Biometrical Journal **42** (2000) 2, 203–214

Stein-Rule Estimation in Mixed Regression Models

SHALABH

Department of Statistics
Panjab University
Chandigarh
India

ALAN T. K. WAN

Department of Management Sciences
City University of Hong Kong
Kowloon
Hong Kong

Summary

This paper considers a Stein-rule mixed regression estimator for estimating a normal linear regression model in the presence of stochastic linear constraints. We derive the small disturbance asymptotic bias and risk of the proposed estimator, and analytically compare its risk with other related estimators. A Monte-Carlo experiment investigates the empirical risk performance of the proposed estimator.

Key words: Mixed regression; Risk; Small disturbance; Stein-rule; Monte-Carlo experiment.

1. Introduction

In regression analysis, we often encounter situations in which incomplete prior information is available in the form of a set of stochastic linear constraints binding the coefficients. In the literature, a range of methods, both frequentist and Bayesian in nature, have been proposed for incorporating such information. A method that has received widespread textbook treatment is the mixed regression estimator (MRE) due to THEIL and GOLDBERGER (1961). The MRE is essentially an operational variant of the best linear unbiased estimator of the regression coefficients. Early studies by NAGAR and KAKWANI (1964) and KAKWANI (1968) examined the bias and moment matrix of the estimator, and SRIVASTAVA (1980) provided an annotated bibliography of much of the subsequent developments. More recent contributions emphasized the exact finite sample properties of the MRE, and extended the method of mixed regression to problems of restricted estimation and prediction, and such models as the random coefficient and seemingly unrelated regression equations models. See, for example, SRIVASTAVA, RAJ, and KUMAR (1984), OHTANI and HONDA (1984a, b), SRIVASTAVA and GILES (1987), SRIVASTAVA and OHTANI (1995), TOUTENBURG and SHALABH (1996), among others. SRIVASTAVA and SRIVASTAVA (1983) crafted the idea of Stein-rule estimation [STEIN (1956); JAMES and STEIN (1961)] in the method of mixed regression, and introduced two families of estimators of the regression coefficients. The authors further analyzed the biases and risks of the estimators under a general quadratic loss function using small disturbance asymptotic theory. They found that one family of estimators has no merit, at least to the given order of approximation, while the other provides estimators that have smaller risk in comparison to the mixed regression estimator under some mild constraints. More recently, through the work of TOUTENBURG and SHALABH (2000), the methods of estimation introduced by SRIVASTAVA and SRIVASTAVA (1983) have been extended to the simultaneous prediction of actual and average values of the dependent variable in a regression model.

A problem concerning the estimators specified by SRIVASTAVA and SRIVASTAVA (1983) and used in TOUTENBURG and SHALABH (2000) is that they are formulated either by substituting the mixed regression estimates for the ordinary least squares estimates in the Stein-rule estimator's formula, or by replacing the ordinary least squares by the Stein-rule estimates in the MRE's expression. They should not, therefore, be considered as Stein-rule estimators in the true sense. In this paper, an alternative formulation is considered, and we argue that this formulation brings together the idea of Stein-rule estimation and the mixed regression method more sensibly than do the estimators considered by SRIVASTAVA and SRIVASTAVA (1983). In Section 2, we describe the mixed regression framework and introduce a Stein-rule mixed regression estimator (SRMRE). Section 3 analyzes the approximate bias, mean squared error matrix and risk under quadratic loss of the SRMRE, along with the conditions for the dominance of the SRMRE over other estimators based on the risk criterion. A Monte-Carlo experiment designed to explore the empirical risk performance of the proposed estimator is described in Section 4. It is found that the SRMRE provides much greater risk reduction, compared to the MRE and the estimators of SRIVASTAVA and SRIVASTAVA (1983), over a wide range of parametric values. An appendix containing the derivation of the main results concludes the paper.

