ $i=1,2$ kısıtlamasına göre rezidü kareler toplamını minimize eden PC tahmin edicisi olsun. O zaman $MSE(\hat{\beta}_{PC}^2) - MSE(\hat{\beta}_{PC}^1)$ nin p.s.d. olmasını sağlayan parametre değerlerinin bulunması için gerek ve yeter koşul $r_2 \leq 1$ olmasıdır.

İspat : Guilkey ve Price (1981) , $r_2 > 1$ olduğunda $\Omega(\hat{\beta}_{PC}^1, \hat{\beta}_{PC}^2)$ nin boş küme olduğunu ve $r_2 = 0$ veya $r_2 = 1$ ve $r_1 = 0$ olduğunda $\Omega(\hat{\beta}_{PC}^1, \hat{\beta}_{PC}^2)$ nin boş küme olmadığını göstermişlerdir. Buna göre teoremin ispatlanması için $r_2 = 1$ ve $r_1 > 0$ olduğunda $MSE(\hat{\beta}_{PC}^2) - MSE(\hat{\beta}_{PC}^1)$ p.s.d. olacak şekilde β ve σ^2 değerlerinin bulunduğu gösterilmesi yeterlidir.

Genelliği kaybetmeden V matrisinin ; V_4 , $k \times (k - (r_1 + r_2 + r_3))$ tipinde olmak üzere $V = [V_1, V_2, V_3, V_4]$ şeklinde parçalandığı kabul edelim.

Eğer $R' = [V_1, V_2, V_3]$ ve $\Lambda_1 = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_{r_1})$ olarak alınırsa o zaman

$$Z = (X'X)^{-1} R' (R(X'X)^{-1} R')^{-1} \text{ ve}$$

$$A = \sigma^2 \begin{bmatrix} \Lambda_1^{-1} & 0 & 0 \\ 0 & -(\lambda_{r_1+1})^{-1} & 0 \\ 0 & 0 & 0 \\ \vdots & \vdots & \vdots \end{bmatrix} + \begin{bmatrix} 0 \\ V_2' \beta \\ V_3' \beta \end{bmatrix} \begin{bmatrix} 0 \\ V_2' \beta \\ V_3' \beta \end{bmatrix}' - \begin{bmatrix} V_1' \beta \\ 0 \\ V_3' \beta \end{bmatrix} \begin{bmatrix} V_1' \beta \\ 0 \\ V_3' \beta \end{bmatrix}'$$

olmak üzere $MSE(\hat{\beta}_{PC}^2) - MSE(\hat{\beta}_{PC}^1) = ZAZ'$ olur. Z nin rankı $r_1 + r_2 + r_3$ olduğundan $MSE(\hat{\beta}_{PC}^2) - MSE(\hat{\beta}_{PC}^1)$ nin p.s.d. olması için gerek ve yeter koşul A nın p.s.d. olmasıdır. β ; V_2 tarafından gerilen alt uzayda bulunan sıfırdan farklı bir vektör olsun.

Bu , $V_1' \beta$ ve $V_3' \beta$ nin her ikisinin de sıfır vektörü olduğunu belirtir. Dolayısıyla

$$A = \begin{bmatrix} \sigma^2 \Lambda_1^{-1} & 0 & 0 \\ 0 & -\sigma^2 (\lambda_{r_1+1})^{-1} + (V_2' \beta)^2 & 0 \\ 0 & 0 & 0 \\ \vdots & \vdots & \vdots \end{bmatrix}$$

olur. $\sigma^2 \Lambda_1^{-1}$, tüm parametre değerleri için p.s.d olduğundan ve

$$-\sigma^2(\lambda_{r+1})^{-1} + (V_2' \beta)^2 \geq 0$$

olacak şekilde bir σ^2 değeri bulunabileceğinden ; $MSE(\hat{\beta}_{PC}^2) - MSE(\hat{\beta}_{PC}^1)$, bazı parametre değerleri için p.s.d. olmalıdır.

Teorem 3.5.1 ; $\hat{\beta}_{PC}^1$ ve $\hat{\beta}_{PC}^2$ katsayıların iki tahmin edicisi ise ve eğer bu iki tahmin edici aynı bileşenlerin çıkarılması (silinmesi) ile elde edilmemişse o zaman, bu tahmin ediciler MSE kriterine göre karşılaştırılmazlar.

Bununla beraber diğer tüm durumlarda ; $MSE(\hat{\beta}_{PC}^2) - MSE(\hat{\beta}_{PC}^1)$ tanımsız bir matris olmayacak şekilde, her zaman bazı parametre değerleri bulunabilir.

3.6. ORR ve PC Tahmin Edicilerinin Karşılaştırılması

Teorem 3.6.1 : $D = \text{diag}(d_1, d_2, \dots, d_k)$ olmak üzere ORR tahmin edicisi $\hat{\beta}(k) = (X'X + VDV')^{-1} X'Y$ ve $\hat{\beta}_{PC}$ PC tahmin edicisi olsun. O zaman

i) $MSE(\hat{\beta}_{PC}) - MSE(\hat{\beta}(k))$ p.s.d. olacak şekilde parametre değerlerinin bulunması için gerek ve yeter koşul $\hat{\beta}_{PC}$ nin , bir bileşeninin silinmesiyle elde edilmesidir.

ii) $MSE(\hat{\beta}(k)) - MSE(\hat{\beta}_{PC})$ p.s.d. olacak şekilde parametre değerlerinin bulunması için gerek ve yeter koşul $\hat{\beta}_{PC}$ nin, bir tanesi dışında tüm bileşenlerinin silinmesiyle elde edilmiş olmasıdır.

İspat (i) : Genelliği kaybetmeden $\hat{\beta}_{PC}$ nin ; $\lambda_1, \lambda_2, \dots, \lambda_r$ lere karşılık gelen bileşenlerin silinmesiyle elde edildiğini kabul edelim.

Dolayısıyla V_1 $k \times r$ tipinde V_2 ise $k \times (k-r)$ tipinde matrisler iken $V = [V_1, V_2]$ olursa o zaman PC tahmin edicisi $\hat{\beta}_{PC}$ $V_1' \hat{\beta}_{PC} = 0$ kısıtlamasını sağlayan RLS tahmin edicisi olur.

$$\Lambda_1 = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_r)$$

$$\Lambda_2 = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_k)$$

$$D_1 = \text{diag}(d_1, d_2, \dots, d_r)$$

$$D_2 = \text{diag}(d_1, d_2, \dots, d_k)$$

olsun. O zaman $MSE(\hat{\beta}_{PC}) - MSE(\hat{\beta}(k))$ nin p.s.d. olması için gerek ve yeter koşul

$$V' \{MSE(\hat{\beta}_{PC}) - MSE(\hat{\beta}(k))\} V = \sigma^2 \begin{bmatrix} 0 & 0 \\ 0 & \Lambda_2^{-1} \end{bmatrix} + \begin{bmatrix} V_1' \beta \beta' V_1 & 0 \\ 0 & 0 \end{bmatrix} - \sigma^2 \begin{bmatrix} (\Lambda_1 + D_1)^{-1} \Lambda_1 (\Lambda_1 + D_1)^{-1} & 0 \\ 0 & (\Lambda_2 + D_2)^{-1} \Lambda_2 (\Lambda_2 + D_2)^{-1} \end{bmatrix} \quad (3.6.1)$$

$$- \begin{bmatrix} (\Lambda_1 + D_1)^{-1} D_1 V_1' \beta \\ (\Lambda_2 + D_2)^{-1} D_2 V_2' \beta \end{bmatrix} \begin{bmatrix} (\Lambda_1 + D_1)^{-1} D_1 V_1' \beta \\ (\Lambda_2 + D_2)^{-1} D_2 V_2' \beta \end{bmatrix}' \text{ nin p.s.d. olmasıdır.}$$

Şimdi iki durumu ele alalım:

a) $r \geq 2$ olduğunu kabul edelim. O zaman (3.6.1) deki matrisin p.s.d. olması için gerek ve yeter koşul

$$-\sigma^2 (\Lambda_1 + D_1)^{-1} \Lambda_1 (\Lambda_1 + D_1)^{-1} + (V_1' \beta) (V_1' \beta)' \quad (3.6.2)$$

nin p.s.d. olmasıdır. Bu ifadedeki ilk matris n.d. olduğundan ve $r \geq 2$ ve $(V_1' \beta)$ $r \times 1$ tipinde olduğundan (3.6.2) deki matris, Lemma 3.1.1 den, p.s.d. olamaz. Bu; $\hat{\beta}_{PC}$ tahmin edicisi iki veya daha fazla bileşenin silinmesiyle elde edilmiş ise o zaman $MSE(\hat{\beta}_{PC}) - MSE(\hat{\beta}(k))$ nin hiçbir zaman p.s.d. olamayacağını belirtir.

b) $r=1$ olsun. β , V_1 tarafından gerilen alt uzayda bulunan sıfırdan farklı herhangi bir vektör olsun. O zaman

$$\left(1 - d_1^2 (\lambda_1 + d_1)^{-2}\right) (V_1' \beta)^2 - \sigma^2 d_1 (\lambda_1 + d_1)^{-2} \quad (3.6.3)$$

negatif olmayan ve

$$\sigma^2 \left\{ \Lambda_2^{-1} - (\Lambda_2 + D_2)^{-1} \Lambda_2 (\Lambda_2 + D_2)^{-1} \right\} \quad (3.6.4)$$

p.s.d. ise (3.6.1) deki matris p.s.d. olur.

(3.6.4) deki matris tüm σ^2 değerleri için p.s.d. dir. β sıfırdan farklı olduğundan (3.6.3) deki ifade, negatif olmayan olacak şekilde σ^2 değerleri bulmak mümkündür. Dolayısıyla $\Omega(\hat{\beta}(k), \hat{\beta}_{PC})$ boş küme değildir.

İspat (ii) : Teoremin (i) kısmının ispatından; $MSE(\hat{\beta}(k)) - MSE(\hat{\beta}_{PC})$ nin p.s.d. olması için gerek ve yeter koşul (3.6.1) deki matrisin n.s.d. olmasıdır. Yine iki durum inceleyelim:

a) $k - r > 1$ olsun. Eğer (3.6.1) deki matris n.s.d. ise o zaman

$$-\sigma^2 \left\{ \Lambda_2^{-1} - (\Lambda_2 + D_2)^{-1} \Lambda_2 (\Lambda_2 + D_2)^{-1} \right\} + \left\{ (\Lambda_2 + D_2)^{-1} D_2 V_2' \beta \right\} \left\{ (\Lambda_2 + D_2)^{-1} D_2 V_2' \beta \right\} \quad (3.6.5)$$

p.s.d. olmalıdır. Bu ifadedeki ilk matris n.d. olduğundan ve $k - r \geq 2$ için $(\Lambda_2 + D_2)^{-1} D_2 V_2' \beta$; $(k - r) \times 1$ tipinde olduğundan; (3.6.5) deki matris Lemma 3.1.1 den dolayı hiçbir zaman p.s.d. olamaz.

Dolayısıyla $\hat{\beta}_{PC}$, (k-1) den az bileşenin silinmesiyle elde edilmiş ise o zaman $\Omega(\hat{\beta}_{PC}, \hat{\beta}(k))$ boş kümedir.

b) $k-r=1$ olsun. β , V_2 tarafından gerilen alt uzayda sıfırdan farklı herhangi bir vektör olsun. O zaman $\sigma^2(\Lambda_1 + D_1)^{-1} \Lambda_1 (\Lambda_1 + D_1)^{-1}$ ps.d. ise ve $-\sigma^2 \{ \lambda_k^{-1} - \lambda_k (\lambda_k + d_k)^{-2} \} + (V_2' \beta)^2$ negatif olmayan ise (3.6.1) deki matris n.s.d. dir. Her iki koşulu da sağlayan β ve σ^2 değerlerinin bulunması mümkündür. Dolayısıyla $\hat{\beta}_{PC}$ bir bileşen dışındaki tüm bileşenlerin silinmesiyle elde edilmiş ise o zaman $\Omega(\hat{\beta}_{PC}, \hat{\beta}(k))$ boş küme değildir.

Teorem 3.6.1, Gunst ve Mason (1976) tarafından verilen sonuç ile çelişir. Gunst ve Mason (1976) tüm $\hat{\beta}_{PC}$ ve $\hat{\beta}(k)$ için $\Omega(\hat{\beta}_{PC}, \hat{\beta}(k))$ ve $\Omega(\hat{\beta}(k), \hat{\beta}_{PC})$ boş kümeler olduğunu söylemiştir. Yukarıdaki sonuçlara göre $\hat{\beta}_{PC}$ ve $\hat{\beta}(k)$ nin MSE kriterine göre karşılaştırılabileceği durumlar vardır. Öte yandan $1 < r < k-1$ olmak üzere; $\hat{\beta}_{PC}$ r tane bileşenin silinmesiyle elde edilmiş ise o zaman $\hat{\beta}_{PC}$ MSE kriterine göre hiçbir ORR tahmin edicisi ile karşılaştırılmaz.

3.7. OLS ve RLS Tahmin Edicilerinin Karşılaştırılması

$$Y = X\beta + \varepsilon, E(Y) = X\beta, Var(Y) = \sigma^2 I$$

lineer regresyon modelini ele alalım. β nin minimum varyanslı yansız tahmin edicisi (MVUE)

$$\hat{\beta} = (X'X)^{-1} X'Y \quad (3.7.1)$$

dir. $\hat{\beta}$ nm mse matrisi

$$MSE(\hat{\beta}) = E\left[(\hat{\beta} - \beta)(\hat{\beta} - \beta)'\right] = \sigma^2(X'X)^{-1} \quad (3.7.2)$$

ve $\hat{\beta}$ nm W – ağırlıklı karesel riski, W pxp tipinde p.d. matris olmak üzere

$$\begin{aligned} \rho_w(\hat{\beta}) &= E\left[(\hat{\beta} - \beta)'W(\hat{\beta} - \beta)\right] \\ &= tr\{WMSE(\hat{\beta})\} \\ &= \sigma^2 tr\{W(X'X)^{-1}\} \end{aligned} \quad (3.7.3)$$

dir. R , mnxp tipinde tam satır ranklı bir matris ve r , mx1 tipinde bir vektör olmak üzere

$$\delta(\beta) = R\beta - r \quad (3.7.4)$$

olsun. $U = (X'X)^{-1}R'$, $T = R(X'X)^{-1}R'$ olmak üzere

$$\beta^* = \hat{\beta} - UT^{-1}\delta(\hat{\beta}) \quad (3.7.5)$$

şeklindeki tahmin ediciyi ele alalım. Eğer $\delta(\beta) = 0$ ise β^* , $M_r = [Y, X\beta, \delta(\beta) = 0, \sigma^2 I]$ kısıtlanmış modeli altında β nm MVUE sidir. Ayrıca β^* , Löwner'in $MSE(\beta^*) \leq MSE(\hat{\beta})$ kısmi sıralamasına göre, $\hat{\beta}$ dan daha iyidir. Yani $MSE(\hat{\beta}) - MSE(\beta^*)$ bir n.n.d matristir. Sonuç olarak bu durumda herhangi bir karesel ağırlıklı riske göre β^* , $\hat{\beta}$ dan daha iyidir. Yani her W için $\rho_w(\beta^*) \leq \rho_w(\hat{\beta})$ dir.

Bununla beraber eğer $\delta(\beta) \neq 0$ ise β^* bir yanlı tahmin edicidir. Bu tahmin edicinin risk matrisi

$$MSE(\beta^*) = \sigma^2 \left[(X'X)^{-1} - UT^{-1}U' \right] + UT^{-1}\delta(\beta)\delta(\beta)'T^{-1}U' \quad (3.7.6)$$

ve ağırlıklı karesel riski

$$\rho_w(\beta^*) = \sigma^2 \text{tr} [W(X'X)^{-1} - UT^{-1}U'] + \delta(\beta)' T^{-1}U'WUT^{-1}\delta(\beta) \quad (3.7.7)$$

şeklindedir. Artık β^* in $\hat{\beta}$ üzerindeki geçerliliği doğru değildir. Bu gibi durumlarda, $\delta(\beta) = 0$ kısıtlamaları yanlış olduğunda β^* , $\hat{\beta}$ ya tercih edilebilir. Toro-Vizcarrando ve Wallace (1968)

$$\lambda(\beta, \sigma^2) = \frac{\delta(\beta)' T^{-1}\delta(\beta)}{2\sigma^2} \quad (3.7.8)$$

olmak üzere $MSE(\beta^*) \leq MSE(\hat{\beta})$ olması için gerek ve yeter koşulun

$$\lambda(\beta, \sigma^2) \leq \frac{1}{2} \quad (3.7.9)$$

olduğunu göstermişlerdir.

Bu sonuç Baksalary ve Pordzik (1985) tarafından, $MSE(\beta^*) \leq MSE(\hat{\beta})$ olması için gerek ve yeter koşulun $m=1$ ve $\lambda(\beta, \sigma^2) \geq \frac{1}{2}$ olduğunun gösterilmesiyle tamamlanmıştır. (3.7.8) deki kriterin en önemli özelliği $\lambda(\beta, \sigma^2)$ nin

$$F = \frac{\delta(\hat{\beta})' T^{-1}\delta(\hat{\beta})}{Y'(I - X(X'X)^{-1}X')Y} \cdot \frac{n-p}{m} \quad (3.7.10)$$

istatistiğinin dağılımının merkezi olmama parametresi olarak alınmasıyla uygun bir hipotezin testi olarak kullanılabilmesidir. Bu da lineer regresyon modeli altında $H_0 : \delta(\beta) = 0$ hipotezinin bilinen testini sağlar.

$F_{k,\ell, \frac{1}{2}}$ dağılımının yani payı k , paydası ℓ serbestlik dereceli ve merkezi olmama parametresi $\frac{1}{2}$ olan, merkezi olmayan F dağılımının kritik noktalarının tablosu Wallace ve Toro-Vizcarrando (1969) tarafından verilmiştir.

(3.7.8) deki kriter $MSE(\beta^*)$ ile $MSE(\hat{\beta})$ nin karşılaştırılmasına dayandığında zor olabilir ve bu durumda kriterin, $\rho_w(\beta^*)$ ve $\rho_w(\hat{\beta})$ nin karşılaştırılmasına dayanan benzeri kullanılabilir. Wallace (1972), Judge ve Bock (1978) in genelleştirilmesiyle

$$\gamma_w(\beta, \sigma^2) = \frac{\delta(\beta)' T^{-1} U' W U T^{-1} \delta(\beta)}{2\sigma^2} \quad (3.7.11)$$

ve

$$K = T^{-1} U' W U T^{-1} \quad (3.7.12)$$

olmak üzere $\rho_w(\beta^*) \leq \rho_w(\hat{\beta})$ eşitsizliğinin sağlanması için gerek ve yeter koşulun

$$\gamma_w(\beta, \sigma^2) \leq \frac{tr(K)}{2} \quad (3.7.13)$$

olduğu bulunur.

$\gamma_w(\beta, \sigma^2)$, lineer regresyon modelinde, bazı hipotezlerin bilinen testini sağlayan bir F istatistiğinin dağılımının merkezi olmama parametresinin bir katı iken (3.7.13) kriteri “işlemsel” olarak adlandırılır.

Genelde (3.7.11) in işlemsel olması çok zormuş gibi görünür. Kaynaklarda bu problemi çözmek için iki yaklaşım kullanılır. Birinci yaklaşımın en önemli noktası, W ağırlık matrisinin özel bir formda seçilmesidir. Özel olarak Wallace (1976), $\gamma_w(\beta, \sigma^2) = \lambda(\beta, \sigma^2)$ ve $tr(K) = m$ olduğunda $W=S$ seçilmesini önermiştir. $F_{k, l, \frac{k}{2}}$ dağılımının kritik noktaları Goodnight ve Wallace (1972) tarafından verilmiştir.

İkinci yaklaşım; (3.7.11) deki gerek ve yeter koşulun ve $\rho_w(\hat{\beta}) \leq \rho_w(\beta^*)$ için tamamlayıcı kısmının, başka koşullarla yer değiştirmesidir. Bu yeni koşullar $\lambda(\beta, \sigma^2)$ ye bağlı olarak sadece yeterli koşullardır. Wallace (1972), Yoncey, Judge ve Bock (1973), Judge ve Bock (1978); $\xi_1 \geq \xi_2 \geq \dots \geq \xi_m > 0$ K nin özdeğerleri olmak üzere

$$\lambda(\beta, \sigma^2) \leq \frac{tr(K)}{2\xi_m} \text{ olduğunda } \rho_w(\hat{\beta}) \leq \rho_w(\beta^*) \text{ olduğunu ve } \lambda(\beta, \sigma^2) \leq \frac{tr(K)}{2\xi_1}$$

olduğunda $\rho_w(\beta^*) \leq \rho_w(\hat{\beta})$ olduğunu göstermişlerdir. Bu yaklaşımın bir sonucu ise lineer regresyon modelinin parametre uzayının

$$\frac{tr(K)}{2\xi_1} \leq \lambda(\beta, \sigma^2) \leq \frac{tr(K)}{2\xi_m} \quad (3.7.14)$$

bölgesinde β^* ve $\hat{\beta}$ nın birbirlerinden daha iyi olduğunun ispat edilememesidir.

Cebirsel açıdan bakıldığında (3.7.14) deki belirsizlik bölgesinin yok olması için gerek ve yeter koşul (3.7.13) de tanımlanan K nın özdeğerlerinin birbirine eşit olmasıdır. Yani herhangi bir $c > 0$ için

$$U'WU = cT \text{ (yani } RS^{-1}WS^{-1}R' = cRS^{-1}R' \text{)} \quad (3.7.15)$$

olmasıdır.

Bu özelliğe bağlı olarak ilginç bir istatistiksel yorum yapılabilir. Bu yorum ise bu özelliğin, (3.7.11) deki kriterin işlemsel olması için gerek ve yeter koşul olduğudur.

Bu; kaynaklarda ; (3.7.15) koşulunu sağlayan W nin seçimine göre $\rho_w(\beta^*)$ ile $\rho_w(\hat{\beta})$ nın karşılaştırılmasında ortaya çıkan problemi çözmek için kullanılan iki yaklaşımın birleştirilmesi olarak düşünülebilir.

Sonuçlar

Aşağıdaki teorem; belirli istenen özelliklere sahip ağırlıklı risk fonksiyonunu belirleyen tüm p.d. W matrislerinin çeşitli karakterizasyonlarını sağlar.

Verilen bir A matrisi için $\mathfrak{R}(A)$, A nın kolon uzayını ; $N(A)$ ise A nın sıfır uzayını gösterebilir. P_A ve Q_A ise A nın $\mathfrak{R}(A)$ ve $N(A)$ üzerindeki ortogonal izdüşümlerini gösterebilir.

Teorem 3.7.1 : W $p \times p$ tipinde p.d. bir matris ve $C = S^{-1}R'(RS^{-1}R')^{-1}RS^{-1}$ olsun. $\xi_1 \geq \xi_2 \geq \dots \geq \xi_m$, CW nin pozitif özdeğerleri olsun. Bu durumda W ağırlıklı karesel riske göre β^* ile $\hat{\beta}$ nin karşılaştırılmasında ortaya çıkan problem için aşağıdaki durumlar denktir.

- (3.7.14) deki belirsizlik bölgesi kaybolur.
- (3.7.11) deki kriter işlemseldir.
- $\xi_1 = \xi_m$, yani CW nin sıfırdan farklı tüm özdeğerleri birbirine eşittir.
- $tr(CW) = m\xi_1$
- $tr(CW) = m\xi_m$
- $[tr(CW)]^2 = mtr[(CW)^2]$
- W ; $c = \frac{tr(CW)}{m}$ olduğu durumda, herhangi bir $c > 0$ için,

$RS^{-1}WS^{-1}R' = CRS^{-1}R'$ eşitliğini sağlar.

h) W nin $W = c(V_1\Lambda^{-1}V_1' + V_1GV_2' + V_2G'V_1' + V_2HV_2')$ şeklinde bir gösterimi vardır. Bu gösterimde c herhangi bir pozitif skalerdir. Ayrıca H , $(p-m)$ dereceli herhangi bir p.d. matris ve G $m \times (p-m)$ tipinde $H - G'\Lambda G$ p.d. olacak şekilde herhangi bir matris olmak üzere ; V_1, V_2 ve Λ , C nin $C = (V_1 : V_2)diag(\Lambda, 0)(V_1 : V_2)'$ şeklinde spektral ayrışımını oluşturur.

i) W nin $W = c(P_{S^{-1}R'} + Q_{S^{-1}R'}Z)S(P_{S^{-1}R'} + Q_{S^{-1}R'}Z)'$ şeklinde bir gösterimi vardır. Bu gösterimde Z , $p \times p$ tipinde $N(RS^{-1}) \cap N(Z') = \{0\}$ ve $\mathfrak{R}(S^{-1}R') \cap \mathfrak{R}(Z'Q_{S^{-1}R'}) = \{0\}$ olacak şekilde herhangi bir matristir.

İspat : A_1A_2 ve A_2A_1 in sıfırdan farklı özdeğerleri aynı olduğundan (Marshall ve Olkin (1979)) $A_1 = UT^{-\frac{1}{2}}$ ve $A_2 = T^{-\frac{1}{2}}U'W$ için (3.7.13) deki K nin özdeğerleri ile CW in sıfırdan farklı özdeğerleri aynıdır. Bu sebeple (3.7.14) ve (3.7.15) den dolayı $(a) \Leftrightarrow (c) \Leftrightarrow (g)$ dir.

$tr(CW) = \xi_1 + \xi_2 + \dots + \xi_m$ olduğundan (d) ve (e) ; CW nin pozitif özdeğerlerinin aritmetik ortalamasının, en küçüklerine veya en büyüklerine eşit olduğunu gösterir. Böylece $(d) \Leftrightarrow (c) \Leftrightarrow (e)$ olduğu açıktır.

(h) ve (i) deki W gösterimi, Baksalay (1984) deki Teorem 2 ve Teorem 3'ün $A = RS^{-1}$, $B = RS^{-1}R'$ ve bilinmeyen p.d. $V = c^{-1}W$ olmak üzere $AVA' = B$ eşitliğine uygulanmasıyla elde edilir.

$b \Rightarrow g$ nin ispatı için R_1 m_1 ranklı $m \times p$ tipinde herhangi bir matris ve r_1 , $m_1 \times 1$ tipinde herhangi bir vektör ve $\delta_1(\beta) = R_1\beta - r_1$ olmak üzere $H_0 : \delta_1(\beta) = 0$ lineer hipotezini göz önüne alalım.

Uygun F istatistiği (3.7.10) daki formda ; m yerine m_1 , $\delta(\hat{\beta})$ yerine $\delta_1(\hat{\beta})$ ve T yerine $T_1 = R_1S^{-1}R_1'$ yazılmasıyla elde edilir. Bu istatistik $\lambda_1(\beta, \sigma^2) = \frac{\delta_1(\beta)'}{2\sigma^2} T_1^{-1} \delta_1(\beta)$

olmak üzere $F_{m_1, n-p, \lambda_1(\beta, \sigma^2)}$ dağılımına sahiptir.

Sonuç olarak (3.7.11) deki kriterin işlemsel olması için gerek ve yeter koşul, her $p \times 1$ tipindeki β vektörü için $\gamma_w(\beta, \sigma^2) = c\lambda_1(\beta, \sigma^2)$ olacak şekilde $c > 0$, R_1 ve r_1 in bulunmasıdır veya eşdeğer olarak her $n \times 1$ tipindeki Y vektörü için

$$\gamma_w(\hat{\beta}, \sigma^2) = c\lambda_1(\hat{\beta}, \sigma^2) \quad (3.7.16)$$

olmasıdır.

$H_0 : \delta_1(\beta) = 0$ hipotezinin testinde kullanılan F istatistiğinin payında ortaya çıkan $2\lambda_1(\hat{\beta}, \sigma^2)$ karesel formu χ^2 dağılımına sahiptir. O zaman (3.7.16) gösterir ki $2c^{-1}\gamma_w(\hat{\beta}, \sigma^2)$ de χ^2 dağılımına sahiptir. Fakat $2c^{-1}\gamma_w(\hat{\beta}, \sigma^2)$ ifadesi $z'Az$ şeklinde karesel formda gösterilebilir. Burada $T = RS^{-1}R'$ ve $U = S^{-1}R'$ olmak üzere

$$z = \sigma^{-1}\delta(\hat{\beta}) \sim N_m(\sigma^{-1}\delta(\beta), T) \quad (3.7.17)$$

ve

$$A = c^{-1}T^{-1}U'WUT^{-1} \quad (3.7.18)$$

dür.

(3.7.17) deki varyans matrisi p.d. olduğundan Carpenter (1950) nin sonucuna göre $z'Az$ nin χ^2 dağılımına sahip olması için gerek ve yeter koşul $ATA = A$ olmasıdır. (3.7.18) açısından bakıldığında ve $RS^{-1}WS^{-1}R'$ nün tekil olmayan olmasından dolayı bu eşitlik (g) deki eşitliğe denktir.

Son olarak (3.7.9) ve (3.7.12) den, eğer g koşulu sağlanırsa o zaman $\gamma_w(\beta, \sigma^2) = c\lambda(\beta, \sigma^2)$ olur. Sonuç olarak $\lambda(\beta, \sigma^2)$, $H_0 : \delta(\beta) = 0$ hipotezinin testinde kullanılan (3.7.10) daki istatistiğin, merkezi olmama parametresi olduğundan dolayı (g) \Rightarrow (b) dir.

Sonuç : β^* ve $\hat{\beta}$ in W ağırlıklı karesel riske göre karşılaştırılması problemi; W nun (3.7.14) deki belirsizlik bölgesi kaybolacak şekildeki seçimine göre değişmez. Bir başka deyişle $RS^{-1}WS^{-1}R' = cRS^{-1}R'$ koşulunu sağlayan her W için $\rho_w(\beta^*) \leq \rho_w(\hat{\beta})$

eşitsizliğinin $\lambda(\beta, \sigma^2) \leq \frac{m}{2}$ ye denk olması yoluyla (3.7.11) deki kriterin işlemsel olması bu problemi değiştirmez.

3.8. ORR ve OLS Tahmin Edicilerinin Karşılaştırılması

Farebrother (1984) ün de aldığı sınıfın bir alternatif ifadesi ,

$$\bar{\beta}(k) = (X'X + kR'R)^{-1} X'Y + k(X'X + kR'R)^{-1} R'r \quad (3.8.1)$$

dir. Bu ifade $F = (X'X + kR'R)^{-1} X'$ ve $f = k(X'X + kR'R)^{-1} R'r$ olmak üzere $\bar{\beta}(k)$ nin $FY + f$ şeklindeki homojen olmayan bir lineer tahmin edeci olarak yazılmasını sağlar. (3.8.1) yazımı ; lineer regresyon modelinde ,

$$\bar{\beta}(k) = (X'X + kR'R)^{-1} (X'X\hat{\beta} + kR'R\hat{\beta}_r) \quad k \geq 0 \quad (3.8.2)$$

formundaki tüm istatistiklerin , mse kriterine göre , lineer tahmin ediciler sınıfında kabul edilebilir olduğunun belirlenmesinde yararlıdır.

Rao (1976) daki Teorem 6.6 nın , β nın tahminine bir uyarlaması olan Baksalary ve Markiewicz (1985) deki Teorem 2 ye göre ; XF nin tüm özdeğerlerinin $[0,1]$ aralığında olması ve f nin $I_p - XF'$ nin sütun uzayında olması için XF matrisinin simetrik olması yeterlidir.

Baksalary, Liski ve Trenkler (1989) daki, Teorem 2 ye göre $\bar{\beta}(k)$ nin kabul edilebilir olması, lineer regresyon modelinin parametre uzayında bazı noktalarda bu tahmin edicinin $\hat{\beta}$ (OLS) dan daha iyi olması gerektiğini gösterir.

OLS ve RLS tahmin edicilerinin ağırlıklı ortalamalarının sınıfı olan (3.8.2) nin ikinci özelliği ;

$$(X'X + H)^{-1} X'Y + h \quad (3.8.3)$$

formundaki tüm kabul edilebilir genel ridge tahmin edicilerini kapsamaktadır. Burada H, herhangi bir n.n.d matris ve kabul edilebilirlik koşuluna göre h, bazı g ler için $h = (X'X + H)^{-1} Hg$ olacak şekilde herhangi bir vektördür. Daha sonra bazı L için $H = L'L$ ise o zaman $R = \left(\frac{1}{\sqrt{k}}\right)L$ ve $\left(\frac{1}{\sqrt{k}}\right)Lg$ nin yazılmasıyla (3.8.1) den (3.8.2) elde edilir.

Daha İyi Olma Kriteri

$$(Y, X\beta, \sigma^2 V(k)) \quad (3.8.4)$$

lineer regresyon modelini ele alalım. Burada $D(Y) = \sigma^2 V(k)$ dır ve

$$V(k) = I_n + \left(\frac{2}{k}\right) XAX' , k \geq 0 \quad (3.8.5)$$

dir. Bu ifadede A , p x p tipinde nnd bir matristir. X in tam sütun ranklı olma varsayımı altında V(k) ile A arasında birebir bir bağıntı vardır.

$W(k) = V(k)^{-1}$ olsun ve Y nin çoklu normal dağılıma sahip olduğunu kabul edelim. (3.8.4) modeli altında $H_0 : R\beta = r$ hipotezinin bilinen testi:

$$F(k) = \frac{[R\tilde{\beta}(k) - r]' [R[X'W(k)X]^{-1} R']^{-1} [R\tilde{\beta}(k) - r]}{Y'[W(k) - W(k)X[X'W(k)X]^{-1} X'W(k)]Y} \cdot \frac{n-p}{m} \quad (3.8.6)$$

istatistiğine dayalı olarak bulunur. Burada $\tilde{\beta}(k)$, β nin BLUE tahmin edicisidir.

$V(k)X$ in sütunları, X in sütunlarının lineer kombinasyonları olduğundan ; Zyskind (1962) den $\tilde{\beta}(k)$ ile β nin OLS tahmin edicisinin aynı olduğu yani $\tilde{\beta}(k) = [X'W(k)X]^{-1} X'W(k)Y = \hat{\beta}$ olduğu görülür (Puntanen ve Styan, 1989).

(3.8.6) istatistiği ; payı m ; paydası $n-p$ serbestlik dereceli ve merkezi olmama parametresi

$$\lambda = (2\sigma^2)^{-1} \delta' [R[X'W(k)X]^{-1} R']^{-1} \delta \quad (3.8.7)$$

olan merkezi olmayan F dağılımından gelmektedir.

β , p boyutlu uzayda değiştiği ; δ , m boyutlu uzayı gerdiğinden dolayı ;

(3.8.7) ve $\lambda(k) = (2\sigma^2)^{-1} \delta' \left[\left(\frac{2}{k} \right) I_m + R(X'X)^{-1} R' \right]^{-1} \delta$ ifadelerinden, her β için

$\lambda(k) = \lambda$ olması için gerek ve yeter koşul

$$\left(\frac{2}{k} \right) I_m + R(X'X)^{-1} R' = R[X'W(k)X]^{-1} R' \quad (3.8.8)$$

olmasıdır.

(3.8.5) formundaki $V(k)$ için $W(k) = V(k)^{-1}$ ifadesi

$$W(k) = I_n - X(X'X)^{-1} X' + X(X'X)^{-1} \left[(X'X)^{-1} + \left(\frac{2}{k} \right) A \right]^{-1} (X'X)^{-1} X'$$

şeklinde yazılabilir ve böylece (3.8.8) eşitliği

$$RAR' = I_m \quad (3.8.9)$$

formunu alır.

(3.8.9) un n.n.d. çözümünün bir genel sunumu mümkündür (Baksalary (1984)

Teorem 1 Bu durumdan bağımsız olmakla birlikte (3.8.9) un özel olarak

$A = R'(RR')^{-2}R$ tarafından sağlandığı görülebilir ve bu düşüncelerle aşağıdaki teorem elde edilir.

Teorem 3.8.1 : $\lambda(k) = (2\sigma^2)^{-1} \delta' \left[\left(\frac{2}{k} \right) I_m + R(X'X)^{-1}R' \right]^{-1} \delta$ şeklindeki $\lambda(k)$

ifadesi ; $V(k) = I_n + \left(\frac{2}{k} \right) XR'(RR')^{-2}RX'$ olmak üzere (3.8.4) modeli altında;

$H_0 : R\beta = r$ hipotezinin bilinen testi için kullanılan (3.8.6) deki F istatistiğinin merkezi olmama parametresidir.

Bu teorem ; k ya bağlı varyans matrisine sahip bir lineer modelde; R verildiğinde k değerinin seçimi probleminin $H_0 : R\beta = r$ hipotezinin testinin terimleri cinsinden yazılabileceğini göstererek (3.8.2) formundaki genelleştirilmiş ridge tahmin edicilerin iyiliğinin ölçülmesine yeni bir bakış getirir.

$k \rightarrow \infty$ iken $V(k) = I_n$ ve $r = 0$ olduğundan, bu teorem Toro-Vizcarrando ve Wallace (1968) tarafından oluşturulan kriterin işlemsel bir yorumunu içerir.

Toro – Vizcarrando ve Wallace (1968) in bu sonucu, Judge ve Bock (1978) tarafından detaylı olarak incelenen preliminary test tahmin edicilerinin oluşturulmasında çok önemlidir.

$m = p$, $R = I_p$ ve $r = 0$ olduğu özel durumda $\bar{\beta}_0(k) = (X'X + kI_p)^{-1}X'Y$ şeklindeki ORR tahmin edicisinin $\hat{\beta}$ (yani OLS) dan daha iyi olması için gerek ve yeter koşul $\lambda_0(k) = (2\sigma^2)^{-1} \beta' \left[\left(\frac{2}{k} \right) I_p + (X'X)^{-1} \right]^{-1} \beta$ olmak üzere $\lambda_0(k) \leq \frac{1}{2}$ olmasıdır.

Bu teoreme göre ; $\lambda_0(k)$; $\left\{ Y, X\beta, \sigma^2 \left[I_n + \left(\frac{2}{k} \right) X'X \right] \right\}$ modeli altında

$H_0 : \beta = 0$ hipotezinin bilinen testini sağlayan

$$F_0 = \frac{\hat{\beta}' \left[\left(\frac{2}{k} \right) I_p + (X'X)^{-1} \right]^{-1} \hat{\beta}}{Y' \left[I_n - X(X'X)^{-1} X' \right] Y} \cdot \frac{n-p}{p}$$

istatistiğinin merkezi olmama parametresidir.

$k \rightarrow 0$ iken $F_0(k) \rightarrow 0$ dır. Dolayısıyla ORR tahmin edicisi $\bar{\beta}_0(k)$, k nin yeterince küçük değerleri için OLS tahmin edicisi $\hat{\beta}$ dan daha iyidir (Theobald (1974)).

