Erratum: Biom. J. 39 (1997) 7, 879

Predictive Performance of the Methods of Restricted and Mixed Regression Estimators

H. Toutenburg

Institut für Statistik
Universität München
Germany

Shalabh

Department of Statistics
University of Jammu
India

Summary

This article considers the problem of simultaneous prediction of actual and average values of the study variable in a linear regression model when a set of linear restrictions binding the regression coefficients is available, and analyzes the performance properties of predictors arising from the methods of restricted regression and mixed regression besides least squares.

1. Introduction

Generally predictions from a linear regression model are made either for the actual values of the study variable or for the average values at a time. However, situations may occur in which one may be required to consider the predictions of both the actual and average values simultaneously. For example, consider the installation of an artificial tooth in patients through a specific device. Here a dentist would like to know the life of a restoration, on the average. On the other hand, a patient would be more interested in knowing the actual life of restoration in his/her case. Thus a dentist is interested in the prediction of average value but he may not completely ignore the interest of patients in the prediction of actual value. The dentist may assign higher weightage to prediction of average values in comparison to the prediction of actual values. Similarly, a patient may give more weightage to prediction of actual values in comparison to that of average values.

   Appreciating the need of simultaneous prediction of actual and average values of the study variable, Shalabh (1995) has proposed a composite target function which considers the prediction of both the actual and average values together. Using such an approach, we expose the traditional prediction methods and analyze

their performance properties when a set of linear restrictions binding the coefficients in a linear regression model is available in the form of prior information.

The organization of this paper is as follows. Section 2 describes the linear regression model and presents the target function for the prediction of actual and average values of the study variable. Section 3 assumes the availability of a set of exact linear restrictions on regression coefficients. Two predictors arising from the methods of least squares (that ignores the restrictions) and restricted regression (that incorporates the restrictions) are presented and their efficiency properties are analyzed. Similarly, in Section 4, we assume the availability of a set of stochastic linear restrictions and study the performance properties of predictors stemming from pure and mixed estimation methods. Finally, some remarks are placed in Section 5.

2. Specification of Model and Target Function

Let us postulate the following linear regression model:

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

where y is an $n \times 1$ vector of n observations on the study variable, X is an $n \times K$ full column rank matrix of n observations on K explanatory variables, β is a column vector of regression coefficients, σ is an unknown scalar and u is an $n \times 1$ vector of disturbances.

It is assumed that the elements of u are independently and identically distributed with mean zero and variance unity.

If b denotes an estimator of β, then the predictor for the values of study variable within the sample is generally formulated as $\hat{T} = Xb$ which is used for predicting either the actual values y or the average values $\mathrm{E}(y) = X\beta$ at a time.

When the situation demands prediction of both the actual and average values together, we may define the following target function

$$T(y) = \lambda y + (1 - \lambda)\,\mathrm{E}(y) = T \tag{2.2}$$

and use $\hat{T} = Xb$ for predicting it where $0 \leq \lambda \leq 1$ is a nonstochastic scalar specifying the weightage to be assigned to the prediction of actual and average values of the study variable; see, e.g. SHALABH (1995).

Next, let us consider the prediction of values of the study variable outside the sample. Accordingly, we assume that a set of n_f values of the explanatory variables is given so that

$$y_f = X_f\beta + \sigma u_f \tag{2.3}$$

where y_f is an $n_f \times 1$ vector of future values of the study variable, X_f is an $n_f \times K$ matrix of values of the explanatory variables and u_f is an $n_f \times 1$ vector of disturbances possessing the same distributional properties as u. Further, elements of u_f and u are stochastically independent.

Now we may construct the predictor as $\hat{T}_f = X_f b$ and the target function as

$$T_f = \lambda y_f + (1 - \lambda) \, E(y_f) . \tag{2.4}$$

It may be observed that the target function permits a kind of unified treatment to the problem of predicting the actual and average values of the study variable through its characterizing scalar λ.

3. Exact Linear Restrictions

Let us suppose that we are given a set of J exact linear restrictions binding the regression coefficients:

$$r = R\beta \tag{3.1}$$

where r is a $J \times 1$ vector and R is a $J \times K$ full row rank matrix.

If these restrictions are ignored, the least squares estimator of β is

$$\hat{\beta} = (X'X)^{-1} X'y \tag{3.2}$$

which may not necessarily obey (3.1). Such is, however, not the case with restricted regression estimator given by

$$\hat{\beta}_R = \hat{\beta} + (X'X)^{-1} R'[R(X'X)^{-1} R']^{-1} (r - R\hat{\beta}) \tag{3.3}$$

which invariably satisfies (3.1).

