

***r-k-d Sınıf Tahmin Edici ile Tahmin ve Kestirim******Estimation and Prediction with r-k-d Class Estimator***Fela ÖZBEY<sup>1</sup>**ÖZET**

Çoklu lineer regresyon modellerinde çoklu iç ilişki olması durumunda, EKK tahmin edicisi yansızlığını korur, ancak varyansı büyük ve mutlak değerce büyük parametre tahminleri vermektedir. Bu nedenle  $X'X$  matrisinin kötü koşulluluğunu ortadan kaldırmaya ve varyansı küçültmeye yönelik çeşitli yanlı tahmin ediciler geliştirilmiş ve önerilmiştir. r-k-d sınıf tahmin edici bunlardan biridir. Bu tahmin edici 7 farklı tahmin ediciyi özel durumlar olarak ihtiva ettiğinden oldukça esnek ve geniş kapsamlı bir tahmin edicidir. Bu çalışmada, r-k-d sınıf tahmin edicinin ampirik çalışmalarda ne denli kullanışlı olduğu gösterilmiştir.

**Anahtar Kelimeler:** Çoklu İç İlişki, r-k-d Sınıf Tahmin Edici, Tahmin, Kestirim, PRESS

**ABSTRACT**

In the presence of multicollinearity the OLS estimator remains unbiased, but the variances and the absolute values of the parameter estimations are tend to be large. Various biased estimators are developed and proposed in order to eliminate ill conditioning of  $X'X$  matrix and reduce the variance. One of them is r-k-d class estimator. This estimator is quite flexible and extensive because it contains 7 different estimators as special cases. The main aim of this paper is to show how usable this estimator is in empirical studies.

**Keywords:** Multicollinearity, r-k-d class estimator, estimation, prediction, PRESS

<sup>1</sup>Araş.Gör.Dr., Çukurova Üniversitesi, İİBF, Ekonometri Bölümü, [fozbey@cu.edu.tr](mailto:fozbey@cu.edu.tr)

## 1.Giriş

Çoklu iç ilişki olması durumunda,  $X'X$  matrisinin kötü koşulluluğunu ortadan kaldırmaya ve varyansı küçültmeye yönelik çeşitli yanlı tahmin ediciler önerilmiştir. Bunlardan en yaygın olarak bilineni Hoerl ve Kennard (1970)'in önerdiği ridge tahmin edicisidir.

Baye ve Parker (1984) temel bileşenler tahmin edicisiyle ridge tahmin edicisini birleştirerek oluşturdukları ve r-k sınıf tahmin edici olarak adlandırdıkları yeni tahmin edici ile TB tahmin edicisine göre daha iyi sonuçlar alınabileceğini göstermişlerdir.

Liu (1993), ridge tahmin edicisi ve Stein tahmin edicisini (Stein,1956) birleştirerek yeni bir tahmin edici önermiştir. Daha sonra bu tahmin edici Akdeniz ve Kaçıranlar (1995) tarafından Liu tahmin edicisi olarak adlandırılmıştır.

Kaçıranlar ve Sakallıoğlu (2001) temel bileşenler tahmin edicisi ve Liu tahmin edicisini birleştirerek r-d sınıf tahmin edicisini önermişlerdir ve bu tahmin edicinin Liu, en küçük kareler ve temel bileşenler tahmin edicilerinden skaler Hata Kareleri Ortalaması (HKO) bakımından daha iyi olduğunu göstermişlerdir.

Liu (2003), parametre tahminlerini daha da iyileştirmek amacıyla Liu tahmin edicisinin kısıtlarını gevşeterek Liu tipi tahmin ediciyi önermiştir ve belli koşullar altında, ridge tahmin edicisine göre daha iyi sonuçlar alınabileceğini göstermiştir.

İnan (2011), EKK, TB, Ridge, Liu, Liu Tipi, r-k sınıf ve r-d sınıf tahmin edicilerini, oldukça esnek ve geniş kapsamlı bir tahmin edici çatısı altında toplayan r-k-d sınıf tahmin ediciyi önermiştir.