3.9. RLS ve ORR Tahmin Edicilerinin Karşılaştırılması

$$Y = X\beta + \varepsilon, E(\varepsilon) = 0, E(\varepsilon\varepsilon') = \sigma^2 I_n \quad (3.9.1)$$

lineer regresyon modelini ele alalım. Parametrelerin $R\beta = r$ kısıtlamasını sağladığı kabul edilsin. Bu kısıtlamada R ve r $m \times p$ ve $m \times 1$ tipinde belirli sayılardan oluşmuş matris olsun ve R nin rankı m olsun. r rastgele olsaydı

$$r = R\beta + \eta, E(\eta) = 0, E(\eta\eta') = \left(\frac{\sigma^2}{k} \right) I_m \quad (3.9.2)$$

yazılabilir ve (3.9.1) ve (3.9.2) birleştirilmiş modelindeki β nin BLUE si olarak

$$\hat{\beta}(k) = (X'X + kR'r)^{-1} (X'Y + kR'r) \quad (3.9.3)$$

tahmin edicisi elde edilebilirdi. Ancak r ve β belirli olduğundan $\hat{\beta}(k)$ nin bu şekilde elde edilmesi ve eşdeğer olarak bayes yöntemi ile elde edilmesi geçersiz olur.

Bununla beraber

$$\hat{\beta}(k) = (X'X + kR'R)^{-1} (X'X\hat{\beta} + kR'R\hat{\beta}_R) \quad (3.9.4)$$

ifadesi ; OLS tahmin edicisi

$$\hat{\beta}_{OLS} = (X'X)^{-1} X'Y \quad (3.9.5)$$

ve $V = (X'X)^{-1}$ ve $\hat{\beta}(0) = \hat{\beta}$ ve $\hat{\beta}(\infty) = \hat{\beta}_R$ olmak üzere

$$\hat{\beta}_{RLS} = \hat{\beta} - VR'(RVR')^{-1}(R\hat{\beta} - r) \quad (3.9.6)$$

şeklinde tanımlanan RLS tahmin edicisinin bir ağırlıklı ortalamasıdır. Çünkü ,

$$\begin{aligned} \hat{\beta}(k) &= \hat{\beta} - k(X'X + kR'R)^{-1} R'(R\hat{\beta} - r) \\ &= \hat{\beta} - kVR'(I_m + kRVR')^{-1}(R\hat{\beta} - r) \end{aligned} \quad (3.9.7)$$

ve

$$(X'X + kR'R)VR' = R'(I_m + kRVR') \quad (3.9.8)$$

dir.

Bundan başka , Kuks ve Olman (1972) ; $\hat{\beta}(k)$ nin ; β nin $(R\hat{\beta} - r)'(R\hat{\beta} - r) \leq \frac{\sigma^2}{k}$ kısıtlamasına göre minimax lineer tahmin edicisi olduğunu belirtmişlerdi.

Lemma 3.9.1 : A , $m \times m$ tipinde simetrik bir p.d. matris ; F $n \times m$ tipinde m ranklı bir matris ; b $m \times 1$ tipinde bir matris ve d bir skaler olsun. O zaman $F[dA - bb']F'$ ifadesinin $(n-m)$ tane sıfır özdeğeri ; $(m-1)$ tane işareti d tarafından belirlenen özdeğeri ve işareti $d - b'A^{-1}b$ tarafından belirlenen bir tane özdeğeri vardır.

İspat : $m \times m$ tipinde tekil olmayan bir P matrisi için $A = PP'$ dir. $n \times (n-m)$ tipindeki $(n-m)$ ranklı bir G matrisi için $[FP \ G]$ tekil olmayan matristir.

$c = P^{-1}b$ olsun. O zaman

$$\begin{bmatrix} dI_m - cc' & 0 \\ 0 & 0_{n-m} \end{bmatrix} \quad (3.9.9)$$

nin $(n-m)$ tane sıfır özdeğeri ; $(m-1)$ tane d ye eşit özdeğeri ve bir tane $d - cc' = d - b'A^{-1}b$ ye eşit özdeğeri olur.

Sylvester ' in kuralına göre

$$F[dA - bb']F' = [FP \quad G] \begin{bmatrix} dI_m & 0 \\ 0 & 0_{n-m} \end{bmatrix} \begin{bmatrix} P'F' \\ G' \end{bmatrix} \quad (3.9.10)$$

un özdeğerlerinin işareti ile (3.9.9) un özdeğerlerinin işareti aynıdır. Böylelikle ispat tamamlanır.

Mse Matrisleri

$\hat{\beta}(k)$ nin mse matrisi ; $\delta = R\beta - r$ olmak üzere

$$MSE(\hat{\beta}(k)) = (X'X + kR'R)^{-1} (\sigma^2 X'X + k^2 R' \delta \delta' R) (X'X + kR'R)^{-1} \quad (3.9.11)$$

şeklinde tanımlanır.

Ayrıca $MSE(\hat{\beta}(0)) = \sigma^2 (X'X)^{-1}$ dir. Öyle ki

$$A = \frac{2\sigma^2}{k} I_m + \sigma^2 RVR' - \delta\delta' \quad (3.9.12)$$

olmak üzere

$$MSE(\hat{\beta}(0)) - MSE(\hat{\beta}(k)) = k^2 (X'X + kR'R)^{-1} R'AR (X'X + kR'R)^{-1} \quad (3.9.13)$$

olur. Daha önce elde edilen sonucun kullanılmasıyla , $MSE(\hat{\beta}(0)) - MSE(\hat{\beta}(k))$ nin en fazla bir tane negatif özdeğere sahip olduğu görülür. Ayrıca $MSE(\hat{\beta}(0)) - MSE(\hat{\beta}(k))$ nin p.s.d olması için gerek ve yeter koşul.

$$(R\beta - r)' \left[\frac{2}{k} I_m + RVR' \right]^{-1} (R\beta - r) \leq \sigma^2 \quad (3.9.14)$$

olmasıdır. Bu ise

$$k \leq \frac{2\sigma^2}{\delta'\delta} \quad (3.9.15)$$

veya

$$(R\beta - r)'(RVR')^{-1}(R\beta - r) \leq \sigma^2 \quad (3.9.16)$$

olduğunda sağlanır.

$\hat{\beta}(\infty)$ nin mse matrisi ; $Q = RVR'$ olmak üzere

$$MSE(\hat{\beta}(\infty)) = \sigma^2 V - \sigma^2 VR'Q^{-1}RV + VR'Q^{-1}\delta\delta'Q^{-1}RV \quad (3.9.17)$$

şeklinde olur öyle ki :

$$\begin{aligned} B &= \sigma^2 Q^{-1} - Q^{-1}\delta\delta'Q^{-1} - kQ^{-1}\delta\delta' - k\delta\delta'Q^{-1} \\ &= \sigma^2 Q^{-1} + k^2\delta\delta' - (Q^{-1} + kI)\delta\delta'(Q^{-1} + kI_m) \end{aligned} \quad (3.9.18)$$

olmak üzere

$$MSE(\hat{\beta}(k)) - MSE(\hat{\beta}(\infty)) = (X'X + kR'R)^{-1} R'BR(X'X + kR'R)^{-1} \quad (3.9.19)$$

olur.

Daha önce elde edilen sonucun kullanılmasıyla $MSE(\hat{\beta}(k)) - MSE(\hat{\beta}(\infty))$ un en fazla bir tane negatif özdeğeri olduğu söylenebilir ve $MSE(\hat{\beta}(k)) - MSE(\hat{\beta}(\infty))$ un p.s.d. olması için gerek ve yeter koşul

$$\delta'(Q^{-1} + kI_m)(\sigma^2 Q^{-1} + k^2\delta\delta')^{-1}(Q^{-1} + kI_m)\delta \leq 1 \quad (3.9.20)$$

olmasıdır.

Ayrıca ,

$$(\sigma^2 Q^{-1} + k^2 \delta \delta')^{-1} = \frac{(\sigma^2 + k^2 \delta' Q \delta) Q - k^2 Q \delta \delta' Q}{\sigma^2 (\sigma^2 + k^2 \delta' Q \delta)} \quad (3.9.21)$$

olduğundan (3.9.20) nin sağlanması için gerek ve yeter koşul

$$k^2 [\delta' Q \delta \delta' Q^{-1} \delta - (\delta' \delta)^2] + 2k\sigma^2 \delta' \delta + \sigma^2 \delta' Q^{-1} \delta \leq \sigma^4 \quad (3.9.22)$$

olmasıdır. (3.9.21) in $k=0$ için sağlanması için gerek ve yeter koşul $\delta' Q^{-1} \delta \leq \sigma^2$ olmasıdır ve $k > 0$ için sağlanması için $\delta' Q^{-1} \delta \leq \sigma^2$ olması yeterli koşuldur. Dolayısıyla (3.9.16) $MSE(\hat{\beta}(k)) - MSE(\hat{\beta}(\infty))$ un p.s.d. olması için gerekli ; $MSE(\hat{\beta}(0)) - MSE(\hat{\beta}(k))$ nın p.s.d. olması için yeterli ve $MSE(\hat{\beta}(0)) - MSE(\hat{\beta}(\infty))$ p.s.d. olması için gerek ve yeter koşuldur. Bu sonuç Toro-Vizcarrondo ve Wallace (1968) dan elde edilir.

Sıfır Tahmin Edicisi

Ridge parametresi k nın nonstokastik olduğu kabul edildiğinden MSE (matris) kriteri kullanılmıştır. Ancak pratikte k rastgele olur ve skaler mse kriteri kullanılır. Yapılan analize göre ridge tipi tahmin edicilerin simülasyon çalışmalarında RLS tahmin edicilerinin kullanılması faydalıdır.

Örneğin $R = I_p$ ve $r=0$ olduğunda RLS tahmin edicisi, mse matrisi $\beta\beta'$ olan sıfır tahmin edicisi olur.

Bu mse matrisinin izinin bulunması ve $\beta\beta'$ ya bölünmesiyle Hoerl, Kennard ve Baldwin (1975) in çalışmasından

$$E \left[\frac{L_1^2(\infty)}{r^2} \right] = \left[\frac{L_1^2(\infty)}{r^2} \right] = 1$$

ve benzer şekilde (3.9.20) ye bölünmesiyle (Hoerl, Kennard (1977)) , Vinod (1976) ın çalışmasından $MSE(\hat{\beta}(0))=0.05$ elde edilir.

3.10. AUGRR ve OLS Tahmin Edicilerinin Karşılaştırılması

$$Y = X\beta + \varepsilon, \varepsilon \sim N(0, \sigma^2 I_n)$$

lineer regresyon modelini ele alalım. $\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_p)$ köşegen elemanları $X'X$ ün özdeğerleri olan bir köşegen matris ve T , $p \times p$ tipinde $T'X'XT = \Lambda$ ve $T'T = TT' = I_p$ koşulunu sağlayan, $X'X$ in özvektörlerinden oluşan bir matris , $Z = XT$, $\gamma = T'\beta$, $Z'Z = T'X'XT = \Lambda$ olmak üzere lineer regresyon modeli

$$Y = Z\gamma + \varepsilon, \varepsilon \sim N(0, \sigma^2 I_n) \quad (3.10.1)$$

şeklinde yazılabilir.

Bu modelde γ nın OLS tahmin edicisi

$$g = (Z'Z)^{-1} Z'Y = \Lambda^{-1} Y \quad (3.10.2)$$

GRR tahmin edicisi $K = \text{diag}(k_1, k_2, \dots, k_p)$ olmak üzere

$$\begin{aligned} \hat{\gamma} &= (\Lambda + K)^{-1} Z'Y \\ &= (\Lambda + K)^{-1} \Lambda g \\ &= (I - (\Lambda + K)^{-1} K) g \end{aligned} \quad (3.10.3)$$

şeklindedir. $\gamma = T'\beta$ olduğundan β nın GRR tahmin edicisi

$$\hat{\beta} = T'\hat{\gamma} = (X'X + TKT')^{-1} X'Y \quad (3.10.4)$$

olarak bulunur. $k = k_1 = k_2 = \dots = k_p$ olduğunda γ nın ORR tahmin edicisi

$$\hat{\gamma} = (\Lambda + kI_p)^{-1} Z'Y \quad (3.10.5)$$

şeklinde yazılabilir.

γ nın AURR tahmin edicisi

$$\begin{aligned} \tilde{\gamma} &= (I + (\Lambda + K)^{-1} K) \hat{\gamma} \\ &= \left\{ I - ((\Lambda + K)^{-1} K) \right\} g \end{aligned} \quad (3.10.6)$$

olarak yazılır.

AURR tahmin edicisi ile GRR ve OLS tahmin edicilerinin karşılaştırılmasını ele alalım. Ridge parametresi k_i nin non stokastik olduğunu kabul edelim.

OLS tahmin edicisi g nin i -nci bileşeninin mse si

$$MSE(g_i) = E(g_i - \gamma_i)^2 = \frac{\sigma^2}{\lambda_i} \quad (i = 1, 2, \dots, p) \quad (3.10.7)$$

olarak yazılır. Buradan OLS tahmin edicisi g nin mse si

$$MSE(g) = E(g - \gamma)'(g - \gamma) = \sum_{i=1}^p \frac{\sigma^2}{\lambda_i} \quad (3.10.8)$$

ve g nin tahmin edici mse si (PMSE)

$$PMSE(g) = E(g - \gamma)' X'X(g - \gamma) = p\sigma^2 \quad (3.10.9)$$

şeklindedir. Benzer şekilde $\hat{\gamma}_i$ nin mse si

$$MSE(\hat{\gamma}_i) = \frac{(\gamma_i^2 k_i^2 + \sigma^2 \lambda_i)}{(\lambda_i + k_i)^2} \quad (i = 1, 2, \dots, p) \quad (3.10.10)$$

olarak yazılabilir.

Lemma 3.10.1 : $MSE(\tilde{y}_i)$;

$$MSE(\tilde{y}_i) = \frac{\sigma^2 \lambda_i (2k_i + \lambda_i)^2}{(\lambda_i + k_i)^4 + \frac{k_i^4 \gamma_i^2}{(\lambda_i + k_i)^4}} \quad (3.10.11)$$

$$= \frac{(\gamma_i^2 k_i^4 + 4\sigma^2 \lambda_i k_i^2 + 4\sigma^2 \lambda_i^2 k_i + \sigma^2 \lambda_i^3)}{(\lambda_i + k_i)^4}$$

şeklindedir.

AUGRR ve GRR tahmin edicilerinin mse lerinin karşılaştırılmalarını inceleyelim. Aşağıdaki teorem Singh ve Chaubey (1987) den elde edilir.

Teorem 3.10.1 :

$$K1_i = \frac{\left[3\sigma^2 - \lambda_i \gamma_i^2 + \left\{ (3\sigma^2 + \lambda_i \gamma_i^2)^2 + 4\sigma^2 \lambda_i \gamma_i^2 \right\}^{\frac{1}{2}} \right]}{4\gamma_i^2} > 0 \quad (3.10.12)$$

olmak üzere

$0 < k_i < K1_i$ için $MSE(\tilde{y}_i) > MSE(\hat{y}_i)$ dir.

$K1_i < k_i < \infty$ için $MSE(\tilde{y}_i) < MSE(\hat{y}_i)$ dir.

İspat : Teorem 3.10.1 , (3.10.10) ve (3.10.11) ifadelerinden elde edilen

$$(\lambda_i + k_i)^4 \{MSE(\tilde{y}_i) - MSE(\hat{y}_i)\} = -\lambda_i k_i \{2\gamma_i^2 k_i^2 - (3\sigma^2 - \lambda_i \gamma_i^2) k_i - 2\sigma^2 \lambda_i\}$$

eşitliğinin kullanılmasıyla ispatlanır.

Şimdi de AUGRR ve OLS tahmin edicilerinin mse lerinin karşılaştırılmasını ele alalım.

Teorem 3.10.2 :

$$K2_i = \frac{\left\{ 2\sigma^2 \lambda_i + (2\sigma^4 \lambda_i^2 + 2\sigma^2 \lambda_i^3 \gamma_i^2)^{\frac{1}{2}} \right\}}{(\lambda_i \gamma_i^2 - \sigma^2)} \quad (3.10.13)$$

olsun. O zaman $0 < k_i < \infty$ için eğer $\lambda_i \gamma_i^2 - \sigma^2 \leq 0$ ise $MSE(\tilde{y}_i) < MSE(g_i)$ olur.

Eğer $\lambda_i \gamma_i^2 - \sigma^2 > 0$ ise o zaman ; $0 < k_i < K2_i$ için $MSE(\tilde{y}_i) < MSE(g_i)$ ve $K2_i < k_i < \infty$ için $MSE(\tilde{y}_i) > MSE(g_i)$ olacak şekilde bir pozitif $K2_i$ vardır.

İspat : Teorem 3.10.2 ; (3.10.7) ve (3.10.11) in kullanılmasıyla elde edilen

$$\lambda_i (\lambda_i + k_i)^4 \{MSE(\tilde{y}_i) - MSE(g_i)\} = k_i^2 \{(\lambda_i \gamma_i^2 - \sigma^2) k_i^2 - 4\sigma^2 \lambda_i k_i - 2\sigma^2 \lambda_i^2\}$$

eşitliği ile ispatlanır. Bundan başka $MSE(\tilde{y}_i)$ nin k_i ye göre türevinin alınmasıyla

$$\frac{dMSE(\tilde{y}_i)}{dk_i} = \frac{4\lambda_i k_i (\gamma_i^2 k_i^2 - 2\sigma^2 k_i - \sigma^2 \lambda_i)}{(\lambda_i + k_i)^5} \quad (3.10.14)$$

eşitliği elde edilir.

(3.10.14) den ; $MSE(\tilde{y}_i)$ nin minimum olduğu optimal k_i değeri

$$k_i^* = \frac{\left\{ \sigma^2 + (\sigma^4 + \sigma^2 \lambda_i \gamma_i^2)^{\frac{1}{2}} \right\}}{\gamma_i^2} = \left(\frac{\sigma^2}{\gamma_i^2} \right) \left[1 + \left\{ 1 + \lambda_i \left(\frac{\gamma_i^2}{\sigma^2} \right) \right\}^{\frac{1}{2}} \right] \quad (3.10.15)$$

bulunur.

(3.10.12) , (3.10.13) ve (3.10.15) den ; eğer $\lambda_i \gamma_i^2 - \sigma^2 > 0$ ise o zaman

$$K1_i < k_i^* < K2_i \quad (3.10.16)$$

olduğu bulunur.

AUORR tahmin edicisi ele alındığında $k = k_1 = k_2 = \dots = k_p$ olmak üzere k nın elde edilmesi için (3.10.15) deki k_i^* nın harmonik ortalaması kullanılabilir. Bu ise

$$k^h = \frac{p\sigma^2}{\sum_{i=1}^p \frac{\gamma_i^2}{\left\{1 + \left(1 + \lambda_i \left(\frac{\gamma_i^2}{\sigma^2}\right)\right)^{\frac{1}{2}}\right\}}} \quad (3.10.17)$$

şeklindedir.

Hoerl , Kennard ve Baldwin (1975) in bahsettiği sebepten dolayı harmonik ortalama regresyon katsayılarına uygulanabilir. (3.10.17) deki k^h , γ ve σ^2 bilinmeyen parametrelerine bağlıdır.

k nın Seçimi

Ridge parametresinin kullanılmasıyla elde edilen ; Hoerl , Kennard ve Baldwin (1975) in önerdiği ve Lawless ve Wang (1976) nın önerdiği işlemsel ridge parametrelerini ele alalım.

Birincisi (3.10.17) deki ; optimal ridge parametrelerinin değerlerinin harmonik ortalaması ; bilinmeyen γ ve σ^2 ye bağlı olduğundan γ ve σ^2 yerine OLS tahminleri

yazılır. (3.10.17)'ye uygun işlemsel ridge parametresi, $\hat{\sigma}^2 = \frac{(Y - Zg)'(Y - Zg)}{(n - p)}$ olmak

üzere

$$k_{HMO} = \frac{p\sigma^2}{\sum_{i=1}^p \left[\frac{g_i^2}{\left\{ 1 + \left(1 + \lambda_i \left(\frac{g_i^2}{\hat{\sigma}^2} \right) \right)^{\frac{1}{2}} \right\}} \right]} \quad (3.10.18)$$

şeklindedir.

Bu ridge parametresi HMO ridge parametresi olarak adlandırılır.

İkincisi Hoerl , Kennard ve Baldwin' in (HKB) ridge parametresi

$$k_{HKB} = \frac{p\hat{\sigma}^2}{g'g} \quad (3.10.19)$$

dir.

Üçüncüsü Lawless ve Wang 'ın (LW) ridge parametresi

$$k_{LW} = \frac{p\hat{\sigma}^2}{\sum_{i=1}^p \lambda_i g_i} \quad (3.10.20)$$

dir.

3.11. RRR Tahmin Edicisinin Mse Kriterine Göre Performansı

RRR tahmin edicisi yanlı bir tahmin edici olduğundan bu tahmin edicinin performansını incelemek için uygun olan kriter mse dir.

MSE kriterine göre RRR tahmin edicisi ile RLS ile OLS tahmin edicilerini karşılaştıralım. İlk olarak $\delta = 0$ koşulunu inceleyelim.

I. DURUM : (Kısıtlamalar Doğru Olduğunda Yani $\delta = 0$ ise)

RR ile RLS nin Karşılaştırılması

RRR tahmin edicisi her zaman yanlı bir tahmin edicidir. Fakat bu durumda RLS tahmin edicisi yansızdır. Tahmin edicilerin performansını örneklem varyansı açısından inceleyelim.

$$V(\beta^*) - V(\beta^*(k)) = \sigma^2 (A - WAW') = \sigma^2 W (kS^{-1}A + kAS^{-1} + k^2S^{-1}AS^{-1}) W' \quad (3.11.1)$$

olur. S^{-1} p.d. ve A ps.d. olduğundan $S^{-1}A$ nın tüm karakteristik kökleri negatif olmayandır, öyle ki $S^{-1}A$ bir nnd matristir.

$S^{-1}AS^{-1}$ p.s.d. olduğundan (3.11.1) den , $k \geq 0$ için $[V(\beta^*) - V(\beta^*(k))]$ p.s.d. dir.

Teorem 3.11.1 : RRR tahmin edicisi yanlıdır. Ancak RLS tahmin edicisi kısıtlamalar doğru olduğunda yansız olur. Bununla beraber RRR tahmin edicisinin örneklem varyansı her zaman RLS tahmin edicisinin örneklem varyansından küçük veya eşittir.

RRR tahmin edicisinin RLS tahmin edicisi ile performanslarını mse kriterine göre karşılaştıralım

P , $P'SP = \Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_p)$ ve $B = P'AP$ psd olacak şekilde bir ortogonal matris olmak üzere $WAW' = P(\Lambda + kI_p)^{-1} \Lambda B (\Lambda + kI_p)^{-1} P'$ dir. Genelliği

kaybetmeden $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p > 0$ olduğunu kabul edelim. Bu durumda b_{ii} , B nin i-nci köşegen elemanı $b_{ii} \geq 0$ olmak üzere

$$\text{tr}(WAW') = \sum_{i=1}^p \frac{\lambda_i^2 b_{ii}}{(\lambda_i + k)^2} \quad (3.11.2)$$

olur. Ayrıca $\alpha = P'\beta$ olmak üzere

$$\beta'S(k)^{-2} \beta = \alpha'(\Lambda + kI_p)^{-2} \alpha = \sum_{i=1}^p \frac{\alpha_i^2}{(\lambda_i + k)^2}$$

dir. (3.11.2) ifadesini RRR tahmin edicisinin mse sinde yerine yazarsak

$$\text{MSE}(\beta^*(k)) = \sigma^2 \sum_{i=1}^p \frac{\lambda_i^2 b_{ii}}{(\lambda_i + k)^2} + k^2 \sum_{i=1}^p \frac{\alpha_i^2}{(\lambda_i + k)^2} \quad (3.11.3)$$

bulunur. Toplam varyans $\delta_1(k)$, k nın bir sürekli ve monoton azalan fonksiyonudur. $\delta_1(k)$ nın k ya göre türevi, $k \rightarrow 0^+$ ve $\lambda_p \rightarrow 0$ iken $-\infty$ a yaklaşır. Ayrıca yanlılığın karesi olan $\delta_2(k)$ ise k nın bir sürekli ve monoton artan fonksiyonudur. $\delta_2(k)$ nın türevi $k \rightarrow 0^+$ iken sıfıra yaklaşır.

(3.11.3) deki eşitliğin her iki tarafının k ya göre türevinin alınması ve elde edilen ifadenin düzenlenmesiyle

$$\frac{d\text{MSE}(\beta^*(k))}{dk} = 2 \sum_{i=1}^p \frac{\lambda_i}{(\lambda_i + k)^3} [k\alpha_i^2 - \sigma^2 \lambda_i b_{ii}] \quad (3.11.4)$$

bulunur.

$$\alpha_i^{*2} = \frac{\alpha_i^2}{\lambda_i b_{ii}} \text{ olarak tanımlanmak üzere } \alpha_i^{*2} \text{, } i=1,2,\dots,p \text{ için } \alpha_i^{*2} \text{ lerin en}$$

büyüğü olsun. Tüm $i=1,2,\dots,p$ ler için $b_{ii} \geq 0$ ve $\lambda_i > 0$ olduğundan, (3.11.4) den

$\text{MSE}(\beta^*(k))$ nın birinci türevinin negatif olması için yeterli koşulun

$$0 < k < \frac{\sigma^2}{\alpha_{\max}^{*2}}$$

olduğu açıkça görülebilir.

Teorem 3.11.2 : Kısıtlamalar doğru olduğunda , $0 < k < \frac{\sigma^2}{\alpha_{\max}^{*2}}$ aralığında, her

zaman, RRR tahmin edicisinin mse si RLS tahmin edicisinin mse sinden küçük olacak şekilde bir k değeri vardır.

$$C = P'S^{-1}R'(RS^{-1}R')^{-1}RS^{-1}P \quad \text{psd} \quad \text{olmak} \quad \text{üzere} \quad B \quad \text{nin} \quad \text{tanımından}$$

$B = P'AP = \Lambda^{-1} - C$ yazılabilir. Böylece c_{ii} , C nin i-nci köşegen elemanı olmak üzere

$$b_{ii} = \frac{1}{\lambda_i} - c_{ii} \leq \frac{1}{\lambda_i}$$

olur. O zaman $\frac{\sigma^2}{\alpha_{\max}^{*2}} \leq \frac{\sigma^2}{\alpha_{\max}^2}$ olduğu açıktır.

Kesin ön bilginin birleştirilmesiyle elde edilen en önemli sonuçlardan biri şudur : MSE kriterine göre incelendiğinde ; RRR tahmin edicisinin RLS tahmin edicisinden daha iyi olmasını sağlayan k değerlerinin aralığı ; ORR tahmin edicisinin OLS tahmin edicisinden daha iyi olmasını sağlayan k değerlerinin aralığı ile karşılaştırıldığında daha dardır.

RRR ve ORR Tahmin Edicilerinin Karşılaştırılması

RRR ve ORR tahmin edicileri, β nin yanlış tahmin edicileridir. Ve kısıtlamalar doğru olduğunda her ikisinin yanlışlığı aynıdır ve $-kS(k)^{-1}\beta$ ya eşittir.

RRR VE ORR tahmin edicilerinin varyans-kovaryans matrislerinin farkı

$$\text{Var}(\hat{\beta}(k)) - \text{Var}(\beta^*(k)) = S(k)^{-1}S^{-1}R'(RS^{-1}R')RS^{-1}S(k)^{-1}$$

şeklindedir ve bu matris $k \geq 0$ için bir p.s.d. matristir. Dolayısıyla örneklem varyansı kriterine göre , RRR tahmin edicisi her zaman ORR tahmin edicisinden daha iyidir. Tahmin edicilerin yanlılıkları ve örneklem varyansları ile ilgili sonuçlar birleştirildiğinde; kısıtlamaların doğru olduğu durumda RRR tahmin edicisinin her zaman ORR tahmin edicisinden daha iyi olduğu söylenebilir.

II. Durum: (Kısıtlamalar Doğru Olmadığında Yani $\delta \neq 0$ ise)

RRR tahmin edicisinin performansı δ ya bağlıdır. Yani kısıtlamaların hangi oranda sağlanmadığına bağlıdır. Bir başka deyişle bu tahmin edicinin mse kriterine göre üstünlüğü, doğru olmadığı kabul edilen kısıtlamalara göre olan yanlılığın miktarına bağlıdır.

Bu durum için, RRR tahmin edicisi mse kriterine göre RLS ve ORR tahmin edicisinden daha iyi olacak şekildeki δ ya bağlı olan k nın mümkün olan değerlerinin seçimini gerektiren koşulları inceleyelim.

RRR Ve RLS Tahmin Edicilerinin Karşılaştırılması

Kısıtlamalar doğru olmadığında RRR ve RLS tahmin edicileri yanlıdır. $WS^{-1} = S(k)^{-1}$ olduğundan RRR tahmin edicisinin mse si

$$MSE(\beta^*(k)) = \sigma^2 \text{tr}(WAW') + [S(k)^{-1}\delta^* - kS(k)^{-1}\beta] [S(k)^{-1}\delta^* - kS(k)^{-1}\beta] \quad (3.11.5)$$

olarak bulunur. Bazı cebirsel işlemlerden sonra bu ifade,

$$MSE(\beta^*(k)) = \sum_{i=1}^p \frac{1}{(\lambda_i + k)^2} [\sigma^2 \lambda_i^2 b_{ii} + k^2 \alpha_i^2 + \delta_i^{*2} - 2k\alpha_i \delta_i^*] \quad (3.11.6)$$

olarak yazılabilir.

(3.11.6) daki eşitliğin k ya göre türevi alınır

$$\frac{dMSE(\beta^*(k))}{dk} = 2 \sum_{i=1}^p \frac{1}{(\lambda_i + k)^3} \left[k(\lambda_i \alpha_i^2 + \alpha_i \delta_i^*) - (\sigma^2 \lambda_i b_{ii} + \delta_i^{*2} + \alpha_i \lambda_i \delta_i^*) \right] \quad (3.11.7)$$

elde edilir. Böylece $MSE(\beta^*(k))$ nın birinci türevinin negatif olması için yeterli koşul bulunabilir.

$$\text{Teorem 3.11.1: } k^* = \frac{\min_i (\sigma^2 \lambda_i^2 b_{ii} + \delta_i^* + \alpha_i \lambda_i \delta_i^*)}{\max_i (\lambda_i \alpha_i^2 + \alpha_i \delta_i^*)} \text{ olarak tanımlansın ve } k$$

pozitif olsun. O zaman RRR tahmin edicisi , $0 < k < k^* < \infty$ için, mse kriterine göre RLS tahmin edicisinden daha iyidir.

RRR Ve ORR Tahmin Edicilerinin Karşılaştırılması

β nın ORR tahmin edicisi $\hat{\beta}(k)$ nın mse si

$$MSE(\hat{\beta}(k)) = \sum_{i=1}^p \frac{1}{(\lambda_i + k)^2} [\sigma^2 \lambda_i + k^2 \alpha_i^2] \quad (3.11.8)$$

dir. (3.11.6) ve (3.11.8) den

$$MSE(\hat{\beta}(k)) - MSE(\beta^*(k)) = \sum_{i=1}^p \frac{1}{(\lambda_i + k)^2} \left[2k\alpha_i \delta_i^* - (\delta_i^{*2} - \sigma^2 \lambda_i + \sigma^2 \lambda_i^2 b_{ii}) \right] \quad (3.11.9)$$

bulunur. $\tilde{k} = \frac{\max_i (\delta_i^{*2} - \sigma^2 \lambda_i + \sigma^2 \lambda_i^2 b_{ii})}{\min_i (2\alpha_i \delta_i^*)}$ olmak üzere. $k > \tilde{k} > 0$

olduğunda (3.11.9) daki eşitliğin sağ tarafı pozitiftir. Dolayısıyla aşağıdaki teorem elde edilir.

$$\text{Teorem 3.11.2: } \tilde{k} = \frac{\max_i (\delta_i^{*2} - \sigma^2 \lambda_i + \sigma^2 \lambda_i^2 b_{ii})}{\min_i (2\alpha_i \delta_i^*)} \text{ olmak üzere RRR tahmin}$$

edicisinin ORR tahmin edicisinden daha küçük mse ye sahip olması için yeterli koşul $0 < \tilde{k} < k < \infty$ olmasıdır.

3.12. Modified Ridge Tahmin Edicisi İle Kısıtlanmış Ridge Regresyon Tahmin Edicisinin Mse Kriterine Göre Karşılaştırılması

Swindel (1976) ön bilgiye dayanan düzenlenmiş bir ridge tahmin edicisi ortaya koymuştur. Ayrıca OLS tahmin edicisinin mse matrisinin, ortaya koyduğu tahmin edicinin mse matrisinden daha büyük olması için gerekli ve yeterli koşulu oluşturmuştur. Bununla beraber Swindel, oluşturduğu tahmin edicinin mse özellikleri ile ORR tahmin edicisinin mse özelliklerini karşılaştırmamıştır. Pliskin (1987), ORR tahmin edicisi ile modified ridge tahmin edicisinin mse matrislerini karşılaştırmıştır. Sarkar (1992), ridge regresyon düşüncesini RLS tahmin edicisine uygulayarak RRR tahmin edicisini oluşturmuştur. Ayrıca, bu tahmin edicinin; kısıtlamalar doğru olduğunda, mse kriterine göre RLS ve ORR tahmin edicilerinden daha iyi olduğunu göstermiştir. Bu tahmin edicinin, kısıtlamalar doğru olmadığında, diğer iki tahmin ediciden daha iyi olduğu koşullar da oluşturulmuştur.

$$Y = X\beta + \varepsilon \quad \varepsilon \sim N(0, \sigma^2 I) \quad (3.12.1)$$

modelinde OLS tahmin edicisi

$$\hat{\beta} = (X'X)^{-1} X'Y \quad (3.12.2)$$

şeklindedir.

$$R\beta = r \quad (3.12.3)$$

lineer kısıtlamalar kümesini ele alalım.

RLS tahmin edicisi $S = X'X$ olmak üzere

$$\beta^* = \hat{\beta} + S^{-1} R'(RS^{-1}R')^{-1} (r - R\hat{\beta}) \quad (3.12.4)$$

şeklindedir.

ORR tahmin edicilerinin ailesi $W = (I_p + kS^{-1})^{-1}$ olmak üzere Hoerl ve Kennard (1970, a,b) tarafından

$$\hat{\beta}(k) = (X'X + kI_p)^{-1} X'Y \quad k \geq 0 \quad (3.12.5)$$

şeklinde tanımlanmıştır.

Sarkar (1992) tarafından tanımlanan RRR tahmin edicisi ; β^* , β nın RLS tahmin edicisi olmak üzere

$$\beta^*(k) = W\beta^* , \quad k \geq 0 \quad (3.12.6)$$

şeklindedir. Swindel (1976); b^* , β nın belirli bir ön tahmin vektörü olmak üzere

$$b(k, b^*) = (X'X + kI_p)^{-1} (X'Y + kb^*) , \quad k \geq 0 \quad (3.12.7)$$

tahmin edicisini tanımlamıştır. Uygulamada b^* , β üzerindeki ön bilgiyi veya hipotezleri yansıtacak şekilde seçilebilir. Genel olarak ; β üzerindeki ön bilgi β_0 ortalama, vektörüne ve V_0 kovaryans matrisine sahip bir p-değişkenli normal dağılımla tanımlanabiliyorsa , β nın Bayes tahmin edicisi

$$\hat{\beta}_B = (X'X + \sigma^2 V_0^{-1})^{-1} (X'Y + \sigma^2 V_0^{-1} \beta_0)$$

şeklinde bulunur.

$\beta^*(k)$ ve $b(k, b^*)$ yanlı tahmin edicilerdir. Bir $\tilde{\beta}$ tahmin edicisinin yanlılığı ve varyansı mse matrisi ile

$$MSE(\tilde{\beta}) = Var(\tilde{\beta}) + Bias(\tilde{\beta})Bias(\tilde{\beta})' \quad (3.12.8)$$

ve

$$Bias(\tilde{\beta}) = E(\tilde{\beta}) - \beta \quad (3.12.9)$$

ve

$$Var(\tilde{\beta}) = E \left[(\tilde{\beta} - E(\tilde{\beta})) (\tilde{\beta} - E(\tilde{\beta}))' \right] \quad (3.12.10)$$

şeklinde bulunur.

β nın bir değeri için ; $MSE(\tilde{\beta}_1) - MSE(\tilde{\beta}_2)$ bir n.n.d matris ise $\tilde{\beta}_2$ tahmin edicisi $\tilde{\beta}_1$ tahmin edicisine tercih edilir. Ancak tersi doğru değildir (Theobald,1974).

$b(k, b^*)$ Ve $\beta^*(k)$ Tahmin Edicilerinin Karşılaştırılması

β^* in varyans kovaryans matrisi

$$Var(\beta^*) = \sigma^2 \left(S^{-1} - S^{-1} R' (RS^{-1} R')^{-1} RS^{-1} \right) \quad (3.12.11)$$

şeklindedir. $\beta^*(k) = W\beta^*$ olduğundan $A = S^{-1} - S^{-1} R' (RS^{-1} R')^{-1} RS^{-1}$ ve $W = (I_p + kS^{-1})^{-1}$ olmak üzere

$$Var(\beta^*(k)) = \sigma^2 W A W' \quad (3.12.12)$$

dir. $\beta^*(k)$ nin beklenen değeri ve yanlışlığı $\delta = r - R\beta$ ve $\delta^* = R' (RS^{-1} R')^{-1} \delta$ olmak üzere

$$E[\beta^*(k)] = W\beta + WS^{-1} R' (RS^{-1} R')^{-1} \delta$$

$$Bias(\beta^*(k)) = WS^{-1} \delta^* - kS(k)^{-1} \beta \quad (3.12.13)$$

olarak bulunur.

(3.12.2) ve (3.12.3) ün birleştirilmesiyle $S(k) = S + kI$ olmak üzere

$$MSE(\beta^*(k)) = \sigma^2 W A W' + [WS^{-1} \delta^* - kS(k)^{-1} \beta] [WS^{-1} \delta^* - kS(k)^{-1} \beta]' \quad (3.12.14)$$

bulunur. Eğer $\delta = 0$ ise yani kısıtlamalar doğru ise o zaman $MSE(\beta^*(k))$

$$MSE(\beta^*(k)) = \sigma^2 W A W' + k^2 S(k)^{-1} \beta \beta' S(k)^{-1} \quad (3.12.15)$$

ifadesine indirgenir.

$b(k, b^*)$ nin varyans-kovaryans matrisi

$$\text{Var}(b(k, b^*)) = \sigma^2 S(k)^{-1} SS(k)^{-1} \quad (3.12.16)$$

olarak verilir. Bundan başka $b(k, b^*)$ nin yanlılığı

$$\text{Bias}(b(k, b^*)) = -kS(k)^{-1} (\beta - b^*) \quad (3.12.17)$$

şeklindedir.

(3.12.16) ve (3.12.17) nin birleştirilmesiyle

$$\text{MSE}(b(k, b^*)) = \sigma^2 S(k)^{-1} SS(k)^{-1} + k^2 S(k)^{-1} (\beta - b^*)(\beta - b^*)' S(k)^{-1} \quad (3.12.18)$$

bulunur.

$\beta^*(k)$ ve $b(k, b^*)$ nin örneklem varyans-kovaryans matrislerine göre performansı

$$\text{Var}(b(k, b^*)) - \text{Var}(\beta^*(k)) = \sigma^2 [S(k)^{-1} SS(k)^{-1} - WAW'] \quad (3.12.19)$$

olarak bulunur.

$W = SS(k)^{-1}$ olduğundan (3.12.19) kullanılarak

$$\text{Var}(b(k, b^*)) - \text{Var}(\beta^*(k)) = \sigma^2 S(k)^{-1} [R'(RS^{-1}R')^{-1}R] S(k)^{-1} \quad (3.12.20)$$

olduğu kolayca görülebilir.

