Restricted regression estimation in measurement error models

Shalabh*, Gaurav Garg, Neeraj Misra

Department of Mathematics and Statistics, Indian Institute of Technology, Kanpur 208016, India

Available online 18 May 2007

Abstract

The problem of consistent estimation of the regression coefficients when some prior information about the regression coefficients is available is considered. Such prior information is expressed in the form of exact linear restrictions. The knowledge of covariance matrix of measurement errors that is associated with explanatory variables is used to construct the consistent estimators. Some consistent estimators are suggested which satisfy the exact linear restrictions also. Their asymptotic properties are derived and analytically analyzed under a multivariate ultrastructural model with not necessarily normally distributed measurement errors. The finite sample properties of the estimators are studied through a Monte-Carlo simulation experiment.

© 2007 Elsevier B.V. All rights reserved.

Keywords: Measurement errors; Exact linear restriction; Ultrastructural model; Loëwner ordering; Non-normal distribution

1. Introduction

The observations in any statistical analysis are assumed to be recorded without any measurement errors. However, this assumption is not satisfied in many applications and the observations are recorded with measurement errors. The presence of measurement errors in the observations disturb the optimal properties of estimators, which are meant for those situations where observations are recorded free of measurement error. In the context of linear regression analysis, the ordinary least squares estimator (OLSE), which is the best linear unbiased estimator of regression parameter becomes biased as well as inconsistent in the presence of measurement errors. Some additional information is required to estimate the parameters consistently in a measurement error model (see, e.g., Cheng and Van Ness, 1999; Fuller, 1987, for more details).

Sometimes, some prior information about the regression coefficients is available from outside sample sources like as from past experience, long association of the experimenter with the experiment, or from some similar kind of studies done in the past, etc. Use of such information may improve upon the efficiency of the estimator (see, e.g., Toutenburg, 1982; Rao and Toutenburg, 1999, for more details). When such information is available in the form of exact linear restrictions, the restricted least squares estimator (RLSE) is used which is more efficient than OLSE in the absence of measurement errors. In the presence of measurement errors, the same RLSE becomes inconsistent and biased estimator of the regression coefficients. So the problem is to search for the estimators which are not only consistent but also satisfy the given exact linear restrictions on regression coefficients in the presence of measurement errors in the data. An attempt is made in this paper to find such estimators using the knowledge of covariance matrix of measurement

* Corresponding author. Tel.: +91 512 2597905; fax: +91 512 2597500.
E-mail addresses: shalab@iitk.ac.in, shalabh1@yahoo.com (Shalabh).
errors associated with explanatory variables and prior information together. An iterative procedure for obtaining the estimators under restrictions using total least squares approach is discussed in Schaffrin (2006).

In general, the measurement errors are assumed to follow a normal distribution. When there is departure from normality, then the statistical inferences derived under the assumption of normal distribution become incorrect and invalid (see, e.g., Srivastava and Shalabh 1997a; b; Shalabh et al., 2004, for more details). What is the effect of departure from normality of measurement errors on the efficiency properties of the estimators is another question addressed in this paper. Only the existence and finiteness of the first four moments of distribution of measurement errors and random error component are assumed without assuming any distributional form. The asymptotic properties of such estimators of regression coefficients which are not only consistent, but also satisfy the given linear restrictions are derived and studied under the specification of multivariate ultrastructural model when measurement errors are not necessarily normally distributed. The multivariate ultrastuctural model is a very general form of multivariate measurement error model which encompasses its structural and functional variants as well as the classical regression model as its particular cases.

The plan of this paper is as follows. The multivariate ultrastuctural model with not necessarily normally distributed measurement errors and random error component is described in Section 2. Utilizing the additional information on the covariance matrix of measurement errors associated with explanatory variables, four consistent estimators are proposed which satisfy the given linear restrictions also. In Section 3, the asymptotic distribution and the asymptotic properties of the estimators are presented and studied. The analysis of the finite sample properties of the estimators through a Monte Carlo simulation experiment are presented in Section 4 followed by the derivation of the results Appendix.

2. The model and the estimators

Assume that a \((n \times 1)\) vector of study variables \(\eta\) and a \((n \times p)\) matrix of explanatory variables \(T = (\xi_{ij})\) are related by

\[
\eta = T\beta, \tag{1}
\]

where \(\beta\) is a \((p \times 1)\) vector of unknown regression coefficients. Suppose that the variables \(\eta\) and \(T\) are unobservable and can only be observed with a \((n \times 1)\) vector of measurement errors \(e = (e_1, e_2, \ldots, e_n)'\) and a \((n \times p)\) matrix of measurement errors \(\Delta = (\delta_{ij})\) as

\[
y = \eta + e \tag{2}
\]

and

\[
X = T + \Delta, \tag{3}
\]

where \(y\) and \(X\) are the observed values of \(\eta\) and \(T\), respectively.