Daha önce, yukarıda adı geçen bu tahmin edicilerin bazıları birbirleriyle farklı çalışmalarda değişik açılardan karşılaştırılmışlardır. Ancak bu karşılaştırmalarda, işlemlerde kolaylık sağlaması veya ispatların yapılabilirliğinin sağlanması gibi nedenlerle tahmin edicilerin  $r$ ,  $k$  ve  $d$  değerlerinin eşit olduğu varsayılmıştır. Oysa ki, her bir tahmin edicinin minimize edilecek fonksiyonu (HKO, PRESS, vs.) farklı olduğundan, fonksiyonu minimize eden  $r$ ,  $k$  ve  $d$  değerleri de farklı olabilecektir. r-k-d sınıf tahmin edicinin işlevlerinden biri de kapsadığı 7 farklı tahmin ediciyi, kendilerine ait optimum  $r$ ,  $k$  ve  $d$  değerleri ile karşılaştırarak, kullanılan veri seti için en iyi sonuçları veren seçmektir. Diğer bir işlevi de, doğal olarak, ihtiva ettiği bu 7 tahmin edicilerin hepsinden daha iyi sonuçlar veren  $r$ ,  $k$  ve  $d$  kombinasyonu var ise, onu ortaya koymaktır.

## 2. r-k-d Sınıf Tahmin Edici

$$y = X\beta + \varepsilon \quad (4.1)$$

şeklindeki genel lineer regresyon modelini göz önüne alalım. Burada  $y$ ,  $n \times 1$  boyutlu cevap değişken vektörü,  $X$ ,  $n \times p$  boyutlu açıklayıcı değişkenler matrisi,  $\beta$ ,  $p \times 1$  boyutlu bilinmeyen parametreler vektörü ve  $\varepsilon$ ,  $n \times 1$  boyutlu,  $iid(0, \sigma^2 I_n)$  varsayımını sağlayan rezidüler vektörüdür. Değişkenlerin,  $XX$  açıklayıcı değişkenler arasındaki korelasyonlar matrisi ve  $X'y$  açıklayıcı değişkenler ile cevap değişken arasındaki korelasyonlar vektörü olacak şekilde standartlaştırıldığı varsayılacaktır.

$T = [t_1, t_2, \dots, t_p]$  matrisi  $XX$  matrisinin özvektörlerinden oluşan  $n \times p$  boyutlu matris olsun.  $T_r$  matrisi,  $T$  matrisinden  $p-r$  kolon silinerek elde edilen matris olsun. Böylece  $T_r'X'XT_r = \Lambda_r = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_r)$  olacaktır. Burada  $\lambda_1, \lambda_2, \dots, \lambda_r$ ,  $XX$  matrisinin en büyük ilk  $r$  tane özdeğeridir. Ayrıca,  $\alpha = T'\beta$  şeklinde tanımlansın.(2.1)'deki  $\beta$ 'nin r-k-d sınıf tahmin edicisi

$$\begin{aligned}\hat{\beta}_{r-k-d} &= T_r(T_r'X'XT_r + kI)^{-1}(T_r'X'y - dT_r'\hat{\beta}_r) \\ &= T_r(T_r'X'XT_r + kI)^{-1}(T_r'X'XT_r - dI)\hat{\alpha}_r\end{aligned}\quad (4.2)$$

dir. Burada  $r \leq p$ ,  $-\infty < k < \infty$ ,  $-\infty < d < \infty$  ve  $\hat{\beta}_r = T_r(T_r'X'XT_r)^{-1}T_r'X'y$  temel bileşenler tahmin edicisidir. Bu haliyle, r-k-d sınıf tahmin edicisi, özel haller olarak kendi içinde 7 farklı tahmin ediciyi kapsayan oldukça genel bir tahmin edicidir. Şöyle ki,

- 1)  $r = p$  olduğunda Liu tipi tahmin edici
- 2)  $r = p$  ve  $k = 1$  olduğunda Liu tahmin edicisi
- 3)  $r = p$  ve  $d = 0$  olduğunda Ridge tahmin edicisi
- 4)  $r = p$  ve  $k = -d$  olduğunda EKK tahmin edicisi
- 5)  $r < p$  ve  $k = 1$  olduğunda  $r-d$  sınıf tahmin edicisi
- 6)  $r < p$  ve  $d = 0$  olduğunda  $r-k$  sınıf tahmin edicisi
- 7)  $r < p$  ve  $k = -d$  olduğunda TB tahmin edicisi

haline dönüşecektir.