Pozitif yarı tanımlı matrislerinin temel özellikleri kullanılarak $[\text{Var}(b(k, b^*)) - \text{Var}(\beta^*(k))]$ nin bir p.s.d. matris olduğu görülebilir ve RRR tahmin edicisinin modified ridge tahmin edicisinden daha küçük örneklem varyans – kovaryans matrisine sahip olduğu sonucuna varılabilir. Şimdi k nin nonstokastik olduğu ve lineer kısıtlamaların doğru olduğu varsayımı altında ; aynı k değeri için hesaplanan $\beta^*(k)$ ve $b(k, b^*)$ tahmin edicilerinin mse matrislerinin karşılaştırılmasını inceleyelim.

Lemma 3.12.1 : (Gruber, 1990) A , $n \times n$ tipinde simetrik bir matris ve a , $n \times 1$ tipinde bir vektör ve d pozitif bir skaler olsun . O zaman $dA - aa'$ nin p.s.d. olması

için gerek ve yeter koşul A nın n.n.d. olması, a nın A nın kolon uzayında olması ve A^+ , A nın Moore-Penrose tersi olmak üzere $a'A^+a \leq d$ olmasıdır.

(3.12.15) ve (3.12.18) kullanılmasıyla

$$\begin{aligned} MSE(b(k, b^*)) - MSE(\beta^*(k)) &= S(k)^{-1} \left[\sigma^2 (S - SAS) + k^2 \left\{ (\beta - b^*)(\beta - b^*)' - \beta\beta' \right\} \right] S(k)^{-1} \\ &= S(k)^{-1} \left[\sigma^2 R'(RS^{-1}R')^{-1}R + k^2 \left\{ (\beta - b^*)(\beta - b^*)' - \beta\beta' \right\} \right] S(k)^{-1} \quad (3.12.21) \end{aligned}$$

olduğu bulunur.

Böylelikle, lineer kısıtlamalar doğru olduğunda aşağıdaki teoremler oluşturulur.

Teorem 3.12.1 : $D = (\beta - b^*)(\beta - b^*)' - \beta\beta'$ olsun. Her iki tahmin edici aynı k değeri için hesaplandığında, eğer D p.s.d. matris ise o zaman $MSE(b(k, b^*)) - MSE(\beta^*(k))$ p.s.d. dir.

İspat : Eğer $b(k, b^*)$ ile $\beta^*(k)$ nın hesaplanmasında aynı k değeri kullanılıyorsa (3.12.21) elde edilir. $R'(RS^{-1}R')^{-1}R$ p.s.d. olduğundan; eğer D p.s.d. matris ise o zaman $MSE(b(k, b^*)) - MSE(\beta^*(k))$ p.s.d. dir.

Teorem 3.12.1 den sezgisel olarak; $\beta^*(k)$ ve $b(k, b^*)$ aynı k değeri için hesaplandığında; $\beta^*(k)$ nın varyans kovaryans matrisinin, $b(k, b^*)$ in varyans kovaryans matrisinden daha küçük olduğu söylenebilir.

$k^2 S(k)^{-1} D S(k)^{-1}$ matrisi; iki tahmin edicinin yanlışlık vektörleri arasındaki farkı yansıtır. Dolayısıyla D nin bir p.s.d. matris olması yeterlidir.

RRR tahmin edicisinin performansı D nin p.s.d. olması koşulu ile karakterize edilmiştir. Bu küme, D , hiçbir zaman boş değildir. Çünkü $\beta = c(\beta - b^*)$ veya

$$b^* = \frac{c-1}{c} \beta \text{ alınırsa yani } |c| < 1 \text{ için } D = \frac{1-c^2}{c^2} \beta\beta' \text{ p.s.d. dir.}$$

$\beta = c(\beta - b^*)$ veya $b^* = \frac{c-1}{c}\beta$ ($|c| < 1$) koşulu ; $\beta^*(k)$ nin yanlılığının $b(k, b^*)$ nin yanlılığından küçük olmasını sağlar.

$\beta = c(\beta - b^*)$ veya eşdeğer olarak $b^* = \frac{c-1}{c}\beta$ koşulu altında , (3.12.7) den aşağıdaki gibi istatistiksel olarak ilginç tahmin ediciler elde edilir.

i) $c \rightarrow 1$ iken $b(k, b^*)$, $\hat{\beta}(k)$ ya yaklaşır. (yani $b^*=0$)

ii) $b^* = \frac{c-1}{c}\beta$ da , β yerine yansız tahmin edici $\hat{\beta}$ nin yazılması ile $c \rightarrow \infty$ ile $\hat{b}^* = \hat{\beta}$ elde edilir. Dolayısıyla (3.12.7) den $b(k, \hat{b}^*) = \hat{\beta}$ elde edilir. Öte yandan $k \rightarrow 0$ iken $b(k, b^*)$ nin $\hat{\beta}$ ya yaklaştığı görülür.

iii) $b^* = \frac{c-1}{c}\beta$ da β yerine RLS tahmin edicisi β^* in yazılması ve $c \rightarrow \infty$ ile $\hat{b}^* = \beta^*$ elde edilir. Dolayısıyla (3.12.7) den

$$b(k, \beta^*) = (X'X + kI_p)^{-1}(X'Y + k\beta^*) , k \geq 0 \quad (3.12.22)$$

elde edilir.

Farebrother (1984) tarafından tanımlanan tahmin edici sınıfını göz önüne alalım.

$$\hat{\beta}(k) = (X'X + kR'R)^{-1}(X'Y + kR'R\beta^*) , k \geq 0 \quad (3.12.23)$$

$R = I_p$ özel durumunda bu , (3.12.22) ye döndüştür.

iv) $b^* = \frac{c-1}{c}\beta$ daki β nin OLS tahmin edicisi olan $\hat{\beta}$ alınırsa $\hat{b}^* = \frac{c-1}{c}\hat{\beta}$

elde edilir. O zaman (3.12.7) den $k \rightarrow \infty$ iken $b(k, \hat{b}^*) = \left(1 - \frac{1}{c}\right)\hat{\beta}$ olarak bulunur.

Açıkça $c > 1$ için shrunken tahmin edici elde edilir.(Mayer ve Willke, 1973).

Teorem 3.12.2 : Eğer $R^+R\beta = \beta$ ve $k^2 < \frac{\sigma^2}{\beta'S^{-1}\beta}$ ise

$MSE(b(k, b^*)) - MSE(\beta^*(k)) \geq 0$ dir.

İspat: $k^2(\beta - b^*)(\beta - b^*)'$ n.n.d. olduğundan; eğer

$\sigma^2 R'(RS^{-1}R')^{-1}R - k^2\beta\beta' \geq 0$ ise o zaman $MSE(b(k, b^*)) - MSE(\beta^*(k))$ p.s.d. dir.

Lemma 3.12.1 den $\sigma^2 R'(RS^{-1}R')^{-1}R - k^2\beta\beta' \geq 0$ olması için gerek ve yeter koşul

$R^+R\beta = \beta$ ve $k^2 < \frac{\sigma^2}{\beta'S^{-1}\beta}$ olmasıdır. Bu ise ispatı tamamlar.

Şimdi , k nın nonstokhostik olduğu ve parametrik kısıtlamaların doğru olmadığı varsayımı altında ; aynı k değeri için hesaplanan $\beta^*(k)$ ve $b(k, b^*)$ tahmin edicilerinin mse matrislerinin karşılaştırılmalarını inceleyelim.

Teorem 3.12.3 : Eğer $R^+R\beta = \beta$ ve $(k\beta - \delta^*)'S^{-1}(k\beta - \delta^*) < \sigma^2$ ise $MSE(b(k, b^*)) - MSE(\beta^*(k)) \geq 0$ dir.

İspat : Kısıtlamalar doğru olmadığında

$$\begin{aligned} MSE(b(k, b^*)) - MSE(\beta^*(k)) &= S(k)^{-1} \left[\sigma^2 R'(RS^{-1}R')^{-1}R + k^2(\beta - b^*)(\beta - b^*)' \right] S(k)^{-1} \\ &\quad - \left[WS^{-1}\delta^* - kS(k)^{-1}\beta \right] \left[WS^{-1}\delta^* - kS(k)^{-1}\beta \right]' \\ &= S(k)^{-1} \left[\sigma^2 R'(RS^{-1}R')^{-1}R + k^2(\beta - b^*)(\beta - b^*)' - (k\beta - \delta^*)(k\beta - \delta^*)' \right] S(k)^{-1} \end{aligned}$$

olarak bulunur.

$k^2(\beta - b^*)(\beta - b^*)'$ n.n.d. olduğundan Lemma 3.12.1 in kullanılmasıyla;

eğer $R^+R\beta = \beta$ ve $(k\beta - \delta^*)S^{-1}(k\beta - \delta^*) < \sigma^2$ ise

$MSE(b(k, b^*)) - MSE(\beta^*(k)) \geq 0$ olduğu bulunur. Böylece ispat tamamlanır.

3.13. Gauss-Markoff Modelindeki Minimum Yanlılığa Sahip En İyi Lineer Tahmin Ediciler

$$Y = X\beta + \varepsilon \quad (3.13.1)$$

genel Gauss-Markov (GGM) lineer modelini ele alalım. Bu modelde $E(\varepsilon) = 0$ ve $Var(\varepsilon) = \sigma^2 G$ dir. X matrisi tam sütun ranklı olmayabilir veya G singüler olabilir. Bu sebepten bu model $(Y, X\beta, \sigma^2 G)$ GGM modeli olarak adlandırılır. Böyle bir genel modelin uygulanabilirliği Rao (1971) tarafından ortaya konulmuştur ve LS nin birleştirilmiş teorisi oluşturulmuştur.

Rao'nun bu modeli , lineer tahmin için kriter oluşturmuş ve β nin lineer parametrik fonksiyonları için lineer tahmin edicilerin bir genel formunu vermiştir. Bu kriterlerden bazıları minimum yanlı lineer tahmin edicilerin (MB-LE) bulunması için veya yeni minimum varyans, minimum yanlı lineer tahmin edici (MV-MB-LE) elde etmek için oluşturulmuştur. Rao (1973 b) ayrıca karesel hata kaybı altında bir lineer tahmin edicinin bayes riskinin minimize edilmesine eşdeğer olan, bir ön bilgi altında mse nin minimize edilmesini ele almıştır. Burada lineer tahminin minimum mse-minimum yanlılık, lineer tahmin (MMS-MB-LE) olarak adlandırılan bir başka kriteri üzerinde durulacaktır.

Lineer Tahmin İçin Bir Kriter

$p'\beta$ lineer fonksiyonunun , $L'Y$ lineer tahmin edicisi tarafından tahmin edildiğini kabul edelim. $L'Y$ nin mse si

$$E[(L'Y - p'\beta)^2] = \sigma^2 L'GL + (X'L - p)' \beta \beta' (X'L - p) \quad (3.13.2)$$

şeklindedir.

Rao (1973) ; W bilinen bir n.n.d. matris olmak üzere , kriterin

$$\begin{aligned}
S &= L'GL + (X'L - p)'W(X'L - p) \\
&= L'(G + XWX')L - 2p'WX'L + p'Wp
\end{aligned} \tag{3.13.3}$$

şeklinde minimize edilmesini ele almıştır. Sonuçta elde edilen tahmin edici en iyi lineer tahmin edici (BLE) olarak adlandırılır.

Chipman (1964) , B n.n.d. olmak üzere minimum yanlı lineer tahmin ediciyi (MB-LE) elde etmek için L ye göre minimize edilmesini ele almıştır. $Var(L'Y) = L'GL$ yi minimize eden bu sınıftaki L yi kısıtlamıştır. BLE ve MB-LE için çözümler Rao (1973-b) tarafından verilmiştir.

MB-LE sınıfında S nin minimize edilmesi ele alınacak ve bu kriter MMS-MB-LE olarak adlandırılacaktır.

Bu noktada dikkat edilmesi gerek husus $W=B$ olduğunda MMS-MBLE ve MV-MB-LE kriterlerinin eşdeğer olduğudur.

Teorem 3.13.1 : $(X'L - p)'B(X'L - p)$ yi minimize eden L nin sınıfında $S = L'GL + (X'L - p)'W(X'L - p)$ nin minimumu

$$L = E_1XWp + E_2XBp \tag{3.13.4}$$

şeklinde elde edilir. Burada B ve W verilen iki n.n.d. matris ve E_1, E_2

$$\begin{bmatrix} G + XWX' & XBX' \\ XBX' & 0 \end{bmatrix}^- = \begin{bmatrix} E_1 & E_2 \\ E_3 & E_4 \end{bmatrix} \tag{3.13.5}$$

eşitliğinden elde edilir ve $(...)^-$ genelleştirilmiş inversi gösterir.

İspat : λ , Lanrange çarpanı vektörü olmak üzere, L nin

$$(G + XWX')L + XBX'\lambda = XWp \tag{3.13.6}$$

$$XBX'L = XBp \tag{3.13.7}$$

eşitliklerini sağladığının gösterilmesi kolaydır.

(3.13.6) , (3.13.7) nin çözülebileceği ve (3.13.6) , (3.13.7) nin (3.13.4) de olduğu gibi her çözümlerinin S nin minimumunu verdiği gözlenebilir.

Hatırlatma 3.13.1 : Parçalanmış bir matrisin genelleştirilmiş tersinin gösteriminin kullanılmasıyla E_1 ve E_2 nin bir seçimi ; $N = M + A^2$ ve $A = XB X'$ olmak üzere

$$E_1 = N^{-1} \left[I - A(AN^{-1}A)^{-1} AN^{-1} \right] \quad E_2 = N^{-1} A(AN^{-1}A)^{-1} \quad (3.13.8)$$

şeklinde verilir.

$p'\beta$ nin, $p'HY$ formundaki MMS-MB-LE olan lineer tahmin edicisi bulunmak istenebilir. Bu problemin genel bir çözümü aşağıdaki teorem ile verilebilir.

Teorem 3.13.2 : H için bir genel çözüm

$$T = trHG H' + tr(HX - I)W(HX - I)' \quad (3.13.9)$$

ifadesi

$$\|HX - I\|^2 = tr(HX - I)B(HX - I)' \quad (3.13.10)$$

ifadesini minimize eden H matrisleri , sınıfındaki minimum olacak şekilde

$$H = WX'E_{1g} + BX'E_{2g} \quad (3.13.11)$$

şeklinde verilir. Burada

$$E_{1g} = E_1 + C_1(I - ME_1 - AE_2) + C_2AE_1 + (I - E_1M - E_2A)D_1 + E_1AD_2 \quad (3.13.12)$$

$$E_{2g} = E_2 + C_1(-ME_2 - AE_4) + C_2(I - AE_2) + (I - E_1M - E_2A)D_3 + E_1AD_4 \quad (3.13.13)$$

şeklindedir . $C_1, C_2, D_1, D_2, D_3, D_4$ rastgele olarak seçilmiştir.

İspat : Λ , Lagrange çarpanlarının matrisi olmak üzere

$$H(G + XWX') + \Lambda XB X' = WX' \quad (3.13.14)$$

ve

$$HXBX' = BX' \quad (3.13.15)$$

eşitliklerinin minimize edilmesi probleminin bir genel çözümünün bulunmasıyla teorem ispatlanabilir. H için çözüm

$$[H : \Lambda] = [WX' : BX'] \begin{bmatrix} G + XWX' & XBX' \\ XBX' & 0 \end{bmatrix} \quad (3.13.16)$$

nın çözülmesiyle elde edilir.

Hatırlatma 3.13.2 : Lineer Bayes tahmin edicisi ; $E(\beta\beta') = \sigma^2 W$ olmak üzere yukarıdaki teoreme benzer çözüme sahiptir.

3.14. Lineer Tahmin Edicilerin Karşılaştırılması

$\tilde{\beta} = CY$ şeklindeki herhangi bir homojen lineer tahmin edicinin

$$MSE(\tilde{\beta}) = (CX - I)\beta\beta'(CX - I)' + \sigma^2 CC' \quad (3.14.1)$$

şeklindeki ikinci dereceden moment matrisini ele alalım. (3.14.1) ifadesi

$$MSE(\tilde{\beta}) = B(\tilde{\beta})B(\tilde{\beta})' + Cov(\tilde{\beta}) \quad (3.14.2)$$

şeklinde yazılabilir. Burada $\tilde{\beta}$ nin yanlılığı

$$B(\tilde{\beta}) = E(\tilde{\beta}) - \beta = (CX - I)\beta \quad (3.14.3)$$

ve varyans-kovaryans matrisi

$$Cov(\tilde{\beta}) = E \left[(\tilde{\beta} - E(\tilde{\beta}))(\tilde{\beta} - E(\tilde{\beta}))' \right] = \sigma^2 CC' \quad (3.14.4)$$

şeklinindedir.

Bu durumda $j=1,2$ için $\tilde{\beta}_j = C_j Y$ şeklindeki iki tahmin edici için

$$\begin{aligned}
MSE(\tilde{\beta}_1) - MSE(\tilde{\beta}_2) &= Cov(\tilde{\beta}_1) - Cov(\tilde{\beta}_2) + B(\tilde{\beta}_1)B(\tilde{\beta}_1)' - B(\tilde{\beta}_2)B(\tilde{\beta}_2)' \\
&= \sigma^2(C_1C_1' - C_2C_2') + B(\tilde{\beta}_1)B(\tilde{\beta}_1)' - B(\tilde{\beta}_2)B(\tilde{\beta}_2)' \quad (3.14.5)
\end{aligned}$$

elde edilir. Şimdi $MSE(\tilde{\beta}_1) - MSE(\tilde{\beta}_2)$ nin p.d. olduğu koşulları inceleyelim. Bu koşullar $C_1C_1' - C_2C_2'$ matrisinin pozitif tanımlılığına bağlıdır.

Yeterli Bir Koşul

$C_1C_1' - C_2C_2'$ nün p.d. olduğunu kabul edelim. Eğer $\sigma^2(C_1C_1' - C_2C_2') - B(\tilde{\beta}_2)B(\tilde{\beta}_2)'$ p.d. ise $MSE(\tilde{\beta}_1) - MSE(\tilde{\beta}_2)$ p.d. dir.

Farebrother (1976) nin bir sonucu uygulanabilir. A bir p.d. matris olsun c sıfırdan farklı bir vektör ve θ pozitif bir skaler olsun. O zaman $\theta A - cc'$ nün p.d. olması için gerek ve yeter koşul $c'A^{-1}c$ nin θ dan küçük olmasıdır. Buradan aşağıdaki teorem elde edilir.

Teorem 3.14.1 : $j=1,2$ için $\tilde{\beta}_j = C_j Y$; $C_1C_1' - C_2C_2'$ p.d. olacak şekilde iki homojen lineer tahmin edici olsun. Bundan başka

$$\beta'(C_2X - I)'(C_1C_1' - C_2C_2')^{-1}(C_2X - I)\beta < \sigma^2 \quad (3.14.6)$$

eşitsizliği sağlansın. O zaman $MSE(\tilde{\beta}_1) - MSE(\tilde{\beta}_2)$ p.d. dir.

Gunst ve Mason (1976) nin yaklaşımından faydalanılarak ; Teorem 3.14.1 iterasyon tahmin edicinin OLS, RR ve shrunken tahmin edicisiyle karşılaştırılmasında kullanılabilir. Ancak dikkat edilmesi gereken husus Teorem 3.14.1 , PC tahmin edicisine uygulanamaz. Çünkü bu durumda $C_1C_1' - C_2C_2'$ p.d. olmaz.

$$X = Q\Omega P \quad (3.14.7)$$

şeklindeki tekil değer ayrışımını ele alalım. Burada Q , $n \times n$ tipinde ortonormal bir matris; P $p \times p$ tipinde ortonormal bir matris ve

$$\Omega = \begin{bmatrix} 1 \\ \Lambda^{\frac{1}{2}} \\ 0 \end{bmatrix} \quad (3.14.8)$$

$n \times p$ tipinde bir matris ve $\Lambda = \text{diag}\{\lambda_j\}_{j=1}^p$ dir.

Tekil değer ayrışımından faydalanılarak OLS tahmin edicisi , RR tahmin edicisi, shrunken tahmin edici , iterasyon tahmin edicisi sırasıyla

$$\begin{aligned} \hat{\beta} &= P'\Lambda^{-1}\Omega'QY \\ \hat{\beta}(k) &= P'(\Lambda + kI)^{-1}\Omega'QY \\ \hat{\beta}_s &= SP'\Lambda^{-1}\Omega'QY \\ \hat{\beta}_{m,\alpha} &= P'[\Delta_{m,\alpha} \quad 0]QY \end{aligned} \quad (3.14.9)$$

olarak yazılabilir. (3.14.9) ifadesinde

$$\begin{aligned} \Delta_{m,\alpha} &= \text{diag} \left\{ \frac{1 - (1 - \alpha\lambda_j)^{m+1}}{\sqrt{\lambda_j}} \right\}_{j=1}^p \\ &= \text{diag} \left\{ \frac{1 - \lambda_j(m, \alpha)}{\sqrt{\lambda_j}} \right\}_{j=1}^p \end{aligned} \quad (3.14.10)$$

ve

$$\lambda_j(m, \alpha) = (1 - \alpha\lambda_j)^{m+1}, \quad j = 1, 2, \dots, p \quad (3.14.11)$$

dır (Trenkler (1978a-b)).

3.15. OLS İle İterasyon Tahmin Edicisinin Karşılaştırılması

$C_1 = P'A^{-1}\Omega'Q$ ve $C_2 = P'[\Delta_{m,\alpha} \quad 0]Q$ yazılırsa

$$C_1C_1' - C_2C_2' = P(\Lambda^{-1} - \Delta_{m,\alpha}^2)P \quad (3.15.1)$$

p.d. matrisi elde edilir. Bundan başka

$$\Lambda_{m,\alpha} = \text{diag}\left\{-\lambda_j(m,\alpha)\right\}_{j=1}^p \quad (3.15.2)$$

olmak üzere $B(\hat{\beta}_{m,\alpha}) = P'\Lambda_{m,\alpha}P\beta$ elde edilir. Sonuç olarak

$$B(\hat{\beta}_{m,\alpha})B(\hat{\beta}_{m,\alpha})' = P'\Lambda_{m,\alpha}P\beta\beta'P'\Lambda_{m,\alpha}P \quad (3.15.3)$$

bulunur. Teorem 3.14.1 in uygulanmasıyla aşağıdaki sonuç elde edilir. Eğer

$$\beta'P'\Lambda_{m,\alpha}(\Lambda^{-1} - \Delta_{m,\alpha}^2)^{-1}\Lambda_{m,\alpha}P\beta < \sigma^2 \quad (3.15.4)$$

ise $MSE(\hat{\beta}) - MSE(\hat{\beta}_{m,\alpha})$ p.d. dir. Ya da eşdeğer olarak eğer

$$\beta'P\text{diag}\left\{\frac{\lambda_j\lambda_j(m,\alpha)}{2-\lambda_j(m,\alpha)}\right\}_{j=1}^p P\beta < \sigma^2 \quad (3.15.5)$$

ise $MSE(\hat{\beta}) - MSE(\hat{\beta}_{m,\alpha})$ p.d. dir. $j=1,2,\dots,p$ için

$$\lim_{m \rightarrow \infty} \frac{\lambda_j\lambda_j(m,\alpha)}{2-\lambda_j(m,\alpha)} = 0 \quad (3.15.6)$$

olduğundan ; tüm $m \geq m_0$ için $MSE(\hat{\beta}) - MSE(\hat{\beta}_{m,\alpha})$ pd olacak şekilde bir m_0 tamsayısı vardır. (3.15.5) eşitsizliği $MSE(\hat{\beta}) - MSE(\hat{\beta}_{m,\alpha})$ nın p.d. olması için de gereklidir (Trenkler, 1978b). Ancak m_0 ; bilinmeyen β ve σ^2 değerlerine bağlıdır. Fakat ; k bilinen bir pozitif sabit olmak üzere ; β ve σ^2 ile ilgili $\beta\beta' \leq k\sigma^2$

şeklinde bir ön bilgiye sahip olduğumuzu kabul edelim. O zaman ; tüm $m \geq \bar{m}$ için $MSE(\hat{\beta}) - MSE(\hat{\beta}_{m,\alpha})$ p.d. olacak şekilde kesin bir \bar{m} sayısı belirlenebilir. \bar{m} sayısı

$$\lambda_p(\bar{m} + 1, \alpha) = (1 - \alpha \lambda_p)^{\bar{m}-1} < \frac{1}{k \lambda_1} \quad (3.15.7)$$

eşitsizliğini sağlayan en küçük sayı olarak seçilir. İspatı (Trenkler (1978-b)) de yapılmıştır.

3.16. ORR İle İterasyon Tahmin Edicisinin Karşılaştırılması

$C_1 = P'(\Lambda + kI)^{-1} \Omega' Q$ ve $C_2 = P'[\Delta_{m,\alpha} \quad 0] Q$ alınmasıyla

$$C_1 C_1' - C_2 C_2' = P'(\Lambda(\Lambda + kI)^{-2} - \Delta_{m,\alpha}^2) P \quad (3.16.1)$$

elde edilir ve

$$C_1 C_1' - C_2 C_2' = P' \text{diag} \left\{ \frac{\lambda_j^2 - (\lambda_j + k)^2 (1 - \lambda_j(m, \alpha))^2}{\lambda_j (\lambda_j + k)^2} \right\}_{j=1}^p P \quad (3.16.2)$$

bulunur.

m nin belirli olduğunu düşünelim. Tüm $k \leq k_0$ için $C_1 C_1' - C_2 C_2'$ p.d. olacak şekilde bir k_0 sabitinin olduğu açıktır . Böylelikle Teorem 3.14.1 in uygulanmasıyla , $MSE(\hat{\beta}_k) - MSE(\hat{\beta}_{m,\alpha})$ nın p.d. olması için yeterli koşul yani ; $k \leq k_0$ ve

$$B(\hat{\beta}_{m,\alpha})' (C_1 C_1' - C_2 C_2')^{-1} B(\hat{\beta}_{m,\alpha}) < \sigma^2 \quad (3.16.3)$$

elde edilir.

Bu yeterli koşulun geçerliliği , β ve σ^2 değerlerinin büyüklüğüne bağlıdır. Dolayısıyla m rastgele olduğunda , $\hat{\beta}_{m,\alpha}$ nın uygun ridge tahmin edicilerinden MSE

kriterine göre potansiyel olarak daha iyi olduğu k değerlerinin bulunduğu sonucuna ulaşılabilir.

$$C_2 C_2' - C_1 C_1' = P' \text{diag} \left\{ \frac{(\lambda_j + k)^2 (1 - \lambda_j(m, \alpha))^2 - \lambda_j^2}{\lambda_j (\lambda_j + k)^2} \right\}_{j=1}^p P \quad (3.16.4)$$

ifadesinden; belirli bir k için tüm $m \geq m^*$ için her zaman $C_2 C_2' - C_1 C_1'$ p.d. olacak şekilde bir m^* sabitinin bulunabileceği sonucuna ulaşılır.

Yine Teorem 3.14.1 den $MSE(\hat{\beta}_{m, \alpha}) - MSE(\hat{\beta}_k)$ nin p.d. olması için yeterli koşul

$$m \geq m^* \text{ ve } B(\hat{\beta}_k)' (C_2 C_2' - C_1 C_1')^{-1} B(\hat{\beta}_k) < \sigma^2 \quad (3.16.5)$$

olarak elde edilir.

β ve σ^2 üzerinde kısıtlama olmadığı düşünüldüğünde, bu yeterli koşulun sağlanacağı göz ardı edilemez. Buradan elde edilebilecek sonuca göre; MSE kriterine göre bir tahmin edicinin diğerinden daha iyi olduğu gözlenemez. β ve σ^2 nin belirli değerlerine göre bir tahmin edici diğerinden daha iyi olabilir ancak öteki değerler için kötü olabilir. β ve σ^2 ile ilgili ek bilgi verilmediğinde her iki tahmin edici eş zamanlı olarak kullanılmalıdır.

3.17. Shrunken Tahmin Edici İle İterasyon Tahmin Edicinin Karşılaştırılması

Shrunken ve iterasyon tahmin edicilerine ait olan $C_1 = sP'\Lambda^{-1}\Omega'Q$ ve $C_2 = P'[\Delta_{m, \alpha} \quad 0]Q$ matrislerini ele alalım.

$$C_1 C_1' - C_2 C_2' = P' \text{diag} \left\{ \frac{s^2 - (1 - \lambda_j(m, \alpha))^2}{\lambda_j} \right\}_{j=1}^p P \quad (3.17.1)$$

olduğu bulunur.

m belirli olduğunda ve $s \rightarrow 1$ iken $C_1 C_1' - C_2 C_2'$ nün p.d. olabileceği sonucuna ulaşılabilir. Bir başka deyişle eğer $0 < s < 1$ ise o zaman tüm $m \geq m'$ için $C_2 C_2' - C_1 C_1'$ p.d. olacak şekilde bir m' sabiti bulunabilir. Sonuçta MSE kriterine göre bir tahmin edicinin diğerinden daha iyi olduğu sonucuna ulaşamayacağı görülür.

Gerekli Bir Kriter: $\hat{\beta}_r = A_r^+ X'Y$ şeklindeki PC tahmin edicisi ile $\tilde{\beta}$ homojen lineer tahmin edicileri karşılaştırılırken ; $Cov(\hat{\beta}_r) - Cov(\tilde{\beta})$ nin singüler veya regüler olabileceği ancak p.d. olamayacağı gösterilebilir. Dolayısıyla Teorem 3.14.1 uygulanamaz. Ancak bazen $H_r = \frac{1}{\sigma^2} (Cov(\hat{\beta}_r) - Cov(\tilde{\beta}))$ ifadesi ;

$$H_r = P' \begin{bmatrix} D_1 & 0 \\ 0 & D_2 \end{bmatrix} P \quad (3.17.2)$$

yazımı ya p.d. ya da n.d. olan D_1 ve D_2 köşegen matrisleri içerecek biçimde yazılabilir. Bu nedenle aşağıdaki teorem oluşturulabilir.

Teorem 3.17.1 : $MSE(\tilde{\beta}_1) - MSE(\tilde{\beta}_2)$ p.d. olacak şekilde $\tilde{\beta}_j = C_j Y$ ($j=1,2$) şeklindeki iki homojen lineer tahmin ediciyi ele alalım. Bundan başka D_1 ve D_2 köşegen matrisler olacak şekilde $H = \frac{1}{\sigma^2} (Cov(\tilde{\beta}_1) - Cov(\tilde{\beta}_2))$ nin

$$H = P' \begin{bmatrix} D_1 & 0 \\ 0 & D_2 \end{bmatrix} P \quad (3.17.3)$$

olarak yazıldığını kabul edelim . $B(\tilde{\beta}_1)B(\tilde{\beta}_1)'$ matrisi ; $\gamma_j \gamma_j'$, D_j ($j=1,2$) ile aynı tipte olmak üzere

$$B(\tilde{\beta}_1)B(\tilde{\beta}_1)' = P' \begin{bmatrix} \gamma_1\gamma_1' & \gamma_1\gamma_2' \\ \gamma_2\gamma_1 & \gamma_2\gamma_2 \end{bmatrix} P \quad (3.17.4)$$

olarak yazılsın. O zaman D_1 veya D_2 p.d. ise ; buradan $\gamma_1' D_1^{-1} \gamma_1 < \sigma^2$ veya $\gamma_2' D_2^{-1} \gamma_2 < \sigma^2$ olur.

İspat : D_1 in p.d. olduğunu kabul edelim.

$$MSE(\tilde{\beta}_1) - MSE(\tilde{\beta}_2) = \sigma^2 H + B(\tilde{\beta}_1)B(\tilde{\beta}_1)' - B(\tilde{\beta}_2)B(\tilde{\beta}_2)' \quad (3.17.5)$$

p.d. olduğundan $\sigma^2 D_1 + \gamma_1\gamma_1'$ nün p.d. olması gerekir. Farebrother'in sonucunun uygulanmasıyla $\gamma_1' D_1^{-1} \gamma_1 < \sigma^2$ elde edilir. Aynı düşünce D_2 için de uygulanır.

3.18. PC İle İterasyon Tahmin Edicinin Karşılaştırılması

Marquardt (1970) PC tahmin edicisini

$$\hat{\beta}_r = A_r^+ X'Y \quad (3.18.1)$$

şeklinde vermiştir. Burada A_r^+ matrisi ; A_r

$$A_r = P' \begin{bmatrix} \Lambda_1 & 0 \\ 0 & 0 \end{bmatrix} P \quad (3.18.2)$$

nın Moore-Penrose tersi olacak şekilde seçilmiştir.

(3.18.2) de $\Lambda_1 = \text{diag}\{\lambda_j\}_{j=1}^r$; $X'X$ in r büyük özdeğerini içerir.

$$X'X = P' \Lambda P = P' \begin{bmatrix} \Lambda_1 & 0 \\ 0 & \Lambda_2 \end{bmatrix} P \quad (3.18.3)$$

şeklinde parçalandığında ; Λ_2 , $X'X$ in ; PC tahmin edicisi tarafından silinen p-r küçük özdeğerini içerir.

$X = Q'\Omega P$ eşitliğinden

$$\hat{\beta}_r = P' \begin{bmatrix} \Lambda_1^{-\frac{1}{2}} & 0 \\ 0 & 0 \end{bmatrix} QY \quad (3.18.4)$$

elde edilir.

$C_1 = X_{m,\alpha}$ ve $C_2 = A_r^+ X'$ nin şeklinde tanımlanmasıyla

$$Cov(\hat{\beta}_{m,\alpha}) - Cov(\hat{\beta}_r) = \sigma^2 P' \left[\Delta_{m,\alpha}^2 - \begin{bmatrix} \Lambda_1^{-1} & 0 \\ 0 & 0 \end{bmatrix} \right] P \quad (3.18.5)$$

in ne p.d. ne de n.d. olduğu sonucuna varılabilir.

Ancak

$$\left\{ \begin{array}{l} d_j = \frac{(1 - \lambda_j(m, \alpha))^2 - 1}{\lambda_j} ; j = 1, 2, \dots, r \text{ iken} \\ \frac{(1 - \lambda_j(m, \alpha))^2}{\lambda_j} ; j = r + 1, \dots, p \text{ iken} \end{array} \right\} \quad (3.18.6)$$

olmak üzere

$$\Delta_{m,\alpha}^2 - \begin{bmatrix} \Lambda_1^{-1} & 0 \\ 0 & 0 \end{bmatrix} = \text{diag}\{d_j\}_{j=1}^p = \begin{bmatrix} D_1 & 0 \\ 0 & D_2 \end{bmatrix} \quad (3.18.7)$$

olduğundan ; D_1 ve D_2 köşegen matrislerinin n.d. ve p.d. olduğu görülür.

$B(\hat{\beta}_{m,\alpha})B(\hat{\beta}_{m,\alpha})'$ nin ; $\gamma = \Lambda_{m,\alpha} P\beta$ olmak üzere

$$B(\hat{\beta}_{m,\alpha})B(\hat{\beta}_{m,\alpha})' = P'\gamma\gamma'P \quad (3.18.8)$$

şeklinde yazılabilir. Veya eşdeğer olarak , $\gamma_j \gamma_j'$, D_j ($j=1,2$) ile aynı tipte ve

$\gamma = [\gamma_1 \quad \gamma_2]$ olmak üzere

şeklindedir ve bu matris $k \geq 0$ için bir p.s.d. matristir. Dolayısıyla örneklem varyansı kriterine göre, RRR tahmin edicisi her zaman ORR tahmin edicisinden daha iyidir. Tahmin edicilerin yanlılıkları ve örneklem varyansları ile ilgili sonuçlar birleştirildiğinde; kısıtlamaların doğru olduğu durumda RRR tahmin edicisinin her zaman ORR tahmin edicisinden daha iyi olduğu söylenebilir.

II. Durum: (Kısıtlamalar Doğru Olmadığında Yani $\delta \neq 0$ ise)

RRR tahmin edicisinin performansı δ ya bağlıdır. Yani kısıtlamaların hangi oranda sağlanmadığına bağlıdır. Bir başka deyişle bu tahmin edicinin mse kriterine göre üstünlüğü, doğru olmadığı kabul edilen kısıtlamalara göre olan yanlılığın miktarına bağlıdır.

Bu durum için, RRR tahmin edicisi mse kriterine göre RLS ve ORR tahmin edicisinden daha iyi olacak şekildeki δ ya bağlı olan k nın mümkün olan değerlerinin seçimini gerektiren koşulları inceleyelim.

RRR Ve RLS Tahmin Edicilerinin Karşılaştırılması

Kısıtlamalar doğru olmadığında RRR ve RLS tahmin edicileri yanlıdır.

$WS^{-1} = S(k)^{-1}$ olduğundan RRR tahmin edicisinin mse si

$$MSE(\beta^*(k)) = \sigma^2 \text{tr}(WAW') + [S(k)^{-1} \delta^* - kS(k)^{-1} \beta]^T [S(k)^{-1} \delta^* - kS(k)^{-1} \beta] \quad (3.11.5)$$

olarak bulunur. Bazı cebirsel işlemlerden sonra bu ifade,

$$MSE(\beta^*(k)) = \sum_{i=1}^p \frac{1}{(\lambda_i + k)^2} \left[\sigma^2 \lambda_i^2 b_{ii} + k^2 \alpha_i^2 + \delta_i^{*2} - 2k \alpha_i \delta_i^* \right] \quad (3.11.6)$$

olarak yazılabilir.

(3.11.6) daki eşitliğin k ya göre türevi alınırsa

$$\frac{dMSE(\beta^*(k))}{dk} = 2 \sum_{i=1}^p \frac{1}{(\lambda_i + k)^3} \left[k(\lambda_i \alpha_i^2 + \alpha_i \delta_i^*) - (\sigma^2 \lambda_i b_{ii} + \delta_i^{*2} + \alpha_i \lambda_i \delta_i^*) \right] \quad (3.11.7)$$

elde edilir. Böylece $MSE(\beta^*(k))$ nın birinci türevinin negatif olması için yeterli koşul bulunabilir.

Teorem 3.11.1: $k^* = \frac{\min_i (\sigma^2 \lambda_i^2 b_{ii} + \delta_i^* + \alpha_i \lambda_i \delta_i^*)}{\max_i (\lambda_i \alpha_i^2 + \alpha_i \delta_i^*)}$ olarak tanımlansın ve k

pozitif olsun. O zaman RRR tahmin edicisi , $0 < k < k^* < \infty$ için, mse kriterine göre RLS tahmin edicisinden daha iyidir.

RRR Ve ORR Tahmin Edicilerinin Karşılaştırılması

β nın ORR tahmin edicisi $\hat{\beta}(k)$ nın mse si

$$MSE(\hat{\beta}(k)) = \sum_{i=1}^p \frac{1}{(\lambda_i + k)^2} [\sigma^2 \lambda_i + k^2 \alpha_i^2] \quad (3.11.8)$$

dir. (3.11.6) ve (3.11.8) den

$$MSE(\hat{\beta}(k)) - MSE(\beta^*(k)) = \sum_{i=1}^p \frac{1}{(\lambda_i + k)^2} \left[2k\alpha_i \delta_i^* - (\delta_i^{*2} - \sigma^2 \lambda_i + \sigma^2 \lambda_i^2 b_{ii}) \right] \quad (3.11.9)$$

bulunur. $\tilde{k} = \frac{\max_i (\delta_i^{*2} - \sigma^2 \lambda_i + \sigma^2 \lambda_i^2 b_{ii})}{\min_i (2\alpha_i \delta_i^*)}$ olmak üzere. $k > \tilde{k} > 0$

olduğunda (3.11.9) daki eşitliğin sağ tarafı pozitifdir. Dolayısıyla aşağıdaki teorem elde edilir.