Further, assume that \(\xi_{ij}\) are randomly distributed with mean \(\mu_{ij}\) and random disturbance \(\phi_{ij}\), \(i = 1, 2, \ldots, n; j = 1, 2, \ldots, p.\) Let \(M = (\mu_{ij})\) and \(\Phi = (\phi_{ij})\) be the \((n \times p)\) matrices so that we can express

\[
T = M + \Phi. \tag{4}
\]

This completes the specification of an ultrastructural model, see Dolby (1976). When \(\xi_{ij}\)’s are identically and independently distributed with same mean, say, \(\mu\), the model reduces to the structural form of the measurement error model. When \(\xi_{ij}\)’s are fixed in nature, i.e., \(\phi_{ij} = 0\) (or equivalently variance of \(\phi_{ij}\) is zero), the model reduces to the functional form of the measurement error model. When \(\delta_{ij} = 0\) (or equivalently variance of \(\delta_{ij}\) is zero), the model reduces to the classical linear regression model which is free from any contamination of measurement errors.

Suppose, the prior information about the regression coefficients is available in the form of \(J (< p)\) exact linear restrictions which can be expressed as

\[
r = R\beta, \tag{5}
\]

where \(r\) is a \((J \times 1)\) known vector and \(R\) is a \((J \times p)\) known full row rank matrix. We assume that \(J < p\). If \(J = p\), then all the regression coefficients can be obtained from the restrictions only and is an uninteresting case. Such prior information arises from outside sample (and not from the sample), e.g., from similar kind of experiments conducted in
the past, past experience or long association of experimenter with the experiment, similar type of studies, etc. A good example of restricted regression is the regression in the well known Cobb–Douglas production function in economics which has a condition of constant returns to scale. Such a condition means that the sum of regression coefficients is unity. Also in analysis of variance, we have a reparametrization condition like \( \sum_i \beta_i = c \) (some constant), so we have \( r = c, R = (1, \ldots, 1) \) and \( J = 1 \) in such a case.

It is assumed that the elements of \( \Delta \) are independently and identically distributed with first four finite moments given by \( 0, \sigma^2, \gamma_3, \gamma_4 \) respectively. Here, \( \gamma_3 \) and \( \gamma_4 \) denote the coefficients of skewness and kurtosis, respectively, of the distribution of \( \delta_{ij} \). The elements of \( \Phi \) are assumed to be independently and identically distributed with mean 0, variance \( \sigma^2 \) and finite third and fourth moments. The elements of \( \epsilon \) are assumed to be independently and identically distributed with mean 0 and variance \( \sigma^2 \). Further, \( \epsilon, \Delta \) and \( \Phi \) are also assumed to be statistically independent.

We assume that \( \lim_{n \to \infty} (1/n)M'M =: \Sigma_\mu \) exists and is non-singular and \( \lim_{n \to \infty} (1/n)M'e_n =: \sigma^2 \) exists and is finite, where \( e_n \) is a \((n \times 1)\) vector of elements unity. This assumption is needed to avoid the presence of any trend in the observations (see, Schneeweiss, 1976, 1991) and for the application of large sample asymptotic approximation theory.

In the classical regression model without measurement errors, the OLSE of \( \beta \) is

\[
b = S^{-1}X'y,
\]

where \( S = X'X \) and the RLSE of \( \beta \) under the restrictions (5) is

\[
b_R = b + S^{-1}R'(RS^{-1}R')^{-1}(r - Rb).
\]

It follows from Lemma 1 in Appendix that, as \( n \to \infty \),

\[
\text{plim} b = (I_p - \sigma^2 \Sigma^{-1})\beta,
\]

and

\[
\text{plim} b_R = [I_p - \sigma^2 \Sigma^{-1}(I_p - R'(R\Sigma^{-1}R')^{-1}R\Sigma^{-1})]\beta,
\]

where \( \Sigma = \Sigma_\mu + \sigma^2 I_p + \sigma^2 I_p \).

Thus both the estimators \( b \) and \( b_R \) are inconsistent for \( \beta \) under (1)–(4). Moreover, \( b \) does not satisfy the given restrictions (5), while \( b_R \) satisfies the restrictions. Under the assumption of normality of random errors in a linear regression model without measurement errors, \( b_R \) is the maximum likelihood estimator of \( \beta \) under the restrictions.

We assume that the common variance \( \sigma^2 \) of measurement errors associated with the explanatory variables is known. Under such an assumption, a consistent estimator of \( \beta \) is

\[
b^{(1)}_\delta = (S - n\sigma^2 I_p)^{-1}X'y,
\]

provided \((S - n\sigma^2 I_p)\) is a positive definite matrix. If this condition is violated, the estimator falls on the boundary of the parameter space (see, Fuller, 1987; Shalabh, 2003). Although, the estimator \( b^{(1)}_\delta \) is consistent for \( \beta \), it does not satisfy the given linear restrictions (5). It is also the maximum likelihood estimator of \( \beta \) under the assumption of normality of measurement errors and known \( \sigma^2 \).