Bu tahmin edicinin en önemli kullanım amaçlarından biri, elimizdeki veri setiyle (2.1) ile verilen modeli tahmin etmek istediğimizde, yukarıda adı geçen 7 farklı tahmin ediciyi performans açısından karşılaştırmak ve o veri seti için en iyi sonucu veren tahmin ediciyi seçmektir. Daha iyi sonuçların elde edilmesi mümkün kılan  $r$ ,  $k$  ve  $d$  kombinasyonu var ise de bu kombinasyonu bulmaktır.

r-k-d sınıf tahmin edicisinin yanlılık parametreleri  $k$  ve  $d$  amacımız doğrultusunda farklı şekillerde seçilebilir. Amacımız en iyi parametre tahminlerini elde etmek olduğunda HKO'yu minimize edecek şekilde  $k$  ve  $d$ 'yi seçebiliriz. İşlem kolaylığı açısından (2.1)'deki modeli

$$y = Z\alpha + \varepsilon \quad (4.3)$$

şeklinde ifade edersek r-k-d sınıf tahmin edicisinin skaler HKO'su

$$SHKO(\hat{\beta}_{r-k-d}) = \sigma^2 \left[ \sum_{i=1}^r \frac{(\lambda_i - d)^2}{\lambda_i (\lambda_i + k)^2} \right] + \sum_{i=1}^r \frac{\alpha_i^2 (d+k)^2}{(\lambda_i + k)^2} + \sum_{i=r+1}^p \alpha_i^2 \quad (4.4)$$

dir. Bunu minimize eden  $d$  değeri

$$\hat{d}_{MSE} = \sum_{i=1}^r \frac{\lambda_i (\sigma^2 - \alpha_i^2 k)}{\lambda_i \alpha_i^2 + \sigma^2} \quad (4.5)$$

ve  $k$  değeri

$$\hat{k}_{MSE} = \sum_{i=1}^r \frac{\sigma^2 (\lambda_i - d)}{\lambda_i} - rd \quad (4.6)$$

dir. Burada  $\hat{d}$ 'nin değeri  $\hat{k}$ 'nin değerine;  $\hat{k}$ 'nin değeri de  $\hat{d}$ 'nin değerine bağlı olduğundan, bu değerler, belli bir başlangıç değeri seçilerek iterasyonla hesaplanır.

Bileşen sayısının seçiminde ise Baye ve Parker (1984)'ın önerdiği şekilde, tüm bileşenlerle başlanarak ve tek tek bileşenler azaltılarak ve en iyi sonucu veren  $r$ ,  $k$  ve  $d$  kombinasyonu araştırılmalıdır.

Amacımız  $y$ 'nin bilinmeyen değerlerini en iyi şekilde kestirmek olduğunda ise  $k$  ve  $d$ 'nin değerlerini Montgomery ve Friedman (1993)'da önerildiği gibi PRESS istatistiğini minimum yapacak şekilde de seçebiliriz.

### 3. Sayısal Örnek

Kullanılacak ikinci veri seti Montgomery ve Friedman(1993)'dan alınan Mil Başına Benzin Tüketimi verisidir. Burada  $Y$  arabaların mil başına harcadıkları benzin,  $X_1$  toplam silindir hacmi,  $X_2$  beygir gücü,  $X_3$  tork,  $X_4$  kompresyon oranı,  $X_5$  arka aks oranı,  $X_6$  karbüratör büyüklüğü,  $X_7$  vitesi sayısı,  $X_8$  arabanın uzunluğu,  $X_9$  arabanın genişliği,  $X_{10}$  arabanın ağırlığı ve  $X_{11}$  vites tipidir. Bu veri seti Montgomery ve Friedman tarafından DUPLEX algoritması kullanılarak ilki 20 tane tahmin amaçlı veri ve diğeri 10 tane kestirim amaçlı veriden oluşmak üzere iki veri setine ayrılmıştır.

