JOURNAL OF COMBINATORICS, INFORMATION & SYSTEM SCIENCES
Vol. 29, Nos. 1–4, 229–238. (2002)

PREDICTION IN RESTRICTED REGRESSION MODELS

SHALABH

Department of Mathematics
Indian Institute of Technology Kanpur, Kanpur-208 016, India
E-mail: shalab@iitk.ac.in; shalabh1@yahoo.com

and

RAM CHANDRA

C.S. Azad University of Agriculture and Technology
Kanpur-208 002, India

This paper has considered the problem of predicting both the actual and average values of study variable in a linear regression model subject to a set of exact linear restrictions on regression coefficients. Three types of predictions arising from restricted regression and Stein-rule methods are presented for the values of study variable within the sample and outside the sample, and their performance properties are analyzed. (JEL: C29)

1. INTRODUCTION

When the coefficients in a linear regression model are subject to a set of exact linear restrictions, it is well documented that the restricted regression estimator of coefficient vector not only obeys the given prior restrictions but also possesses the property of minimum variance in the class of linear and unbiased estimators. Such an estimator, however, does not perform well when the aim is to predict the average values of study variable within the sample, see, e.g., Mittelhammer and Conway (1984) and Tracy and Srivastava (1995). These authors have employed the philosophy of Stein-rule estimation and have presented biased and nonlinear estimators possessing superior predictive performance than the linear and unbiased restricted regression estimator.

As Tracy and Srivastava (1995) have confined their attention to the prediction of average values of study variable within the sample, a natural question arises about their performance when the aim is to predict the actual values of study variable within the sample. Further, we may be interested to know aboutu their performance when the aim is to predict the values outside the sample, for instance, for the purpose of forecasting and preparing policy prescriptions. This article is an attempt to answer these questions.

Generally, predictions of study variable in linear models are obtained either for actual values or for average-values but not for both simultaneously. Situations may arise in practice where it is desirable to predict both the actual and average values of the study variable at the same time; see, e.g., Zellner (1994) and Shalabh (1995) for some illustrative examples. For this purpose, Shalabh (1995) has presented an interesting framework that possesses sufficient flexibility and permits assignment of possibly unequal weight to predictions for actual and average values of study variable. The present investigations have been carried out under such a framework.

The organization of this paper is as follows. Section 2 describes the linear regression model along with a set of exact linear restrictions binding the coefficients. The restricted regression estimator and two families of estimators emerging from Stein-rule philosophy are presented. Based on these, vectors of predictions for the values of study variable are formulated. Section 3 deals with the performance properties of predictions within the sample while Section 4 reports similar investigations for predictions outside the sample. Section 5 places some concluding remarks. Finally, the derivation of main results is outlined in the Appendix.

2. Restricted Regression Model and The Predictors

Consider the following linear regression model:

$$Y = X\beta + \sigma u \tag{1}$$

where Y is a $n \times 1$ vector of n observations on the study variable, X is an $n \times p$ full column rank matrix of n observations on p explanatory variables, β is a $p \times 1$ vector of regression coefficients, σ is an unknown scalar and u is an $n \times 1$ vector of disturbances assumed to be identically and independently distributed, each following a normal probability law with mean zero and variance unity.

Further, the regression coefficients are subject to the following set of exact linear restrictions:

$$r = R\beta \tag{2}$$

where r is a $j \times 1$ vector and R is a $j \times p$ full row rank matrix with known elements.

The restricted regression estimator of β is given by

$$\hat{\beta}_r = b - (X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}(Rb - r) \tag{3}$$

where $b = (X'X)^{-1}X'Y$ is the unrestricted estimator in the sense that it does not utilize the restrictions (2).

The estimator (3) obeys the restrictions (2) and is the best linear unbiased estimator of β.

Employing the philosophy of Stein-rule estimation, Mittelhammer and Conway (1984) have presented a class of nonlinear and biased estimators. An equivalent but more useful form of it has been obtained by Tracy and Srivastava (1995).

$$\hat{\beta}_{MC} = \hat{\beta}_r - a_{MC} \frac{(Y - Xb)'(Y - Xb)}{b'Gb} (X'X)^{-1} Gb \tag{4}$$

where $G = (X'X) - R'[R(X'X)^{-1}R']^{-1}R$ and a_{MC} is any positive and non-stochastic scalar characterizing the estimator.

