1978

Evaluation of the Mean Squared Error of Certain Generalized Ridge Estimators

H. D. Vinod

Aman Ullah

Krishna Kadiyala

20156

RESEARCH REPORT 7832

EVALUATION OF THE MEAN SQUARED ERROR OF CERTAIN GENERALIZED RIDGE ESTIMATORS

by

H. D. Vinod
Aman Ullah
Krishna Kadiyala

November, 1978

Evaluation of the Mean Squared Error of Certain Generalized Ridge Estimators

H. D. Vinod
AT&T

Aman Ullah
University of Western Ontario

Krishna Kadiyala
Algoma University College

ABSTRACT

This paper studies the mean squared error (MSE) properties of Hoerl and Kennard's (1970) generalized Ridge Regression (RR) estimators. We make extensive use of so called G( ) functions to provide both exact and asymptotic approximations to the MSE. The results of a limited simulation are also reported.

1. INTRODUCTION

Vinod's (1978a) survey notes that the performance of the "generalized" RR is poor compared to that of the "ordinary" RR in several simulations. This seems to be related to the well known "admissibility" of ordinary least squares (OLS) in the unit dimensional case mentioned in Stein (1956), Bunke (1975), among others.

Definition of an admissible estimator

Let a px1 vector $\underset{\sim}{b}$ be an estimator of $\underset{\sim}{\beta}$, and let $WMSE(\underset{\sim}{b})=E(\underset{\sim}{b}-\underset{\sim}{\beta})'\underset{\sim}{D}(\underset{\sim}{b}-\underset{\sim}{\beta})$, be the "weighted" MSE, i.e., a quadratic loss function for a specified diagonal matrix $\underset{\sim}{D}$. The estimator $\underset{\sim}{b}$ is said to be admissible under $WMSE(\underset{\sim}{b})$ if there exists no other estimator $\underset{\sim}{b}^*$ such that $WMSE(\underset{\sim}{b}^*) \leqslant WMSE(\underset{\sim}{b})$ with strict inequality for at least one value of $\underset{\sim}{\beta}$.

Stein (1956) gives several earlier references to the result that the usual maximum likelihood (OLS) estimator of $\underset{\sim}{\beta}$ is admissible for p=1. In "generalized" RR a shrinkage type modification is performed for each dimension separately by choosing the biasing parameters k_i ($i=1,..,p$). The admissibility of OLS for $p=1$ implies that any operational formulas for the choice of individual k_i (i.e. which do not depend on unknown parameters) cannot reduce the MSE of OLS everywhere in the parameter space. By contrast, the so called Stein-Rule or related shrinkage methods can reduce the MSE of OLS for $p>2$ simply because three or more choices of k_i are somehow coordinated in these methods.

In section 2 the biasing parameter k_i used for generalized ridge estimator (GRE) is expressed in terms of two constants f_1 and f_2. Hoerl and Kennard's (1970) first step in their iterative GRE amounts to choosing $f_1=1$ and $f_2=0$. Similarly, Hemmerle and Brantle's (1978) proposal when no constraints are imposed amounts to a choice $f_1=f_2=1$. Other forms of GRE's proposed in the literature are also seen to be special choices for f_1 and f_2. Hence these are called members of our double f class (DFC) estimators.

Section 3 studies the theoretical bias and MSE of DFC using the so called g( ) functions based on confluent hypergeometric functions. The resulting exact results and asymptotics are stated as functions of f_1 and f_2. Section 4 gives proofs and Section 5 briefly reports the results of a limited simulation. The Appendices give the theory of g functions adapted to our context.

2. THE MODEL

Let us write the standard linear regression model as

$$\underset{\sim}{y} = \underset{\sim}{X}\underset{\sim}{\beta} + \underset{\sim}{u} \tag{2.1}$$

where $\underset{\sim}{y}$ is a Tx1 vector of observations on the dependent variable, $\underset{\sim}{X}$ is a Txp matrix of p explanatory variables, $\underset{\sim}{\beta}$ is a px1 vector of unknown regression coefficients, and $\underset{\sim}{u}$ is a Tx1 vector of unknown disturbances.

We state the following conventional assumptions:

Assumption 1 - The matrix of explanatory variables is nonstochastic and of rank p.

Assumption 2 - The disturbance vector $\underset{\sim}{u}$ is distributed as multivariate normal with mean vector zero and covariance matrix $\sigma^2 \underset{\sim}{I}$, i.e.,

$$\underset{\sim}{u} \sim N(\underset{\sim}{0}, \sigma^2 \underset{\sim}{I})$$

Assumption 3 - The sample size T is greater than the total number of explanatory variables p in (2.1).

The model in (2.1) can be written in the following canonical form:

$$\underset{\sim}{y} = \underset{\sim}{Z}\underset{\sim}{\alpha} + \underset{\sim}{u}, \tag{2.2}$$

where

$$\underset{\sim}{Z} = \underset{\sim}{X}\underset{\sim}{G}, \underset{\sim}{\alpha} = \underset{\sim}{G}'\underset{\sim}{\beta}. \tag{2.3}$$

and $\underset{\sim}{G}$ is a pxp matrix of normalized eigenvectors corresponding to the eigenvalues of $\underset{\sim}{X}'\underset{\sim}{X}$ such that

$$\underset{\sim}{G}'\underset{\sim}{G} = \underset{\sim}{G}\underset{\sim}{G}' = \underset{\sim}{I} \tag{2.4}$$

and

$$\underset{\sim}{Z}'\underset{\sim}{Z} = \underset{\sim}{\Lambda}, \underset{\sim}{X}'\underset{\sim}{X} = \underset{\sim}{G}\underset{\sim}{\Lambda}\underset{\sim}{G}' \tag{2.5}$$

where $\underset{\sim}{\Lambda}$ is a pxp diagonal matrix of eigenvalues of $\underset{\sim}{X}'\underset{\sim}{X}$ as

$$\underset{\sim}{\Lambda} = Diag.\ [\lambda_1, \cdots, \lambda_p]. \tag{2.6}$$

The ordinary least squares (OLS) estimator of $\underset{\sim}{\alpha}$ in (2.2) is

$$\underset{\sim}{a} = \underset{\sim}{\Lambda}^{-1}\underset{\sim}{Z}'\underset{\sim}{y}, \tag{2.7}$$

and its mean, variance and the mean squared error (MSE) are, respectively, given as

$$E\underset{\sim}{a} = \underset{\sim}{\alpha}, V(\underset{\sim}{a}) = \sigma^2\underset{\sim}{\Lambda}^{-1}, \tag{2.8}$$

and

$$MSE(\underset{\sim}{a}) = \underset{\sim}{\sigma}^2\sum_{i=1}^{b}\lambda_i^{-1} = E(\underset{\sim}{a}-\underset{\sim}{\alpha})'(\underset{\sim}{a}-\underset{\sim}{\alpha}). \tag{2.9}$$

Further, we note from (2.3) that

$$\underset{\sim}{a} = \underset{\sim}{G'}\underset{\sim}{b} \tag{2.10}$$

such that

$$MSE(\underset{\sim}{a}) = MSE(\underset{\sim}{b}), \tag{2.11}$$

where $\underset{\sim}{b} = (\underset{\sim}{X'}\underset{\sim}{X})^{-1}\underset{\sim}{X'}\underset{\sim}{y}$ is the OLS estimator of $\underset{\sim}{\beta}$ in (2.1).

The generalized ridge estimator (GRE) of $\underset{\sim}{\alpha}$ in (2.1) given by Hoerl and Kennard (1970a), is formed by adding $k_1, \cdots, k_p$, the "additive eigenvalue inflation factors" to $\lambda_1,..,\lambda_p$. Specifically, it can be written as

$$\begin{aligned} \underset{\sim}{a}^{GRE} &= [\underset{\sim}{\Lambda} + \underset{\sim}{K}]^{-1}\underset{\sim}{Z'}\underset{\sim}{y} \\ &= \underset{\sim}{\Delta}\underset{\sim}{a}, \end{aligned} \tag{2.12}$$

where $\underset{\sim}{K}$ and $\underset{\sim}{\Delta}$ are pxp diagonal matrices given, respectively, as

$$\underset{\sim}{K} = Diag\ [k_1, \cdots, k_p], \tag{2.13}$$

$$\underset{\sim}{\Delta} = Diag\ [\delta_1, \cdots, \delta_p], \tag{2.14}$$

where $\delta_i = \lambda_i(\lambda_i+k_i)^{-1}, for\ i=1,...,p$; and $\underset{\sim}{a}$ is the OLS estimator, (2.7), of the canonical model. For the original model (2.1) the GRE of $\underset{\sim}{\beta}$ is written as

$$\underset{\sim}{b}^{GRE} = \underset{\sim}{G}\underset{\sim}{a}^{GRE} = (\underset{\sim}{X'}\underset{\sim}{X}+\underset{\sim}{G}\underset{\sim}{K}\underset{\sim}{G'})^{-1}\underset{\sim}{X'}\underset{\sim}{y}. \tag{2.15}$$

In the particular case when $k_1=k_2=..=k_p=k$ we have $\underset{\sim}{K}=k\underset{\sim}{I}$, and (2.12) is the "ordinary" ridge estimator (ORE) given by Hoerl and Kennard (1970).

