Faculty Working Papers

A COMPARISON OF TRADITIONAL AND STEIN-RULE ESTIMATORS UNDER SQUARED ERROR LOSS

George G. Judge and Mary E. Bock

#204

College of Commerce and Business Administration
University of Illinois at Urbana-Champaign

FACULTY WORKING PAPERS

College of Commerce and Business Administration

University of Illinois at Urbana-Champaign

September 6, 1974

A COMPARISON OF TRADITIONAL AND STEIN-RULE ESTIMATORS UNDER SQUARED ERROR LOSS

George G. Judge and Mary E. Bock

#204

28 August, 1974

A COMPARISON OF TRADITIONAL AND STEIN-RULE ESTIMATORS UNDER WEIGHTED SQUARED ERROR LOSS

George G. Judge and Mary Ellen Bock

University of Illinois at Urbana-Champaign

Using a weighted squared error loss measure of goodness maximum likelihood, pre test and Stein rule estimators are analytically compared and the conditions necessary for one estimator to be superior to another are developed.

1. Introduction

For the problem of estimating the mean vector of a multivariate normal distribution or more explicitly for our purposes, the problem of estimating the K dimensional parameter vector $\underline{\theta}$ for the orthonormal regression model

$$(1.1) \qquad \underline{y} = X\underline{\beta} + \underline{e} = XS^{-\frac{1}{2}}S^{\frac{1}{2}}\underline{\beta} + \underline{e} = Z\underline{\theta} + \underline{e},$$

where $\underline{y}$ and $\underline{e}$ are (T x 1) normal vectors with means $Z\underline{\theta}$ and $\underline{o}$, respectively, and covariance $\sigma^2 I_T$, the following results have been obtained under a total squared error loss measure of goodness, with risk $E[(\underline{\tilde{\theta}} - \underline{\theta})'(\underline{\tilde{\theta}} - \underline{\theta})]$; i) James and Stein [1961] developed a Stein-rule estimator, $\underline{\theta}^* = (1 - c^*/u)\underline{\hat{\theta}}$, which, for $K \geq 3$ and $0 \leq c^* \leq 2(K-2)(T-K)K^{-1}[T-K+2]^{-1}$, where $u = \underline{\hat{\theta}}'\underline{\hat{\theta}}/K\hat{\sigma}^2 = F_{(K,T-K)}$ is the likelihood ratio test statistic, dominated the conventional maximum likelihood estimator, $\underline{\hat{\theta}} - Z'\underline{y}$); ii) Baranchik [1964] and Stein [1966] developed a positive part variant of the Stein-rule estimator, $\underline{\theta}^+ = I_{[c^*, \infty)}(u)(1 - c^*/u)\underline{\hat{\theta}}$, where $I_{[c^*, \infty)}(u)$ is a zero-one indicator function, that dominated the James and Stein estimator, $\underline{\theta}^*$; iii) Strawderman [1971] developed minimax rules which were admissible but did not dominate the Stein rule positive part estimator, iv) Sclove, Morris, and Radhakrishnan [1972] showed the pre test estimator, $\underline{\hat{\hat{\theta}}} = I_{[c, \infty)}(u)\underline{\hat{\theta}}$, an estimator often used in econometric work, is uniformly inferior or dominated by a modified version of the positive part Stein-rule estimator, $\underline{\theta}^{++} = I_{[c, \infty)}(u)I_{[c^*, \infty)}(u)(1 - c^*/u)\underline{\hat{\theta}}$, and v) Efron and Morris [1973], using empirical Bayes ideas, showed that the Stein positive rule estimator, $\underline{\theta}^+$, is a member of a class of good rules that have Bayesian

properties.

These results apply to the $\underline{\Theta}$ space where the coefficient variances and weights are equal. In practice many situations may arise where the coefficient variances are not all equal or we wish to weight certain elements of $\underline{\theta}$ differently than others. Therefore, the purpose of this paper is to summarize the conditions necessary for the James and Stein estimator [1961] and its variants to dominate the conventional and pre test estimators for the regression statistical model under a weighted squared error loss measure of goodness. It is noted that under situations usually found in econometric paractice, where nonexperimental data and model building procedures are employed, these conditions may often not be fulfilled.

2. Estimator Comparisons

To compare the estimators we use the weighted risk function

$$(2.1)\quad E[(\underline{\tilde{\beta}} - \underline{\beta})'W(\underline{\tilde{\beta}} - \underline{\beta})] = E[(\underline{\tilde{\Theta}} - \underline{\Theta})'S^{-\frac{1}{2}}WS^{-\frac{1}{2}}(\underline{\tilde{\Theta}} - \underline{\Theta})],$$

where W is a $(K \times K)$ symmetric positive definite matrix and from (1.1), $\underline{\Theta} = S^{\frac{1}{2}}\beta$. If as a special case the weight matrix is $W = I_K$ then the weight matrix in the $\underline{\Theta}$ space is $S^{-\frac{1}{2}}S^{-\frac{1}{2}} = S^{-1} = (X'X)^{-1}$ and for expository purposes this is the case we will consider, although it should be understood that the results we present are valid for any positive definite W matrix.