3.1 Prediction within the Sample

Employing (3.2) and (3.3), we get the following two predictors for the values of the study variable within the sample:

$$\hat{T} = X\hat{\beta} , \tag{3.4}$$

$$\hat{T}_R = X\hat{\beta}_r . \tag{3.5}$$

It is easy to see from (2.2) that both the predictors are weakly unbiased in the sense that

$$E_\lambda(\hat{T} - T) = 0 , \tag{3.6}$$

$$E_\lambda(\hat{T}_R - T) = 0 . \tag{3.7}$$

Further, the predictive variance covariance matrices are given by

$$V_\lambda(\hat{T}) = E(\hat{T} - T) (\hat{T} - T)'$$

$$= \sigma^2 [\lambda^2 I_n + (1 - 2\lambda) P] , \tag{3.8}$$

where $P = X(X'X)^{-1}X' = XX^+$ and $X^+ = (X'X)^{-1}X'$,

$$V_\lambda(\hat{T}_R) = \mathrm{E}(\hat{T}_R - T)\,(\hat{T}_R - T)'$$
$$= \sigma^2[\lambda^2 I_n + (1 - 2\lambda)\,P - (1 - 2\lambda)\,X^{+'}R'[R(X'X)^{-1}R']^{-1}RX^+]$$

whence we find

$$V_\lambda(\hat{T}) - V_\lambda(\hat{T}_R) = \sigma^2(1 - 2\lambda)\,X^{+'}R'[R(X'X)^{-1}R']^{-1}RX^+ \qquad (3.9)$$

which is a nonnegative definite matrix as far as $(1 - 2\lambda) \geq 0$.

It is thus seen that the predictor $\hat{T}_R$ is at least as superior as $\hat{T}$ according to the criterion of variance covariance matrix when λ is less than 0.5. In other words, if prediction of actual values of the study variable is assigned relatively lesser weight in comparison to the prediction of average values, incorporating the restrictions in the estimation of β for constructing predictions is surely a better procedure than ignoring the restrictions. On the other hand if higher weightage is to be assigned to the prediction of actual values in comparison to the prediction of average values, the incorporation of restrictions in the estimation procedure may not serve any useful purpose. An interesting implication of this observation is that efficient estimation of regression coefficients may not necessarily lead to efficient predictions of values of study variable.

Let us now examine the predictive risks associated with the two predictors $\hat{T}$ and $\hat{T}_R$. When the aim is simply to predict the average values of the study variable $(\lambda = 0)$, the predictive risks of $\hat{T}$ and $\hat{T}_R$ are $\sigma^2 K$ and $\sigma^2(K - J)$ respectively. Similarly, when the aim is solely to predict the actual values $(\lambda = 1)$, the predictive risks of $\hat{T}$ and $\hat{T}_R$ are $\sigma^2(n - K)$ and $\sigma^2(n - K + J)$ respectively. Thus the predictions based on $\hat{T}$ will have smaller risk property for average values of the study variable rather than for its actual values if n exceeds $2K$. If n exceeds $2(K - J)$, then the predictions based on $\hat{T}_R$ have this property. Just the reverse will be true, i.e., predictions will be more suitable for actual values of the study variable than for its average values when n is less than $2K$ in case of $\hat{T}$ and $2(K - J)$ in case of $\hat{T}_R$.

4. Stochastic Linear Restrictions

Let us be given a set of following stochastic linear restrictions connecting the regression coefficients:

$$r = R\beta + v \qquad (4.1)$$

where v is a $J \times 1$ random vector with mean vector $\mathbf{0}$ and variance covariance matrix Ψ which is positive definite and known.

Further, it is assumed that u and v are stochastically independent.

When the prior restrictions (4.1) are ignored and least squares method is applied to (2.1), we obtain the pure regression estimator of β as $\hat{\beta}$ given by (3.2). When

the prior restrictions are incorporated, we get the mixed regression estimator

$$\hat{\beta}_M = (X'X + s^2 R' \Psi^{-1} R)^{-1} (X'y + s^2 R' \Psi^{-1} r) \tag{4.2}$$

where $s^2 = (y - X\hat{\beta})' (y - X\hat{\beta})/(n - K)$ is an unbiased estimator of σ^2; see THEIL and GOLDBERGER (1961).