A major problem associated with GRE of (2.12) is the determination of the unknown matrix $\underset{\sim}{K}$. A solution to this problem and related ramifications are studied in the following subsection.

2.1 Determination of $\underset{\sim}{K}$

Let us write the MSE of a_i, the i-th component of the OLS estimator a, from (2.7) as

$$MSE(a_i) = \sigma^2/\lambda_i, i = 1,\dots,p. \tag{2.16}$$

Similarly, it can be easily verified that

$$MSE(a_i^{GRE}) = \frac{k_i^2\alpha_i^2}{(\lambda_i+k_i)^2} + \frac{\sigma^2\lambda_i}{(\lambda_i+k_i)^2}, \tag{2.17}$$

where a_i^{GRE} is the i-th component of $\underset{\sim}{a}^{GRE}$.

Thus

$$MSE(a_i^{GRE}) - MSE(a_i) = \frac{k_i\lambda_i(k_i\alpha_i^2-2\sigma^2) - \sigma^2k_i^2}{\lambda_i(\lambda_i+k_i)^2}. \tag{2.18}$$

It follows from (2.18) that for

$$0 < k_i < \frac{2\sigma^2}{\alpha_i^2-\sigma^2\lambda_i^{-1}}, \; (\lambda_i\alpha_i^2>\sigma^2) \tag{2.19}$$

$MSE(a_i^{GRE})$ is smaller than $MSE(a_i)$. In fact a sufficient condition under which $MSE(a_i^{GRE}) < MSE(a_i)$ is that

$$0<k_i < 2\sigma^2\alpha_i^{-2} \tag{2.20}$$

The upper bound of k_i in (2.20), however, is more conservative than in (2.19). An alternative conservative range of k_i, given in Hoerl and Kennard (1970), is

$$0<k_i < \sigma^2\alpha_i^{-2} \tag{2.21}$$

It is to be noted that $k_i = \sigma^2/\alpha_i^2$ is the optimal value of k_i for which $MSE(a_i^{GRE})$ is minimum.

The k_i in (2.19) (2.20) and (2.21) for which GRE dominates OLS can be written in a compact and general form as follows:

$$k_i = \frac{f_1\sigma^2}{\alpha_i^2 - f_2\sigma^2\lambda_i^{-1}},\ i=1,...,p. \tag{2.22}$$

where f_1 *and* f_2 *are arbitrary scalars which could be stochastic or non-stochastic. For* $0<f_1<2$, *and* $f_2=1$ *(2.22) becomes (2.19) and similarly (2.20) and (2.21) are special cases of (2.22).*

We note that, since k_i *in (2.22) depends on the unknown parameters* α_i *and* σ^2 *it cannot be used to determine GRE in (2.12) or (2.15). However, we can use unbiased estimators* a_i *and* s^2 *for* α_i *and* σ^2, *respectively. The* k_i *in (2.22) can then be determined as*

$$\hat{k}(f_1,f_2)_i = \hat{k}_i = \frac{f_1 s^2}{a_i^2 - f_2 s^2\lambda_i^{-1}};s^2 = \frac{1}{n}\hat{u}'\hat{u},\hat{u} = y - Za \tag{2.23}$$

For $f_1=1$ *and* $f_2=0$ *we get*

$$(\hat{k}_{1,0})i = s^2/a_i^2 \tag{2.24}$$

which is suggested by Hoerl and Kennard (1970) as their first iteration. Further for $f_1=2$ *and* $f_2=2$ *we get*

$$(\hat{k}_{2,2})i = 2s^2/(a_i^2 - 2s^2\lambda_i^{-1}) \tag{2.25}$$

based on Stein's "unbiased" estimate of MSE discussed in Vinod (1977). Also for $f_1=2$ *and* $f_2=1$ *we obtain*

$$(\hat{k}_{2,1})_i = 2s^2/(a_i^2 - s^2\lambda_i^{-1}), \tag{2.26}$$

which corresponds to the upper bound of k_i *in (2.19), and is motivated in Vinod (1977). Recently, Hemmerle and Brantle (1978) study* $f_1=f_2=1$. *based on their minimization of an "unbiased" estimate of MSE.*

Although for some fixed positive k_i *given by (2.22) the MSE*$(\underset{\sim}{a}^{GRE})$ *is smaller than that of OLS, this result may not hold for stochastic* k_i *given by (2.23). In fact our following section shows that for Hoerl and Kennard's (1970) choice* $f_1=1$ *and* $f_2=0$ *given in (2.24) the MSE of GRE is high, and GRE does not dominate OLS. Similar lack of dominance can be shown for (2.25) and (2.26). We have noted earlier that for the case when* $p=1$ *it is well known that the OLS estimator is "admissible", i.e., cannot be dominated everywhere. The following section suggests a procedure for comparing MSE's for alternative choices of* f_1 *and*

f_2 values.

2.2 A family of Double f-class Estimators

If we substitute (2.23) in (2.12) we can write a family of "double f-class" (DFC) estimators as

$$\underset{\sim}{a}^{DFC}(f_1,f_2) = \underset{\sim}{\hat{\Delta}}\,\underset{\sim}{a}, \tag{2.27}$$

where

$$\underset{\sim}{\hat{\Delta}} = Diag.\ [\hat{\delta}_1,\ldots,\hat{\delta}_p], \tag{2.28}$$

is a pxp matrix of shrinkage factors such that for $i=1,\ldots,p$,

$$\hat{\delta}_i = \lambda_i(\lambda_i+\hat{k}_i)^{-1} = [1-\frac{f_1 s^2}{\lambda_i a_i^2+(f_1-f_2)s^2}]. \tag{2.29}$$

As noted before f_1 and f_2 are arbitrary scalars which could be stochastic or nonstochastic. The estimator in (2.27) is an "operational" GRE. The i-th component of $\underset{\sim}{a}^{DFC}(f_1,f_2)$ can be written as

$$a_i^{DFC}(f_1,f_2) = \hat{\delta}_i a_i = [1-\frac{f_1 s^2}{\lambda_i a_i^2+(f_1-f_2)s^2}]a_i. \tag{2.30}$$

It is interesting to note that for $f_1=0$, the double f-class estimator reduces to the OLS estimator. Further, for $f_1=1$ and $f_2=0$

$$a_i^{DFC}(1,0) = [1-\frac{s^2}{\lambda_i a_i^2+s^2}]a_i, \tag{2.31}$$

which is the i-th component of the Hoerl and Kennard's (1970a) estimator having $\hat{k}_i$ given in (2.24). Also, $a^{DFC}(2,1)$ for $f_1=2$ and $f_2=1$ is an estimator motivated by Vinod (1977) having $\hat{k}_i$ given in (2.26). Clearly, no DFC estimator can dominate OLS since OLS is known to be admissible in this (one parameter) case.

For $f_2=f_1$, the i-th component of $\underset{\sim}{a}^{DFC}(f_1,f_2)$ can be written as

$$a_i^{DFC}(f_1,f_1) = [1-\frac{f_1 s^2}{\lambda_i a_i^2}]a_i. \tag{2.32}$$

It can be easily verified that (Hint: $a_i/a_i^2 = 1/a_i$) the moments of any order do not exist when $f_1=f_2$, because the random variable a_i in the denominator can be zero. Two particular cases of this are $f_1=f_2=2$ motivated in Vinod (1977), and $f_1=f_2=1$ suggested in Hemmerle and Brantle (1978).

The above discussion indicates that the estimators $\underset{\sim}{a}^{DFC}(1,0)$, $\underset{\sim}{a}^{DFC}(2,1)$, $\underset{\sim}{a}^{DFC}(1,1)$, and $\underset{\sim}{a}^{DFC}(2,2)$ suggested in the literature so far do not improve over OLS. In the following section, therefore, we analyze the exact and approximate bias, moment matrix and MSE of the double f-class estimator. Further, we obtain f_1 and f_2 for which "approximate" MSE of DFC would be smaller than that of OLS in a certain region of the parameter space.

