IMPROVED ESTIMATION OF THE DISTURBANCE VARIANCE IN A LINEAR REGRESSION MODEL

BY

ALAN E. GELFAND and DIPAK K. DEY

TECHNICAL REPORT NO. 418

JULY 11, 1989

PREPARED UNDER CONTRACT

N00014-89-J-1627 (NR-042-267)

FOR THE OFFICE OF NAVAL RESEARCH

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

IMPROVED ESTIMATION OF THE DISTURBANCE VARIANCE IN A LINEAR REGRESSION MODEL

BY

ALAN E. GELFAND and DIPAK K. DEY

TECHNICAL REPORT NO. 418
JULY 11, 1989

Prepared Under Contract
N00014-89-J-1627 (NR-042-267)
For the Office of Naval Research

Herbert Solomon, Project Director

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

1. Introduction.

Consider the linear regression model

$$Y = X\beta + e\ , \tag{1.1}$$

where Y is an $(n \times 1)$ vector of observations on a dependent variable, X is an $(n \times r)$ nonstochastic matrix (we assume full column rank for convenience) of observations on r explanatory variables, β is an $(r \times 1)$ vector of regression coefficients, and e is an $(n \times 1)$ vector of disturbances, normally distributed with zero expectation and covariance matrix $\sigma^2 I$. Estimators of β improving upon the least squares estimator,equivalently the maximum likelihood estimator, have been extensively discussed. See, for example, Judge and Bock (1978). Improved estimation of the disturbance variance σ^2 seems to have been generally overlooked. The usual estimator of σ^2, $(Y-\hat{Y})^t(Y-\hat{Y})/(n-r)$ is best unbiased but is inadmissible under squared error loss (SEL), $L(\sigma^2,\hat{\sigma}^2) = (\sigma^2 - \hat{\sigma}^2)^2$. It is immediately dominated by the best invariant estimator based upon the error sum of squares,

$$(Y-\hat{Y})^t(Y-\hat{Y})/(n-r+2). \tag{1.2}$$

However, (1.2) is also inadmissible. In the sequel, we develop a class of estimators of σ^2 which arise naturally in a regression model and dominate (1.2) in terms of the mean square error (MSE), $EL(\sigma^2,\hat{\sigma}^2)$. The roots of this problem date to Stein (1964), who showed that if $X_1,\ldots,X_n$ are $N(\mu,\sigma^2)$, then $\Sigma(X_i-\bar{X})^2/(n+1)$, the best invariant estimator of σ^2 based upon $S = \Sigma(X_i - \bar{X})^2$, is inadmissible, i.e., for fixed $\mu = \mu_0$,

$$\delta(S,\bar{X}) = \min\{\Sigma(X_i-\mu_0)^2/(n+2),\ \Sigma(X_i-\bar{X})^2/(n+1)\}$$

dominates the best invariant estimator. The interpretation is that for some samples the best invariant estimator is too large. Interestingly, $\delta(S,\bar{X})$ does not dominate $\Sigma(X_i-\mu_0)^2/(n+2)$. The latter does better in a neighborhood of μ_0. Brown (1968) subsequently extended this idea to a different class of dominating estimators. For further reference, see Brewster and Zidek (1974), Strawderman (1974), and Gelfand and Dey (1986). The key point is that $(S,\bar{X})$ is a version of the sufficient statistic for this problem, that $\bar{X}$ contains information about σ^2, and that estimators based upon S only will be inadmissible. In the regression model (1.1), a version of the sufficient statistic is $(\hat{\beta} = (X'X)^{-1}X'Y,\ SSE = (Y - X\hat{\beta})'(Y - X\hat{\beta}))$ which suggests why (1.2) will be inadmissible. Our development in extending Stein's work is briefly alluded to in Klotz, Milton and Zacks (1969, p. 1392) in the context of variance components. In Section 2, we offer a general development of dominating estimators of σ^2 for this type of problem. In Section 3, we look specifically at the example of the regression model in (1.1). We note that these results will be applicable to more complicated econometric models.