2. Model Specification and Estimators

Consider the following linear regression model,

$$y = X\beta + \sigma u\,; \qquad u \sim N(0, I) \tag{2.1}$$

where y is a $n \times 1$ vector of observations on the dependent variable, X is a $n \times p$ full column rank matrix of n observations on p explanatory variables, β is a $p \times 1$

parameter vector, u is a $n \times 1$ vector of disturbances and σ is a scalar. The extraneous information is available in the form of a set of G stochastic linear constraints binding the regression coefficients given by

$$r = R\beta + \nu\,; \qquad \nu \sim N(0, \Psi) \tag{2.2}$$

where R is non-stochastic, $G \times p$ and of rank G $(< p)$, r is $G \times 1$, Ψ is assumed to be known and positive definite, and $E(u\nu') = 0$.

The ordinary least squares (OLS) or minimum variance unbiased estimator of β is,

$$b = (X'X)^{-1}\, X'y\,, \tag{2.3}$$

which is dominated, under a squared error loss measure, by the Stein-rule estimator,

$$b_s = \left(1 - \frac{k}{n-p+\alpha}\,\frac{(y-Xb)'\,(y-Xb)}{b'X'Xb}\right) b\,, \tag{2.4}$$

when $p > 2$ and $0 \le k \le 2(p-2)$, where α is any scalar such that $n - p + \alpha$ is positive. A feature of b and b_s is that both estimators ignore the extraneous information given in (2.2). Now, writing (2.1) and (2.2) compactly, we obtain

$$\begin{bmatrix} y \\ r \end{bmatrix} = \begin{bmatrix} X \\ R \end{bmatrix} \beta + \begin{bmatrix} \sigma u \\ \nu \end{bmatrix}. \tag{2.5}$$

That is, $y^* = Z\beta + w$, where w has a variance-covariance matrix given by $V(w) = \sigma^2 \begin{pmatrix} I & 0 \\ 0 & \Psi/\sigma^2 \end{pmatrix} = \sigma^2\Omega$. If $V(w)$ is known, then β can be estimated using the generalized least squares estimator,

$$\hat{\beta} = \left(X'X + \sigma^2 R'\Psi^{-1}R\right)^{-1} \left(X'y + \sigma^2 R'\Psi^{-1} r\right). \tag{2.6}$$

The estimator $\hat{\beta}$ is non-operational. The mixed regression estimator of THEIL and GOLDBERGER (1961) is just the feasible counterpart to $\hat{\beta}$, namely,

$$b_m = \left(X'X + s^2 R'\Psi^{-1}R\right)^{-1} \left(X'y + s^2 R'\Psi^{-1} r\right), \tag{2.7}$$

where $s^2 = \dfrac{(y-Xb)'(y-Xb)}{n-p+\alpha}$ is an estimator of σ^2. Crafting the idea of Stein-rule in mixed regression, SRIVASTAVA and SRIVASTAVA (1983) introduced the estimators,

$$b^*_{ms} = \left[1 - \frac{k}{n-p+\alpha}\,\frac{(y-Xb_m)'\,(y-Xb_m)}{b'_m X'Xb_m}\right] b_m \tag{2.8}$$

and

$$b_{sm}^* = \left(X'X + \frac{(y - Xb_s)'(y - Xb_s)}{n - p + \alpha} R'\Psi^{-1}R\right)^{-1} \times \left(X'y + \frac{(y - Xb_s)'(y - Xb_s)}{n - p + \alpha} R'\Psi^{-1}r\right) \qquad (2.9)$$

The authors showed that b_{sm}^* is asymptotically equivalent to the MRE in the sense that the asymptotic approximations for the bias vectors and MSE matrix of b_{sm}^* to order $O(\sigma^4)$ are the same as those of the MRE; while b_{ms}^* performs better in terms of risk than the MRE if

$$0 < k^* < 2(p-2)/(n-p+2)\,, \qquad (2.10)$$

provided that $p > 2$, where $k^* = k/(n - p + \alpha)$. Notwithstanding these developments, both b_{sm}^* and b_{ms}^* are somewhat arbitrary in nature as these estimators are obtained either by substituting b_m for b in (2.4), or by replacing b by the Stein-rule estimator b_s in (2.7). Indeed, given (2.5) and the intimate connection between b and b_m , a remedy is to consider, instead, the formulation,

