Ann. Inst. Statist. Math.
Vol. 41, No. 4, 717–724 (1989)

COMPARISONS AMONG SOME ESTIMATORS IN MISSPECIFIED LINEAR MODELS WITH MULTICOLLINEARITY

NITYANANDA SARKAR

Economic Research Unit, Indian Statistical Institute, 203, Barrackpore Trunk Road, Calcutta - 700 035, India

(Received January 28, 1988; revised March 7, 1989)

Abstract. In this paper we deal with comparisons among several estimators available in situations of multicollinearity (e.g., the $r-k$ class estimator proposed by Baye and Parker, the ordinary ridge regression (ORR) estimator, the principal components regression (PCR) estimator and also the ordinary least squares (OLS) estimator) for a misspecified linear model where misspecification is due to omission of some relevant explanatory variables. These comparisons are made in terms of the mean square error (mse) of the estimators of regression coefficients as well as of the predictor of the conditional mean of the dependent variable. It is found that under the same conditions as in the true model, the superiority of the $r-k$ class estimator over the ORR, PCR and OLS estimators and those of the ORR and PCR estimators over the OLS estimator remain unchanged in the misspecified model. Only in the case of comparison between the ORR and PCR estimators, no definite conclusion regarding the mse dominance of one over the other in the misspecified model can be drawn.

Key words and phrases: Misspecification, multicollinearity, ordinary ridge regression estimator, principal components regression estimator, $r-k$ class estimator.

1. Introduction

Econometricians, while dealing with the problem of choice of a proper model, face, *inter alia*, the problems of multicollinearity and misspecification of the model. The statistical consequences of multicollinearity in a linear regression model have been studied in great detail (see Judge *et al.* ((1980), Chapter 12) for detailed discussion of this problem). It is well known that in situations of multicollinearity, it becomes difficult to obtain precise estimates of the separate effects of the variables involved in the

regression model; the method of least squares produces large sampling variances of the estimated regression coefficients, which in turn gives rise to the possibility that otherwise significant coefficients may be dropped from the analysis improperly.

In order to circumvent these problems, many alternative estimators have been suggested by researchers. These often yield point estimates (of parameters) superior to those provided by the traditional procedures under a variety of loss functions. These procedures include, for example, the ordinary ridge regression (see for instance Hoerl and Kennard (1970), Vinod (1978) and Vinod and Ullah (1981) for relevant discussions) and the principal components regression (see Farebrother (1972) and Fomby *et al.* (1978) etc. in this context). Although detailed sampling properties of these estimators are mostly unknown, these continue to remain ad-hoc solutions to the problem of multicollinearity. In fact, there have been many works concerning the efficiencies of these estimators vis-a-vis the ordinary least squares estimator. Further, Baye and Parker (1984) proposed the $r-k$ class estimator which includes the ordinary least squares (OLS) estimator, the ordinary ridge regression (ORR) estimator and the principal components regression (PCR) estimator as special cases, and compared its performance to the PCR estimator by the mean square error criterion. Nomura and Ohkubo (1985) extended this further, and compared the performance of the $r-k$ class estimator with the OLS and ORR estimators by the criteria of mean square error (mse) of the regression coefficients as well as those of the predictor of $E(y/X)$. The principal results concerning these estimators, viz. the $r-k$ class estimator, the ORR and PCR estimators, are that these estimators are generally biased, and that under specific conditions on the parameters involved, there exist specific values for the constant k (or a range of values for k) for which the $r-k$ class estimator has smaller mse value than the ORR as well as the PCR estimators. There are similar results concerning the mse dominance of the ORR and PCR estimators over the OLS estimator as well. But all these comparisons of superiority of one of the estimators over the others have been carried out for what could be termed "true models".