In the comparing the estimators, $\underline{\tilde{\theta}}$, will be said to dominate, $\underline{\bar{\theta}}$, if

$$(2.2)\quad E[(\underline{\tilde{\theta}} - \underline{\theta})'S^{-\frac{1}{2}}WS^{-\frac{1}{2}}(\underline{\tilde{\theta}} - \underline{\theta})] - E[(\underline{\bar{\theta}} - \underline{\theta})'S^{-\frac{1}{2}}WS^{-\frac{1}{2}}(\underline{\bar{\theta}} - \underline{\theta})] \leq 0,$$

for all $\underline{\theta}$, with strict inequality holding for some $\underline{\theta}$.

From the work of Bock, Yancey and Judge [1973] and Bock [1974], we may write, for the case $W = I_K$, the risk function for the James and Stein estimator $\underline{\Theta}^*$, as

$$(2.3)\qquad E[(\underline{\theta}^*-\underline{\theta})S^{-\frac{1}{2}}S^{-\frac{1}{2}}(\underline{\theta}^*-\underline{\theta})] = \sigma^2 trS^{-1} + \sigma^2 trS^{-1}c^*(T-K)$$

$$E\{(K+2H)^{-1}(K-2+2H)^{-1}\,[c^*(T-K+2)-2(K-2)$$

$$+ (\underline{\theta}'S^{-1}\underline{\theta}/\underline{\theta}'\underline{\theta}(trS^{-1}))\quad 2H(c^*(T-K+2)$$

$$- 2(\underline{\theta}'\underline{\theta}(trS^{-1})/\underline{\theta}'S^{-1}\underline{\theta} - 2))]\}$$

where H is a Poisson random variable with parameter $\underline{\theta}'\underline{\theta}/\sigma^2$.

In order for the risk of the James and Stein estimator, $\underline{\theta}^*$, to be equal to or less than the risk of the maximum likelihood estimator, $\hat{\underline{\theta}}$, over the entire range of the parameter space, $\underline{\Theta}$, the expression between the {} brackets must be zero or negative. For this to occur and thus $E[(\hat{\underline{\theta}}-\underline{\theta})'S^{-1}(\hat{\underline{\theta}}-\underline{\theta})] - E[(\underline{\theta}^*-\underline{\theta})'S^{-1}(\underline{\theta}^*-\underline{\theta})] \geq 0$, it must be true that

$$(2.4)\qquad \sum_{i=1}^{K} d_i/d_L = trS^{-1}/d_L > 2$$

$$(2.5)\qquad 0 \leq c^* \leq 2(T-K)(trS^{-1}\,d_L^{-1} - 2)/K(T-K+2),$$

where the d_i are the roots of S^{-1}, with d_L being the largest. For the general weighted case these roots would be associated with the matrix $S^{-\frac{1}{2}}W\,S^{-\frac{1}{2}}$.

These conditions imply that under the weighted squared error loss, the superiority of the James and Stein estimator, $\underline{\theta}^*$, relative to the conventional estimator, $\hat{\underline{\theta}}$, depends not only on the number of explanatory variables or hypotheses, as was one of the requirements for the traditional unweighted loss in the $\underline{\theta}$ space, but also on whether or not trS^{-1} or the sum of characteristic roots divided by the largest characteristic root of S^{-1} is equal to or larger than 2. If $trS^{-1}/d_L \leq 2$, then for no value of $c > 0$ does the James-Stein dominate the least squares estimator.

Since the degree of collinearity of the columns of the design matrix X is related to the magnitude of the roots of S^{-1}, and ill conditioned X'X matrix may affect whether or not the risk functions cross at some point in the parameter space and thus has a direct impact on the choice of estimator. An analysis of the design matrix for several auto regressive and conventional estimated economic relationships suggests that the condition $trS^{-1}/d_L > 2$ may often not be fulfilled in econometric work and therefore the James and Stein estimator does not fulfill the minimax criterion.