2. Development of Dominating Estimators.

Suppose we observe several independent chi-square random variables as $S_0 \sim \sigma^2\chi^2_{n_0}$ and $S_i \sim \sigma^2\chi^2_{n_i,\lambda_i}$, $i = 1,\ldots,p$, i.e., S_0 is central chi-square with n_0 as d.f. and the S_i are noncentral chi-square with n_i d.f. and noncentrality parameter λ_i. Define $c_j = \sum_{i=0}^{j} n_i + 2$, $T_j = c_j^{-1} \sum_{i=0}^{j} S_i$ and finally

$$\delta_j = \min(T_0, T_1, \ldots, T_j), \; j = 0,1,\ldots,p. \tag{2.1}$$

Then we have

Theorem 2.1. In estimating σ^2 under squared error loss, the following holds:

$$\delta_0 << \delta_1 << \delta_2 << \ldots << \delta_p \, ,$$

where $\delta_i << \delta_j$ means δ_j dominates δ_i.

Proof. Since the MSE of δ_j is $E_{\sigma^2}(\sigma^2 - \delta_j)^2 = \sigma^4 \, E_1(1 - \delta_j)^2$, $j = 0,\ldots,p$, without loss of generality we take $\sigma^2 = 1$. We may consider $S_i | L_i \sim \chi^2_{n_i + 2L_i}$ where $L_i \sim Po(\lambda_i)$ and given $L_1,\ldots,L_p$ the S_i, $i = 0,\ldots,p$, are conditionally independent. Moreover, the variables $U_j = c_j T_j / c_{j+1} T_{j+1}$, $j = 0,1,\ldots,p-1$, are also conditionally independent and for fixed j, $U_0,\ldots,U_{j-1}$ are conditionally independent of T_j.

Thus for any estimator of the form

$$c_j h(U_0, U_1, \ldots, U_{j-1}) T_j \tag{2.2}$$

we may write its MSE at $\sigma^2 = 1$ as

$$E(c_j h(U_0,U_1,\ldots,U_{j-1})T_j - 1)^2$$

$$= E[E\{(h(U_0,U_1,\ldots,U_{j-1})c_j T_j - 1)^2 | L_1,\ldots,L_j\}]$$

$$= E[(c_j - 2 + 2\sum_{i=1}^{j} L_i)(c_j + 2\sum_{i=1}^{j} L_i) E\{h^2(U_0,U_1,\ldots,U_{j-1}) | L_1,\ldots,L_j\}$$

$$- 2(c_j - 2 + 2\sum_{i=1}^{j} L_i) E\{h(U_0,U_1,U_{j-1}) | L_1,\ldots,L_j\}] + 1$$

(using the fact that given $L_1,\ldots,L_j$, $T_j \sim \chi^2_{c_j - 2 + 2\sum_{i=1}^{j} L_i}$

$$= E[\{(h(U_0,U_1,\ldots,U_{j-1}) - (c_j + 2\sum_{i=1}^{j} L_i)^{-1}\}^2 (c_j - 2 + 2\sum_{i=1}^{j} L_i)(c_j + 2\sum_{i=1}^{j} L_i)$$

$$+ 2(c_j + 2\sum_{i=1}^{j} L_i)^{-1}] \tag{2.3}$$

where the expectation in (2.3) is over the U_i given L_i and then over L_i.

From (2.3) we see that replacing h by

$$h^* = \min(h, c_j^{-1}) \tag{2.4}$$

yields an estimator $c_j h^* T_j$ which dominates $c_j h T_j$.