Our main interest is to find an estimator which is consistent as well as satisfies the linear restrictions and clearly any of the estimators \( b_R \) and \( b^{(1)}_\delta \) does not satisfy both of these properties simultaneously. Moreover, it is well known that the inclusion of prior information in the form of exact restrictions in a classical linear regression model leads to more efficient estimators under Loewner ordering of their covariance matrices (see, e.g., Rao and Toutenburg, 1999, p. 139). This motivates to find improved estimators of regression coefficients under the measurement error model by incorporating the prior information. It may be noted that it is not possible to obtain the maximum likelihood estimators of regression coefficients in our set up by directly maximizing the likelihood function under measurement error models with restrictions because we are not assuming any distributional form of the measurement errors. So we attempt to derive some estimators using other alternative approaches.

Assuming that \((S - n\sigma^2 I_p)\) is a positive definite matrix, we consider an approach based on the philosophy of least squares (or similar to maximum likelihood under the normally distributed measurement errors) using the weighted sum
of squares due to errors. We note that the naive least squares cost function is

\[ Q = \| y - X\beta \|^2 \]

and the corrected cost function is

\[ Q_c = \| y - T\beta \|^2 - n\sigma^2 \| \beta \|^2 \]

with the property that \( E(Q_c | y, T) = Q \). Here \( \| \cdot \| \) is a vector norm defined as the square root of the sum of squares of elements of the vector. So we propose to minimize \( Q_c \) subject to the restrictions \( r = R\beta \) using the Lagrangian multiplier method. This yields the following estimator of \( \beta \):

\[ b_{\delta}^{(m)} = b_{\delta}^{(1)} + (S - n\sigma^2 I_p)^{-1} R' R_{S\delta}^{-1} (r - R b_{\delta}^{(1)}), \]

where \( R_{S\delta} = R(S - n\sigma^2 I_p)^{-1} R' \). Clearly, \( plim b_{\delta}^{(m)} = \beta \) and obviously \( R b_{\delta}^{(m)} = r \), i.e., \( b_{\delta}^{(m)} \) is consistent as well as satisfies the linear restrictions (5). Interestingly enough, minimizing \( Q_c \) is equivalent to minimizing the weighted sum of squares due to error \( (b_{\delta}^{(1)} - \beta)'(S - n\sigma^2 I_p)(b_{\delta}^{(1)} - \beta) \) subject to the restrictions \( R\beta = r \) with weight matrix \( (S - n\sigma^2 I_p) \) under the least squares approach which yields the same estimator as \( b_{\delta}^{(m)} \). Such an estimator can be termed as restricted least squares-like (or restricted maximum likelihood-like under normal distribution of measurement errors) estimator.

Another natural approach to derive an estimator of \( \beta \) is to follow the procedure of obtaining the feasible version of an estimator. We notice that the estimator \( Rb_{\delta} \) satisfies the linear restrictions (5) but is inconsistent for \( \beta \). This inconsistency arises due to the involvement of OLSE \( b \) of \( \beta \) which is inconsistent for \( \beta \) in the presence of measurement errors. Thus in order to obtain an estimator of \( \beta \) that is consistent as well as satisfies the given linear restrictions (5), we propose to replace the inconsistent estimator \( b \) in \( Rb_{\delta} \) by a consistent estimator \( b_{\delta}^{(1)} \). This yields the following estimator:

\[ b_{\delta}^{(2)} = b_{\delta}^{(1)} + S^{-1} R' R_{S\delta}^{-1} (r - R b_{\delta}^{(1)}), \]

where \( R_{S\delta} = RS^{-1} R' \). Clearly, \( plim b_{\delta}^{(2)} = \beta \) and \( R b_{\delta}^{(2)} = r \), i.e., \( b_{\delta}^{(2)} \) is consistent as well as satisfy the linear restrictions (5). Interestingly enough, we observe that estimator (10) can also be obtained by minimizing the weighted sum of squares due to error \( (b_{\delta}^{(1)} - \beta)'S(b_{\delta}^{(1)} - \beta) \) subject to the restrictions \( R\beta = r \) with weight matrix \( S \).