**Tablo 3.1a.** Tahmin amaçlı veri seti:

Y	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11
19.7	318	140	255	8.5	2.71	2	3	215.3	76.3	4370	1
11.2	440	215	330	8.2	2.88	4	3	184.5	69	4215	1
18.3	351	143	255	8	3	2	3	199.9	74	3890	1
21.5	360	180	290	8.4	2.45	2	3	214.2	76.3	4250	1
20.3	140	83	109	8.4	3.4	2	4	168.8	69.4	2700	0
36.5	85.3	80	83	8.5	3.89	2	4	160.6	62.2	2009	0
18.9	350	165	260	8	2.56	4	3	200.3	69.9	3910	1
30.4	96.9	75	83	9	4.3	2	5	165.2	65	2320	0
13.9	351	148	243	8	3.25	2	3	215.5	78.5	4540	1
14.9	440	215	330	8.2	2.71	4	3	231	79.7	5185	1
21.5	171	109	146	8.2	3.22	2	4	170.4	66.9	2655	0
17.8	302	129	220	8	3	2	3	199.9	74	3890	1
17.8	350	155	250	8.5	3.08	4	3	196.7	72.2	3910	1
16.4	318	145	255	8.5	2.45	2	3	197.6	71	3660	1
23.5	231	110	175	8	2.56	2	3	179.3	65.4	3050	1
31.9	96.9	75	83	9	4.3	2	5	165.2	61.8	2275	0
14.4	500	190	360	8.5	2.73	4	3	224.1	79.8	5290	1
22.1	231	110	175	8	2.56	2	3	179.3	65.4	3020	1
17.0	350	170	275	8.5	2.56	4	3	199.6	72.9	2860	1
20.0	250	105	185	8.25	2.73	1	3	196.7	72.2	3510	1

**Tablo 3.1b.** Kestirim amaçlı veri seti

Y	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11
20.1	225	95	170	8.4	2.8	1	3	194	71.8	3365	0
34.7	89.7	70	81	8.2	3.9	2	4	155.7	64	1905	0
16.5	350	155	250	8.5	3.1	4	3	195.4	74.4	3885	1
19.7	258	110	195	8	3.1	1	3	171.5	77	3375	1
13.3	460	223	366	8	3	4	3	228	79.8	5430	1
13.8	360	195	295	8.3	3.2	4	3	209.3	77.4	4215	1
21.5	262	110	200	8.5	2.6	2	3	179.3	65.4	3180	1
16.5	350	165	255	8.5	2.7	4	3	185.2	69	3660	1
13.3	351	148	243	8	3.3	2	3	216.1	78.5	4715	1
23.9	133.6	96	120	8.4	3.9	2	5	171.5	63.4	2535	0

Hesaplamalara başlamadan önce değişkenler standartlaştırılmıştır. Tahmin amaçlı veri setinin  $XX$  matrisinin özdeğerleri  $\lambda_1 = 7.8929$ ,  $\lambda_2 = 1.4237$ ,  $\lambda_3 = 0.8371$ ,  $\lambda_4 = 0.3934$ ,  $\lambda_5 = 0.1652$ ,  $\lambda_6 = 0.1373$ ,  $\lambda_7 = 0.0650$ ,  $\lambda_8 = 0.0472$ ,  $\lambda_9 = 0.0265$ ,  $\lambda_{10} = 0.0108$ ,  $\lambda_{11} = 0.0009$  dur. Koşul sayısı  $\kappa = \sqrt{\frac{\lambda_1}{\lambda_{11}}} \cong 93.65$  çoklu doğrusal bağlantının olduğunu gösterir.

İlk olarak, bu veri setinin tahmin amaçlı 20 gözlemlik değerleri kullanılarak (2.1)'deki model için EKK, TB, Ridge, Liu, Liu tipi, r-k sınıf, r-d sınıf ve r-k-d sınıf tahmin ediciler kullanılarak her biri için HKO istatistiği kriter alınarak optimum  $r$ ,  $k$  ve  $d$  kombinasyonları bulunmuştur. Sonuçlar Tablo 3.2.'de verilmiştir.

