ETLA ELINKEINOELÄMÄN TUTKIMUSLAITOS
THE RESEARCH INSTITUTE OF THE FINNISH ECONOMY
Lönnrotinkatu 4 B, 00120 Helsinki 12, Finland, tel. 601322

Keskusteluaiheita
Discussion papers

TIMO TERÄSVIRTA

RESTRICTED SUPERIORITY OF LINEAR HOMOGENEOUS ESTIMATORS OVER ORDINARY LEAST SQUARES*

No. 83 4.6.1981

* I am grateful to Ilkka Mellin for his helpful comments.

Abstract. In this paper the quadratic risk of a homogeneous linear estimator and the ordinary least squares estimator is compared in subsets of the parameter space and conditions for strong (matrix risk) and weak restricted superiority of homogeneous linear estimators over OLS in an ellipsoid are established. In particular, the restricted superiority of the minimax estimator of Kuks and Olman (1972) and the restricted least squares estimator over OLS are investigated, and it is shown that there always exists a minimax estimator superior to OLS in any arbitrary ellipsoid. Finally, the theoretical results of the paper are illustrated by an example.

Keywords: biased estimation; homogeneous linear estimator; minimax estimation; restricted least squares; restricted superiority; ridge regression.

1. Introduction

The idea of applying biased estimators to the estimation of parameters in linear models arose from the need to improve estimation when multicollinearity rendered ordinary least squares (OLS) estimates too uncertain. It was therefore natural to compare the performance of these estimators with the OLS estimator. Necessary and sufficient conditions based on quadratic loss and different loss matrices were constructed to indicate the superiority of various biased estimator over OLS at a certain point (β,σ^2) of the parameter space, cf. e.g. Swamy and Mehta (1978), Teräsvirta (1981a), Toro-Vizcarrondo and Wallace (1968), Wallace (1972) and Yancey et al. (1974).

The above-mentioned conditions relate to estimators which do not dominate OLS. There also exist estimators dominating OLS under quadratic loss and a suitable loss matrix, cf. e.g. Draper and Van Nostrand (1979), Judge and Bock (1978) and Sclove (1968) for James-Stein estimators, and Alam and Hawkes (1978) and Casella (1980) for ridge estimators.

Rao (1976) has derived necessary and sufficient conditions for an estimator to be admissible under quadratic loss and a positive definite loss matrix. Hoffmann (1977, 1980) has studied a more restrictive case, in which the regression coefficients were contained in an ellipsoid centred in the origin, and has discussed admissibility of estimators in that subset of the original parameter space.

In this paper the interest is focussed on the question of improving OLS in a subset of the original parameter space. As in Kuks and Olman (1972) and Hoffmann (1977), we choose an ellipsoid as our subset but proceed without reference to admissibility. This approach will give us further insight into situations in which certain biased estimators, not dominating OLS, do improve estimation as compared to OLS.

The outline of the paper is as follows: The general results concerning linear homogeneous estimators and strong restricted superiority are presented in Section 2 while the following section is related to weak restricted superiority. Sections 4 to 7 are devoted to the minimax estimator of Kuks and Olman (1972), which itself is a linear homogeneous estimator, and its special case, the restricted least squares estimator. The results are illustrated by way of an example in Section 8.

2. Restricted strong superiority

Assume a linear model

$$y = X\beta + \varepsilon, \ E\varepsilon = 0, \ \mathrm{cov}(\varepsilon) = \sigma^2 I \tag{2.1}$$

where y and ε are $n \times 1$ vectors, X is an $n \times p$ matrix of full rank p and uncorrelated with ε, and β is a $p \times 1$ parameter vector. Define a homogeneous linear estimator of β as $b_D = Dy$. One frequently applied criterion for superiority of b_D over the OLS estimator $b = (X'X)^{-1}X'y$ is the quadratic risk. Using it

b_D can be defined to be (strongly) superior to b at (β,σ^2) if and only if the difference

$$\Delta(b_D,\beta,A) = R(b,\beta,A) - R(b_D,\beta,A) \geq 0$$

for all $A \geq 0$, where

$$R(\tilde{b},\beta,A) = E(\tilde{b} - \beta)'A(\tilde{b} - \beta) = \mathrm{tr}\, A\, MSE(\tilde{b}) \qquad (2.2)$$

with

$$MSE(\tilde{b}) = E(\tilde{b} - \beta)(\tilde{b} - \beta)', \quad \tilde{b} = b_D, b.$$

