Toutenburg, Shalabh:

# Estimation of Regression Coefficients Subject to Exact Linear Restrictions when some Observations are Missing and Balanced Loss Function is Used

Sonderforschungsbereich 386, Paper 163 (1999)

# Estimation of Regression Coefficients Subject to Exact Linear Restrictions when some Observations are Missing and Balanced Loss Function is Used

H. Toutenburg
Institut für Statistik,
Universität München

Shalabh
Department of Statistics
Panjab University

July 26, 1999

**Abstract**

This article considers a linear regression model when a set of exact linear restrictions binding the coefficients is available and some observations on the study variable are missing. Estimators for the vectors of regression coefficients are presented and their superiority properties with respect to the criteria of the variance covariance matrix and the risk under balanced loss functions are analyzed.

**Key Words** Balanced loss function, exact restrictions, missing observations

## 1 Introduction

For analyzing the efficiency properties of any estimation procedure for the coefficients in a linear regression model, the performance criteria is either the precision of estimates or the goodness of fitted model. In several practical situations, as pointed out by Zellner (1994), it may be desirable to use both the criteria simultaneously assigning possibly unequal weights. This has led Zellner (1994) to introduce the balanced loss function following a quadratic structure for characterizing the losses. Such a loss function is indeed a convex linear combination of the residual sum of squares and the weighted sum of estimation errors. It also permits a kind of unified treatment to the two criteria for the performance analysis.

Application of balanced loss function to some specific problems in linear regression models has provided some illuminating findings; see, e.g., Giles, Giles and Ohtani (1996), Ohtani (1998), Ohtani, Giles and Giles (1997) Wan (1994) and Zellner (1994). All these investigations assume that the available data set is complete and there is no missing observation- Such a specification may not be tenable in actual practice, and some observations may be unavailable; see, e.g., Little and Rubin (1987) for an interesting exposition. In this article, we assume that some observations on the study variable are missing. Now the estimation of regression coefficients by least squares method using all the observations provides essentially the same estimators as obtained by an application of the least squares method to complete observations alone; see, e.g., Rao and Toutenburg (1995, Chap. 8). Thus incomplete observations play absolutely no role, and no gain in the efficiency is achieved despite their use is in estimation procedure. This result may take a pleasent turn when some additional information about the model is available. This is the point of investigation here.

Let us assume that the prior information about the model consists of the specification of some exact restrictions binding the regression coefficients. Such a prior information may be derived from past experience of similar investigations and/or from the exhibition of stability of estimates of regression coefficients in repeated studies and/or from some extraneous sources and/or from some theoretical considerations; see, e.g., Judge, Griffiths, Hill, Lütkepohl and Lee (1985, Chap.3) and Rao and Toutenburg (1995, Chap. 5). Incorporation of such a prior information into the estimation procedure, it is well documented, leads to generally efficient estimation of regression coefficients provided that there is no missing observation in the data set. This article examines the truthfulness of this result when some observations on the study variable are missing and the performance criteria are the variance covariance matrix and the risk under balanced loss function.

The organization of this article is as follows. In Section2, we describe the model and present the estimators. Section 3 reports the superiority comparisons according to the criterion of variance covariance matrix while Section 4 compares the estimators with respect to the criterion of risk under balanced loss function and discusses the optimal choice of the estimator. Finally, some concluding remarks are placed in Section 5.

# 2 Model Specification And Estimators

Let us postulate a linear regression model in which there are $n_c$ complete and $n_m$ incomplete observations:

$$Y_c = X_c\beta + \sigma\epsilon_c \tag{2.1}$$

$$Y_{mis} = X_m\beta + \sigma\epsilon_m \tag{2.2}$$

where $Y_c$ and $Y_{mis}$ denote column vectors of observations on the study variable, $X_c$ and $X_m$ are matrices of observations on $K$ explanatory variables, $\epsilon_c$ and $\epsilon_m$ are column vectors of disturbances, $\beta$ denotes the column vector of unknown regression coefficients and $\sigma$ is an unknown scalar.