Omission of some relevant explanatory variables in regression models is quite common in applied works. The consequences of such omissions on standard inferential problems have been widely studied. But the consequences of such omissions (henceforth to be referred to as misspecification and linear models with such omissions as misspecified models) for linear regression models with multicollinearity have not yet been examined. In other words, it becomes an interesting investigation to find out if these alternative estimators would still remain superior in misspecified models in situations when these are so in true models. This paper attempts to examine precisely this point. The comparisons are made among these estimators themselves as well as with the OLS estimator. Indeed, this study

is expected to shed light on whether in actual analysis we really have to differentiate between working with a true model and a misspecified model, insofar as comparison by means of the mse criterion is concerned. The paper has been arranged as follows: In Section 2 we describe the model and then define the $r-k$ class estimator. The comparisons among the estimators (including the OLS) of the regression coefficients and those of the predictor of $E(y/X)$ by means of the mse criterion are done in Sections 3 and 4, respectively. The paper ends with conclusions in Section 5.

2. The model

We consider the true model as given by

$$y = X\beta + Z\gamma + \varepsilon , \tag{2.1}$$

where y is an $(n \times 1)$ vector of observations on the dependent variable, X is an $(n \times p)$ matrix of non-stochastic variables of rank p, Z is another $(n \times q)$ matrix of non-stochastic variables of rank q $(p + q < n)$, β and γ are the corresponding $(p \times 1)$ and $(q \times 1)$ vectors of parameters associated with X and Z, respectively, and ε is an $(n \times 1)$ vector of error terms with mean zero and variance-covariance matrix $\sigma^2 I_n$. We now assume that the misspecified model is one where the set of q regressors have been omitted from the true model in (2.1) and is given by

$$y = X\beta + u , \tag{2.2}$$

where $u = Z\gamma + \varepsilon$. The error term u of the misspecified model is distributed with mean vector $Z\gamma$ and variance-covariance matrix $\sigma^2 I_n$.

Let $T = (t_1, t_2,\ldots, t_p)$ be an orthogonal matrix with $T'X'XT = \Lambda$ being diagonal, and $T_r = (t_1, t_2,\ldots, t_r)$ where $r \leq p$. Obviously, then, $T_r'X'XT_r = \Lambda_r =$ diagonal $(\lambda_1, \lambda_2,\ldots, \lambda_r)$. Also, $T_{p-r}'X'XT_{p-r} = \Lambda_{p-r} =$ diagonal $(\lambda_{r+1}, \lambda_{r+2},\ldots, \lambda_p)$ where $T_{p-r} = (t_{r+1}, t_{r+2},\ldots, t_p)$.

Baye and Parker (1984) proposed a general estimator for β, and called it the $r-k$ class estimator for β. This estimator $b_r^*(k)$ for β, in the context of the misspecified model in (2.2), is given as

$$b_r^*(k) = T_r(T_r'X'XT_r + kI_r)^{-1}T_r'X'y, \quad k \geq 0 . \tag{2.3}$$

This estimator is a general estimator which includes the ordinary least squares (OLS) estimator, the ordinary ridge regression (ORR) estimator and the principal components regression (PCR) estimator as special cases. In fact, for the misspecified model in (2.2), these special cases of the $r-k$ class estimator are as follows:

(i) $b_p^*(0) = b^* = (X'X)^{-1}X'y$ is the OLS estimator.

(ii) $b_p^*(k) = b^*(k) = (X'X + kI_p)^{-1}X'y$ is the ORR estimator.
(iii) $b_r^*(0) = b_r^* = T_r(T_r'X'XT_r)^{-1}T_r'X'y$ is the PCR estimator.

3. The mean square error comparison of the estimators of regression coefficients

As stated in the previous section, the $r-k$ class estimator of β in the misspecified model is given as

$$(3.1) \qquad b_r^*(k) = T_r(T_r'X'XT_r + kI_r)^{-1}T_r'X'y$$
$$= T_r(\Lambda_r + kI_r)^{-1}T_r'X'(X\beta + Z\gamma + \varepsilon) .$$

Let $S_r(k)^{-1}$ be the inverse of the matrix $S_r(k) = (\Lambda_r + kI_r)$. Then

$$(3.2) \qquad E(b_r^*(k)) = T_rS_r(k)^{-1}T_r'X'X\beta + T_rS_r(k)^{-1}T_r'\delta ,$$

where $\delta = X'Z\gamma$.