In this paper, we need the following

Definition. *An estimator* $\tilde{b}_1$ *is (strongly) superior to* $\tilde{b}_2$ *in* $\mathcal{B}(\beta_0,T,d) = \mathcal{B} = \{\beta:(\beta-\beta_0)'T(\beta-\beta_0) \leq \sigma^2 d^{-1},\ T > 0,\ d > 0\}$ *if and only if*

$$R(\tilde{b}_2,\beta,A) - R(\tilde{b}_1,\beta,A) \geq 0 \qquad (2.3)$$

for all $\beta \in \mathcal{B}$ *and* $A \geq 0$.

Later on, the superiority thus defined will also be called strong restricted superiority of $\tilde{b}_1$ over $\tilde{b}_2$ in $\mathcal{B}$.

In order to investigate the restricted superiority of b_D over b, write

$$MSE(b_D) = \sigma^2 DD' + H\beta\beta'H' \qquad (2.4)$$

where $H = DX - I$. Choose temporarily $A = aa'$ where a is a non-zero p x 1 vector. Then

$$\Delta(b_D, \beta, A) = a'[\sigma^2(U - DD') - H\beta\beta'H']a$$

$$= \sigma^2 a'(U - DD')a - (a'H\beta)^2 \qquad (2.5)$$

where $U = [u_{ij}] = (X'X)^{-1}$. Approximating (2.5) from below in $\mathcal{B}$ yields, since $T > 0$,

$$\Delta(b_D, \beta, aa')$$

$$= \sigma^2 a'(U - DD')a - [a'HT^{-1/2}T^{1/2}(\beta - \beta_0) + a'H\beta_0]^2$$

$$\geq \sigma^2 a'(U - DD')a - [|a'HT^{-1/2}T^{1/2}(\beta - \beta_0)| + |a'H\beta_0|]^2$$

$$\geq \sigma^2 a'(U - DD')a - [(a'HT^{-1}H'a)^{1/2}\sigma d^{-1/2}$$

$$+ (\beta_0'H'H\beta_0)^{1/2}(a'a)^{1/2}]^2$$

$$\geq \sigma^2 a'(U - DD')a - r^2 a'a \qquad (2.6)$$

where

$$r = \lambda_{max}^{1/2}(HT^{-1}H')\sigma d^{-1/2} + (\beta_0'H'H\beta_0)^{1/2}$$

and $\lambda_{max}(Y)$ denotes the largest eigenvalue of Y. The second inequality in (2.6) follows from the minimization in $\mathcal{B}$ and the Cauchy-Schwarz inequality, while for the third one this inequality and a result in Rao (1973, p. 62) are needed. A necessary and sufficient condition for the last expression in (2.6) to be non-negative in $\mathcal{B}$ for all $a \neq 0$ is

$$\sigma^2(U - DD') - r^2 I \geq 0 \tag{2.7}$$

or, equivalently,

$$\lambda_{min}^{1/2}(U - DD') - [\lambda_{max}^{1/2}(HT^{-1}H')d^{-1/2} + (\beta_0'H'H\beta_0/\sigma^2)^{1/2}] \geq 0. \tag{2.8}$$

But then, if (2.6) holds in $\mathcal{B}$ for all $a \neq 0$, then $\Delta(b_D, \beta, A) \geq 0$ in $\mathcal{B}$ for all $A \geq 0$, cf. Bunke (1975). Note that (2.8) is a sufficient but not necessary condition for strong restricted superiority.

Since $T > 0$, and $H \neq 0$ if b_D is biased, a necessary condition for (2.7) to hold is that $U - DD' > 0$, and, even more, if $\beta_0 \neq 0$ we have to require that

$$\lambda_{min}(U - DD') \geq \beta_0'H'H\beta_0/\sigma^2.$$

If X'X and DD' do not have the same eigenvectors, it is generally unlikely that $U - DD' > 0$. In practice, however, the eigenvectors are often the same; this is the case for instance for the ridge estimator of Hoerl and Kennard (1970), the shrinkage estimator cb, $c > 0$ (Mayer and Willke, 1973), and the principal component estimator (Gunst and Mason, 1977). Of these, however, $U - DD' > 0$ does not hold for the principal component estimator.

