Department of Economics
UNIVERSITY OF CANTERBURY
CHRISTCHURCH, NEW ZEALAND

UNBIASED ESTIMATION OF THE
MEAN SQUARED ERROR
FEASIBLE GENERALISED RIDGE REGRESSION ESTIMATOR

V. K. SRIVASTAVA AND D. E. A. GILES

Discussion Paper

No.8907

Department of Economics University of Canterbury
Christchurch, New Zealand

Discussion Paper No. 8907

October 1989

UNBIASED ESTIMATION OF THE
MEAN SQUARED ERROR
FEASIBLE GENERALISED RIDGE REGRESSION ESTIMATOR

V. K. SRIVASTAVA
University of Lucknow
and
D. E. A. GILES
University of Canterbury

UNBIASED ESTIMATION OF THE

MEAN SQUARED ERROR OF THE

FEASIBLE GENERALISED RIDGE REGRESSION

ESTIMATOR

V.K. Srivastava
Department of Statistics
Lucknow University
Lucknow 226007
INDIA

AND

D.E.A. Giles
Department of Economics
University of Canterbury
Christchurch 1
NEW ZEALAND

Abstract

The paper derives an exact unbiased estimator of the mean squared error of the feasible generalised ridge regression estimator for a linear regression coefficient. This provides the basis for calculating the standard error of such an estimator in a meaningful way.

1.   INTRODUCTION

Hoerl and Kennard (1970) suggest the method of ridge regression to deal with the problem of multicollinear data. Subsequently, a variety of ridge and ridge-type estimators have been proposed and their properties derived and compared. However, the practical application of these estimators has been limited by the fact that they produce only point estimates - in general, appropriate formulae for computing the associated "standard errors" are unavailable.

The reason for this is that the expressions for the exact first and second moments of these estimators are extremely complex, and obtaining unbiased estimators of these moments (as are needed to derive standard errors which are meaningful in finite samples) seems to be a non-trivial task.

Although asymptotic standard errors for the feasible ridge regression estimator are easily obtained, they are of limited use. In particular, as they coincide with the least squares standard errors, no distinction can be drawn between the relative estimated efficiencies of the two estimators if this approach is pursued.

In the spirit of results obtained for the Stein-rule estimator by Carter et al. (19**), this paper derives an unbiased estimator of the exact mean squared error of the feasible ridge estimator. The result obtained is of a simple form and is readily computed. This facilitates the computation of meaningful standard errors which have a proper finite-sample justification. We deal only with the generalised feasible ridge estimator, but in principle the same analytical approach could be applied to other related estimators.

2. GENERALISED RIDGE REGRESSION

Consider the model

$$(2.1) \qquad y = X\beta + u ; \qquad u \sim N(0, \sigma^2 I_n)$$

where X is (n×p), and for convenience (and without loss of generality) we take the model to be in canonical form, so that $X'X = \Lambda = \text{diag.}(\lambda_i)$. The Ordinary Least Squares (OLS) estimator of β is

$$(2.2) \qquad b = (X'X)^{-1}X'y = \Lambda^{-1}X'y$$

and this unbiased estimator has covariance matrix

$$(2.3) \qquad V(b) = \sigma^2 \Lambda^{-1} ,$$

for which an unbiased estimator is

$$(2.4) \qquad \hat{V}(b) = s\Lambda^{-1}/\nu ,$$

where $\nu = (n-p)$ denotes degrees of freedom, and $s = (y-Xb)'(y-Xb)$ is the residual sum of squares.

The generalised ridge regression estimator of β is

$$(2.5) \qquad \hat{\beta} = (\Lambda+K)^{-1}X'y ,$$

where $K = \text{diag.}(k_i)$, and the k_i's are biasing parameters. The ith element of $\hat{\beta}$ is

$$(2.6) \qquad \hat{\beta}_i = \lambda_i b_i / (\lambda_i + k_i) ,$$

where b_i is the ith element of b.

