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

Keskusteluaiheita
Discussion papers

Timo Teräsvirta

STRONG SUPERIORITY OF HETEROGENEOUS LINEAR ESTIMATORS

No. 137 15 June 1983

Paper prepared for the 1983 American Statistical Association Meeting, Toronto, 15-18 August. This is a revised and somewhat abridged version of the first four sections of my ETLA Discussion Paper No. 127. The revision consists of correcting minor errors and extending and clarifying results of the earlier paper on superiority comparisons between two restricted least squares estimators.

Abstract. This is a revised and somewhat abridged version of the first four sections of an earlier paper called "Superiority comparisons of heterogeneous linear estimators" (ETLA Discussion Paper No. 127). The previous results concerning comparisons between two restricted least squares estimators have been slightly extended in the present paper. Some minor errors have been corrected as well.

Keywords. biased estimation, heterogeneous linear estimator, mixed estimator, restricted least squares, ridge regression

1. Introduction

During the last few years, a variety of biased estimators have been proposed alongside the previous ones like the restricted least squares (RLS) estimator for estimating the parameter vector of the general linear model. The performance of these estimators has been compared to that of the ordinary least squares (OLS) estimator mostly by using superiority criteria based on quadratic risk. Among the first examples of such comparisons are the superiority condition for the RLS estimator to be superior to the OLS estimator (Toro-Vizcarrondo and Wallace, 1968) and the dominance results for James-Stein estimators, for discussion see e.g. Judge et al. (1980) and Vinod and Ullah (1981).

Fewer results have been available on comparisons between biased linear estimators. However, such comparisons have also been made in econometric literature . For instance the problem whether to omit unobservables or substitute proxy variables for them in linear models is equivalent to comparing two different biased estimators. Hocking et al. (1976) have compared certain homogeneous linear estimators with each other. More recently, Trenkler (1980), Teräsvirta (1982a) and Trenkler and Trenkler (1983) have compared general homogeneous linear estimators using the generalised mean square error as the superiority criterion. Teräsvirta (1981a) has in particular discussed the relationship between the mixed and minimax estimators on that basis. Guilkey and Price (1981) have carried out comparisons between RLS estimators. Price (1982) has included various homogeneous linear estimators in his comparisons but without a general framework.

In this paper, a general framework is set up for comparing heterogeneous linear estimators, see also Teräsvirta (1982b). The special cases discussed in the literature can thereafter be treated in a straightforward fashion by applying the general theorem. The comparisons are based on the concept of strong superiority of an estimator over another.

Two applications will be considered here. One of them is a comparison between two ridge estimators, while the other consists of comparing restricted least squares estimators. For more discussion and examples the reader is referred to Teräsvirta (1982b).

2. Preliminaries

Consider a linear model

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

where y and ε are $n \times 1$ stochastic vectors, X is an $n \times p$ fixed matrix with rank$(X) = p$, β is a $p \times 1$ vector of regression coefficients, and σ^2 is the error variance. Define two linear heterogenous estimators of β as $b_j = D_j y + h_j$, $= 1,2$, where D_j is a fixed $n \times p$ matrix and h_j a fixed $p \times 1$ vector. In this paper, the interest will be focussed upon the conditions under which one of these estimators is better than the other. Following established practice we consider this problem using quadratic risk functions. The strong superiority of b_2 over b_1 (cf. also Toro-Vizcarrondo and Wallace, 1968) at a single point (β,σ^2) in the parameter space is defined as follows:

Definition. *Estimator* b_2 *is strongly superior to* b_1 *at* (β,σ^2) *if and only if*

$$E(b_1 - \beta)'A(b_1 - \beta) \geq E(b_2 - \beta)'A(b_2 - \beta) \tag{2.2}$$

for all loss matrices $A \geq 0$.[1)]

This definition is equivalent to requiring that the difference of two MSE matrices

$$MSE(b_1) - MSE(b_2) \geq 0 \text{ where } MSE(b_j) = E(b_j - \beta)(b_j - \beta)',$$

cf. Theobald (1974).

Less restrictive definitions for superiority can be constructed by relaxing the restriction that the inequality (2.2) has to be valid for all non-negative definite loss matrices simultaneously. They are not considered here. For discussion see e.g. Wallace (1972) and Judge et al. (1980, pp. 24-26).

