Çukurova Üniversitesi İİBF Dergisi

Cilt:17 Sayı:1 Haziran 2013
ss.1-15

Bazı Yanlı Tahmin Edicilerinin Karşılaştırılması: Bir Monte Carlo Çalışması

Comparison of Some Biased Estimators: A Monte Carlo Study

Fela ÖZBEY[1]

ÖZET

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

Anahtar Kelimeler: Yanlı Tahmin Edici, Hata Kareleri Ortalaması, GCV, Kestirim.

ABSTRACT

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

Keywords: Biased Estimator, Mean Squared Error, GCV, Prediction.

[1] Araş.Gör.Dr., Çukurova Üniversitesi, İİBF, Ekonometri Bölümü, fozbey@cu.edu.tr

1.GİRİŞ

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

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

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

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

En Küçük Kareler (EKK) tahmin edicisi lineer regresyon modellerinin tahmininde oldukça yaygın olarak kullanılmaktadır. EKK tahmin edicinin kullanımına yönelim, çoğunlukla yansız oluşundan ve yansız tahmin ediciler arasında en küçük varyansa sahip oluşundandır. Ancak, EKK tahmin edicisi, özellikle açıklayıcı değişkenler arasında doğrusal bir ilişki var ise, belli yönlerde oldukça değişken olabilir.

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

Çoklu iç ilişki sorunu ile baş etmek için kullanılan en eski tekniklerden birisi Temel Bileşenler (TB) regresyonudur (Massy, 1965). Bu yöntem, açıklayıcı değişkenlerin kolon uzayının bir alt kümesini seçip kullanmaya ve yanıt değişkenin bu alt küme üzerine izdüşümüne dayanmaktadır. Bir diğer popüler örnek Hoerl ve Kennard (1970) tarafından önerilen Ridge tahmin edicidir. Ridge tahmin edici, yansız bir tahmin ediciye, parametre tahminlerinin ve kestirimlerinin HKO'sunu küçültmek amacıyla yanlılık eklenilmesinin en bilinen örneğidir. Liu (1993) tarafından önerilen yanlı tahmin edici, Ridge tahmin edicisinin özel bir haline Stein tahmin edicinin (Stein tahmin edici için bkz: Stein, 1956) eklenmesi ile elde edilmiş bir modifikasyondur. Bu tahmin edicinin, HKO'sunun yanlılık parametresinin doğrusal bir fonksiyonu olması nedeniyle, yanlılık parametresinin seçimi daha kolaydır; bu bakımdan da Ridge tahmin edicisine göre üstündür. Baye ve Parker (1984), yukarıda adı geçen TB ve Ridge tahmin edicilerini birleştirerek, r-k sınıf tahmin edicisini oluşturmuşlardır. Bu tahmni edici, EKK, TB ve Ridge tahmin edicilerini ihtiva eden daha geniş kapsamlı bir sınıfı oluşturmaktadır. Liu tahmin edicinin Ridge tahmin ediciye üstünlüğünü göz önünde bulundurarak Kaçıranlar ve Sakallıoğlu (2001) TB tahmin ediciyi Liu tahmin edici ile birleştirerek r-k sınıf tahmin ediciye alternatif olarak r-d sınıf tahmin ediciyi önermişlerdir. Bu tahmin edici de EKK, TB ve Liu tahmin edicilerini ihtiva etmektedir. Bu çalışmada, bu iki sınıf tahmin edicilerin kestirim performansları kendi aralarında karşılaştırılırken, ihtiva ettikleri tahmin edicilere üstünlükleri de incelenecektir.

2. TAHMİN EDİCİLER VE GCV İSTATİSTİKLERİ

Çoklu doğrusal regresyon modeli

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{2.1}$$

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

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

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

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

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

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

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

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

şeklinde tanımlanır. Burada, $\mathbf{T} = \left[\mathbf{t}_1, \mathbf{t}_2, \ldots \mathbf{t}_p\right]$, $\mathbf{T}'\mathbf{X}'\mathbf{X}\mathbf{T} = \boldsymbol{\Lambda}$ olacak şekilde ortogonal bir matristir. Yani $\mathbf{T}$, $\mathbf{X}'\mathbf{X}$ matrisinin $p \times p$ boyutlu öz vektörleri matrisidir. $\boldsymbol{\Lambda} = diag\left(\lambda_1, \lambda_2, \ldots, \lambda_p\right)$, $\mathbf{X}'\mathbf{X}$ matrisinin öz değerleri matrisidir ve öz değerler azalan sıradadır. $\mathbf{T}_r$ matrisi, $\mathbf{T}$ matrisinin $p - r$ tane kolonu silindikten sonra kalan kısmı ve $\boldsymbol{\Lambda}_r = diag(\lambda_1, \lambda_2, \ldots, \lambda_r)$ dir.