$$b_{ms} = \left(1 - \frac{k}{n - p + \alpha} \frac{(y^* - Zb_m)'\,\hat{\Omega}^{-1}(y^* - Zb_m)}{b_m' Z' \hat{\Omega}^{-1} Z b_m}\right) b_m\,, \qquad (2.11)$$

which we refer to as the Stein-rule mixed regression estimator hereafter, where $\hat{\Omega} = \begin{pmatrix} I & 0 \\ 0 & \Psi/s^2 \end{pmatrix}$. The close analogy between b_s and b_{ms} is obvious. Being based on model (2.5), the SRMRE provides a more formal framework for incorporating the idea of Stein-rule in mixed regression than do the estimators b_{sm}^* and b_{ms}^*. Now, it is easily verified that b_{ms} can be equivalently expressed as,

$$b_{ms} = \left[1 - \frac{k}{n - p + \alpha} \left(\frac{(y - Xb_m)'(y - Xb_m) + s^2(r - Rb_m)'\,\Psi^{-1}(r - Rb_m)}{b_m' X' X b_m + s^2 b_m' R' \Psi^{-1} R b_m}\right)\right] b_m. \qquad (2.12)$$

Comparing the expressions of b_{ms}^* and b_{ms}, it is interesting that b_{ms} looks like an extended version of b_{ms}^*. In the next section, we compare the performance of b_m, b_{ms}^* and b_{ms} using small disturbance asymptotic theory. The properties of b_{sm}^* require no further discussion given the results of SRIVASTAVA and SRIVASTAVA (1983).

3. Properties of Estimators

In view of the results of SRIVASTAVA and SRIVASTAVA (1983), we select $\alpha = 2$ in comparing the properties of b^*_{ms} and b_{ms}. Now, KAKWANI (1968) showed that the MRE is unbiased, and it can be shown (see, for example, TOUTENBURG and SHALABH (1996)) that the variance-covariance matrix of the MRE is,

$$V(b_m) = \sigma^2 (X'X)^{-1} - \sigma^4 \left(\frac{n-p}{n-p+2}\right) (X'X)^{-1} R'\Psi^{-1} R (X'X)^{-1} \tag{3.1}$$

to order $O(\sigma^4)$. For the estimator b^*_{ms}, SRIVASTAVA and SRIVASTAVA (1983) showed that the bias vector, to order $O(\sigma^2)$, is given by,

$$\begin{aligned} B\left(b^*_{ms}\right) &= E\left(b^*_{ms} - \beta\right) \\ &= -\sigma^2 \left(\frac{n-p}{n-p+2}\right) \left(\frac{k}{\beta' X' X \beta}\right) \beta \end{aligned} \tag{3.2}$$

and the difference in MSE matrix between b_m and b^*_{ms} is

$$\begin{aligned} D(b_m; b^*_{ms}) &= E[(b_m - \beta)\,(b_m - \beta)'] - E[\left(b^*_{ms} - \beta\right)\,\left(b^*_{ms} - \beta\right)'] \\ &= 2\sigma^4 \left(\frac{n-p}{n-p+2}\right) \left(\frac{k}{\beta' X' X \beta}\right) \left[(X'X)^{-1} - \frac{4+k}{2\beta' X' X \beta}\, \beta\beta'\right] \end{aligned} \tag{3.3}$$

to order $O(\sigma^4)$.

Now, it is shown in the Appendix that the bias vector of b_{ms} to order $O(\sigma^2)$ is given by

$$\begin{aligned} B(b_{ms}) &= E(b_{ms} - \beta) \\ &= -\sigma^2 \left(\frac{(n-p)\,(n-p+2+G)}{(n-p+2)^2}\right) \left(\frac{k}{\beta' X' X \beta}\right) \beta \end{aligned} \tag{3.4}$$

and the difference in MSE matrix between b_m and b_{ms}, to order $O(\sigma^4)$, is

$$\begin{aligned} D\left(b_m; b_{ms}\right) &= E(b_m - \beta)\,(b_m - \beta)' - E(b_{ms} - \beta)\,(b_{ms} - \beta)' \\ &= 2\sigma^4 \left(\frac{(n-p)\,(n-p+2+G)}{(n-p+2)^2}\right) \left(\frac{k}{\beta' X' X \beta}\right) \\ &\quad \times \left[(X'X)^{-1} - \frac{h}{2\beta X' X \beta}\, \beta\beta'\right], \end{aligned} \tag{3.5}$$

where

$$h = 4 + k\left[1 + \frac{G}{n-p+2}\left(1 + \frac{2}{n-p+2+G}\right)\right]. \tag{3.6}$$