3. THE BIAS AND MSE OF DFC

3.1 The Exact Results

First, we use (2.30) to write the sampling error of the DFC estimator in (2.27) as

$$\underset{\sim}{a}^{DFC}(f_1,f_2) - \underset{\sim}{\alpha} = (\underset{\sim}{a} - \underset{\sim}{\alpha}) - f_1^* \underset{\sim}{D}\underset{\sim}{a}, \tag{3.1}$$

where $\underset{\sim}{D}$ is a pxp diagonal matrix which can be written as

$$\underset{\sim}{D} = Diag.\ [\underset{\sim}{y}'\underset{\sim}{M}\underset{\sim}{y}/\underset{\sim}{y}'\underset{\sim}{L}_1\underset{\sim}{y}, \ldots, \underset{\sim}{y}'\underset{\sim}{M}\underset{\sim}{y}/\underset{\sim}{y}'\underset{\sim}{L}_p\underset{\sim}{y}], \tag{3.2}$$

and $\underset{\sim}{M}$ and $\underset{\sim}{L}_i$, i=1,...,p, are TxT matrices given by

$$\underset{\sim}{M} = \underset{\sim}{I} - \underset{\sim}{Z}\underset{\sim}{\Lambda}^{-1}\underset{\sim}{Z}', \quad \underset{\sim}{L}_i = f_2^*\underset{\sim}{M} + \underset{\sim}{Z}\underset{\sim}{D}_i\underset{\sim}{Z}', \tag{3.3}$$

where D_i is the pxp diagonal matrix of zero elements except the i-th which is $1/\lambda_i$; and where we denote

$$f_1^* = f_1/n \text{ and } f_2^* = (f_1 - f_2)/n. \tag{3.4}$$

We also note that $\underset{\sim}{M}$ is an idempotent matrix of rank n=T-p. From (3.1) we can write the sampling error of the i-th component as

$$(a_i^{DFC} - \alpha_i) = (a_i - \alpha_i) - f_1^* d_i a_i, \tag{3.5}$$

where the notation $a_i^{DFC}(f_1,f_2)$ *is simplified as* a_i^{DFC}*, and where*

$$d_i = \underset{\sim}{y}'\underset{\sim}{M}\underset{\sim}{y}/\underset{\sim}{y}'\underset{\sim}{L}_i\underset{\sim}{y}, \tag{3.6}$$

is the ith diagonal element of the matrix $\underset{\sim}{D}$ *of (3.2).*

Secondly, according to Assumption 2 we observe that

$$\underset{\sim}{y} \sim N(\underset{\sim}{\bar{y}}, \sigma^2 \underset{\sim}{I}), \underset{\sim}{\bar{y}} = \underset{\sim}{Z}\underset{\sim}{\alpha}. \tag{3.7}$$

From (3.1) we note that to derive the bias and MSE of the double f-class estimators in (3.5) we require the expectations of $d_i a_i, d_i a_i^2$*, and* $d_i^2 a_i^2$*. To obtain these expectations we first derive* Ed_i *and* Ed_i^2 *(see Appendix 1) by using Lemma 1 of Sawa (1972, p. 658), and then use the Ullah and Nagar technique (see Section 4) to get* $Ed_i a_i, Ed_i a_i^2$ *and* $Ed_i^2 a_i^2$*.*

Before stating results, we introduce the following notations and functions for the sake of simplicity of exposition.

$$g_{i,\mu,\nu} = G(1-f_2^*, \theta_i, \frac{n+1}{2}+\mu, \frac{n}{2}+\nu) \tag{3.8}$$

where

$$\theta_i = \frac{\underset{\sim}{\alpha}'\underset{\sim}{Z}\underset{\sim}{L}_i\underset{\sim}{Z}\underset{\sim}{\alpha}}{2\sigma^2} = \frac{\lambda_i \alpha_i^2}{2\sigma^2}. \tag{3.9}$$

All the results stated below hold under the assumptions 1 to 3 of sections 2 and for nonstochastic values of f_1 *and* f_2*. We require* $f_2^* = \left(f_1 - f_2\right)/n$ *to be strictly positive, because for nonstochastic* $f_2^* = o$ *or* $f_2^* < o$ *the moments do not exist as indicated in the Appendix. We are now ready to state the following results.*

THEOREM 1. The exact bias of an element of the double f-class estimator exists for $f_1 - f_2 > 0$*, and it is given as*

$$E(a_i^{DFC} - \alpha_i) = -\frac{1}{2} f_1 \alpha_i g_{i,2,1} \tag{3.10}$$

where $g_{i,2,1}$ *is as given in (3.8) for* $\mu = 2$ *and* $\nu = 1$*.*

Corollary 1. The following results regarding the exact bias of an element of $\underset{\sim}{a}^{DFC}(f_1,f_2)$ *is true.*

a. *The exact relative bias of* a_i^{DFC} *for a given sample size lies in the following range.*

$$-f_1(n+3)^{-1} \leqslant E(a_i^{DFC}-\alpha_i)\alpha_i^{-1} \leqslant 0 \tag{3.11}$$

if $f_1 > o$, *for* $f_1 < 0$ *the inequality will be reversed.*

b. *The exact relative bias of* a_i^{DFC} *is an increasing function of the noncentrality parameter* θ_i *if* $f_1 > o$. *For* $f_1 < o$ *it is a decreasing function of* θ_i.

Note: The results in the above corollary follow by noting that $g_{i,2,1} > 0$ and $\frac{\partial}{\partial \theta_i} g_{i,2,1} < 0$ according to (A.9) of the Appendix. Further, for given T

$$E(a_i^{DFC}-\alpha_i)\alpha_i^{-1}$$

tends to zero in the limit as $\theta_i \rightarrow \infty$, and it tends to $-f_1/n+3$, as $\theta_i \rightarrow 0$, by using (3.10), (A.12) and (A.6) of the Appendix A. This then gives the result in (3.11).

THEOREM 2. The exact MSE of an element of the double f-class estimator exists for $f_1 - f_2 > 0$ *and it is given as*

$$E(a_i^{DFC}-\alpha_i)^2 = \frac{\sigma^2}{\lambda_i} + \alpha_i^2 [f_1(g_{i,2,1} - g_{i,3,1}) + + \frac{(n+2)}{4n} f_1^2 (g_{i,3,2} - g_{i,4,2})] + \frac{\sigma^2}{\lambda_i} [-f_1 g_{i,2,1} + + \frac{(n+2)}{4n} f_1^2 (g_{i,2,2} - g_{i,3,2})]. \tag{3.12}$$

Note: Using (3.12) the MSE of $\underset{\sim}{a}^{DFC}(f_1,f_2)$ can be obtained as

$$E(\underset{\sim}{a}^{DFC}-\underset{\sim}{\alpha})'(\underset{\sim}{a}^{DFC}-\underset{\sim}{\alpha}) = \sum_{i=1}^{p} E(a_i^{DFC}-\alpha_i)^2. \tag{3.13}$$

3.2 Large θ_i Asymptotic Expansion

We now present the asymptotic expansions of the bias and MSE of the double f-class estimator in terms of the inverse of θ_i. These results help in analyzing the complicated expressions of the exact bias and MSE given in (3.10) and (3.12), respectively. We note, however, that the results in Theorems 3 and 4 make sense for sufficiently large θ_i, which according to (3.9) mean relatively "small sigma" in Kadane's (1970, 1971) sense. Thus, the terms of order θ_i^{-1}, θ_i^{-2} are the same as the terms of order σ^2, σ^4, etc. respectively. We can now state the following theorems.

THEOREM 3. The asymptotic expansion of the bias of an element of the double f-class estimator in (3.10) up to the order $1/\theta_i$ is given as

$$E(a_i^{DFC}-\alpha_i)^2 = -\frac{f_1}{2}\frac{\alpha_i}{\theta_i}. \tag{3.14}$$

THEOREM 4. The asymptotic expansion of the MSE of an element of $1/\theta_i^3$ in (3.12) up to the order $1/\theta_i^3$ is given as

$$E(a_i^{DFC}-\alpha_i)^2 = \frac{\sigma^2}{\lambda_i}+\frac{\alpha_i^2 f_1}{4\theta_i^2 n}A_1+\frac{\alpha_i^2 f_1}{8\theta_i^3 n}[3A_1-2(f_1-f_2)A_2], \tag{3.15}$$

where $A_1=f_1(n+2)+2n$, $A_2=\frac{n+2}{n}[f_1(n+4)+3n]$, and $f_1-f_2>o$.

Clearly, (3.15) implies that the asymptotic expansion of the MSE $[\underset{\sim}{a}^{DFC}(f_1,f_2)]$ is obtained by a summation of the right hand side of (3.15) over the range i=1,..,p.

Using the definition $\theta_i=\lambda_i\alpha_i^2/2\sigma^2$ we may rewrite (3.15) as

$$MSE(a_i^{DFC}) = \sigma^2\lambda_i^{-1}+P_1 f_1 A_1+P_2 f_1 A_1-P_3(f_1-f_2)f_1 P_4 A_3, \tag{3.16}$$

where $P_1=\sigma^4/n\lambda_i^2\alpha_i^2$, $P_2=3\sigma^6/n\lambda_i^3\alpha_i^4$, $P_3=2\sigma^6/n\lambda_i^3\alpha_i^4$, $P_4=(n+2)/n$, *and* $A_3=f_1(n+4)+3n$. The notation P_i is chosen to suggest positive quantities since $P_i>o$ for i=1,..,4.

Now, this MSE is less than the $MSE(a)=\sigma^2\lambda_i^{-1}$ provided we have

$$f_1(\theta_i+\frac{3}{2})\ \ P_4^{-1}(A_1/A_3)<f_1(f_1-f_2), \tag{3.17}$$

where f_1 is not cancelled from both sides because it is not assumed to be positive. Now, verify that no choice of f_1 and f_2 (two finite real numbers subject to $f_1-f_2>0$) can satisfy (3.17) everywhere in the parameter space; which is also an implication of "admissibility" of OLS for p=1 mentioned earlier.