Tracy and Srivastava (1995) have presented another class of nonlinear and biased estimators for β:

$$\hat{\beta}_{TS} = \hat{\beta}_r - a_{TS} \frac{(Y - X\hat{\beta}_r)'(Y - X\hat{\beta}_r)}{\hat{\beta}_r' G \hat{\beta}_r} (X'X)^{-1} G \hat{\beta}_r \tag{5}$$

where a_{TS} is any positive and non-stochastic scalar characterizing the estimator.

It is easy to see that the three estimators (3), (4) and (5) satisfy the restrictions (2). Using these, we can formulate the following three vectors of predictions for the values of study variable within the sample:

$$p_r = X\hat{\beta}_r, \quad p_{MC} = X\hat{\beta}_{MC}, \quad p_{TS} = X\hat{\beta}_{TS} \tag{6}$$

Similarly, for the prediction of values outside the sample, let us assume that we are given a set of n_f values of the explanatory variables and

$$Y_f = X_f\beta + \sigma u_f \tag{7}$$

where Y_f is an $n_f \times 1$ vector of n_f unobserved values of study variable, X_f is $n_f \times p$ a matrix of n_f pre-specified values of explanatory variables and u_f is $n_f \times 1$ vector of disturbances which have the same distributional properties as those of u in (1).

Thus the predictions are

$$p_{fr} = X_f\hat{\beta}_r, \quad p_{fMC} = X_f\hat{\beta}_{MC}, \quad p_{fTS} = X_f\hat{\beta}_{TS} \tag{8}$$

which can be used for predicting the actual and average values of the study variable outside the sample.

3. Performance Properties of Predictions Within the Sample

When the aim is to predict the average values ($X\beta$) of the study variable within the sample, Tracy and Srivastava (1995) have demonstrated that the predictions p_{MC} and p_{TS} are superior to the predictions p_r with respect to the criterion of total mean squared error when

$$0 < a_{MC} < 2\left(\frac{p - j - 2}{n - p + 2}\right) \tag{9}$$

$$0 < a_{TS} < 2\left(\frac{p-j-2}{n-p+j+2}\right) \tag{10}$$

provided that $(p-j)$ exceeds 2. Similarly, if we compare p_{MC} and p_{TS} assuming $a_{MC} = a_{TS} = a$, it is seen that p_{TS} is superior to p_{MC} when

$$0 < a < \left(\frac{p-j-2}{n-p+\frac{j}{2}+1}\right) \tag{11}$$

provided that $(p-j)$ exceeds 2.

Now a natural question arises related to performance of predictions when they are used for actual values of study variable rather than for average values. Also one may sometimes use these for predicting the actual and average values together; see, e.g., Zellner (1994) and Shalabh (1995) for few examples. We therefore define the following target function:

$$T = \lambda Y + (1-\lambda)E(Y) \tag{12}$$

where λ is a scalar between 0 and 1, the choice of which depends upon the weight to be given to the predictions of actual values in comparison to average values; see Shalabh (1995).

It is easy to see that p_r is weakly unbiased for T in the sense that

$$E(p_r - T) = 0 \tag{13}$$

while p_{MC} and p_{TS} are not.

Next, we observe that the total mean squared error of p_r is given by

$$\begin{aligned} M(p_r) &= E(p_r - T)'(p_r - T) \\ &= \sigma^2[(1-2\lambda)(p-j) + \lambda^2 n] \end{aligned} \tag{14}$$

For p_{MC} and p_{TS}, we present small disturbance asymptotic approximations.

THEOREM 1. *When disturbances are small, the differences in total mean squared errors up to order $O(\sigma^4)$ are given by*

$$\begin{aligned} D_{MC} &= M(p_r) - M(p_{MC}) \\ &= \sigma^4 \frac{a_{MC}(n-p)}{\beta' G\beta}[2(1-\lambda)(p-j-2) - (n-p+2)a_{MC}] \end{aligned} \tag{15}$$

$$\begin{aligned} D_{TS} &= M(p_r) - M(p_{TS}) \\ &= \sigma^4 \frac{a_{TS}(n-p+j)}{\beta' G\beta}[2(1-\lambda)(p-j-2) - (n-p+j+2)a_{TS}] \end{aligned} \tag{16}$$

These results are derived in the Appendix.

It is obvious from the above results that both p_{MC} and p_{TS} fail to beat p_r so long as they are used for predicting the actual values of the study variable within the sample ($\lambda = 1$). Such is, however, not the case when the aim is to predict either the average values ($\lambda = 0$) or both the actual and average values together ($0 < \lambda < 1$). In all such cases, p_{MC} is superior to p_r when

$$0 < a_{MC} < 2(1-\lambda)\left(\frac{p-j-2}{n-p+2}\right) \tag{17}$$

provided that $(p-j)$ exceeds 2. Similarly, p_{TS} is superior to p_r when

$$0 < a_{TS} < 2(1-\lambda)\left(\frac{p-j-2}{n-p+j+2}\right) \tag{18}$$

provided that $(p-j)$ exceeds 2.