Teorem 3.11.2: $\tilde{k} = \frac{\max_i (\delta_i^{*2} - \sigma^2 \lambda_i + \sigma^2 \lambda_i^2 b_{ii})}{\min_i (2\alpha_i \delta_i^*)}$ olmak üzere RRR tahmin

edicisinin ORR tahmin edicisinden daha küçük mse ye sahip olması için yeterli koşul $0 < \tilde{k} < k < \infty$ olmasıdır.

3.12. Modified Ridge Tahmin Edicisi İle Kısıtlanmış Ridge Regresyon Tahmin Edicisinin Mse Kriterine Göre Karşılaştırılması

Swindel (1976) ön bilgiye dayanan düzenlenmiş bir ridge tahmin edicisi ortaya koymuştur. Ayrıca OLS tahmin edicisinin mse matrisinin, ortaya koyduğu tahmin edicinin mse matrisinden daha büyük olması için gerekli ve yeterli koşulu oluşturmuştur. Bununla beraber Swindel, oluşturduğu tahmin edicinin mse özellikleri ile ORR tahmin edicisinin mse özelliklerini karşılaştırmamıştır. Pliskin (1987), ORR tahmin edicisi ile modified ridge tahmin edicisinin mse matrislerini karşılaştırmıştır. Sarkar (1992), ridge regresyon düşüncesini RLS tahmin edicisine uygulayarak RRR tahmin edicisini oluşturmuştur. Ayrıca, bu tahmin edicinin; kısıtlamalar doğru olduğunda, mse kriterine göre RLS ve ORR tahmin edicilerinden daha iyi olduğunu göstermiştir. Bu tahmin edicinin, kısıtlamalar doğru olmadığında, diğer iki tahmin ediciden daha iyi olduğu koşullar da oluşturulmuştur.

$$Y = X\beta + \varepsilon \quad \varepsilon \sim N(0, \sigma^2 I) \quad (3.12.1)$$

modelinde OLS tahmin edicisi

$$\hat{\beta} = (X'X)^{-1} X'Y \quad (3.12.2)$$

şeklindedir.

$$R\beta = r \quad (3.12.3)$$

lineer kısıtlamalar kümesini ele alalım.

RLS tahmin edicisi $S = X'X$ olmak üzere

$$\beta^* = \hat{\beta} + S^{-1} R' (RS^{-1} R')^{-1} (r - R\hat{\beta}) \quad (3.12.4)$$

şeklindedir.

ORR tahmin edicilerinin ailesi $W = (I_p + kS^{-1})^{-1}$ olmak üzere Hoerl ve Kennard (1970, a,b) tarafından

$$\hat{\beta}(k) = (X'X + kI_p)^{-1} X'Y \quad k \geq 0 \quad (3.12.5)$$

şeklinde tanımlanmıştır.

Sarkar (1992) tarafından tanımlanan RRR tahmin edicisi ; β^* , β nın RLS tahmin edicisi olmak üzere

$$\beta^*(k) = W\beta^* , \quad k \geq 0 \quad (3.12.6)$$

şeklindedir. Swindel (1976); b^* , β nın belirli bir ön tahmin vektörü olmak üzere

$$b(k, b^*) = (X'X + kI_p)^{-1} (X'Y + kb^*) , \quad k \geq 0 \quad (3.12.7)$$

tahmin edicisini tanımlamıştır. Uygulamada b^* , β üzerindeki ön bilgiyi veya hipotezleri yansıtmak şekilde seçilebilir. Genel olarak ; β üzerindeki ön bilgi β_0 ortalama, vektörüne ve V_0 kovaryans matrisine sahip bir p-değişkenli normal dağılımla tanımlanabiliyorsa , β nın Bayes tahmin edicisi

$$\hat{\beta}_B = (X'X + \sigma^2 V_0^{-1})^{-1} (X'Y + \sigma^2 V_0^{-1} \beta_0)$$

şeklinde bulunur.

$\beta^*(k)$ ve $b(k, b^*)$ yanlı tahmin edicilerdir. Bir $\tilde{\beta}$ tahmin edicisinin yanlılığı ve varyansı mse matrisi ile

$$MSE(\tilde{\beta}) = Var(\tilde{\beta}) + Bias(\tilde{\beta})Bias(\tilde{\beta})' \quad (3.12.8)$$

ve

$$Bias(\tilde{\beta}) = E(\tilde{\beta}) - \beta \quad (3.12.9)$$

ve

$$Var(\tilde{\beta}) = E \left[(\tilde{\beta} - E(\tilde{\beta})) (\tilde{\beta} - E(\tilde{\beta}))' \right] \quad (3.12.10)$$

şeklinde bulunur.

β nin bir değeri için ; $MSE(\tilde{\beta}_1) - MSE(\tilde{\beta}_2)$ bir n.n.d matris ise $\tilde{\beta}_2$ tahmin edicisi $\tilde{\beta}_1$ tahmin edicisine tercih edilir. Ancak tersi doğru değildir (Theobald,1974).

$b(k, b^*)$ Ve $\beta^*(k)$ Tahmin Edicilerinin Karşılaştırılması

β^* in varyans kovaryans matrisi

$$Var(\beta^*) = \sigma^2 \left(S^{-1} - S^{-1} R' (RS^{-1} R')^{-1} RS^{-1} \right) \quad (3.12.11)$$

şeklindedir. $\beta^*(k) = W\beta^*$ olduğundan $A = S^{-1} - S^{-1} R' (RS^{-1} R')^{-1} RS^{-1}$ ve $W = (I_p + kS^{-1})^{-1}$ olmak üzere

$$Var(\beta^*(k)) = \sigma^2 W A W' \quad (3.12.12)$$

dir. $\beta^*(k)$ nin beklenen değeri ve yanlılığı $\delta = r - R\beta$ ve $\delta^* = R' (RS^{-1} R')^{-1} \delta$ olmak üzere

$$\begin{aligned} E[\beta^*(k)] &= W\beta + WS^{-1} R' (RS^{-1} R')^{-1} \delta \\ Bias(\beta^*(k)) &= WS^{-1} \delta^* - kS(k)^{-1} \beta \end{aligned} \quad (3.12.13)$$

olarak bulunur.

(3.12.2) ve (3.12.3) ün birleştirilmesiyle $S(k) = S + kI$ olmak üzere

$$MSE(\beta^*(k)) = \sigma^2 W A W' + [WS^{-1} \delta^* - kS(k)^{-1} \beta] [WS^{-1} \delta^* - kS(k)^{-1} \beta]' \quad (3.12.14)$$

bulunur. Eğer $\delta = 0$ ise yani kısıtlamalar doğru ise o zaman $MSE(\beta^*(k))$

$$MSE(\beta^*(k)) = \sigma^2 W A W' + k^2 S(k)^{-1} \beta \beta' S(k)^{-1} \quad (3.12.15)$$

ifadesine indirgenir.

$b(k, b^*)$ nin varyans-kovaryans matrisi

$$\text{Var}(b(k, b^*)) = \sigma^2 S(k)^{-1} SS(k)^{-1} \quad (3.12.16)$$

olarak verilir. Bundan başka $b(k, b^*)$ nin yanlılığı

$$\text{Bias}(b(k, b^*)) = -kS(k)^{-1}(\beta - b^*) \quad (3.12.17)$$

şeklindedir.

(3.12.16) ve (3.12.17) nin birleştirilmesiyle

$$\text{MSE}(b(k, b^*)) = \sigma^2 S(k)^{-1} SS(k)^{-1} + k^2 S(k)^{-1} (\beta - b^*)(\beta - b^*)' S(k)^{-1} \quad (3.12.18)$$

bulunur.

$\beta^*(k)$ ve $b(k, b^*)$ nin örneklem varyans-kovaryans matrislerine göre performansı

$$\text{Var}(b(k, b^*)) - \text{Var}(\beta^*(k)) = \sigma^2 [S(k)^{-1} SS(k)^{-1} - WAW'] \quad (3.12.19)$$

olarak bulunur.

$W = SS(k)^{-1}$ olduğundan (3.12.19) kullanılarak

$$\text{Var}(b(k, b^*)) - \text{Var}(\beta^*(k)) = \sigma^2 S(k)^{-1} [R'(RS^{-1}R')^{-1}R] S(k)^{-1} \quad (3.12.20)$$

olduğu kolayca görülebilir.

Pozitif yarı tanımlı matrislerinin temel özellikleri kullanılarak $[\text{Var}(b(k, b^*)) - \text{Var}(\beta^*(k))]$ nin bir p.s.d. matris olduğu görülebilir ve RRR tahmin edicisinin modified ridge tahmin edicisinden daha küçük örneklem varyans - kovaryans matrisine sahip olduğu sonucuna varılabilir. Şimdi k nin nonstokhastik olduğu ve lineer kısıtlamaların doğru olduğu varsayımı altında ; aynı k değeri için hesaplanan $\beta^*(k)$ ve $b(k, b^*)$ tahmin edicilerinin mse matrislerinin karşılaştırılmasını inceleyelim.

Lemma 3.12.1 : (Gruber, 1990) A , $n \times n$ tipinde simetrik bir matris ve a , $n \times 1$ tipinde bir vektör ve d pozitif bir skaler olsun . O zaman $dA - aa'$ nin p.s.d. olması

için gerek ve yeter koşul A nın n.n.d. olması, a nın A nın kolon uzayında olması ve A^+ , A nın Moore-Penrose tersi olmak üzere $a'A^+a \leq d$ olmasıdır.

(3.12.15) ve (3.12.18) kullanılmasıyla

$$\begin{aligned} MSE(b(k, b^*)) - MSE(\beta^*(k)) &= S(k)^{-1} \left[\sigma^2 (S - SAS) + k^2 \{ (\beta - b^*)(\beta - b^*)' - \beta\beta' \} \right] S(k)^{-1} \\ &= S(k)^{-1} \left[\sigma^2 R'(RS^{-1}R')^{-1}R + k^2 \{ (\beta - b^*)(\beta - b^*)' - \beta\beta' \} \right] S(k)^{-1} \quad (3.12.21) \end{aligned}$$

olduğu bulunur.

Böylelikle, lineer kısıtlamalar doğru olduğunda aşağıdaki teoremler oluşturulur.

Teorem 3.12.1 : $D = (\beta - b^*)(\beta - b^*)' - \beta\beta'$ olsun. Her iki tahmin edici aynı k değeri için hesaplandığında, eğer D p.s.d. matris ise o zaman $MSE(b(k, b^*)) - MSE(\beta^*(k))$ p.s.d. dir.

İspat : Eğer $b(k, b^*)$ ile $\beta^*(k)$ nın hesaplanmasında aynı k değeri kullanılıyorsa (3.12.21) elde edilir. $R'(RS^{-1}R')^{-1}R$ p.s.d. olduğundan; eğer D p.s.d. matris ise o zaman $MSE(b(k, b^*)) - MSE(\beta^*(k))$ p.s.d. dir.

Teorem 3.12.1 den sezgisel olarak; $\beta^*(k)$ ve $b(k, b^*)$ aynı k değeri için hesaplandığında; $\beta^*(k)$ nın varyans kovaryans matrisinin, $b(k, b^*)$ nın varyans kovaryans matrisinden daha küçük olduğu söylenebilir.

$k^2 S(k)^{-1} D S(k)^{-1}$ matrisi; iki tahmin edicinin yanlılık vektörleri arasındaki farkı yansıtır. Dolayısıyla D nin bir p.s.d. matris olması yeterlidir.

RRR tahmin edicisinin performansı D nin p.s.d. olması koşulu ile karakterize edilmiştir. Bu küme, D , hiçbir zaman boş değildir. Çünkü $\beta = c(\beta - b^*)$ veya

$$b^* = \frac{c-1}{c} \beta \text{ alınırsa yani } |c| < 1 \text{ için } D = \frac{1-c^2}{c^2} \beta\beta' \text{ p.s.d. dir.}$$

$\beta = c(\beta - b^*)$ veya $b^* = \frac{c-1}{c}\beta$ ($|c| < 1$) koşulu ; $\beta^*(k)$ nin yanlılığının $b(k, b^*)$ nin yanlılığından küçük olmasını sağlar.

$\beta = c(\beta - b^*)$ veya eşdeğer olarak $b^* = \frac{c-1}{c}\beta$ koşulu altında , (3.12.7) den aşağıdaki gibi istatistiksel olarak ilginç tahmin ediciler elde edilir.

i) $c \rightarrow 1$ iken $b(k, b^*)$, $\hat{\beta}(k)$ ya yaklaşır. (yani $b^*=0$)

ii) $b^* = \frac{c-1}{c}\beta$ da , β yerine yansız tahmin edici $\hat{\beta}$ nin yazılması ile $c \rightarrow \infty$ ile $\hat{b}^* = \hat{\beta}$ elde edilir. Dolayısıyla (3.12.7) den $b(k, \hat{b}^*) = \hat{\beta}$ elde edilir. Öte yandan $k \rightarrow 0$ iken $b(k, b^*)$ nin $\hat{\beta}$ ya yaklaştığı görülür.

iii) $b^* = \frac{c-1}{c}\beta$ da β yerine RLS tahmin edicisi β^* in yazılması ve $c \rightarrow \infty$ ile $\hat{b}^* = \beta^*$ elde edilir. Dolayısıyla (3.12.7) den

$$b(k, \beta^*) = (X'X + kI_p)^{-1}(X'Y + k\beta^*) , k \geq 0 \quad (3.12.22)$$

elde edilir.

Farebrother (1984) tarafından tanımlanan tahmin edici sınıfını göz önüne alalım.

$$\hat{\beta}(k) = (X'X + kR'R)^{-1}(X'Y + kR'R\beta^*) , k \geq 0 \quad (3.12.23)$$

$R = I_p$ özel durumunda bu , (3.12.22) ye dönüşür.

iv) $b^* = \frac{c-1}{c}\beta$ daki β nin OLS tahmin edicisi olan $\hat{\beta}$ alınrsa $\hat{b}^* = \frac{c-1}{c}\hat{\beta}$

elde edilir. O zaman (3.12.7) den $k \rightarrow \infty$ iken $b(k, \hat{b}^*) = \left(1 - \frac{1}{c}\right)\hat{\beta}$ olarak bulunur.

Açıkça $c > 1$ için shrunken tahmin edici elde edilir.(Mayer ve Willke, 1973).

Teorem 3.12.2 : Eğer $R^+R\beta = \beta$ ve $k^2 < \frac{\sigma^2}{\beta'S^{-1}\beta}$ ise $MSE(b(k, b^*)) - MSE(\beta^*(k)) \geq 0$ dir.

İspat: $k^2(\beta - b^*)(\beta - b^*)'$ n.n.d. olduğundan; eğer $\sigma^2 R'(RS^{-1}R')^{-1}R - k^2\beta\beta' \geq 0$ ise o zaman $MSE(b(k, b^*)) - MSE(\beta^*(k))$ p.s.d. dir. Lemma 3.12.1 den $\sigma^2 R'(RS^{-1}R')^{-1}R - k^2\beta\beta' \geq 0$ olması için gerek ve yeter koşul $R^+R\beta = \beta$ ve $k^2 < \frac{\sigma^2}{\beta'S^{-1}\beta}$ olmasıdır. Bu ise ispatı tamamlar.

Şimdi , k nın nonstokhostik olduğu ve parametrik kısıtlamaların doğru olmadığı varsayımı altında ; aynı k değeri için hesaplanan $\beta^*(k)$ ve $b(k, b^*)$ tahmin edicilerinin mse matrislerinin karşılaştırılmalarını inceleyelim.

Teorem 3.12.3 : Eğer $R^+R\beta = \beta$ ve $(k\beta - \delta^*)'S^{-1}(k\beta - \delta^*) < \sigma^2$ ise $MSE(b(k, b^*)) - MSE(\beta^*(k)) \geq 0$ dir.

İspat : Kısıtlamalar doğru olmadığında

$$\begin{aligned} MSE(b(k, b^*)) - MSE(\beta^*(k)) &= S(k)^{-1} \left[\sigma^2 R'(RS^{-1}R')^{-1}R + k^2(\beta - b^*)(\beta - b^*)' \right] S(k)^{-1} \\ &\quad - \left[WS^{-1}\delta^* - kS(k)^{-1}\beta \right] \left[WS^{-1}\delta^* - kS(k)^{-1}\beta \right]' \\ &= S(k)^{-1} \left[\sigma^2 R'(RS^{-1}R')^{-1}R + k^2(\beta - b^*)(\beta - b^*)' - (k\beta - \delta^*)(k\beta - \delta^*)' \right] S(k)^{-1} \end{aligned}$$

olarak bulunur.

$k^2(\beta - b^*)(\beta - b^*)'$ n.n.d. olduğundan Lemma 3.12.1 in kullanılmasıyla; eğer $R^+R\beta = \beta$ ve $(k\beta - \delta^*)'S^{-1}(k\beta - \delta^*) < \sigma^2$ ise $MSE(b(k, b^*)) - MSE(\beta^*(k)) \geq 0$ olduğu bulunur. Böylece ispat tamamlanır.

3.13. Gauss-Markoff Modelindeki Minimum Yanlılığa Sahip En İyi Lineer Tahmin Ediciler

$$Y = X\beta + \varepsilon \quad (3.13.1)$$

genel Gauss-Markov (GGM) lineer modelini ele alalım. Bu modelde $E(\varepsilon) = 0$ ve $Var(\varepsilon) = \sigma^2 G$ dir. X matrisi tam sütun ranklı olmayabilir veya G singüler olabilir. Bu sebepten bu model $(Y, X\beta, \sigma^2 G)$ GGM modeli olarak adlandırılır. Böyle bir genel modelin uygulanabilirliği Rao (1971) tarafından ortaya konulmuştur ve LS nin birleştirilmiş teorisi oluşturulmuştur.

Rao'nun bu modeli , lineer tahmin için kriter oluşturmuş ve β nin lineer parametrik fonksiyonları için lineer tahmin edicilerin bir genel formunu vermiştir. Bu kriterlerden bazıları minimum yanlı lineer tahmin edicilerin (MB-LE) bulunması için veya yeni minimum varyans, minimum yanlı lineer tahmin edici (MV-MB-LE) elde etmek için oluşturulmuştur. Rao (1973 b) ayrıca karesel hata kaybı altında bir lineer tahmin edicinin bayes riskinin minimize edilmesine eşdeğer olan, bir ön bilgi altında mse nin minimize edilmesini ele almıştır. Burada lineer tahminin minimum mse-minimum yanlılık, lineer tahmin (MMS-MB-LE) olarak adlandırılan bir başka kriteri üzerinde durulacaktır.

Lineer Tahmin İçin Bir Kriter

$p'\beta$ lineer fonksiyonunun , $L'Y$ lineer tahmin edicisi tarafından tahmin edildiğini kabul edelim. $L'Y$ nin mse si

$$E[(L'Y - p'\beta)^2] = \sigma^2 L'GL + (X'L - p)' \beta \beta' (X'L - p) \quad (3.13.2)$$

şeklindedir.

Rao (1973) ; W bilinen bir n.n.d. matris olmak üzere , kriterin

$$\begin{aligned}
S &= L'GL + (X'L - p)'W(X'L - p) \\
&= L'(G + XWX')L - 2p'WX'L + p'Wp
\end{aligned} \tag{3.13.3}$$

şeklinde minimize edilmesini ele almıştır. Sonuçta elde edilen tahmin edici en iyi lineer tahmin edici (BLE) olarak adlandırılır.

Chipman (1964) , B n.n.d. olmak üzere minimum yanlı lineer tahmin ediciyi (MB-LE) elde etmek için L ye göre minimize edilmesini ele almıştır. $Var(L'Y) = L'GL$ yi minimize eden bu sınıftaki L yi kısıtlamıştır. BLE ve MB-LE için çözümler Rao (1973-b) tarafından verilmiştir.

MB-LE sınıfında S nin minimize edilmesi ele alınacak ve bu kriter MMS-MB-LE olarak adlandırılacaktır.

Bu noktada dikkat edilmesi gerek husus $W=B$ olduğunda MMS-MBLE ve MV-MB-LE kriterlerinin eşdeğer olduğudur.

Teorem 3.13.1 : $(X'L - p)'B(X'L - p)$ yi minimize eden L nin sınıfında $S = L'GL + (X'L - p)'W(X'L - p)$ nin minimumu

$$L = E_1XWp + E_2XBp \tag{3.13.4}$$

şeklinde elde edilir. Burada B ve W verilen iki n.n.d. matris ve E_1, E_2

$$\begin{bmatrix} G + XWX' & XBX' \\ XBX' & 0 \end{bmatrix}^{-1} = \begin{bmatrix} E_1 & E_2 \\ E_3 & E_4 \end{bmatrix} \tag{3.13.5}$$

eşitliğinden elde edilir ve $(...)^{-}$ genelleştirilmiş inversi gösterir.

İspat : λ , Lanrange çarpanı vektörü olmak üzere, L nin

$$(G + XWX')L + XBX'\lambda = XWp \tag{3.13.6}$$

$$XBX'L = XBp \tag{3.13.7}$$

eşitliklerini sağladığının gösterilmesi kolaydır.

(3.13.6) , (3.13.7) nin çözülebileceği ve (3.13.6) , (3.13.7) nin (3.13.4) de olduğu gibi her çözümlerinin S nin minimumunu verdiği gözlenebilir.

Hatırlatma 3.13.1 : Parçalanmış bir matrisin genelleştirilmiş tersinin gösteriminin kullanılmasıyla E_1 ve E_2 nin bir seçimi ; $N = M + A^2$ ve $A = XB X'$ olmak üzere

$$E_1 = N^{-1} \left[I - A(AN^{-1}A)^{-1} AN^{-1} \right] \quad E_2 = N^{-1} A(AN^{-1}A)^{-1} \quad (3.13.8)$$

şeklinde verilir.

$p'\beta$ nin, $p'HY$ formundaki MMS-MB-LE olan lineer tahmin edicisi bulunmak istenebilir. Bu problemin genel bir çözümü aşağıdaki teorem ile verilebilir.

Teorem 3.13.2 : H için bir genel çözüm

$$T = trHGH' + tr(HX - I)W(HX - I)' \quad (3.13.9)$$

ifadesi

$$\|HX - I\|^2 = tr(HX - I)B(HX - I)' \quad (3.13.10)$$

ifadesini minimize eden H matrisleri , sınıfındaki minimum olacak şekilde

$$H = WX'E_{1g} + BX'E_{2g} \quad (3.13.11)$$

şeklinde verilir. Burada

$$E_{1g} = E_1 + C_1(I - ME_1 - AE_2) + C_2AE_1 + (I - E_1M - E_2A)D_1 + E_1AD_2 \quad (3.13.12)$$

$$E_{2g} = E_2 + C_1(-ME_2 - AE_4) + C_2(I - AE_2) + (I - E_1M - E_2A)D_3 + E_1AD_4 \quad (3.13.13)$$

şeklinde dir . $C_1, C_2, D_1, D_2, D_3, D_4$ rastgele olarak seçilmiştir.

İspat : Λ , Lagrange çarpanlarının matrisi olmak üzere

$$H(G + XWX') + \Lambda XB X' = WX' \quad (3.13.14)$$

ve

$$HXBX' = BX' \quad (3.13.15)$$

eşitliklerinin minimize edilmesi probleminin bir genel çözümünün bulunmasıyla teorem ispatlanabilir. H için çözüm

$$[H : \Lambda] = [WX' : BX'] \begin{bmatrix} G + XWX' & XBX' \\ XBX' & 0 \end{bmatrix} \quad (3.13.16)$$

nın çözülmesiyle elde edilir.

Hatırlatma 3.13.2 : Lineer Bayes tahmin edicisi ; $E(\beta\beta') = \sigma^2 W$ olmak üzere yukarıdaki teoreme benzer çözüme sahiptir.

3.14. Lineer Tahmin Edicilerin Karşılaştırılması

$\tilde{\beta} = CY$ şeklindeki herhangi bir homojen lineer tahmin edicinin

$$MSE(\tilde{\beta}) = (CX - I)\beta\beta'(CX - I)' + \sigma^2 CC' \quad (3.14.1)$$

şeklindeki ikinci dereceden moment matrisini ele alalım. (3.14.1) ifadesi

$$MSE(\tilde{\beta}) = B(\tilde{\beta})B(\tilde{\beta})' + Cov(\tilde{\beta}) \quad (3.14.2)$$

şeklinde yazılabilir. Burada $\tilde{\beta}$ nin yanlılığı

$$B(\tilde{\beta}) = E(\tilde{\beta}) - \beta = (CX - I)\beta \quad (3.14.3)$$

ve varyans-kovaryans matrisi

$$Cov(\tilde{\beta}) = E\left[(\tilde{\beta} - E(\tilde{\beta}))(\tilde{\beta} - E(\tilde{\beta}))'\right] = \sigma^2 CC' \quad (3.14.4)$$

şeklinde dir.

Bu durumda $j=1,2$ için $\tilde{\beta}_j = C_j Y$ şeklindeki iki tahmin edici için

$$\begin{aligned}
MSE(\tilde{\beta}_1) - MSE(\tilde{\beta}_2) &= Cov(\tilde{\beta}_1) - Cov(\tilde{\beta}_2) + B(\tilde{\beta}_1)B(\tilde{\beta}_1)' - B(\tilde{\beta}_2)B(\tilde{\beta}_2)' \\
&= \sigma^2(C_1C_1' - C_2C_2') + B(\tilde{\beta}_1)B(\tilde{\beta}_1)' - B(\tilde{\beta}_2)B(\tilde{\beta}_2)' \quad (3.14.5)
\end{aligned}$$

elde edilir. Şimdi $MSE(\tilde{\beta}_1) - MSE(\tilde{\beta}_2)$ nin p.d. olduğu koşulları inceleyelim. Bu koşullar $C_1C_1' - C_2C_2'$ matrisinin pozitif tanımlılığına bağlıdır.

Yeterli Bir Koşul

$C_1C_1' - C_2C_2'$ nün p.d. olduğunu kabul edelim. Eğer $\sigma^2(C_1C_1' - C_2C_2') - B(\tilde{\beta}_2)B(\tilde{\beta}_2)'$ p.d. ise $MSE(\tilde{\beta}_1) - MSE(\tilde{\beta}_2)$ p.d. dir.

Farebrother (1976) nin bir sonucu uygulanabilir. A bir p.d. matris olsun c sıfırdan farklı bir vektör ve θ pozitif bir skaler olsun. O zaman $\theta A - cc'$ nün p.d. olması için gerek ve yeter koşul $c'A^{-1}c$ nin θ dan küçük olmasıdır. Buradan aşağıdaki teorem elde edilir.

Teorem 3.14.1 : $j=1,2$ için $\tilde{\beta}_j = C_j Y$; $C_1C_1' - C_2C_2'$ p.d. olacak şekilde iki homojen lineer tahmin edici olsun. Bundan başka

$$\beta'(C_2X - I)'(C_1C_1' - C_2C_2')^{-1}(C_2X - I)\beta < \sigma^2 \quad (3.14.6)$$

eşitsizliği sağlansın. O zaman $MSE(\tilde{\beta}_1) - MSE(\tilde{\beta}_2)$ p.d. dir.

Gunst ve Mason (1976) nin yaklaşımından faydalanılarak ; Teorem 3.14.1 iterasyon tahmin edicinin OLS, RR ve shrunken tahmin edicisiyle karşılaştırılmasında kullanılabilir. Ancak dikkat edilmesi gereken husus Teorem 3.14.1 , PC tahmin edicisine uygulanamaz. Çünkü bu durumda $C_1C_1' - C_2C_2'$ p.d. olmaz.

$$X = Q\Omega P \quad (3.14.7)$$

şeklindeki tekil değer ayrışımını ele alalım. Burada Q , $n \times n$ tipinde ortonormal bir matris; P $p \times p$ tipinde ortonormal bir matris ve

$$\Omega = \begin{bmatrix} 1 \\ \Lambda^2 \\ 0 \end{bmatrix} \quad (3.14.8)$$

$n \times p$ tipinde bir matris ve $\Lambda = \text{diag}\{\lambda_j\}_{j=1}^p$ dir.

Tekil değer ayrışımından faydalanılarak OLS tahmin edicisi , RR tahmin edicisi, shrunken tahmin edici , iterasyon tahmin edicisi sırasıyla

$$\begin{aligned} \hat{\beta} &= P'\Lambda^{-1}\Omega'QY \\ \hat{\beta}(k) &= P'(\Lambda + kI)^{-1}\Omega'QY \\ \hat{\beta}_s &= SP'\Lambda^{-1}\Omega'QY \\ \hat{\beta}_{m,\alpha} &= P'[\Delta_{m,\alpha} \quad 0]QY \end{aligned} \quad (3.14.9)$$

olarak yazılabilir. (3.14.9) ifadesinde

$$\begin{aligned} \Delta_{m,\alpha} &= \text{diag} \left\{ \frac{1 - (1 - \alpha\lambda_j)^{m+1}}{\sqrt{\lambda_j}} \right\}_{j=1}^p \\ &= \text{diag} \left\{ \frac{1 - \lambda_j(m,\alpha)}{\sqrt{\lambda_j}} \right\}_{j=1}^p \end{aligned} \quad (3.14.10)$$

ve

$$\lambda_j(m,\alpha) = (1 - \alpha\lambda_j)^{m+1}, \quad j = 1, 2, \dots, p \quad (3.14.11)$$

dir (Trenkler (1978a-b)).

3.15. OLS İle İterasyon Tahmin Edicisinin Karşılaştırılması

$C_1 = P'A^{-1}\Omega'Q$ ve $C_2 = P'[\Lambda_{m,\alpha} \quad 0]Q$ yazılırsa

$$C_1C_1' - C_2C_2' = P(\Lambda^{-1} - \Delta_{m,\alpha}^2)P \quad (3.15.1)$$

p.d. matrisi elde edilir. Bundan başka

$$\Lambda_{m,\alpha} = \text{diag}\{-\lambda_j(m,\alpha)\}_{j=1}^p \quad (3.15.2)$$

olmak üzere $B(\hat{\beta}_{m,\alpha}) = P'\Lambda_{m,\alpha}P\beta$ elde edilir. Sonuç olarak

$$B(\hat{\beta}_{m,\alpha})B(\hat{\beta}_{m,\alpha})' = P'\Lambda_{m,\alpha}P\beta\beta'P'\Lambda_{m,\alpha}P \quad (3.15.3)$$

bulunur. Teorem 3.14.1 in uygulanmasıyla aşağıdaki sonuç elde edilir. Eğer

$$\beta'P'\Lambda_{m,\alpha}(\Lambda^{-1} - \Delta_{m,\alpha}^2)^{-1}\Lambda_{m,\alpha}P\beta < \sigma^2 \quad (3.15.4)$$

ise $MSE(\hat{\beta}) - MSE(\hat{\beta}_{m,\alpha})$ p.d. dir. Ya da eşdeğer olarak eğer

$$\beta'P \text{diag}\left\{\frac{\lambda_j\lambda_j(m,\alpha)}{2-\lambda_j(m,\alpha)}\right\}_{j=1}^p P\beta < \sigma^2 \quad (3.15.5)$$

ise $MSE(\hat{\beta}) - MSE(\hat{\beta}_{m,\alpha})$ p.d. dir. $j=1,2,\dots,p$ için

$$\lim_{m \rightarrow \infty} \frac{\lambda_j\lambda_j(m,\alpha)}{2-\lambda_j(m,\alpha)} = 0 \quad (3.15.6)$$

olduğundan ; tüm $m \geq m_0$ için $MSE(\hat{\beta}) - MSE(\hat{\beta}_{m,\alpha})$ pd olacak şekilde bir m_0 tamsayısı vardır. (3.15.5) eşitsizliği $MSE(\hat{\beta}) - MSE(\hat{\beta}_{m,\alpha})$ nın p.d. olması için de gereklidir (Trenkler, 1978b). Ancak m_0 ; bilinmeyen β ve σ^2 değerlerine bağlıdır. Fakat ; k bilinen bir pozitif sabit olmak üzere ; β ve σ^2 ile ilgili $\beta\beta' \leq k\sigma^2$

şeklinde bir ön bilgiye sahip olduğumuzu kabul edelim. O zaman ; tüm $m \geq \bar{m}$ için $MSE(\hat{\beta}) - MSE(\hat{\beta}_{m,\alpha})$ p.d. olacak şekilde kesin bir \bar{m} sayısı belirlenebilir. \bar{m} sayısı

$$\lambda_p(\bar{m} + 1, \alpha) = (1 - \alpha \lambda_p)^{\bar{m}-1} < \frac{1}{k \lambda_1} \quad (3.15.7)$$

eşitsizliğini sağlayan en küçük sayı olarak seçilir. İspatı (Trenkler (1978-b)) de yapılmıştır.

3.16. ORR İle İterasyon Tahmin Edicisinin Karşılaştırılması

$C_1 = P'(\Lambda + kI)^{-1} \Omega' Q$ ve $C_2 = P'[\Delta_{m,\alpha} \quad 0] Q$ alınmasıyla

$$C_1 C_1' - C_2 C_2' = P'(\Lambda(\Lambda + kI)^{-2} - \Delta_{m,\alpha}^2) P \quad (3.16.1)$$

elde edilir ve

$$C_1 C_1' - C_2 C_2' = P' \text{diag} \left\{ \frac{\lambda_j^2 - (\lambda_j + k)^2 (1 - \lambda_j(m, \alpha))^2}{\lambda_j (\lambda_j + k)^2} \right\}_{j=1}^p P \quad (3.16.2)$$

bulunur.

m nin belirli olduğunu düşünelim. Tüm $k \leq k_0$ için $C_1 C_1' - C_2 C_2'$ p.d. olacak şekilde bir k_0 sabitinin olduğu açıktır . Böylelikle Teorem 3.14.1 in uygulanmasıyla , $MSE(\hat{\beta}_k) - MSE(\hat{\beta}_{m,\alpha})$ nın p.d. olması için yeterli koşul yani ; $k \leq k_0$ ve

$$B(\hat{\beta}_{m,\alpha})' (C_1 C_1' - C_2 C_2')^{-1} B(\hat{\beta}_{m,\alpha}) < \sigma^2 \quad (3.16.3)$$

elde edilir.

Bu yeterli koşulun geçerliliği , β ve σ^2 değerlerinin büyüklüğüne bağlıdır. Dolayısıyla m rastgele olduğunda , $\hat{\beta}_{m,\alpha}$ nın uygun ridge tahmin edicilerinden MSE

kriterine göre potansiyel olarak daha iyi olduğu k değerlerinin bulunduğu sonucuna ulaşılabilir.

$$C_2 C_2' - C_1 C_1' = P' \text{diag} \left\{ \frac{(\lambda_j + k)^2 (1 - \lambda_j(m, \alpha))^2 - \lambda_j^2}{\lambda_j (\lambda_j + k)^2} \right\}_{j=1}^p P \quad (3.16.4)$$

ifadesinden; belirli bir k için tüm $m \geq m^*$ için her zaman $C_2 C_2' - C_1 C_1'$ p.d. olacak şekilde bir m^* sabitinin bulunabileceği sonucuna ulaşılır.

Yine Teorem 3.14.1 den $MSE(\hat{\beta}_{m,\alpha}) - MSE(\hat{\beta}_k)$ nin p.d. olması için yeterli koşul

$$m \geq m^* \text{ ve } B(\hat{\beta}_k)' (C_2 C_2' - C_1 C_1')^{-1} B(\hat{\beta}_k) < \sigma^2 \quad (3.16.5)$$

olarak elde edilir.

β ve σ^2 üzerinde kısıtlama olmadığı düşünüldüğünde, bu yeterli koşulun sağlanacağı göz ardı edilemez. Buradan elde edilebilecek sonuca göre; MSE kriterine göre bir tahmin edicinin diğerinden daha iyi olduğu gözlenemez. β ve σ^2 nin belirli değerlerine göre bir tahmin edici diğerinden daha iyi olabilir ancak öteki değerler için kötü olabilir. β ve σ^2 ile ilgili ek bilgi verilmediğinde her iki tahmin edici eş zamanlı olarak kullanılmalıdır.

3.17. Shrunken Tahmin Edici İle İterasyon Tahmin Edicinin Karşılaştırılması

Shrunken ve iterasyon tahmin edicilerine ait olan $C_1 = sP'\Lambda^{-1}\Omega'Q$ ve $C_2 = P'[\Delta_{m,\alpha} \ 0]Q$ matrislerini ele alalım.

$$C_1 C_1' - C_2 C_2' = P' \text{diag} \left\{ \frac{s^2 - (1 - \lambda_j(m, \alpha))^2}{\lambda_j} \right\}_{j=1}^p P \quad (3.17.1)$$

olduğu bulunur.

m belirli olduğunda ve $s \rightarrow 1$ iken $C_1 C_1' - C_2 C_2'$ nün p.d. olabileceği sonucuna ulaşılabılır. Bir başka deyişle eğer $0 < s < 1$ ise o zaman tüm $m \geq m'$ için $C_2 C_2' - C_1 C_1'$ p.d. olacak şekilde bir m' sabiti bulunabilir. Sonuçta MSE kriterine göre bir tahmin edicinin diğerinden daha iyi olduğu sonucuna ulaşılamayacağı görülmür.

Gerekli Bir Kriter: $\hat{\beta}_r = A_r^+ X' Y$ şeklindeki PC tahmin edicisi ile $\tilde{\beta}$ homojen lineer tahmin edicileri karşılaştırılırken ; $Cov(\hat{\beta}_r) - Cov(\tilde{\beta})$ nin singüler veya regüler olabileceği ancak p.d. olamayacağı gösterilebilir. Dolayısıyla Teorem 3.14.1 uygulanamaz. Ancak bazen $H_r = \frac{1}{\sigma^2} (Cov(\hat{\beta}_r) - Cov(\tilde{\beta}))$ ifadesi ;

$$H_r = P' \begin{bmatrix} D_1 & 0 \\ 0 & D_2 \end{bmatrix} P \quad (3.17.2)$$

yazımı ya p.d. ya da n.d. olan D_1 ve D_2 köşegen matrisleri içerecek biçimde yazılabilir. Bu nedenle aşağıdaki teorem oluşturulabilir.

Teorem 3.17.1 : $MSE(\tilde{\beta}_1) - MSE(\tilde{\beta}_2)$ p.d. olacak şekilde $\tilde{\beta}_j = C_j Y$ ($j=1,2$) şeklindeki iki homojen lineer tahmin ediciyi ele alalım. Bundan başka D_1 ve D_2 köşegen matrisler olacak şekilde $H = \frac{1}{\sigma^2} (Cov(\tilde{\beta}_1) - Cov(\tilde{\beta}_2))$ nin

$$H = P' \begin{bmatrix} D_1 & 0 \\ 0 & D_2 \end{bmatrix} P \quad (3.17.3)$$

olarak yazıldığını kabul edelim . $B(\tilde{\beta}_1) B(\tilde{\beta}_1)'$ matrisi ; $\gamma_j \gamma_j'$, D_j ($j=1,2$) ile aynı tipte olmak üzere

$$B(\tilde{\beta}_1)B(\tilde{\beta}_1)' = P' \begin{bmatrix} \gamma_1\gamma_1' & \gamma_1\gamma_2' \\ \gamma_2\gamma_1' & \gamma_2\gamma_2' \end{bmatrix} P \quad (3.17.4)$$

olarak yazılsın. O zaman D_1 veya D_2 p.d. ise ; buradan $\gamma_1' D_1^{-1} \gamma_1 < \sigma^2$ veya $\gamma_2' D_2^{-1} \gamma_2 < \sigma^2$ olur.