For large n the term $P_4^{-1}\ A_1/A_3$ may be approximated by $(f_1+2)/(f_1+3)$. Now, for Hoerl and Kennard's (1970) (first iteration) choice $f_1 = 1$ and $f_2 = 0$ the condition for superiority over OLS for large n becomes

$$(\theta_i + \frac{3}{2})\ \frac{3}{4} < 1, \tag{3.18}$$

which will not hold. Hemmerle and Brantle's (unconstrained) choice $f_1 = f_2 = 1$ was already shown to have infinite MSE. In the present context (large n) the condition for it to reduce the MSE of OLS is

$$(\theta_i + \frac{3}{2})\ \frac{3}{4} < 0, \tag{3.19}$$

which is clearly impossible (since $\theta_i > 0$). For large θ_i, $f_1 > o$, and large n, (3.17) will be satisfied provided f_2 is large and negative. Since θ_i can be infinite whereas f_2 must be finite (3.17) will be violated for large enough θ_i. Thus, our procedure will not reduce the MSE of OLS for all possible values of θ_i, as implied by "admissibility" results. In a simulation in Section 5 we fix $f_1=1$, choose f_2 to be large and negative, and find that this choice reduces the MSE of OLS. When some information about θ_i is available in advance, it may be used to determine appropriate choices of f_1 and f_2.

4. PROOFS OF THEOREMS 1 AND 2

In this section we shall give the proofs of Theorems 1 and 2 stated in Section 3.1.

4.1 Proof of Theorem 1

Let us consider the expectation on both sides of (3.5). Using (2.8) we write

$$E(a_i^{DFC} - \alpha_i) = -f_1^{*} E d_i a_i, \tag{4.1}$$

where $f_1^* = f_1/n$, and $d_i = \underset{\sim}{y}'\underset{\sim}{M}\underset{\sim}{y}/\ \underset{\sim}{y}'\ \underset{\sim i}{L}\ \underset{\sim}{y}$, for i=1,..,p from (3.6) above. To obtain the expectation of $d_i a_i$ on the right of (4.1), we write

$$E\, d_i a_i = \sigma\ \underset{\sim}{\iota}'\ \underset{\sim}{\Lambda}^{-1}\ \underset{\sim}{Z}'\ E\underset{\sim}{z}\, d_i \tag{4.2}$$

where ι' is a 1xp vector of zeros except the i^{th} element which is 1, and

$$\underset{\sim}{z} = \underset{\sim}{y}/\ \sigma \sim N(\underset{\sim}{\bar{z}}, \underset{\sim}{I}),\ \underset{\sim}{\bar{z}} = \underset{\sim}{Z}\underset{\sim}{\alpha}/\sigma, \tag{4.3}$$

by using (3.7). Now, we rewrite

$$d_i = \underset{\sim}{y}'\underset{\sim}{M}\underset{\sim}{y}/\underset{\sim}{y}'\underset{\sim i}{L}\underset{\sim}{y} = \underset{\sim}{z}'\ \underset{\sim}{M}\underset{\sim}{z}/\underset{\sim}{z}'\ \underset{\sim i}{L}\ \underset{\sim}{z}. \tag{4.4}$$

Now we note that

$$E\underset{\sim}{z}\, d_i = E(\underset{\sim}{z} - \underset{\sim}{\bar{z}})d_i + \underset{\sim}{\bar{z}}\, Ed_i, \tag{4.5}$$

$$= \frac{1}{(2\pi)^{T/2}} \int_z (\underset{\sim}{z} - \underset{\sim}{\bar{z}})d_i \exp - \frac{1}{2}\{(\underset{\sim}{z} - \underset{\sim}{\bar{z}})'(\underset{\sim}{z} - \underset{\sim}{\bar{z}})\}d\underset{\sim}{z} + \underset{\sim}{\bar{z}}\, Ed_i,$$

$$= \frac{\partial}{\partial \underset{\sim}{\bar{z}}}\ Ed_i + \underset{\sim}{\bar{z}}\, Ed_i,$$

where

$$Ed_i = \frac{n}{2} G(1 - f_2^*,\ \theta_i,\ \frac{n+1}{2} + 1,\ \frac{n}{2} + 1), \tag{4.6}$$

has been obtained from (B.14) of Appendix B by substituting $k = 1 - f_2^*$ and x by

$$\theta_i = \frac{\underset{\sim}{\alpha}'\underset{\sim}{Z}'\underset{\sim i}{L}\ \underset{\sim}{Z}\underset{\sim}{\alpha}}{2\sigma^2} = \frac{1}{2}\ \frac{\lambda_i \alpha_i^2}{\sigma^2} = \frac{\underset{\sim}{\bar{z}}'\underset{\sim i}{L}\ \underset{\sim}{\bar{z}}}{2} \tag{4.7}$$

Next, noting the fact that

$$\frac{\partial}{\partial \underset{\sim}{\bar{z}}}\ Ed_i = (\frac{\partial}{\partial \theta_i}\ Ed_i)\ \frac{\partial \theta_i}{\partial \underset{\sim}{\bar{z}}} \tag{4.8}$$

and using (A.9) of the appendix, (4.5) can be written as

$$E\underset{\sim}{z}d_i = \frac{n}{2}[g_{i,2,1} - g_{i,1,1}]\underset{\sim}{Z}\underset{\sim}{D}_i\underset{\sim}{Z}'\underset{\sim}{\bar{z}} + \frac{n}{2}\underset{\sim}{\bar{z}}g_{i,1,1}, \tag{4.9}$$

where $\underset{\sim}{D}_i$ is given in (3.3); further, substituting (4.9) in (4.2) we get

$$Ed_ia_i = \frac{n\alpha_i}{2}G(1 - f_2^*, \theta_i, \frac{n+1}{2} + 2, \frac{n}{2} + 1). \tag{4.10}$$

Finally, using (4.10) in (4.1) we obtain the result stated in Theorem 1.

4.2 Proof of Theorem 2

Using (3.5) we write the MSE of the i^{th} component of a_i^{DFC} as

$$E(a_i^{DFC} - \alpha_i)^2 = E(a_i - \alpha_i)^2 + 2f_1^*\alpha_iEa_id_i - 2f_1^*Ea_i^2d_i + f_1^{*2}Ea_i^2d_i^2 \tag{4.11}$$

The first term on the right hand of (4.11) is σ^2/λ_i as given in (2.8). Next, considering the second term we note from (4.10) that

$$E\alpha_ia_id_i = \frac{n\alpha_i^2}{2}G(1 - f_2^*, \theta_i, \frac{n+1}{2} + 2, \frac{n}{2} + 1). \tag{4.12}$$

Now, taking the third term on the right hand of (4.11) we write

$$Ed_ia_i^2 = \sigma^2\underset{\sim}{\iota}'\underset{\sim}{\Lambda}^{-1}\underset{\sim}{Z}'E(\underset{\sim}{\bar{z}}\underset{\sim}{\bar{z}}'d_i)\underset{\sim}{Z}\underset{\sim}{\Lambda}^{-1}\underset{\sim}{\iota} \tag{4.13}$$

where $\underset{\sim}{z}$ and d_i are as defined in (4.3) and (4.4), respectively. Using the procedure in (4.5), we note that

$$\begin{aligned} E\underset{\sim}{\bar{z}}\underset{\sim}{\bar{z}}'d_i &= E\{(\underset{\sim}{z}-\underset{\sim}{\bar{z}})(\underset{\sim}{z}-\underset{\sim}{\bar{z}})' + (\underset{\sim}{z}-\underset{\sim}{\bar{z}})\underset{\sim}{\bar{z}}' + \underset{\sim}{\bar{z}}(\underset{\sim}{z}-\underset{\sim}{\bar{z}})' + \underset{\sim}{\bar{z}}\underset{\sim}{\bar{z}}'\}d_i \\ &= \frac{\partial^2}{\partial\underset{\sim}{\bar{z}}\partial\underset{\sim}{\bar{z}}'}Ed_i + 2\underset{\sim}{\bar{z}}\frac{\partial}{\partial\underset{\sim}{\bar{z}}'}Ed_i + (\underset{\sim}{\bar{z}}\underset{\sim}{\bar{z}}' + \underset{\sim}{I})Ed_i \\ &= \underset{\sim}{Z}\underset{\sim}{D}_i\underset{\sim}{Z}'\underset{\sim}{\bar{z}}\underset{\sim}{\bar{z}}'\underset{\sim}{Z}\underset{\sim}{D}_i\underset{\sim}{Z}'\frac{\partial^2}{\partial\theta_i^2}Ed_i + (2\underset{\sim}{\bar{z}}\underset{\sim}{\bar{z}}' + \underset{\sim}{I})\underset{\sim}{Z}\underset{\sim}{D}_i\underset{\sim}{Z}'\frac{\partial}{\partial\theta_i}Ed_i + (\underset{\sim}{\bar{z}}\underset{\sim}{\bar{z}}' + \underset{\sim}{I})Ed_i, \end{aligned} \tag{4.14}$$

where Ed_i is as given in (4.6). Further, using (A.9) and (A.10) of Appendix A, (4.14) can be simplified, and (4.13) can be written as

$$Ed_i a_i^2 = \frac{n}{2}[\alpha_i^2 g_{i,3,1} + \frac{\sigma^2}{\lambda_i} g_{i,2,1}]. \tag{4.15}$$