**Tablo 3.2.** Mil Başına Benzin Tüketimi veri seti için TB, r-k sınıf, r-d sınıf ve r-k-d sınıf tahmin edicilerinin her bir  $r$  seçimine karşılık gelen yanlılık parametreleri ve HKO değerleri

r	TB	r-d sınıf		r-k sınıf		r-k-d sınıf		
	HKO	d	HKO	k	HKO	k	d	HKO
1	10.3177	-0.8700	10.3177	0.1171	10.3177	0.0579	0.0583	10.3177
2	10.3250	0.2993	10.3207	1.0689	10.3209	0.5372	0.5560	10.3202
3	10.3356	0.4594	10.3204	1.4224	10.3215	0.0357	0.7613	10.3194
4	10.3297	-0.2176	10.3060	0.6176	10.3069	1.8326	-0.6162	10.3056
5	10.3750	-0.1520	10.3014	0.5821	10.3020	1.2254	-0.2276	10.3013
6	10.4086	-0.2135	10.2855	0.4060	10.2878	1.5400	-0.4017	10.2851
7	10.4340	-0.3351	10.2253	0.1761	10.2429	2.4350	-0.9194	10.2232
8	9.9228	-0.6069	9.7129	0.0288	9.7679	5.7703	-3.6084	9.7055
9	8.4509	-0.7358	8.2041	0.0089	8.2876	10.3952	-7.7187	8.1947
10	5.2500*	-0.7765	4.8184*	0.0033	4.9529	13.4315	10.4699	4.8081
	EKK	Liu		Ridge		Liu Tipi		
	HKO	d	HKO	k	HKO	k	d	HKO
11	14.6368	-0.4058	5.9931	0.0033	4.2919*	0.0033	0.0000	4.2919*

\*Tahmin edicinin minimum HKO değeridir.

Yukarıdaki tabloda da görüldüğü gibi TB ve r-d sınıf tahmin edicileri için optimum  $r$  değeri 10, r-k ve r-k-d sınıf tahmin edicileri için ise optimum  $r$  değeri 11'dir. Veri seti için bu tahmin edicilerle elde edilebilen en küçük HKO değeri 4.2919'dur. Bu değer,  $r = 11$ ,  $k = 0.0033$  ve  $d = 0$  kombinasyonu ile elde edilmektedir. Bu da bize bu veri seti için, yukarıdaki tahmin ediciler arasında HKO kriterine göre en iyi sonuçların Ridge tahmin ediciyle elde edildiğini göstermektedir.

Tablo 3.3'te Görüldüğü gibi TB, r-k sınıf ve r-d sınıf tahmin edicileri için optimum  $r$  değeri 1, r-k-d sınıf tahmin edicisi için ise optimum  $r$  değeri 5'tir. Veri seti için bu tahmin edicilerle elde edilebilen en küçük PRESS değeri 0.2571'dir. Bu değer,  $r = 5$ ,  $k = 0$  ve  $d = 0.8534$  kombinasyonu ile elde edilmektedir. Bu da bize bu veri seti için, yukarıdaki tahmin ediciler arasında PRESS kriterine göre en iyi sonuçların r-k-d sınıf tahmin ediciyle elde edildiğini göstermektedir.

**Tablo 3.3.** Mil Başına Benzin Tüketimi veri seti için TB, r-k sınıf, r-d sınıf ve r-k-d sınıf tahmin edicilerinin her bir  $r$  seçimine karşılık gelen yanlılık parametreleri ve PRESS değerleri