In particular, writing δ_{j-1} in the form (2.2), we have

$$h(U_0,\ldots,U_{j-1}) = \beta(U_0,\ldots,U_{j-2})U_{j-1}$$

where

$$\beta(U_0,\ldots,U_{j-1}) = \begin{cases} \prod_{i=k}^{j-2} U_i/c_k & \text{if } \prod_{i=k}^{k'-1} U_i < c_k/c_{k'},\ k < k', \text{ and} \\ & \phantom{\text{if }} \prod_{i=k'}^{k-1} U_i > c_{k'}/c_k,\ k' < k, \\ c_{j-1}^{-1} & \text{if } \prod_{i=k}^{j-2} U_i \geq c_k/c_{j-1},\ k = 1,2,\ldots,j-2, \end{cases}$$

i.e., $\beta = \prod_{i=k}^{j-2} U_i/c_k = T_k/c_{j-1}T_{j-1}$ if $T_k = \min_{1\leq i\leq j-1} T_i$, $k < j-1$,

and $\beta = c_{j-1}^{-1}$ if $T_{j-1} = \min_{1\leq i\leq j-1} T_i$.

Finally, using h^* in (2.4), we obtain $c_j h^* T_j = \min(c_j h T_j, T_j)$ $= \min(\delta_{j-1}, T_j) = \delta_j$ dominating $c_j h T_j = \delta_{j-1}$.

<u>Remark 2.1</u>. When $p = 1$, we obtain the fact that

$$\delta_1 = \min(S_0/(n_0+2),\ (S_0+S_1)/(n_0+n_1+2))$$

dominates $S_0/(n_0+2)$, which includes Stein's (1964) result. We note, possibly counter to one's intuition, that δ_0 does not dominate $(S_0+S_1)/(n_0+n_1+2)$. In particular, when the noncentrality parameter λ_1

is very small, the MSE of δ_0 is only slightly smaller than $2/(n_0+2)$ (see, for example, Brown (1968)), while that of $(S_0+S_1)/(n_0+n_1+2)$ is only slightly greater than $2/(n_0+n_1+2)$. Stated another way, δ_1 is too small in the estimation of σ^2 when λ_1 is small. Figure 1 illustrates the situation.

More generally, the estimator δ_j as defined in (2.1) dominates T_0 but not $T_1,\ldots,T_j$.

<u>Remark 2.2</u>. Theorem 2.1 establishes the inadmissibility of nontrivial scale preserving estimators of the form (2.2). It can be generalized to the estimation of σ^m by extending the discussion in Gelfand and Dey (1986). We omit the details here.

Next we state as Theorem 2.2, an extension of Theorem 2.1 for $p = 1$. The extension for general p is apparent.

<u>Theorem 2.2</u>. Suppose $S_0 \sim \sigma^2\chi^2_{n_0}$, $S_1 \sim (\sigma^2+\tau^2)\chi^2_{n_1}$ and S_0, S_1 are independent. Using the notation of Theorem 2.1, in estimating σ^2, δ_1 dominates δ_0 under SEL.

<u>Proof</u>. We may think of S_1 as arising from $S_1|W_1 = w_1 \sim \sigma^2\chi^2_{n_1,w_1}$, where $W_1 \sim (\tau^2/2\sigma^2)\chi^2_{n_1}$, so that the resulting marginal distribution of S_1 is $(\sigma^2+\tau^2)\chi^2_{n_1}$.

Hence, by Theorem 2.1, regardless of the given W_1, $\delta_0 << \delta_1$ whence integrating over W_1 yields the result.

<u>Remark 2.3</u>. Theorem 2.2 and its extension to general p finds immediate application to the estimation of the error variance in balanced variance components models. See, Klotz, Milton, and Zacks (1969) for discussion in the one-way layout.

Remark 2.4. As in Remark 2.2, Theorem 2.2 can be generalized to the estimation of σ^m.