4.1 Prediction within the Sample

Based on the pure and mixed regression estimators of β, one can formulate the predictor $\hat{T}$ defined by (3.4) and the following predictor

$$\hat{T}_M = X\hat{\beta}_M . \tag{4.3}$$

Following Kakwani (1968), it can be easily seen that the predictor $\hat{T}_M$ is weakly unbiased for the target function T specified by (2.2) in the sense that $E(\hat{T}_M - T) = 0$ provided that the mean of the predictor $\hat{T}_M$ is finite and the elements of u follow a symmetric probability distribution. The exact predictive risk will, however, be difficult to derive unless we assume any specific form of the symmetric distribution for disturbances.

If we assume normality of disturbances, the exact expression for predictive variance covariance matrix can be derived but it will be sufficiently intricate and may not permit us to draw any clear inference; see, e.g., SWAMY and MEHTA (1969). We therefore propose to employ an asymptotic approximation of it employing the small disturbance asymptotic theory.

Result 1: When disturbances are normally distributed, the predictive variance covariance matrix of $\hat{T}_M$ to order $O(\sigma^4)$ is given by

$$V_\lambda(\hat{T}_M) = E(\hat{T}_M - T) (\hat{T}_M - T)'$$
$$= \sigma^2 [\lambda^2 I_n + (1 - 2\lambda) P] - \sigma^4 \left[1 - 2 \left(\lambda + \frac{1}{n - K} \right) \right] X^{+'} R' \Psi^{-1} R X^+ . \tag{4.4}$$

This result is derived in the Appendix.

Comparing (4.4) with (3.8), it is observed that the predictor $\hat{T}_M$ is better than $\hat{T}$ with respect to the criterion of predictive variance covariance matrix so long as

$$\lambda < \left(\frac{1}{2} - \frac{1}{n - K} \right) \tag{4.5}$$

which implies that the predictor based on mixed regression estimation is preferable when prediction of actual values of study variable receives smaller weight in comparison to the prediction of average values. However, if prediction of average values is at least as important as the prediction of actual values, it may not be worthwhile to take into account the prior restrictions in the estimation of regression coefficients.

4.2 *Prediction outside the Sample*

When the aim is to predict the values of study variable outside the sample, the following predictor arising from mixed regression estimation may be used:

$$\hat{T}_{fM} = X_f \hat{\beta}_M. \tag{4.6}$$

Assuming multivariate normality of u and u_f, it is easy to see that $\hat{T}_{fM}$ is unbiased for T_f, i.e.,

$$\mathrm{E}(\hat{T}_{fM} - T_f) = 0. \tag{4.7}$$

For the predictive risk of $\hat{T}_{fM}$, we present an asymptotic approximation employing the small disturbance asymptotic theory.

Result 2: When disturbances are normally distributed, the predictive variance covariance matrix associated with $\hat{T}_{fM}$ upto order $O(\sigma^4)$ is given by

$$
\begin{aligned}
V_\lambda(\hat{T}_{fM} &= \mathrm{E}(\hat{T}_{fM} - T_f)(\hat{T}_{fM} - T_f)' \\
&= \sigma^2[\lambda^2 I_{nf} + X_f(X'X)^{-1} X_f'] \\
&\quad - \sigma^4\left(1 - \frac{2}{n-K}\right) X_f(X'X)^{-1} R' \Psi^{-1} R(X'X)^{-1} X_f
\end{aligned}
\tag{4.8}
$$

which is derived in the Appendix.

Comparing (4.8) with $V_\lambda(\hat{T}_f) = \sigma^2[\lambda^2 I_{nf} + X_f(X'X)^{-1} X_f']$ the risk of the predictor $\hat{T}_f$ which does not incorporate the prior stochastic restrictions, we observe that $\hat{T}_{fM}$ is superior to $\hat{T}_f$ as long as $(n - K)$, the excess of the number of observations over the number of regression coefficients, exceeds 2 which is indeed a very mild constraint.

Setting $\lambda = 0$ and $\lambda = 1$ in (4.8), it is observed that like $\hat{T}_f$ the predictor $\hat{T}_{fM}$ has better performance when used for predicting average values of the study variable.