If we set $\lambda = 0$ in (17) and (18), we get the conditions derived by Tracy and Srivastava (1995) on the basis of exact expressions for the total mean squared error. Further, it is interesting to note that the ranges of characterizing scalars for p_r to be dominated by p_{MC} and p_{ST} are decreasing functions of λ. In other words, ranges have a shrinking tendency when we increase the weight assigned to prediction of actual values in relation to the prediction of average values.

If we assume $a_{MC} = a_{TS} = a$ (say) following Tracy and Srivastava (1995), it is seen from (15) and (16) that

$$M(p_{MC}) - M(p_{TS}) = \frac{2\sigma^4 aj}{\beta' G\beta}\left[(1-\lambda)(p-j-2) - \left(n-p+\frac{j}{2}+1\right)a\right] \tag{19}$$

which is positive when

$$a < (1-\lambda)\left(\frac{p-j-2}{n-p+\frac{j}{2}+1}\right) \tag{20}$$

provided that $(p-j)$ is greater than 2. Thus p_{TS} yields better predictions in comparison to p_{MC} so long as the condition (20) is satisfied, ignoring the uninteresting case of $\lambda = 1$ as then both p_{MC} and p_{TS} are inferior to p_r. When the inequality (20) holds with a reversed sign, the opposite is true, i.e., p_{MC} is better than p_{TS}.

If the same choice a of the characterizing scalar is taken for both p_{MC} and

p_{TS} and $0 \le \lambda < 1$, it is thus observed that p_{TS} is superior to both p_r and p_{MC} so long as the condition (20) is satisfied. If the choice of the characterizing scalar violates the condition (20) but its value is smaller than the upper bound in (17), p_{MC} is better than p_r and p_{TS}.

4. Performance Properties of Predictions outside the Sample

When the aim is to predict the values of a study variable outside the sample, we may take the target function, in the spirit of (12), as follows

$$T_f = \lambda Y_f + (1-\lambda) E(Y_f) \tag{21}$$

where Y_f is specified by (7).

Let us consider the three vectors p_{fs}, p_{fMC} and p_{fTS} of predictions defined in (8).

It is easy to see that

$$E(p_{fr} - T_f) = 0 \tag{22}$$

so that p_{fr} is weakly unbiased for T_f. Such is not the case with p_{fMC} and p_{fTS} which are biased. Further, the total mean squared error of p_{fr} is

$$\begin{aligned} M(p_{fr}) &= E(p_{fr} - T_f)'(p_{fr} - T_f) \\ &= \sigma^2(\lambda^2 n + g) \end{aligned} \tag{23}$$

where
$$g = tr(X'X)^{-1}G(X'X)^{-1}X_f'X_f \tag{24}$$

THEOREM 2. *When disturbances are small, the differences in the total mean squared errors up to order* $O(\sigma^4)$ *are given by*

$$\begin{aligned} D_{fMC} &= M(p_{fr}) - M(p_{fMC}) \\ &= \sigma^4 \frac{(n-p)a_{MC}}{\beta' G \beta} [2(g-2k) - a_{MC}k(n-p+2)] \end{aligned} \tag{25}$$

$$\begin{aligned} D_{fTS} &= M(p_{fr}) - M(p_{fTS}) \\ &= \sigma^4 \frac{(n-p+j)a_{TS}}{\beta' G \beta} [2(g-2k) - a_{TS}k(n-p+j+2)] \end{aligned} \tag{26}$$

where
$$k = \frac{\beta' G(X'X)^{-1}X_f'X_f(X'X)^{-1}G\beta}{\beta' G \beta} \tag{27}$$

These results are derived in the Appendix.

The above results reveal some interesting observations. Firstly, both the differences (25) and (26) do not involve λ. This implies that the efficiency gain or loss, at least to the order of our approximation, remains same whether the predictions are used for actual values or average values or a weighted combination of them. Secondly, p_{fMC} is preferable over p_{fr} when

$$0 < a_{MC} < 2\left(\frac{\frac{q}{k} - 2}{n - p + 2}\right) \tag{28}$$

while p_{fTS} is preferable over p_{fr} when

$$0 < a_{TS} < 2\left(\frac{\frac{q}{k} - 2}{n - p + j + 2}\right) \tag{29}$$

provided that the ratio (g/k) exceeds 2.