İspat : D_1 in p.d. olduğunu kabul edelim.

$$MSE(\tilde{\beta}_1) - MSE(\tilde{\beta}_2) = \sigma^2 H + B(\tilde{\beta}_1)B(\tilde{\beta}_1)' - B(\tilde{\beta}_2)B(\tilde{\beta}_2)' \quad (3.17.5)$$

p.d. olduğundan $\sigma^2 D_1 + \gamma_1\gamma_1'$ nün p.d. olması gerekir. Farebrother'ın sonucunun uygulanmasıyla $\gamma_1' D_1^{-1} \gamma_1 < \sigma^2$ elde edilir. Aynı düşünce D_2 için de uygulanır.

3.18. PC İle İterasyon Tahmin Edicinin Karşılaştırılması

Marquardt (1970) PC tahmin edicisini

$$\hat{\beta}_r = A_r^+ X'Y \quad (3.18.1)$$

şeklinde vermiştir. Burada A_r^+ matrisi ; A_r

$$A_r = P' \begin{bmatrix} \Lambda_1 & 0 \\ 0 & 0 \end{bmatrix} P \quad (3.18.2)$$

nm Moore-Penrose tersi olacak şekilde seçilmiştir.

(3.18.2) de $\Lambda_1 = \text{diag}\{\lambda_j\}_{j=1}^r$; $X'X$ in r büyük özdeğerini içerir.

$$X'X = P'\Lambda P = P' \begin{bmatrix} \Lambda_1 & 0 \\ 0 & \Lambda_2 \end{bmatrix} P \quad (3.18.3)$$

şeklinde parçalandığında ; Λ_2 , $X'X$ in ; PC tahmin edicisi tarafından silinen $p-r$ küçük özdeğerini içerir.

$X = Q'\Omega P$ eşitliğinden

$$\hat{\beta}_r = P' \begin{bmatrix} \Lambda_1^{-\frac{1}{2}} & 0 \\ 0 & 0 \end{bmatrix} QY \quad (3.18.4)$$

elde edilir.

$C_1 = X_{m,\alpha}$ ve $C_2 = A_r^+ X'$ nin şeklinde tanımlanmasıyla

$$Cov(\hat{\beta}_{m,\alpha}) - Cov(\hat{\beta}_r) = \sigma^2 P' \left[\Delta_{m,\alpha}^2 - \begin{bmatrix} \Lambda_1^{-1} & 0 \\ 0 & 0 \end{bmatrix} \right] P \quad (3.18.5)$$

in ne p.d. ne de n.d. olduğu sonucuna varılabilir.

Ancak

$$\left\{ \begin{array}{l} d_j = \frac{(1 - \lambda_j(m, \alpha))^2 - 1}{\lambda_j} ; j = 1, 2, \dots, r \text{ iken} \\ \frac{(1 - \lambda_j(m, \alpha))^2}{\lambda_j} ; j = r + 1, \dots, p \text{ iken} \end{array} \right\} \quad (3.18.6)$$

olmak üzere

$$\Delta_{m,\alpha}^2 - \begin{bmatrix} \Lambda_1^{-1} & 0 \\ 0 & 0 \end{bmatrix} = \text{diag}\{d_j\}_{j=1}^p = \begin{bmatrix} D_1 & 0 \\ 0 & D_2 \end{bmatrix} \quad (3.18.7)$$

olduğundan ; D_1 ve D_2 köşegen matrislerinin n.d. ve p.d. olduğu görülür.

$B(\hat{\beta}_{m,\alpha})B(\hat{\beta}_{m,\alpha})'$ nin ; $\gamma = \Lambda_{m,\alpha} P\beta$ olmak üzere

$$B(\hat{\beta}_{m,\alpha})B(\hat{\beta}_{m,\alpha})' = P'\gamma\gamma'P \quad (3.18.8)$$

şeklinde yazılabilir. Veya eşdeğer olarak , $\gamma_j \gamma_j'$, D_j ($j=1,2$) ile aynı tipte ve

$\gamma = [\gamma_1 \quad \gamma_2]$ olmak üzere

$$\mathbf{B}(\hat{\beta}_{m,\alpha})\mathbf{B}(\hat{\beta}_{m,\alpha})' = P' \begin{bmatrix} \gamma_1\gamma_1' & \gamma_1\gamma_2' \\ \gamma_2\gamma_1' & \gamma_2\gamma_2' \end{bmatrix} P \quad (3.18.9)$$

şeklinde yazılabilir. Böylelikle Teorem 3.17.1 in kullanılmasıyla eğer $\hat{\beta}_r$, MSE kriterine göre $\hat{\beta}_{m,\alpha}$ dan daha iyi ise o zaman

$$\gamma_2' D_2^{-1} \gamma_2 < \sigma^2 \quad (3.18.10)$$

sonucuna ulaşılabilir.

Ancak bu koşul γ_2 ve σ^2 nin değerlerine bağlı olarak, geçerli olabilir veya olmayabilir. Dolayısıyla MSE kriterine göre $\hat{\beta}_r$ nin tüm $\hat{\beta}_{m,\alpha}$ dan daha iyi olduğu söylenemez.

Eğer $MSE(\hat{\beta}_r) - MSE(\hat{\beta}_{m,\alpha})$ p.d. ise

$$\gamma_1' D_1^{-1} \gamma_1 < \sigma^2 \quad (3.18.11)$$

sonucuna ulaşılabilir

Bununla beraber bu eşitsizlik sağlanabilir veya sağlanamaz. Dolayısıyla ; β ve σ^2 üzerinde bir ön bilgi olmadan MSE kriterine göre $\hat{\beta}_{m,\alpha}$ nın tüm β ve σ^2 değerlerine göre $\hat{\beta}_r$ den daha iyi olduğu söylenemez.

İterasyon tahmin edicisinin; ridge, shrunken ve PC tahmin edicileri ile karşılaştırılmasında görülür ki MSE kriterine göre hiçbiri diğerine tercih edilemez. Öte yandan diğer yanlı tahmin ediciler gibi iterasyon tahmin edicisi de OLS tahmin edicisinden daha iyidir. Özellikle $\beta'\beta/\sigma^2$ oranı için bir üst sınır bulunduğu iterasyon tahmin edicisi OLS den daha iyidir. Bu oran ; lineer regresyon modelindeki yanlı lineer tahmin edicilerin performansı için anahtar rolü oynar.

İterasyon tahmin edicisini; ridge tahmin edicisine karşı ciddi bir rakip yapan sebepler vardır.

i) İterasyon tahmin edicisini oluşturmak için matris tersinin bulunması gerekmez. X tam sütun ranklı olmasa bile ; 0 ile $\frac{1}{tr(X'X)}$ arasında bir α nın seçimi ; m artarken $\hat{\beta}_{m,\alpha}$ nın yakınsak olmasını sağlar.

ii) $X_0 = \alpha X'$ ve $X_{m+1} = (2I - X_m X) X_m$ tanımlanmasıyla $X_{2^m - 1, \alpha} = X_m$ (Ben-İsrael/Greville (1974)) olduğu gösterilebilir. Bu bağıntı $\hat{\beta}_{m,\alpha}$ nın hesaplanması yönteminin kontrol edilmesinde ve hızlandırılmasında çok yararlıdır.

iii) İterasyon izinin araştırılması ; yani $\hat{\beta}_{m,\alpha}$ nın katsayılarının ve rezidü kareler toplamı $\phi_{m,\alpha} = (Y - X\hat{\beta}_{m,\alpha})'(Y - X\hat{\beta}_{m,\alpha})$ nin şeklinin incelenmesi ; m nin artan değerleri için ridge izinin analizine bir alternatiftir.

iv) Ridge tahmininde k nın seçimi için kullanılan yöntemler iterasyon tahmin prosedürüne uygulanabilir.

3.19. Homojen Lineer Tahmin Edicilerin Karşılaştırılması

Terasvirta'nın araştırmalarının β nın genel lineer ve lineer olmayan tahmin edicilerine genelleştirilebileceği üzerinde durulacaktır. Bundan başka hata vektörünün sadece skaler kovaryans matrisi durumu incelenmeyecektir. MSE matrislerine göre bazı yanlı ve yansız tahmin edicilerin birbiriyle karşılaştırılmaları ele alınacaktır.

Terasvirta'nın Sonuçlarının Genelleştirilmesi

Terasvirta'nın varsayımlarını genelleştirerek ; X $n \times p$ tipinde bir matris, V $n \times n$ tipinde p.d. bir matris olmak üzere ve X , ε ile ilişkili olmamak üzere

$$Y = X\beta + \varepsilon , E(\varepsilon) = 0 , Cov(\varepsilon) = \sigma^2 V$$

lineer modelini ele alalım.

b_1 ve b_2 , β nin herhangi iki tahmin edicisi olsun. Amaç b_j nin homojen lineer tahmin edici olduğunu kabul etmeden

$$MSE(b_j) = E[(b_j - \beta)(b_j - \beta)'] \quad (3.19.1)$$

şeklindeki MSE matrisleri farkını incelemektir.

Ayrıca $b_j = C_j Y + c_j$ formundaki tahmin ediciler veya lineer olmayan tahmin ediciler ele alınacaktır. Kovaryans matrisi

$$Cov(b_j) = E[(b_j - E(b_j))(b_j - E(b_j))'] , j = 1,2 \quad (3.19.2)$$

yanlılık

$$Bias(b_j) = E(b_j) - \beta , j = 1,2 \quad (3.19.3)$$

şeklinde gösterilecektir.

Böylece bilinen bir eşitlik

$$MSE(b_j) = Cov(b_j) + Bias(b_j)Bias(b_j)' , j = 1,2 \quad (3.19.4)$$

elde edilir.

$$C = Cov(b_1) - Cov(b_2) \quad (3.19.5)$$

ve

$$B_{21} = Bias(b_2)Bias(b_2)' - Bias(b_1)Bias(b_1)' \quad (3.19.6)$$

olsun. C nin p.d. olması ; $C = TT'$ olacak şekilde regüler bir T matrisinin olduğunu belirtir. Böylece

$$MSE(b_1) - MSE(b_2) = T(I - T^{-1}B_{21}T'^{-1})T' \quad (3.19.7)$$

nin p.d. olması için gerek ve yeter koşulun ; $I - T^{-1}B_{21}T'^{-1}$ in p.d. olması olduğu sonucuna ulaşırız. Bunun olması için gerek ve yeter koşul $T^{-1}B_{21}T'^{-1}$ in tüm özdeğerlerinin 1 den küçük olmasıdır.

$g = T^{-1}Bias(b_1)$ ve $h = T^{-1}Bias(b_2)$ denirse ve Terasvirta'nın Lemması uygulanırsa aşağıdaki genelleme bulunur.

Teorem 3.19.1 : $C = Cov(b_1) - Cov(b_2)$ nin p.d. olduğunu kabul edelim. O zaman $MSE(b_1) - MSE(b_2)$ nin p.d. olması için gerek ve yeter koşul

$$\rho_j = Bias(b_j)' C^{-1} Bias(b_j), j=1,2 \quad (3.19.8)$$

ve

$$\eta = Bias(b_1)' C^{-1} Bias(b_2) \quad (3.19.9)$$

olmak üzere

$$\frac{1}{2}(\rho_2 - \rho_1) + \sqrt{\frac{1}{4}(\rho_2 - \rho_1)^2 + \rho_1\rho_2 - \eta^2} < 1 \quad (3.19.10)$$

olmasıdır.

Sonuç 3.19.1 : $Cov(b_1) - Cov(b_2)$ nin p.d. olduğunu kabul edelim. O zaman $\rho_2 < 1$ ise $MSE(b_1) - MSE(b_2)$ p.d. dir. Bundan başka eğer b_1, β için yansız ise $\rho_2 < 1$ koşulu , $MSE(b_1) - MSE(b_2)$ nin p.d. olması için gerek ve yeter koşul olacaktır.

X tam sütun ranklı olsun. Önemli bir durum olan tahmin edicilerden örneğin b_1 in $b = (X'V^{-1}X)^{-1} X'V^{-1}Y$ şeklindeki OLS tahmin edicisi olmasını ele alalım. O zaman aşağıdaki sonuç ortaya çıkar.

Sonuç 3.19.2 : b_2, β nin bir başka tahmin edicisi olmak üzere $(X'V^{-1}X)^{-1} - \frac{1}{\sigma^2} Cov(b_2)$ nin p.d. olduğunu kabul edelim. O zaman aşağıdaki iki varsayım eşdeğerdir.

i) $MSE(b) - MSE(b_2)$ p.d. dir.

$$\text{ii) } Bias(b_2)' \left[(X'V^{-1}X)^{-1} - \frac{1}{\sigma^2} Cov(b_2) \right]^{-1} Bias(b_2) < \sigma^2 \quad (3.19.11)$$

Önceki sonuçları ileride uygulamak için $V = \sigma_2 I$ olduğunu kabul edelim. Ayrıca

$\Omega = \begin{bmatrix} \Lambda^{\frac{1}{2}} & 0 \end{bmatrix}'$ nxp tipinde bir matris , P ve Q ortogonal matrisler olmak üzere

bağımsız değişkenler matrisi $X = Q'\Omega P$ (tekil değer ayrışımı) formunda yazılmış olsun. Λ köşegen matrisi , $X'X$ in özdeğerlerinden oluşmuş ve $X'X = P'\Lambda P$ dir.

L , pxp tipinde negatif olmayan elemanlardan oluşan bir köşegen matris ve C , px1 tipinde stokastik olmayan bir vektör olmak üzere

$$\tilde{\beta} = P' \begin{bmatrix} L^{\frac{1}{2}} & 0 \end{bmatrix} QY + C \quad (3.19.12)$$

şeklindeki lineer tahmin edicilerden oluşan ϕ_c sınıfını ele alalım. Uygun k ve c nin seçilmesiyle , ϕ_c nin kaynaklardaki birçok homojen yanlı tahmin edicileri , örneğin ridge, shrunken, iterasyon tahmin edicileri içerdiği görülür. Ayrıca OLS tahmin edicisi de bu sınıfa aittir. ϕ_c nin

$$\tilde{\beta}_j = P' \begin{bmatrix} L_j^{\frac{1}{2}} & 0 \end{bmatrix} QY + r_j , j = 1,2 \quad (3.19.13)$$

şeklinde iki tahmin edicisini ele alalım.

$L_1 - L_2$ nin regüler olduğunu kabul ederek

$$F = (L_1 - L_2)^{-1} , D_j = (L_j \Lambda)^{\frac{1}{2}} - I$$

$$\gamma = P\beta , \delta_j = Pc_j , j = 1,2$$

oluşturalım.

Bazı cebirsel işlemlerden sonra

$$\rho_j = \frac{1}{\sigma^2} \left[\gamma F D_j^2 \gamma + 2\gamma D_j' F \delta_j + \delta_j' F \delta_j \right] \quad (3.19.14)$$

ve

$$\eta = \frac{1}{\sigma^2} \left[\gamma D_1' F D_2 \gamma + \gamma F (D_1 \delta_2 + D_2 \delta_1) + \delta_1' F \delta_2 \right] \quad (3.19.15)$$

olarak bulunur.

Bu formüllerle MSE kriterine göre bir tahmin edicinin diğerinden daha iyi olup olmadığına karar vermek daha kolaydır. Örneğin, eğer $\tilde{\beta}_1$, yansız ve F p.d. ise Sonuç 3.19.2 den $MSE(\tilde{\beta}_1) - MSE(\tilde{\beta}_2)$ nin p.d. olması için gerek ve yeter koşulun

$$\gamma F D_2^2 \gamma + 2\gamma D_2' F \delta_2 + \delta_2' F \delta_2 < \sigma^2 \quad (3.19.16)$$

olduğu söylenebilir.

3.20. Minimum Risk Tahmin Edicileri

$b = (X'V^{-1}X)^{-1} X'V^{-1}Y$ şeklindeki genelleştirilmiş LS tahmin edicisine bir alternatif olarak ; $H = \beta^* \beta^{*'} + \sigma^2 (X'V^{-1}X)^{-1}$ ve β^* px1 tipinde stokastik olmayan bir vektör olmak üzere

$$b^* = \beta^* \beta^{*'} H^{-1} b \quad (3.20.1)$$

şeklindeki minimum koşul hata karesi ortalaması tahmin edicisini (MCMSE) ele alabiliriz.

$W_\mu = \frac{1}{\sigma^2} X'V^{-1}X + \mu \Delta^{-1}$; μ negatif olmayan bir skaler ; $\bar{\beta}$ px1 tipinde bir vektör ve Δ^{-1} pxp tipinde bir p.d. matris olmak üzere minimum ortalama risk lineer tahmin edicisi (MARLE)

$$b_\mu = W_\mu^{-1} \left(\frac{1}{\sigma^2} X'V^{-1}XY + \mu \Delta^{-1} \bar{\beta} \right) \quad (3.20.2)$$

şeklindedir.

Yazımları basitleştirmek için aşağıdaki kısaltmaları yapalım .

$$S = X'V^{-1}X , \varepsilon = \beta^{*'} S \beta^*$$

$$A_\mu = W_\mu S^{-1} W_\mu , \phi = \beta^{*'} A_\mu \beta^*$$

$$H^{-1} = \frac{1}{\sigma^2} \left[S - \frac{1}{(\sigma^2 + \varepsilon)} S \beta^* \beta^{*'} S \right] \text{ olsun . MCMSE alternatif olarak}$$

$$b^* = \frac{1}{\sigma^2 + \varepsilon} \beta^* \beta^{*'} X'V^{-1}Y \quad (3.20.3)$$

şeklinde yazılabilir.

$$b_\mu \text{ nün yanlılığı hesaplanırken } I - \frac{1}{\sigma^2 W_\mu^{-1} S} = \mu W_\mu^{-1} \Delta^{-1} \text{ bağıntısı kullanılır.}$$

3.21. GLS İle Minimum Risk Tahmin Edicilerinin Karşılaştırılması

Swamy -Mehta (1977), MSE matrislerinin farkının n.n.d olmaları için gerekli koşulları ortaya koymuşlardır. Cauchy-Schwarz eşitsizliği kullanılmadan da Swamy - Mehta (1977) nin elde ettiği sonuçlara benzer sonuçlar elde edilir.

Teorem 3.21.1 : $MSE(b) - MSE(b^*)$ nin p.d. olması için gerek ve yeter koşul

$$\beta' \left[S - \frac{2}{\sigma^2 + 2\varepsilon} S \beta^* \beta^{*'} S \right] \beta < \sigma^2 \quad (3.21.1)$$

olmasıdır.

| | GLS b | MCMSE b^* | MARLE b_μ |
|-------------------|-------------------|--------------------------------------------------------------------------------|----------------------------------------------------|
| Yanlılık | 0 | $\left(\frac{1}{\sigma^2 + \varepsilon} \beta^* \beta^{*'} S - I\right) \beta$ | $\mu W_\mu^{-1} \Delta^{-1} (\bar{\beta} - \beta)$ |
| Kovaryans Matrisi | $\sigma^2 S^{-1}$ | $\frac{\sigma^2 \varepsilon}{(\sigma^2 + \varepsilon)^2} \beta^* \beta^{*'}$ | $\frac{1}{\sigma^2} A_\mu^{-1}$ |

Tablo 3.21.1 : Üç tahmin edicinin yanlılıkları ve kovaryans matrisleri

İspat : Tablo 3.21.1 den

$$\left[S^{-1} - \frac{1}{\sigma^2} \text{Cov}(b^*) \right]^{-1} = S + \frac{\varepsilon}{\sigma^4 + 2\sigma^2 \varepsilon} S \beta^* \beta^{*'} S \quad (3.21.2)$$

nin bir p.d. matris olduğu görülebilir. Buradan

$$\text{Bias}(b^*)' \left[S^{-1} - \frac{1}{\sigma^2} \text{Cov}(b^*) \right]^{-1} \text{Bias}(b^*) = \beta' \left[S - \frac{2}{\sigma^2 + 2\varepsilon} S \beta^* \beta^{*'} S \right] \beta \quad (3.21.3)$$

elde edilir. Bu eşitlikten ve Sonuç 3.19.2 den faydalanılarak kabul edilen denklik elde edilir.

Teorem 3.21.1 : $\mu > 0$ olduğunu kabul edelim. O zaman $MSE(b) - MSE(\mu)$ nün p.d. olması için gerek ve yeter koşul

$$(\bar{\beta} - \beta)' \left[\frac{2}{\mu \sigma^2} \Delta + S^{-1} \right]^{-1} (\bar{\beta} - \beta) < \sigma^2 \quad (3.21.4)$$

olmasıdır.

İspat : Tablo 3.21.1 den faydalanılarak

$$\begin{aligned} \left[S^{-1} - \frac{1}{\sigma^2} \text{Cov}(b_\mu) \right]^{-1} &= \left[S^{-1} - \frac{1}{\sigma^4} A_\mu^{-1} \right]^{-1} = W_\mu \left[A_\mu - \frac{1}{\sigma^4} S \right]^{-1} W_\mu \\ &= \frac{1}{\mu^2} W_\mu \Delta \left[\frac{2}{\mu \sigma^2} \Delta + S^{-1} \right]^{-1} \Delta W_\mu \end{aligned} \quad (3.21.5)$$

in p.d. olduğu sonucuna varılabilir.

Yine Tablo 3.21.1 den elde edilen $\text{Bias}(b_\mu) = \mu W_\mu^{-1} \Delta^{-1} (\bar{\beta} - \beta)$ eşitliği ve Sonuç 3.19.2 den yola çıkılarak istenen denklik elde edilir.

3.22. Mcmse İle Marle' nin Karşılaştırılması

Belirli koşullar altında MSE ye göre b^* , b_μ den daha iyidir. Her ne kadar Swamy-Mehta (1976) böyle bir karşılaştırma yapmanın zor olduğunu belirtse de Sonuç 3.19.1 kullanılarak, b^* in b_μ den daha iyi olması için yeterli bir koşul oluşturulabilir.

μ nün belirli olduğunu kabul edelim. O zaman öklid normu yeterince küçük olan tüm β^* için

$$2\mu\beta^{*'} \Delta^{-1} \beta^* + \mu^2 \sigma^2 \beta^{*'} \Delta^{-1} S^{-1} \Delta^{-1} \beta^* < \frac{\sigma^2}{\varepsilon} + 2 \quad (3.22.1)$$

elde edilir.

Teorem 3.22.1 : $\sigma^2 \beta' A_\mu \beta < 1$ olduğunu kabul edelim. O zaman $MSE(b_\mu) - MSE(b^*)$ p.d. olacak şekilde bir b^* tahmin edicisi vardır.

İspat : (3.22.1) deki eşitsizliği sağlayan bir β^* vektörünü seçelim. O zaman

$$B = \frac{(\sigma^2 + \varepsilon)^2}{\sigma^4 \varepsilon} A_\mu^{-1} - \beta^* \beta^{*'} \quad (3.22.2)$$

olmak üzere

$$Cov(b_{\mu}) - Cov(b^*) = \frac{\sigma^2 \varepsilon}{(\sigma^2 + \varepsilon)^2} B \quad (3.22.3)$$

elde edilir.

Farebrother (1976) dan , B nin p.d. olması için gerek ve yeter koşul $\sigma^2 \phi < \frac{(\sigma^2 + \varepsilon)^2}{\sigma^2 \varepsilon}$ olmasıdır. Ancak β^* üzerindeki , $Cov(b_{\mu}) - Cov(b^*)$ ın p.d. olduğu sonucuna ulaşabileceğimiz (3.22.1) kısıtlamasından dolayı $\sigma^2 \phi$ nin

$$\beta^{*'} \left[\frac{1}{\sigma^2 S} + 2\mu \Delta^{-1} + \mu^2 \sigma^2 \Delta^{-1} S^{-1} \Delta^{-1} \right] \beta^*$$

olduğu görülebilir.

Bunun tersi

$$w = \frac{\sigma^4 \varepsilon}{(\sigma^2 + \varepsilon)^2 - \sigma^4 \varepsilon \phi} \quad (3.22.4)$$

olmak üzere

$$[Cov(b_{\mu}) - Cov(b^*)]^{-1} = \sigma^2 [A_{\mu} + w A_{\mu} \beta^* \beta^{*'} A_{\mu}] \quad (3.22.5)$$

olarak verilebilir.

$$G = Bias(b^*)' [Cov(b_{\mu}) - Cov(b^*)]^{-1} Bias(b^*) \quad (3.22.6)$$

ve $\varphi = \frac{1}{(\sigma^2 + \varepsilon)}$ ın yazılmasıyla ve bazı işlemlerin yapılmasıyla

$$G = \sigma^2 \beta' \left[\phi \varphi^2 (1 + w \phi) S \beta^* \beta^{*'} - 2\varphi (1 + w \phi) S \beta^* \beta^{*'} A_{\mu} + A_{\mu} + w A_{\mu} \beta^* \beta^{*'} A_{\mu} \right] \beta \quad (3.22.7)$$

elde edilir.

$\beta^* \rightarrow 0$ olsun. O zaman ϕ , φ ve w nin sonlu limitleri vardır. Sonuç olarak G, $\sigma^2 \beta' A_{\mu} \beta$ ya yaklaşır ve varsayımdan dolayı bu 1 den küçüktür.

Sonuç : Teorem 3.19.1 e göre, kovaryans matrisleri bilindiğinde, lineer tahmin ediciler ile diğer tahmin ediciler arasında bir karşılaştırma yapılması mümkün olabilir. Judge ve Bock (1978) tarafından araştırılan pre-test tahmin edicisi lineer değildir ve eğer hata vektörünün çoklu normal dağılıma sahip olduğu kabul ediliyorsa ; bu tahmin edicinin kovaryans matrisi kesin olarak hesaplanabilir. Ancak LS ve pre-test tahmin edicileri karşılaştırılırken kovaryans matrislerinin farkının tekil olmayan olduğu gösterilemez ve böylece Teorem 3.19.1 bu durumda geçerli olmaz. Bununla beraber Judge ve Bock tarafından yapılan uygun mse matrislerinin farkına dayanan bir karşılaştırma yapılması mümkündür. Lineer olmayan diğer tahmin ediciler, Teorem 3.19.1 in uygulaması için uygun değildir. Çünkü onların kovaryans matrisleri çok karmaşıktır ve bunlar sadece asimptotik olarak bulunur.

3.23. Heterojen Lineer Tahmin Edicilerin Karşılaştırılması

$Y = X\beta + \varepsilon$, $\varepsilon \approx N(0, \sigma^2 I)$ (1.2.1) lineer regresyon modelini ele alalım. D_j belirli bir $p \times n$ tipinde matris ve h_j belirli bir $p \times 1$ tipinde vektör ve $j=1,2$ olmak üzere β nın $b_j = D_j Y + h_j$ şeklindeki iki heterojen tahmin edicisi tanımlansın. Hangi koşullarda bu tahmin edicilerden birinin diğerinden daha iyi olduğunu inceleyelim. b_1 in b_2 üzerindeki güçlü üstünlüğü (strong superiority) karesel risk fonksiyonunun kullanılmasıyla tanımlansın. Toro-Vizcorrondo ve Wallace (1968) parametre uzayında bir tek (β, σ^2) noktasında aşağıdaki tanımları yapmışlardır.

Tanım 3.23.1 : b_2 tahmin edicisinin , (β, σ^2) de b_1 den güçlü üstün olması için gerek ve yeter koşul tüm $A \geq 0$ kayıp matrisleri için

$$E(b_1 - \beta)' A (b_1 - \beta) \geq E(b_2 - \beta)' A (b_2 - \beta)$$

olmasıdır. $MSE(b_j) = E(b_j - \beta)(b_j - \beta)'$ olmak üzere $\Delta_{12} = MSE(b_1) - MSE(b_2)$

oluşturalım. O zaman bu tanım $\Delta_{12} \geq 0$ a denk olur.

Üstünlük için daha az kısıtlayıcı tanımlar ; koşulun tüm n.n.d. matrisleri için eş zamanlı olarak geçerli olması gerekliliğinin kaldırılmasıyla oluşturulur.

Böyle bir tanım aşağıdaki gibidir.

Tanım 3.23.2 : b_2 tahmin edicisinin (β, σ^2) de b_1 den zayıf üstün (weak superior) olması için gerek ve yeter koşul

$$E(b_1 - \beta)' X'X(b_1 - \beta) \geq E(b_2 - \beta)' X'X(b_2 - \beta)$$

olmasıdır.

Bu durumda X ve β verildiğinde, kayıp ile $X\beta$ nin tahmini için kullanılan tahmin edicinin performansı ve Y nin koşullu beklenen değeri bağlantılıdır. Bu tanım ile yararlı sonuçlar elde edildiğinden kullanılabilir. Wallace (1972) Tanım 3.23.2 yi ikinci zayıf mse kriteri olarak adlandırmıştır.

Her iki tanım da; bir zamanda, parametre uzayının bir tek noktasındaki üstünlük ile ilgilidir.

Üstünlük Koşulları

İlk olarak güçlü üstünlüğü ele alalım. Genelliği kaybetmeden, MSE matrisi ; kovaryans ve yanlılık ifadelerine ayrıştırılabilir. Dolayısıyla , $j=1,2$ için $H_j = D_j X - I$ ve $d_j = H_j \beta + h_j$ olmak üzere $MSE(b_j) = \sigma^2 D_j D_j' + d_j d_j'$ dür. Bundan başka $j=1,2$ için, K $p \times r$ tipinde rank $(K) = r$ olan bir matris ; L $r \times r$ tipinde simetrik bir matris ve f_j $r \times 1$ tipinde vektör olmak üzere

$$C = K L K' , d_j = K f_j , j = 1,2 \quad (3.23.1)$$

şeklinde bir ayrışımın yapıldığını kabul edelim. Bu ayrışım , tahmin edicilerin tekil kovaryans matrisleriyle karşılaştırılmasında yararlıdır.

$\sigma^2 L + f_1 f_1'$ p.s.d. olmak üzere ve $N[A]$, A nın sıfır uzayı olmak üzere eğer

$N[f_2] = N[f_1]$ veya $N[\sigma^2 L + f_1 f_1'] \subset N[f_2]$ ise karşılaştırmalar benzer olur. Özel durumlar göz önüne alınarak aşağıdaki teorem oluşturulur.

Teorem 3.23.1 : (1.2.1) modelini ve $j=1,2$ için $b_j = D_j Y + h_j$ şeklindeki iki heterojen lineer tahmin ediciyi ele alalım. (3.23.1) ayrışımının olduğunu ve $N[f_2] \neq N[f_1]$ ve $\sigma^2 L + f_1 f_1' > 0$ olduğunu kabul edelim. O zaman b_2 nin b_1 den güçlü üstün olması için gerek ve yeter koşul

$$f_2' (\sigma^2 L + f_1 f_1')^{-1} f_2 \leq 1 \quad (3.23.2)$$

olmasıdır.

Eğer $L > 0$ kabul edilirse güçlü üstünlük $f_j = f_j' L^{-1} f_j$, $j = 1,2$ olmak üzere

$$\sigma^2 \left\{ f_{22} - f_{21}^2 (\sigma^2 + f_{11})^{-1} \right\} \leq 1 \quad (3.23.3)$$

ifadesine eşdeğer olur. Öte yandan eğer $L < 0$ ise b_2 nin b_1 den güçlü üstün olması için gerek ve yeter koşul L nin bir skaler olması ve

$$\sigma^2 \ell_{11} + f_1^2 - f_2^2 \geq 0 \quad (3.23.4)$$

eşitsizliğinin sağlanmasıdır.

İspat : (3.23.1) ayrışımının olduğu kabul edilirse ,
 $\Delta_{12} = \sigma^2 C + d_1 d_1' - d_2 d_2' \geq 0$ ifadesi $\sigma^2 L + f_1 f_1' - f_2 f_2' \geq 0$ (3.23.5) ifadesine eşdeğer olur.

Bu durumda eğer $\sigma^2 L + f_1 f_1' > 0$ ise (3.23.5) ifadesi (3.23.2) ye eşdeğer olur (Farebrother (1975)). Eğer $L > 0$ ise (3.23.3) den (3.23.2) elde edilir. Eğer $L < 0$ ise ; Guilkey ve Price (1981) deki ; (3.23.5) in, sadece L bir skaler olduğunda yani $r=1$ olduğunda sağlandığını belirten lemma dan bulunur.

L matrisi ; Teorem 3.23.1 uygulandığında kilit pozisyonundadır. Eğer L matrisi ne p.d. ne de bir skaler ise geriye kalan tek olasılık, $r-1$ tane pozitif ve bir tane

sıfır veya negatif özdeğer olduğudur.

(3.23.3) den , $L > 0$ olduğunda $\Delta_{12} \geq 0$ olması için bir yeterli koşulun $\sigma^{-2} f_{22} \leq 1$ (Trenkler-1980) olduğu görülür. Eğer b_1 yansız olacak şekilde $f_1 = 0$ ise, bu koşul da yeterlidir.

Tanım 3.232 den zayıf üstünlük için aşağıdaki teorem elde edilir.

Teorem 3.23.2 : (1.2.1) lineer modelini ve b_1, b_2 tahmin edicilerini ele alalım. O zaman b_2 nin b_1 den zayıf üstün olması için gerek ve yeter koşul

$$\sigma^{-2} \left(d_2' X' X d_2 - d_1' X' X d_1 \right) \leq tr(X' X C)$$

olmasıdır.

Örnekler

$j=1,2$ için , R $m \times p$ tipinde belirli bir matris , $\text{rank}(R) = m$, r $m \times 1$ tipinde stokastik bir vektör, $k_j > 0$ olmak üzere $b_R(k_j) = (X'X + k_j R'R)^{-1} (X'Y + k_j R'r)$ şeklindeki iki mixed tahmin ediciyi ele alalım (Theil ve Goldberger, 1961). Eğer $E(r - R\beta) \neq 0$ ise o zaman $b_R(k_2)$ nin $b_R(k_1)$ den güçlü üstün olması için gerekli bir koşul $k_2 > k_1$ olmasıdır. Bu ayrıca $b_R^*(k_j) = (X'X + k_j R'R)^{-1} X'Y$, $j = 1,2$ şeklindeki iki minimax tahmin edici karşılaştırıldığında da doğrudur (Kuks ve Olman, 1972). Eğer $k_2 > k_1$ ise üstünlük koşulu için gerek ve yeter koşul Teorem 3.23.1 den elde edilir.

Şimdi RLS tahmin edicisini ve $j=1,2$ için $\tilde{r}_j = \tilde{R}_j \beta$ şeklindeki iki lineer kısıtlama kümesini ele alalım. Uygun RLS tahmin edicileri : $\hat{s}_j = \tilde{r}_j - \tilde{R}_j b$, $b = UX'Y$, $U = (X'X)^{-1}$ olmak üzere $b_{\tilde{R}_j} = b + U \tilde{R}_j' \left(\tilde{R}_j U \tilde{R}_j' \right)^{-1} \hat{s}_j$, $j = 1,2$

şeklindedir. $Cov(b_{\tilde{R}_j}) = \sigma^2 \left(U - U\tilde{R}_j' \left(\tilde{R}_j U \tilde{R}_j' \right)^{-1} \tilde{R}_j' U \right)$ olur. Teorem 3.23.1 i bu duruma uygulamaya çalışalım.

$r_3 = R_3 \beta$, her iki küme için ortak olan kısıtlamaların bir alt kümesi olacak şekilde

$$\tilde{R}_j = \begin{bmatrix} R_j \\ R_3 \end{bmatrix}, \tilde{r}_j = \begin{bmatrix} r_j \\ r_3 \end{bmatrix}, \tilde{s}_j = \begin{bmatrix} s_j \\ s_3 \end{bmatrix} = \begin{bmatrix} r_j - R_j \beta \\ r_3 - R_3 \beta \end{bmatrix} \quad (3.23.6)$$

şeklindeki blok bölmenin olduğunu kabul edelim. Kısıtlamaların doğru olmadığını varsayalım. r_j , $m \times 1$ tipinde bir vektör ve R_j $m \times p$ tipinde bir matris ve $j=1,2$ olsun. Eğer $m_1 > 0$ ise $m_1 + m_2 + m_3 \leq p$ ve eğer $m_1 = 0$ ise $m_2 + m_3 < p$ olduğunu kabul edelim. Bundan başka $rank \begin{pmatrix} R_1' & R_2' & R_3' \end{pmatrix} = m_1 + m_2 + m_3$ olsun . $m_j = 0$ olması kısıtlamaların j -nci kümesinin olmadığını gösterir.

(3.23.6) daki blok bölmenin kullanılmasıyla

$$D_{j \cdot 3} = R_j U R_j' - R_j U R_3' \left(R_3 U R_3' \right)^{-1} R_3 U R_j' = \sigma^2 R_j Cov(b_{R_3}) R_j' , \quad j=1,2$$

ve $B = I - R_3' \left(R_3 U R_3' \right)^{-1} R_3 U$ olmak üzer

$$C = UB \left(R_2' D_{2 \cdot 3}^{-1} R_2 - R_1' D_{1 \cdot 3}^{-1} R_1 \right) B' U$$

elde edilir. Burada b_{R_3} , $r_3 = R_3 \beta$ kısıtlamasına dayalı RLS tahmin edicisidir. C genellikle tanımsızdır. Blok bölmenin uygulanmasıyla

$$d_j = U \left\{ B R_j' D_{j \cdot 3}^{-1} s_j + \left(I - B R_j' D_{j \cdot 3}^{-1} R_j U \right) R_3' \left(R_3 U R_3' \right)^{-1} s_3 \right\}, \quad j=1,2 \quad (3.23.7)$$

elde edilir.

(3.23.7) ve C nin karşılaştırılmasıyla ; $s_3 = 0$ olmadıkça Teorem 3.23.1 in

uygulanamayacağı açıktır. Bu varsayımı yapalım ve $R' = \begin{pmatrix} R_1' & R_2' \end{pmatrix}$ tanımlayalım.

$$K = UBR', \quad L = \text{diag}\{-D_{11.3}^{-1}, D_{22.3}^{-1}\}$$

$$f_1 = \begin{bmatrix} D_{11.3}^{-1} \\ 0 \end{bmatrix} s_1 \quad \text{ve} \quad f_2 = \begin{bmatrix} 0 \\ D_{22.3}^{-1} \end{bmatrix} s_2$$

olarak seçelim. Teorem 3.23.1 in sıfır uzay varsayımları sağlanır. L matrisi tekil olmayan bir matristir. Fakat tanımsızdır, $\text{rank}(K) = m_1 + m_2$ ve negatif özdeğerlerin sayısı m_1 dir.

$$\sigma^2 L + f_1 f_1' = \text{diag}\{-\sigma^2 D_{11.3}^{-1} + D_{11.3}^{-1} s_1 s_1' D_{11.3}^{-1}, \sigma^2 D_{22.3}^{-1}\} \quad (3.23.8)$$

yazalım ve $m_1 = 1$ olarak seçelim . (3.23.8) in p.d. olması için gerek ve yeter koşul

$$\sigma^{-2} s_1' D_{11.3}^{-1} s_1 > 1 \quad (3.23.9)$$

olmasıdır. Bu da yansız tahmin edici b_{R_3} ün $b_{\tilde{R}_1}$ den güçlü üstün olması için bir koşuldur. Eğer $m_1 > 1$ ise o zaman (3.23.9) yine gerekli koşuldur. Ancak artık yeterli değildir.