It is thus seen from (3.2) and (3.4) that the estimator b^*_{ms} is less biased in magnitude than b_{ms}. It is also observed the elements of the bias vector of both estimators have signs which are opposite to those of corresponding elements in β. Now, to analyze (3.3) and (3.5) further, we employ the following two lemmas for any $p \times p$ positive definite matrix G and a column vector g:

Lemma 1: The matrix $(G - gg')$ is non-negative definite if and only if $g'G^{-1}g \leq 1$ [see, for instance, YANCEY, JUDGE and BOCK (1974)].

Lemma 2: The matrix $(gg' - G)$ cannot be non-negative definite for $p > 1$ [see, for example, GUILKEY and PRICE (1981)].

Using Lemma 1, we observe that the matrix expression given in (3.3) is positive definite implying the superiority of b^*_{ms} over b_m if and only if $\left(\frac{4+k}{2\beta'X'X\beta}\right)\beta'X'X\beta < 1$, which cannot hold true. Similarly, for (3.5), the matrix expression is positive definite if and only if $\left(\frac{h}{2\beta'X'X\beta}\right)\beta'X'X\beta < 1$, which also cannot be true. In other words, neither b^*_{ms} nor b_{ms} dominates the mixed regression estimator b_m. Now, using Lemma 2, we find that the negative of (3.3) and (3.5) cannot be positive definite for p greater than one. That is, b_m cannot be superior to b^*_{ms} or b_{ms} for any regression model with more than one coefficient. Thus, under the mean square error matrix criterion, b_m neither dominates nor is dominated by b^*_{ms} or b_{ms}. Such is, however, not the case if we consider the risk function,

$$C(\hat{\beta}) = E[(\hat{\beta} - \beta)' Q(\hat{\beta} - \beta)] \tag{3.7}$$

under a weighted squared error loss structure with weight matrix Q, where $\hat{\beta}$ is any estimator of β, and evaluate the estimators with respect to this risk to order $O(\sigma^4)$. Now, for $Q = X'X$, it can be shown, in a manner parallel to that of SRIVASTAVA and SRIVASTAVA (1983), that the risks of b_m and b^*_{ms} are, respectively,

$$C(b_m) = \sigma^2 p - \sigma^4 \left(\frac{n-p}{n-p+2}\right) tr(X'X)^{-1} R'\Psi^{-1}R\,, \tag{3.8}$$

and

$$C(b^*_{ms}) = \sigma^2 p - \sigma^4 \left(\frac{n-p}{n-p+2}\right) \times \left\{ tr(X'X)^{-1} R'\Psi^{-1}R - \frac{k}{\beta'X'X\beta}[k - 2(p-2)] \right\}. \tag{3.9}$$

Hence the estimator b^*_{ms} has smaller risk than b_m whenever,

$$0 < k < 2(p-2)\,, \tag{3.10}$$

provided that p exceeds 2, and the largest amount of risk reduction is achieved when $k = p - 2$. SRIVASTAVA and SRIVASTAVA (1983) used $\alpha = 0$ in deriving the risks of b_m and b^*_{ms}, and this accounts for the minor differences between the risk expressions given there and equations (3.8) and (3.9).

Now, the risk of b_{ms} can be written as,

$$C(b_{ms}) = \sigma^2 p - \sigma^4 \left(\frac{n-p}{n-p+2}\right) \left\{ tr(X'X)^{-1} R' \Psi^{-1} R - \left(\frac{k}{\beta' X' X \beta}\right) \times \left[\left(\frac{n-p+2+G}{n-p+2}\right) \times \left(k \left(1 + \frac{G}{n-p+2} \left(1 + \frac{2}{n-p+2+G}\right)\right) - 2(p-2)\right)\right]\right\}. \tag{3.11}$$