On the other hand, $\lambda_{max}(HT^{-1}H')$ can on average be expected to grow with increasing bias so that a badly biased estimator b_D is not likely to satisfy (2.8). Increasing the size of $\mathcal{B}(\beta_0,T,d)$ by decreasing d naturally makes it less likely for (2.8) to hold, other things equal. An increase in the length of β_0 has a similar effect.

If $\mathcal{B}(\beta_0,T,d)$ is centred in the origin, i.e., $\beta_0 = 0$, (2.8) becomes

$$d^{-1} \leq \lambda_{min}(U - DD')/\lambda_{max}(HT^{-1}H') \qquad (2.9)$$

giving an explicit upper limit to the size of the ellipsoid. Note that (2.9) is now both a necessary and sufficient condition for strong restricted superiority. It does not depend on σ^2 so that the r.h.s. can be determined from data for given T. Nevertheless, the ellipsoids depend on σ^2, so that the variance has to be estimated if we want to get an idea of the β's for which b_D is superior to b in a particular application.

3. Restricted weak superiority

The strong criterion for restricted superiority can be replaced by weaker criteria if desired. If our goal is to predict y, then replacing A by X'X in (2.2) would be appropriate, cf. e.g. Wallace (1972). Then, following (2.3), we can define $\tilde{b}_1$ to be weakly superior to $\tilde{b}_2$ in $\mathcal{B}$ if and only if

$$\min_{\beta\in\mathcal{B}}[R(\tilde{b}_2,\beta,X'X) - R(\tilde{b}_1,\beta,X'X)] \geq 0. \tag{3.1}$$

Choosing $\tilde{b}_2 = b$, $\tilde{b}_1 = b_D$ and applying (2.3) with $A = X'X$ we obtain

$$\begin{aligned}
&\Delta(b_D,\beta,X'X) \\
&= \sigma^2 p - (\sigma^2 \mathrm{tr}X'XDD' + \beta'H'X'XH\beta) \\
&= \sigma^2(p - \mathrm{tr}X'XDD') - \beta'T^{1/2}T^{-1/2}H'X'XHT^{-1/2}T^{1/2}\beta.
\end{aligned} \tag{3.2}$$

Now, making use of the spectral decomposition of X'X and proceeding as above, a sufficient condition for (3.2) to be non-negative in $\mathcal{B}$ can be found. Since its form is slightly complicated and perhaps not very illustrative, we do not give it here but rather concentrate on the case $\beta_0 = 0$.

For $\beta\in\mathcal{B}_0 = \mathcal{B}(0,T,d)$ we then have

$$\Delta(b_D, \beta, X'X)$$

$$\geq \sigma^2 (p - \mathrm{tr}X'XDD') - \sigma^2 d^{-1} \lambda_{max}(T^{-1/2}H'X'XHT^{-1/2}) \quad (3.3)$$

cf. Rao (1973, p. 62). The r.h.s. of (3.3) is non-negative in B_0, if and only if

$$d^{-1} \leq \lambda_{max}^{-1}(T^{-\frac{1}{2}}H'X'XHT^{-\frac{1}{2}})(p - \mathrm{tr}X'XDD'). \quad (3.4)$$

A necessary condition for (3.4) to hold is $\mathrm{tr}X'XDD' \leq p$. This is of course a weaker condition than $U - DD' > 0$; it is for instance seen to be valid for the principal component estimator.

If the superiority comparison is based upon the mean square error as is often done in practice, then the inequality corresponding to (3.3) becomes

$$\Delta(b_D, \beta, I)$$

$$\geq \sigma^2 (\mathrm{tr}\, U - \mathrm{tr}\, DD') - \sigma^2 d^{-1} \lambda_{max}(T^{-\frac{1}{2}}H'HT^{-\frac{1}{2}}) . \quad (3.5)$$

The m.s.e. of b_D is thus smaller than that of b in B_0 if and only if

$$d^{-1} \leq \lambda_{max}^{-1}(T^{-\frac{1}{2}}H'HT^{-\frac{1}{2}})\mathrm{tr}(U - DD') . \quad (3.6)$$

Conditions (3.4) and (3.6) do not depend on σ^2 as they correspond to ellipsoids with centres in the origin.