For the next approach, we observe that the estimator \( Rb_{\delta} \) satisfy the linear restrictions but it is inconsistent for \( \beta \), i.e., \( plim b_{\delta} \neq \beta \). So it suggests that \( Rb_{\delta} \) can be transformed into a consistent estimator of \( \beta \) satisfying the linear restrictions (5) by adjusting its probability in limit. So following such an approach, we derive another consistent estimator of \( \beta \) by using \( plim((1/n)S) \) and adjusting the \( plim(b_{R\delta}) \) for its inconsistency as

\[ b_{\delta}^{(3)} = [I_p - n\sigma^2 (I_p - S^{-1} R' R_{S\delta}^{-1} R) S^{-1}]^{-1} b_R. \]

Clearly, \( plim b_{\delta}^{(3)} = \beta \). Let \( U = I_p - n\sigma^2 (I_p - S^{-1} R' R_{S\delta}^{-1} R) S^{-1} \), therefore \( b_{\delta}^{(3)} = U^{-1} b_R \). Note that \( RU = R \), which implies \( R = RU^{-1} \) and therefore \( R b_{\delta}^{(3)} = RU^{-1} b_R = Rb_{\delta} = r \). Hence the estimator \( b_{\delta}^{(3)} \) satisfies the linear restrictions (5) and is also a consistent estimator of \( \beta \).

The estimators \( b_{\delta}^{(m)} \) and \( b_{\delta}^{(2)} \) are derived by minimizing the weighted sum of squares due to error under the linear restrictions. Alternatively, the unweighted sum of squares due to errors can also be minimized under the linear restrictions. So finally, we propose to minimize \( (b_{\delta}^{(1)} - \beta)'(b_{\delta}^{(1)} - \beta) \) with respect to \( \beta \) subject to the linear restrictions \( R\beta = r \) using the Lagrangian multiplier’s method. This yields the following estimator of \( \beta \):

\[ b_{\delta}^{(4)} = b_{\delta}^{(1)} + R' (R R')^{-1} (r - R b_{\delta}^{(1)}). \]

Again, \( b_{\delta}^{(4)} \) is also consistent as well as satisfies the linear restrictions (5).

Now we have four different consistent estimators of \( \beta \) that satisfy the given linear restrictions (5) also and they are derived following different statistical philosophies.
3. Asymptotic properties of the estimators

The asymptotic distributions of the estimators $b_{\delta}^{(m)}$, $b_{\delta}^{(2)}$, $b_{\delta}^{(3)}$ and $b_{\delta}^{(4)}$ are obtained in this section. Some conditions are obtained under which one estimator dominates the other one in the sense of closeness to the parameter in stochastic order.

From (30), (34), (40) and (42) in Appendix (for detailed derivation see Appendix), we have

$$\sqrt{n}(b_{\delta}^{(l)} - \beta) = A_l h + \frac{1}{\sqrt{n}} Z_l, \quad l = m, 2, 3, 4,$$

where

$$A_m = (I_p - Q_T \Sigma_T) \Sigma_T^{-1},$$
$$A_2 = (I_p - Q \Sigma_X) \Sigma_T^{-1},$$
$$A_3 = (I_p - \sigma_0^2 (\Sigma_X^{-1} - Q))^{-1} (\Sigma_X^{-1} - Q),$$
$$A_4 = [I_p - R'(RR')^{-1} R] \Sigma_T^{-1},$$
$$Q_T = \Sigma_T^{-1} R'(R \Sigma_T^{-1} R')^{-1} R \Sigma_T^{-1},$$
$$Q = \Sigma_X^{-1} R'(R \Sigma_X^{-1} R')^{-1} R \Sigma_X^{-1},$$
$$\Sigma_X = \Sigma_T + \sigma_0^2 I_p,$$
$$\Sigma_T = \frac{1}{n} M'M + \sigma_0^2 I_p,$$
$$h = \frac{1}{\sqrt{n}} [(M'\delta + \Phi'\delta + \Delta'\delta) - (M'\Delta + \Phi'\Delta)\beta - (\Delta'\Delta - n\sigma_0^2 I_p)\beta]$$

and $Z_l$ are random vectors, such that $Z_l = O_p(1), l = m, 2, 3, 4$.

Let $A_l^* := \lim_{n \to \infty} A_l; \quad l = m, 2, 3, 4$. Then we find that

$$A_m^* = (I_p - Q_{T0} \Sigma_{T0}) \Sigma_{T0}^{-1},$$
$$A_2^* = (I_p - Q_0 \Sigma) \Sigma_{T0}^{-1},$$
$$A_3^* = (I_p - \sigma_0^2 (\Sigma_{T0}^{-1} - Q_0))^{-1} (\Sigma_{T0}^{-1} - Q_0),$$
$$A_4^* = [I_p - P_R] \Sigma_{T0}^{-1},$$

where

$$\Sigma_{T0} = (\Sigma + \sigma_0^2 I_p),$$
$$Q_{T0} = \Sigma_{T0}^{-1} R'(R \Sigma_{T0}^{-1} R')^{-1} R \Sigma_{T0}^{-1},$$
$$Q_0 = \Sigma_{T0}^{-1} R'(R \Sigma_{T0}^{-1} R')^{-1} R \Sigma_{T0}^{-1},$$
$$P_R = R'(RR')^{-1} R,$$

and both the matrices, viz. $(I_p - P_R)$ and $P_R$ are idempotent.