3. Conditions for strong superiority

For the purposes of this paper, it is convenient to write the MSE matrix as a decomposition into covariance and bias:

$$MSE(b_j) = E(b_j - \beta)(b_j - \beta)' = \sigma^2 D_j D_j' + d_j d_j'$$

---

1) $A \geq 0$ means A is a non-negative definite matrix while $A > 0$ means that A is positive definite.

where

$$d_j = H_j\beta + h_j \text{ with } H_j = D_jX - I, \quad j = 1,2 .$$

Set $C = D_1D_1' - D_2D_2'$ so that

$$\Delta_{12} = MSE(b_1) - MSE(b_2) = \sigma^2C + d_1d_1' - d_2d_2' . \qquad (3.1)$$

As pointed out above, b_2 is strongly superior to b_1 if and only if (3.1) is non-negative definite. Assume that we have the following decomposition

$$C = KLK', \quad d_j = Kf_j, \quad j = 1,2 \qquad (3.2)$$

where K is $p \times r$, $r \leq p$, L is $r \times r$, and f_j is $r \times 1$, $j = 1,2$. This decomposition is useful whenever we want to compare estimators with singular covariance matrices. The difference (3.1) can now be written as

$$\Delta_{12} = K(\sigma^2L + f_1f_1' - f_2f_2')K' .$$

It is well-known that $\Delta_{12} \geq 0$ if and only if

$$\sigma^2L + f_1f_1' - f_2f_2' \geq 0 . \qquad (3.3)$$

Let us first exclude the trivial possibility that $L \geq 0$ and $f_2 = \alpha f_1$, $|\alpha| < 1$. This means that we do not consider any estimator b_2 with both smaller variance and bias than b_1; a very rare case in practice. The assumption $L \geq 0$ is retained as yet. For (3.3) to hold it is then necessary that $\sigma^2L + f_1f_1' > 0$. This last assumption implies that either $\text{rank}(L) = r-1$ and f_1 is linearly independent of the columns of L or that L is non-singular. In the latter case L may be either indefinite with exactly one negative eigenvalue or positive definite. In both cases, (3.3) is equivalent to the following condition, cf. Farebrother (1976),

$$f_2'(\sigma^2 L + f_1 f_1')^{-1} f_2 \leq 1 . \tag{3.4}$$

If we can assert that $L > 0$ then, using a well-known matrix identity, (3.4) can be written in the form

$$\sigma^{-2}\{ f_{22} - f_{21}^2(\sigma^2 + f_{11})^{-1} \} \leq 1 \tag{3.5}$$

where

$$f_{ij} = f_i' L^{-1} f_j , \quad i,j = 1,2 .$$

This is the main result of this section. A corresponding condition for two homogeneous linear estimators when $K = I$ and $L > 0$ is to be found in Teräsvirta (1982a), and Trenkler and Trenkler (1983). From (3.5), a sufficient but generally not necessary condition for (3.3) to hold when $L > 0$ is

$$\sigma^{-2} f_{22} \leq 1 \tag{3.6}$$

see also Trenkler (1980). If b_1 is unbiased then $f_1 = 0$ and (3.6) is necessary as well.

Assume now that $L < 0$. A lemma in Guilkey and Price (1981) states that (3.3) can then be valid only if L is a scalar, i.e. if $r = 1$. Then $\Delta_{12} \geq 0$ if and only if, in obvious notation,

$$\sigma^2 l_{11} + f_1^2 - f_2^2 \geq 0 . \tag{3.7}$$

Note that if $L > 0$, then

$$\sigma^2 L + f_1 f_1' - f_2 f_2' \leq 0 \tag{3.8}$$

is a necessary condition for b_1 to be strongly superior to b_2.

Reverse now the rôles of b_1 and b_2 in the Definition, so that $-L > 0$ and (3.3) becomes

$$-\sigma^2 L + f_2 f_2' - f_1 f_1' \geq 0 \, . \tag{3.9}$$

Then it is seen from (3.8) that if $r = 1$, the necessary condition is also sufficient while this is not so when $r > 1$.