4. Special case: the minimax estimator

In this section the above theory will be applied to the minimax estimator

$$b_R(k) = (X'X + kR'R)^{-1}X'y \tag{4.1}$$

of Kuks and Olman (1972), see also Bibby and Toutenburg (1977). In (4.1) R is an $m \times p$ matrix with rank m and constant $k > 0$. If $R = I$ then (4.1) is the ridge estimator and, more generally, if $R'R$ has the same eigenvectors as $X'X$, then (4.1) is called the generalized ridge estimator. Estimator (4.1) has the optimal property that it has the smallest maximum quadratic risk among homogeneous linear estimators in $\mathcal{B}(0,R'R,k)$ for all $A \geq 0$, cf. Bunke (1975). This property implies that $b_R(k)$ must at any rate be superior to b in $\mathcal{B}(0,R'R,k)$, and it may thus be of interest to consider the situation more generally in $\mathcal{B}(\beta_0,T,d)$.

To begin with, for $b_D = b_R(k)$ we have

$$U - DD' = UR'S_k(2k^{-1}I + RUR')S_kRU$$

where $S_k = (k^{-1}I + RUR')^{-1}$, see Teräsvirta (1981b). As $H = -UR'S_kR$, (2.6) can be written as

$$\Delta(b_R(k),\beta,aa')$$

$$\geq \sigma^2h'(2k^{-1}I + RUR')h - \tilde{r}^2h'h \geq 0 \tag{4.2}$$

where $h = S_k RUa$ and

$$\tilde{r} = \lambda_{max}^{1/2}(RT^{-1}R')\sigma d^{-1/2} + (s_0's_0)^{1/2}$$

with $s_0 = R\beta_0$. Since $a \neq 0$ implies $h \neq 0$, our problem is to find out when does (4.2) hold for all $h \neq 0$. A necessary and sufficient condition for this is seen to be

$$\sigma^2(2k^{-1}I + RUR') - \tilde{r}^2 I \geq 0.$$

which is equivalent to

$$\sigma[2k^{-1} + \lambda_{min}(RUR')]^{1/2} \geq \tilde{r} . \tag{4.3}$$

Further elaboration of (4.3) yields

$$d^{-1/2} \leq \lambda_{max}^{-1/2}(RT^{-1}R')\{[2k^{-1} + \lambda_{min}(RUR')]^{1/2} - (s_0's_0/\sigma^2)^{1/2}\} . \tag{4.4}$$

A necessary condition for (4.4) to hold at least for some $d > 0$ is that the expression in braces be positive, i.e., that

$$s_0's_0/\sigma^2 \leq 2k^{-1} + \lambda_{min}(RUR') . \tag{4.5}$$

Whether (4.4) is valid or not again depends on $\|s_0\|$, T and d, and in this case also on R and k. Setting $s_0 \equiv 0$, and $R = T = I$, (4.4) has the form

$$d^{-1} \leq 2k^{-1} + \lambda_{min}(U)$$

which holds for all X'X if $d^{-1} \leq 2k^{-1}$. Thus the ridge estimator $b_I(k)$ is always superior to OLS in $\mathcal{B}(0,I,k/2)$.

5. Existence of restricted superior minimax estimator

It can be seen that if $k \to 0$ the r.h.s. of (4.4) increases monotonically beyond any preset bound, and we have the following

Theorem. *Assume linear model* (2.1) *and minimax estimator* (4.1). *Then there always exists such a* $k > 0$ *that* (4.1) *is superior to OLS in* $\mathcal{B}(\beta_0,T,d)$ *for arbitrary* $d > 0$ *and* $T > 0$.

Hoerl and Kennard (1970) proved for the ridge estimator ($R \equiv I$) that there always existed a $k > 0$ such that the m.s.e. of the ridge estimator was smaller than that of the least squares estimator. Teräsvirta (1981a) demonstrated that a similar result holds for the mixed estimator of Theil and Goldberger (1961) closely related to (4.1), when quadratic risk with arbitrary $A \geq 0$ is used as the measure of superiority. These results refer to a particular (β,σ^2), whereas the above theorem concerns an ellipsoid $\mathcal{B}(\beta_0,T,d)$ which can be arbitrarily large.

The existence theorem of Hoerl and Kennard has been used as an argument for the ridge estimator, but the problem is that in practice we do not know whether the chosen ridge constant k leads to a sufficiently small risk at (β,σ^2). The result (4.4) contains σ^2 and is thus not free from unknown parameters either. However, if $s_0 \equiv 0$, then (4.4) ceases to be dependent on σ^2.