Hence the estimator b_{ms} dominates b_m whenever

$$0 < k < 2(p-2) \left[1 + \frac{G}{n-p+2} \left(1 + \frac{2}{n-p+2+G}\right)\right]^{-1} \tag{3.12}$$

provided that $p > 2$, and the largest reduction in risk arises when

$$k = (p-2) \left[1 + \frac{G}{n-p+2} \left(1 + \frac{2}{n-p+G+2}\right)\right]^{-1}. \tag{3.13}$$

Using (3.9) and (3.11), we find that the estimates b_{ms} has smaller risk than b^*_{ms} whenever

$$0 < k < 2(p-2) \left[2 + \frac{G+2}{n-p+2}\right]^{-1} \tag{3.14}$$

provided that $p > 2$. It can be easily seen that the largest amount of reduction in risk of b_{ms} over b^*_{ms} is achieved when

$$k = (p-2) \left[2 + \frac{G+2}{n-p+2}\right]^{-1} \tag{3.15}$$

provided that p exceeds 2.

4. Monte-Carlo Results

The investigation of the small sample properties of the various estimators is based on the following Monte Carlo experiment using the SHAZAM econometric package version 8. The data for X are generated such that $X'X = I$. We set $\sigma^2 = 1$ or 10, and consider $n = 20$, 60 and $p = 4$, 10. The prior information of interest is a single linear stochastic constraint $\beta_1 + \nu = r$, where β_1 is the first element of β. Therefore, the variance matrix Ψ is a scalar and we consider $\Psi = 1, 10$. We let $\tau = (\beta_1 - r)^2$ and the risks are evaluated as functions of τ. Further, we

Table 1

Relative risks of estimators as percentages of p, the risk of the OLS estimator scaled by σ^2

	τ	b_m	b^*_{ms}	b_{ms}
$n = 20$,	0.00	0.9004	0.3388	0.2855
$p = 10$,	0.10	0.9005	0.3383	0.2620
$\Psi = 1$,	0.25	0.9007	0.3380	0.2542
$\sigma^2 = 10$	0.70	0.9011	0.3373	0.2546
	1.60	0.9020	0.3364	0.2794
	2.20	0.9026	0.3360	0.2993
	5.50	0.9058	0.3336	0.3975
	259.7	11.5847	2.9615	10.9031
$n = 20$,	0.00	0.9018	0.3373	0.3053
$p = 10$,	0.10	0.9019	0.3369	0.3013
$\Psi = 10$,	0.25	0.9020	0.3366	0.2985
$\sigma^2 = 10$	0.70	0.9024	0.3361	0.2926
	1.60	0.9032	0.3353	0.2846
	2.20	0.9037	0.3349	0.2806
	5.50	0.9064	0.3331	0.2677
	259.7	11.1471	3.0027	6.7969
$n = 20$,	0.00	0.7466	0.6941	0.4676
$p = 4$,	0.10	0.7467	0.6975	0.4404
$\Psi = 10$,	0.25	0.7471	0.6983	0.4308
$\sigma^2 = 10$	0.70	0.7479	0.6880	0.4437
	1.60	0.7497	0.6488	0.3739
	2.20	0.7508	0.6267	0.3636
	5.50	0.7574	0.5598	0.3581
	259.7	5.1103	2.1079	4.4058
$n = 60$,	0.00	0.8988	0.2662	0.2262
$p = 10$,	0.10	0.8989	0.2659	0.1939
$\Psi = 1$,	0.25	0.8990	0.2657	0.1802
$\sigma^2 = 10$	0.70	0.8995	0.2651	0.1708
	1.60	0.9004	0.2641	0.1876
	2.20	0.9010	0.2634	0.2058
	5.50	0.9042	0.2603	0.3091
	259.7	11.437	2.0690	10.7904
$n = 20$,	0.00	0.9832	0.4762	0.4761
$p = 10$,	0.10	0.9835	0.4761	0.4755
$\Psi = 10$,	0.25	0.9838	0.4760	0.4751
$\sigma^2 = 1$	0.70	0.9844	0.4757	0.4744
	1.60	0.9855	0.4751	0.4735
	2.20	0.9862	0.4747	0.4731
	5.50	0.9898	0.4729	0.4712
	259.7	12.3872	4.7878	11.2128

choose the optimal value of k given in (3.13) for b_{ms} and $k = p - 2$ for the estimator b^*_{ms}. Each part of the experiment is based on 2500 replications.