The following theorem gives the asymptotic distribution of $b_{\delta}^{(m)}$, $b_{\delta}^{(2)}$, $b_{\delta}^{(3)}$ and $b_{\delta}^{(4)}$. The proof of the theorem is given in Appendix.
Theorem 1. The asymptotic distribution of $\sqrt{n}(b_\delta^{(l)} - \beta)$ is normal with mean vector 0 and covariance matrix $A_\delta^T \Omega A_\delta$, $l = m, 2, 3, 4$, where
\[
\Omega = \sigma^2 \Sigma + \sigma^2 G (\text{tr} \beta \beta') \Sigma + \sigma^2 \beta \beta' + N,
\] (19)
and
\[
N = \gamma_1 \sigma^3 (f(\sigma_{12} \beta_{12}'), \beta_{12}') + (f(\sigma_{12} \beta_{12}'))^T + \gamma_2 \sigma^4 f(I_p, \beta \beta'),
\]
the function $f : \mathbb{R}^{p \times p} \times \mathbb{R}^{p \times p} \rightarrow \mathbb{R}^{p \times p}$ is defined as
\[
f(Z_1, Z_2) = Z_1(Z_2 * I_p), \quad Z_1, Z_2 \in \mathbb{R}^{p \times p},
\] (20)
where $*$ denotes the Hadamard product operator of matrices and $\mathbb{R}^{p \times p}$ is the collection of all $(p \times p)$ real matrices.

In (19), $N$ represents the contribution of non-normality of the distribution of measurement error $\delta_{ij}$. It is to be noted that the covariance matrices of the asymptotic distributions of $\sqrt{n}(b_\delta^{(l)} - \beta)$, $l = m, 2, 3, 4$ are affected by the skewness and kurtosis of the distribution of $\delta_{ij}$'s only and not by the non-normality effects of any of the distributions of $e_{ij}$'s and $\phi_{ij}$'s. It is clear from $N$ that the assumption of normal distribution for measurement errors may lead to an under or over reporting of the asymptotic covariance matrix of $b^{(m)}$, $b^{(2)}$, $b^{(3)}$ and $b^{(4)}$ when measurement errors do not necessarily follow the normal distribution. The magnitude of the departure from normality depends on the values of coefficients of skewness and kurtosis. Since the mean of asymptotic distribution of $\sqrt{n}(b_\delta^{(l)} - \beta)$, $l = m, 2, 3, 4$, is zero, it follows that all the estimators $b^{(m)}$, $b^{(2)}$, $b^{(3)}$ and $b^{(4)}$ are asymptotically unbiased.

Now we compare the performance of different estimators under the criterion of Loewner ordering. It may be noted that an estimator $b^{(l)}$ is better than other estimator $b^{(k)}$ ($l \neq k = m, 2, 3, 4$), at least asymptotically, in Loewner ordering when the difference $A_\delta^T \Omega A_\delta - A^{(k)}_\delta \Omega A^{(k)}_\delta$ is a positive definite matrix. Since $b^{(m)}$ has more appeal, so we propose to compare the performance of $b^{(m)}$ with $b^{(2)}$, $b^{(3)}$ and $b^{(4)}$. We use the notation here that for the two square matrices $G_1$ and $G_2$, $G_1 \leq_L G_2$ means $G_1 - G_2$ is a negative definite matrix and $G_1 \geq_L G_2$ means $G_1 - G_2$ is a positive definite matrix under the criterion of Loewner ordering.

The estimator $b^{(m)}$ is better than $b^{(2)}$ under the criterion of Loewner ordering when
\[
Q_{T0} \Psi Q_{T0} - Q_0 \Psi Q_0 \leq_L (Q_{T0} - Q_0) \Psi \Sigma^{-1} + \Sigma^{-1} \Psi (Q_{T0} - Q_0),
\] (21)
where $\Psi = \Sigma \Psi_0 \Sigma$ and $\Psi_0 = \Sigma Q_0 \Sigma^{-1}$. The reverse holds true, i.e., $b^{(2)}$ is better than $b^{(m)}$ when (21) holds true with a reverse inequality sign. Both $b^{(m)}$ and $b^{(2)}$ are equally efficient if $Q_{T0} = Q_0$ which holds true if $\Sigma_{T0} = \Sigma_0$, or if $\sigma^2 \delta = 0$. But such a condition cannot hold true in our set up and is an uninteresting case.

Similarly, $b^{(m)}$ is better than $b^{(3)}$ under the criterion of Loewner ordering when
\[
\Sigma^{-1} Q_{T0} \Sigma^{-1} (Q_{T0} \Psi Q_{T0}) - \Sigma^{-1} (Q_{T0} \Sigma^{-1} Q_0 \Sigma Q_0 \Sigma) \Sigma^{-1}
\leq_L (\Psi \Sigma^{-1} Q_{T0} + Q_{T0} \Psi \Sigma^{-1}) - (\Sigma^{-1} Q_0 \Sigma Q_0 \Sigma) \Sigma^{-1}
\]
(22)
where $\Sigma_0 = (\Sigma + \sigma^2 \delta \Sigma Q_0)$. The reverse holds true, i.e., $b^{(3)}$ is better than $b^{(m)}$ when (22) holds true with a reverse inequality sign.