A noteworthy point is that if d is chosen small (or B large), then as a rule k has to be very close to zero for (4.4) to hold. That means that the subsequent minimax estimator does not deviate much from the OLS estimator and the price paid for the restricted dominance of the minimax over the OLS estimator is that the improvement in estimation due to the minimax estimator remains minor. This also seems to be a characteristic feature of some non-linear estimators dominating the least squares estimator. Several Monte Carlo studies indicate, cf. e.g. Dempster et al. (1977), Gunst and Mason (1977) and Lawless (1978), that the gains from the use of James-Stein estimators dominating the least squares, when normality of errors is assumed and the loss matrix is X'X, are generally small. This seems to be true at least if the predictors are not nearly orthogonal, see Thisted (1977).

6. Weak restricted superiority of minimax estimator

Next, we adapt results of Section 3 to the minimax estimator assuming that $s_0 \equiv 0$. Then, from (3.3) and (3.4) we have

$\Delta(b_R(k),\beta,\ X'X) \geq 0$ in $B(0,T,d)$ if and only if

$$d^{-1} \leq \lambda_{max}^{-1}(T^{-1/2}R'S_kRUR'S_kRT^{-1/2})\,tr(2k^{-1}I + RUR')S_kRUR'S_k \qquad (6.1)$$

The r.h.s. of (6.1) is always positive; note that S_k and RUR' have the same eigenvectors.

Correspondingly, from (3.5) and (3.6) it can be concluded that $\Delta(b_R(k),\beta,\ I) \geq 0$ in $B(0,T,d)$ is equivalent to

$$d^{-1} \leq \lambda_{max}^{-1}(T^{-\frac{1}{2}}R'S_kRU^2R'S_kRT^{-\frac{1}{2}})\,trUR'S_k(2k^{-1}I + RUR')S_kRU\ . \qquad (6.2)$$

As in (6.1), the r.h.s. of (6.2) is always positive so that the existence of at least one ellipsoid of weak restricted superiority is always guaranteed.

7. Restricted least squares estimator

Letting $k \to \infty$ in (4.4) we obtain a necessary and sufficient condition for the restricted least squares estimator with restriction(s) $R\beta = 0$ to be strongly superior to b in $B(\beta_0,T,d)$. A condition corresponding to (4.5) becomes

$$s_0's_0/\sigma^2 \leq \lambda_{min}(RUR') \qquad (7.1)$$

respectively. Note that (7.1) is a sufficient but not necessary condition for the superiority of b_R over b at (β_0,σ^2), but it is needed for the existence of a whole superiority ellipsoid with its centre at β_0. If (7.1) holds, then we obtain from (4.4)

$$d^{-1/2} \leq \lambda_{max}^{-1/2}(RT^{-1}R')[\lambda_{min}^{1/2}(RUR') - (s_0's_0/\sigma^2)^{1/2}]. \tag{7.2}$$

If β_0 is chosen to lie in the hyperplane $R\beta = 0$ then (7.1) is automatically satisfied and (7.2) has the simple form

$$d^{-1} \leq \lambda_{max}^{-1}(RT^{-1}R')\lambda_{min}(RUR'). \tag{7.3}$$

Inequality (7.3) is both necessary and sufficient for $\Delta(b_R,\beta,A) \geq 0$ to hold for all $A \geq 0$ in B_0. By definition, b_R is superior to b whenever $R\beta = 0$. Condition (7.3) indicates the situation when instead of this hyperplane we consider the set of ellipsoids with their centres in $\{\beta\colon R\beta = 0\}$.

Taking the weak predictive superiority criterion, assuming that $s_0 \equiv 0$ and letting $k \to \infty$ in (6.1) yields

$$d^{-1} \leq m\lambda_{max}^{-1}(T^{-\frac{1}{2}}R'(RUR')^{-1}RT^{-\frac{1}{2}}). \tag{7.4}$$

If we are using the m.s.e. as our superiority criterion and retain the assumption $s_0 \equiv 0$, we finally obtain from (6.2)

$$d^{-1} \leq \lambda_{max}^{-1}(T^{-\frac{1}{2}}R'(RUR')^{-1}RU^2R'(RUR')^{-1}RT^{-\frac{1}{2}})\,tr(RUR')^{-1}RU^2R'. \tag{7.5}$$