Eğer (3.23.9) sağlanıyorsa ve $m_1 = 1$ ise o zaman Teorem (3.23.1) uygulanabilir ve ((3.23.9) verildiğinde) $b_{\tilde{R}_2}$ nin $b_{\tilde{R}_1}$ den güçlü üstün olması için gerek ve yeter koşul

$$\sigma^{-2} s_2' D_{22.3}^{-1} s_2 \leq 1 \quad (3.23.10)$$

olmasıdır.

(3.23.10) eşitsizliği , $b_{\tilde{R}_2}$ nin b_{R_3} den güçlü üstün olması için bir koşuldur.

Eğer $m_1 = 0$ ise $K = UBR_2'$, $L = D_{22.3}^{-1}$ ve $f_2 = D_{22.3}^{-1} s_2$ seçilebilir. Teorem 3.23.1 in uygulanmasıyla (3.23.10) elde edilir. Çünkü $L > 0$ ve $\text{rank}(K) = m_2$ dir.

Sonuç 3.23.1 : (1.2.1) lineer modelini göz önüne alalım. m_3 tane $r_3 = R_3 \beta$

ortak kısıtlamaya sahip $b_{\tilde{R}_1}$ ve $b_{\tilde{R}_2}$ RLS tahmin edicilerini ele alalım. $rank\left(R_1' R_2' R_3'\right) = m_1 + m_2 + m_3$ ve $s_3 = 0$ olduğunu yani ortak kısıtlamaların doğru olduğunu kabul edelim.

Eğer $m_1 = 1$ ise $b_{\tilde{R}_2}$ nin $b_{\tilde{R}_1}$ den güçlü üstün olması için gerek ve yeter koşul

- i) Yansız tahmin edici b_{R_3} ün $b_{\tilde{R}_1}$ den güçlü üstün olması
- ii) $b_{\tilde{R}_2}$ nin b_{R_3} den güçlü üstün olmasıdır.

Eğer $m_1 = 0$ ise $R_3 = \tilde{R}_1$ dir ve üstünlük koşulu (3.23.10) dur. Eğer $m_1 > 1$ ise üstünlük koşulu yoktur.

Bu sonuç bize, yanlı RLS tahmin edicisinin aynı tipteki bir başka tahmin edici ile hangi koşullar altında karşılaştırıldığını gösterir.

Doğru olmayan lineer kısıtlamalardan birinin kaldırılması , tahminin geçerliliğini arttırmalı ve sonuçta elde edilen tahmin edici yansız olmalıdır. b_{R_3} yansız tahmin edicisi, yeni $r_2 = R_2\beta$ kısıtlamalarının eklenmesiyle düzenlenebilir. Bu kısıtlamaların doğru olması gerekmez. Ancak (3.23.10) için gerekli bir koşul ; b_{R_2} nin b den güçlü üstün olması yani $Q = \sigma^{-2} s_j' \left(R_j U R_j' \right)^{-1} s_j$ olmak üzere $Q_2 \leq 1$ olmasıdır. Eğer $s_3 = 0$ ise, Gauss-Markov teoremine göre b_{R_3} , b den güçlü üstün olur, öyle ki b_{R_2} nin b üzerindeki güçlü üstünlüğü ; $b_{\tilde{R}_2}$ nin b_{R_3} üzerindeki güçlü üstünlüğü için yeterli değildir. Bu (3.23.10) un

$$Q_2 + \sigma^{-2} s_2' \left(R_2 U R_2' \right)^{-1} R_2 U R_3' D_{33.2}^{-1} R_3 U R_2' \left(R_2 U R_2' \right)^{-1} s_2 \leq 1$$

şeklinde yazılmasıyla görülür.

Guilkey ve Price (1981, Teorem 3) ayrıca iki RLS tahmin edicisinin karşılaştırılmasını ele almıştır. Sonuçları ise s_3 ü hiçbir şekilde kısıtlamaz.

Problemin temeli ise teoremlerinin ispatının (i) kısmındaki,

$MSE(b_{\tilde{R}_1}) - MSE(b_{\tilde{R}_2}) \geq 0$ olması için gerek ve yeter koşulun yanlış olmasıdır.

Şimdi de, iki tahmin edicinin, $m_3 = 0$ olacak şekilde ortak kısıtlamalar içermediğini kabul edelim. O zaman $D_{11:3} = R_1UR_1'$ ve $D_{22:3} = R_2UR_2'$ olur.

Sonuç 3.23.2 : (1.2.1) lineer modelini ve b_{R_1} , b_{R_2} RLS tahmin edicilerini ele alalım. $R' = \begin{pmatrix} R_1' & R_2' \end{pmatrix}$ nün sütunları lineer bağımsız olsun. Eğer $m_1 = 1$ ise b_{R_2} nin b_{R_1} den güçlü üstün olması için gerek ve yeter koşul

i) OLS tahmin edicisi b nin b_{R_1} den güçlü üstün olması

ii) b_{R_2} nin b den güçlü üstün olmasıdır.

Eğer $m_1 > 1$ ise güçlü üstünlük için koşul oluşmaz. Guilkey ve Price (1981) benzer bir sonuç elde etmişlerdir (Teorem 4). Bununla birlikte R_1 ve R_2 üzerinde daha kesin ek kısıtlamalar getirmişlerdir. Sonuç 3.23.2 den bunun gerekli olmadığı görülür.

PC tahmin edicisi, RLS tahmin edicisinin bir özel halidir (Judge ve ark., 1985)

Burada $X'X$ ün özdeğerlerinin pozitif olduğunu varsayalım.

Sonuç 3.23.3 : (1.2.1) deki lineer modeli ve b_{R_1} ve b_{R_2} PC tahmin edicilerini ele alalım. Birinci tahmin edici için bir temel bileşenin çıkarıldığını; benzer şekilde ikinci tahmin edici için m_2 temel bileşenin çıkarıldığını kabul edelim. O zaman b_{R_2} nin b_{R_1} den güçlü üstün olması için gerek ve yeter koşul

i) b nin b_{R_1} den güçlü üstün olması

ii) b_{R_2} nin b den güçlü üstün olmasıdır.

Her iki PC tahmin edicisi için çıkarılan temel bileşenler ortak ise o zaman üstünlük koşulu oluşmaz. Genel olarak β nın, çıkarılan temel bileşenlere ortogonal

olduğu varsayımı yapılamaz. Yani $s_3 = 0$ varsayımı yapılamaz.

Guilkey ve Price (1981) deki sonuçlara dayanan ; Price (1982) deki iki PC tahmin edici için üstünlük koşulunda s_3 ün doğru olmadığı kısıtlaması yoktur.

Yukarıdaki örnekler; iki RLS tahmin edicisi karşılaştırıldığında, güçlü üstünlük sonuçlarının uygulamada ne kadar kısıtlı olduğunu gösterir. Genel durumda, eğer her bir tahmin edici diğer tahmin edicinin kısıtlamalarından lineer bağımsız birden fazla kısıtlamaya sahip ise üstünlük sonuçları elde edilemez.

Eğer iki tahmin ediciden biri diğerinden güçlü üstün ise yansız RLS tahmin edicisi (ortak kısıtlamalar olmadığında) veya OLS tahmin edicisinin bu iki tahmin ediciyi ayırdığı gözlenebilir.

Uygulama açısından bakıldığında zayıf üstünlük daha kullanışlı bir kavramdır.

b_{R_2} ve b_{R_1} ele alalım ve $m_3 = 0$ olduğu kabul edelim. O zaman $j=1,2$ için ve $tr(X'XC) = m_2 - m_1$ olur. b_{R_2} nin b_{R_1} üzerindeki zayıf üstünlüğü, Teorem 3.23.2 nin sonucundan yararlanılarak şu şekilde oluşturulabilir.

Sonuç 3.23.4 : b_{R_1} ve b_{R_2} RLS tahmin edicilerini ele alalım. b_{R_2} tahmin edicisinin b_{R_1} den zayıf üstün olması için gerek ve yeter koşul

$$Q_2 - Q_1 \leq m_2 - m_1 \quad (3.23.11)$$

olmasıdır.

Sonuç 3.23.4 , üstünlük için , “ b_{R_2} deki lineer kısıtlamalar en az b_{R_1} deki kadar olmalıdır” şeklinde bir gerekli koşul belirtir. Eğer $m_1 = 0$ ise (3.23.11) ; Wallace (1972) nin b_{R_2} nin b üzerindeki üstünlüğü için ikinci zayıf MSE koşulu olur.

(1.2.1) de $X = (X_1 \ X_2)$ ve $\beta = \begin{pmatrix} \beta_1' & \beta_2' \end{pmatrix}$ yazalım ve X_j nin $p_j \times 1$ tipinde ve $rank(X_j) = p_j$ olan bir matris olduğunu kabul edelim. Z_2 ise X_2 yerine

yazılan (proxy) vekil değişkenler matrisi olsun. Ayrıca $b_2 = \begin{pmatrix} b_{21}' & b_{22}' \end{pmatrix}$ karşılık gelen OLS tahmin edicisi olsun. Bundan başka $U_1 = \left(X_1' X_1 \right)^{-1}$ olmak üzere $b_1 = \begin{pmatrix} b_{11}' & 0' \end{pmatrix} = \left(Y' X_1 U_1 \quad 0' \right)'$ olsun ve b_{11} ise β_1 in çıkarılmış değişkenler (OV) tahmin edicisi olsun. O zaman $F = Z_2' M_1 Z_2 > 0$ ve $M_1 = I - X_1 U_1 X_1'$ olmak üzere

$$-C = \begin{bmatrix} X_1' X_1 & X_1' Z_2 \\ Z_2' X_1 & Z_2' Z_2 \end{bmatrix}^{-1} - \begin{bmatrix} U_1 & 0 \\ 0 & 0 \end{bmatrix}$$

$$= \begin{bmatrix} -U_1 X_1' Z_2 \\ I \end{bmatrix} F^{-1} \begin{pmatrix} -Z_2' X_1 U_1 I \\ \end{pmatrix} \geq 0 \quad (3.23.12)$$

olduğu görülmüştür.

Biri gözlenmemiş (gözlenemeyen) iki bağımsız değişken olduğunda , değişkenlerde hata olması durumunda ; McCallum (1972) ve Wickens (1972) proxy (vekil) değişkenin kullanılmasını önermişlerdir. Çünkü böylelikle ; β_1 skalerinin tahmininde yanlılık, OV tahmin edicisi ile kıyaslandığında azalır.

Ancak o zaman ; (3.23.12) , proxy (vekil) değişkenler belirli (sabit) kabul edildiğinde b_2 nin bileşenlerinin hiçbirinin varyansının , b_1 in karşılık gelen bileşenlerinin varyansından asla küçük olamayacağını belirtir.

Varyansın öneminin üzerinde durulması ve yanlılığın göz ardı edilmesi sonucunda X_2 nin çıkarılması ve proxy (vekil) değişkenlerin kullanılmaması gerekir. Bu ise McCallum ve Wickens'in sonucu ile taban tabana zıt bir sonuçtur. Ayrıca güçlü veya zayıf üstünlüğün karşılaştırmalar için uygun bir temel oluşturduğu da bir gerçektir.

Güçlü üstünlük ele alındığında ; (3.23.12) ; genel durumda çıkarılabilecek tek güçlü koşulun ; “çıkarılmış alternatif” (omission alternative) b_1 in, “vekil alternatif” (proxy alternative) b_2 ye olan üstünlüğü için olduğunu belirtir.

$$d_1 = \begin{bmatrix} U_1 X_1' X_2 \\ -I \end{bmatrix} \beta_2 \text{ ve}$$

$$d_2 = \begin{bmatrix} U_1 X_1' (I - Z_2 F^{-1} Z_2' M_1) X_2 \\ F^{-1} Z_2' M_1 X_2 - I \end{bmatrix} \beta_2 = d_1 - \begin{bmatrix} U_1 X_1' Z_2 \\ -I \end{bmatrix} F^{-1} Z_2' M_1 X_2 \beta_2$$

hesaplayalım ve $N[d_1] \neq N[d_2]$ olacak şekilde Z_2 yi ele alalım. d_1 ve d_2 nin (3.23.12) ye göre karşılaştırılmasıyla aşağıdaki sonuç elde edilir.

Sonuç 3.23.5 : (1.2.1) lineer modelini ele alalım. X_2 nin çıkarıldığı b_1 tahmin edicisi ile ; X_2 yerine Z_2 nin yazıldığı b_2 tahmin edicisinin karşılaştırılmasını ele alalım. Z_2 nin ; b_1 ve b_2 nin yanlılık vektörleri lineer bağımsız olacak şekilde olduğunu kabul edelim. O zaman bu iki tahmin ediciden hiç biri diğerinden güçlü üstün değildir. Bu ayrıca b_{11} ve b_{21} karşılaştırıldığında da doğrudur.

Bu sonuç farklı düşüncelerle Frost (1979) un elde ettiği sonuçlarla uyumludur.

Zayıf üstünlük ile ilgilenildiğinde tahminin mse si (MSEP) nin tanımı, proxy (vekil) değişkenlerin kullanılmasının uyumlu olacak şekilde düzenlenmelidir.

$\tilde{X} = (X_1 \ X_2)$ olmak üzere

$$MSEP(b_2) = E(b_2 - \beta)' \tilde{X}' \tilde{X} (b_2 - \beta)$$

tanımlayalım. Teorem 3.23.2 nin kullanılmasıyla aşağıdaki sonuç elde edilir.

Sonuç 3.23.6 : Sonuç 3.23.5 deki tahmin edicileri ele alalım. O zaman b_1 in b_2 den zayıf üstün olması için gerek ve yeter koşul

$$\sigma^{-2} \beta_2' X_2' M_1 Z_2 F^{-1} Z_2' M_1 X_2 \beta_2 \leq p_2 \quad (3.23.13)$$

olmasıdır.

$p_1 = p_2 = 1$ seçilmesiyle (3.23.13) ifadesi , Ohtani (1981) tarafından verilen üstünlük koşuluna indirgenir.

Şimdi RLS tahmin edicilerinin karşılaştırılmaları ile ilgili sonuç çıkarma problemlerini ele alalım . (3.23.13) ün geçerliliğinin test edilmesi , Toro-Vizcarrondo ve Wallace (1968) in sonuçlarının bir genelleştirilmesidir. $r_3 = R_3\beta$ nın sağlandığını kabul edelim. Bu durumda (3.23.13) altında $\hat{s}_{2,3} = r_2 - R_2b_{R_3}$ olmak üzere $F_2 = \hat{\sigma}_3^{-2} \hat{s}'_{2,3} D_{22,3}^{-1} \hat{s}_{2,3}$ ve $\hat{\sigma}_3^2 = (n - p + m_3)^{-1} (Y - Xb_{R_3})' (Y - Xb_{R_3})$ ün merkezi olmayan $F\left(m_2, n - p + m_3, \frac{1}{2}\right)$ dağılımına sahip olduğu gösterilebilir.

Yeni lineer kısıtlamalar dizisinin ortaya çıkma olasılığını ve tahminin geçerliliğini kanıtlayıp kanıtlamadığının testini ele alalım. O zaman güçlü üstünlük kullanılamaz. Çünkü tahminin kanıtlanması (desteklenmesi) için , eğer kısıtlamaların sayısı artıyorsa , önceki kısıtlamaların tamamının geçerli olması gerekir. Bununla beraber zayıf üstünlük durumunda ; Sonuç 3.23.4 deki varsayımlara ek varsayımlar yapılarak ilginç sonuçlar elde edilebilir. Lineer kısıtlamaların ikinci kümesinin birincinin içinde olduğunu yani $R_2\beta = r_2$ ifadesinin $R_1\beta = r_1$ olduğunu belirttiğini ve $m_1 < m_2$ olduğunu kabul edelim. Dolayısıyla G, $m_1 \times m_2$ tipinde m_1 ranklı bir matris olmak üzere $R_1 = GR_2$ ve $r_1 = Gr_2$ olur ve aşağıdaki teorem oluşturulur.

Teorem 3.23.3 : (1.2.1) lineer modelini ele alalım. $j=1,2$ olmak üzere $R_j\beta = r_j$ lineer kısıtlamalar kümesi olsun. Burada R_j , $m_j \times p$ tipinde matris , $m_1 < m_2 < p$ ve $R_1 = GR_2$ ve $r_1 = Gr_2$ dir.

$\hat{\sigma}^2 = (n - p)^{-1} Y'(I - XUX')Y$; $\hat{s}_1 = G\hat{s}_2$ ve $\hat{s}_2 = r_2 - R_2b$ olmak üzere $\hat{Q}_j = \hat{\sigma}^{-2} \hat{s}'_j (R_j U R_j')^{-1} \hat{s}_j$ oluşturalım. O zaman

$$F_{2/1} = (m_2 - m_1)^{-1} (\hat{Q}_2 - \hat{Q}_1) \quad (3.23.14)$$

ifadesi $m_2 - m_1$ ve n-p serbestlik dereceli ve $(Q_2 - Q_1)/2$ merkezi olmama parametrelili

merkezi olmayan F dağılımına sahiptir.

İspat : $F_{2/1}$ in ilk payını ele alalım ve $j=1,2$ için $A_j = XUR_j' \left(R_j UR_j' \right)^{-1} R_j UX'$

olmak üzere $q_j = \varepsilon' A_j \varepsilon$ tanımlayalım. O zaman $R_1 = GR_2$ olduğundan ,
 $A_1 A_2 = A_2 A_1 = A_1$ olur öyle ki $A_2 - A_1$ idempotenttir ve
 $rank(A_2 - A_1) = tr(A_2 - A_1) = m_2 - m_1$ dir. O zaman $\sigma^{-2}(q_2 - q_1) \approx \chi^2(m_2 - m_1)$ olur
(Rao, 1965).

Buradan $\hat{q} = \sigma^{-2} \left\{ \hat{s}_2' \left(R_2 UR_2' \right)^{-1} \hat{s}_2 - \hat{s}_1' \left(R_1 UR_1' \right)^{-1} \hat{s}_1 \right\}$ ifadesi $m_2 - m_1$

serbestlik dereceli ve $\left(\frac{1}{2} \right) E\hat{q} = \frac{(Q_2 - Q_1)}{2}$ merkezi olmama parametrelili merkezi olmayan χ^2 dağılımına sahiptir (Rao 1965). $(I - XUX')A_j = 0$, $j=1,2$ dir.

Normallik varsayımı yapıldığından bu $\hat{\sigma}^2$ nin paydan bağımsız olduğunu belirtir.

Teorem 3.23.3 den (3.23.11) in test edilebilir bir hipotez olduğu ve geçerli olduğu varsayımı yapıldığında (3.23.14) ün merkezi olmayan

$F\left(m_2 - m_1, n - p, \frac{(m_2 - m_1)}{2} \right)$ dağılımına sahip olduğu bulunur.

$R_j = G_j R_{j+1}$, $r_j = G_j r_{j+1}$, $j = 1, 2, \dots, h-1$ ve $m_1 < m_2 < \dots < m_h \leq p$ olmak üzere $R_j \beta = r_j$ şeklinde bir lineer kısıtlamalar dizisi bulunsun. İlk hipotezi $H_1 : Q_1 \leq m_1$ yani b_{R_1} in b den zayıf üstün olduğu hipotezini ele alalım. Bu hipotez altında

$F_1 = m_1^{-1} \hat{Q}_1$ ifadesi merkezi olmayan $F\left(m_1, n - p, \frac{m_1}{2} \right)$ dağılımına sahiptir (Wallace, 1972).

H_1 kabul edilsin. O zaman H_1 e bağlı olarak $H_{2/1} : Q_2 - Q_1 \leq m_2 - m_1$ hipotezi

test edilebilir. Teorem 3.23.3 , $H_{2/1}$ altında $F_{2/1}$ istatistiğinin $F\left(m_2 - m_1, n - p, \frac{(m_2 - m_1)}{2}\right)$ dağılımına sahip olduğunu belirtir.

Genelde $H_{j+1/j} : Q_{j+1} - Q_j \leq m_{j+1} - m_j$ nin testi ; $H_{j+1/j}$ altında

$$F\left(m_{j+1} - m_j, n - p, \frac{(m_{j+1} - m_j)}{2}\right)$$

dağılımına sahip olan $F_{j+1/j} = (m_{j+1} - m_j)^{-1}(\hat{Q}_{j+1} - \hat{Q}_j)$ nin kullanılmasıyla yapılır.

Testler ilk reddedilen hipoteze ulaşıncaya kadar devam ettirilir.

Lineer hipotezlerin bir dizisinin bu şekilde test edilmesi Anderson (1971, 1962) nin optimal prosedürünün bir genelleştirilmesidir. Kısıtlamaların $(H_j : Q_j = 0)$ geçerliliği yerine sadece bu kısıtlamaların kullanılmasının PMSE yi indirgeyip indirgemediği test edilir. Burada F istatistikleri aynı stokastik $\hat{\sigma}^2$ paydasına sahip olduklarından asimptotik olarak ilişkisizdirler. Bu sebepten test dizisinin kesin boyutu , ayrı ayrı testlerin boyutları kullanılarak bulunur.

3.24. Lineer Regresyonda En İyi Homojen Ve Heterojen Tahmin Edicilerin Karşılaştırılması

Klasik lineer regresyon modelinde bir tahmin edicinin , homojen tahmin ediciler sınıfında optimal olması için MSE sinin minimum olması gerektiği bilinir. Bu tahmin edici ile ilgili problem bilinmeyen parametreler içermesi ve bir tahmin edici olarak güvenle ele alınamamasıdır. Bunu gerçekten işlemsel yapmak için bir çok düzenlenmiş versiyon vardı.

Bu şekildeki tahmin edicilerden biri Farebrother (1975) in oluşturduğu tahmin edicidir. Bu tahmin edicinin geniş örneklem özellikleri Dwivedi ve Srivastava(1978)

tarafından incelenmiştir. Stahlecker ve Trenkler (1985) tahmin amacı ile ; eğer yapılabiliyorsa ; katsayı vektörü ile ilgili bir ön bilginin örneklem bilgisi ile birleştirilmesini ve işlemsel versiyonu ; LS tahmin edicisi ile sabit ön bilgi vektörünün bir konveks kombinasyonu olan bir heterojen tahmin edicinin oluşturulmasını önermişlerdir.

Tracy ve Srivastava (1994); Stahlecker ve Trenkler (1985) in heterojen tahmin edicisinin geniş örneklem özelliklerini incelemişler ; bu tahmin ediciyi Farebrother'ın homojen tahmin edicisi ile karşılatırmışlar ve heterojen tahmin edicinin homojen tahmin ediciden daha iyi olduğu koşulları incelemişlerdir.

Ayrıca Tracy ve Srivastava (1994) ; uygun versiyonu LS tahmin edicisinin bir konveks kombinasyonu olan bir başka heterojen tahmin edici ve bu tahmin edicinin ortalama vektörünü oluşturmuşlar ve geniş örneklem özelliklerini incelemişlerdir.

Tahmin Ediciler Ve Özellikleri

$Y = X\beta + \varepsilon$ (1.2.1) lineer regresyon modelinde β nın BLUE'si OLS tahmin edicisi $\hat{\beta}$ idi. Bununla beraber eğer yansızlık dikkate alınmazsa yanlı tahmin ediciler de kullanılabilir. Theil (1958), Toutenburg (1908) ve Rao (1971) in önerdiği gibi minimum MSE kriterinin kullanılmasıyla yeni tahmin ediciler oluşturulabilir. Bu şekilde oluşturulan minimum MSE tahmin edicisi , β nın en iyi lineer homojen tahmin edicisidir ve

$$\beta^* = \frac{\beta'X'Y}{\sigma^2 - \beta'X'X\beta} \beta = \beta\beta'X'(X\beta\beta'X' + \sigma^2 I_n)^{-1} Y \quad (3.24.1)$$

şeklindedir. β^* ile ilgili problem ; β ve σ^2 bilinmeyen parametreler içerdiğinden bir tahmin edici olmaması ve dolayısıyla pratikte yararlı olmamasıdır. Bunu işlemsel yapmak için Rao (1973) ; β^* a β ve σ^2 nin ön bilgisinin eklenmesini önermiştir.

Farebrother (1975) ise

$$s^2 = \frac{1}{n-p} (Y - X\hat{\beta})' (Y - X\hat{\beta}) \quad (3.24.2)$$

olmak üzere β ve σ^2 yerine bunların OLS tahmin edicileri olan $\hat{\beta}$ ve s^2 nin yazılmasını önermiştir.

Farebrother'dan faydalanılarak uygun optimal β^* tahmin edicisi

$$\hat{\beta}_F = \frac{\hat{\beta}' X' Y}{s^2 + \hat{\beta}' X' X \hat{\beta}} \hat{\beta} \quad (3.24.3)$$

şeklinde bulunur.

$O_p(\dots)$; olasılıktaki dereceyi göstermek üzere Dwivedi ve Srivastava (1978) ; $\hat{\beta}_F$ tahmin edicisinin yanlılık vektörü için $O_p(n^{-1})$ dereceye kadar ; MSE matrisi için $O_p(n^{-2})$ dereceye kadar

$$\begin{aligned} \text{Bias}(\hat{\beta}_F) &= E(\hat{\beta}_F - \beta) = -\frac{\sigma^2}{\beta' X' X \beta} \beta \\ E(\hat{\beta}_F - \beta)(\hat{\beta}_F - \beta)' &= \sigma^2 (X' X)^{-1} + \frac{\sigma^4}{\beta' X' X \beta} \left[\frac{5}{\beta' X' X \beta} \beta \beta' - 2(X' X)^{-1} \right] \end{aligned} \quad (3.24.4)$$

şeklinde geniş örneklem asimptotik yaklaşımları üzerinde çalışmışlardır. Ayrıca bunun OLS tahmin edicisi ile karşılaştırılmasını yapmışlardır.

Stahlecker ve Trenkler (1985) ; katsayı vektörü β ile ilgili ön bilginin ; β nin tahmini için MSE kriterine göre örneklem ön bilgisi ile kullanılmasını önermişlerdir.

Ayrıca

$$\beta_{OPT} = \rho\beta + (1 - \rho)c \quad (3.24.5)$$

tahmin edicisini ileri sürmüşlerdir. Burada

$$\rho = \frac{\beta'X'Y}{\sigma^2 + \beta'X'X\beta} \quad (3.24.6)$$

ve c $n \times 1$ tipinde bir kolon vektörü ; β üzerindeki ön bilgiyi yansıtan değerler vektörüdür. β üzerindeki ön bilgi ; β nin eski değerleri veya β hakkındaki bir tahmindir.

β_{OPT} ; bir minimum MSE tahmin edicisidir ve β nin en iyi heterojen lineer tahmin edicisi olarak adlandırılabilir. Yine ; β ve σ^2 bilinmediğinden β_{OPT} , optimal lmasına rağmen pratikte kullanışlı değildir. Bu sebepten Stahlecker ve Trenkler (1985); Farebrother (1975) in yaptığı gibi ; β ve σ^2 yerine ; β^* da $\hat{\beta}$ ve s^2 yazılmasını önermişlerdir. Böylelikle β_{OPT} in işlemsel versiyonu

$$\hat{\rho} = \frac{\hat{\beta}'X'Y}{s^2 + \hat{\beta}'X'X\hat{\beta}} = \frac{\hat{\beta}'X'X\hat{\beta}}{s^2 + \hat{\beta}'X'X\hat{\beta}}$$

olmak üzere

$$\hat{\beta}_{ST} = \hat{\rho}\hat{\beta} + (1 - \hat{\rho})c \quad (3.24.7)$$

şeklinde elde edilir.

Simulasyon çalışmasıyla $\hat{\beta}_{ST}$ nin davranışı üzerinde çalışmışlardır. Tracy ve Srivastava (1994) ; $\lim_{n \rightarrow \infty} \frac{1}{n} X'X$ in var olduğunu ve simetrik bir p.d. matris olduğunu kabul ederek $\hat{\beta}_{ST}$ tahmin edicisinin örneklem özelliklerini incelemişlerdir.

Teorem 3.24.1 : Eğer hatalar simetrik dağılıma sahip ise , $\hat{\beta}_{ST}$ in $O_p(n^{-1})$ dereceye kadar olan yanlılık vektörü ve $O_p(n^{-2})$ dereceye kadar olan mse matrisi

$$\text{Bias}(\hat{\beta}_{ST}) = E(\hat{\beta}_{ST} - \beta) = -\frac{\sigma^2}{\beta'X'X\beta}(\beta - c)$$

$$\text{MSE}(\hat{\beta}_{ST}) = E(\hat{\beta}_{ST} - \beta)(\hat{\beta}_{ST} - \beta)' \quad (3.24.8)$$

$$= \sigma^2(X'X)^{-1} + \frac{\sigma^4}{\beta'X'X\beta} \left[\frac{1}{\beta'X'X\beta} \{(\beta - c)(\beta - c)' + 2\beta(\beta - c)' + 2(\beta - c)\beta'\} - 2(X'X)^{-1} \right]$$

olarak bulunur.

$\hat{\beta}_{ST}$ tahmin edicisinin avantajı ; Stahlecker ve Trenkler (1985) in de belirttiği gibi ; OLS tahmin edicisi $\hat{\beta}$ nın bir ağırlıklı ortalaması ve ön bilgi vektörü c nin hem örneklem bilgisine hem de ön bilgiye ağırlık kazandırmasıdır. $\hat{\rho} \rightarrow 1$ iken OLS tahmin edicisi $\hat{\beta}$ daha çok önem kazanır. $\hat{\rho} \rightarrow 0$ iken de $\hat{\beta}$, c ön bilgi vektörüne küçülür. Ancak bu çok inandırıcı değildir. Örneğin $\hat{\rho} \rightarrow 0$ iken $\hat{\beta}_{ST}$ tahmin edicisi ön bilgiye bağlıdır. Bu ön bilgi iki tipte olabilir.

- i) β nın bazı geçmiş değerleri
- ii) Geçmiş tecrübelerle dayanılarak β üzerinde yapılan bir tahmin

Her iki durumda da ön bilgi vektörü c , ön bilgiye benzeyebilir veya benzemez.

Eğer sadece bir tahmin değeri ise tamamen yanlış olabilir. Eğer geçmişte karşılaşılan bir değer ise ; örneklem bilgisi ile yansıtılan β nın davranışı ile uyuşmayabilir.

Dolayısıyla ; ön bilginin örneklem bilgisine uygunluğunu test edecek bir prosedür oluşturulmadıkça ; $\hat{\rho} \rightarrow 0$ iken $\hat{\beta}_{ST}$ nin ön bilgiye küçültülmesi yanıltıcı olabilir. Olası bir yol ; c nin belirli bir vektör olarak değil de stokastik bir vektör olarak alınmasıdır. Basit bir seçim ise $\bar{\beta} = \frac{\ell' \hat{\beta}}{\ell' \ell} = \frac{1}{\rho} \sum_{i=1}^{\rho} b_i$ şeklindedir ve burada $\ell = I_{\rho}$ dir.

$\hat{\beta}_{ST}$ nin $\bar{\beta}\ell$ ortalama vektörüne küçültülmesi ; $\hat{\beta}_{ST}$ nin $\hat{\rho} \rightarrow 0$ iken belirli bir c vektörüne küçültülmesinden daha uygundur. Böylece

$$\hat{\beta}_{TS} = \hat{\rho}\beta + (1 - \hat{\rho})\bar{\beta}\ell \quad (3.24.9)$$

tahmin edicisi elde edilir.

Teorem 3.24.2 : $\hat{\beta}_{TS}$ tahmin edicisinin $O_p(n^{-1})$ dereceye kadar olan yanlılık vektörü ve $O_p(n^{-2})$ dereceye kadar olan MSE matrisi , $\bar{\beta} = \frac{\ell'\beta}{\ell'\ell}$ olmak üzere

$$\begin{aligned} \text{Bias}(\hat{\beta}_{TS}) &= E(\hat{\beta}_{TS} - \beta) = -\frac{\sigma^2}{\beta'X'X\beta}(\beta - \bar{\beta}\ell) \\ \text{MSE}(\hat{\beta}_{TS}) &= E(\hat{\beta}_{TS} - \beta)(\hat{\beta}_{TS} - \beta)' \quad (3.24.10) \\ &= \sigma^2(X'X)^{-1} + \frac{\sigma^4}{\beta'X'X\beta} \left[\frac{1}{\beta'X'X\beta} \left\{ (\beta - \bar{\beta}\ell)(\beta - \bar{\beta}\ell)' + 2\beta(\beta - \bar{\beta}\ell)' + 2(\beta - \bar{\beta}\ell)\beta' \right\} \right. \\ &\quad \left. - 2(X'X)^{-1} + \frac{1}{\rho} \left\{ (X'X)^{-1} \ell\ell' + \ell\ell'(X'X)^{-1} \right\} \right] \end{aligned}$$

şeklindedir.

Tahmin Edicilerin Karşılaştırılmaları

Homojen tahmin edici $\hat{\beta}_F$ ve heterojen tahmin ediciler $\hat{\beta}_{ST}$ ve $\hat{\beta}_{TS}$ yanlı tahmin edicilerdir.

$\hat{\beta}_F$ nin yanlılığının işareti , β nin elemanlarının işareti ile zıttır ve $\hat{\beta}_{ST}$ ve $\hat{\beta}_{TS}$ nin yanlılığının işareti $(\beta - \ell)$ ve $(\beta - \bar{\beta}\ell)$ vektörlerinin sapmalarının işaretine bağlıdır.

$\hat{\beta}_{ST}$ nin yanlılığı , ön bilgi vektörü olan c , β ya yakın olduğunda küçük olacaktır. Ayrıca $\hat{\beta}_{ST}$ tahmin edicisi ; c , β ya eşit olduğunda asimptotik olarak yansız olacaktır. $\hat{\beta}_{ST}$ nin yanlılığı ; β nin elemanları , ortalama değerleri olan $\bar{\beta}$ ye yakın olduğunda küçük olacaktır. Bu tahmin edici β nin tüm elemanları eşit olduğunda asimptotik olarak yansız olacaktır.

Bu tahmin edicilerin geçerlilikleri , Q kayıp matrisi p.d. ve simetrik bir matris olmak üzere, ağırlıklı MSE kriterine göre karşılaştırılmak istendiğinde, bu üç tahmin edicinin $O_p(n^{-2})$ dereceye kadar ağırlıklı mse lerinin

$$\begin{aligned} WMSE(\hat{\beta}_F) &= E(\hat{\beta}_F - \beta)'(\hat{\beta}_F - \beta) \\ &= \sigma^2 trQ(X'X)^{-1} + \frac{\sigma^4}{\beta'X'X\beta} \left[\frac{5\beta'Q}{\beta'X'X\beta} - 2trQ(X'X)^{-1} \right] \end{aligned} \quad (3.24.11)$$

$$\begin{aligned} WMSE(\hat{\beta}_{ST}) &= E(\hat{\beta}_{ST} - \beta)'(\hat{\beta}_{ST} - \beta) \\ &= \sigma^2 trQ(X'X)^{-1} + \frac{\sigma^4}{\beta'X'X\beta} \left[\frac{\{(\beta - c)'Q(\beta - c) + 4\beta'Q(\beta - c)\}}{\beta'X'X\beta} - 2trQ(X'X)^{-1} \right] \end{aligned} \quad (3.24.12)$$

$$\begin{aligned} WMSE(\hat{\beta}_{TS}) &= E(\hat{\beta}_{TS} - \beta)'Q(\hat{\beta}_{TS} - \beta) \\ &= \sigma^2 trQ(X'X)^{-1} + \frac{\sigma^4}{\beta'X'X\beta} \left[\frac{(\beta - \bar{\beta}\ell)'Q(\beta - \bar{\beta}\ell) + 4\beta'Q(\beta - \bar{\beta}\ell)}{\beta'X'X\beta} \right. \\ &\quad \left. - 2trQ(X'X)^{-1} + \frac{2\ell'Q(X'X)^{-1}\ell}{p} \right] \end{aligned} \quad (3.24.13)$$

olduğu görülür. $\hat{\beta}_{ST}$ ile $\hat{\beta}_F$ karşılaştırılırken

$$WMSE(\hat{\beta}_F) - WMSE(\hat{\beta}_{ST}) = \frac{\sigma^4}{\beta'X'X\beta} (6\beta - c)' Qc \quad (3.24.14)$$

bulunur ve bu ifade $(6\beta - c)' Qc$ pozitif olduğu sürece pozitiftir.

Dolayısıyla heterojen $\hat{\beta}_{ST}$ tahmin edicisi ;

$$(6\beta - c)Qc > 0 \quad (3.24.15)$$

olduğu sürece homojen $\hat{\beta}_F$ tahmin edicisinden daha iyidir.

$\hat{\beta}_{TS}$ ve $\hat{\beta}_F$ karşılaştırılmak istendiğinde

$$WMSE(\hat{\beta}_F) - WMSE(\hat{\beta}_{TS}) = \frac{\sigma^4}{\beta'X'X\beta} \left[\frac{(6\beta - \bar{\beta}l)Ql\bar{\beta}}{\beta'X'X\beta} - \frac{2l'Q(X'X)^{-1}l}{l'l} \right] \quad (3.24.16)$$

elde edilir.

$\hat{\beta}_{TS}$ nin daha iyi veya daha kötü olma koşullarının elde edilmesi zordur.

$c = \bar{\beta}l$ alındığında, iki heterojen tahmin edici arasında ilginç bir karşılaştırma yapılabilir. Bu, β nin tüm p elemanları için tahminin aynı seçilmesi anlamına gelir.

Dolayısıyla

$$WMSE(\hat{\beta}_{TS}) - WMSE(\hat{\beta}_{ST}) = \frac{2\sigma^4 l'Q(X'X)^{-1}l}{\beta'X'X\beta l'l} > 0 \quad (3.24.17)$$

elde edilir. Bu ise $\bar{\beta}$ nin bir \bar{b} tahminin heterojen tahmin edici formatında kullanılmasıyla geçerlilikte bir kayıp olduğunu belirtir.

3.25. Bir Regresyon Modelinde Stokastik Olarak Kısıtlanmış Lineer Tahmin Edicilerinin Karşılaştırılması

$Y = X\beta + \varepsilon$ (1.2.1) lineer regresyon modelini ele alalım.

$$r_j = R_j\beta + v_j, \text{Cov}(v_j) = V_j, j = 1, 2 \quad (3.25.1)$$

formundaki iki stokastik kısıtlama kümesi altında β nın tahmin edicilerini inceleyelim. Burada r_j , $m_j \times 1$ tipinde bilinen bir vektör ; R_j , m_j ranklı $m_j \times p$ tipinde bilinen bir matris ; v_j bilinmeyen rastgele hata vektörü ve V_j ise $m_j \times m_j$ tipinde bilinen bir n.n.d. matristir. (Ayrıca $\text{Cov}(v_j) = 0$ $j=1,2$ olduğunu kabul edelim.)