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

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

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

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

$$\hat{\boldsymbol{\beta}}_{Liu} = (\mathbf{X}'\mathbf{X}+\mathbf{I})^{-1}(\mathbf{X}'\mathbf{y}+d\hat{\boldsymbol{\beta}}_{EKK})$$

$$= (\mathbf{X}'\mathbf{X}+\mathbf{I})^{-1}(\mathbf{X}'\mathbf{X}+d\mathbf{I})\hat{\boldsymbol{\beta}}_{EKK}, \quad 0 < d < 1; \qquad (2.5)$$

d yanlılık parametresidir. $d = 1$ iken $\hat{\boldsymbol{\beta}}_{Liu} = \hat{\boldsymbol{\beta}}_{EKK}$. Bu tahmin edicinin, HKO bakımından EKK tahmin edicisinden daha iyi olduğu ve d 'nin doğrusal bir fonksiyonu olması nedeniyle, yanlılık parametresinin seçimi daha kolay olduğu, bu bakımdan da Ridge tahmin edicisine göre üstün olduğu Liu (1993) tarafından gösterilmiştir.

Baye ve Parker (1984) özel durumlarda EKK, ridge tahmin edici ve temel bileşenler (TB) tahmin ediciyi veren r-k sınıf tahmin ediciyi tanımlamışlardır:

$$\hat{\boldsymbol{\beta}}_{r-k} = \mathbf{T}_r\left(\mathbf{Z}_r'\mathbf{Z}_r + k\mathbf{I}_r\right)^{-1}\mathbf{Z}_r'\mathbf{y} \quad , \quad k \geq 0 . \qquad (2.6)$$

Burada k yanlılık parametresidir. r-k sınıf tahmin edicinin, temel bileşenlerden daha küçük HKO değerine sahip olduğunu göstermişlerdir. Bu tahmin edicinin performansı farklı kriterlere göre Nomura ve Ohkubo (1985), Sarkar (1989), Sarkar (1996) ve Özkale ve Kaçıranlar (2008) de ele alınmıştır.

r-k sınıf tahmin ediciye alternatif olarak, Kaçıranlar ve Sakallıoğlu (2001) özel durumlarda EKK, Liu tahmin edici ve TB tahmin ediciyi veren r-d sınıf tahmin ediciyi tanımlamışlardır:

$$\hat{\boldsymbol{\beta}}_{r-d} = \mathbf{T}_r\left(\mathbf{Z}_r'\mathbf{Z}_r + \mathbf{I}_r\right)^{-1}(\mathbf{Z}_r'\mathbf{y} + d\hat{\boldsymbol{\alpha}}_{TB}) \quad , \quad -\infty < d < \infty . \qquad (2.7)$$

r-d sınıf tahmin edicinin, HKO yönünden, EKK, TB ve Liu tahmin edicilerinden daha iyi olduğu Kaçıranlar ve Sakallıoğlu (2001) tarafından gösterilmiştir.

Özbey ve Kaçıranlar (2010) yukarıda adı geçen tüm tahmin edicilerin yanlılık parametresini PRESS istatistiğini minimize edecek şekilde seçerek, kestirim performanslarını karşılaştırmışlardır. r-k ve r-d sınıf tahmin edicilerin ihtiva ettikleri diğer tahmin edicilerden daha küçük PRESS istatistikleri verdiği gözlemlemiştir.

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

r-k ve r-d sınıf tahmin edicilerinin HKO ve PRESS istatistikleri minimize edildiğinde, diğer tahmin edicilere göre daha iyi tahmin ve kestirim performansı gösterdiği literatürdeki çalışmalarda ele alınmıştır. Bu çalışmada, bu tahmin edicilerin GCV istatistiklerini minimize eden yanlılık parametreleri seçilerek, bu koşullar altında r-k ve r-d tahmin edicilerin tahmin ve kestirim performansları değerlendirilecektir.

Bir tahmin edicinin HKO'su, bu tahmin edicinin, tahmin edilecek parametreye ortalama uzaklığıdır.

$$\mathrm{HKO}(\hat{\boldsymbol{\beta}}) = E(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^2. \tag{2.8}$$

GCV istatistiği ise

$$\mathrm{GCV} = \frac{n\|(\mathbf{I}-\mathbf{H})\mathbf{y}\|^2}{(Tr(\mathbf{I}-\mathbf{H}))^2} \tag{2.9}$$

ile tanımlanır. Burada $\mathbf{H}$, $\mathbf{y}$ vektörünü $\hat{\mathbf{y}}$ vektörüne dönüştüren şapka matrisidir (hat matrix). Dolayısıyla (2.2) ile verilen EKK tahmin edicisi için GCV istatistiği:

$$\mathrm{GCV}_{EKK} = \frac{n\sum_{i=1}^{n}(y_i - \mathbf{x}_i(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y})^2}{(\sum_{i=1}^{n}(1-\mathbf{x}_i(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i'))^2} \tag{2.10}$$

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

$$\mathrm{GCV}_{TB} = \frac{n\sum_{i=1}^{n}(y_i - \mathbf{z}_{ri}(\mathbf{Z}_r'\mathbf{Z}_r)^{-1}\mathbf{Z}_r'\mathbf{y})^2}{(\sum_{i=1}^{n}(1-\mathbf{z}_{ri}(\mathbf{Z}_r'\mathbf{Z}_r)^{-1}\mathbf{z}_{ri}'))^2}\,. \tag{2.11}$$