It is assumed that observations in $Y_{mis}$ are missing. Further, the elements of vectors $\epsilon_c$ and $\epsilon_m$ are independently and identically distributed with mean zero and variance unity.

Besides the observations, let us assume to be given a set of $J$ exact linear restrictions binding the regression coefficients:

$$r = R\beta \tag{2.3}$$

where the $J \times 1$ vector $r$ and $J \times K$ matrix $R$ contains known elements. It is assumed that redundant restrictions are absent so that $R$ has full row rank.

From (2.1), we observe that the unrestricted least squares estimator of $\beta$ is given by

$$b = (X_c'X_c)^{-1}X'Y_c \tag{2.4}$$

which may not necessarily obey the restrictions (2.3). Such is, however, not the case with restricted least squares estimator:

$$b_R = b + (X_c'X_c)^{-1}R'[R(X_c'X_c)^{-1}R']^{-1}(r - Rb) \tag{2.5}$$

which does satisfy the restrictions.

The estimators $b$ and $b_R$ can now be used to find the predicted or imputed values $X_m b$ and $X_m b_R$ for $Y_{mis}$; see, e.g., Toutenburg and Shalabh (1996) for their predictive performance. Now using these in (2.2) and applying least squares method, we find the following estimators of $\beta$:

$$\begin{aligned} \hat{\beta} &= (X_c'X_c + X_m'X_m)^{-1}(X_c'Y_c + X_m'X_m b) & (2.6)\\ &= b \\ \hat{\beta}_R &= (X_c'X_c + X_m'X_m)^{-1}(X_c'Y_c + X_m'X_m b_R) & (2.7)\\ &= b + (X_c'X_c + X_m'X_m)^{-1}X_m'X_m(X_c'X_c)^{-1} \\ &\qquad R'[R(X_c'X_c)^{-1}R']^{-1}(r - Rb). \end{aligned}$$

Observing that

$$(X_c'X_c + X_m'X_m)^{-1}X_m'X_m = \mathbf{I}_p - (X_c'X_c + X_m'X_m)^{-1}X_c'X_c \tag{2.8}$$

we can express

$$\hat{\beta}_R = (X_c'X_c + X_m'X_m)^{-1}(X_c'X_c b + X_m'X_m b_R) \tag{2.9}$$

whence it follows that the estimator $\hat{\beta}_R$ is a matrix weighted average of unrestricted and restricted estimators.

The equivalence of $\hat{\beta}$ and $b$ is a well-known result implying that use of imputed values derived from complete observations but ignoring the prior information has no impact on the estimation of regression coefficients.

Further, it may be observed that both the estimators $\hat{\beta}$ and $\hat{\beta}_R$ do not satisfy the prior restrictions (2.3). To overcome this problem, we may assume for a moment that $y_{mis}$ is known, and then apply the method of restricted least squares to (2.1), (2.2) and (2.3). The thus obtained estimator of $\beta$ will obviously contain $Y_{mis}$ which can now be replaced by its imputed value. This proposition provides the following two estimators of $\beta$:

$$\begin{aligned} \tilde{\beta} &= \hat{\beta} + (X_c'X_c + X_m'X_m)^{-1}R'[R(X_c'X_c + X_m'X_m)^{-1}R']^{-1}(r - R\hat{\beta}) \\ &= b + (X_c'X_c + X_m'X_m)^{-1}R'[R(X_c'X_c + X_m'X_m)^{-1}R']^{-1}(r - Rb) & (2.10)\\ \tilde{\beta} &= \hat{\beta}_R + (X_c'X_c + X_m'X_m)^{-1}R'[R(X_c'X_c + X_m'X_m)^{-1}R']^{-1}(r - R\hat{\beta}_R) & (2.11) \end{aligned}$$

which clearly obey the prior restrictions.