These same conditions or requirements ($trS^{-1}/d_L > 2$) and (2.5) must also hold in order for the James and Stein positive part estimator, $\underline{\theta}^+$, and the extension of the Sclove modified positive part Stein-rule or preliminary test estimator, $\underline{\theta}^{++}$, to dominate the least squares, $\hat{\underline{\theta}}$, and conventional preliminary test, $\hat{\hat{\underline{\theta}}}$, estimators, respectively. This means the appearance of three or more regressors and a suitably small c* do not insure, as they did in the unweighted loss (prediction in the $\underline{\beta}$ space) case, that the risk of these various extensions of the Stein-rule estimators, will be less over the entire parameter space than conventional and pre test estimators. As a consequence the condition (2.4) suggests a new rule for determining the range of d_i or in another sense the coefficient weights or degree of collinearity permissible among the columns in the design matrix X to permit these extensions of the Stein-rule estimators to dominate the conventional sampling theory estimators.

3. Concluding Remarks

As we change from the traditional statistical model usually analyzed in the literature when considering Stein-rule estimators to that of the more general regression case with a weighted squared error loss function, the following analytical results hold:

i) In order for the extended Stein-rule estimators specified in this paper to dominate conventional estimators, the ratio of the sum of the characteristic roots of the (X'X) matrix, to the largest root of

this matrix must be equal to or greater than 2.

ii) If this ratio is not greater than two then some members of the family of potential risk functions for extensions of the Stein-rule estimators, in a given problem, cross the risk function of the least squares estimator;

iii) If conditions (2.4 and 2.5) are fulfilled, the Sclove modified Stein-rule (pre test) estimator, for the general model, is uniformly superior to the conventional pre test estimator over the parameter space. Like the conventional pre test estimator however, its risk function crosses that of the least squares estimator for large values of the critical value c or small values of α; the Stein-rule only yields a minimax estimator for smaller values of c or larger values of α than are ordinarily used in practice.

iv) The incidence of collinearity among the columns of the X matrix to the extent that the X'X matrix is "borderline" full rank, and one root is small relative to the other roots, means that, the conventional pre test and the family of Stein-rule estimators are superior (smaller risk) to the least squares estimator only over a small interval of the parameter space and are inferior (larger risk) over a large, and in some cases, infinite range of the parameter space.

The analytical results summarized herein suggest that unless the researcher has great confidence that his linear hypotheses are true, under the risk measure we have employed when collinearity among the columns of the design matrix is present, there is much to lose and very little to gain by broadening the class of estimators and using the two stage pre test or Stein-rule procedures.

Finally we note the rules discussed in this paper, with the exception of those by Strawderman [1971] do not satisfy the conditions necessary for a generalized Bayes estimator and thus are not admissible. One important area of work for the

future is to evaluate the performance of these estimators under alternative specifications and loss functions. In regard to alternative measures of performance it should be noted that Bednarek-Kozek [1973], using as his measure of goodness the risk matrix $E[S^{-\frac{1}{2}}(\tilde{\underline{\theta}} - \underline{\theta})(\tilde{\underline{\theta}} - \underline{\theta})'S^{-\frac{1}{2}}]$, has, my making use of the work of Stein [1955], demonstrated that the least squares estimator is admissible.

4. References

Baranchik, A. J. (1964): "Multiple Regression and the Estimation of the Mean of a Multivariate Normal Distribution," Stanford University Technical Report, No. 51.

Bednarek-Kozek, B. (1973): "On Estimation in the Multidimensional Gaussian Model," Zastosowania Matenatyki, 13:511-20.

Bock, M. E. (1974): "Certain Minimax Estimators of the Mean of a Multivariate Normal Distribution," Ph.D. Thesis, University of Illinois.

Bock, M. E., T. A. Yancey, and G. G. Judge (1973): "The Statistical Consequences of Preliminary Test Estimators in Regression," Journal of American Statistical Association, 68:109-116.

Efron, B. and C. Morris (1973): "Stein's Estimation Rule and Its Competitors--An Empirical Bayes Approach," Journal of American Statistical Association, 68:117-130.

James, W. and C. Stein (1961): "Estimation with Quadratic Loss," Proceedings of Fourth Berkeley Symposium on Mathematical Statistic Problems, University of California Press, Berkeley, pp. 361-79.

Sclove, S. S., C. Morris, and R. Radhakrishnan (1972): "Non Optimality of Preliminary-Test Estimators for the Multinormal Mean," Annals of Mathematical Statistics, 43:1481-90.

Stein, C. (1966): "An Approach to the Recovery of Interblock Information in Balanced Incomplete Block Designs," Research Papers in Statistics: Festschrift for J. Neyman, F. N. David (ed.), John Wiley and Sons, Inc., New York, pp. 351-66.

Stein C. (1955): "Inadmissibility of the Usual Estimator for the Mean of a Multivariate Normal Distribution," Proceedings of the Third Berkeley Symposium, University of California Press, Berkeley, 1:197-206.

Strawderman, W. E. (1971): "Proper Bayes Minimax Estimates of the Multivariate Normal Mean," Annals of Mathematical Statistics, 42: 385-88.