The mean square error (mse) of $b_r^*(k)$, denoted by MSE $(b_r^*(k))$, is obtained as

$$(3.3) \quad \text{MSE}\,(b_r^*(k)) = E[(b_r^*(k) - \beta)'(b_r^*(k) - \beta)]$$
$$= \beta'(T_rS_r(k)^{-1}T_r'X'X - I)'(T_rS_r(k)^{-1}T_r'X'X - I)\beta$$
$$+ \sigma^2 \text{ tr } (S_r(k)^{-1}\Lambda_rS_r(k)^{-1}) + \delta'T_rS_r(k)^{-2}T_r'\delta$$
$$= \text{MSE}\,(b_r(k)) + \delta'T_rS_r(k)^{-2}T_r'\delta ,$$

where MSE $(b_r(k)) = \{\beta'(T_rS_r(k)^{-1}T_r'X'X - I)'(T_rS_r(k)^{-1}T_r'X'X - I)\beta + \sigma^2 \cdot \text{tr } (S_r(k)^{-1}\Lambda_rS_r(k)^{-1})\}$ is the mse of the $r-k$ class estimator of β, if the model in (2.2) was, in fact, the true model. The mse expressions for the OLS, ORR and PCR estimators can thus be obtained as special cases of MSE $(b_r^*(k))$ from (3.3) by appropriate choice of the parameters r and k.

It should be noted at this stage that in the mse expression above, the first term refers to the mse of the estimator (of β) considered if the model in (2.2) was, in fact, not misspecified, and the second term, i.e., the quadratic form in δ, appears due to misspecification of the model.

It is known (see Baye and Parker (1984) and Nomura and Ohkubo (1985) for instance) that under certain conditions, the $r-k$ class estimator is superior to the OLS, ORR and PCR estimators for the true model over some specified range of values for k. For example, Nomura and Ohkubo ((1985), p. 2493) have stated in their Corollary 2 that if $\sum_{i\in\bar{N}_r} (\alpha_i^2 - \sigma^2/\lambda_i) \leq 0$, then MSE $(b_r(k)) <$ MSE (b) for $0 < k \leq \left(2\sigma^2 \Big/ \sum_{i\in N_r} \alpha_i^2 \right)$, where $N_r = \{1, 2, \ldots, r\}$, $\bar{N}_r = \{r+1, r+2, \ldots, p\}$ and α_i is the i-th element of the vector $\alpha = T'\beta$. We

now examine if the superiority of the $r-k$ class estimator over the other estimators would still be maintained under the same conditions and the same specified ranges of values for k even when the model is misspecified. For the purpose of this comparison (by the mse criterion) between any two estimators of β in the misspecified model, it is then obviously sufficient to compare the two relevant quadratic forms in δ only.

3.1 *The $r-k$ class estimator vs. the OLS estimator*

We first consider comparison between the $r-k$ class estimator and the OLS estimator of β for the model given in (2.2) by means of the mse criterion. To that end, we first obtain the expression for MSE (b^*), the mse of the OLS estimator b^*, by substituting $k=0$ and $r=p$ in (3.3). Then, by subtracting this from (3.3), we get

$$\text{(3.4)}\quad \text{MSE}\,(b_r^*(k)) - \text{MSE}\,(b^*) = \text{MSE}\,(b_r(k)) - \text{MSE}\,(b) + \delta' S^{-1}(ST_rS_r(k)^{-2}T_r'S - I)S^{-1}\delta\,,$$

where MSE $(b) = \sigma^2$ tr (S^{-1}) and $S^{-1} = TS_p(0)^{-1}T' = (X'X)^{-1}$.