Similarly, considering the fourth term on the right hand side of (4.11) we first note that

$$E\underset{\sim}{\bar{z}}\underset{\sim}{\bar{z}}'d_i^2 = \underset{\sim}{Z}\underset{\sim}{D}_i\underset{\sim}{Z}'\underset{\sim}{\bar{z}}\underset{\sim}{\bar{z}}'\underset{\sim}{Z}\underset{\sim}{D}_i\underset{\sim}{Z}' \frac{\partial^2}{\partial\theta_i^2} Ed_i^2 + (2\,\underset{\sim}{\bar{z}}\underset{\sim}{\bar{z}}' + \underset{\sim}{I})\,\underset{\sim}{Z}\underset{\sim}{D}_i\underset{\sim}{Z}' \frac{\partial}{\partial\theta_i} Ed_i^2 \tag{4.16}$$

$$+ (\underset{\sim}{\bar{z}}\underset{\sim}{\bar{z}}' + \underset{\sim}{I})Ed_i^2$$

where

$$Ed_i^2 = \frac{n(n+2)}{4}[g_{i,1,2} - g_{i,2,2}], \tag{4.17}$$

has been obtained from (B.16) of Appendix B by replacing k by $1 - f_2^*$ and x by θ_i. Now, using the partial derivatives of G given in (A.9) and in (A.10) of Appendix A we can obtain (4.16) and hence $Ea_i^2 d_i^2$ as

$$Ea_i^2 d_i^2 = \sigma^2\, \underset{\sim}{\iota}'\,\underset{\sim}{\Lambda}^{-1}\underset{\sim}{Z}'\,E(\underset{\sim}{\bar{z}}\underset{\sim}{\bar{z}}'d_i^2)\underset{\sim}{Z}\underset{\sim}{\Lambda}^{-1}\underset{\sim}{\iota} \tag{4.18}$$

$$= \frac{n(n+2)}{4}[\alpha_i^2(g_{i,3,2} - g_{i,4,2}) + \frac{\sigma^2}{\lambda_i}(g_{i,2,2} - g_{i,3,2})]$$

Finally, substituting (2.8), (4.12), (4.15) and (4.18) in (4.11) the result stated in Theorem 2 follows.

5. A SIMULATION

We used two (Hald's and Gorman-Toman's) well-known data structures from previously published simulations of RR in Hoerl, Kennard and Baldwin (1975), Lawless and Wang (1976), Vinod (1978b) among others. We simulate "ordinary" as well as "generalized" RR for a useful comparison. These have p=4, T=13 and p=10, T=36 respectively. For the purpose of this simulation we consider a further transformation of (2.1).

$$\underset{\sim}{y} = \underset{\sim}{X}\beta + \underset{\sim}{u} = \underset{\sim}{H}\eta + \underset{\sim}{u} \tag{5.1}$$

where $\underset{\sim}{H} = \underset{\sim}{X}\underset{\sim}{G}\underset{\sim}{\Lambda}^{1/2}$, $\underset{\sim}{X}'\underset{\sim}{X} = \underset{\sim}{G}\underset{\sim}{\Lambda}\underset{\sim}{G}'$ (as before), $\underset{\sim}{H}'\underset{\sim}{H} = \underset{\sim}{I}$, and the unknown parameters are defined by

$$\underset{\sim}{\eta} = \underset{\sim}{\Lambda}^{\frac{1}{2}}\underset{\sim}{G}'\underset{\sim}{\beta} = \underset{\sim}{\Lambda}^{\frac{1}{2}}\underset{\sim}{\alpha}, \tag{5.2}$$

where $\underset{\sim}{\alpha}$ is defined in (2.3).

Note that the MSE of an estimator $\hat{\underset{\sim}{\beta}}$ of $\underset{\sim}{\beta}$ is a weighted sum of the MSE of an estimator $\hat{\underset{\sim}{\eta}}$ of $\underset{\sim}{\eta}$ with weights λ_i^{-1}.

$$E(\hat{\underset{\sim}{\beta}}-\underset{\sim}{\beta})'(\hat{\underset{\sim}{\beta}}-\underset{\sim}{\beta}) = E(\hat{\underset{\sim}{\eta}}-\underset{\sim}{\eta})'\underset{\sim}{\Lambda}^{-1}(\hat{\underset{\sim}{\eta}}-\underset{\sim}{\eta}) \tag{5.3}$$

Let $L^2 = \underset{\sim}{\eta}'\underset{\sim}{\eta}$ denote the squared length of the "true" parameter vector $\underset{\sim}{\eta}$. Analogous to (3.9) let $\theta = \Sigma\theta_i$ denote the non-centrality parameter defined by $2\theta = L^2/\sigma^2$. We choose seven values of θ in the interval (0.5, 1250) leading to seven choices of $\underset{\sim}{\eta}$ for the simulation. It may be argued that a large part of the action insofar as which of the estimators dominates OLS is taking place on a "sphere" selected to have the same L^2 and σ^2. Hence we select $\sigma^2 = 1$ and $L^2 = 2\theta$ in all our experiments.

Given L^2 we choose p random numbers from a uniform distribution in the range -1 to +1 and call them w_i. Then $\eta_i = Lw_i/\left(\Sigma w_i^2\right)^{1/2}$ satisfy $L^2 = \underset{\sim}{\eta}'\underset{\sim}{\eta}$. Thus, the "true" η_i are selected from a scaled unit cube. This conforms with the practitioner's notion that the true regression coefficients can be anywhere in the selected range. The alternative procedure of selecting η_i from a scaled unit ball of radius L was not adopted.

Next, a vector $\underset{\sim}{u}$ of T random normal deviates with mean zero and unit variance is created using a "super duper" random number generator, Marsaglia, et. al. (1973).

The T (=13 or 36) elements of $\underset{\sim}{y}$ are obtained 500 times for each choice of $\underset{\sim}{\eta}$ (based on θ) for the two data structures (p=4, p=10) by adding $\underset{\sim}{u}$ as in (5.1). The OLS estimator of $\underset{\sim}{\eta}$ is $\underset{\sim}{\eta}^o = (\underset{\sim}{H}'\underset{\sim}{H})^{-1}\underset{\sim}{H}'\underset{\sim}{y} = \underset{\sim}{H}'\underset{\sim}{y}$ is then obtained without repeated matrix inversions. Various modifications of OLS compared in our simulation include two ordinary ridge estimators: The first is abbreviated by HKB, for Hoerl, Kennard and Baldwin's (1975) $k = ps^2/\Sigma a_i^{o2}$; and the second by LAWW for Lawless and Wang's (1976) $k=ps^2/\Sigma a_i^{02}\lambda_i$. The well-known "positive part" Stein-Rule estimator, Stein (1956), is abbreviated as STN+.

Bhattacharya's (1966) estimator (BH+) is interpreted in Vinod (1978b) as a complicated generalized ridge estimator. It is guaranteed to have a lower MSE than OLS, and does not belong to the class of double f-class estimators.

A simulation of the DFC (f_1,f_2) estimator requires a careful choice of f_1 *and* f_2 values in addition to those in the literature. An "optimal" choice of f_1 and f_2 values may be based on a minimization of (3.16) subject to the inequality constraint (3.17). The solution based on Kuhn-Tucker conditions depends on n, λ_i and θ_i values; and seems to be too complicated to be useful. Upon simplifying $n \doteq n+1 \doteq \cdots \doteq n+10$, and $\theta \doteq \theta+1 \doteq \cdots \doteq \theta+10$, we find that f_1 may be found by solving the following cubic equation: $f_1^3+4f_1^2+8f_1-1=0$, whose only real solution is $f_1=2.3733 \doteq 2$. The corresponding "largest acceptable" f_2 can be shown to be

$$f_2 = f_1 - \theta_i(f_1+2)/(f_1+3). \tag{5.4}$$

In other words a DFC estimator with $f_1 \doteq 2$, and f_2 given by (5.4) will have the same MSE as OLS. Since the above simplification involving "largest acceptable" f_2 may not be satisfactory, we derive an alternative f_2 as follows. Consider the unconstrained Lagrangian minimization of (3.16) by differentiating it with respect to f_1, setting the derivative equal to zero and solve for f_2 as a function of θ and f_1 to yield

$$f_2(\theta_i,f_1) = \frac{3f_1^2(n+4)+6f_1 n}{2f_1(n+4)+3n} - (\theta_i+\frac{3}{2})\,\frac{2f_1(n+2)+2n}{2f_1(n+4)+3n}. \tag{5.5}$$

Since θ is unknown in practice, we shall simply fix it at 1000 for this simulation. Now, for $f_1=2$ and $\theta=1000$ and n=7 (5.4) implies $f_2=-775$ which is applicable to the p=4 example and abbreviated as DFC(2,-775) in Table 1. The choice $f_1=1$, $f_2=-749$ or DFC(1,-749) considered for the p=4 example also appears in Table 1. Similarly, for the p=10 example, equation (5.4) leads to DFC(1,-781) and DFC(2,-825) estimators which appear in Table 2.