Also, $b^{(m)}$ is better than $b^{(4)}$ under the criterion of Loewner ordering when
\[
Q_{T0} \Psi Q_{T0} - P_R \Psi_0 P_R \leq_L (\Psi \Sigma^{-1} Q_{T0} + Q_{T0} \Psi \Sigma^{-1}) - (\Psi_0 P_R + P_R \Psi_0).
\] (23)
The reverse holds true, i.e., $b^{(4)}$ is better than $b^{(m)}$ when (23) holds true with a reverse inequality sign.

Another interesting question arises is how the prior information improves the estimators in the sense of their variability. First we note that the asymptotic distribution of $\sqrt{n}(b_\delta^{(l)} - \beta)$ is normal with mean vector 0 and covariance matrix $A_1 \Omega A_1^T$, where $A_1 = (\Sigma - \sigma^2 \delta I_p)^{-1}$. The use of prior information leads to gain in efficiency as long as the difference
We choose the following set up for simulation:

\[ A_l \Omega A_l' - A_l \Omega A_l' = m, 2, 3, 4 is a positive definite matrix. For illustration, we compare \( b_\delta^{(m)} \) and \( b_\delta^{(2)} \) with \( b_\delta^{(1)} \). We find that the use of prior information leads to gain in efficiency in the sense of Löwner ordering in \( b_\delta^{(m)} \) and \( b_\delta^{(2)} \) when

\[ Q_{T0} \Sigma \Psi_0 (\Sigma Q_{T0} - I_\rho) < \Psi_0 \Sigma Q_{T0} \]

and

\[ \Psi \Sigma^{-1} Q_0 + Q_0 \Sigma^{-1} \Psi < Q_0 \Psi Q_0, \]

respectively. Similar conditions for other estimators can also be derived.

We have tried to explore the dominance conditions in finite samples through a Monte-Carlo simulation which are reported in the next section.

The following theorem gives a sufficient condition for the dominance of the estimators \( b_\delta^{(m)}, b_\delta^{(2)}, b_\delta^{(3)} \) and \( b_\delta^{(4)} \) over each other. The proof of the theorem is given in Appendix.

**Theorem 2.** If \((A_l A_l' - A_k A_k')\) is a positive semi-definite matrix then for every \( \alpha > 0 \),

\[
\lim_{n \to \infty} P \{ \| \sqrt{n}(b_\delta^{(l)} - \beta) \| \leq \alpha \} \leq \lim_{n \to \infty} P \{ \| \sqrt{n}(b_\delta^{(k)} - \beta) \| \leq \alpha \},
\]

for \( l, k = m, 2, 3, 4 \) and \( l \neq k \), where \( A_l \)'s are given in (14) where \( \| \cdot \| \) is a vector norm defined as the square root of the sum of squares of elements of the vector.

4. Monte-Carlo simulation study

We have considered the large sample asymptotic approximation theory to study the behavior of the distributions of the estimators \( b_\delta^{(m)}, b_\delta^{(1)}, b_\delta^{(2)}, b_\delta^{(3)} \) and \( b_\delta^{(4)} \). The large sample theory gives the idea of the behavior of the distributions of the estimators in the central part of the distributions only. So in order to study the overall distribution of the estimators and in finite samples, we conducted a Monte-Carlo simulation using MATLAB. In order to study the effect of skewness and kurtosis of the distributions of measurement errors and random errors, we adopt the following distributions:

(i) normal distribution (having no skewness and no kurtosis),

(ii) Student’s \( t \) distribution (having kurtosis only) and

(iii) gamma distribution (having both skewness and kurtosis).

We choose the following set up for simulation:

\[
\begin{align*}
\beta &= \begin{pmatrix}
2.2 \\
1.1 \\
3.0 \\
4.2 \\
2.5 \\
\end{pmatrix}, \\
R &= \begin{pmatrix}
0.8 & 0.6 & 0.7 & 0.9 & 0.8 \\
0.2 & 0.7 & 0.4 & 0.7 & 0.8 \\
0.6 & 0.4 & 0.6 & 0.1 & 0.4 \\
0.5 & 0 & 0.8 & 0.9 & 0.4 \\
\end{pmatrix}, \\
\end{align*}
\]

and \( r = R \beta \).