In order to study the definiteness of the quadratic form in δ in the above expression in (3.4), it may be noted that all the characteristic roots of the matrix $ST_rS_r(k)^{-2}T_r'S$ are same as that of $T_r'ST_rS_r(k)^{-2}T_r'ST_r = \Lambda_rS_r(k)^{-2}\cdot\Lambda_r$, since $T_r'T_r = I_r$. But,

$$\begin{aligned}\Lambda_rS_r(k)^{-2}\Lambda_r &= [(I+k\Lambda_r^{-1})(I+k\Lambda_r^{-1})]^{-1}\\ &= [I+B_r]^{-1}\,,\end{aligned}$$

where $B_r = 2k\Lambda_r^{-1} + k^2\Lambda_r^{-2}$ is a non-negative definite (n.n.d.) matrix for all $k \geq 0$. Thus, all the roots of $(I+B_r)$ are ≥ 1 and hence the roots of the matrix $[I+B_r]^{-1}$ are all ≤ 1. One can then conclude (cf. Rao (1974), p. 70) that $(ST_rS_r(k)^{-2}T_r'S - I)$ is a non-positive definite (n.p.d.) matrix. So, (MSE $(b_r^*(k))$ − MSE (b^*)) is negative, if (MSE $(b_r(k))$ − MSE (b)) is so. We thus have the following theorem.

THEOREM 3.1. *Suppose* $\sum_{i\in\bar{N}_r}(\alpha_i^2 - \sigma^2/\lambda_i) \leq 0$ *and* $0 < k \leq 2\sigma^2 \Big/ \sum_{i\in N_r}\alpha_i^2$. *Then,* MSE $(b_r^*(k)) <$ MSE (b^*).

It is therefore established that if the $r-k$ class estimator is mse superior to the OLS estimator in the true model, then it remains so in the misspecified model as well. Obviously, if the $r-k$ class estimator is inferior to the OLS estimator by the mse criterion in the true model, then no conclusion can be drawn regarding superiority of one over the other in the misspecified model.

3.2 *The $r-k$ class estimator vs. the ORR estimator*

By substituting $r = p$ in (3.3), we find the expression for MSE $(b^*(k))$, the mse of the ORR estimator $b^*(k)$, and hence, we have

$$\text{(3.5)} \quad \text{MSE}\,(b_r^*(k)) - \text{MSE}\,(b^*(k)) = \text{MSE}\,(b_r(k)) - \text{MSE}\,(b(k)) + \delta'[T_r S_r(k)^{-2} T_r' - S(k)^{-2}]\delta\,,$$

where $\text{MSE}\,(b(k)) = \{k^2\beta' S(k)^{-2}\beta + \sigma^2 \text{ tr } (S(k)^{-2}X'X)\}$ and $S(k)^{-1} = TS_p(k)^{-1} \cdot T' = (X'X + kI_p)^{-1}$. As in the previous case, we find from Nomura and Ohkubo (1985) that under certain conditions, the $r-k$ class estimator is superior to the ORR estimator in some range of values for k in the true model. To see if this superiority of the $r-k$ class estimator is still retained in the misspecified model, it is enough to check if the quadratic form in δ in (3.5) is negative. Now,

$$\begin{aligned}
&\delta'[T_r S_r(k)^{-2} T_r' - S(k)^{-2}]\delta \\
&\quad = \delta'[T_r S_r(k)^{-2} T_r' - T(\Lambda + kI_p)^{-2} T']\delta \\
&\quad = -\,\delta' T_{p-r}(\Lambda_{p-r} + kI_{p-r})^{-2} T_{p-r}'\delta\,,
\end{aligned}$$

and this is always n.p.d. for all $k \geq 0$. We thus have the following theorem.

THEOREM 3.2. *If the $r-k$ class estimator is superior to the ORR estimator in the true model, then the same holds in the misspecified model also. If, however, the $r-k$ class estimator is not superior to the ORR estimator in the true model, no definite conclusion can then be drawn.*

3.3 *The $r-k$ class estimator vs. the PCR estimator*

As in the previous two cases, we can find the expression for MSE (b_r^*), the mse of the PCR estimator b_r^*, as a special case of MSE $(b_r^*(k))$ in (3.3), where $k = 0$. Now, proceeding as before we can state the following theorem for this case.