If k_i is fixed then $\hat{\beta}_i$ has bias given by

$$(2.7) \qquad B(\hat{\beta}_i) = - \left(\frac{k_i \beta_i}{\lambda_i + k_i} \right)$$

and Mean Squared Error (MSE):

$$(2.8) \qquad M(\hat{\beta}_i) = \frac{\lambda_i \sigma^2 + k_i^2 \beta_i^2}{(\lambda_i + k_i)^2} .$$

In this case, the choice of k_i which minimizes the MSE of $\hat{\beta}_i$ is $k_i = (\sigma^2/\beta_i^2)$. Hoerl and Kennard (1970) suggest a feasible version of this optimal ridge estimator:

$$(2.9) \qquad \hat{\beta}_i^* = \left[\lambda_i b_i^2/(\lambda_i b_i^2 + s/\nu)\right] b_i \; .$$

Clearly, $\hat{\beta}_i^*$ no longer minimises MSE – some of its exact finite sample properties are derived by Dwivedi et al. (1980).

For applied work, a major limitation of $\hat{\beta}_i^*$ is that it provides only a point estimate of the regression coefficient. This problem is resolved in various ad hoc ways in practice (e.g. Judge et al. (1988, pp.878-882)). One possibility is to use the OLS standard error $(s/\nu\lambda_i)^{\frac{1}{2}}$ as a standard error for $\hat{\beta}_i^*$. This is legitimate asymptotically, but has no finite-sample justification and provides no information about the gains or losses in measured efficiency when ridge regression is used in preference to OLS. Another approach that is sometimes adopted involves using estimators of β_i and σ^2 to estimate k_i (as in (2.9)) and hence (2.7) and (2.8). Although "standard errors" for the $\hat{\beta}_i^*$ estimators can be generated in this way, their obvious limitation is that (2.7) and (2.8) are no longer the correct expressions for the estimator's bias and mean squared error once k_i is estimated.

Complete inference based on the feasible ridge estimator requires knowledge of its sampling distribution. Dwivedi et al. (1980) derive the first two moments of $\hat{\beta}_i^*$, but the expressions concerned are complex and unobservable. However, if unbiased estimators of these two moments can be derived, then we have a legitimate basis for constructing ridge regression standard

errors. This task is considered in the next section. Our objective is to provide exact finite sample measures, and hence resolve the difficulties currently facing an applied researcher using feasible generalised ridge regression.

3.   UNBIASED ESTIMATION OF THE BIAS AND MEAN SQUARED ERROR

From (2.9), we observe that the bias of $\hat{\beta}_i^*$ is

$$(3.1) \qquad B(\hat{\beta}_i^*) \;=\; E\left[(b_i - \beta_i) \;-\; \frac{sb_i}{\nu\lambda_i b_i^2 + s}\right]$$

$$\;=\; -\,E\left[\frac{sb_i}{\nu\lambda_i b_i^2 + s}\right] \;,$$

from which it follows that an unbiased estimator of the bias is

$$(3.2) \qquad \hat{B}(\hat{\beta}_i^*) \;=\; -\left(\frac{b_i}{\nu f_i + 1}\right)$$

where $f_i = \left(\dfrac{\lambda_i b_i^2}{s}\right)$. It may be noticed that unbiased estimation of the bias does not require normality of the disturbances.

Similarly, the mean squared error of $\hat{\beta}_i^*$ is

$$(3.3) \qquad M(\hat{\beta}_i^*) \;=\; E(b_i - \beta_i)^2 + E\left[\frac{sb_i}{\nu\lambda_i b_i^2 + s}\right]^2 - 2E\left[\frac{sb_i(b_i - \beta_i)}{\nu\lambda_i b_i^2 + s}\right]$$

$$\;=\; \frac{\sigma^2}{\lambda_i} + E\left[\frac{sb_i}{\nu\lambda_i b_i^2 + s}\right]^2 - 2E\left[\frac{sb_i(b_i - \beta_i)}{\nu\lambda_i b_i^2 + s}\right].$$

An unbiased estimator of the first term on the right hand side of (3.3) is

4

(3.4)
$$\left(\frac{s}{\nu\lambda_i} \right) ,$$

while an unbiased estimator for the second term is

(3.5)
$$\left(\frac{s}{\nu\lambda_i} \right) \left[\frac{\nu f_i}{(\nu f_i + 1)^2} \right].$$

For the third term, we observe that b_i has a normal distribution with mean β_i and variance (σ^2/λ_i) while (s/σ^2) has a χ^2-distribution with ν degrees of freedom, independent of b_i. Using this result and integrating by parts:

(3.6)
$$E\left[\frac{s b_i (b_i - \beta_i)}{\nu\lambda_i b_i^2 + s} \right] = \frac{\sigma^2}{\lambda_i} E\left[s \frac{\partial}{\partial b_i} \left(\frac{b_i}{\nu\lambda_i b_i^2 + s} \right) \right]$$

$$= -\frac{\sigma^2}{\lambda_i} E\left[\frac{s(\nu\lambda_i b_i^2 - s)}{(\nu\lambda_i b_i^2 + s)^2} \right].$$

Now suppose that $g(s, b_i)$ is absolutely continuous in s and is such that

(3.7)
$$E[s\, g(s, b_i)] = E\left[\frac{s b_i (b_i - \beta_i)}{\nu\lambda_i b_i^2 + s} \right].$$

Integrating the left side of (3.7) by parts and using (3.6), we get

$$\sigma^2 E\left[\nu\ g(s,b_i) + 2s\ \frac{\partial g(s,b_i)}{\partial s}\right] = E\left[\frac{sb_i(b_i-\beta_i)}{\nu\lambda_i b_i^2+s}\right]$$

$$= -\frac{\sigma^2}{\lambda_i} E\left[\frac{s(\nu\lambda_i b_i^2-s)}{(\nu\lambda_i b_i^2+s)^2}\right]$$

or

$$E\left[\nu\ g(s,b_i) + 2s\ \frac{\partial g(s,b_i)}{\partial s} + \frac{s(\nu\lambda_i b_i^2-s)}{\lambda_i(\nu\lambda_i b_i^2+s)^2}\right] = 0.$$

As (s,b_i) are jointly complete suficient statistics for (σ^2,β_i), this last equation suggests

$$(3.9) \qquad \nu\ g(s,b_i) + 2s\ \frac{\partial g(s,b_i)}{\partial s} + \frac{s(\nu\lambda_i b_i^2-s)}{\lambda_i(\nu\lambda_i b_i^2+s)^2} = 0.$$

This is a linear differential equation of first order and first degree, with solution

$$(3.10) \qquad g(s,b_i) = -\left(\frac{1}{2}\right) s^{-\frac{\nu}{2}} \int_0^s \frac{(\nu\lambda_i b_i^2-x)x^{\frac{\nu}{2}}}{\lambda_i(\nu\lambda_i b_i^2+x)^2}\ dx$$

$$= -\left(\frac{1}{2\lambda_i}\right) \int_0^1 \frac{(\nu f_i-x)x^{\frac{\nu}{2}}}{(\nu f_i+x)^2}\ dx,$$

which provides the functional form of $g(s,b_i)$.

Substituting (3.10) in (3.7) we observe that an unbiased estimator of the third term on the right hand side of (3.3) is

$$(3.11) \qquad - \left(\frac{s}{\lambda_i} \right) \int_0^1 \left[\frac{(\nu f_i - x) x^{\frac{\nu}{2}}}{(\nu f_i + x)^2} \right] dx.$$

Using (3.4), (3.5) and (3.11), an unbiased estimator of the MSE of $\hat{\beta}_i^*$ is

$$(3.12) \qquad \hat{M}(\hat{\beta}_i^*) = \left(\frac{s}{\nu \lambda_i} \right) \left[1 + \frac{\nu f_i}{(\nu f_i + 1)^2} + \nu\, J(\nu, f_i) \right]$$

where $\quad J(\nu, f_i) = \int_0^1 \left[\frac{(\nu f_i - x) x^{\frac{\nu}{2}}}{(\nu f_i + x)^2} \right] dx.$

The square root of $\hat{M}(\hat{\beta}_i^*)$ gives a meaningful standard error for $\hat{\beta}_i^*$. This expression is simple, especially when compared with its population counterpart (see Dwivedi et al. (1980, p.206)). The value of $J(\nu, f_i)$ is easily determined and $\hat{M}(\hat{\beta}_i^*)$ is expressed in terms of quantities readily obtained from conventional regression output.

Finally, (3.12) depends on the regressors through λ_i and is derived on the assumption that X'X has been diagonalised. However, considering the orthogonal transformations that would be used to achieve such a diagonalisation it is easily seen that $\hat{V}(b)$ in (2.4) depends on the data in a corresponding way; that "t-ratios" based on (2.9) and (3.12) are independent of this assumption; and that the comparative evaluations considered in the next section are perfectly general in the sense that they are not in fact limited to the case of a diagonalised design matrix.