The above results can be formulated as

Theorem 1. *Assume linear model* (2.1) *and ~~two~~ heterogeneous linear estimators* $b_j = D_j y + h_j$, $j = 1,2$. *Set* $C = D_1 D_1' - D_2 D_2'$, *assume decomposition* (3.2) *and furthermore that* $\sigma^2 L + f_1 f_1' > 0$. Then b_2 *is strongly superior to* b_1 *if and only if* (3.4) *holds. If it is assumed that* $L > 0$ *then the strong superiority is equivalent to* (3.5). *On the other hand, if* $L < 0$ *then* b_2 *is strongly superior to* b_1 *if and only if* L *is a scalar and* (3.7) *is valid.*

In practice, $L > 0$ (or $L < 0$) seems to be a standard situation. In the following we shall also have an example of the case in which L is non-singular but indefinite. Assumption $L \geq 0$ combined with the rank and linear independence conditions obviously remains a more theoretical possibility.

4. Examples

4.1. Mixed and ridge estimators

Assume that we use stochastic prior information

$$r = R\beta + \phi_1 \tag{4.2}$$

where r is an $m \times 1$ stochastic vector, R is an $m \times p$ fixed matrix with rank $m \leq p$, and it is also assumed that $\phi_1 \sim N(0,(\sigma^2/k_1)I)$, $k_1 > 0$. Suppose that in reality this information is biased so that

$$Er = R\beta + s \tag{4.3}$$

where $s \neq 0$, see Theil and Goldberger (1961), Yancey et al. (1974) and Teräsvirta (1981b). Combining (4.2) with the sample information (2.1) yields the mixed estimator

$$b_R(k_1) = (X'X + k_1R'R)^{-1}(X'y + k_1R'r) .$$

Compare this with another mixed estimator $b_R(k_2)$ where R and (4.3) are the same as above but the uncertainty of prior information is altered in such a way that ϕ_1 in (4.2) is replaced by $\phi_2 \sim N(0,(\sigma^2/k_2)I)$, $k_2 > 0$. To find out when $b_R(k_2)$ is strongly superior to $b_R(k_1)$, we need

$$C = UR'(S_{k_2} - S_{k_1})RU$$

where

$$S_{k_j} = (k_j^{-1}I + RUR')^{-1}, \quad j = 1,2$$

see Teräsvirta (1981b). As $d_j = UR'S_{k_j}s$, we can choose $K = UR'$. Since for two pd matrices A and B, $A - B > 0$ implies $B^{-1} - A^{-1} > 0$ we conclude that $L = S_{k_1} - S_{k_2} > 0$ if and only if $k_2 > k_1$. If $k_1 = k_2$ then $L = 0$. Thus we can improve upon $b_R(k_1)$ only by choosing $k_2 > k_1$ if $m > 1$. When $k_2 \to \infty$, $b_R(k_2)$ converges towards the restricted least squares (RLS) estimator b_R. Thus, for some combinations of X, β and σ^2, a mixed estimator can be improved upon by a RLS estimator.

For two minimax estimators (Kuks and Olman, 1972) $b_I^*(k_j) = (X'X + k_jR'R)^{-1} X'y$, $j = 1,2$, we have

$$\Delta_{12} = \sigma^2 UR'(S_{k_2}T_{k_2}^{-1}S_{k_2} - S_{k_1}T_{k_1}^{-1}S_{k_1})RU + UR'S_{k_1}R\beta\beta'R'S_{k_1}RU - UR'S_{k_2}R\beta\beta'R'S_{k_2}RU \tag{4.3}$$

where

$$T_{k_j} = (2k_j^{-1}I + RUR')^{-1}, \quad j = 1,2 .$$

Matrices S_{k_1}, S_{k_2}, T_{k_1} and T_{k_2} have the same eigenvectors, and since the eigenvalues of $S_kT_k^{-1}S_k$ are monotonously increasing functions of k, it follows that in (4.3), $S_{k_2}T_{k_2}^{-1}S_{k_2} - S_{k_1}T_{k_1}^{-1}S_{k_1} > 0$ if and only if $k_2 > k_1$. When $R = I$, the minimax estimator is simply the ridge estimator. Thus if $p > 1$, a ridge estimator $b_I^*(k_1)$ with a fixed ridge parameter k_1 can in some parts of the parameter space be improved upon by increasing the value of the ridge parameter. The same is not possible by decreasing its value.