Burada $\mathbf{z}_{ri}$, $\mathbf{Z}_r$ matrisinin i. satırından oluşan vektördür. (2.4) ile tanımlanan Ridge tahmin edicinin GCV istatistiği

$$\mathrm{GCV}_{R} = \frac{n\sum_{i=1}^{n}(y_i - \mathbf{x}_i(\mathbf{X}'\mathbf{X}+k\mathbf{I})^{-1}\mathbf{X}'\mathbf{y})^2}{(\sum_{i=1}^{n}(1-\mathbf{x}_i(\mathbf{X}'\mathbf{X}+k\mathbf{I})^{-1}\mathbf{x}_i'))^2}\,, \tag{2.12}$$

(2.5) ile verilen Liu tahmin edicinin GCV istatistiği:

$$\mathrm{GCV}_{L} = \frac{n\sum_{i=1}^{n}(y_i - \mathbf{x}_i(\mathbf{X}'\mathbf{X}+\mathbf{I})^{-1}(\mathbf{X}'\mathbf{X}-d\mathbf{I})(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y})^2}{(\sum_{i=1}^{n}(1-\mathbf{x}_i(\mathbf{X}'\mathbf{X}+\mathbf{I})^{-1}(\mathbf{X}'\mathbf{X}-d\mathbf{I})(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i'))^2} \tag{2.13}$$

(2.6) ile verilen r-k sınıf tahmin edicinin GCV istatistiği:

$$\mathrm{GCV}_{r-k}=\frac{n\sum_{i=1}^{n}(y_i-\mathbf{z}_{ri}(\mathbf{Z}_r'\mathbf{Z}_r+k\mathbf{I}_r)^{-1}\mathbf{Z}_r'\mathbf{y})^2}{(\sum_{i=1}^{n}(1-\mathbf{z}_{ri}(\mathbf{Z}_r'\mathbf{Z}_r+k\mathbf{I}_r)^{-1}\mathbf{z}_{ri}'))^2}, \quad (2.14)$$

ve (2.7) ile tanımlanan r-d sınıf tahmin edicinin GCV istatistiği:

$$\mathrm{GCV}_{r-d}=\frac{n\sum_{i=1}^{n}(\mathbf{y}-\mathbf{Z}_r(\mathbf{Z}_r'\mathbf{Z}_r+\mathbf{I})^{-1}(\mathbf{Z}_r'\mathbf{Z}_r-d\mathbf{I}_r)(\mathbf{Z}_r'\mathbf{Z}_r)^{-1}\mathbf{Z}_r'\mathbf{y})^2}{(\sum_{i=1}^{n}(1-\mathbf{z}_{ri}(\mathbf{Z}_r'\mathbf{Z}_r+\mathbf{I}_r)^{-1}(\mathbf{Z}_r'\mathbf{Z}_r-d\mathbf{I}_r)(\mathbf{Z}_r'\mathbf{Z}_r)^{-1}\mathbf{z}_{ri}'))^2} \quad (2.15)$$

şeklinde tanımlanır.

3.r-k VE r-d SINIF TAHMİN EDİCİLERİN KESTİRİM PERFORMANSLARININ KARŞILAŞTIRILMASI

EKK tahmin edicisine alternatif olarak tanımlanan yanlı tahmin edicilerden ikisi de r-k ve r-d sınıf tahmin edicileridir. Bunların hata kareleri ortalaması (HKO) kriterine göre karşılaştırılmaları literatürde ele alınmıştır. Bu bölümde, birbirine alternatif bazı yanlı tahmin edicilerin, yanlılık parametrelerinin GCV istatistiklerini minimize edecek şekilde seçilerek performansları karşılaştırılacaktır.

3.1. Monte Carlo Simülasyonları

Literatürdeki çalışmalarda, birçok tahmin edici özellikle HKO yönünden karşılaştırılmıştır. Simülasyon çalışmaları da daha çok HKO karşılaştırılmalarına yöneliktir. Ancak, yazarın bilgisine göre, r-k ve r-d tahmin edicileri GCV kriterine göre karşılaştıran çalışmalar literatürde yoktur. Bu çalışmada farklı değişken ve gözlem sayı kullanılarak deney farklı sayılarda tekrarlanmıştır. Bu nedenle, bu çalışmada elde edilen sonuçlar, uygulamalı çalışmalar yapan araştırmacılara tahmin edici seçimi konusunda yardımcı olacaktır.

Bu çalışmada açıklayıcı değişkenler, McDonald ve Galarneau (1975), Kibria (2003) ve Güler ve Kaçıranlar(2009)'ın çalışmasındakilere benzer şekilde türetilmiştir:

$$x_{ij} = (1-\eta^2)^{1/2} z_{1ij} + \eta z_{1ip+1}, \quad i = 1,2,...,n, \quad j = 1,2,..., p\,. \tag{3.1}$$

Burada z_{1ij} standart normal görünürde rastsal (pseudo-random) sayılardır. Bu durumda x_j ve x_k ($j \neq k$) gibi herhangi iki açıklayıcı değişken arasındaki korelasyon η^2 dir. η sırasıyla 0.8, 0.9 ve 0.99 alınarak deneyler üç farklı korelasyon ilişkisi türetilerek tekrarlanmıştır. Deneyler 100, 500 ve 1000 tekrarlama ile yapılmıştır. Deneydeki açıklayıcı değişken sayısı $n = 30$ için $p = 5$ ve $n = 50$ için $p = 7$ seçilmiştir. Açıklayıcı değişkenler deneyin başlangıcında üretilmiştir ve deney sırasında sabit tutulmuştur.