THEOREM 3.3. *If the $r-k$ class estimator is mse dominant over the PCR estimator in the true model, then there always exists a positive value of k for which the $r-k$ class estimator has a smaller mse value than that for the PCR estimator in the misspecified model as well.*

Finally, we compare the performance of any two of the OLS, ORR and PCR estimators. To this end, we look at the differences between the two relevant mse expressions and then simplify them to obtain the following theorem.

THEOREM 3.4. *If the mse of the OLS estimator is greater than each of the mse's of the ORR and PCR estimators in the true model, then the same will be true in the misspecified model also. However, no such difinite conclusion can be drawn (in the misspecified model) in comparing mse superiority between the ORR and PCR estimators, even when such a superiority between the two estimators exists in the true model.*

4. The mean square error comparison of predictor

In this section we compare the performance of the $r-k$ class estimator and the three other estimators, viz., the OLS, ORR and PCR estimators by means of the criterion of mse of the predictor of $E(y/X)$, defined as

$$\text{MSE}\,(y_r^*(k)) = E[(b_r^*(k)-\beta)'X'X(b_r^*(k)-\beta)]\,.$$

Since the results as well as the proofs are similar to those in the last section, we omit the derivations and state these results in the following theorem.

THEOREM 4.1. *The $r-k$ class estimator dominates the OLS, ORR and PCR estimators by the criterion of mse of the predictor of $E(y/X)$ in the misspecified model, if the dominance of the $r-k$ class estimator over the other estimators holds for the true model.*

Insofar as comparisons among the OLS, ORR and PCR estimators by this criterion are concerned, it can easily be shown that the results are exactly the same as stated in Theorem 3.4, and hence these are not explicitly stated for this case.

5. Conclusions

An attempt to compare the performance of several well-known estimators available for regression models with multicollinearity in situations of misspecification of the model, has been made in this paper. The criteria used for the purpose of these comparisons are the mean square error of the regression coefficients as well as of the predictor of $E(y/X)$. It has been found that misspecification of the model does not alter the relative performances of these estimators so long as the same holds for the true model; the only exception to this is the comparison between the ordinary ridge regression estimator and the principal components regression estimator. Nothing definite can be said on the superiority of one of these two estimators over the other in the misspecified model even if this is possible for the true model.

Acknowledgement

The author is thankful to the referee for suggestions which have improved the presentation in the paper.

REFERENCES

Baye, M. R. and Parker D. F. (1984). Combining ridge and principal component regression: A money demand illustration, *Comm. Statist. A—Theory Methods*, **13**, 197–205.

Farebrother, R. W. (1972). Principal component estimators and minimum mean square error criteria in regression analysis, *Rev. Econom. Statist.*, **54**, 322–336.

Fomby, T. B., Hill, R. C. and Johnson, S. R. (1978). An optimality property of principal components regression, *J. Amer. Statist. Assoc.*, **73**, 191–193.

Hoerl, A. E. and Kennard, R. W. (1970). Ridge regression: Biased estimation of non-orthogonal problems, *Technometrics*, **12**, 55–67.

Judge, G. G., Griffiths, W. E., Hill, R. C. and Lee, T. C. (1980). *The Theory and Practice of Econometrics*, Wiley, New York.

Nomura, M. and Ohkubo, T. (1985). A note on combining ridge and principal component regression, *Comm. Statist. A—Theory Methods*, **14**, 2489–2493.

Rao, C. R. (1974). *Linear Statistical Inference and Its Applications*, Wiley Eastern Private Limited, New Delhi.

Vinod, H. D. (1978). A survey of ridge regression and related techniques for improvements over ordinary least squares, *Rev. Econom. Statist.*, **60**, 121–131.

Vinod, H. D. and Ullah, A. (1981). *Recent Advances in Regression Methods*, Marcel Dekker, New York.