4.2. Restricted least squares estimators

In the previous sub-section the set (Er, R) was kept unchanged throughout. Here we start from two separate sets of linear restrictions

$$\tilde{r}_j = \tilde{R}_j\beta, \quad j = 1,2$$

where $\tilde{r}_j$ is now deterministic. The corresponding RLS estimators are

$$b_{\tilde{R}_j} = b + U\tilde{R}_j'(R_jU\tilde{R}_j')^{-1}\hat{\tilde{s}}_j, \quad j = 1,2 \tag{4.4}$$

where $\hat{\tilde{s}}_j = \tilde{r}_j - \tilde{R}_jb$. We shall derive conditions for strong superiority of $b_{\tilde{R}_2}$ over $b_{\tilde{R}_1}$ at (β,σ^2). Assume the block division

$$\tilde{R}_j = \begin{bmatrix} R_j \\ R_3 \end{bmatrix}, \ \tilde{r}_j = \begin{bmatrix} r_j \\ r_3 \end{bmatrix}, \ \tilde{s}_j = \begin{bmatrix} s_j \\ s_3 \end{bmatrix} = \begin{bmatrix} r_j - R_j\beta \\ r_3 - R_3\beta \end{bmatrix}, \ j = 1,2 \tag{4.5}$$

so that $r_3 = R_3\beta$ is a subset of restrictions common for both sets. The rows of R_1 are not linearly dependent of the rows of R_2. Let r_j be an $m_j \times 1$ vector, and R_j an $m_j \times p$ matrix, $1 < m_1 + m_2 + m_3 \leq p$, $j = 1,2$, and assume rank $(R_1', R_2', R_3') = m_1 + m_2 + m_3$. Let $m_j = 0$ symbolise the absence of the j^{th} set of restrictions. We have

$$C = U\tilde{R}_2' (\tilde{R}_2U\tilde{R}_2')^{-1}\tilde{R}_2U - U\tilde{R}_1' (\tilde{R}_1U\tilde{R}_1')^{-1}\tilde{R}_1U \tag{4.6}$$

and

$$d_j = U\tilde{R}_j'(\tilde{R}_jU\tilde{R}_j')^{-1}\tilde{s}_j, \quad j = 1,2\,. \tag{4.7}$$

Using (4.5) we can write

$$U\tilde{R}_j'(\tilde{R}_jU\tilde{R}_j')^{-1}\tilde{R}_jU = UR_3'(R_3UR_3')^{-1}R_3U + UB R_j' D_{jj\cdot3}^{-1} R_j B'U, \ j = 1,2 \tag{4.8}$$

where

$$D_{jj\cdot 3} = R_jUR_j' - R_jUR_3'(R_3UR_3')^{-1} R_3UR_j' = \sigma^{-2}R_j\mathrm{cov}(b_{R_3})R_j', \quad j = 1,2 \tag{4.9}$$

and

$$B = I - R_3' (R_3UR_3')^{-1} R_3U .$$

The first term on the r.h.s. of (4.8) is the contribution of the restrictions $r_3 = R_3\beta$ to the covariance matrix of $b_{\tilde{R}_j}$, whereas the second term represents the remaining contribution of $r_j = R_j\beta$ after purging out the effect of $r_3 = R_3\beta$. Making use of (4.8) in (4.6) yields

$$C = UB(R_2'D_{22\cdot 3}^{-1} R_2 - R_1' D_{11\cdot 3}^{-1} R_1) B'U . \tag{4.10}$$

The matrix in parentheses in (4.10) is generally indefinite. Conforming to the block division in (4.5) we also have

$$d_j = U \{ BR_j'D_{jj\cdot 3}^{-1}s_j + (I - BR_j'D_{jj\cdot 3}^{-1}R_jU)R_3'(R_3UR_3')^{-1}s_3 \}, \quad j = 1,2 . \tag{4.11}$$