5. Some Remarks

Considering the problem of simultaneous predictions of actual and average values of the study variable in linear regression models subject to a set of linear restrictions connecting the regression coefficients, it is found that the least squares procedure ignoring the restrictions continues to remain unbeaten as long as larger weight is given to the prediction of actual values in relation to the prediction of average values. Incorporation of prior restrictions into the estimation procedure is reasonable only when the opposite is true, i.e., prediction of actual values receives smaller weight in comparison to the prediction of average values. Interestingly enough, all the three estimation procedures (viz., pure regression, restricted regres-

sion and mixed regression) are found to be more suitable, under some mild constraint, in the sense of yielding predictions with smaller variability when the aim is to predict the average values of the study variable rather than its actual values.

The above findings relate to the prediction of values within the sample which essentially may reflect the success of an estimation procedure. These findings do not carry over to the case of prediction outside the sample such as forecasting. Although all the predictors continue to remain weakly unbiased (assuming normality), their performances with respect to the criterion of predictive variance covariance matrix change. It is observed that utilization of prior restrictions is always a better strategy than ignoring them irrespective of the weightage assigned to the prediction of actual and average values. However, each of the three predictors is seen to provide more efficient predictions for average values rather than for actual values just as in case of prediction within the sample.

We have confined our attention to three simple traditional predictors. It will be interesting to analyze the properties of other predictors; see, e.g., TOUTENBURG and TRENKLER (1990), TOUTENBURG, TRENKLER and LISKI (1992) and RAO and TOUTENBURG (1995).

A Appendix: Derivation of Results

From (2.1) and (4.2), we can express

$$
\begin{aligned}
(\boldsymbol{b}_M - \boldsymbol{\beta}) &= [\boldsymbol{X}'\boldsymbol{X} + \sigma^2 \left(\frac{\boldsymbol{u}'\boldsymbol{M}\boldsymbol{u}}{n - K} \right) \boldsymbol{R}'\boldsymbol{\Psi}^{-1}\boldsymbol{R}]^{-1} \left[\sigma \boldsymbol{X}'\boldsymbol{u} + \sigma^2 \left(\frac{\boldsymbol{u}'\boldsymbol{M}\boldsymbol{u}}{n - K} \right) \boldsymbol{R}'\boldsymbol{\Psi}^{-1}\boldsymbol{v} \right] \\
&= \sigma(\boldsymbol{X}'\boldsymbol{X})^{-1} \boldsymbol{X}'\boldsymbol{u} + \sigma^2 \left(\frac{\boldsymbol{u}'\boldsymbol{M}\boldsymbol{u}}{n - K} \right) (\boldsymbol{X}'\boldsymbol{X})^{-1} \boldsymbol{R}'\boldsymbol{\Psi}^{-1}\boldsymbol{v} \\
&\quad - \sigma^3 \left(\frac{\boldsymbol{u}'\boldsymbol{M}\boldsymbol{u}}{n - K} \right) (\boldsymbol{X}'\boldsymbol{X})^{-1} \boldsymbol{R}'\boldsymbol{\Psi}^{-1}\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1} \boldsymbol{X}'\boldsymbol{u} + O_p(\sigma^4)
\end{aligned}
$$

$$(A.1)$$

where $\boldsymbol{M} = \boldsymbol{I}_n - \boldsymbol{P}$.

Observing that

$$(\boldsymbol{P}_M - \boldsymbol{T}) = \boldsymbol{X}(\boldsymbol{b}_M - \boldsymbol{\beta}) - \sigma\lambda\boldsymbol{u}$$

we can write

$$(\boldsymbol{P}_M - \boldsymbol{T})(\boldsymbol{P}_M - \boldsymbol{T})' = \sigma^2\boldsymbol{G}_2 + \sigma^3(\boldsymbol{G}_3 + \boldsymbol{G}_3') + \sigma^4(\boldsymbol{G}_4 + \boldsymbol{G}_4') + O_p(\sigma^5),$$

$$(A.2)$$

where

$$
\begin{aligned}
\boldsymbol{G}_2 &= [(1 - \lambda)\,\boldsymbol{I}_n - \boldsymbol{M}]\,\boldsymbol{u}\boldsymbol{u}'[(1 - \lambda)\,\boldsymbol{I}_n - \boldsymbol{M}], \\
\boldsymbol{G}_3 &= \left(\frac{\boldsymbol{u}'\boldsymbol{M}\boldsymbol{u}}{n - K} \right) [(1 - \lambda)\,\boldsymbol{I}_n - \boldsymbol{M}]\,\boldsymbol{u}\boldsymbol{v}'\boldsymbol{\Psi}^{-1}\boldsymbol{R}\boldsymbol{X}^{+},
\end{aligned}
$$

$$G_4 = \frac{1}{2} \left(\frac{u'Mu}{n-K} \right)^2 X^{+'} R' \Psi^{-1} v v' \Psi^{-1} RX^+ ,$$

$$- \left(\frac{u'Mu}{n-K} \right) [(1-\lambda) I_n - M] uu'X^{+'} R' \Psi^{-1} RX^+ .$$