4. NUMERICAL EVALUATION

The numerical evaluation of (3.12) for any choice of ν and f_i is straightforward - we have used the FORTRAN routines QROMB, TRAPZD and POLINT given by Press et al. (1986) to implement Romberg's method to determine $J(\nu, f_i)$. The relative <u>estimated</u> efficiency of b_i to $\hat{\beta}_i^*$ is $e = \hat{M}(\hat{\beta}_i^*)/\hat{V}(b_i)$, where $\hat{V}(b_i) = (s/\nu\lambda_i)$, and values of e are given in Table 1 for various choices of ν and $t_i = (\nu f_i)^{\frac{1}{2}}$. The latter quantity is the t-ratio for testing if $\beta_i = 0$, so this parameterisation facilitates practical prescriptions.

We see that $e < 1$ for combinations of small values of ν and t_i. Small values of t_i are characteristic of the application of OLS in the context of collinear data, and our results indicate that in such cases the OLS standard errors understate the precision of the estimates. If $t_i \le 0.5$ then $\hat{\beta}_i^*$ is estimated to be more efficient than b_i <u>for any</u> ν; and conversely if $t_i \ge 0.8$ then b_i is estimated to be more efficient than $\hat{\beta}_i^*$ <u>for any</u> ν.

From (2.9), $\hat{\beta}_i^* = b_i t_i^2/(1+t_i^2)$, so as $\nu \to \infty$ or $f_i \to \infty$, $\hat{\beta}^* \to b_i$ and $e \to 1$. This is supported by the results in Table 1 and also when e is tabulated in terms of ν and f_i values. The results in Table 1 also indicate that for the majority of combinations of ν and t_i, b_i is estimated to be efficient relative to $\hat{\beta}_i^*$, and the maximum such efficiency gain is 46%. Regardless of ν, this maximum gain arises for t_i between 1.2 and 1.6, and generally for $t_i = 1.4$ ($t_i^2 \cong 2$).

As noted already, in practice $\hat{V}(b_i)$ is often used to produce standard errors to use in association with $\hat{\beta}_i^*$. This can be justified asymptotically, but in finite samples this can now be improved upon through the use of $\hat{M}(\hat{\beta}_i^*)$. Comparing these two

8

approaches to measuring the precision of $\hat{\beta}_i^*$, our values for e show that generally the use of OLS standard errors in conjunction with ridge regression point estimates involves an overstatement of the precision of these estimates, relative to that indicated by the more appropriate measure derived in this paper. The figures in Table 1 may be viewed as correction factors to compensate for this overstatement - multiplying an OLS standard error by $e^{\frac{1}{2}}$ yields an appropriate standard error to use with the ridge estimator.

It is interesting to note that, in terms of <u>actual</u> MSE, $\hat{\beta}_i^*$ is efficient relative to b_i if $(\lambda_i \beta_i^2 / 2\sigma^2) < 1$. While it may be tempting to estimate the left hand side of this expression to see if the inequality holds empirically, it is clear that this strategy can be very misleading. For example, substituting the least squares estimators of β_i and σ^2 gives the condition $t_i^2 < 2$. Adjusting for the bias introduced by estimating in this way yields the condition $(\nu-2)t_i^2/\nu < 3$. That this is an unsatisfactory procedure can be seen by noting that these last two inequalities involve the estimation of a condition relating to the <u>true</u> relative efficiencies of the two estimators. In contrast, the information in Table 1 relates to the relative <u>estimated</u> efficiencies of $\hat{\beta}_i^*$ and b_i. Dealing with the inequality relating to the actual MSE's of $\hat{\beta}_i^*$ and b_i properly involves testing if $(\lambda_i \beta_i^2 / 2\sigma^2) < 1$ and choosing the estimator accordingly. This "preliminary-test" estimation strategy is discussed by Srivastava and Giles (1984).