We simulated a few additional choices of f_1 and f_2 values which are not reported in our tables to save space. In general, our theoretical conclusions are supported by the simulations. For example, choices where $f_1=f_2$, (2.32), having infinite MSE in theory, fail to dominate OLS in simulation. We also find that negative choices of f_1 accompanied by either a negative or a large positive choice of f_2 fail to dominate OLS. By analogy with the "positive part" Stein-Rule estimator we imposed the non-negativity constraint on the shrinkage factors in our simulation which is equivalent to the constraint in Hemmerle and Brantle (1978).

For each estimator, including OLS a weighted sum of squares of errors (WSSE) for j^{th} choice of $\underset{\sim}{u}$ is

$$WSSE_j = \sum_{i=1}^{b}(\hat{\eta}_i - \eta_i)\lambda_i^{-1}, \tag{5.6}$$

for j=1,...,500, for the 500 replications. Their average, $MSE = \Sigma_j \, WSSE_j/500$, is reported as the upper figure in Tables 1 and 2 for the p=4 and p=10 structures respectively. Now, let us define $SDE = \Sigma_j \, (WSSE_j - MSE)^2/500$. To assess the sampling variability over the 500 replications we report in the last row, marked "variability" in both tables, the value of $SDE/\sqrt{500}$ associated with the OLS estimator. Similar SDE values for other estimators are not reported for brevity. Had we simulated 10,000 times we would have SDE/100 as a measure of variability which is an order of magnitude smaller than ours. The cost of such a large sample would be high, although it would offer somewhat greater comparative precision among estimators.

In Tables 1 and 2 the upper figure is the MSE value, and the lower figure is the percent of times WSSE of the estimator is *strictly* less than the WSSE of OLS (%) in 500 replications. Although the MSE performance of STN+ is good, it should be remembered that it leaves the signs and relative magnitudes of OLS regression coefficients unchanged; and therefore cannot help solve "wrong sign" type practical problems. Furthermore, the % associated with STN+ in Table 1 for θ=100 and 450 is slightly below 50%. The fact that the performance of "ordinary" ridge estimators HKB and LAWW is poorer here than in simulations reported in the literature may be attributed to the presence of weights λ_i^{-1} in our (5.3) and (5.5). The theoretical result that BH+ guarantees a lower MSE than OLS in all situations is generally supported by high percent numbers in both tables. The MSE performance of

BH+ is impressive. It is clear that Hemmerle and Brantle's (1978) choice DFC(1,1) or Hoerl and Kennard's choice DFC (1,0) cannot be recommended on the basis of this simulation since % values are often very low. Our DFC(2,-775) and DFC(2,-825) in Tables 1 and 2 respectively always have a lower MSE than OLS, with the % almost never falling below 50%.

Clearly, there is room for a more ambitious simulation of choices of f_1 and f_2, and a wider range of choices of the eigenvalue spectrums before definitive conclusions can be reached. The practitioner may simply choose f_1=2 and let f_2 be chosen by graphical RIDGE TRACE methods, making sure that f_2 does not exceed the "largest acceptable" value given in (5.4).

1. MSE and Percent times WSSE is strictly less than the WSSE of OLS for p=4 structure[a]

Estimator	0.5	4.5	θ 24.5	40.5	100	450	1250
OLS	590.07[b] 0.	615.68 0.	620.37 0.	638.60 0.	665.05 0.	642.03 0.	698.67 0.
HKB	155.81 61.	1024.43 20.	1093.68 18.	3677.22 2.	50717.7 0.	97145.2 0.	47207.9 0.
STN+	341.44 72.	543.44* 54.	585.21 60.	618.77* 50.	664.82 49.	642.58 48.	698.09 53.*
BH+	588.91 96.*	615.42 78.*	620.32 63.*	638.57 61.*	665.03 60.*	642.03 53.*	698.67 53.*
LAWW	156.01 61.	1024.26 20.	1089.02 19.	3556.82 2.	50964.8 0.	127047. 0.	32689.5 0.4
DFC(1,1)	156.52 61.	1040.19 20.	1151.02 18.	3981.35 1.	60169.1 0.	174703. 0.	17773.4 0.
DFC(1,0)	155.69* 61.	1020.13 20.	1124.61 18.	3837.11 2.	43962.8 0.	84728.9 0.	82904.0 0.
DFC(2,-775)	587.10 80.	612.50 59.	617.38* 61.	635.12 51.	660.81* 50.	640.76* 48.	695.04* 52.
DFC(1,-749)	588.53 80.	614.03 59.	618.82 61.	636.79 51.	662.75 50.	641.08 48.	696.48 53.*
Variability	37.45	39.31	39.87	40.38	47.46	39.44	42.83

a The eigenvalues λ_i are respectively 2.2357, 1.5761, 0.18661 and 0.0016238.

b The upper figure is the MSE, the lower figure is the %.

* The most favorable number in each column is indicated by a * (having low MSE, high %).

2. MSE and Percent times WSSE is strictly less than the WSSE of OLS for p=10 structure[a]

Estimator	θ 0.5	4.5	24.5	40.5	100	450	1250
OLS	33.18[b] 0.	31.91 0.	33.71 0.	32.85 0.	32.68 0.	32.70 0.	34.06 0.
HKB	6.08 99.	23.93 57.	87.37 7.	33.60 46.	21.77 82.	45.72 28.	50.02 30.
STN+	9.35 99.	20.07* 83.	30.97 56.	29.78* 74.	30.86 82.	32.44* 54.	33.99 50.
BH+	16.12 100.*	22.91 96.	30.69* 83.*	31.42 66.	32.05 73.	32.48 56.*	33.98 60.*
LAWW	5.96* 99.	23.99 57.	78.57 10.	42.52 30.	22.50 85.	58.41 20.	42.25 36.
DFC(1,1)	6.57 98.	30.48 41.	145.54 7.	57.87 27.	29.28 49.	113.76 16.	41.13 40.
DFC(1,0)	8.42 98.	22.92 57.	72.06 19.	33.67 46.	21.34* 67.	75.65 19.	40.77 40.
DFC(2,-825)	33.02 100.*	31.75 99.	33.55 69.	32.70 84.*	32.52 87.	32.57 56.*	33.94* 51.
DFC(1,-781)	33.09 100.*	31.83 99.*	33.62 70.	32.77 84.*	32.60 88.*	32.63 56.*	33.98 52.
Variability	1.12	.97	1.01	.98	1.03	1.03	1.08

a The eigenvalues λ_i are 3.6864, 1.5496, 1.2966, 1.0515, 0.94467, 0.65709, 0.36047, 0.23176, 0.14792 and 0.074018.

b The upper figure is the MSE and the lower figure the %.

* The most favorable number in each column is indicated by a *.

APPENDIX

A. HYPERGEOMETRIC AND G FUNCTIONS

The hypergeometric *functions* ${}_1F_1$ *and* ${}_2F_1$ have the following power series representations [Slater (1960)],

$$ {}_1F_1(a;c;x) = \frac{\Gamma c}{\Gamma a} \sum_{m=0}^{\infty} \frac{\Gamma(a+m)}{\Gamma(c+m)} \frac{x^m}{m!}, \; c > 0, |x| < \infty \tag{A.1} $$

$$ {}_2F_1(a,b;c;x) = \frac{\Gamma c}{\Gamma a \Gamma b} \sum_{m=0}^{\infty} \frac{\Gamma(a+m)\Gamma(b+m)}{\Gamma(c+m)} \frac{x^m}{m!}, \; c > 0, \; |x| < 1 \tag{A.2} $$

Similarly, the power series representation of the function G,

$$ G(k,x;a,c) = \int_{-\infty}^{0} h(t;k,x,a,c)dt \tag{A.3} $$

where $x > 0, a \geqslant b, -1 \leqslant k \leqslant 1$ and

$$ h(t;k,x,a,c) = \frac{2 \, exp[2tx/(1-2t)]}{(1-2t)^{a-c} \, (1-2tl-k)^c} \tag{A.4} $$

is given by [see Sawa (1972, p. 678)].

$$ G = e^{-x} \frac{\Gamma(a-1)}{\Gamma(c)} \sum_{h=0}^{\infty} (k)^h \frac{\Gamma(c+h)}{\Gamma(a+h)} \, {}_1F_1 \, (a-1;a+h,x), \; a>1. \tag{A.5} $$

An alternative representation is

$$ G = e^{-x} \sum_{h=0}^{\infty} \frac{x^h}{h!} \frac{\Gamma(a-1+h)}{\Gamma(a+h)} \, {}_2F_1 \, (1,c;a+h;k), \; a>1. \tag{A.6} $$

For k=1, (A.5) and (A.6) reduce to

$$ G = e^{-x} \frac{\Gamma(a-c-1)}{\Gamma(a-c)} \, {}_1F_1(a-c-1;a-c;x), \; a-c > 1. \tag{A.7} $$