Using (2.8), we observe that

$$
\begin{aligned}
(r - R\hat{\beta}_R) &= r - Rb - R(X_c'X_c + X_m'X_m)^{-1} \\
&\qquad X_m'X_m(X_c'X_c)^{-1}R'[R(X_c'X_c)^{-1}R'](r - Rb) \\
&= r - Rb - R[(X_c'X_c)^{-1} - (X_c'X_c + X_m'X_m)^{-1}R']^{-1} \\
&\qquad R'[R(X_c'X_c)^{-1}R'](r - Rb) \\
&= R(X_c'X_c + X_m'X_m)^{-1}R'[R(X_c'X_c)^{-1}R']^{-1}(r - R_b)
\end{aligned}
$$

whence it follows that

$$\tilde{\beta}_R = b_R. \tag{2.12}$$

Thus we have four distinct estimators $b$, $b_R$, $\hat{\beta}_R$ and $\tilde{\beta}$ of $\beta$ out of which only $b_R$ and $\tilde{\beta}$ satisfy the prior restrictions while the remaining two estimators $b$ and $\hat{\beta}_R$ may not necessarily satisfy them. Further, the estimators $b$ and $b_R$ fail to utilize the incomplete observations while the estimators $\hat{\beta}_R$ and $\tilde{\beta}$ use all the available observations.

# 3 Variance Covariance Matrix

Employing (2.1) and (2.3), it can be easily verified from (2.4), (2.5), (2.9), and (2.10) that all the four estimators $b$, $b_R$, $\hat{\beta}_R$ and $\tilde{\beta}$ are unbiased for $\beta$. Further, if we write

$$S_c = \sigma^2(X_c'X_c)^{-1} \tag{3.1}$$
$$S = \sigma^2(X_c'X_c + X_m'X_m)^{-1} \tag{3.2}$$

their variance covariance matrices are given by

$$\mathrm{V}(b) = S_c \tag{3.3}$$
$$\mathrm{V}(b_R) = S_c - S_cR'(RS_cR')^{-1}RS_c \tag{3.4}$$
$$\mathrm{V}(\hat{\beta}_R) = S_c - S_cR'(RS_cR')^{-1}RS_c + SR'(RS_cR')^{-1}RS \tag{3.5}$$
$$
\begin{aligned}
\mathrm{V}(\tilde{\beta}) &= S_c - SR'(RSR')^{-1}RS_c - S_cR'(RSR')^{-1}RS \\
&\quad + SR'(RSR')^{-1}RS_cR'(RSR')^{-1}RS.
\end{aligned} \tag{3.6}
$$

Recalling that the estimators $b$ and $\hat{\beta}_R$ do not satisfy the prior restrictions (2.3), it is seen that

$$
\begin{aligned}
\mathrm{D}(b;\hat{\beta}_R) &= \mathrm{V}(b) - \mathrm{V}(\hat{\beta}_R) \\
&= S_cR'(RS_cR')^{-1}RS_c - SR'(RS_cR')^{-1}RS.
\end{aligned} \tag{3.7}
$$

As $S^{-1} - S_C^{-1} = \sigma^2 X_m'X_m$ is a nonnegative definite matrix, the matrix expression (3.7) is also nonnegative definite implying the superiority of $\hat{\beta}_R$ over $b$ with respect to the criterion of variance covariance matrix.

Similarly, if we compare $b_R$ and $\tilde{\beta}$ which satisfy the prior restrictions (2.3), we observe that

$$
\begin{aligned}
\mathrm{D}(\tilde{\beta};b_R) &= \mathrm{V}(\tilde{\beta}) - \mathrm{V}(b_R) \\
&= G'RS_cR'G
\end{aligned} \tag{3.8}
$$

where

$$G = (RSR')^{-1}RS - (RS_cR')^{-1}RS_c. \tag{3.9}$$

Clearly, the matrix expression (3.8) is nonnegative definite. This means that the estimator $b_R$ is superior to $\tilde{\beta}$.

Next, let us compare the estimators $b$ and $b_R$ which fail to utilize the incomplete observations of the data set. It is seen from (3.3) and (3.4) that

$$\begin{aligned} \mathrm{D}(b; b_R) &= \mathrm{V}(b) - \mathrm{V}(b_R) \\ &= S_cR'(RS_cR')^{-1}RS_c \end{aligned} \tag{3.10}$$

which is a nonnegative definite matrix implying the superiority of restricted estimator $b_R$ over the unrestricted estimator $b$. This is a well documented result.