A selection of results being representative of the general pattern is given in Table 1. For ease of comparison, the risks are scaled by σ^2 and are expressed as percentages of p, the OLS estimator's risk. At least for the experimental settings that we have considered, the MRE is uniformly dominated by b_{ms} and b^*_{ms}, but neither b_{ms} nor b^*_{ms} dominates each other in all parts of the parameter space. It is found that for small to moderate values of τ, b_{ms} is invariably better than b^*_{ms} on the basis of risk. In particular, if Ψ is small, the risk reduction from using b_{ms} over b^*_{ms} or the MRE can be substantial. On the other hand, b^*_{ms} can have smaller risk than b_{ms} when τ is large, though over this region the OLS estimator dominates both estimators. Other things being equal, increasing Ψ increases the region such that b_{ms} has the smallest risk, though the risk magnitude of b_{ms} also increases as Ψ increases. It is also found that increasing τ results in increasing risk values of the MRE, while for b_{ms} and b^*_{ms}, the risks decline with τ if τ is of small or moderate values, and then increase monotonically for large values of τ. Furthermore, increasing n reduces the risks of the MRE, b_{ms} and b^*_{ms}, *ceteris paribus*. In general, risk results involving $\sigma^2 = 1$ and $\sigma^2 = 10$ yield the same qualitative comparisons. For the case of $\sigma^2 = 1$, the agreement between the Monte-Carlo and analytical results is sufficiently close to offer some assurance about the validity of the asymptotic results derived in the previous section.

Acknowledgements

Research for this paper was supported in part by a strategic grant from the City University of Hong Kong. The authors are grateful to Viren Srivastava, Hikaru Hasegawa, the editor and two anonymous referees for their illuminating comments on an earlier version of this paper. The usual disclaimers apply.

Appendix: Derivation of results

In this section, we provide details of the derivation of equations (3.4) and (3.5). From (2.12) and using $\alpha = 2$ as in the rest of the paper, we can write,

$$b_{ms} = b_m - \frac{k}{n-p+2} f b_m \,, \tag{A.1}$$

where

$$f = \frac{(y - Xb_m)'\,(y - Xb_m) + s^2 (r - Rb_m)'\,\Psi^{-1}(r - Rb_m)}{b'_m X' X b_m + s^2 b'_m R' \Psi^{-1} R b_m} \tag{A.2}$$

From TOUTENBURG and SHALABH (1996), we have,

$$\begin{aligned} b_m &= \beta + \sigma (X'X)^{-1} X'u + \sigma^2 \left(\frac{u'Mu}{n-p+2} \right) (X'X)^{-1} R' \Psi^{-1} \nu + O_p\left(\sigma^3\right) \\ &= \beta + \sigma e_1 + \sigma^2 e_2 + O_p(\sigma^3) \end{aligned} \tag{A.3}$$

where $M = I - X(X'X)^{-1} X'$, $e_1 = (X'X)^{-1} X'u$ and $e_2 = \dfrac{u'Mu(X'X)^{-1} R'\Psi^{-1} v}{n-p+2}$. Therefore,

$$\begin{aligned} f &= \left(\frac{\sigma^2}{\beta' X' X \beta} \right) \left(\frac{u'Mu}{n-p+2} \right) \left[n - p + 2 + \nu' \,\Psi \nu - 2\sigma \nu' \Psi^{-1} R (X'X)^{-1} X'u \right] \\ &\quad \times \left[1 - 2\, \frac{\sigma \beta' X' u}{\beta' X' X \beta} - \frac{\sigma^2 (u'Mu)\, (\beta' R' \Psi^{-1} R \beta)}{(n-p+2)\, (\beta' X' X \beta)} + 4\, \frac{(\sigma \beta' X' u)^2}{(\beta' X' X \beta)^2} \right] \end{aligned} \tag{A.4}$$

