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 A SHRINKAGE ESTIMATOR WITH A FIXED SHRINKAGE FACTOR

No. 84 4.6.1981

Abstract. In this paper the existence of a shrinkage estimator (Mayer and Willke, 1973) superior to the OLS estimator in an arbitrary ellipsoid in the parameter space is demonstrated. Properties of these superior estimators are discussed and compared with the corresponding features of James-Stein estimators.

Keywords: biased estimation; James-Stein estimator; linear model; restricted superiority; shrinkage estimator.

1. Introduction

In estimating parameters of linear models, James-Stein estimators have quite often been advocated since, although biased, they dominate the ordinary least squares (OLS) estimator under quadratic risk and certain other conditions. However, in simulation experiments conducted by several researchers the reduction in risk as compared to OLS has turned out be small most of the time.

The James-Stein estimators are of type $\hat{c}b$, where b is the OLS estimator and $\hat{c}$ is stochastic and depends on the observations. Mayer and Willke (1973) suggested a biased shrinkage estimator of type cb, $0 \leq c < 1$, where c is fixed. This estimator does not dominate OLS, but in a sense, for a suitable choice of c, it can be shown to come close to outright dominance. This paper investigates the conditions under which that happens and studies certain properties of James-Stein estimators in the light of the present findings.

2. Concepts and results

Assume a linear model

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

where y and ε are $n \times 1$, X is an $n \times p$ matrix of full rank and uncorrelated with ε and β is a $p \times 1$ parameter vector. Assume that the vector of regression coefficients is estimated by two estimators $\tilde{b}_1$ and $\tilde{b}_2$.

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

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

for all $\beta \in B(\beta_0,T,d)$ *and* $A \geq 0$. $R(\tilde{b}_j,\beta,A)$ *is the quadratic risk of* $\tilde{b}_j$ *with loss matrix* A.

Note that d can be regarded as the size parameter of $B(\beta_0,T,d)$. As $d \to 0$, the size of the ellipsoid increases beyond any preset limit.

Theorem 1. *A linear homogeneous estimator* $b_D = Dy$ *is strongly superior to the least squares estimator* $b = UX'y$ *in* $B(\beta_0,T,d)$ *if and only if*

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

where

$$U = (X'X)^{-1},\ H = DX - I$$

and $\lambda_{min}(Y)$ *and* $\lambda_{max}(Y)$ *are the smallest and the largest eigenvalue of* Y.

For a proof, see Teräsvirta (1981).

Corollary. *Estimator* $b_D = Dy$ *is strongly superior to* b *in* $B(0,T,d)$ *if and only if*

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

3. Shrinkage estimator

Consider the shrinkage estimator $b_c = cb$, $0 \leq c < 1$ (Mayer and Willke, 1973). Since this estimator shrinks to zero as $c \to 0$, we may be interested in its superiority over OLS in ellipsoids centred in the origin. According to the Corollary, b_c is strongly superior to b in $B(0,T,d)$ if and only if

$$d^{-1} \leq (1 + c)(1 - c)^{-1} h \tag{3.1}$$

where

$$h = \lambda_{min}(U)\lambda_{min}(T).$$

The r.h.s. of (3.1) is an increasing function of c and grows to infinity as $c \to 1$. Thus, for any fixed pair T and d, there is a $c < 1$ such that (3.1) holds. This can be formulated as

Theorem 2. *Consider the set of shrinkage estimators* $B = \{b_c : b_c = cb, 0 \leq c < 1\}$ *for* β *in* (2.1). *There always exists such an estimator* $b_{c'} \in B$ *that* $b_{c'}$ *is strongly superior to* b *in* $B(0,T,d)$.

In fact, the set of shrinkage estimators satisfying (3.1) is $B_d = \{b_c : b_c = cb, (1 - dh)(1 + dh)^{-1} \leq c < 1\}$. For $d^{-1} \leq h$, B_d equals B. If the ellipsoid is large (d small), then c has to be close to one for $b_c \in B_d$.

If, in the Definition, instead of Λ, we prefer X'X as the loss matrix in the quadratic risk function and define weak restricted superiority in $B(0,T,d)$ accordingly, we have, from Teräsvirta

(1981), that b_c is weakly superior to b in $B(0,d,T)$ if and only if

$$d^{-1} \leq p(1+c)(1-c)^{-1}\lambda_{max}^{-1}(T^{-\frac{1}{2}}X'XT^{-\frac{1}{2}}). \qquad (3.2)$$

When T = I so that the ellipsoid are spheres, the r.h.s. of (3.2) equals p times the r.h.s. of (3.1).

4. Discussion

The shrinkage estimator b_c never dominates the least squares. However, there always exists a shrinkage estimator which is superior to b in an arbitrarily large $B(0,T,d)$. The larger the ellipsoid, the smaller is the set of superior shrinkage estimators and the closer are the OLS and the superior shrinkage estimates to each other.

In this paper, c has been assumed fixed. A class of shrinkage estimators with $\hat{c}$ stochastic, so-called James-Stein estimators, has been widely studied in the literature, for discussion see e.g. Judge and Bock (1978) and Draper and Van Nostrand (1979). It is well-known that, under quadratic risk with loss matrix X'X and $p \geq 3$, the James-Stein estimator dominates OLS. In the simulation studies it also seems to have a consistently smaller mean square error than OLS, see for instance Dempster et al. (1977), Gunst and Mason (1977) and Lawless (1978), but its dominance has not been proved in that case. On the other hand, for a proper choice of c, b_c was shown to be strongly superior to OLS in an arbitrarily large ellipsoid. This might lend some support to the idea that, analogously, there exist shrinkage estimators of type $\hat{c}b$ which dominate b under quadratic risk for loss matrices other than X'X.

The simulation studies further indicate that the gains from the use of James-Stein estimators instead of OLS are relatively small. The results of this paper show that if $B(0,T,d)$ is chosen very large then c has to be close to one to guarantee restricted superiority, which again leads only to minor improvements over OLS. It would seem that dominance over OLS, or restricted superiority if the ellipsoid is very large, are such strong requirements that the overall improvement in performance cannot be very substantial, and that estimators superior to OLS in much smaller ellipsoids may yield clearly larger improvements in those parts of the parameter space.

The above-mentioned simulation studies also show that the gains achieved by James-Stein estimators depend on the structure of X'X. They seem to be larger when X'X is near-orthogonal than in the presence of strong multicollinearity, for discussion see Draper and Van Nostrand (1979) and Thisted (1977). Assuming T = I for simplicity and studying (3.1) and (3.2) shows that, keeping d fixed, these conditions are satisfied for lower values of c if $\lambda_{max}(X'X)$ is small than if it is large. In the former case a $b_c \in B_d$ can lead to a larger average gain than in the latter, which is in line with the observed behaviour of James-Stein estimators.

References

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.

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

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

Teräsvirta, T. (1981). Restricted superiority of linear homogeneous estimators over ordinary least squares. Research Institute of the Finnish Economy, Discussion Paper No. 83.

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