Another way of considering the practical implications of our results is to compare t_i with $t_i^* = \hat{\beta}_i^*/(M(\hat{\beta}_i^*))^{\frac{1}{2}}$, the ridge-regression analogue of the usual t-ratio. The quantity $\gamma_i^* =$

(t_i^*/t_i) is the factor by which an OLS t-ratio should be scaled to convert it to its ridge regression counterpart. Of course, t_i^* is not t-distributed, but it is still a measure of some practical interest. Values of γ_i^* appear in Table 2. We see that, especially for small values of ν and/or t_i, a t-ratio based on OLS is misleading in the sense that it suggests substantially greater estimation precision than would be inferred from the value of t_i^*.

5.   CONCLUDING REMARKS

This paper solves the problem of presenting a meaningful measure of the precision associated with a point estimate obtained by feasible generalised ridge regression. The practices generally adopted by applied workers are unsatisfactory because either they involve the estimation of a quantity which does not actually measure the population precision; or they are based on expressions which are uninformative because they are valid only asymptotically. In contrast, we present a simple expression for an unbiased estimator of the ridge estimator's mean squared error which is exact in finite samples and whose square root provides a meaningful standard error to be associated with a ridge regression coefficient. The tabulated evaluations indicate the extent to which standard errors and "t-ratios" based on least squares results may differ from those based on the procedure proposed here. The calculations needed to compute the proposed standard error are trivial, and can easily be incorporated into ridge regression packages.

<u>October 1989</u>

Table 1: Values of $e = \hat{M}(\hat{\delta}_i^2)/\hat{V}(b_i)$

ν \ t_i	0.50	0.60	0.70	0.80	0.90	1.00	1.20	1.40	1.60	1.80	2.00	2.50	3.00	4.00	5.00
1	0.97	1.14	1.25	1.32	1.37	1.39	1.40	1.37	1.34	1.30	1.27	1.20	1.15	1.09	1.06
2	0.77	1.01	1.17	1.29	1.36	1.41	1.44	1.42	1.39	1.35	1.32	1.24	1.18	1.11	1.07
3	0.64	0.90	1.10	1.25	1.34	1.40	1.45	1.44	1.41	1.38	1.34	1.26	1.20	1.12	1.08
4	0.55	0.83	1.05	1.21	1.32	1.39	1.45	1.45	1.43	1.39	1.35	1.27	1.21	1.13	1.09
5	0.49	0.78	1.01	1.18	1.30	1.37	1.45	1.46	1.44	1.40	1.36	1.28	1.21	1.13	1.09
6	0.44	0.74	0.98	1.15	1.28	1.36	1.44	1.46	1.44	1.41	1.37	1.28	1.22	1.14	1.09
7	0.41	0.71	0.95	1.13	1.26	1.35	1.44	1.46	1.44	1.41	1.37	1.29	1.22	1.14	1.09
8	0.38	0.69	0.93	1.12	1.25	1.34	1.44	1.46	1.45	1.41	1.38	1.29	1.23	1.14	1.09
9	0.36	0.67	0.91	1.10	1.24	1.34	1.44	1.46	1.45	1.42	1.38	1.29	1.23	1.14	1.10
10	0.35	0.65	0.90	1.09	1.23	1.33	1.43	1.46	1.45	1.42	1.38	1.30	1.23	1.14	1.10
12	0.32	0.63	0.88	1.07	1.22	1.32	1.43	1.46	1.45	1.42	1.39	1.30	1.23	1.15	1.10
14	0.30	0.61	0.86	1.06	1.21	1.31	1.42	1.46	1.45	1.42	1.39	1.30	1.24	1.15	1.10
16	0.29	0.60	0.85	1.05	1.20	1.30	1.42	1.46	1.45	1.42	1.39	1.31	1.24	1.15	1.10
18	0.28	0.59	0.84	1.04	1.19	1.30	1.42	1.46	1.45	1.42	1.39	1.31	1.24	1.15	1.10
20	0.27	0.58	0.83	1.04	1.19	1.29	1.42	1.45	1.45	1.43	1.39	1.31	1.24	1.15	1.10
25	0.26	0.56	0.82	1.02	1.18	1.29	1.41	1.45	1.45	1.43	1.39	1.31	1.24	1.15	1.10
30	0.25	0.55	0.81	1.01	1.17	1.28	1.41	1.45	1.45	1.43	1.40	1.31	1.24	1.15	1.10
35	0.24	0.55	0.80	1.01	1.16	1.28	1.41	1.45	1.45	1.43	1.40	1.31	1.24	1.15	1.10