8. Example

In order to illustrate the above theory, we shall discuss the following example. Consider a four-variable linear model from Hald (1952, p. 647). The number of observations is 13, the correlation matrix of predictors being

$$\hat{P} = \begin{bmatrix} 1.000 & 0.228 & -0.824 & -0.245 \\ & 1.000 & -0.139 & -0.972 \\ & & 1.000 & 0.029 \\ & & & 1.000 \end{bmatrix} .$$

Multicollinearity can be regarded as a problem in this data set. Estimating the regression parameters of the scaled variables by OLS yields

$$\hat{y} = \underset{(0.28)}{0.61x_1} + \underset{(0.71)}{0.53x_2} + \underset{(0.30)}{0.043x_3} - \underset{(0.74)}{0.16x_4}$$

while the residual variance $\hat{\sigma}^2 = 0.024$ and the squared sum of estimated coefficients $b'b = 0.67$. The figures in parentheses are estimated standard deviations, indicating that some parameters have become rather inaccurately estimated.

To improve the estimation, two ridge estimators were chosen, one (HKB) suggested by Hoerl et al. (1975) and the other (LW) by Lawless and Wang (1976). Another alternative was the restricted least squares (RLS) estimator with the restriction $\beta_4 = 0$.

The results of Alam and Hawkes (1978) and Casella (1980) show that, under certain conditions, HKB and LW estimators dominate OLS in terms of the m.s.e. In this example, the necessary condition given in Alam and Hawkes (1978) is met for both estimators while the sufficient one is not, because

$$\mathrm{tr}U^2 - 2\lambda_{max}(U^2) < 0.$$

Thus, outright dominance cannot be conjectured so that the example is not trivial.

The performance of the three estimators was judged using spheres $\mathcal{B}(0,I,d)$ and treating the estimated value of k as fixed. By substituting $\hat{\sigma}^2$ for the unknown σ^2 it was possible to approximate $\beta'\beta \leq \sigma^2 d^{-1}$ in order to find spheres centred in the origin, such that all β's belonging to them would be estimated more accurately by these biased alternatives than by OLS.

The results are in Table 1. When the strong superiority criterion (4.4) with $\beta_0 = 0$ is applied it is doubtful whether HKB can be considered to improve estimation ($b'b = 0.67$, $b'_{HKB}b_{HKB} = 0.41$, $b'_{LW}b_{LW} = 0.52$) while LW no doubt does. Omitting β_4 (using zero as the estimator of β_4) does improve estimation of all parameters in a larger sphere than HKB. When the p.m.s.e. (6.1) or the m.s.e. (6.2

criterion is used, both ridge estimators could be judged as acceptable substitutes for the OLS with little risk for the researcher being worse off by using them instead of sticking to the OLS. For the RLS estimator all the three criteria lead to $d^{-1} \leq u_{44} = 21.8$, and in the light of the estimation results it is quite possible that $\beta'\beta \leq \hat{\sigma}^2 u_{44} = 0.51$ in this example.

It may also be noted that the James and Stein estimator $b_{JS} = \hat{c}b$ used for comparison resulted in $\hat{c} = 0.9967$ so that the James-Stein and OLS estimates were practically identical.

9. Final remarks

The results in the previous sections do not provide the researcher with any clear idea of how *much* biased estimators improve the estimation if they do. For instance, if β_4 in the preceding example were very close to zero then the above RLS estimator would obviously be a reasonable choice, although it improves estimation in a smaller sphere than the two ridge estimators according to both weak criteria of superiority. However, even as things are, the results are useful in charting areas in the parameter space in which certain homogeneous linear estimators can be thought of as reasonable substitutes to OLS.

Table 1. *Upper limits for* d^{-1} *and corresponding values of* $\hat{\sigma}^2 d^{-1}$ *for three biased estimators from three criteria of restricted superiority, using data from Hald (1952)*

Estimator	Criterion					
	(4.4) or (7.3)		(6.1) or (7.4)		(6.2) or (7.5)	
	d^{-1}	$\hat{\sigma}^2 d^{-1}$	d^{-1}	$\hat{\sigma}^2 d^{-1}$	d^{-1}	$\hat{\sigma}^2 d^{-1}$
HKB ($\hat{k}$ = 0.140)	14.4	0.34	69.5	1.63	61.9	1.45
LW ($\hat{k}$ = 0.008)	250	5.88	306	7.18	297	6.97
RLS (β_4 = 0)	21.8	0.51	21.8	0.51	21.8	0.51

References

Alam, K. and J.S. Hawkes (1978). Estimation of regression coefficients. *Scandinavian Journal of Statistics* 5, 169-172.