Açıklayıcı değişken

$$\mathbf{y} = \mathbf{X\beta} + \boldsymbol{\varepsilon} \tag{3.2}$$

şeklinde tanımlanarak türetilmiştir. Burada $\boldsymbol{\beta}$ vektörü, $\mathbf{X'X}$ matrisinin en büyük özdeğerine karşılık gelen normalleştirilmiş özvektör olarak seçilmiştir. (bknz. Newhouse ve Oman, 1971) $\boldsymbol{\varepsilon} \sim N(0,1)$ görünürde rastsal sayılardan oluşmaktadır.

Yukarıda belirtilen koşullar altında gerçekleştirilen deneyde GCV istatistikleri minimize edilerek yanlılık parametreleri seçilmiştir. Ayrıca her tahmin edici için skaler HKO

$$HKO(\hat{\boldsymbol{\beta}}) = \frac{1}{MCT} \sum_{i=1}^{MCT} (\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta})'(\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}) \tag{3.3}$$

ile elde edilmiştir. Burada $\hat{\boldsymbol{\beta}}_i$, $\boldsymbol{\beta}$ parametre vektörü için i. tekrarda elde edilen tahmin değeridir. MCT ise Monte Carlo tekrarlarının sayısıdır. Simülasyonlar MATLAB 7.0 programı kullanılarak yapılmıştır. Rastsal sayılar üretilirken başlangıç değeri (tohum) 45324762 alınmıştır.

Tablo 3.1.1: $n = 30$ ve $p = 5$ için Elde Edilen GCV ve HKO istatistikleri

p=5, r =1, n=30													
η	istatistik	EKK			TB			r-k			r-d		
		TS=100	TS=500	TS=1000	TS=100	TS=500	TS=1000	TS=100	TS=500	TS=1000	TS=100	TS=500	TS=1000
0.80	**GCV**	1.1711	1.1617	1.1553	1.0060	0.9991	0.9994	1.0056	0.9987	0.9990	1.0056	0.9987	0.9990
	HKO	0.4527	0.4560	0.4770	0.0097	0.0120	0.0119	0.0096	0.0123	0.0122	0.0096	0.0123	0.0122
0.90	**GCV**	1.1719	1.1616	1.1557	1.0064	0.9986	0.9995	1.0061	0.9983	0.9992	1.0061	0.9983	0.9992
	HKO	0.8483	0.8495	0.8897	0.0075	0.0096	0.0094	0.0074	0.0098	0.0095	0.0074	0.0098	0.0095
0.99	**GCV**	1.1723	1.1615	1.1562	1.0065	0.9981	0.9995	1.0063	0.9978	0.9993	1.0063	0.9978	0.9993
	HKO	8.0445	8.0064	8.3952	0.0058	0.0075	0.0072	0.0057	0.0076	0.0073	0.0057	0.0076	0.0073
p=5, r =2, n=30													
η	istatistik	EKK			TB			r-k			r-d		
		TS=100	TS=500	TS=1000	TS=100	TS=500	TS=1000	TS=100	TS=500	TS=1000	TS=100	TS=500	TS=1000
0.80	**GCV**	1.1711	1.1617	1.1553	1.0488	1.0359	1.0358	1.0356	1.0229	1.0228	1.0311	1.0182	1.0181
	HKO	0.4527	0.4560	0.4770	0.0733	0.0872	0.0881	0.0493	0.0629	0.0635	0.0440	0.0589	0.0596
0.90	**GCV**	1.1719	1.1616	1.1557	1.0496	1.0357	1.0360	1.0273	1.0139	1.0142	1.0074	0.9930	0.9930
	HKO	0.8483	0.8495	0.8897	0.1275	0.1521	0.1541	0.0603	0.0821	0.0836	0.0490	0.0749	0.0768
0.99	**GCV**	1.1723	1.1615	1.1562	1.0501	1.0353	1.0361	1.0068	0.9936	0.9945	0.8985	0.8723	0.8650
	HKO	8.0445	8.0064	8.3952	1.1446	1.3677	1.3937	0.2808	0.4677	0.4882	3.7221	4.1069	4.2481
p=5, r =3, n=30													
η	istatistik	EKK			TB			r-k			r-d		
		TS=100	TS=500	TS=1000	TS=100	TS=500	TS=1000	TS=100	TS=500	TS=1000	TS=100	TS=500	TS=1000
0.80	**GCV**	1.1711	1.1617	1.1553	1.0836	1.0752	1.0745	1.0477	1.0389	1.0385	1.0279	1.0178	1.0183
	HKO	0.4527	0.4560	0.4770	0.1766	0.1778	0.1800	0.0968	0.1018	0.1017	0.0769	0.0850	0.0833
0.90	**GCV**	1.1719	1.1616	1.1557	1.0847	1.0747	1.0744	1.0314	1.0215	1.0217	0.9794	0.9658	0.9694
	HKO	0.8483	0.8495	0.8897	0.3232	0.3252	0.3305	0.1232	0.1307	0.1296	0.1395	0.1570	0.1479
0.99	**GCV**	1.1723	1.1615	1.1562	1.0854	1.0741	1.0742	1.0020	0.9911	0.9926	0.9078	0.8895	0.8960
	HKO	8.0445	8.0064	8.3952	3.0161	3.0399	3.1036	0.6441	0.7339	0.7122	6.9077	7.8378	7.5332