From (4.10) and (4.11) it is obvious that Theorem 1 does not apply unless $s_3 = 0$. We make that assumption and define $R' = (R_1' \; R_2')$. Now choose $K = UBR'$, $L = \mathrm{diag}\{-D_{11\cdot 3}^{-1}, D_{22\cdot 3}^{-1}\}$, $f_1 = \begin{bmatrix} D_{11\cdot 3}^{-1} \\ 0 \end{bmatrix} s_1$ and $f_2 = \begin{bmatrix} 0 \\ D_{22\cdot 3}^{-1} \end{bmatrix} s_2$. Matrix L is non-singular but indefinite, and the number of negative eigenvalues equals m_1. Write

$$\sigma^2 L + f_1f_1' = \mathrm{diag}\{-\sigma^2 D_{11\cdot 3}^{-1} + D_{11\cdot 3}^{-1}s_1s_1'D_{11\cdot 3}^{-1}, \sigma^2 D_{22\cdot 3}^{-1}\} \tag{4.12}$$

and choose $m_1 = 1$. The necessary and sufficient condition for (4.12) to be positive definite is then

$$\sigma^{-2}s_1'D_{11\cdot 3}^{-1}s_1 > 1. \tag{4.13}$$

It is a condition for the unbiased estimator b_{R_3} to be strongly superior to $b_{\tilde{R}_1}$. If $m_1 > 1$, (4.13) is only a necessary but not sufficient condition. If (4.13) holds and $m_1 = 1$ then Theorem 1 applies, and the necessary and sufficient condition for $b_{\tilde{R}_2}$ to be strongly superior to $b_{\tilde{R}_1}$ is

$$\sigma^{-2} s_2' D_{22\cdot 3}^{-1} s_2 \leq 1. \qquad (4.14)$$

Inequality (4.14) is a condition for $b_{\tilde{R}_2}$ to be strongly superior to b_{R_3}. If $m_1 = 0$, we can choose $K = UBR_2'$, $L = D_{22\cdot 3}^{-1}$ and $f_2 = D_{22\cdot 3}^{-1} s_2$. Applying Theorem 1 then yields (4.14) because $L > 0$. We have

<u>Corollary 1.</u> *Assume linear model* (2.1) *and two restricted least squares estimators* $b_{\tilde{R}_1}$ *and* $b_{\tilde{R}_2}$ *with* m_3 *common restrictions* $r_3 = R_3\beta$. *Assume that* $\text{rank}(R_1' \; R_2' \; R_3') = m_1 + m_2 + m_3 \leq p$ *and that* $s_3 = 0$, *i.e., that the common restrictions are true. If* $m_1 = 1$, $b_{\tilde{R}_2}$ *is strongly superior to* $b_{\tilde{R}_1}$ *if and only if* (i) *the unbiased estimator* b_{R_3} *is strongly superior to* $b_{\tilde{R}_1}$ *and* (ii) $b_{\tilde{R}_2}$ *is strongly superior to* b_{R_3}. *If* $m_1 = 0$, $R_3 = \tilde{R}_1$ *and the superiority condition is* (4.14). *If* $m_1 > 1$, *no strong superiority condition exists.*

The corollary tells us under which circumstances a RLS can be improved upon by another estimator of the same type. Removing one linear restriction must then increase the estimation accuracy and the resulting estimator must be unbiased. The unbiased estimator b_{R_3} may then be improved upon by adding new, not necessarily true restrictions $r_2 = R_2\beta$. Guilkey and Price (1981, Theorem 3) consider the same problem, but here the results appear in the correct form.

Assume next that the two RLS estimators do not contain common restrictions so that $m_3 = 0$. Then condition (4.13) becomes

$$\sigma^{-2}s_1'(R_1UR_1')^{-1}s_1 > 1. \tag{4.15}$$

It follows from Theorem 1 that (4.15) is a necessary and sufficient condition for the OLS estimator b to be strongly superior to b_{R_1} when $m_1 = 1$. Similarly, (4.14) has the form

$$\sigma^{-2}s_2'(R_2UR_2')^{-1}s_2 \leq 1. \tag{4.16}$$

Inequality (4.16) is of course the necessary and sufficient condition for b_{R_2} to be strongly superior to b, cf. Toro-Vizcarrondo and Wallace (1968). We have obtained

Corollary 2. *Assume linear model* (2.1) *and two restricted least squares estimators* b_{R_1} *and* b_{R_2}. *The columns of* $R' = (R_1'\ R_2')$ *are assumed linearly independent. If* $m_1 = 1$, b_{R_2} *is strongly superior to* b_{R_1} *if and only if* (i) *the OLS estimator* b *is strongly superior to* b_{R_1} *and* (ii) b_{R_2} *is strongly superior to* b. *If* $m_1 > 1$, *no superiority condition exists.*

Guilkey and Price (1981) have a similar result (Theorem 4). However, they have imposed rather strict additional restrictions on R_1 and R_2 which are not needed here.