We choose a matrix \( M \) of mean values. For our chosen \( M \),

\[
\frac{1}{n} M'M = \begin{pmatrix}
10.1044 & 6.5674 & 7.3198 & 5.2212 & 8.5015 \\
6.5674 & 8.9555 & 6.1732 & 5.4899 & 7.8445 \\
5.2212 & 4.4971 & 6.2650 & 5.8162 & \\
8.5015 & 7.8445 & 7.2151 & 5.8162 & 10.9155 \\
\end{pmatrix}
\]

when \( n = 25 \).
Empirical bias (EB) of the estimators when $\epsilon_i \sim \text{Normal}(0, \sigma_1^2)$, $\phi_{ij} \sim \text{Normal}(0, \sigma_2^2)$ and $d_{ij} \sim \text{Normal}(0, \sigma_3^2)$.

When $\sigma_1^2 = 0.4$, $\sigma_2^2 = 0.4$, $\sigma_3^2 = 0.4$

- $b_{\hat{\delta}}^{(m)}$ $b_{\hat{\delta}}^{(2)}$ $b_{\hat{\delta}}^{(3)}$ $b_{\hat{\delta}}^{(4)}$
- $0.0004$ $0.0008$ $0.0006$ $0.0015$ $-0.0007$ $-0.0001$ $-0.0002$ $0.0005$
- $-0.0034$ $-0.0071$ $-0.0059$ $-0.0134$ $0.0063$ $0.0011$ $0.0019$ $0.0042$
- $-0.0009$ $-0.0019$ $-0.0016$ $-0.0037$ $0.0017$ $0.0003$ $0.0005$ $0.0011$
- $-0.0014$ $-0.0030$ $-0.0025$ $-0.0056$ $0.0026$ $0.0005$ $0.0008$ $-0.0017$
- $0.0046$ $0.0096$ $0.0079$ $0.0182$ $-0.0085$ $-0.0015$ $-0.0026$ $0.0056$

When $\sigma_1^2 = 0.4$, $\sigma_2^2 = 0.4$, $\sigma_3^2 = 1.0$

- $b_{\hat{\delta}}^{(m)}$ $b_{\hat{\delta}}^{(2)}$ $b_{\hat{\delta}}^{(3)}$ $b_{\hat{\delta}}^{(4)}$
- $-0.0003$ $0.0014$ $0.0002$ $0.0026$ $0.0010$ $0.0011$ $0.0002$ $0.0023$
- $0.0030$ $-0.0081$ $-0.0017$ $-0.0191$ $-0.0089$ $-0.0077$ $-0.0015$ $-0.0179$
- $0.0008$ $-0.0021$ $-0.0005$ $-0.0071$ $-0.0024$ $-0.0019$ $-0.0004$ $-0.0063$
- $0.0013$ $-0.0027$ $-0.0007$ $-0.0093$ $-0.0037$ $-0.0021$ $-0.0007$ $-0.0085$
- $-0.0041$ $0.0096$ $0.0023$ $0.0034$ $0.0120$ $0.0082$ $0.0020$ $0.0290$

When $\sigma_1^2 = 1.0476$, $\sigma_2^2 = 1.0476$, $\sigma_3^2 = 1.0476$

- $b_{\hat{\delta}}^{(m)}$ $b_{\hat{\delta}}^{(2)}$ $b_{\hat{\delta}}^{(3)}$ $b_{\hat{\delta}}^{(4)}$
- $-0.0004$ $-0.0010$ $-0.0003$ $-0.0040$ $0.0002$ $0.0003$ $-0.0003$ $0.0015$
- $0.0039$ $0.0095$ $0.0026$ $0.0365$ $-0.0014$ $-0.0032$ $0.0013$ $-0.0142$
- $0.0011$ $0.0026$ $0.0007$ $0.0099$ $-0.0004$ $-0.0009$ $0.0005$ $-0.0009$
- $0.0016$ $0.0040$ $0.0011$ $0.0153$ $-0.0006$ $-0.0013$ $0.0007$ $-0.0059$
- $-0.0053$ $-0.0129$ $-0.0035$ $-0.0492$ $0.0019$ $0.0043$ $-0.0018$ $0.191$

The experiment is repeated 5000 times and the bias vectors and MSE matrices of the estimators are estimated empirically. Keeping in mind the length of the paper, only few results related with $b_{\hat{\delta}}^{(m)}$, $b_{\hat{\delta}}^{(2)}$, $b_{\hat{\delta}}^{(3)}$ and $b_{\hat{\delta}}^{(4)}$ are presented here in Tables 1–6.