(devam ediyor)

Tablo 3.1.1: $n = 30$ ve $p = 5$ için Elde Edilen GCV ve HKO istatistikleri (devam)

p=5, r =4, n=30													
η	istatistik	EKK			TB			r-k			r-d		
		TS=100	TS=500	TS=1000	TS=100	TS=500	TS=1000	TS=100	TS=500	TS=1000	TS=100	TS=500	TS=1000
0.80	**GCV**	1.1711	1.1617	1.1553	1.1172	1.1122	1.1110	1.0517	1.0453	1.0446	1.0114	1.0019	1.0030
	HKO	0.4527	0.4560	0.4770	0.3245	0.3138	0.3173	0.1431	0.1451	0.1459	0.1131	0.1190	0.1171
0.90	**GCV**	1.1719	1.1616	1.1557	1.1176	1.1120	1.1112	1.0301	1.0223	1.0224	0.9614	0.9483	0.9529
	HKO	0.8483	0.8495	0.8897	0.6090	0.5833	0.5905	0.1840	0.1885	0.1885	0.3063	0.3241	0.3063
0.99	**GCV**	1.1723	1.1615	1.1562	1.1176	1.1116	1.1114	0.9973	0.9873	0.9888	0.9211	0.9095	0.9173
	HKO	8.0445	8.0064	8.3952	5.8057	5.5049	5.5814	1.0440	1.1057	1.0912	8.7664	8.2214	7.2626
p=5, r =5, n=30													
η	istatistik	EKK			TB			r-k			r-d		
		TS=100	TS=500	TS=1000	TS=100	TS=500	TS=1000	TS=100	TS=500	TS=1000	TS=100	TS=500	TS=1000
0.80	**GCV**	1.1711	1.1617	1.1553	1.1711	1.1617	1.1553	1.0642	1.0547	1.0507	0.9969	0.9883	0.9883
	HKO	0.4527	0.4560	0.4770	0.4527	0.4560	0.4770	0.1620	0.1698	0.1793	0.1500	0.1610	0.1662
0.90	**GCV**	1.1719	1.1616	1.1557	1.1719	1.1616	1.1557	1.0373	1.0272	1.0244	0.9519	0.9439	0.9468
	HKO	0.8483	0.8495	0.8897	0.8483	0.8495	0.8897	0.2006	0.2139	0.2294	0.5067	0.5051	0.4897
0.99	**GCV**	1.1723	1.1615	1.1562	1.1723	1.1615	1.1562	0.9995	0.9883	0.9878	0.9274	0.9216	0.9259
	HKO	8.0445	8.0064	8.3952	8.0445	8.0064	8.3952	1.0499	1.1909	1.3233	11.0923	8.4063	7.6046

Tablo 3.1.2: $n = 50$ ve $p = 7$ için Elde Edilen GCV ve HKO istatistikleri

p=7, r=2, n=50													
η	istatistik	EKK			TB			r-k			r-d		
		TS=100	TS=500	TS=1000	TS=100	TS=500	TS=1000	TS=100	TS=500	TS=1000	TS=100	TS=500	TS=1000
0.80	**GCV**	1.1418	1.1384	1.1393	1.0240	1.0214	1.0216	1.0193	1.0166	1.0168	1.0181	1.0153	1.0155
	HKO	0.3649	0.3675	0.3656	0.0345	0.0342	0.0356	0.0263	0.0268	0.0279	0.0245	0.0252	0.0263
0.90	**GCV**	1.1417	1.1383	1.1393	1.0240	1.0214	1.0216	1.0147	1.0119	1.0122	1.0075	1.0045	1.0049
	HKO	0.6860	0.6906	0.6869	0.0615	0.0604	0.0633	0.0354	0.0360	0.0380	0.0280	0.0291	0.0312
0.99	**GCV**	1.1415	1.1381	1.1392	1.0240	1.0213	1.0216	1.0014	0.9983	0.9988	0.8937	0.9048	0.9039
	HKO	6.5115	6.5506	6.5163	0.5596	0.5458	0.5758	0.1690	0.1760	0.1944	1.6716	1.7626	1.8259
p=7, r=3, n=50													
η	istatistik	EKK			TB			r-k			r-d		
		TS=100	TS=500	TS=1000	TS=100	TS=500	TS=1000	TS=100	TS=500	TS=1000	TS=100	TS=500	TS=1000
0.80	**GCV**	1.1418	1.1384	1.1393	1.0485	1.0436	1.0429	1.0326	1.0277	1.0271	1.0237	1.0187	1.0184
	HKO	0.3649	0.3675	0.3656	0.0770	0.0813	0.0847	0.0455	0.0500	0.0519	0.0349	0.0398	0.0413
0.90	**GCV**	1.1417	1.1383	1.1393	1.0485	1.0437	1.0430	1.0225	1.0177	1.0174	0.9944	0.9894	0.9904
	HKO	0.6860	0.6906	0.6869	0.1416	0.1491	0.1557	0.0575	0.0652	0.0677	0.0497	0.0588	0.0602
0.99	**GCV**	1.1415	1.1381	1.1392	1.0485	1.0437	1.0431	1.0018	0.9969	0.9974	0.9267	0.9201	0.9260
	HKO	6.5115	6.5506	6.5163	1.3233	1.3890	1.4522	0.2457	0.3281	0.3405	3.4757	3.8106	3.5984
p=7, r=4, n=50													
η	istatistik	EKK			TB			r-k			r-d		
		TS=100	TS=500	TS=1000	TS=100	TS=500	TS=1000	TS=100	TS=500	TS=1000	TS=100	TS=500	TS=1000
0.80	**GCV**	1.1418	1.1384	1.1393	1.0675	1.0669	1.0662	1.0372	1.0360	1.0355	1.0166	1.0144	1.0144
	HKO	0.3649	0.3675	0.3656	0.1426	0.1359	0.1392	0.0717	0.0692	0.0710	0.0507	0.0490	0.0501
0.90	**GCV**	1.1417	1.1383	1.1393	1.0673	1.0669	1.0663	1.0230	1.0216	1.0214	0.9798	0.9759	0.9771
	HKO	0.6860	0.6906	0.6869	0.2666	0.2528	0.2590	0.0931	0.0883	0.0909	0.1162	0.1153	0.1134
0.99	**GCV**	1.1415	1.1381	1.1392	1.0671	1.0669	1.0664	0.9983	0.9961	0.9965	0.9432	0.9353	0.9384
	HKO	6.5115	6.5506	6.5163	2.5239	2.3821	2.4394	0.4726	0.4420	0.4615	3.2181	4.0026	3.7184