Bibby, J. and H. Toutenburg (1977). *Prediction and improved estimation in linear models.* New York: John Wiley.

Bunke, O. (1975). Minimax linear, ridge and shrunken estimators for linear parameters. *Mathematische Operationsforschung und Statistik 6,* 697-701.

Casella, G. (1980). Minimax ridge regression estimation. *Annals of Statistics 8,* 1036-1056.

Dempster, A.P., M. Schatzoff and N. Wermuth (1977). A simulation study of alternatives to ordinary least squares. *Journal of the American Statistical Association 72,* 77-91.

Draper, N.R. and R.C. Van Nostrand (1979). Ridge regression and James-Stein estimation: Review and comments. *Technometrics 21,* 451-465.

Gunst, R.F. and R.L. Mason (1977). Biased estimation in regression: an evaluation using mean square error. *Journal of the American Statistical Association 72,* 616-628.

Hald, A. (1952). *Statistical methods with engineering applications.* New York: John Wiley.

Hoerl, A.E. and R.W. Kennard (1970). Ridge regression: Biased estimation for non-orthogonal problems. *Technometrics 12,* 55-67.

Hoerl, A.E., R.W. Kennard and K.F. Baldwin (1975). Ridge regression: some simulations. *Communications in Statistics 4,* 105-123.

Hoffmann, K. (1977). Admissibility of linear estimators with respect to restricted parameter sets. *Mathematische Operationsforschung und Statistik, Series Statistics, 8,* 425-438.

Hoffmann, K. (1980). Admissible improvements of the least squares estimator. *Mathematische Operationsforschung und Statistik, Series Statistics, 11,* 373-388.

Judge, G.G. and M.E. Bock (1978). *The statistical implications of pre-test and Stein-rule estimators in econometrics.* Amsterdam: North-Holland.

Kuks, J. and V. Olman (1972). Minimaksnaja linejnaja ocenka koefficientov regressii. *Eesti NSV teaduste akadeemia toimetised 21,* 66-72.

Lawless, J.F. (1978). Ridge and related estimation procedures: theory and practice. *Communications in Statistics A7*, 139-164.

Lawless, J.F. and P. Wang (1976). A simulation study of ridge and other regression estimators. *Communications in Statistics 5*, 307-323.

Mayer, L.S. and T.A. Willke (1973). On biased estimation in linear models. *Technometrics 15*, 497-508.

Rao, C.R. (1973). *Linear statistical inference and its applications*, 2nd edition. New York: John Wiley.

Rao, C.R. (1976). Estimation of parameters in a linear model. *Annals of Statistics 4*, 1023-1037.

Sclove, S.L. (1968). Improved estimators for coefficients in linear regression. *Journal of the American Statistical Association 63*, 596-606.

Swamy, P.A.V.B. and J.S. Mehta (1977). A note on minimum average risk estimators for coefficients in linear models. *Communications in Statistics A6*, 1181-1186.

Teräsvirta, T. (1981a). Some results on improving the least squares estimation of linear models by mixed estimation. *Scandinavian Journal of Statistics 8*, 33-38.

Teräsvirta, T. (1981b). A comparison of mixed and minimax estimators of linear models. *Communications in Statistics A10* (forthcoming).

Theil, H. and A.S. Goldberger (1961). On pure and mixed statistical estimation in economics. *International Economic Review 2*, 65-78.

Thisted, R.A. (1977). Comment. *Journal of the American Statistical Association 72*, 102-103.

Toro-Vizcarrondo, C. and T.D. Wallace (1968). A test of the mean square error criterion for restrictions in linear regression. *Journal of the American Statistical Association 63*, 558-572.

Wallace, T.D. (1972). Weaker criteria and tests for linear restrictions in regression. *Econometrica 40*, 689-698.

Yancey, T.A., G.G. Judge and M.E. Bock (1974). A mean square error test when stochastic restrictions are used in regression. *Communications in Statistics 3*, 755-768.