As a result, we obtain,

$$\begin{aligned} fb_m &= f[(b_m - \beta) + \beta] \\ &= \left(\frac{\sigma^2}{\beta' X' X \beta} \right) \left(\frac{u'Mu}{n-p+2} \right) \Big[(n - p + 2 + v' \Psi^{-1} \nu)\, \beta \\ &\quad + \sigma (n - p + 2 + \nu' \Psi^{-1} \nu)\, (X'X)^{-1} X'u \\ &\quad - \frac{2\sigma (n - p + 2 + \nu' \Psi^{-1} \nu)\, \beta' X' u \beta}{\beta' X' X \beta} \\ &\quad - 2\sigma \nu' \Psi^{-1} R (X'X)^{-1} X' u \beta \Big] \quad + O_p(\sigma^4)\,. \end{aligned} \tag{A.5}$$

Making use of this result in (A.1), we observe that, to order $O(\sigma^3)$,

$$b_{ms} = b_m + \sigma^2 l_1 + \sigma^3 l_2\,, \tag{A.6}$$

where

$$l_1 = - \frac{k}{\beta' X' X \beta}\, \frac{(n - p + 2 + \nu' \Psi^{-1} \nu)\, (u'Mu)\, \beta}{(n-p+2)^2} \tag{A.7}$$

and

$$\begin{aligned} l_2 &= \Big\{ 2\nu' \Psi^{-1} R (X'X)^{-1} X' u \beta - (n - p + 2 + \nu' \Psi^{-1} \nu) \\ &\quad \times \left[(X'X)^{-1} - 2 \frac{\beta \beta'}{\beta' X' X \beta} \right] X'u \Big\}\, \frac{k u' M u}{\beta' X' X \beta (n-p+2)^2} \end{aligned} \tag{A.8}$$

Hence, up to order $O(\sigma^2)$, the bias vector of b_{ms} can be written as,

$$B(b_{ms}) = E(b_m - \beta) - \sigma^2\, \frac{kE(u'Mu)\, [n - p + 2 + E(\nu' \Psi^{-1} v)]}{(n-p+2)^2\, \beta' X' X \beta}\, \beta\,, \tag{A.9}$$

by virtue of stochastic independence of u and ν. Now, it is readily shown that $E(u'Mu) = n - p$ and $E(\nu' \Psi^{-1} \nu) = G$. Furthermore, $E(b_m - \beta) = 0$, as $b_m - \beta$ is an odd function of disturbances [see KAKWANI (1968)]. Hence,

$$B(b_{ms}) = -\sigma^2 \left(\frac{(n-p)\,(n-p+2+G)}{(n-p+2)^2}\right)\left(\frac{k}{\beta' X' X \beta}\right)\beta\,, \qquad \text{(A.10)}$$

which is equation (3.4).

Next, consider the MSE matrix of b_{ms} to order $O(\sigma^4)$. Using (A.1), it is readily shown that the MSE matrix of b_{ms}, up to order $O(\sigma^4)$, can be written as,

$$E[(b_{ms}-\beta)(b_{ms}-\beta)'] = E[(b_m-\beta)(b_m-\beta)'] + \sigma^3 E\big(e_1 l_1' + l_1 e_1'\big) + \sigma^4 E\big(e_2 l_1' + e_1 l_2' + l_1 e_2' + l_1' l_1 + l_2 e_1'\big)\,. \qquad \text{(A.11)}$$

Note that,

$$E(e_1 l_1') = E(l_1 e_1') = 0\,, \qquad \text{(A.12)}$$

$$E(e_2 l_1') = (l_1 e_2') = 0\,, \qquad \text{(A.13)}$$

$$E(e_1 l_2') = -\frac{k(n-p+2+G)}{\beta' X' X \beta (n-p+2)^2}\, E\left[u' M u (X'X)^{-1} X' u u' X \times \left((X'X)^{-1} - \frac{2\beta\beta'}{\beta' X' X \beta}\right)\right]$$

$$= -\frac{k(n-p)\,(n-p+2+G)}{\beta' X' X \beta (n-p+2)^2}\left((X'X)^{-1} - \frac{2\beta\beta'}{\beta' X' X \beta}\right), \qquad \text{(A.14)}$$