r	TB	r-d sınıf		r-k sınıf		r-k-d sınıf		
	PRESS	d	PRESS	k	PRESS	k	d	PRESS
1	0.3100*	0.9409	0.3100*	0.0528	0.3100*	0.0263	0.0264	0.3100
2	0.3177	0.7942	0.3171	0.2121	0.3168	0.1105	0.1123	0.3168
3	0.3558	-0.0448	0.3408	0.6957	0.3393	0	1.4302	0.3169
4	0.3738	-0.1233	0.3407	0.7666	0.3400	0.1600	0.5555	0.3317
5	0.4399	-0.7468	0.3327	0.9484	0.3466	0	0.8534	0.2571*
6	0.5886	-0.9787	0.3186	1.0581	0.3517	0.0360	0.6522	0.2648
7	0.5874	-1.1388	0.3145	1.0539	0.3518	0.2813	0.8465	0.2855
8	0.6425	-0.2576	0.3457	1.0026	0.3493	0.7488	0.2898	0.3442
9	0.7626	-0.1618	0.3469	0.9924	0.3493	0.9107	0.1586	0.3467
10	0.6359	0.0736	0.3484	0.9783	0.3488	0.9948	-0.0727	0.3484
	EKK	Liu		Ridge		Liu Tipi		
	PRESS	d	PRESS	k	PRESS	k	d	PRESS
11	0.7102	0.0633	0.3486	0.9784	0.3488	0.9966	-0.0641	0.3486

\*Tahmin edicinin minimum PRESS değeridir.

**Tablo 3.4a.** Mil Başına Benzin Tüketimi veri seti için TB, r-k sınıf, r-d sınıf ve r-k-d sınıf tahmin edicileri PRESS istatistiğine dayanılarak seçilen optimum  $r$ ,  $k$  ve  $d$  değerler ile elde edilen parametre tahminleri

	TB	r-k sınıf	r-d sınıf	r-k-d sınıf
r	1	1	1	5
k		0.0528		0
d			0.9409	0.8534
$\hat{\beta}_1$	-0.1057	-0.1050	-0.1050	-0.1822
$\hat{\beta}_2$	-0.0988	-0.0982	-0.0982	-0.0385
$\hat{\beta}_3$	-0.1052	-0.1045	-0.1045	-0.1144
$\hat{\beta}_4$	0.0558	0.0554	0.0554	-0.1133
$\hat{\beta}_5$	0.0882	0.0876	0.0876	-0.1896
$\hat{\beta}_6$	-0.0584	-0.0580	-0.0580	0.0396
$\hat{\beta}_7$	0.0963	0.0957	0.0957	0.1536
$\hat{\beta}_8$	-0.0986	-0.0980	-0.0980	-0.0369
$\hat{\beta}_9$	-0.0917	-0.0911	-0.0911	0.1584
$\hat{\beta}_{10}$	-0.0996	-0.0989	-0.0989	-0.1606
$\hat{\beta}_{11}$	-0.0967	-0.0961	-0.0961	-0.5510
$R^2$	0.7341	0.7341	0.7341	0.5893
PRESS	0.3100	0.3100	0.3100	0.2571
RSSP	0.1968	0.1977	0.1977	0.2960

Bir sonraki aşama olarak her tahmin edicinin optimum  $r$ ,  $k$  ve  $d$  değerleri ile modelin parametreleri tahmin edildi. Daha sonra bu parametre tahminleri kullanılarak veri setindeki kestirimi amaçlanan  $y$  değerleri için kestirimler elde edildi ve elde edilen kestirimler ile gerçek değerler arasındaki farkların kareleri toplamı hesaplandı. Sonuçlar Tablo 3.4a., Tablo 3.4b., Tablo 3.5a. ve Tablo 3.5b.'de verilmiştir.

**Tablo 3.4b.** Mil Başına Benzin Tüketimi veri seti için EKK, Ridge, Liu ve Liu Tipi tahmin edicileri PRESS istatistiğine dayanılarak seçilen optimum  $k$  ve  $d$  değerler ile elde edilen parametre tahminleri

	EKK	Ridge	Liu	Liu tipi
$k$		0.9784		0.9966
$d$			0.0633	-0.0641
$\hat{\beta}_1$	-1.5741	-0.1203	-0.2116	-0.2133
$\hat{\beta}_2$	-0.4580	-0.0920	-0.1149	-0.1154
$\hat{\beta}_3$	1.4987	-0.1032	-0.0014	0.0003
$\hat{\beta}_4$	0.0839	0.0960	0.0945	0.0946
$\hat{\beta}_5$	0.6251	0.0782	0.1126	0.1132
$\hat{\beta}_6$	-0.0240	-0.0560	-0.0538	-0.0538
$\hat{\beta}_7$	-0.6516	0.0735	0.0277	0.0269
$\hat{\beta}_8$	1.5734	-0.0275	0.0729	0.0748
$\hat{\beta}_9$	-1.3603	-0.1115	-0.1897	-0.1912
$\hat{\beta}_{10}$	-0.1897	-0.0900	-0.0961	-0.0963
$\hat{\beta}_{11}$	-0.5389	-0.0576	-0.0883	-0.0888
$R^2$	0.9020	0.7437	0.7625	0.7629
PRESS	0.7102	0.3488	0.3486	0.3486
RSSP	1.1452	0.2399	0.2424	0.2425