Further, for k=0

$$G = e^{-x} \frac{\Gamma(a-1)}{\Gamma(a)} \,{}_1F_1(a-1;a;x). \tag{A.8}$$

The partial derivatives of G with respect to x can be written as

$$\frac{\partial}{\partial x} G(k,x;a,c) = G(k,x;a+1,c) - G(k,x;a,c) \tag{A.9}$$

$$\frac{\partial^2}{\partial x^2} G(k,x;a,c) = G(k,x;a+2,c) - 2G(k,x;a+1,c) + G(k,x;a,c) \tag{A.10}$$

and so on. Also we note from (A.3) and (A.5) that

$$\frac{\partial}{\partial k} G(k,x;a+1,c) > 0 \text{ and } \frac{\partial}{\partial x} G(k,x;a,c) < 0. \tag{A.11}$$

Finally, for large x, the asymptotic expansion of the function G in (A.5), up to order $\frac{1}{x^4}$, is given by [see Sawa (1972, p. 667)].

$$\begin{aligned} G(k,x;a,c) = {} & \frac{1}{x} + (ck-a+2)\frac{1}{x^2} + [c(c+1)k^2 - 2c(a-2)k \\ & + (a-2)(a-3)]\frac{1}{x^3}, \\ & + Ec(c+1)(c+2)k^3 - 3c(c+1)(a^{-2})k^3 \\ & + 3c(a-2)(a-3)k - (a-2)(a-3)(a-4)]\frac{1}{x4} \end{aligned} \tag{A.12}$$

B. EVALUATION OF SOME EXPECTATIONS REQUIRED IN SECTION 4

Let $\underset{\sim}{z}$ by a Tx1 normally distributed random vector such that

$$E\underset{\sim}{z} = \underset{\sim}{\bar{z}} \text{ and } E(\underset{\sim}{z}-\underset{\sim}{\bar{z}})(\underset{\sim}{z}-\underset{\sim}{\bar{z}})' = \underset{\sim}{I}_T \tag{B.1}$$

Further, consider $\underset{\sim}{M}$ as a TxT idempotent matrix with rank n<T, and the matrix

$$\underset{\sim}{L}_i = f_2^{*} \underset{\sim}{M} + \underset{\sim}{Z}^{*} \underset{\sim}{D}_i \underset{\sim}{Z}' \tag{B.2}$$

where $\underset{\sim}{D}_i$ is a pxp diagonal matrix of zeros except the i^{th} element which is $\frac{1}{\lambda_i}$. These are as given in (3.3). L_i is non-negative definite for $f_2^{*} \geqslant 0$.

The joint moment generating function $M(t_1, t_2)$ of the quadratic forms $\underset{\sim}{z}' \underset{\sim}{L}_i \underset{\sim}{z}$ and $\underset{\sim}{z}' \underset{\sim}{M} \underset{\sim}{z}$ can be written as

$$M(t_1, t_2) = E exp[t_1 \underset{\sim}{z}' \underset{\sim}{L}_i \underset{\sim}{z} + t_2 \underset{\sim}{z}' \underset{\sim}{M} \underset{\sim}{z}] \tag{B.3}$$

$$= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} exp[t_1 \underset{\sim}{z}' \underset{\sim}{L}_i \underset{\sim}{z} + t_2 \underset{\sim}{z}' \underset{\sim}{M} \underset{\sim}{z}] f(z) d\underset{\sim}{z}$$

where $f(\underset{\sim}{z})$ represents the multivariate normal density of $\underset{\sim}{z}$ with mean vector $\underset{\sim}{\bar{z}}$ and covariance matrix $\underset{\sim}{I}$.

Since $\underset{\sim}{M}$ is an idempotent matrix of rank n < T, and $\underset{\sim}{M}$ and $\underset{\sim}{L}_i$ commute we can always obtain an orthogonal matrix $\underset{\sim}{P}$ such that the orthogonal transformation of the matrix

$$\underset{\sim}{Q}_i = \underset{\sim}{I} - 2 t_1 \underset{\sim}{L}_i - 2 t_2 \underset{\sim}{M}, \quad i = 1, \ldots, p \tag{B.4}$$

can be written as $\underset{\sim}{\Lambda}_i^{*}$, which is introduced here to denote

$$\underset{\sim}{P}' \underset{\sim}{Q}_i \underset{\sim}{P} = \begin{bmatrix} & \underset{\sim}{I}_{i-1} & 0 & 0 & 0 \\ & 0 & (1-2t_1) & 0 & 0 \\ & 0 & 0 & \underset{\sim}{I}_{p-i} & 0 \\ 0 & 0 & 0 & (1-2t_1 f_2^{*} - 2t_2) \underset{\sim}{I}_n \end{bmatrix} . \tag{B.5}$$

Since $\underset{\sim}{Z}' \underset{\sim}{Z} = \underset{\sim}{\Lambda}$ from (2.5), note that $\underset{\sim}{Q}_i$ is a TXT diagonal matrix. Thus, if we restrict the domains of t_1 and t_2 as

$$2 t_1 < 1, \textit{ and } 2 t_1 f_2^{*} + 2 t_2 < 1, \tag{B.6}$$

it follows that $\underset{\sim}{Q}_i$ is a non-negative definite matrix.

We can now simplify (B.3) as

$$M(t_1,t_2) = \exp \frac{1}{2}[\bar{z}'Q_i^{-1}\bar{z} - \bar{z}'\bar{z}]|Q_i^{-1}|^{1/2} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(z)dz \quad \text{(B.7)}$$

where $g(z)$ is multivariate normal density of z with mean vector $Q_i^{-1}\bar{z}$ and variance covariance matrix Q_i^{-1}. Finally, using (B.5) and noting that the integral value on the right of (B.7) is unity we obtain

$$M(t_1,t_2) = exp \frac{1}{2}[\bar{z}'Q_i^{-1}\bar{z} - \bar{z}'\bar{z}] \;/\; (1-2t_1)^{\frac{1}{2}}(1-2t_1f_2^* - 2t_2)^{\frac{n}{2}} \quad \text{(B.8)}$$

In this case where the matrix M is such that

$$M\bar{z} = 0 \quad \text{(B.9)}$$

we note

$$Q_i\bar{z} = (I - 2t_1 Z D_i Z')\bar{z} \quad \text{(B.10)}$$

In this case (B.8) is simplified as

$$M(t_1,t_2) = exp\left[\frac{2t_1\theta_i}{1-2t_1}\right] / \; (1-2t_1)^{\frac{1}{2}}(1-2t_1f_2^* - 2t_2)^{\frac{n}{2}} \quad \text{(B.11)}$$

where θ_i is as defined in (4.7). The following derivatives of (B.11) can then be easily verified.

$$\left.\frac{\partial M(t_1,t_2)}{\partial t_2}\right|_{t_2=0} = \frac{n}{2} h(t_1; 1-f_2^*, \theta_i, \frac{n+1}{2}+1, \frac{n}{2}+1), \quad \text{(B.12)}$$

$$\left.\frac{\partial^2 M(t_1,t_2)}{\partial t_2^2}\right|_{t_2=0} = \frac{n(n+2)}{2}\, h(t_1; 1-f_2^*, \theta_i, \frac{n+1}{2}+2, \frac{n}{2}+2), \quad \text{(B.13)}$$

and h( ) is as defined in (A.4).

Finally, using (B.12), (B.13) and (A.3) we obtain the following expectations for $0 \leqslant f_2^* \leqslant 2$ (See Sawa (1972), Williams (1941)).

$$E(\frac{\underset{\sim}{z}'\underset{\sim}{M}\underset{\sim}{z}}{\underset{\sim}{z}'\underset{\sim}{L}_i\underset{\sim}{z}}) = \int_{-\infty}^{0}\left[\frac{\partial M(t_1,t_2)}{\partial t_2}\right]_{t_2=0} dt_1 \tag{B.14}$$

$$= \frac{n}{2}\, G(1-f_2^{\bullet},\theta_i;\frac{n+1}{2}+1,\frac{n}{2}+1)$$

$$E(\frac{\underset{\sim}{z}'\underset{\sim}{M}\underset{\sim}{z}}{\underset{\sim}{z}'\underset{\sim}{L}_i\underset{\sim}{z}})^2 = \int_{-\infty}^{0} -t_1\left[\frac{\partial^2 M(t_1,t_2)}{\partial t_2^2}\right]_{t_2=0} dt_1 \tag{B.15}$$

$$= \frac{n+2}{2}\left[G(1-f_2^{\bullet},\theta_i;\frac{n+1}{2}+1,\frac{n}{2}+2)\right. \tag{B.16}$$

$$\left. - G(1-f_2^{\bullet},\theta_i;\frac{n+1}{2}+2,\frac{n}{2}+2)\right]$$

where use has been made of

$$2t_1 nh(t_1;1-f_2^{\bullet},\theta_i,\frac{n+1}{2}+2,\frac{n}{2}+2) = \frac{\partial}{\partial\theta_i}h(t_1;1-f_2^{\bullet},\theta_i,\frac{n+1}{2}+1,\frac{n}{2}+2) \tag{B.17}$$

and (A.9).