Utilizing the multivariate normality of u and stochastic independence of u and v, it can be easily verified that

$$E(G_2) = (1-\lambda)^2 I_n - (1-2\lambda) M ,$$

$$E(G_3) = 0 ,$$

$$E(G_4) = -\frac{1}{2} \left[(1-2\lambda) - \left(\frac{2}{n-K} \right) \right] X^{+'} R' \Psi^{-1} RX^+ .$$

Taking expectation of both the sides of (A.2) and using the above results, we obtain the expression (4.4).

For the expression (4.8), we notice that

$$(P_{fM} - T_f) = X_f(b_M - \beta) - \sigma\lambda u_f$$

whence, using (A.1), it follows that

$$(P_{fM} - T_f)(P_{fM} - T_f)' = \sigma^2 G_2^* + \sigma^3 (G_3^* + G_3^{'*}) + \sigma^4 (G_4^* + G_4^{'*}) + O_p(\sigma^5)$$

$$\text{(A.3)}$$

where

$$G_2^* = [X_f(X'X)^{-1} X'u - \lambda u_f] [X_f(X'X)^{-1} X'u - \lambda u_f]' ,$$

$$G_3^* = \left(\frac{u'Mu}{n-K} \right) [X_f(X'X)^{-1} X'u - \lambda u_f] v' \Psi^{-1} R(X'X)^{-1} X_f ,$$

$$G_4^* = \frac{1}{2} \left(\frac{u'Mu}{n-K} \right)^2 X_f(X'X)^{-1} R' \Psi^{-1} v v' \Psi^{-1} R(X'X)^{-1} X_f'$$

$$- \left(\frac{u'Mu}{n-K} \right) [X_f(X'X)^{-1} X'u - \lambda u_f] u'X(X'X)^{-1} R' \Psi^{-1} R(X'X)^{-1} X_f' .$$

By virtue of the multivariate normality of u and u_f along with stochastic independence of u, u_f and v, we observe that

$$E(G_2^*) = \lambda^2 I_{nf} + X_f(X'X)^{-1} X_f' ,$$

$$E(G_3^*) = 0 ,$$

$$E(G_4^*) = -\frac{1}{2} \left(1 - \frac{2}{n-K} \right) X_f(X'X)^{-1} R' \Psi^{-1} R(X'X)^{-1} X_f .$$

Taking expectations on both the sides of (A.3) and substituting the above results, we find the expression (4.8).

References

KAKWANI, N. C., 1968: Note on the unbiasedness of mixed regression estimation. *Econometrica* **36**, 610–611.

RAO, C. R. and TOUTENBURG, H., 1995: *Linear Models: Least Squares and Alternatives*. Springer, New York.

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 International Statistical Institute* **56**, 1375–1390.

SWAMY, P. A. V. B. and MEHTA, J. S., 1969: On Theil's mixed regression estimator. *Journal of the American Statistical Association* **66**, 273–276.

THEIL, H. and GOLDBERGER, A. S., 1961: On pure and mixed estimation in econometrics. *International Economic Review* **2**, 65–78.

TOUTENBURG, H. and TRENKLER, G., 1990: Mean square error matrix comparisons of optimal and classical predictors and estimators in linear regression. *Computational Statistics and Data Analysis* **10**, 297–305.

TOUTENBURG, H., TRENKLER, G., and LISKI, E. P., 1992: Optimal erstimation methods under weakened linear restrictions in regression. *Computational Statistics and Data Analysis* **14**, 527–536.

H. TOUTENBURG
Institut für Statistik
Universität München
D-80799 München
Germany

SHALABH
Department of Statistics
University of Jammu
Jammu 180004
India

7,879

Erratum

TOUTENBURG, H., SHALABH: Predictive Performance of the Methods of Restricted and Mixed Regression Estimators. *Biometrical Journal* **38** (7), 951–959.

In the Appendix (pp. 957–958), the quantities b_M, P_M and P_{fM} should be read as $\hat{\beta}_M$, $\hat{T}_M$ and $\hat{T}_{fM}$, respectively.

<div align="right">Received, 29 July 1997</div>