Similarly, if we compare the estimators $\hat{\beta}$ and $\tilde{\beta}$ which utilize the entire data set, we observe from (3.5) and (3.6) that

$$\begin{aligned} \mathrm{D}(\tilde{\beta}; \hat{\beta}_R) &= \mathrm{V}(\tilde{\beta}) - \mathrm{V}(\hat{\beta}_R) \\ &= G'RS_cR'G - SR'(RS_cR')^{-1}RS. \end{aligned} \tag{3.11}$$

It is, however, difficult to determine whether the matrix expression (3.11) is nonnegative definite or not. Consequently, no general conclusion can be drawn regarding the superiority of one estimator over the other with respect to the criterion of variance covariance matrix in this case.

Finally, looking at the expression (3.3), (3.4), (3.5) and (3.6), it is interesting to observe that the variance covariance matrix of $b$ exceeds the variance covariance matrices of $b_R$, $\hat{\beta}_R$ and $\tilde{\beta}$ by a nonnegative definite matrix implying the superiority of the three estimators over $b$.

# 4 Risk Function Under Balanced Loss

If $\beta^*$ is any estimator of $\beta$, the risk under balanced loss function is defined by

$$\begin{aligned} \rho(\beta^*) &= wE[(Y_c - X_c\beta^*)'(Y_c - X_c\beta^*) + (Y_{mis} - X_m\beta^*)'(Y_{mis} - X_m\beta^*)] \\ &\quad + (1-w)E[(\beta^* - \beta)'(X_c'X_c + X_m'X_m)(\beta^* - \beta)] \end{aligned} \tag{4.1}$$

where $w$ is a nonnegative constant not exceeding 1 and can be regarded as weight being assigned to the criterion of goodness of fitted model in comparison to the criterion of precision of estimator; see Zellner (1994).

Expressing

$$\rho(\beta^*) = \sigma^2 \,\mathrm{tr}\, S^{-1}\,\mathrm{V}(\beta^*) + w[\sigma^2(n_c + n_m) - 2\sigma\,\mathrm{E}\{(\beta^* - \beta)(\epsilon_c'X_c + \epsilon_m'X_m)\}]$$

and using (3.3), (3.4), (3.5) and (3.6), it is easy to obtain the following results

$$\rho(b) = \sigma^2[\mathrm{tr}\, S^{-1}S_c + w(n_c + n_m - 2K)] \tag{4.2}$$

$$\rho(b_R) = \sigma^2[\mathrm{tr}\, S^{-1}S_c - Jf + w(n_c + n_m - 2K + 2J)] \tag{4.3}$$

$$\rho(\hat{\beta}_R) = \sigma^2[\mathrm{tr}\, S^{-1}S_c - J(f-g) + w(n_c + n_m - 2K + 2J - 2Jg)] \tag{4.4}$$

$$\rho(\tilde{\beta}) = \sigma^2[\mathrm{tr}\, S^{-1}S_c - Jh + w(n_c + n_m - 2K + 2J)] \tag{4.5}$$

where

$$\begin{aligned} f &= \frac{1}{J}\,\mathrm{tr}(RS_cR')^{-1}RS_cS^{-1}S_cR' \\ g &= \frac{1}{J}\,\mathrm{tr}(RS_cR')^{-1}RSR' \\ h &= \frac{1}{J}\,\mathrm{tr}(RSR')^{-1}RS_cR'. \end{aligned} \tag{4.6}$$