(1.2.1) deki modelden örneklem bilgisi ve (3.25.1) deki stokastik kısıtlama, genişletilmiş bir lineer model olarak yazılabilir. Bu ilk defa Theil ve Goldberger (1961) tarafından yapılmıştır ve bu genişletilmiş modelden elde edilen OLS tahmin edicisini bir “mixed tahmin edicisi” olarak adlandırmışlardır. Mixed tahmin edicileri teorisi ayrıntılı olarak Toutenburg (1982) tarafından ortaya konulmuştur.

(3.25.1) formundaki kısıtlamaya sahip (1.2.1) modelindeki β nın minimum varyanslı lineer tahmin edicisi

$$\hat{\beta}_j = \hat{\beta} + S^{-1}R_j'U_j^{-1}(r_j - R_j\hat{\beta}), j = 1, 2 \quad (3.25.2)$$

şeklinde ifade edilebilir. Burada $S = X'X$ ve $\hat{\beta} = S^{-1}X'Y$ (1.2.1) deki model için OLS tahmin edicisidir.

Eğer $E(r_j) = R_j\beta$ ise $\hat{\beta}_j$ tahmin edicisi yansızdır. $\hat{\beta}_j$ nin varyans kovaryans matrisi

$$\text{Cov}(\hat{\beta}_j) = \sigma^2 \left(S^{-1} - S^{-1}R_j'U_j^{-1}R_jS^{-1} \right) \quad (3.25.3)$$

formundadır.

Baksalary (1984) $\hat{\beta}_1$ ve $\hat{\beta}_2$ tahmin edicilerini $R_1 = R_2$ özel durumunda

karşılaştırmıştır. Srivastava ve Agnihotri (1980) ise $\hat{\beta}_1$ ve $\hat{\beta}_2$ tahmin edicilerini varyans kovaryans matrisine göre karşılaştırmıştır.

Herhangi bir ranka sahip V_j için (3.25.2) geçerlidir oysa Theil ve Goldberger'in orijinal formülü sadece tam ranklı V_j için geçerlidir. Bu sebepten (3.25.2) formülü, Theil'in formülünden daha kullanışlıdır. Çünkü V_j singüler olabilir ; kısıtlamalar kısmen stokastik, kısmen non stokastik olabilir. $V_j = 0$ özel durumunda non stokastik olarak kısıtlanmış LS tahmin edicisi elde edilir.

Sonuçlar

$\lambda_1(\dots)$ bir matrisin en büyük özdeğerini ve A^+ ise A matrisin Moore-Penrose tersi olsun

Teorem 3.25.1 : $\hat{\beta}_1$ ve $\hat{\beta}_2$ tahmin edicileri için $Cov(\hat{\beta}_1) - Cov(\hat{\beta}_2)$ nin n.n.d. olması için gerek ve yeter koşul

$$i) R_1 = PR_2$$

$$ii) \lambda_1 \left[\left(R_2' U_2^{-1} R_2 \right)^+ R_1 U_1^{-1} R_1 \right] \leq 1$$

olacak şekilde $m_1 \times m_2$ tipinde bir P matrisinin olmasıdır.

İspat : (3.25.3) den

$$S^{-1} \left(R_2' U_2^{-1} R_2 - R_1' U_1^{-1} R_1 \right) S^{-1} \quad (3.25.4)$$

nnd olduğundan $Cov(\hat{\beta}_1) - Cov(\hat{\beta}_2)$ n.n.d. dir.

S tam ranklı olduğundan (3.25.4) in sağlanması için gerek ve yeter koşul

$$R_2' U_2^{-1} R_2 - R_1' U_1^{-1} R_1 \quad (3.25.5)$$

ifadesinin n.n.d. olmasıdır.

Tanımdan dolayı $R_j' U_j^{-1} R_j$ matrislerinin n.n.d. olduğu açıktır.

Stenplok (1985) in 1. teoreminden (3.25.5) varsayımı ; $M(R_j' U_j^{-1} R_j)$,

$R_j' U_j^{-1} R_j$ nin sütun uzayı olmak üzere

$$M(R_1' U_1^{-1} R_1) \subset M(R_2' U_2^{-1} R_2) \quad (3.25.6)$$

ve

$$\lambda_1 \left[(R_2' U_2^{-1} R_2)^+ R_1' U_1^{-1} R_1 \right] \leq 1$$

olduğunda sağlanır. Ancak $M(R_j' U_j^{-1} R_j) = M(R_j')$ dir ve bu ifade

$M(R_1') \subset M(R_2')$ olduğunu belirtir.

Tanımdan $M(R_1') \subset M(R_2')$ dir ve bu teoremin (i) kısmına eşdeğerdir.

Trenkler (1986) (i) nin ; $Cov(\hat{\beta}_1) - Cov(\hat{\beta}_2)$ n.n.d . olduğu varsayımından elde edildiğini göstermiştir. Teoremden birçok önemli sonuç elde edilir.

İlk olarak β nin non stokastik olarak kısıtlanmış tahminleri ile stokastik olarak kısıtlanmış tahmin edicisinin karşılaştırılmasını inceleyelim.

Sonuç 3.25.1 : $\hat{\beta}_2$; $r_2 = R_2 \beta$ kısıtlamasına göre β nin RLS tahmin edicisi ve

$\hat{\beta}_1$ ise stokastik olarak kısıtlanmış bir tahmin edici olsun . $\hat{\beta}_1$ ve $\hat{\beta}_2$ için $Cov(\hat{\beta}_1) - Cov(\hat{\beta}_2)$ nin n.n.d. olması için gerek ve yeter koşul $R_1 = PR_2$ olacak şekilde bir P matrisinin olmasıdır.

İspat : $V_2 = 0$ olduğunda teoremin (ii) bölümünün sağlandığını gösterilmesi yeterlidir. Bu durumda; teoremden $M(R_1') \subset M(R_2')$ ve $R_2'(R_2S^{-1}R_2)^{-1}R_2S^{-1}$; $M(R_2')$ üzerine bir projektör olduğundan

$$U_2 = R_2S^{-1}R_2' \text{ ve } R_2'(R_2S^{-1}R_2)^{-1}R_2S^{-1}R_1' = R_1' \quad (3.25.7)$$

olur. Yazım kolaylığı için $A_j = R_jS^{-1}R_j'$ alalım . (3.25.7) den

$$S^{-1}R_1' = \left(R_2'A_2^{-1}R_2\right)^+ R_1'$$

olur ve buradan da

$$S^{-1}R_1'U_1^{-1}R_1 = \left(R_2'A_2^{-1}R_2\right)^- R_1'U_1^{-1}R_1 \quad (3.25.8)$$

olur . $U^{-\frac{1}{2}} ; \left(U^{-\frac{1}{2}}\right)^2 = U^{-1}$ olacak şekilde olmak üzere

$$\lambda_1\left(S^{-1}R_1'U_1^{-1}R_1\right) = \lambda_1\left(A_1U_1^{-1}\right)$$

ve

$$\lambda_1\left(U_1U_1^{-1}\right) = \lambda_1\left(U_1^{-\frac{1}{2}}A_1U_1^{-\frac{1}{2}} + U_1^{-\frac{1}{2}}\sigma^{-2}V_1U_1^{-\frac{1}{2}}\right)$$

olduğuna dikkat edilmelidir. $U_1^{-\frac{1}{2}}A_1U_1^{-\frac{1}{2}}$ ve $U_1^{-\frac{1}{2}}V_1U_1^{-\frac{1}{2}}$ n.n.d. olduklarından

$$\lambda_1(U_1 U_1^{-1}) = 1 \geq \lambda_1\left(U_1^{-\frac{1}{2}} A_1 U_1^{-\frac{1}{2}}\right) = \lambda_1(A_1 U_1^{-1})$$

olur. Böylelikle ispat tamamlanır.

Sonuç 3.25.1 ,stokhastik olarak kısıtlanmış tüm $\hat{\beta}$ tahmin edicileri için geçerlidir ve $V_1 = 0$ özel durumunda da geçerlidir.

Sonuç (3.25.2) : Nonstokhastik olarak kısıtlanmış LS tahmin edicileri $\hat{\beta}_1$ ve $\hat{\beta}_2$ için $(V_1 = V_2 = 0)$ $Cov(\hat{\beta}_1) - Cov(\hat{\beta}_2)$ nin n.n.d. olması için gerek ve yeter koşul $R_1 = PR_2$ olacak şekilde bir P matrisini bulunmasıdır.

Sonuç 3.25.2 , Trenkler (1987) tarafından farklı bir teknikle ispatlanmıştır. Eğer $R_1 = R_2$ ise

$$V_1 - V_2 \quad (3.25.9)$$

n.n.d. olduğunda $Cov(\hat{\beta}_1) - Cov(\hat{\beta}_2)$ n.n.d. olur.

Ayrıca bu sonuç Baksalary (1984) tarafından elementer bir yolla ispatlanmıştır.

(3.25.4) ün sağlanması için gerek ve yeter koşul

$$U_2^{-1} - U_1^{-1} \quad (3.25.10)$$

in n.n.d. olmasıdır. Ancak $U_1 - U_2$ n.n.d. olduğunda (3.25.10) sağlanır ve bu da (3.25.9) a eşdeğerdir.

3.26. Linear Modellerde Mixed Ve Minimax Tahmin Edicilerinin Karşılaştırılması

Regresyon parametreleri ile ilgili ön bilginin lineer modellere dahil edilmesinde kullanılan iki temel yol vardır. Model parametreleri tahmin edildiğinde ; lineer ön bilgi,

örneklem gözlemleri ile aynı durumda olan hayali (uydurma) gözlemler olarak formüle edilip modele dahil edilebilir. Bu yaklaşım ilk olarak Durbin (1953) tarafından önerildi. Theil ve Goldberger (1961), bu yaklaşım üzerinde çalıştı ; bugünkü genel formuna getirdi ve elde edilen tahmin ediciyi mixed tahmin edici olarak adlandırdı.

Lineer olmayan ön bilgi de lineer modellere dahil edilebilir. Regresyon katsayılarının vektörünün, katsayı uzayında bilinen bir elipsoidde bulunduğu veya regresyon katsayılarının vektörünün , alt uzaydaki bilinen bir elipsoiddeki izdüşümü ile ilgili bir ön bilgi olarak kabul ediliyorsa, bu bilgi , tahmin kullanılabilir. Kayıp matrisinin rankı 1 olduğunda ön kısıtlamalar altında karesel riskin maksimumunun minimize edilmesiyle minimax tahmin edicisi elde edilir.

Aynı tahmin edici ayrıca; katsayı vektörünün yukarıda bahsedilen elipsoid üzerinde veya içinde olduğu ile ilgili kısıtlamaya göre hata karelerinin toplamını da minimize eder. Bu varsayım ; ridge regresyonda , katsayı uzayında elipsoidin küreye dönüşmesi durumunda Meeter (1966) tarafından ortaya konulan sonuca dayanır.

Mixed ve minimax tahmin edicilerinden bazen aynı tahminler elde edilir. Ancak stokastik özellikleri farklıdır.

$Y = X\beta + \varepsilon$ (1.2.1) lineer modelini ele alalım.

$$\beta \in B = \left\{ \beta : \beta'R'R\beta \leq \frac{\sigma^2}{c} \right\}, c > 0 \quad (3.26.1)$$

olsun. Lineer ön bilginin

$$r = R\beta + \phi, \phi \approx N \left[0, \left(\frac{\sigma^2}{k} \right) I \right] \quad (3.26.2)$$

şeklinde olduğunu kabul edelim. Burada r ve ϕ $m \times 1$ tipinde stokastik vektörler, ϕ gözlenemeyen hata vektörü , r ise her zaman sıfır olarak gözlenen uydurma bir vektördür. Bundan başka $rank(R) = m$ olduğunu kabul edelim. Böyle bir durum ön bilgi

“ $R\beta$ sifıra yakın ” ifadesi stokhastik lineer formda verildiğinde ortaya çıkar. $R\beta \neq 0$ in olasılığı (3.26.1) e göre 1 dir. O zaman (1.2.1) ve (3.26.2) ye dayanan mixed tahmin edicisi

$$b_R(k) = (X'X + kR'R)^{-1} (X'Y + kR'r) \quad (3.26.3)$$

şeklindedir. Mixed tahmin edicisi yanlıdır. Çünkü (3.26.2) yanlı bir bilgidir ve gerçekte

$$\phi \sim N \left(-R\beta, \left(\frac{\sigma^2}{k} \right) I \right) \text{ dir.}$$

Eğer ön bilgi (3.26.1) şeklinde ise o zaman a mxl tipinde bir vektör ve $A = aa'$ kayıp matrisi olmak üzere, karesel risk altında, A kayıp matrisli minimax tahmin edicisi

$$b_R^*(c) = (X'X + cR'R)^{-1} X'Y \quad (3.26.4)$$

şeklindedir.

$R=I$ olduğunda (3.26.4) deki ifadeden Hoerl ve Kennard'ın ridge tahmin edicisi elde edilir. r nin sıfır olarak gözlemlendiği kabul edilirse (3.26.3) ve (3.26.4) den aynı tahmin ediciler elde edilir.

Ancak r nin stokhastik yapısından dolayı elde edilen tahmin edicilerin stokhastik özellikleri aynı değildir. r ve ϕ stokhastik ve β nonstokhastik olduğundan ; Taylor (1974) ün yaptığı gibi (3.26.3) de $r=0$ yazılarak (3.26.4) in elde edilmesi mantıklı değildir. Bu son işlem ancak (3.26.2) de $k \rightarrow \infty$ olursa ve limitte $\phi = 0$ olursa mümkündür . (3.26.2) de $k \rightarrow \infty$ ve (3.26.1) de $c \rightarrow \infty$ ise iki tahmin edici aynı olur ve limit alınarak elde edilen tahmin edici RLS tahmin edicisi olur.

Mixed ve minimax tahmin edicilerin birbirleriyle karşılaştırılması için bazı kuralların olması gereklidir. Karşılaştırmalar karesel kayıp fonksiyonları terimleri cinsinden yapılacaktır.

Tahmin ediciler

$$R(b_j, \beta, A) = E(b_j - \beta)' A (b_j - \beta)$$

karesel risk fonksiyonu kullanılarak karşılaştırılacaktır. Her $A \geq 0$ için b_1 in b_2 den daha iyi olması için gerek ve yeter koşul $R(b_1, \beta, A) \leq R(b_2, \beta, A)$ olmasıdır.

Dikkat edilmesi gereken husus $rank(A) > 1$ olduğunda (3.26.4) bir minimax tahmin edicisi değildir (Kuks ve Olman (1872), Lauter (1975) ve Hoffmann (1979)).

$R\beta \neq 0$ ve r sabit şekilde sıfır olarak gözlemlendiğinde, (3.26.2) yanlı olacak şekilde (3.26.1) in sağlandığı kabul edilmiştir. Bununla beraber (3.26.1) in sağlanması, risk karşılaştırması için gerekli değildir. Bu varsayım minimax tahmin edicisinin oluşturulması için gereklidir.

Minimax Tahmin Edicinin, Ols Tahmin Edici Ve Mixed Tahmin Edici Üzerindeki Üstünlüğü

Minimax ve mixed tahmin edicilerinin karşılaştırılmaları için bu tahmin edicilerin risklerinin bulunması gerekmektedir.

$s = R\beta$ olmak üzere mixed tahmin edicisinin riski ;

$$U = (X'X)^{-1} \text{ ve } S_k = (k^{-1}I + RUR')^{-1}$$

olmak üzere

$$R(b_R(k), \beta, A) = trA[\sigma^2 U - \sigma^2 UR'S_k RU + UR'S_k SS'S_k RU] \quad (3.26.5)$$

şeklindedir.

Minimax tahmin edicisinin riski benzer şekilde

$$S_c = (c^{-1}I + RUR')^{-1}$$

olmak üzere

$$R(b_R^*(c), \beta, A) = \text{tr}A[\sigma^2 U - \sigma^2 UR'S_c(2c^{-1}I + RUR')S_c RU + UR'S_c s s' S_c RU] \quad (3.26.6)$$

şeklinde elde edilir.

(3.26.5) ve (3.26.6) nın karşılaştırılmasından önce ; bunların OLS tahmin edicisi β nın riski ile karşılaştırılmalarını ele alalım.

(3.26.3) deki mixed tahmin edicisinin tüm $A \geq 0$ için OLS tahmin edicisinden daha küçük riske sahip olması için gerek ve yeter koşul

$$\left(\frac{1}{\sigma^2}\right) s' S_k s \leq 1 \quad (3.26.7)$$

olmasıdır. (Yancey ve ark.(1974)).

Minimax tahmin edicisinin tüm $A \geq 0$ için OLS tahmin edicisinden daha iyi olması için gerek ve yeter koşul

$$\left(\frac{1}{\sigma^2}\right) s'(2c^{-1}I + RUR')^{-1} s \leq 1 \quad (3.26.8)$$

olmasıdır (Swamy ve Mehta (1977) ve Farebrother, 1978).

$c=2k$ seçilmesiyle (3.26.7) ve (3.26.8) aynı olur. Eğer belirli bir k değeri için (3.26.7) sağlanıyorsa ; o zaman minimax tahmin edicisinin OLS tahmin edicisinden daha iyi olması için yeterli koşul $0 < c \leq 2k$ olmasıdır.

Yansız Ön Bilgi Olduğunda Karşılaştırma

Daha önceki tanıma göre minimax tahmin edicisinin mixed tahmin edicisinden daha iyi olması için gerek ve yeter koşul tüm $A \geq 0$ için $R(b_R(k), \beta, A) - R(b_R^*(c), \beta, A)$ farkının negatif olmamasıdır.

$RUR'S_c = I - c^{-1}S_c$ olduğundan $B = S_c - S_k + c^{-1}S_c^2$ olmak üzere

$$\begin{aligned} & R(b_R(k), \beta, A) - R(b_R^*(c), \beta, A) = \\ & \quad \text{tr}AUR' \left[\sigma^2 (2c^{-1}S_c^2 + S_cRUR'S_c - S_k) - (S_c ss'S_c - S_k ss'S_k) \right] RU \\ & = \text{tr}AUR' \left[\sigma^2 B - (S_c ss'S_c - S_k ss'S_k) \right] RU \quad (3.26.9) \end{aligned}$$

elde edilir.

(3.26.9) daki farkın tüm $A \geq 0$ için negatif olmayan olması için gerek ve yeter koşul parantez içindeki matrisin negatif olmayan olmasıdır (Theobald (1974)).

İlk olarak $s=0$ olan özel durumu ele alalım. Bu durumda tüm $A \geq 0$ için (3.26.9) un negatif olmayan olması için gerek ve yeter koşul $B \geq 0$ olmasıdır. S_c ve S_k matrisleri aynı özvektörlere sahiptir. C_R , sütunları RUR' nün özvektörleri olan bir matris ve Λ_R , köşegen elemanları RUR' nün özdeğerleri olan köşegen matris olmak üzere; $C_R \Lambda_R C_R' = RUR'$ şeklindeki spektral ayrışımının kullanılmasıyla

$$B = C_R \left[(c^{-1}I + \Lambda_R)^{-1} - (k^{-1}I + \Lambda_R)^{-1} + c^{-1}(c^{-1}I + \Lambda_R)^{-2} \right] C_R' \quad (3.26.10)$$

elde edilir.

Eğer (3.26.10) negatif olmayan ise parantez içindeki matrislerin tüm özdeğerlerinin negatif olmayan olması gerekir. Yani $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m > 0$, RUR' nün özdeğerleri olmak üzere

$$(c^{-1} + \Lambda_j)^{-1} - (k^{-1} + \Lambda_j)^{-1} + c^{-1}(c^{-1} + \Lambda_j)^{-2} \geq 0, \quad j = 1, 2, \dots, m \quad (3.26.11)$$

olması gerekir. Bu ise $c^{-2} - 2c^{-1}k^{-1} - k^{-1}\lambda_j \leq 0$, $j = 1, 2, \dots, m$ şeklinde yazılabilir.

Karşılık gelen eşitliğin c^{-1} e göre çözülmesiyle

$$c^{-1} = k^{-1} \left(1 \pm (1 + k\lambda_j)^{\frac{1}{2}} \right)$$

elde edilir. c negatif olmayan olması gerektiğinde köklerden biri daima negatif olacağından

$$c \geq k \left(1 + (1 + k\lambda_j)^{\frac{1}{2}} \right)^{-1}, \quad j = 1, 2, \dots, m$$

elde edilir.

O zaman $B \geq 0$ olması için gerek ve yeter koşul

$$c \geq k \left(1 + (1 + k\lambda_m)^{\frac{1}{2}} \right)^{-1} \quad (3.26.12)$$

olmasıdır.

(3.26.12) den tüm X ve R için $c \geq \frac{k}{2}$ olduğunda $B \geq 0$ olduğunu görürüz..

Genel Durum

Bu noktada genel duruma dönelim. Tüm $A \geq 0$ için (3.26.9) daki farkın negatif olmayan olması için gerek ve yeter koşul

$$\sigma^2 B - (S_c ss' S_c - S_k ss' S_k) \geq 0 \quad (3.26.13)$$

olmasıdır.

Lemma 3.26.1 : h ve g $m \times 1$ tipinde iki vektör olsun. O zaman $hh' - gg'$ tanımsızdır ve sıfırdan farklı özdeğerleri

$$\lambda = -2^{-1} (g'g - h'h) \pm \left[4^{-1} (g'g - h'h)^2 + g'gh'h - (g'h)^2 \right]^{\frac{1}{2}}$$

şeklindedir.

Lemma (3.26.1) den , $c \neq k$ için eğer $S_c ss' S_c - S_k ss' S_k$ skaler değilse daima

tanımsızdır. Dolayısıyla (3.26.13) ün sağlanması için gerek ve yeter koşul $B > 0$ olmasıdır.

Bu koşulun sağlandığı kabul edilerek T tekil olmayan matris olmak üzere $B = TT'$ şeklinde tanımlanırsa (3.26.13) koşulu

$$\sigma^2 I - (T^{-1}S_c s s' S_c' T'^{-1} - T^{-1}S_k s s' S_k' T'^{-1}) \geq 0$$

şeklinde yeniden yazılabilir.

Bu eşitsizliğin sağlanması için gerek ve yeter koşul $T^{-1}S_c s s' S_c' T'^{-1} - T^{-1}S_k s s' S_k' T'^{-1}$ in özdeğerlerinin σ^2 den büyük olmasıdır.

$h = T^{-1}S_k s$ ve $g = T^{-1}S_c s$ yazılarak lemmanın uygulanmasıyla

$$\left(\frac{1}{\sigma^2} \right) \left\{ 2^{-1} (s' S_k B^{-1} S_k s - s' S_c B^{-1} S_c s) + \left[4^{-1} (s' S_k B^{-1} S_k s - s' S_c B^{-1} S_c s)^2 + s' S_c B^{-1} S_c s s' S_k B^{-1} S_k s - (s' S_c B^{-1} S_k s)^2 \right]^{1/2} \right\} \leq 1$$

(3.26.14)

şeklinde gerek ve yeter koşul elde edilir.

Daha önce belirtildiği gibi $B > 0$ olduğu kabul edilmiştir. Dolayısıyla (3.26.14) sadece (3.26.12) nin geçerli olduğu koşulda sağlar.

Bu genel sonuçtan özel sonuçlar elde etmek zordur. Bununla beraber $c=k$ olduğunda işlemler kolaylaşır. O zaman tüm $A \geq 0$ için X , R ve s den bağımsız olarak (3.26.9) ifadesinden

$$R(b_R(k), \beta, A) - R(b_R^*(k), \beta, A) = \text{tr} A U R' (\sigma^2 k^{-1} S_c^2) R U \geq 0$$

yazılabilir. Böylelikle, tüm s ler için $\beta' R' R \beta = \frac{\sigma^2}{k}$ tipindeki ön bilginin örneklem

bilgisiyle birleştirilmesiyle ; r sıfır olarak gözlenmiş ve $\phi \sim N\left(0, \left(\frac{\sigma^2}{k}\right)I\right)$ olmak üzere

$r = R\beta + \phi$ şeklindeki yanlı bilgiden elde edilen tahminlerden daha iyi tahminler elde edilir. Bu durum ayrıca (3.26.3) ün yansız olduğu $R\beta = 0$ durumunda da gözlenebilir.

(3.26.14) oldukça karmaşık bir ifade olmasına rağmen ; tüm s ler için geçerli olursa o zaman

$$\left(\frac{1}{2\sigma^2}\right)\left[s'S_k B^{-1}S_k s - s'S_c B^{-1}S_c s + \left|s'S_k B^{-1}S_k s - s'S_c B^{-1}S_c s\right|\right] \leq 1 \quad (3.26.15)$$

ifadesi tüm s ler için doğru olur. Çünkü $s'S_c B^{-1}S_c s s'S_k B^{-1}S_k s - (s'S_c B^{-1}S_c s)^2 \geq 0$ dır.

Ancak o zaman $k = c$ durumu dışında ; tüm s ler için (3.26.15) in geçerli olması için gerek ve yeter koşul tüm s ler için

$$s'S_k B^{-1}S_k s - s'S_c B^{-1}S_c s < 0 \quad (3.26.16)$$

olmasıdır. (3.26.10) ve (3.26.11) nin kullanılmasıyla (3.26.16) eşitsizliği $j = 1, 2, \dots, m$ için

$$\left[\left(c^{-1} + \lambda_j\right)^{-1} - \left(k^{-1} + \lambda_j\right)^{-1} + c^{-1}\left(c^{-1} + \lambda_j\right)^{-2}\right]^{-1} \left[\left(k^{-1} + \lambda_j\right)^{-2} - \left(c^{-1} + \lambda_j\right)^{-2}\right] < 0 \quad (3.26.17)$$

şeklinde yazılabilir. $B > 0$ olduğu kabul edildiğinden , $j = 1, 2, \dots, m$ olmak üzere (3.26.17) nin tüm λ_j ler için sağlanması için gerek ve yeter koşul parantezlerin içindeki ikinci terimin pozitif olmasıdır ki bu $c > k$ olduğunu belirtir. Dolayısıyla tüm s ler için (3.26.16) ; (3.26.14) ün sağlanması için gerekli koşuldur fakat yeterli koşul değildir. Ancak bu ; tüm s ler için sağlanmayan ; (3.26.14) ün sağlanması için gerek ve yeter koşuldandaha kullanışlıdır.

Bir Bağımsız Değişkenli Lineer Model

(1.2.1) deki modelde bir bağımsız değişken olduğu kabul edilirse koşullar daha netlik kazanacaktır. Bu koşulda özel olarak $s'S_c B^{-1} S_c s s' S_k B^{-1} S_k s = (s'S_c B^{-1} S_k s)^2$ olur ve dolayısıyla, (3.26.14), (3.26.15) deki forma dönüşür. Buna göre (3.26.15) in sağlanması için gerek ve yeter koşul $c > k$ olmasıdır. Bu koşul ayrıca tek bağımsız değişkenli durumda, tüm s ler için, minimax tahmin edicisinin mixed tahmin edicisinden daha iyi olması için gerek ve yeter koşuldur. Bu üstünlük $c=k$ için de oluşturulabilir. Şöyle ki $u = (X'X)^{-1}$ olmak üzere $k \left(1 + (1 + ku)^{\frac{1}{2}}\right)^{-1} < c < k$ için; $c \geq k > 0$ iken (3.26.15) tüm s ler için geçerli olduğunda; (3.26.15); $b_R^*(c)$ nin $b_R(k)$ dan daha iyi olması için gerek ve yeter koşuldur.

$c=k$ olduğunda, tüm s ler için minimax tahmin edicisinin, mixed tahmin edicisine olan üstünlüğü açık iken; genel durumda yani birden fazla bağımsız değişken olduğunda $c > k$ koşulu, minimax tahmin edicisinin mixed tahmin edicisinden daha iyi olması için gerekli koşuldur ancak yeterli koşul değildir.

3.27. Lineer Regresyonda Admissible Ve Minimax - Lineer Tahmin Ediciler Arasındaki İlişki

$p \times n$ tipindeki reel matrislerin kümesi $M_{p,n}$ ile, $p \times p$ tipindeki reel matrisler ise M_p ile gösterilsin. $L \in M_{p,n}$ matrisinin transpozunu L' dır. L nin ve L' nün sütunları tarafından gerilen lineer alt uzay $R(L)$ ile gösterilir. L^{-1} ve L^- ise L nin tersi ve genelleştirilmiş tersidir. $L \in M_n$ matrisinin izi $tr(L)$ ile gösterilir.

$$M = \{Y, X\beta, \sigma^2 I\} \quad (3.27.1)$$

lineer regresyon modelindeki β parametresinin lineer tahmininin ele alalım. Bu modelde β nın BLUE si $S = X'X$ olmak üzere $\hat{\beta} = S^{-1}X'Y$ şeklindedir.

β nın homojen olmayan tüm lineer tahmin edicilerinin ailesi olan \mathcal{L} , yani

$$\mathcal{L} = \{A\hat{\beta} + d : A \in M_p, d \in M_{p,1}\} \quad (3.27.2)$$

ailesini ele alalım .

$L \in M_{p,n}$ ve $d \in M_{p,1}$ olmak üzere $LY + d$ şeklindeki bir lineer tahmin edici, eğer $R(L') \subseteq R(X)$ ise $A\hat{\beta} + d$ şeklinde yazılabilir.

Varyansda belli bir azalma sağlamak için ufak bir yanlılığa izin vererek , parametre uzayının belli bir bölgesinde β nın daha kesin (geçerli) bir tahmin edicisini elde etmek için bazı alternatif prosedürler uygulanır.

Yanlı tahmindeki problem , $\hat{\beta}$ için uygun bir $A \in M_p$ çarpan matrisinin bulunmasıdır. (3.27.2) deki ifade , lineer tahmini bu açıdan sıkıntıya sokar . Judge ve Yancey (1986) , Toutenburg (1982) ve Trenkler (1981) in bu konu üzerinde çalışmaları vardır.

Tahmin edicilerin geçerliliği incelenirken mse leri veya MSE matrisleri incelenir.

Verilen bir $\tilde{\beta} = A\hat{\beta} + d \in \mathcal{L}$ için ağırlıklı mse (WMSE) ; $C \in M_p$ p.d. olmak üzere

$$R(\beta; \tilde{\beta}, C) = E \left[(\tilde{\beta} - \beta)' C (\tilde{\beta} - \beta) \right] \quad (3.27.3)$$

$$= \sigma^2 tr(CAS^{-1}A') + [(I - A)\beta + d]' C [(I - A)\beta + d]$$

dir. Varyans matrisi kriterine göre en az $\hat{\beta}$ kadar iyi olan lineer tahmin edicilerin ailesi

$\angle(\hat{\beta})$ ile gösterilir ve

$$\angle(\hat{\beta}) = \{\tilde{\beta} = A\hat{\beta} + d \in \angle : \text{Var}(\tilde{\beta}) \leq \text{Var}(\hat{\beta})\} \quad (3.27.4)$$

şeklindedir. Burada $\text{Var}(\tilde{\beta}) \leq \text{Var}(\hat{\beta})$ eşitsizliği $\text{Var}(\hat{\beta}) - \text{Var}(\tilde{\beta})$ nın n.n.d. olduğunu yani $\text{Var}(\tilde{\beta})$ nin Löwner'in kısmi sıralamasına göre $\text{Var}(\hat{\beta})$ dan daha iyi olduğunu gösterir.

Uygun olarak $\text{Var}(\tilde{\beta})$ n.n.d. ise $\text{Var}(\tilde{\beta}) \geq 0$ olarak ve eğer $\text{Var}(\hat{\beta})$ p.d. ise $\text{Var}(\hat{\beta}) > 0$ olarak yazılabilir.

Minimax Lineer Tahmin

(3.27.1) M modeli altında, (3.27.3) riskine göre β nin minimax lineer tahmin edicileri (MILE) ailesini ele alalım. Admissible tahmin ediciler ve \angle arasında yapısal bir ilişki vardır. (3.27.3) deki risk fonksiyonu R^p de yani p boyutlu öklid uzayında sınırsızdır. Ancak her belirli $k \in R$ ve $T \in M_p$ için

$$B_k(T) = \{\beta : \beta^T T \beta \leq k, k > 0, T > 0\} \quad (3.27.5)$$

kümesi üzerinde sınırlıdır. Eğer belirli bir k ve T için $\beta \in B_k(T)$ alınrsa o zaman minimax prensibi risk ile ilgili düşüncelere uygulanabilir.

Tanım 3.27.1 : Bir $\tilde{\beta}$ tahmin edicisi eğer \angle de $\sup_{\beta \in B_k(T)} R(\beta; \tilde{\beta}, C)$ yi minimize ediyorsa yani $\min_{\tilde{\beta} \in \angle} \sup_{\beta \in B_k(T)} R(\beta; \tilde{\beta}, C) = \sup_{\beta \in B_k(T)} R(\beta; \tilde{\beta}, C)$ ise o zaman $\tilde{\beta}$, β nin

\angle için de bir minimax lineer tahmin edicisi (MILE) olarak adlandırılır.

Bir tahmin edicinin bir MILE olması için gereken özellik tüm olası tahmin edicilerin sınıfına bağlıdır. \angle içindeki minimax lineer tahmin edicilerin sıfır yanlılığa sahip tahmin edicilere benzer olduğu gösterilebilir (Toutenburg (1982)).

Bu sebepten homojen lineer tahmin ediciler sınıfına kısıtlanmasında (yani $d \equiv 0$ olmak üzere $A\hat{\beta} + d \in \mathcal{L}$) bir kayıp olmaz. $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > 0$ ve $1 \leq r \leq p$ olmak üzere verilen bir $\Sigma = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_r) \in M_r$ matrisi için

$$M_p(\Sigma) = \{U\Sigma V' \in M_p : U'U = V'V = I_r \in M_r\} \quad (3.27.6)$$

matris kümesi tanımlansın . Burada $U, V \in M_{p,r}$ ve $I_r \in M_r$ dir. Dolayısıyla U ve V nin sütunları ortogonaldir. U ve V nin kare matrisler olmaları gerekmez ve ortogonal matrislerdir. $U\Sigma V'$ nin rankının $r \leq p$ olduğu açıktır. Verilen bir $B \in M_p(\Sigma)$ matrisi için (3.27.6) ayrışımı , B nin tekil değer ayrışımı olarak bilinir. Bu sebepten $M_p(\Sigma)$, verilen bir Σ matrisi için aynı $\sigma_1, \sigma_2, \dots, \sigma_p$ tekil değerlerine sahip $p \times p$ tipindeki matrislerin kümesidir. Eğer $r < p$ ise son $p-r$ singüler değer sıfırdır. $\Sigma = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_r)$ nin köşegen elemanlarının azalan sırada olduğu kabul edilmiştir.

$\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > 0$ olmak üzere bir Σ matrisi ve bir p.d. $C \in M_p$ matrisi verilsin. Buna göre $C^{\frac{1}{2}}$, C nin simetrik karekökü olmak üzere tahmin edici sınıfı

$$\zeta(\Sigma, C) = \left\{ A\hat{\beta} \in \mathcal{L} : A \in M_p, C^{\frac{1}{2}}AC^{-\frac{1}{2}} \in M_p(\Sigma) \right\} \quad (3.27.7)$$

şeklinde tanımlanır.

Teorem 3.27.1. de , M (3.27.1) modeli altında , (3.27.3) risk fonksiyonuna göre β nin $\zeta(\Sigma, C)$ içindeki minimax tahmin edicilerinin sınıfı oluşturulur. Minimax tahmin edicilerin kümesi $\zeta(\Sigma, C)$ ile gösterilir.

Teorem 3.27.1 : $1 \leq r \leq p$ olmak üzere $\Sigma = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_r) > 0$ ve C , (3.27.3) risk fonksiyonundaki ağırlık matrisi olmak üzere $\zeta(\Sigma, C)$ içinde (3.27.3) risk fonksiyonuna göre bir MILE olması için gerek ve yeter koşul $T_1 \geq T_2 \geq \dots \geq T_p$, $C^{\frac{1}{2}}S^{-1}C^{\frac{1}{2}}$ nin özdeğerleri olmak üzere

$$CA = A'C \quad (3.27.8)$$

ve

$$tr(A^2 S^{-1} C) = \sum_{j=1}^r \sigma_j^2 T_{p+1-j} \quad (3.27.9)$$

olmasıdır.

$$\text{İspat : } \tilde{\beta} = A\hat{\beta} \text{ için (3.27.3) riski } \tilde{A} = \left(I - C^{-\frac{1}{2}} A' C^{\frac{1}{2}} \right) \left(I - C^{\frac{1}{2}} A C^{-\frac{1}{2}} \right) \text{ olmak}$$

üzere

$$R(\beta; \tilde{\beta}, C) = \sigma^2 tr(CAS^{-1}A') + \beta' C^{\frac{1}{2}} \tilde{A} C^{\frac{1}{2}} \beta \quad (3.27.10)$$

şeklindedir.

Minimax tahminin bilinen teorisine göre (Toutenburg (1982)) $\lambda_1(\tilde{A})$, \tilde{A} nin en büyük özdeğeri ve $S = X'X$ olmak üzere

$$\sup_{\beta \in B_k(C)} R(\beta; \tilde{\beta}, C) = \sigma^2 tr(CAS^{-1}A') + k\lambda_1(\tilde{A}) \quad (3.27.11)$$

dir. MILE, (3.27.11) i minimize eden $\tilde{\beta} = A\hat{\beta} \in \zeta(\Sigma, C)$ tahmin edicilerinin bulunmasıyla elde edilir.

İlk olarak $\lambda_1(\tilde{A})$ nin minimize edilmesini ele alalım. $C^{\frac{1}{2}} A C^{-\frac{1}{2}} \in M_p(\Sigma)$ olduğundan, $V\Sigma U'$ (3.27.6) da tanımlanan matrislerin kümesi olmak üzere \tilde{A} ; $\tilde{A} = (I - V\Sigma U')(I - U\Sigma V') = (I + V\Sigma^2 V') - (V\Sigma U' + U\Sigma V')$ şeklinde yazılabilir.

Her ortogonal $U, V \in M_{p,r}$ için, en büyük özdeğerin kısıtlanmış maksimum olarak karakterizasyonu ile (Bellman, 1970)

$$\lambda_1(\tilde{A}) \geq \lambda_1(I + V\Sigma^2 V') - \lambda_1(V\Sigma U' + U\Sigma V')$$

eşitsizliği elde edilir. Tüm ortogonal $V \in M_{p,r}$ için $\lambda_1(I + V\Sigma^2 V') = 1 + \sigma_1^2$ olduğundan;

$U, V \in M_{p,r}$ ortogonal matrisler olmak üzere

$$\min_{U,V} \lambda_1(\tilde{A}) \geq 1 + \sigma_1^2 - \max_{U,V} (V\Sigma U' + U\Sigma V') = 1 + \sigma_1^2 - 2\sigma_1 = (1 - \sigma_1)^2 \quad (3.27.12)$$

elde edilir. (3.27.12) deki eşitliğin , $U=V$ olduğunda her $V \in M_{p,r}$ için elde edilebileceği açıktır. Böylelikle (3.27.8) e denk olan

$$C^{\frac{1}{2}} A C^{-\frac{1}{2}} = C^{-\frac{1}{2}} A' C^{\frac{1}{2}}$$

ifadesi elde edilir.