Thirdly, comparing p_{MC} and p_{fTS} for the special case $a_{MC} = a_{TS} = a$ (say), it follows from (25) and (26) that p_{fTS} is preferable over p_{fMC} when

$$a < 2\left(\frac{\frac{q}{k} - 2}{n - p + \frac{j}{2} + 1}\right) \tag{30}$$

provided that (g/K) is greater than 2. Just the reverse is true, i.e., p_{fMC} is better than p_{fTS} when the inequality (30) holds with a reversed sign.

As the quantity k involves unknown regression coefficients, the conditions (28)–(30) for preference of one over the other hardly serve any useful purpose in any practical situation. To overcome this limitation, we observe that

$$C_{\min} \le k \le C_{\max} \tag{31}$$

where $C_{\min}$ and $C_{\max}$ denote the minimum and maximum characteristic roots of the matrix $((X'X)^{-1}G(X'X)^{-1}X_f'X_f$; see, e.g., Rao (1973).

Thus the conditions (28) and (29) will be satisfied as long as

$$0 < a_{MC} < 2\left(\frac{gC_{\max}^{-1} - 2}{n - p + 2}\right) \tag{32}$$

$$0 < a_{TS} < 2\left(\frac{gC_{\max}^{-1} - 2}{n - p + j + 2}\right) \tag{33}$$

provided that $gC_{\max}^{-1}$ is larger than 2.

Similarly, from (30), p_{fTS} is better than p_{fMC} at least as long as

$$a<2\left(\frac{gC_{\max}^{-1}-2}{n-p+\frac{j}{2}+1}\right) \tag{34}$$

provided that $gC_{\max}^{-1}$ is larger than 2.

On the other hand, p_{fMC} is better than p_{fTS} at least so long as

$$a>2\left(\frac{gC_{\min}^{-1}-2}{n-p+\frac{j}{2}+1}\right) \tag{35}$$

provided that $gC_{\min}^{-1}$ exceeds 2 which surely holds true if $gC_{\max}^{-1}$ is greater than 2.

5. Concluding Remarks

We have considered the problem of prediction of actual and average values of study variable in a linear regression model subject to a set of exact linear restrictions. The predictions arising from restricted regression estimation and Stein-rule estimation procedures are then considered and three varieties of predictions are described. Analyzing their performance properties, it is observed that restricted regression method yields unbiased predictions while Stein-rule method does not. We have then compared the predictions with respect to the criterion of total mean squared error.

When the aim is to predict the actual values of study variable within the sample, our investigations have revealed that predictions based on restricted regression method are invariably superior to the predictions based on Stein-rule method. This is however not true when the aim is to predict the average values alone or both the actual and average values simultaneously provided the number of regression coefficients less the number of restrictions on them is greater than two. In such cases, we have obtained the conditions under which one method yields superior predictions in comparison to the other. These conditions, it may be remarked, are simple and easy to check.

When the aim is to predict the values of study variable outside the sample, the relative performance of one method with respect to some other method remains unaltered whether one is interested in actual values only or average values or both together. We have also deduced conditions for the superiority of one method over the other in providing efficient predictions. These conditions can be easily verified in any given application and shed light on the choice of scalar characterizing the estimator.

APPENDIX

In order to derive the results in Theorem 1, we observe from (1), (2), (3) and (12) that

$$(p_r - T) = \sigma[X(X'X)^{-1}G(X'X)^{-1}X' - \lambda I_n]u \tag{A.1}$$

Next, we observe that

$$\begin{aligned}
&(p_r - T)'(p_r - T) - (p_{MC} - T)'(p_{MC} - T) \\
&\quad = 2a_{MC}\frac{(Y - Xb)'(Y - Xb)}{b'Gb}\, b'G(X'X)^{-1}X'(p_r - T) \\
&\quad\quad - \left[\frac{a_{MC}(Y - Xb)'(Y - Xb)}{b'Gb}\right]^2 b'G(X'X)^{-1}Gb \\
&\quad = \frac{a_{MC}Y'MY}{Y'X(X'X)^{-1}G(X'X)^{-1}X'Y}\,[2Y'X(X'X)^{-1}G(X'X)^{-1}X'(p_r - T) - a_{MC}Y'MY]
\end{aligned}$$

where $M = I_n - X(X'X)^{-1}X'$.