Returning to the setting of Theorem 2.1, let $\alpha = (\alpha_1,\ldots,\alpha_p)$ be any permutation of the integers $1,\ldots,p$ and let $c_j^\alpha = n_0 + \sum_{i=1}^{j} n_{\alpha_i} + 2$. Also, define $T_j^\alpha = (c_j^\alpha)^{-1}(\sum_{i=1}^{j} S_{\alpha_i} + S_0)$, where as before $S_0 \sim \sigma^2\chi^2_{n_0}$, $S_{\alpha_i} \sim \sigma^2\chi^2_{n_{\alpha_i},\lambda_{\alpha_i}}$ and S_0 and the S_{α_i} are all independent. Finally, define

$$\delta_j^\alpha = \min(T_0, T_1^\alpha,\ldots,T_j^\alpha) \ . \tag{2.6}$$

Then Theorem 2.1 implies that in estimating σ^2 under SEL, for each α $\delta_0 << \delta_1^\alpha << \delta_2^\alpha << \ldots << \delta_p^\alpha$. Hence we obtain p! estimators, δ_p^α each defined by a permutation of $1,\ldots,p$. These estimators are order dependent (i.e., dependent upon the specification of a particular permutation) and a natural question to ask is how to combine these to construct a permutation invariant estimator. A first thought is $\delta^* = \min \delta_p^\alpha$. The discussion in Remark 2.1 shows that δ^* will be "too small" and will not dominate δ_p^α. A better choice will be

$$\overline{\delta} = (p!)^{-1}\Sigma\delta_p^\alpha \tag{2.7}$$

where the summation is over all permutations of $1,\ldots,p$. In particular, we have

Theorem 2.3. If the MSE of δ_p^α is constant, say m, for all α, then $\delta_p^\alpha << \overline{\delta}$ for any permutation $\alpha = (\alpha_1,\ldots,\alpha_p)$.

Proof. The MSE of $\bar{\delta}$ is

$$E(\bar{\delta} - 1)^2 = E[(p!)^{-1} \Sigma(\delta_p^\alpha - 1)]^2$$
$$= (p!)^{-1} m + (p!)^{-2} \sum_{\alpha \neq \alpha'} E(\delta_p^\alpha - 1)(\delta_p^{\alpha'} - 1)$$
$$= m + (p!)^{-2} \sum_{\alpha \neq \alpha'} [E(\delta_p^\alpha - 1)(\delta_p^{\alpha'} - 1) - m] < m,$$

since by the Cauchy-Schwartz inequality

$$E(\delta_p^\alpha - 1)(\delta_p^{\alpha'} - 1) \leq \{E(\delta_p^\alpha - 1)^2 \, E(\delta_p^{\alpha'} - 1)^2\}^{\frac{1}{2}} = m.$$

This completes the proof of the theorem.

3. Application to Linear Regression.

In this section we will use the improved estimators of σ^2 as developed in Section 2, in a linear regression context. Consider the linear model (1.1) and suppose we are interested in testing the hypothesis $H'\beta = \xi$. Let R_0^2, R_1^2 be the full model and reduced model error sum of squares respectively, i.e.,

$$R_0^2 = \min_\beta (Y - X\beta)'(Y - X\beta)$$

and

$$R_1^2 = \min_{H'\beta=\xi} (Y - X\beta)'(Y - X\beta).$$

Then, whether or not the hypothesis is true, it follows that (see, e.g., Rao (1973) for details) $R_0^2 \sim \sigma^2\chi^2_{n-p}$ and is independent of $R_1^2 - R_0^2$ $\sim \sigma^2\chi^2_{k,\lambda}$ where k = rank (H) and λ is the resulting noncentrality parameter. Thus from Remark 2.1, it follows that

$$\delta_1 = \min\left(\frac{R_0^2}{n-r+2}, \frac{R_1^2}{n-r+k-2}\right) \tag{3.1}$$

dominates $R_0^2/(n-r+2)$ which is the best invariant estimator of σ^2 of the form cR_0^2.

The estimator 3.1 can also be viewed as a preliminary test estimator for testing the null hypothesis that $H'\beta = \xi$ vs. the alternative $H_1 : H'\beta \neq \xi$. For a definition and discussion of preliminary test estimators, see Judge and Bock (1978).