As $(S^{-1} - S_c^{-1})$ and $(S_c - S)$ are nonnegative matrices, it can be easily verified that

$$\begin{aligned} f &= \frac{1}{J}\,\mathrm{tr}(RS_cR')^{-1}RS_cS^{-1}S_cR' && (4.7) \\ &\geq \frac{1}{J}\,\mathrm{tr}(RS_cR')^{-1}RS_cR' = 1 \\ g &= \frac{1}{J}\,\mathrm{tr}(RS_cR')^{-1}RSR' && (4.8) \\ &\leq \frac{1}{J}\,\mathrm{tr}(RS_cR')^{-1}RS_cR' = 1 \\ h &= \frac{1}{J}\,\mathrm{tr}(RSR')^{-1}RS_cR' && (4.9) \\ &\geq \frac{1}{J}\,\mathrm{tr}(RSR')^{-1}RSR' = 1. \end{aligned}$$

When the criterion for comparing the estimators is the precision of estimators, it follows from setting $w = 0$ in (4.2), (4.3), (4.4), and (4.5) that $\tilde{\beta}$ is the best estimator when $h$ is greater than $f$. If $h$ is less than $f$, the estimator $b_R$ emerges to be the best choice.

Similarly, if we compare the estimators with respect to the criterion of goodness of the fitted model, we find that $\tilde{\beta}$ turns out to be the best choice so long as $h$ exceeds the maximum of 2 and $(f+g)$ while $\hat{\beta}_R$ is the best choice as long as $(f+g)$ exceeds the maximum of 2 and $h$. The estimator $b$, however, remains unbeaten so long as $(f+g)$ and $h$ both are less than 2.

When both the criteria of the precision of estimation and the goodness of fitted model are equally important so that $w = 0.5$, it is observed that the estimator $\tilde{\beta}$ is optimal if $h$ is greater than $f$. When $h$ is smaller than $f$, the optimal choice falls on $b_R$ and $\hat{\beta}_R$ which are equally good.

When the precision of estimators and the goodness of fitted model are assigned unequal weights in the risk function, the estimator $b$ is the best choice when

$$(f-g) < 2w(1-g); \quad h < 2w. \tag{4.10}$$

Similarly, the estimator $b_R$ performs best when

$$h < f; \quad w < \frac{1}{2}. \tag{4.11}$$

The condition for the optimality of $\hat{\beta}_R$ is

$$(f-g) > \max(h, 2w) - 2wg; \quad w > \frac{1}{2} \tag{4.12}$$

Table 4.1: Optimal choice of estimator among the estimators $b$, $b_R$, $\hat{\beta}_R$ and $\tilde{\beta}$ according to the risk under balanced loss function.

| | Optimal choice |
|---|---|
| $w = 0$<br>Precision of estimation | $b_R$ if $h < f$<br>$\tilde{\beta}$ if $h > f$ |
| $w = 1$<br>Goodness of fitted model | $b$ if $\max(h, f + g) < 2$<br>$\hat{\beta}_R$ if $\max(2, h) < (f + g)$<br>$\tilde{\beta}$ if $\max(2, f + g) < h$ |
| $w = \frac{1}{2}$<br>Equal weights to the two criteria | $b_R$ and $\hat{\beta}_R$ if $h < f$<br>$\tilde{\beta}$ if $h > f$ |
| $0 < w < \frac{1}{2}$<br>Larger weight to the precision of estimation | $b_R$ if $h < f$<br>$\tilde{\beta}$ if $h > f$ |
| $\frac{1}{2} < w < 1$<br>Smaller weight to the precision of estimation | $b$ if $w > \frac{1}{2} \max\left(h, \frac{f-g}{1-g}\right)$<br>$\hat{\beta}_R$ if $\frac{f-g}{2(1-g)} < w \frac{h-f+g}{2g}$; $h > \left(\frac{f-g}{1-g}\right)$<br>$\tilde{\beta}$ if $w < \frac{1}{2} \min\left(h, \frac{h-f+g}{g}\right)$; $h > f$ |

while similar condition for $\tilde{\beta}$ is given by

$$h > \max(f, 2w). \tag{4.13}$$

All such conditions are compiled and presented in the Table.