Tablo 3.4a. ve Tablo 3.4b.'de RSSP ile verilen istatistik, gerçekleşen  $y$  değerleri ile kestirilen  $y$  değerleri arasındaki farkların kareleri toplamıdır.



**Tablo 3.5a.** Mil Başına Benzin Tüketimi veri seti için TB, r-k sınıf, r-d sınıf ve r-k-d sınıf tahmin ediciler ile elde edilen kestirimler

n	y	$\hat{y}$			
		TB	r-k sınıf	r-d sınıf	r-k-d sınıf
1	0.0394	0.1152	0.1145	0.1145	0.3305
2	0.7865	0.4407	0.4377	0.4377	0.4487
3	-0.1448	-0.1143	-0.1135	-0.1135	-0.2042
4	0.0189	0.0099	0.0099	0.0099	0.0006
5	-0.3086	-0.4176	-0.4148	-0.4148	-0.2860
6	-0.2830	-0.2198	-0.2184	-0.2184	-0.2194
7	0.1110	0.0531	0.0527	0.0527	-0.1023
8	-0.1448	-0.1008	-0.1001	-0.1001	-0.1807
9	-0.3086	-0.1835	-0.1823	-0.1823	-0.1814
10	0.2339	0.4171	0.4143	0.4143	0.3943
$RSSP = \sum_{i=1}^{10} (y_i - \hat{y}_i)^2$		0.1968	0.1977	0.1977	0.2960

**Tablo 3.5b.** Mil Başına Benzin Tüketimi veri seti için EKK, Ridge, Liu ve Liu tipi tahmin ediciler ile elde edilen kestirimler

n	y	$\hat{y}$			
		EKK	Ridge	Liu	Liu tipi
1	0.0394	0.4494	0.1121	0.1331	0.1335
2	0.7865	0.5400	0.3734	0.3833	0.3836
3	-0.1448	-0.2648	-0.0860	-0.0973	-0.0975
4	0.0189	-0.7874	-0.0231	-0.0709	-0.0719
5	-0.3086	-0.1291	-0.3869	-0.3695	-0.3694
6	-0.2830	-0.0588	-0.1904	-0.1817	-0.1817
7	0.1110	0.0836	0.0753	0.0753	0.0754
8	-0.1448	-0.2858	-0.0748	-0.0882	-0.0884
9	-0.3086	-0.1134	-0.1696	-0.1656	-0.1656
10	0.2339	0.5663	0.3699	0.3815	0.3819
$RSSP = \sum_{i=1}^{10} (y_i - \hat{y}_i)^2$		1.1452	0.2399	0.2424	0.2425

Her ne kadar r-k-d sınıf tahmin edici ile elde edilen RSSP değeri diğer tahmin edicilerle elde edilen RSSP değerlerinden daha büyük ise de, birbirine en yakın PRESS ve RSSP değerleri r-k-d sınıf tahmin edicisiyle elde edilmiştir. Bu da bize r-k-d tahmin edicisiyle, diğerlerine göre daha stabil parametre tahminleri elde edilmiş olduğunu göstermektedir.