Note that (4.13) and (4.14) are testable hypotheses under the condition $s_3 = 0$. If we want to test the strong superiority of $b_{\tilde{R}_2}$ over $b_{\tilde{R}_1}$ when

$m_1 = 1$, the hypotheses are ordered. This is because (4.13) is necessary for applying Theorem 1. Define

$$F_j = \hat{\sigma}_3^{-2} m_j^{-1} \hat{s}_j' D_{jj.3}^{-1} \hat{s}_j \ , \quad j = 1,2 \tag{4.17}$$

where $\hat{\sigma}_3^2 = (n - p + m_3)^{-1}(y - Xb_{R_3})'(y - Xb_{R_3})$ and $\hat{s}_j = r_j - R_j b$. It is obvious from (4.9) that under (4.13) statistic F_1 follows a non-central F distribution with one and $n - p + m_3$ degrees of freedom and non-centrality parameter 1/2, cf. also Toro-Vizcarrondo and Wallace (1968). Similarly under (4.14), F_2 has a non-central $F(m_2, n - p + m_3, 1/2)$ distribution. When testing (4.13), low values of F_1 cause a rejection of the hypothesis while in the case of (4.14) high values of F_2 should indicate the superiority of $b_{\tilde{R}_2}$ over b (and $b_{\tilde{R}_1}$ if (4.13) was tested and accepted).

The principal component (PC) estimator is a special case of the RLS estimator, see e.g. Judge et al. (1980, pp. 468-471). In this paper the eigenvalues of X'X have been assumed positive. Then none of the data-specific linear independent restrictions inherent in the PC estimator can be exactly valid. From Corollary 2 we obtain

Corollary 3. *Assume linear model* (2.1) *and two PC estimators* b_{R_1} and b_{R_2}. *Assume that in the former one exactly one principal component is omitted whereas in the latter the same happens to* m_2 *other principal components. Then* b_{R_2} *is strongly superior to* b_{R_1} *if and only if* (i) b *is strongly superior to* b_{R_1} *and* (ii) b_{R_2} *is strongly superior to* b.

If some of the omitted principal components are common to the both PC estimators under comparison then no strong superiority condition can be established. This is because, in earlier notation, $s_3 \neq 0$. Thus the superiority condition for two PC estimators in Price (1982) is incorrect.

References

Farebrother, R.W. (1976). Further results on the mean square error of ridge regression. *Journal of the Royal Statistical Society* B 38, 248-250.

Guilkey, D.K. and J.M. Price (1981). On comparing restricted least squares estimators. *Journal of Econometrics* 15, 397-404.

Hocking, R.R., F.M. Speed and M.J. Lynn (1976). A class of biased estimators in linear regression. *Technometrics* 18, 425-437.

Judge, G.G., W.E. Griffiths, R.C. Hill and T.-C. Lee (1980). *The theory and practice of econometrics*. New York: Wiley.

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

Price, J.M. (1982). Comparisons among regression estimators under the generalized mean square error criterion. *Communications in Statistics* A11, 1965-1984.

Teräsvirta, T. (1981a). A comparison of mixed and minimax estimators of linear models. *Communications in Statistics* A10, 1765-1778.

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

Teräsvirta, T. (1982a). Superiority comparisons of homogeneous linear estimators. *Communications in Statistics* A11, 1595-1601.

Teräsvirta, T. (1982b). Superiority comparisons of heterogeneous linear estimators. ETLA Discussion Paper No. 127.

Theil, H. and A.S. Goldberger (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* B 36, 103-106.

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

Trenkler, G. (1980). Generalized mean square error comparisons of biased regression estimators. *Communications in Statistics* A9, 1247-1259.

Trenkler, G. and D. Trenkler (1983). A note on superiority comparisons of homogeneous linear estimators. *Communications in Statistics* A12, 799-808.

Vinod, H.D. and A. Ullah (1981). *Recent advances in regression methods.* New York: Dekker.

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.