Let us now consider general p. We presume a sequence of nested hypothesis as given below:

$$H_0 : X\beta \in \mathcal{M}(X) \qquad \text{with } \dim(\mathcal{M}(X)) = r,$$

$$H_i : X\beta \in \mathcal{S}_i \subset \mathcal{M}(X) \text{ with } \dim(\mathcal{S}_i) = k_i,\ i = 1,\ldots,p,$$

where $\mathcal{M}(X)$ denotes the linear manifold of X (i.e., the vector space generated by columns of X), $\mathcal{S}_1 \subset \mathcal{S}_2 \subset \ldots \subset \mathcal{S}_p$, and $k_1 > k_2 > \ldots > k_p$.

Now define

$$R_i^2 = \min_{X\beta\in\mathcal{S}_i} (Y - X\beta)'(Y - X\beta),\ i = 1,\ldots,p,$$

and again let R_0^2 be the full model error sum of squares. Thus it follows that

$$S_i = R_{i-1}^2 - R_i^2 \sim \sigma^2\chi^2_{n_i,\lambda_i},\ i = 1,\ldots,p,$$

and S_i's are independent and also independent of $S_0 = R_0^2 \sim \sigma^2\chi^2_{n_0}$ where $n_0 = n - r$, $n_1 = r - k_1$, $n_i = k_{i-1} - k_i$, $i = 2,\ldots,p$. Thus we are in the framework of Section 2. Applying Theorem 2.1, we obtain the improved estimator of σ^2 as

$$\delta_p = \min\left(\frac{R_0^2}{n-r+2}, \frac{R_1^2}{n+2-k_1}, \frac{R_2^2}{n+2-k_2}, \ldots, \frac{R_p^2}{n+2-k_p}\right). \tag{3.2}$$

Computation of (3.2) should present no problem since the R_i^2 are obtained in fitting the nested models.

In the special case where we are looking at the r explanatory variables individually, we have r! sequences in which the variables can be removed. For any particular sequence, α, using the notation of (2.6) in (3.2), we obtain

$$\delta_p^\alpha = \min\left\{ \frac{R_0^2}{n-r+2}, \frac{R_{\alpha_1}^2}{n-r+1}, \ldots, \frac{R_{\alpha_p}^2}{n+2} \right\}. \tag{3.3}$$

In obtaining an estimator of σ^2 we then have two possibilities.

(i) If, on the basis of prior experience or theoretical grounds, we have a particular sequence in which the variables are to be entered, hence removed, then this sequence provides (3.3). If, however, this sequence arises from some formal variable selection procedure, then α is data dependent whence the resultant (3.3) does not meet the assumptions of Theorem 2.1, so that no claims can be made for its MSE performance. In fact, such a δ_p^α is nearly $\min_\alpha \delta_p^\alpha$ which from remarks after (2.6) will likely be "too small."

(ii) Calculate $\bar{\delta}$ as in (2.7). We may argue that the MSE's of the δ_p^α are likely to be close. First we expect that each of the δ_p^α achieves small improvement in MSE over $R_0^2/(n-r+2)$ (see, e.g., Brown 1968 for some empirical evidence) and second, the sequence of denominators in (3.3) is the same regardless of α. Theorem 2.3 thus encourages $\bar{\delta}$.

References

Brewster, J.F., and Zidek, J., 1974, Improving on equivariant estimators, Annals of Statistics 2, 21-38.

Brown, L.D., 1968, Inadmissibility of the usual estimators of scale parameters in problems with unknown location and scale parameters, Annals of Mathematical Statistics 39, 29-48.

Gelfand, A.E., and Dey, D.K., 1986, On the estimation of a variance ratio, Unpublished manuscript.