C. EXISTENCE OF THE MOMENTS OF $\underset{\sim}{a}^{DFC}$

First, we note that the multiple integral

$$\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\frac{z_1^{m_1-1}\cdots z_2^{m_T-1}}{(a_1z_1^2+\ldots+a_Tz_T^2)^r}\exp[-\frac{1}{2}\sum_1^T(z_t-\theta_t)^2]dz_1,\cdots,dz_T \tag{C.1}$$

where a_t, $m_t(t=1,\ldots,T)$ and r are positive, converges if $2r < m_1+\ldots+m_T$ (See Gradshteyn, et. al. (1965)).

Now let us write from (2.6) and (4.4)

$$d_ia_i = \frac{\underset{\sim}{y}'\underset{\sim}{M}\underset{\sim}{y}}{\underset{\sim}{y}'\underset{\sim}{L}_i\underset{\sim}{y}}\underset{\sim}{\iota}'(\underset{\sim}{\Lambda})^{-1}\underset{\sim}{Z}'\underset{\sim}{y} = \underset{\sim}{\iota}'\underset{\sim}{A}[\frac{\underset{\sim}{z}^{*\prime}\underset{\sim}{R}_1\underset{\sim}{z}^{*}}{\underset{\sim}{z}^{*\prime}\underset{\sim}{R}_2\underset{\sim}{z}^{*}}\cdot\underset{\sim}{z}_{-1}] \tag{C.2}$$

$$= \underset{\sim}{\iota}'\underset{\sim}{A}\left[\frac{z_{1p+1}^2+\ldots+z_{1T}^2}{z_{1i}^2+f_2^{\bullet}z_{1p+1}^2+\ldots+f_2^{\bullet}z_{1T}^2}\cdot\underset{\sim}{z}_{-1}^{\bullet}\right]$$

where $\underset{\sim}{z}^{*} = \underset{\sim}{P}'\underset{\sim}{y}/\sigma$, *and*

$$\underset{\sim 1}{R} = \underset{\sim}{P}'\underset{\sim}{M}\underset{\sim}{P} \text{ and } \underset{\sim 2}{R} = \underset{\sim}{P}'\underset{\sim i}{L}\underset{\sim}{P}. \tag{C.3}$$

Further, the notation $\underset{\sim 1}{z}^{*}$ *in (C.2) is based on arranging the columns of the TxT orthogonal matrix* $\underset{\sim}{P}$ *as* $\underset{\sim}{P} = [\underset{\sim 1}{P}\underset{\sim 2}{P}]$, *where* $\underset{\sim 1}{P}$ *is a Txp matrix and* $\underset{\sim 2}{P}$ *is a Txn matrix, and accordingly partitioning* $\underset{\sim}{z}^{*\prime} = [\underset{\sim 1}{z}^{*\prime}\underset{\sim 2}{z}^{*\prime}] = [z_{11} \cdots z_{1p}, z_{1p+1} \cdots z_{1T}]$, *where* $\underset{\sim 1}{z}^{*}$ *and* $\underset{\sim 2}{z}^{*}$ *are px1 and T-px1 vectors, respectively. We have written*

$$(\underset{\sim}{\Lambda})^{-1}\underset{\sim}{Z}'\underset{\sim}{y} = (\underset{\sim}{\Lambda})^{-1}\underset{\sim}{Z}'\underset{\sim}{P}\underset{\sim}{P}'\underset{\sim}{y} = \underset{\sim}{A}\underset{\sim 1}{z}^{*},\ A = \sigma(\underset{\sim}{\Lambda})^{-1}\underset{\sim}{Z}'\underset{\sim 1}{P} \tag{C.4}$$

because $(\underset{\sim}{\Lambda})^{-1}\underset{\sim}{Z}'\underset{\sim}{M} = (\underset{\sim}{\Lambda})^{-1}\underset{\sim}{Z}'\underset{\sim 2}{P}\underset{\sim 2}{P}' = 0$ *by using (C.3).*

We can write (C.2) as

$$d_i a_i = \underset{\sim}{\iota}'\underset{\sim}{A}(z_{1p+1}^2 + \cdots + z_{1T}^2)\left[\frac{\bar{z}_1}{z_{1i}^2 + f_2^* z_{1p+1}^2 + \cdots + f_2^* z_{1T}^2}\right] \tag{C.5}$$

Using (C.1), and noting that $\underset{\sim}{z} \sim N(\underset{\sim}{P}'\underset{\sim}{Z}\underset{\sim}{\alpha}, \underset{\sim}{I},)$ *it can easily be verified that for each of the elements appearing in (C.2)*

$$E\left[\frac{z_{1t}^2\, z_{1t'}}{z_{1i}^2 + z_{1p+1}^2 + \cdots + z_{1T}^2}\right],\ \begin{matrix} t = p+1,\ldots,T \\ t' = 1,\ldots,p \end{matrix} \tag{C.6}$$

converges for $f_2^* > o$ *and, therefore,* $E\, d_i a_i$ *exists.*

Since $d_i a_i$ *is required in* $E\, a_i^{DFC}$ *and* $E\, a_i^2 d_i^2$ *in the risk function of* a_i^{DFC} *(see Section 4) we note that the first two moments of* a_i^{DFC} *exist.*

REFERENCES

[1] *Bhattacharya, P. K., (1966) "Estimating the Mean of a Multivariate Normal Population with General Quadratic Loss Function," Annals of Mathematical Statistics,* 37, 1818-1825.

[2] Bunke, O., (1975) "Improved Inference in Linear Models with Additional Information", *Mathematische Operationsforshung Und Statistik,* 6, 817-829.

[3] Gradshteyn, I. S. and Ryzhik, I. M., (1965). *Tables of Integrals Series and Products.* New York: Academic Press.

[4] Hemmerle, W. J. and Brantle, T. F., (1978), "Explicit and Constrained Generalized Ridge Estimation" *Technometrics,* 20, 109-120.

[5] Hoerl, A. E. and Kennard, R. W., (1970a) "Ridge Regression: Biased Estimations for Non-orthogonal Problems," *Technometrics,* 12, No. 1, 55-67.

[6] Hoerl, A. E. and Kennard, R. W., (1970b) "Ridge Regression: Applications to Non-orthogonal Problems," *Technometrics,* 12, No. 1, 69-82.

[7] Hoerl, A. E., Kennard, R. W., and Baldwin, K. F., (1975) "Ridge Regression: Some Simulations," *Communications in Statistics,* 4, 105-123.

[8] Kadane, J. B., (1970) "Testing Overidentifying Restrictions when the Disturbances are Small," *Journal of the American Statistical Association,* 65, 182-85.

[9] Kadane, J. B., (1971) "Comparison of k-class Estimators when the Disturbances are Small," *Econometrica,* 39, 723-738.

[10] Lawless, J. F. and Wang, P. (1976) "A Simulation Study of Ridge and Other Regression Estimators," *Communications in Statistics,* Ser: A, 5 (No. 4), 307-323.

[11] Marsaglia, G., Ananthanarayanan, K., and Paul, N., (1973). "A Super Duper Random Number Generator", unpublished documentation, School of Computer Science, McGill University, Montreal, Canada.

[12] Nagar, A. L. and Ullah, A., (1973) "Note on the Approximate Skewness and Kurtosis of the

Two-Stage Least Squares Estimator," *Indian Economic Review*, 8, 70-80.

[13] Sawa, T., (1972) "Finite-Sample Properties of the k-class Estimators," *Econometrica*, 40, 653-80.

[14] Scolve, S. L., (1968) "Improved Estimators for Coefficients in Linear Regression," *Journal of the American Statistical Association*, 63, 599-606.

[15] Slater, L. J., (1960) *Confluent Hypergeometric Functions*, Cambridge: University Press.

[16] Stein, C., (1956) "Inadmissibility of the Usual Estimator for the Mean of a Multivariate Normal Distribution," *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, 1, 197-206.

[17] Theil, H., (1963) "On the Use of Incomplete Prior Information in Regression Analysis," *Journal of the American Statistical Association*, 58, 401-414.

[18] Ullah, A., and Nagar, A. L., (1974) "The Exact Mean of the Two Stage Least Squares Estimator of the Structural Parameters in an Equation Having Three Endogenous Variables, *Econometrica*, Vol. 42, No. 2, 749-758.

[19] Ullah, A., and Ullah, S., (1978) "Double K-Class Estimators of Coefficients in Linear Regression," *Econometrica*, 46, 705-722.

[20] Vinod, H. D., (1976) "Application of New Ridge Regression Methods to a Study of Bell System Scale Economies," *Journal of the American Statistical Association, 71*, 835-841.

[21] Vinod, H. D., (1977) "Estimating the Largest Acceptable k and a Confidence Interval for Ridge Regression Coefficients," Presented at the Econometric Society European Meeting in Vienna, September 8, 1977.

[22] Vinod, H. D., (1978a) "A Survey of Ridge Regression and Related Techniques for Improvements over Ordinary Least Squares", *Review of Economics and Statistics*, 60, 121-131.

[23] Vinod, H. D., (1978b) "A Ridge Estimator Whose MSE Dominates OLS", *International Economic Review*, 19, 727-737.

[24] Williams, J. A., (1941) "Moments of the Ratio of the Mean Square Successive to the Mean

Square Difference in Samples from a Normal Universe," *Annals of Mathematical Statistics,* 12, 239-41.