Table 2: Values of $\gamma_i^* = (t_i^*/t_i)$

ν \ t_i	0.50	0.60	0.70	0.80	0.90	1.00	1.20	1.40	1.60	1.80	2.00	2.50	3.00	4.00	5.00
1	0.20	0.28	0.37	0.45	0.52	0.59	0.70	0.78	0.83	0.87	0.90	0.94	0.97	0.98	0.99
2	0.18	0.27	0.36	0.44	0.52	0.59	0.71	0.79	0.85	0.89	0.92	0.96	0.98	0.99	1.00
3	0.16	0.25	0.35	0.44	0.52	0.59	0.71	0.80	0.86	0.90	0.93	0.97	0.98	0.99	1.00
4	0.15	0.24	0.34	0.43	0.51	0.59	0.71	0.80	0.86	0.90	0.93	0.97	0.99	1.00	1.00
5	0.14	0.23	0.33	0.42	0.51	0.59	0.71	0.80	0.86	0.90	0.93	0.97	0.99	1.00	1.00
6	0.13	0.23	0.32	0.42	0.51	0.58	0.71	0.80	0.86	0.91	0.94	0.98	0.99	1.00	1.00
7	0.13	0.22	0.32	0.42	0.50	0.58	0.71	0.80	0.86	0.91	0.94	0.98	0.99	1.00	1.00
8	0.12	0.22	0.32	0.41	0.50	0.58	0.71	0.80	0.86	0.91	0.94	0.98	1.00	1.00	1.00
9	0.12	0.22	0.31	0.41	0.50	0.58	0.71	0.80	0.86	0.91	0.94	0.98	1.00	1.00	1.00
10	0.12	0.21	0.31	0.41	0.50	0.58	0.71	0.80	0.87	0.91	0.94	0.98	1.00	1.00	1.00
12	0.11	0.21	0.31	0.40	0.49	0.57	0.71	0.80	0.87	0.91	0.94	0.98	1.00	1.00	1.00
14	0.11	0.21	0.31	0.40	0.49	0.57	0.70	0.80	0.87	0.91	0.94	0.98	1.00	1.00	1.00
16	0.11	0.20	0.30	0.40	0.49	0.57	0.70	0.80	0.87	0.91	0.94	0.98	1.00	1.00	1.00
18	0.11	0.20	0.30	0.40	0.49	0.57	0.70	0.80	0.87	0.91	0.94	0.99	1.00	1.00	1.00
20	0.10	0.20	0.30	0.40	0.49	0.57	0.70	0.80	0.87	0.91	0.94	0.99	1.00	1.00	1.00
25	0.10	0.20	0.30	0.39	0.49	0.57	0.70	0.80	0.87	0.91	0.94	0.99	1.00	1.00	1.00
30	0.10	0.20	0.30	0.39	0.48	0.57	0.70	0.80	0.87	0.91	0.94	0.99	1.00	1.00	1.00
35	0.10	0.20	0.29	0.39	0.48	0.57	0.70	0.80	0.87	0.91	0.95	0.99	1.00	1.00	1.00

REFERENCES

Carter, R.A.L., Srivastava, M.S., Srivastava, V.K. and Ullah, A. (19**). Unbiased Estimation of the MSE Matrix of Stein-Rule Estimators, Confidence Ellipsoids and Hypothesis Testing. Econometric Theory (forthcoming).

Dwivedi, T.D., Srivastava, V.K. and Hall, R.L. (1980). Finite Sample Properties of Ridge Estimators. Technometrics, 22, 205-121.

Hoerl, A.E. and Kennard, R.W. (1970). Ridge Regression: Biased Estimation for Non-Orthogonal Problems. Techometrics, 12, 55-67.

Judge, G.G., Hill R.C., Griffiths, W.E. and Lee T-C. (1988). Introduction to the Theory and Practice of Econometrics (Wiley, New York).

Press, W.H., Flannery, B.P., Teukolsky, S.A. and Vetterling, W.T. (1986). Numerical Recipies: The Art of Scientific Computing (Cambridge University Press, Cambridge).

Srivastava, V.K. and Giles, D.E.A. (1984). A Pre-Test General Ridge Regression Estimator: Exact Finite-Sample Properties. Australian Journal of Statistics, 26, 323-336.

LIST OF DISCUSSION PAPERS*