Substituting (1) and (A.1) on the right hand side and retaining terms up to order $O(\sigma^4)$ only, we find

$$D_{MC} = \frac{a_{MC}}{\beta'G\beta}(\sigma^3 f_3 - \sigma^4 f_4)$$

$$\begin{aligned}
\text{where}\quad f_3 &= 2(1 - \lambda)E[u'Mu \cdot \beta'G(X'X)^{-1}X'u] \\
&= 0 \\
f_4 &= E[u'Mu \cdot u'\{a_{MC}M - 2(1 - \lambda)X(X'X)^{-1}G(X'X)^{-1}X' \\
&\quad + \frac{4(1 - \lambda)}{\beta'G\beta}X(X'X)^{-1}G\beta\beta'G(X'X)^{-1}X'\}u] \\
&= (n - p)[(n - p + 2)a_{MC} - 2(1 - \lambda)(p - j - 2)]
\end{aligned}$$

This leads to the result (15) of Theorem 1. The other result (16) can be similarly derived.

For the results in Theorem 2, we notice that

$$(p_{fr} - T_f) = \sigma[X_f(X'X)^{-1}G(X'X)^{-1}X'u - \lambda u_f] \tag{A.2}$$

Further, we have

$$\begin{aligned}
&(p_{fr} - T_f)'(p_{fr} - T_f) - (p_{fMC} - T_f)'(p_{fMC} - T_f) \\
&\quad = 2a_{MC}\frac{(Y - Xb)'(Y - Xb)}{b'Gb}\, b'G(X'X)^{-1}X_f'(p_{fr} - T_f) \\
&\quad\quad - \left[\frac{a_{MC}(Y - Xb)'(Y - Xb)}{b'Gb}\right]^2 b'G(X'X)^{-1}X_f'X_f(X'X)^{-1}Gb
\end{aligned}$$

$$= \frac{a_{MC} Y'MY}{Y'X(X'X)^{-1}G(X'X)^{-1}X'Y} [2Y'X(X'X)^{-1}G(X'X)^{-1}X_f'(p_{fr} - T_f)$$

$$- a_{MC} Y'MY \frac{Y'X(X'X)^{-1}G(X'X)^{-1}X_f'X_f(X'X)^{-1}G(X'X)^{-1}X'Y}{Y'X(X'X)^{-1}G(X'X)^{-1}X'Y}$$

Using (A.2) and retaining terms up to order $O(\sigma^4)$ only, we can express

$$D_{fMC} = \frac{a_{MC}}{\beta' G \beta} (\sigma^3 e_3 + \sigma^4 e_4)$$

where

$$e_3 = E[2u'Mu\{\beta'G(X'X)^{-1}X_f'X_f(X'X)^{-1}G(X'X)^{-1}X'u - 2\lambda\beta'G(X'X)^{-1}X_f'u_f\}]$$

$$e_4 = E\Bigg[u'Mu\Bigg\{2u'X(X'X)^{-1}G(X'X)^{-1}X_f'X_f(X'X)^{-1}$$
$$\times\left(G - \frac{2}{\beta'G\beta} G\beta\beta'G\right)(X'X)^{-1}X'u - 2\lambda u'X(X'X)^{-1}\left(G - \frac{2}{\beta'G\beta} G\beta\beta'G\right)$$
$$\times (X'X)^{-1}X_f'u_f - a_{MC}ku'Mu\Bigg\}\Bigg]$$

with k defined by (27).

Employing the distributional properties of u and u_f, it is easy to verify that

$$e_3 = 0$$
$$e_4 = (n-p)[2(g-2k) - a_{MC}k(n-p+2)]$$

whence we find the result (25) stated in Theorem 2.

The result (26) of Theorem 2 can be similarly derived.

References

[2] R.C. Mittelhammer and R.K. Conway (1984), "On the admissibility of restricted least squares estimators", *Communications in Statistics*, **13**, 1135–1145.

[2] C.R. Rao (1973), *Linear Statistical Inference and Its Applications*, John Wiley & Sons, New York.

[3] Shalabh (1995), "Performance of Stein-rule procedure for simultaneous prediction of actual and average values of study variable in linear regression model", *Bulletin of the Fiftieth Session of International Statistical Institute (Netherlands)*, 1375–1390.

[4] D.S. Tracy and A.K Srivastava (1995), "Inadmissibility of restricted estimators in linear regression models", *Journal of Quantitative Economics*, **11**, 127–133.

[5] A. Zellner (1994), "Bayesian and non-Bayesian estimation using balanced loss functions", in: S.S. Gupta and J.O. Berger (Eds.), *Statistical Theory and Related topics-V*, Springer-Verlag, New York.