(3.27.11) deki iz ifadesini ele alalım. İzin değişme özelliği ve $C^{\frac{1}{2}} A C^{-\frac{1}{2}} \in M_p$

nın $C^{\frac{1}{2}} A C^{-\frac{1}{2}} = U \Sigma V'$ şeklindeki tekil değer ayrışımının kullanılmasıyla

$$tr(C A S^{-1} A') = tr\left(U \Sigma V' C^{\frac{1}{2}} S^{-1} C^{\frac{1}{2}} V \Sigma U'\right) = tr\left(V \Sigma^2 V' C^{\frac{1}{2}} S^{-1} C^{\frac{1}{2}}\right)$$

elde edilir. $tr(C A S^{-1} A')$, U ya bağlı değildir. Sonuç olarak (3.27.11) i V nin fonksiyonu olarak minimize etmek yeterlidir. İspatın sonucunda

$$\min_V tr\left(V \Sigma^2 V' C^{\frac{1}{2}} S^{-1} C^{\frac{1}{2}}\right) = \sum_{j=1}^r \sigma_j^2 T_{p+1-j}$$

bulunur (Anderson 1984). Bu ise (3.27.9) koşulunu oluşturur.

(3.27.8) ve (3.27.9) koşulları k nin değerine bağlı değildir. Dolayısıyla $\tilde{\zeta}(\Sigma, C)$ ile gösterilen $\zeta(\Sigma, C)$ içindeki $\beta \in B_k(C)$ minimax lineer tahmin ediciler kümesi ; her $k > 0$ için aynıdır. Buna rağmen β sınırlı bir kümeye kısıtlanmıştır.

Bir $\beta \in B_k(C)$, $k > 0$ kümesi önerilmeden , $\zeta(\Sigma, C)$ içindeki bir MILE den bahsedilebilir.

Teorem 3.27.2. : Verilen bir köşegen $\Sigma > 0$ için (3.27.3) riskine göre β nin $\zeta(\Sigma, C)$ deki bir lineer tahmin edicisi $A\hat{\beta}$, (3.27.8) , (3.27.9) koşullarını sağlıyorsa MILE dir.

C , ağırlık matrisinin özel durumlarının , istatistiksel uygulamalarda önemi vardır. (3.27.3) de $C = I$ nin yerine yazılmasıyla mse kriteri yani $R(\beta; \tilde{\beta}, I) = MSE(\tilde{\beta})$ elde edilir.

Bu olayda (3.27.8) ve (3.27.9) koşulları ; $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p > 0$, S nin özdeğerleri olmak üzere

$$A = A' \quad (3.27.8^0)$$

ve

$$tr(AS^{-1}A) = \sum_{j=1}^r \sigma_j^2 \lambda_{p+1-j}^{-1} \quad (3.27.9^0)$$

formunu alır.

(3.27.9⁰) eşitliği , $V'S^{-1}V = diag(\lambda_1^{-1}, \lambda_2^{-1}, \dots, \lambda_p^{-1})$ olduğunda sağlanır. Bu ise A ve S nin yer değiştirebilen simetrik matrisler olduğu anlamına gelir. GRE gibi shrinkge tahmin edicileri bu tiptedir (Liski 1979 , Obenchain 1978).

Bir başka özel durum $C \equiv S$ yazılmasıyla elde edilir. $R(\beta; \tilde{\beta}, S)$ riski aynı zamanda tahmin edici mse olarak adlandırılır. O zaman (3.27.8) koşulu $SA = A'S$ formunu alır. Verilen bir Σ için , (3.27.9) daki iz bir sabittir. (ve $= tr(\Sigma^2)$ ifadesine eşittir.) Sonuç olarak $\zeta(\Sigma, S)$ minmax lineer tahmin ediciler sınıfı için karakterizasyon elde edilmiş olur.

Sonuç 3.27.1 : $1 \leq r \leq p$ olmak üzere belirli bir $\Sigma = diag(\sigma_1, \sigma_2, \dots, \sigma_r) > 0$ verilmiş olsun. $R(\beta; A\hat{\beta}, S)$ riskine göre β nin $\zeta(\Sigma, S)$ deki $A\hat{\beta} \in \zeta(\Sigma, S)$ şeklindeki bir lineer tahmin edicisinin bir MILE olması için gerek ve yeter koşul

$$SA = A'S \quad (3.27.13)$$

olmasıdır.

3.28. Mixed Regresyon Tahmin Edicilerinin Karşılaştırılması

$Y = X\beta + \varepsilon$ lineer modelini ele alalım.

$$\tilde{r}_j = \tilde{R}_j \beta + \phi_j , E(\tilde{r}_j - \tilde{R}_j \beta) = \tilde{s}_j \quad (3.28.1)$$

$$\text{Cov}(\phi_j) = \left(\frac{\sigma^2}{k_j} \right) V_j, \text{Cov}(\varepsilon \phi_j') = 0 \quad j = 1, 2$$

iki lineer önbilgi kümesi olsun. Burada $j = 1, 2$ için \tilde{r}_j ve ϕ_j , $\tilde{m}_j \times 1$ tipinde stokastik vektörler; \tilde{R}_j , $\tilde{m}_j \times p$ tipinde belirli bir matris ve $\text{rank}(\tilde{R}_j) = \tilde{m}_j$; V_j p.d. ve k_j negatif olmayan bir skalerdir.

Theil ve Goldberger (1961) in, (1.2.1) ve (3.28.1) e dayanan ve $\tilde{s}_j = 0$ kabul edildiği mixed tahmin edicileri

$$\tilde{b}_j(k_j) = \left(X'X + k_j \tilde{R}_j' V_j^{-1} \tilde{R}_j \right)^{-1} \left(X'Y + k_j \tilde{R}_j' V_j^{-1} \tilde{r}_j \right), \quad j = 1, 2 \quad (3.28.2)$$

şeklindedir.

Bu tahmin edicilerin MSE matrisleri $U = (X'X)^{-1}$ ve $\tilde{S}_j = \left(k_j^{-1} V_j^{-1} + \tilde{R}_j U \tilde{R}_j' \right)^{-1}$

olmak üzere

$$\begin{aligned} \text{MSE}(\tilde{b}_j(k_j)) &= E(\tilde{b}_j(k_j) - \beta)(\tilde{b}_j(k_j) - \beta)' \quad (3.28.3) \\ &= \sigma^2 \left(U - U \tilde{R}_j' \tilde{S}_j \tilde{R}_j U \right) + U \tilde{R}_j' \tilde{S}_j \tilde{s}_j \tilde{s}_j' \tilde{S}_j \tilde{R}_j U \quad j = 1, 2 \end{aligned}$$

şeklindedir. Terasvirta (1986); β nin bir \tilde{b}_2 tahmin edicisinin \tilde{b}_1 tahmin edicisinden daha iyi olmasını aşağıdaki şekilde tanımlamıştır.

Tüm $A \geq 0$ kayıp matrisleri için

$$E(\tilde{b}_1 - \beta)' A (\tilde{b}_1 - \beta) - E(\tilde{b}_2 - \beta)' A (\tilde{b}_2 - \beta) \geq 0 \quad (3.28.4)$$

ise \tilde{b}_2 tahmin edicisi \tilde{b}_1 tahmin edicisinden daha iyidir.

Bu ifadeye eşdeğer olan bir başka form ise Theobald (1974) tarafından $\Delta_{12} = \text{MSE}(\tilde{b}_1) - \text{MSE}(\tilde{b}_2) \geq 0$ şeklinde verilmiştir.

Eğer \tilde{b}_2 , \tilde{b}_1 den daha iyi ise Freund ve Trenkler (1986) bunu \tilde{b}_2 nin \tilde{b}_1 den daha kötü olmaması olarak adlandırmıştır.

Mixed Tahmin Edicilerinin Karşılaştırılması

Freund ve Trenkler (1986) ; mixed tahmin edicilerinin karşılaştırılması için aşağıdaki teoremi ispatlamışlardır.

Teorem 3.28.1 : $\tilde{s}_1 = 0$ ve P ise $\tilde{R}_1 = P\tilde{R}_2$ ve

$$Cov(\phi_1) = PCov(\phi_2)P' \quad (3.28.5)$$

olacak şekilde bir $\tilde{m}_1 \times \tilde{m}_2$ tipinde matris olsun . Bu durumda aşağıdaki varsayımlar eşdeğerdir.

i) $\tilde{b}_2(k_2)$; $\tilde{b}_1(k_1)$ den daha iyidir.

ii) $P\tilde{s}_2 = 0$ ve $\tilde{s}_2' \tilde{S}_2 \tilde{s}_2 \leq \sigma^2$.

Bu teorem ; bir yanlı $\tilde{b}_2(k_2)$ mixed tahmin edicisinin bir yansız $\tilde{b}_1(k_1)$ mixed tahmin edicisinden daha iyi olması için gerek ve yeter koşulu verir.(3.28.5) deki varsayımı ele alalım. \tilde{R}_2 nin son \tilde{m}_1 satırı \tilde{R}_1 e eşit olacak ve $P = \begin{pmatrix} 0 & I \end{pmatrix}$ olacak şekilde

$\tilde{m}_2 \geq \tilde{m}_1$ ve $\tilde{R}_2 = \begin{pmatrix} \tilde{T}' & \tilde{R}_1' \end{pmatrix}$ olduğunu kabul edelim. Bundan başka $V_j = \begin{pmatrix} \sigma^2 \\ k_j \end{pmatrix} I_{\tilde{m}_j}$

($j = 1,2$) olduğunu kabul edelim. O zaman (3.28.5) $k_1 = k_2$ olduğunu belirtir.

Bu da (3.28.5) in daha güçlü bir kısıtlama olduğunu gösterir. Bu örnekte ise bu ; ön bilgileri farklı kovaryanslara sahip tahmin edicilerin karşılaştırılmalarını imkansız hale getirmektedir.

Benzer sonuçlar Terasvirta (1986) nin ispatladığı bir teoremin uygulanmasıyla elde edilebilir.

$\tilde{b}_j = D_j Y + h_j$ nin MSE matrisi ; $j = 1,2$ için $d_j = H_j \beta + h_j$ ve $H_j = D_j X - I$

olmak üzere ; kovaryans ve yanlılığın bir toplamı olarak

$$MSE(\tilde{b}_1) = \sigma^2 D_1 D_1' + d_1 d_1'$$

şeklinde yazılsın. $\Delta_{12} = \sigma^2 C + d_1 d_1' - d_2 d_2'$ olacak şekilde $C = D_1 D_1' - D_2 D_2'$ oluşturulur. Teoremden ; daha iyi olma sonuçları için gerekli olan

$$C = K L K' , d_j = K f_j , j = 1, 2 \quad (3.28.6)$$

ayrışımı kullanılır. Burada K $p \times r$ tipinde r ranklı bir matris ; L $r \times r$ tipinde simetrik bir matris ; f_j , $r \times 1$ tipinde bir matris ve $j = 1, 2$ dir. (3.28.6) ayrışımı ; C nin tekil ve $r - 1 \leq \text{rank}(C) \leq r < p - 1$ olduğu koşullar için gereklidir.

(3.28.6) daki temel düşünce ; Δ_{12} nin p.s.d. olması için gerek ve yeter koşul bulma problemini daha kolay bir probleme dönüştürmektir.

Benzer bir problem ; $\Delta_{12} = \sigma^2 L + f_1 f_1' - f_2 f_2'$ nün n.n.d. olması için gerek ve yeter koşul bulma problemidir. Bunun bilinen çözümü $\sigma^2 L + f_1 f_1' > 0$ olması ile sağlanan (3.28.7) dir.

Çünkü L $r \times r$ tipindedir ve son koşul $\text{rank}(L) \geq r - 1$ olduğunu belirtir. Orijinal problem de bu şekilde çözülmüş olur. Çünkü Δ_{12} nin p.s.d. olması için gerek ve yeter koşul $\bar{\Delta}_{12}$ nin n.n.d. olmasıdır. (3.28.6) koşulu daha kısıtlayıcıdır ; d_j , C nin kolon uzayında olmalıdır. Ancak eğer C tekil ve $\text{rank}(C) < p - 1$ olduğunda , böyle bir ayrışım yoksa o zaman Δ_{12} genellikle tanımsızdır.

Eğer f_2 , f_1 e lineer bağımlı ise aşağıdaki teoreme ihtiyaç yoktur.

Teorem 3.28.2 : (Terasvirta (1984 , 1986))

(1.2.1) deki modeli $j = 1, 2$ olmak üzere $\tilde{b}_j = D_j Y + h_j$ şeklindeki iki lineer tahmin ediciyi ele alalım. (3.28.6) daki ayrışımın olduğunu kabul edelim. Bundan başka f_2 ; f_1 den bağımsız ve $\sigma^2 L + f_1 f_1' > 0$ olsun. O zaman \tilde{b}_2 nin \tilde{b}_1 den daha iyi olması için gerek ve yeter koşul

$$f_2' \left(\sigma^2 L + f_1 f_1' \right)^{-1} f_2 \leq 1 \quad (3.28.7)$$

olmasıdır.

Eğer $L > 0$ ise o zaman daha iyi olma koşulu ; $j = 1,2$ için $f_{ij} = f_i' L^{-1} f_j$ olmak üzere

$$\sigma^2 \left\{ f_{22} - f_{21}^2 (\sigma^2 + f_{11})^{-1} \right\} \leq 1 \quad (3.28.8)$$

ifadesine eşdeğerdir.

Öte yandan $L < 0$ ise \tilde{b}_2 nin \tilde{b}_1 den daha iyi olması için gerek ve yeter koşul, L nin bir skaler ve $\sigma^2 l_{11} + f_1^2 - f_2^2 \geq 0$ olmasıdır.

Teorem 3.28.2 ; (3.28.2) tipindeki tahmin edicilerin karşılaştırılması için uygundur.

3.29. Genelleştirilmiş Liu Tahmin Edicisinin Bilinen Liu Tahmin Edicisine Üstünlüğü

$Y = X\beta + \varepsilon$ (1.2.1) lineer modelini ele alalım. Çoklu içilişkililik olduğunda bu problemi çözmek için kullanılan metodlardan biri Liu tahminidir. Liu tahmin edicisi (OLE), d , $0 < d < 1$ şeklinde bir parametre olmak üzere

$$\hat{\beta}_d = (X'X + I)^{-1} (X'Y + d\hat{\beta}) = (X'X + I)^{-1} (X'X + dI)\hat{\beta} \quad (3.29.1)$$

şeklindedir.

Genelleştirilmiş Liu tahmin edicisi (GLE) Liu tarafından $D = \text{diag}(d_1, d_2, \dots, d_p)$ ve $i = 1, 2, \dots, p$ için $0 < d_i < 1$ olmak üzere

$$\hat{\beta}_{Gd} = (X'X + I)^{-1} (X'Y + D\hat{\beta}) = (X'X + I)^{-1} (X'X + D)\hat{\beta} \quad (3.29.2)$$

şeklinde tanımlanmıştır.

GLE Ve OLE Tahmin Edicilerinin Mse Matrisleri

İlk olarak (1.2.1) modelini

$$Y = \beta_0 1 + \tilde{X}\beta_1 + \varepsilon \quad (3.29.3)$$

şeklinde yazalım. Burada değişkenler $\tilde{X}'\tilde{X}$ bir korelasyon matrisi formunda olacak şekilde standartlaştırılmıştır. $\tilde{X}'Y$ her bir açıklayıcı değişkenin korelasyon katsayılarından oluşan vektör ve $[1 \ 1 \ \dots \ 1]'$ dir. O zaman \bar{Y} , β_0 'n tahmin edilmesinde kullanılabilir.

Dolayısıyla sadece β_1 'in tahmininin ele alınması gerekir. $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p > 0$, $\tilde{X}'\tilde{X}$ ün özdeğerleri ve q_1, q_2, \dots, q_p ise karşılık gelen özvektörler olsun. $Q = [q_1 \ q_2 \ \dots \ q_p]$ ve $\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_p)$ olsun. Bu durumda $Z = XQ$ ve $\alpha = Q'B_1$ olmak üzere (3.29.3)

$$Y = \beta_0 1 + Z\alpha + \varepsilon \quad (3.29.4)$$

şeklinde yazılabilir. (3.29.4) modelinde α 'nın OLS tahmin edicisi

$$\hat{\alpha} = \Lambda^{-1}Z'Y \quad (3.29.5)$$

şeklinde ve bu yansız tahmin edicinin varyans-kovaryans matrisi

$$\text{Var}(\hat{\alpha}) = \sigma^2 \Lambda^{-1} \quad (3.29.6)$$

şeklinde. Benzer şekilde α 'nın Liu ve genelleştirilmiş Liu tahmin edicileri

$$\hat{\alpha}_d = (\Lambda + I)^{-1} (Z'Y + d\hat{\alpha}) = (\Lambda + I)^{-1} (\Lambda + dI)\hat{\alpha} \quad (3.29.7)$$

ve

$$\hat{\alpha}_{Gd} = (\Lambda + I)^{-1} (Z'Y + D\hat{\alpha}) = (\Lambda + I)^{-1} (\Lambda + D)\hat{\alpha} \quad (3.29.8)$$

şeklinde yeniden yazılabilir.

$W = (\Lambda + I)^{-1}(\Lambda + D)$ olsun . Böylece (3.29.8) den $\hat{\alpha}_{Gd} = W\hat{\alpha}$ olur. $E(\hat{\alpha}) = \alpha$ olduğundan $E(\hat{\alpha}_{Gd}) = W\alpha$ olur. $\hat{\alpha}_{Gd}$ tahmin edicisinin yanlılığı ve varyans-kovaryans matrisi

$$\text{Var}(\hat{\alpha}_{Gd}) = E[\hat{\alpha}_{Gd} - E(\hat{\alpha}_{Gd})][\hat{\alpha}_{Gd} - E(\hat{\alpha}_{Gd})]' = \sigma^2 W\Lambda^{-1}W' \quad (3.29.9)$$

$$\text{Bias}(\hat{\alpha}_{Gd}) = E(\hat{\alpha}_{Gd}) - \alpha = (W - I)\alpha \quad (3.29.10)$$

şeklindedir.

$\hat{\alpha}_{Gd}$ tahmin edicisinin MSE matrisi

$$\begin{aligned} \text{MSE}(\hat{\alpha}_{Gd}) &= E[(\hat{\alpha}_{Gd} - \alpha)(\hat{\alpha}_{Gd} - \alpha)'] = \text{Var}(\hat{\alpha}_{Gd}) + [\text{Bias}(\hat{\alpha}_{Gd})][\text{Bias}(\hat{\alpha}_{Gd})]' \\ &= \sigma^2 W\Lambda^{-1}W' + (I - D)(\Lambda + I)^{-1}\alpha\alpha'(\Lambda + I)^{-1}(I - D) \end{aligned} \quad (3.29.11)$$

şeklindedir. (3.29.1)'den

$$\begin{aligned} \text{MSE}(\hat{\alpha}_{Gd}) &= \sigma^2 \text{tr}(W\Lambda^{-1}W') + \text{tr}[(I - D)(\Lambda + I)^{-1}\alpha\alpha'(\Lambda + I)^{-1}(I - D)] \\ &= \sigma^2 \sum_{i=1}^p \frac{(\lambda_i + d_i)^2}{\lambda_i(\lambda_i + 1)^2} + \sum_{i=1}^p \frac{(1 - d_i)^2 \alpha_i^2}{(\lambda_i + 1)^2} \end{aligned} \quad (3.29.12)$$

elde edilir. Liu tahmin edicisi için $d_1 = d_2 = \dots = d_p = d$ dir. Dolayısıyla

$$\text{MSE}(\hat{\alpha}_d) = \sigma^2 \sum_{i=1}^p \frac{(\lambda_i + d)^2}{\lambda_i(\lambda_i + 1)^2} + \sum_{i=1}^p \frac{(1 - d)^2 \alpha_i^2}{(\lambda_i + 1)^2} \quad (3.29.13)$$

olur.

GLE nin OLE den Üstünlüğü

Chawla ve Jain (1988) ; MSE kriterini kullanarak GRR nin ORR den daha iyi olabileceğini göstermişlerdir. Benzer olarak GLE nin OLE den , belirli koşullarda daha iyi olduğu gösterilebilir.

Lemma 3.29.1 : $0 < d < \frac{\lambda_i(\alpha_i^2 - \sigma^2)}{\sigma^2 + \lambda_i\alpha_i^2}$ için $\sigma^2 \frac{(\lambda_i + d)^2}{\lambda_i(\lambda_i + 1)^2} + \frac{(1-d)^2\alpha_i^2}{(\lambda_i + 1)^2}$ bir

monoton azalan fonksiyondur.

İspat :

$$g_i(d) = \sigma^2 \frac{(\lambda_i + d)^2}{\lambda_i(\lambda_i + 1)^2} + \frac{(1-d)^2\alpha_i^2}{(\lambda_i + 1)^2} \quad (3.29.14)$$

olsun. d ye göre türev alındığında

$$g_i'(d) = \frac{2[\sigma^2(\lambda_i + d) + \lambda_i\alpha_i^2(d-1)]}{\lambda_i(\lambda_i + 1)^2} \quad (3.29.15)$$

olur.

$0 < d < \frac{\lambda_i(\alpha_i^2 - \sigma^2)}{\sigma^2 + \lambda_i\alpha_i^2}$ olduğunda $g_i'(d) < 0$ olduğundan lemma ispatlanmış olur.

Aşağıdaki teoremden GLE tahmin edicisinin , mse ye göre OLE tahmin edicisinden daha iyi olduğu ispatlanır.

Teorem 3.29.1 : Eğer $d < d_i < \frac{\lambda_i(\alpha_i^2 - \sigma^2)}{\sigma^2 + \lambda_i\alpha_i^2}$ ise $MSE(\hat{\alpha}_{GD}) < MSE(\hat{\alpha}_d)$ dir.

İspat : Lemma 3.29.1 den $d < d_i < \frac{\lambda_i(\alpha_i^2 - \sigma^2)}{\sigma^2 + \lambda_i\alpha_i^2}$ ise

$$\frac{\sigma^2(\lambda_i + d_i)^2}{\lambda_i(\lambda_i + 1)^2} + \frac{(1-d_i)^2\alpha_i^2}{(\lambda_i + 1)^2} < \frac{\sigma^2(\lambda_i + d)^2}{\lambda_i(\lambda_i + 1)^2} + \frac{(1-d)^2\alpha_i^2}{(\lambda_i + 1)^2} \quad (3.29.16)$$

dir. i üzerinde 1 den p ye kadar toplam alındığında (3.29.16) ifadesi ; tüm i için

$$d < d_i < \frac{\lambda_i(\alpha_i^2 - \sigma^2)}{\sigma^2 + \lambda_i\alpha_i^2} \text{ iken}$$

$$\sigma^2 \sum_{i=1}^p \frac{(\lambda_i + d_i)^2}{\lambda_i(\lambda_i + 1)^2} + \sum_{i=1}^p \frac{(1-d_i)^2 \alpha_i^2}{(\lambda_i + 1)^2} < \sigma^2 \sum_{i=1}^p \frac{(\lambda_i + d)^2}{\lambda_i(\lambda_i + 1)^2} + \sum_{i=1}^p \frac{(1-d)^2 \alpha_i^2}{(\lambda_i + 1)^2}$$

ifadesine indirgenir. Dolayısıyla tüm i için $d < d_i < \frac{\lambda_i(\alpha_i^2 - \sigma^2)}{\sigma^2 + \lambda_i\alpha_i^2}$ iken

$MSE(\hat{\alpha}_{Gd}) < MSE(\hat{\alpha}_d)$ elde edilir. Böylece ispat tamamlanmış olur.

KAYNAKLAR

- AKDENİZ, F. , KAÇIRANLAR, S. (1994).** “On the Almost Unbiased Generalized Liu Estimator and Unbiased Estimation of the Bias and MSE”
- ANDERSON, T.W., (1962).** “The Choice of Degree of a Polynomial Regression as a Multiple Decision Problem.” *Annals of Mathematical Statistics*, 33, 255-265
- ANDERSON, T.W. (1971).** “The Statistical Analysis of Time Series”. Wiley, New York
- ANDERSON, T.W. (1984).** “An Introduction to Multivariate Statistical Analysis”. Wiley, New York
- BAKSALARY, J.K. (1984).** “Nonnegative Definite and Positive Definite Solutions to the Matrix Equation $AXA^* = B$ ”. *Linear and Multilinear Algebra*, 16: 133-139.
- BAKSALARY, J.K. (1990).** “A Note on Generalized Ridge Estimator”. *Commun. Statist.-Theory Meth.*, 19(8), 2871-2877.
- BAKSALARY, J.K. , LISKI, E.P , TRENKLER, G. (1989).** “Mean Square Error Matrix Improvements and Admissibility of Linear Estimators”. *J. Statist. Plann. Inference*, 23, 313-325.
- BAKSALARY, J.K. , MARKIEWIEZ, A. (1985).** “Admissible Linear Estimators in Restricted Linear Models”. *Linear Algebra Appl.*, 70, 9-19.
- BAKSALARY, J.K. , PORDZIK, P.R. (1985).** “A Note on Using Linear Restrictions in a Gauss-Markov Model”. *Statistica (Bologna)*, 45: 699-709.
- BAKSALARY, J.K., PORDZIK, P.R. (1990).** “A Note on Comparing Unrestricted and Restricted Least Squares Estimators”. *Linear Algebra and Its Applications*, 127: 371-378.
- BEN-ISRAEL, A. , GREVILLE, T.N.E. (1974).** “Generalized inverses: Theory and Applications”. New York: John Wiley and Sons.
- CARPENTER, O. (1950).** “Note on the Extension of Graig’s Theorem to Noncentral Variates”. *Ann. Math. Statist.* 21: 455-457.
- CHAUBEY, Y.P. (1982).** “Best Minimum Bias Linear Estimators in Gauss-Markoff Model”. *Commun. Statist.-Theor. Meth.*, 11(17), 1959-1963

- CHAWLA, J.S. , JAIN, R.K. (1988).** “Superiority of Generalized Ridge Estimator Over Ordinary Ridge Estimator”. *Metron*, Vol XL VI-N, 1-4, 141-146
- CHIPMAN, J.S. (1964).** “On the Least Squares With insufficient Observations “. *J. Amer. Statist. Assoc.*, 59, 1078-1111.
- DURBIN, J. (1953).** “A Note on Regression When There is Extraneous information About One of the Coefficients”. *J. Amer. Statist. Assoc.*, 48, 799-808.
- DWIVEDI, T.D. , SRIVASTAVA, V.K. (1978).** “On the Minimum Mean Squared Error Estimators in a Regression Model”. *Commun. Statist.-Theor. Meth.*,7, 487-494.
- FAREBROTHER, R.W. (1975).** “The Minimum Mean Square Error Linear Estimator and Ridge Regression”. *Technometrics* 17, 127-128.
- FAREBROTHER, R.W. (1976).** “Further Results on the Mean Square Error of Ridge Regression”. *J.R. Statist. Soc. B*, 38, 248-250.
- FAREBROTHER, R.W. (1978).** “Partitioned Ridge Regression”. *Technometrics* 20, 121-122
- FAREBROTHER, R.W. (1984).** “The Restricted Least Squares Estimator and Ridge Regression”. *Commun. Statist.-Theor. Meth.*, 13(2), 191-196.
- FREUND, TRENKLER (1986).**
- FROST, P.A. (1979).** “Proxy Variables and Specification Bias”. *Review of Economics and Statistics*, 61, 323-325.
- GOODNIGHT, J. , WAALACE, T.D. (1972).** “Operational Techniques and Tables for Making Weak MSE Tests for Restrictions in Regression”. *Econometrica*, 40: 699-709.
- GRUBER, M.H.J. (1990).** “Regression Estimators A Comparative Study”. *Academic Press. inc.*, San Diego.
- GUILKEY, D.K. , PRICE, J.M. (1981).** “On Comparing Restricted Least Squares Estimators”. *Journal of Econometrics* 15, 397-404.
- GUNST, R.F. , MASON, R.L. (1976).** “Generalized Mean Square Error Properties of Regression Estimators”. *Communications in Statistics A5*, 1501-1508.

- HOERL, A.E. , KENNARD, R.W. (1970,a).** “Ridge Regression: Biased Estimation for Nonorthogonal Problems”. *Technometrics* 12, 55-67.
- HOERL, A.E. , KENNARD, R.W. (1970,b).** “Ridge Regression: Applications to Nonorthogonal Problems”. *Technometrics* 12, 69-82.
- HOERL, A.E. , KENNARD, R.W. (1976).** “Ridge Regression: Iterative Estimation of the Biasing Parameter”. *Commun. Statist. A5*, 77-88.
- HOERL, A.E. , KENNARD, R.W. (1977).**
- HOERL, A.E. , KENNARD, R.W. , BALDWIN, K.F. (1975).** “Ridge Regression: Some Simulation”. *Communications in Statistics*, 4, 105-123.
- HOFFMAN, K. (1979).** “Characterization of Minimax Linear Estimators in Linear Regression”. *Math. Operationsforsch.Statist.Ser. Statist.* 10, 19-26.
- JUDGE, G.G. , BOCK, M.E. (1978).** “The Statistical Implications of Pre-Test and Stein-Rule Estimators in Econometrics”. *North Holland Publishing Company* Amsterdam, New York.
- JOHNSON, S.R. , REIMER, S.C. , ROTHROCK, T.P. (1972).** “Principal Components and the Problem of Multicollinearity”. *Metroeconomica* 25, 306-314.
- JUDGE, G.G. , YANCEY, T.A. (1986).** “Improved Methods of Inference in Econometrics”. *Stud. Math. Managerial Econom.* 34(H. Theil and H.Glejser Eds.) North-Holland, Amsterdam.
- KAÇIRANLAR, S. , AKDENİZ, F. (1995).** “Superiority of Generalized Liu Estimator Over Ordinary Liu Estimator”. *Jour. inst. Math. and Com. Sci.* Vol-6, No-1, 1-4.
- KAÇIRANLAR, S. , SAKALLIOĞLU, S. , AKDENİZ, F. (1998).** “Mean Squared Error Comparisons of the Modified Ridge Regression Estimator and the Restricted Ridge Regression Estimator”. *Commun. Statist.-Theor. Meth.*, 27(1), 131-138.
- KADIYALA, K. (1984).** “A Class of Almost Unbiased and Efficient Estimators of Regression Coefficients”. *Economics Letters*, 16, 293-296.
- KEJIAN, L. (1993).** “A New Class of Biased Estimate in Linear Regression”. *Commun. Statist.-Theor. Meth.*, 22(2), 393-402.

- KENDALL, M.G.** (1965). "A Course in Multivariate Analysis". *Charles Griffin and Company Limited*, London.
- KUKS, J. , OLMAN, V.** (1972). "Minimaksnaja Linejnaja Ocenka Koefficientov Regressii". *Eesti NSV Teaduste Akadeemia Toimetised*, 21, 66-72.
- LAUTER, H.** (1975). "A Minimax Linear Estimator for Linear Parameters Under Restrictions in Form of Inequalities". *Math. Operationforsch. Statist.* 6, 689-695.
- LAWLESS, J.F. , WANG, P.** (1976). "A Simulation Study of Ridge and Other Regression Estimators". *Communications in Statistics-Theory and Methods*, 14, 1589-1604.
- LEE, K. , CAMPBELL, D.B.** (1985). "Selecting the Optimum k in Ridge Regression". *Communications in Statistics-Theory and Methods*, 15, 1589-1604.
- LEE, B.M.S. , TRIVEDI, P.K.** (1982). "Error Misspecification and Properties of the Simple Ridge Estimator". *Commun. Statist.-Theor. Meth.*, 11(14), 1615-1624.
- LISKI, E.P.** (1979). "On Reduced Risk Estimation in Linear Models". *Ph. D. Thesis, Acta Univ. Tamper. Ser. A* 105
- LISKI, E.P.** (1989). "Comparing Stochastically Restricted Linear Estimators in a Regression Model". *Biom J.* 31 ,3, 313-316.
- LISKI, E.P.** (1992). "On a Relationship Between Minimax-Linear and Admissible Estimators in Linear Regression". *Linear Algebra and its Applications* 176: 1069-120
- MCCALLUM, B.T.** (1972). "Relative Asymptotic Bias From Errors of Omission and Measurement". *Econometrica*, 40, 757-758.
- MEETER, D.A.** (1966). "On a Theorem Used in Nonlinear Least Squares". *SIAM J. Appl. Math.* 14, 1176-1179.
- MILLER, R.G.** (1974). "An Unbalanced Jack-Knife". *Ann. Statist.* 2, 880-891.
- MONTGOMERY, D.C. , PECK, E.A.** (1991). "Introduction to Linear Regression Analysis". *John Wiley and Sons*, New York.

- NOMURA, M.** (1988). "On the Almost Unbiased Ridge Regression Estimator". *Commun. Statist.-Simula.* 17(3), 729-743.
- OBENCHAIN, R.L.** (1978). "Good and Optimal Ridge Estimators". *Ann. Statist.* 6: 1111-1121.
- OHTANI, K.** (1981). "On the Use of a Proxy Variable in Prediction: An MSE Comparison". *Review of Economics and Statistics*, 63, 627-628.
- OHTANI, K.** (1986). "On Small Sample Properties of the Almost Unbiased Generalized Ridge Estimator". *Communications in Statistics-Theory and Methods*, 15, 1571-1578.
- PLISKIN, J.L.** (1987). "A Ridge-Type Estimator and Good Prior Means". *Commun. Statist.-Theor. Meth.*, 16(12), 3429-3437.
- PRICE, J.M.** (1982). "Comparisons Among Regression Estimators Under the Generalized Mean Square Error Criterion". *Commun. Statist.-Theor. Meth.*, 11(17), 1965-1984.
- PUNTANEN, S. , STYAN, G.P.H.** (1989). "The Equality of the Ordinary Least Squares Estimator and the Best Linear Unbiased Estimator". *Amer. Statist.*, 43, 153-164.
- SARKAR, N.** (1992). "A New Estimator Combining the Ridge Regression and the Restricted Least Squares Methods of Estimation". *Commun. Statist.-Theor. Meth.*, 21(7), 1987-2000.
- SINGH, B. , CHAUBEY, Y.P.** (1987). "On Some Improved Ridge Estimators". *Statistische Hefte*, 28, 53-67.
- SINGH, B. , CHAUBEY, Y.P. , DWIVEDI, T.D.** (1986). "An Almost Unbiased Ridge Estimator". *Sankhya: The indian Journal of Statistics*, B48, 342-346.
- SWAMY, P.A.V.B. , MEHTA, J.S.** (1977). "A Note on Minimum Average Risk Esimators for Coefficients in Linear Models". *Communications in Statistics* A6, 1181-1186.
- SWINDEL, B.F.** (1976). "Good Ridge Estimators Based on Prior Information". *Communications in Statistics*, A5, 1065-1075.
- TAYLOR, W.E.** (1974). "Smoothness Priors and Stochastic Prior Restrictions in Distributed Lag Estimation". *internat. Econom. Rev.*, 15, 803-804.

- TERASVIRTA, T.** (1981). "Some Results on improving the Least Squares Estimation of Linear Models by Mixed Estimation". *Scandinavian Journal of Statistics*, 8, 33-38.
- TERASVIRTA, T.** (1986). "Superiority Comparisons of Heterogeneous Linear Estimators". *Communications in Statistics A, Theory and Methods* 15, 1319-1336.
- TERASVIRTA, T.** (1988). "Superiority Comparisons Between Mixed Regression Estimators". *Commun. Statist.-Theor. Meth.*, 17(10), 3537-3546.
- TERASVIRTA, T.** (1998).
- THEIL, H.** (1958). "Economics Forecasts and Policy". *North-Holland*, Amsterdam.
- THEIL, H. , GOLDBERGER, A.S.** (1961). "On Pure and Mixed Statistical Estimation in Economics". *International Economic Review*, 2, 65-78.
- THEOBALD, C.M.** (1974). "Generalizations of Mean Square Error Applied to Ridge Regression ". *Journal of the Royal Statistical Society, Ser B*, 36, 103-106.
- TORO-VIZCARRONDO, C. , WALLACE, T.D.** (1968). "A Test of the Mean Square Error Criterion for Restrictions in Linear Regression". *Journal of the American Statistical Association*, 63, 558-572.
- TOUTENBURG, H.** (1982). "Prior Information in Linear Models". *Wiley*, Chichester.
- TRACY, D.S. , SRIVASTAVA, A.K.** (1994). "Comparisons of Operational Variants of Best Homogeneous and Heterogeneous Estimators in Linear Regression". *Commun. Statist.-Theor. Meth. A* 10(17), 1765-1778.
- TRENKLER, G.** (1978-b). "Mean Square Error and the Iteration Estimator". *Submitted for Publication*.
- TRENKLER, G.** (1980). "Generalized Mean Square Error Comparisons of Biased Regression Estimators". *Communications in Statistics*, A9, 1247-1259.
- TRENKLER, G.** (1981). "Biased Estimators in the Linear Regression Model". *Oegeschlager, Gunn and Hain*, Cambridge, Mass.

- TRENKLER,G.** (1986). "Mean Square Error Matrix Comparisons Between Mixed Estimators". *Statistica XLVI*, 493-501.
- TRENKLER, G. , TRENKLER, D.** (1983). "A Note on Superiority Comparisons of Homogeneous Liner Estimators". *Communications in Statistics*, A12, 799-808.
- ULLAH, A. , VINOD, H.D. , KADIYALA, K.** (1981). "A Family of Improved Shrinkage Factors for the Ordinary Ridge Regression in: E.G. Charatsis, ed., Proceedings of the Econometric Society European Meeting 1979". *North-Holland, Amsterdam*.
- VINOD, H.D.** (1976). "A Survey of Ridge Regression and Related Techniques for Improvements Over Ordinary Least Squares". *The Review of Economics and Statistics* 60, 121-131.
- VINOD, H.D. , ULLAH, A.** (1981). "Recent Advances in Regression Methods". *Marcel Dekker, New York*.
- WALACE, T.D.** (1972). "Weaker Criteria and Tests for Linear Restrictions in Regression". *Econometrica*, 40, 689-698.
- WALLACE, T.D. , TORO-VIZCARRONDO, C.** (1969). "Tables for the Mean Square Error Test for Exact Linear Restrictions in Regression". *Jour. Amer. Statist. Assoc.*, 64: 1649-1663.
- WICKENS, M.R.** (1972). "A Note on the Use of Proxy Variables". *Econometrica*, 40, 759-761.
- YANCEY, T.A. , JUDGE, G.G. , BOCK, M.E.** (1973). "Wallace's Weak Mean Square Error Criterion for Testing Linear Restrictions in Regression: A Tighter Bound". *Econometrica*, 41, 1203-1206.
- YANCEY, T.A. , JUDGE, G.G. , BOCK, M.E.** (1974). "A Mean Square Error Test When Stochastic Restrictions are Used in Regression". *Comm.Statist.*, 3, 755-768.

ÖZGEÇMİŞ

27 Haziran 1975 tarihinde Hatay'da doğdum . İlk , Orta ve Lise öğrenimimi Hatay , Bolu, Kahramanmaraş ve Adana'da tamamladım. Üniversite öğrenimime Ç.Ü. Fen-Edebiyat Fakültesi Matematik Bölümünde Ekim-1993 döneminde başlayıp Haziran-1997 döneminde mezun oldum . Nisan 1998 de Adana'da öğretmenliğe başladım . Eylül-1998 de Ç.Ü. Fen-Edebiyat Fakültesi Matematik Bölümünde yüksek lisans eğitimine başladım . Halen Milli Eğitim Bakanlığında öğretmenlik yapmaktayım .