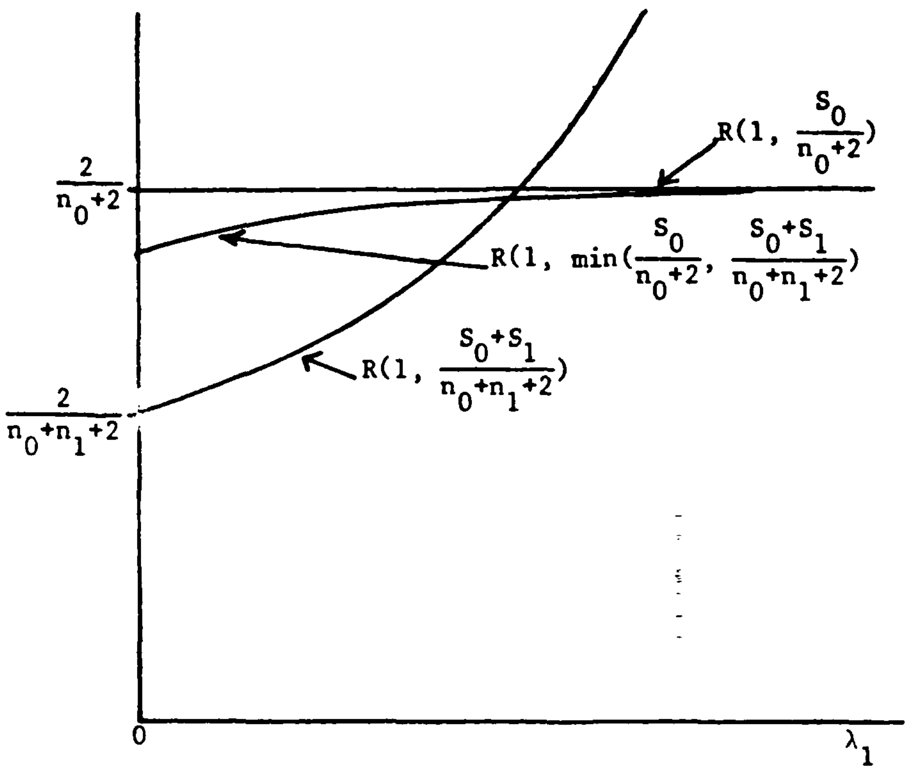
Judge, G.G., and Bock, M.E., 1978, The statistical implications of pre-test and Stein-rule estimators in econometrics, Amsterdam: North Holland.

Klotz, J.H., Milton, R.C., and Zacks, S., 1969, Mean square efficiency of estimators of variance components, Journal of the American Statistical Association 64, 1383-1402.

Rao, C.R., 1973, Linear Statistical Inference and Its Applications, New York: John Wiley and Sons.

Stein, C., 1964, Inadmissibility of the usual estimator for the variance of a normal distribution with unknown mean, Annals of the Institute of Statistical Mathematics 42, 385-388.

Strawderman, W.E., 1974, Minimax estimation of powers of the variance of normal population, Annals of Statistics 2, 190-198.



$R(1, \frac{S_0}{n_0+2})$
$\frac{2}{n_0+2}$
$R(1, \min(\frac{S_0}{n_0+2}, \frac{S_0+S_1}{n_0+n_1+2}))$
$R(1, \frac{S_0+S_1}{n_0+n_1+2})$
$\frac{2}{n_0+n_1+2}$
0
λ_1

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER 418	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Improved Estimation Of The Disturbance Variance In A Linear Regression Model		5. TYPE OF REPORT & PERIOD COVERED TECHNICAL REPORT 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Alan E. Gelfand and Dipak K. Dey		8. CONTRACT OR GRANT NUMBER(s) N00014-89-J-1627
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics Stanford University Stanford, CA 94305		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR-042-267
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Statistics & Probability Program Code 1111		12. REPORT DATE July 11, 1989 13. NUMBER OF PAGES 13
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

APPROVED FOR PUBLIC RELEASE: DISTRIBUTION UNLIMITED

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

regression models; variance estimation

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

In this paper, we consider estimation of the disturbance variance in a linear regression model. A new class of improved estimators is obtained by extending results dating to Stein (1964). These estimators dominate the ordinary least square estimator under squared error loss.