$$E(l_2 e_1') = E((e_1 l_2')') = E(e_1 l_2') \qquad \text{(A.15)}$$

and

$$E(l_1 l_1') = \frac{k^2 (n-p)}{(\beta' X' X \beta)^2 (n-p+2)^3} \times E\Big[(n-p+2)^2 + (\nu' \Psi^{-1} \nu)^2 + 2(n-p+2)\,(\nu' \Psi^{-1} \nu)\Big] \times \beta\beta' = \frac{k^2 (n-p)}{(\beta' X' X \beta)^2\,(n-p+2)^3}\,[(n-p+2+G)^2 + 2G]\,\beta\beta'\,. \qquad \text{(A.16)}$$

Substituting (A.12) – (A.16) in (A.11), we obtain, up to order $O(\sigma^4)$,

$$D(b_m; b_{ms}) = E(b_m-\beta)\,(b_m-\beta)' - E(b_{ms}-\beta)\,(b_{ms}-\beta)' = 2\sigma^4 \left(\frac{(n-p)\,(n-p+2+G)}{(n-p+2)^2}\right)\left(\frac{k}{\beta' X' X \beta}\right)\left[(X'X)^{-1} - \frac{h}{2\beta X' X \beta}\,\beta\beta'\right], \qquad \text{(A.17)}$$

which is equation (3.5).

Finally, the expressions for the risk functions of b_{ms} and b^*_{ms} can be straightforwardly obtained by multiplying their respective MSE matrices by the weight matrix of the loss function and then taking the traces of the resulting matrices.

References

Guilkey, D. K. and Price, J. M., 1981: On comparing restricted least squares estimators. *Journal of Econometrics* **15**, 397–404.

James, W. and Stein, C., 1961: Estimation with quadratic loss. *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability.* University of California Press, Berkeley, 361–379.

Kakwani, N. C., 1968: A note on the unbiasedness of mixed regression estimation. *Econometrica* **36**, 610–611.

Nagar, A. L. and Kakwani, N. C., 1964: The bias and moment matrix of a mixed regression estimator. *Econometrica* **32**, 389–402.

Ohtani, K. and Honda, Y., 1984a: Small sample properties of the mixed regression estimator. *Journal of Econometrics* **26**, 375–385.

Ohtani, K. and Honda, Y., 1984b: On small sample properties of the mixed regression predictor under mis-specification. *Communications in Statistics: Theory and Methods* **13**, 2817–2825.

Srivastava, V. K., 1980: Estimation of linear single equation and simultaneous equation models under stochastic linear constraints: an annotated bibliography. *International Statistical Review* **48**, 79–82.

Srivastava, V. K. and Giles, D. E. A., 1987: *Seemingly Unrelated Regression Equations Models: Estimation and Inference*. Marcel Dekker, New York.

Srivastava, V. K. and Ohtani, K., 1995: A comparison of interval constrained least squares and mixed regression estimators. *Communications in Statistics: Theory and Methods* **24**, 395–413.

Srivastava, V. K., Raj, B. and Kumar, K., 1984: Estimation of a random coefficient model under linear stochastic constraints. *Annals of the Institute of Statistical Mathematics* **36**, 395–401.

Srivastava, V. K. and Srivastava, A. K., 1983: Improved estimation of coefficients in regression models with incomplete prior information. *Biometrical Journal* **25**, 775–782.

Stein, C., 1956: Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability.* University of California Press, Berkeley, 197–206.

Theil, H. and Goldberger, A. S., 1961: On pure and mixed statistical estimation in economics. *International Economic Review* **2**, 65–78.

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

Toutenburg, H. and Shalabh, 2000: Improved predictions in linear regression models with stochastic linear constraints. *Biometrical Journal* **42**, 71–86.

Yancey, T. A., Judge, G. G. and Bock, M. E., 1974: A mean square error test when stochastic restrictions are used in regression. *Communications in Statistics: Theory and Methods* **3**, 755–768.

A. T. K. Wan
Department of Management Sciences
City University of Hong Kong
Tat Chee Avenue
Kowloon
Hong Kong

Received, May 1999
Revised, December 1999
Accepted, December 1999