(devam ediyor)

Tablo 3.1.2: $n = 50$ ve $p = 7$ için Elde Edilen GCV ve HKO istatistikleri (devam)

p=7, r=5, n=50													
η	istatistik	EKK			TB			r-k			r-d		
		TS=100	TS=500	TS=1000	TS=100	TS=500	TS=1000	TS=100	TS=500	TS=1000	TS=100	TS=500	TS=1000
0.80	**GCV**	1.1418	1.1384	1.1393	1.0885	1.0902	1.0892	1.0413	1.0416	1.0409	1.0094	1.0064	1.0065
	HKO	0.3649	0.3675	0.3656	0.2158	0.2029	0.2072	0.0919	0.0871	0.0896	0.0635	0.0618	0.0633
0.90	**GCV**	1.1417	1.1383	1.1393	1.0884	1.0901	1.0892	1.0241	1.0238	1.0234	0.9736	0.9677	0.9698
	HKO	0.6860	0.6906	0.6869	0.4052	0.3801	0.3877	0.1180	0.1093	0.1132	0.1816	0.1896	0.1839
0.99	**GCV**	1.1415	1.1381	1.1392	1.0882	1.0900	1.0893	0.9969	0.9954	0.9957	0.9476	0.9400	0.9450
	HKO	6.5115	6.5506	6.5163	3.8470	3.6014	3.6679	0.6188	0.5410	0.5721	3.7133	4.1930	3.6972
p=7, r=6, n=50													
η	istatistik	EKK			TB			r-k			r-d		
		TS=100	TS=500	TS=1000	TS=100	TS=500	TS=1000	TS=100	TS=500	TS=1000	TS=100	TS=500	TS=1000
0.80	**GCV**	1.1418	1.1384	1.1393	1.1166	1.1158	1.1150	1.0485	1.0467	1.0463	1.0051	0.9985	0.9999
	HKO	0.3649	0.3675	0.3656	0.2808	0.2753	0.2786	0.1033	0.1031	0.1041	0.0755	0.0829	0.0805
0.90	**GCV**	1.1417	1.1383	1.1393	1.1163	1.1156	1.1150	1.0279	1.0259	1.0260	0.9745	0.9639	0.9677
	HKO	0.6860	0.6906	0.6869	0.5279	0.5168	0.5222	0.1294	0.1290	0.1294	0.2309	0.2712	0.2489
0.99	**GCV**	1.1415	1.1381	1.1392	1.1159	1.1153	1.1150	0.9976	0.9947	0.9956	0.9566	0.9439	0.9502
	HKO	6.5115	6.5506	6.5163	5.0132	4.9004	4.9445	0.6413	0.6450	0.6375	3.2879	4.4104	3.7533
p=7, r=7, n=50													
η	istatistik	EKK			TB			r-k			r-d		
		TS=100	TS=500	TS=1000	TS=100	TS=500	TS=1000	TS=100	TS=500	TS=1000	TS=100	TS=500	TS=1000
0.80	**GCV**	1.1418	1.1384	1.1393	1.1418	1.1384	1.1393	1.0518	1.0482	1.0489	1.0007	0.9923	0.9940
	HKO	0.3649	0.3675	0.3656	0.3649	0.3675	0.3656	0.1199	0.1234	0.1220	0.0982	0.1152	0.1095
0.90	**GCV**	1.1417	1.1383	1.1393	1.1417	1.1383	1.1393	1.0290	1.0254	1.0263	0.9744	0.9636	0.9665
	HKO	0.6860	0.6906	0.6869	0.6860	0.6906	0.6869	0.1505	0.1565	0.1531	0.2854	0.3483	0.3222
0.99	**GCV**	1.1415	1.1381	1.1392	1.1415	1.1381	1.1392	0.9967	0.9929	0.9943	0.9594	0.9476	0.9523
	HKO	6.5115	6.5506	6.5163	6.5115	6.5506	6.5163	0.7568	0.8282	0.7882	3.4418	4.4134	3.9499