#### 4.Sonuç

Çoklu lineer regresyon modellerinde çoklu iç ilişki olması durumunda, EKK tahmin edicisi yansızlığını korur, ancak varyansı büyür ve mutlak değerce büyük parametre tahminleri vermektedir. Bu nedenle  $X'X$  matrisinin kötü koşulluluğunu ortadan kaldırmaya ve varyansı küçültmeye yönelik çeşitli yanlı tahmin ediciler geliştirilmiş ve önerilmiştir. Bu tahmin ediciler arasında seçim yapılırken ise çeşitli kriterlere dayanılarak karşılaştırmalar yapılmaktadır. Ancak bu karşılaştırmalarda, işlemlerde kolaylık sağlaması veya ispatların yapılabilirliğinin sağlanması gibi nedenlerle tahmin edicilerin  $r$ ,  $k$  ve  $d$  değerlerinin eşit olduğu varsayılmıştır. Oysa ki, her bir tahmin edicinin minimize edilecek fonksiyonu (HKO, PRESS, vs.) farklı olduğundan, fonksiyonu minimize eden  $r$ ,  $k$  ve  $d$  değerleri de farklı olabilecektir. r-k-d sınıf tahmin edicinin işlevlerinden biri de kapsadığı 7 farklı tahmin ediciyi, kendilerine ait optimum  $r$ ,  $k$  ve  $d$  değerleri ile karşılaştırarak, kullanılan veri seti için en iyi sonuçları veren seçmektir. Diğer bir işlevi de, doğal olarak, ihtiva ettiği bu 7 tahmin edicilerin hepsinden daha iyi sonuçlar veren  $r$ ,  $k$  ve  $d$  kombinasyonu var ise, onu ortaya koymaktır.

En iyi parametre tahminlerini veren tahmin edici seçilmek istendiğinde en uygun kriter hata kareler ortalaması (HKO)'dır. İyi bir kestirim performansı istendiğinde ise PRESS istatistiği minimize edilebilir. Burada her iki kritere göre r-k-d sınıf tahmin edici ve onun ihtiva ettiği diğer 7 tahmin edici ile optimum  $r$ ,  $k$  ve  $d$  kombinasyonları elde edilmiş ve performansları karşılaştırılmıştır. Uygulamanın ilk kısmında açıkça görülmektedir ki, bir veri setini kullanarak (2.1)'deki model tahmin edilmek istendiğinde EKK, TB, Ridge, Liu, Liu tipi, r-k sınıf, r-d sınıf ve r-k-d sınıf tahmin ediciler ayrı ayrı kullanılarak tahminler elde etmeye ve bunların performanslarını karşılaştırıp en iyisini seçmeye çalışmak yerine, modeli sadece r-k-d sınıf tahmin edicisiyle tahmin etmek yeterlidir. Çünkü r-k-d sınıf tahmin edicisi, bu 7 tahmin edici arasından optimum sonucu veren tahmin ediciyi zaten önermektedir.

Uygulamanın ikinci aşamasında ise görülmektedir ki, adı geçen 7 tahmin ediciden daha optimum sonuçların elde edilmesi mümkünse r-k-d sınıf tahmin edici bize bunu sağlayan  $r$ ,  $k$  ve  $d$  kombinasyonunu önermektedir.

#### KAYNAKLAR

- BAYE, M. R., and PARKER, D. F., 1984. Combining Ridge and Principal Component Regression: A Money Demand Illustration. *Comm. Statist. Theory Methods*, 13, 2, 197-205.
- HOERL, A. E., and KENNARD, R. W., 1970. Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics*, 12, 1, 55-67.

- İNAN, D., 2011. Combining the Liu-Type Estimator and the Principal Component Regression Estimator: A Comparative Simulation Study. European Young Statisticians Meeting, 17th EYMS. Lisbon, Portugal.
- KAÇIRANLAR, S., and SAKALLIOĞLU, S., 2001. Combining the Liu Estimator and the Principal Component Regression Estimator. *Comm. Statist. Theory Methods*, 30, 12, 2699-2705.
- LIU, K., 1993. A New Class of Biased Estimate in Linear Regression. *Comm. Statist. Theory Methods*, 22, 2, 393-402.
- LIU, K., 2003. Using Liu-Type Estimator to Combat Collinearity. *Comm. Statist. Theory Methods*, 32, 5, 1009-1020.
- MONTGOMERY, D. C., and FRIEDMAN D. J., 1993. Prediction Using Regression Models with Multicollinear Predictor Variables. *IIE Transactions*, 25 (3), 73-85.
- 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.