# 5 Some Concluding Remarks

We have considered the problem of estimating the regression coefficients when some observations on the study variable are missing and a set of exact restrictions binding the regression coefficients is available. Several strategies have been formulated which have led to four distinct estimators of coefficient vector. The first estimator $b$ takes into account neither the incomplete observations nor the prior restrictions. The second estimator $b_R$ is the traditional restricted estimator which satisfies the prior restrictions but fails to utilize the incomplete observations. Similarly, the third estimator $\hat{\beta}_R$ succeeds in utilizing the incomplete observations but may not necessarily satisfy the prior restrictions. This estimator incidently turns out to be the matrix weighted average of $b$ and $b_R$. The fourth estimator $\tilde{\beta}$, however, meets both the desirable requirements, i.e., it uses the entire data and it satisfies the prior restrictions too.

All the four estimators are found to estimate the regression coefficients unbiasedly. Comparing them with respect to the criterion of variance covariance matrix, it is seen that the unrestricted estimator $b$ is dominated by the remaining three estimators. This means that discarding the incomplete observations as well as the prior restrictions is not a good strategy at all.

Next, comparing the estimators $b_R$ and $\hat{\beta}_R$ representing the two strategies of discarding incomplete observations but incorporating the prior restrictions and utilizing the incomplete observations but the estimator may not necessarily obey the prior restrictions, it is interesting to find that the first strategy is better than the second one for the efficient estimation of regression coefficients.

Treating $\tilde{\beta}$ as representative of the strategy that utilizes all the available observations and prior restrictions together, it is surprising that this strategy is no better than the one which ignores the incomplete observations but incorporates the prior restrictions. So far as the comparison with the strategy that uses the incomplete observations but may provide an estimator not satisfying the prior restrictions is concerned, no clear inference can be drawn. In other words, this strategy may be sometimes preferable and sometimes not.

Finally, we have compared the estimators according to their performance under balanced loss function which combines the two popular criteria, viz., precision of estimation and goodness of fitted model. Considering five interesting situations, we have identified the optimal choice of estimators and have presented the results in a tabular form for convenience. It may be observed from a look at the conditions that they are simple to use in practice and are helpful in making an appropriate choice among the competing estimators.

It may be remarked that if we take the performance criterion as the variance covariance matrix, the estimator arising from simultaneous utilization of incomplete observations and prior restrictions is uniformly superior to the estimator that ignores the incomplete observations as well as the prior restrictions. Further, conditions for its superiority over the estimators that use either the incomplete observations or prior restrictions but not both are hard to deduce. This is, however, not true when we take the performance criterion as risk under balanced loss function which is indeed a combination of the precision of estimation and the goodness of fitted model. In this case, conditions for its superiority are available and easy to verify.

# References


Giles, J. A., Giles, D. E. A. and Ohtani, K. (1996). The exact risk of some pre-test and Stein-type regression estimators under balanced loss, *Communications in Statistics, Part A—Theory and Methods* **25**: 2901–2929.

Judge, G. G., Griffiths, W. E., Hill, R. C., Lütkepohl, H. and Lee, T.-C. (1985). *The theory and practice of econometrics*, 2 edn, Wiley, New York.

Little, R. J. A. and Rubin, D. B. (1987). *Statistical Analysis with Missing Data*, Wiley, New York.

Ohtani, K. (1998). The exact risk of weighted average of the OLS and Stein-rule estimators in regression under balanced loss, *Economics Letter* pp. 35–45.

Ohtani, K., Giles, D. E. A. and Giles, J. A. (1997). The exact risk performance of a pre-test estimator in a heteroskedastic linear regression model under the balanced loss function, *Econometric Reviews* pp. 119–130.

Rao, C. R. and Toutenburg, H. (1995). *Linear Models: Least Squares and Alternatives*, Springer, New York.

Toutenburg, H. and Shalabh (1996). Predictive performance of the methods of restricted and mixed regression estimators, *Biometrical Journal* **38**: 951–959.

Wan, A. T. K. (1994). Risk comparison of the inequality constrained least squares and other related estimators under balanced loss, *Economics Letter* pp. 203–210.

Zellner, A. (1994). Bayesian and non-bayesian estimation using balanced loss functions, *in* S. S. Gupta and J. O. Berger (eds), *Statistical Decision Theory and Related Topics V*, Springer, Berlin, pp. 377–390.