Daha önce de belirtildiği gibi bileşen sayısı 1 iken r-d ve r-k sınıf tahmin ediciler kendilerini TB tahmin edicisine dönüştürme eğilimindedirler. Bunu göstermek amacıyla $n=30$ ve $p=5$ için yapılan simülasyonların tablosunda (Tablo 3.1.1) $r=1$ iken elde edilen sonuçlar tabloya eklenmiştir. $n=50$ ve $p=7$ için yapılan simülasyonların tablosunda (Tablo 3.1.2) ise benzer sonuçlar elde edildiğinden sonuçlar eklenmemiştir.

Tablo 3.1.1 ve Tablo 3.1.2 incelendiğinde genel olarak beklenildiği gibi çoklu bağlantının şiddeti arttıkça (η değeri arttıkça) HKO'nun da arttığı gözlemlenmektedir. Çoklu bağlantının şiddeti arttığı halde HKO'nun azaldığı tek durum bileşen sayısının 1 olduğu durumdur. Bunun nedeni ise çoklu bağlantının şiddeti arttıkça istikrarlı bileşenin varyansa katkısının artmasıdır. Dolayısıyla bu bileşenin bağımsız değişkeni açıklama gücü artmaktadır.

Genel olarak bileşen sayısı ne olursa olsun en küçük GCV istatistikleri r-d tahmin edici ile elde edilmiştir.

$\eta=0.80$ iken bileşen sayısı ne olursa olsun ($r>2$ için), o bileşen sayısın için r-d sınıf tahmin edici ile en küçük GCV ve HKO istatistikleri elde edilmiştir. Bu tahmin edici için bileşen sayısı artıkça HKO istatistiğinin değeri artsa da GCV istatistiğinin değeri gittikçe azalmıştır. Dolayısıyla en küçük GCV istatistiği tüm bileşenlerin kullanıldığı (yani r-d tahmin edicinin Liu tahmin edicisine dönüştüğü) durumda elde edilmiştir. Ancak elde edilen HKO istatistiği, daha az bileşenin kullanıldığı r-k ve r-d sınıf tahmin edicileri ile elde edilenlerden daha yüksektir. Dolayısı ile çoklu bağlantının şiddeti nispeten daha az olduğu durumlarda belli sayıda bileşen seçilerek r-d sınıf tahmin edicinin kullanılması hem iyi kestirim hem de iyi tahmin sonuçları elde edilmesini sağlar.

$\eta=0.90$ iken bileşen sayısı ne olursa olsun yine en küçük GCV istatistikleri r-d sınıf tahmin edici ile elde edilmiştir. Yine bileşen sayısı artıkça GCV istatistiğinin değeri gittikçe azalmış ve en küçük GCV istatistiği r-d tahmin edicinin Liu tahmin edicisine dönüştüğü durumda elde edilmiştir. Bileşen sayısı az iken o bileşen sayısı için en küçük HKO istatistikleri r-d sınıf tahmin edici ile elde edilirken, bileşen sayısı arttıkça o bileşen sayısı için elde edilen en küçük HKO istatistiği r-k sınıf tahmin edici ile elde edilmiştir. Bileşen sayısı arttıkça r-d ve r-k sınıf tahmin edicileri ile elde edilen HKO istatistikleri arasındaki fark da giderek artmıştır.

$\eta=0.99$ iken bileşen sayısı ne olursa olsun en küçük GCV istatistikleri r-d sınıf tahmin edici ile, en küçük HKO istatistikleri ise r-k sınıf tahmin edici ile elde edilmiştir. Önceki durumlardan farklı olarak r-d sınıf tahmin edicinin GCV istatistikleri giderek azalmak yerine bileşen sayısı 2 iken ($r=2$) en küçük değeri almaktadır ve bileşen sayısı arttıkça GCV istatistiği de artmaktadır. Bileşen sayısı arttıkça TB, r-k ve r-d sınıf tahmin edicilerin HKO istatistikleri de artmaktadır ancak, r-k sınıf tahmin edicinin HKO istatistiğindeki artış diğerlerine kıyasla çok küçüktür.

Sonuç

Bileşen sayısı 1 olduğunda r-k ve r-d sınıf tahmin ediciler kendilerini TB tahmin edicisine dönüştürme eğilimindedirler. Daha fazla sayıda bileşen kullanıldığında r-k ve r-d sınıf tahmin ediciler ile daha küçük GCV istatistikleri elde edilmiştir. Çoklu bağlantının çok şiddetli olmadığı durumlarda Liu tahmin ediciyle daha küçük GCV istatistikleri elde edilmiş olsa da HKO istatistikleri r-k ve r-d sınıf tahmin edicileri ile elde edilen HKO istatistiklerinden daha büyüktür. Bu da kestirim performansının daha iyi olmasına rağmen parametre tahmininde r-k ve r-d sınıf tahmin edicilerine göre zayıf kaldığını göstermektedir. Çoklu bağlantı sorunu çok şiddetli olduğunda daha az bileşen kullanılarak r-d sınıf tahmin edici ile daha küçük GCV istatistikleri elde etmek mümkündür. Ancak, elde edilen HKO istatistiklerinin değerleri çok büyük (bazen EKK tahmin edicisi ile elde edilenden bile daha büyük) olmaktadır. Dolayısıyla r-d sınıf tahmin edicisinin tahmin performansının çoklu bağlantının şiddeti ve bileşen sayısına bağlı olarak değiştiği söylenebilir. Buna karşılık r-k sınıf tahmin edicisi ile elde edilen GCV ve HKO istatistikleri bileşen sayısı ve çoklu bağlantının şiddetinden daha az etkilenmektedir. Ayrıca r-k sınıf tahmin edicisi r-d sınıf tahmin edicisinden sonra en küçük GCV istatistiklerine sahiptir. Dolayısı ile r-k sınıf tahmin edicisi ile hem iyi hem de istikrarlı kestirim ve tahminler elde edilmektedir. Ampirik çalışmalarda, tek amaç kestirim olduğunda, r-d sınıf tahmin edicisinin kullanımı önerilebilir. Ancak istenilen hem iyi kestirim hem de iyi parametre tahminleri elde etmek ise, tercih r-k sınıf tahmin edicisinden yana olmalıdır.

Kaynaklar

ALLEN, D. M. (1971), Mean Square Error of Prediction as a Criteration for selection of Variables, Technometrics, 13: 469-475.

ALLEN, D. M. (1974), The Relationship Between Variable Selection and Data Augmentation and a Method for Prediction, Technometrics, 16: 125-127.

BAYE, M. R. ve PARKER D. F. (1984), Combining Ridge and Principal Component Regression: A Money Demand Illustration, Comm. Statist. Theory Methods, 13 (2): 197-205.

GOLUB,G. H., HEATH, M. ve WAHBA, G. (1979), Generalized Cross-Validation as a Method for Choosing a Good Ridge Parameter. Technometrics, 21: 215-223

GÜLER, H. ve KAÇIRANLAR, S. (2009), A Comparision of Mixed and Ridge Estimators of Linear Models, Communications in Statistics – Simulation and Computation, 38: 368-401.

HOERL, A. E ve KENNARD, R. W. (1970), Ridge regression: biased estimation for nonorthogonal problems, Technometrics, 12 (1): 55-76.

KAÇIRANLAR, S. VE SAKALLIOĞLU S. (2001), Combining the Liu Estimator and the Principal Component Regression Estimator, Communications in Statistics – Theory and Methods, 30 (12): 2699-2705.

KIBRIA, B.M.G. (2003), Performance of Some New Ridge Regression Estimators, Communications in Statistics – Simulation and Computation, 32: 419-435

LIU,K. (1993), A New Class of Biased Estimate in Linear Regression. Communications in Statistics – Theory and Methods., 22 (2): 393-402

MALLOWS, C.L. (1973a), Data Analysis in a Regression Context. University of Kentucky Conference on Regression with a Large Number of Predictor Variables, Department of Statistics, University of Kentucky.

MALLOWS, C.L. (1973b), Some Comments on C_p. Technometrics, 15: 661-675

MASSY W.F. (1965), "Principal components regression in exploratory Statistical Research. " Journal of the American Statistical Association, 60: 234-256.

McDONALD, G.C. ve GALARNEAU, D.I. (1975), A Monte Carlo Evaluation of Some Ridge-Type Estimators, Journal of the American Statistical Association, 30: 407-416.

MONTGOMERY, D.C. ve FRIEDMAN, D.J. (1993), Prediction Using Regression Models with Multicollinear Predictor Variables. IIE Transactions, 25 (3): 73-85

NEWHOUSE, J.P. ve OMAN, S.D. (1971), An Evaluation of Ridge Estimators, Rand Corporation R-716-PR.

NOMURA, M. ve OHKUBO, T. (1985), A Note on Combining Ridge and Principal Component Regression, Communications in Statistics – Theory and Methods, 14: 2489-2493.

ÖZBEY, F. ve KAÇIRANLAR, S. (2010), Bazı Yanlı Tahmin Edicilerin Kestirim Performanslarının Karşılaştırılması. X. Ekonometri ve İstatatistik Sempozyumu, Sakarya.

ÖZKALE, M. R. ve KAÇIRANLAR, S. (2008), Comparisons of the r-k Class Estimator to the Ordinary Least Squares Estimator under the Pitman's Closeness Criteration, Statistical Papers, 49: 503-512

SARKAR, N. (1989).,Comparisions Among Some Estimators in Misspecified Linear Models with Multicollinearity, Ann. Inst. Statist. Math., 41: 717-724.

SARKAR, N. (1996). Mean Square Error Matrix Comparision of Some Estimators in Linear Regressions with Multicollinearity, Statistics and Probability Letters, 30: 133-138

STEIN C. 1956. Inadmissibility of usual estimator for the mean of a multivariate normal distribution. Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 1: 197-206.