First we analyze the behavior of empirical absolute bias of the estimators in finite samples with respect to sample sizes and different distributions of measurement errors from Tables 1–3. It is observed that the empirical absolute bias of $b_{\hat{\delta}}^{(m)}$, $b_{\hat{\delta}}^{(2)}$, $b_{\hat{\delta}}^{(3)}$ and $b_{\hat{\delta}}^{(4)}$ decreases as the sample size increases. The values of empirical absolute bias are smaller under the sample size $n = 45$ than under $n = 25$ which shows that the estimators under consideration are asymptotically unbiased even for this sample size. This property of the estimators is also evident from the asymptotic results. The estimator $b_{\hat{\delta}}^{(m)}$ and/or $b_{\hat{\delta}}^{(3)}$ have smaller empirical absolute bias than that of the estimators $b_{\hat{\delta}}^{(2)}$ and $b_{\hat{\delta}}^{(4)}$ have. There is no significant difference between the empirical absolute bias of the estimators $b_{\hat{\delta}}^{(m)}$ and $b_{\hat{\delta}}^{(3)}$. The estimator $b_{\hat{\delta}}^{(4)}$ has the largest empirical absolute bias in both the small and large samples. The fluctuations in the empirical absolute bias of $b_{\hat{\delta}}^{(m)}$ and $b_{\hat{\delta}}^{(3)}$ are least in comparison to $b_{\hat{\delta}}^{(2)}$ and $b_{\hat{\delta}}^{(4)}$ when the sample size and values of various measurement error variances change. The empirical absolute bias of all the estimators increases when the measurement error variances are increased under all the distributions of the measurement errors under consideration. The magnitudes of empirical bias of all the three estimators are almost the same under normal and $t$ distributions. The difference arises when the degrees of freedom associated with the $t$-distribution (which changes the value of coefficient of kurtosis) are changed. On the other hand, the magnitudes of the empirical bias of all the three estimators are large under the gamma distributed measurement errors in comparison to the corresponding values under thenormally and $t$-distributed measurement errors.


The experiment is repeated 5000 times and the bias vectors and MSE matrices of the estimators are estimated empirically. Keeping in mind the length of the paper, only few results related with $b_{\hat{\delta}}^{(m)}$, $b_{\hat{\delta}}^{(2)}$, $b_{\hat{\delta}}^{(3)}$ and $b_{\hat{\delta}}^{(4)}$ are presented here in Tables 1–6.
Comparison of the empirical absolute bias of different estimators under the $t$ and gamma distributed measurement errors is higher than the corresponding absolute bias under the normally distributed measurement errors. Under the same value of measurement error variances, the absolute bias of the corresponding estimators under the normal, $t$ and gamma distributed measurement errors is significantly different in the small and large samples both. The effect of

**Table 2**

Empirical bias (EB) of the estimators when $e_i \sim t_{(12)}$, $\phi_{ij} \sim t_{(12)}$ and $\delta_{ij} \sim t_{(12)}$

<table>
<thead>
<tr>
<th>$n = 25$</th>
<th>$n = 45$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma^2_e = 0.4$, $\sigma^2_{\phi} = 0.4$, $\sigma^2_\delta = 0.4$</td>
<td>$\sigma^2_e = 0.4$, $\sigma^2_{\phi} = 0.4$, $\sigma^2_\delta = 1.0$</td>
</tr>
<tr>
<td>$-0.0002$</td>
<td>$0.0008$</td>
</tr>
<tr>
<td>$0.0019$</td>
<td>$-0.0078$</td>
</tr>
<tr>
<td>$0.0005$</td>
<td>$-0.0021$</td>
</tr>
<tr>
<td>$0.0008$</td>
<td>$-0.0033$</td>
</tr>
<tr>
<td>$-0.0026$</td>
<td>$0.0105$</td>
</tr>
</tbody>
</table>

**Table 3**

Empirical bias (EB) of the estimators when $e_i \sim \text{Gamma}(5, \sqrt{\sigma^2_e}/5)$, $\phi_{ij} \sim \text{Gamma}(5, \sqrt{\sigma^2_{\phi}}/5)$ and $\delta_{ij} \sim \text{Gamma}(5, \sqrt{\sigma^2_\delta}/5)$

<table>
<thead>
<tr>
<th>$n = 25$</th>
<th>$n = 45$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma^2_e = 0.4$, $\sigma^2_{\phi} = 0.4$, $\sigma^2_\delta = 0.4$</td>
<td>$\sigma^2_e = 0.4$, $\sigma^2_{\phi} = 0.4$, $\sigma^2_\delta = 1.0$</td>
</tr>
<tr>
<td>$-0.0007$</td>
<td>$0.0100$</td>
</tr>
<tr>
<td>$0.0068$</td>
<td>$-0.0088$</td>
</tr>
<tr>
<td>$0.0019$</td>
<td>$-0.0024$</td>
</tr>
<tr>
<td>$0.0029$</td>
<td>$-0.0037$</td>
</tr>
<tr>
<td>$-0.0092$</td>
<td>$0.0118$</td>
</tr>
</tbody>
</table>

$\sigma^2_e = 0.4$, $\sigma^2_{\phi} = 0.4$, $\sigma^2_\delta = 1.0$
departure from the normal distribution is clearly evident on the empirical absolute bias of the estimators $b^{(m)}$, $b^{(2)}$, $b^{(3)}$ and $b^{(4)}$.

Next, we analyze the empirical mean squared error matrices (EM) of these estimators from Tables 4–6. Observing the diagonal elements of the EM, it is noticed that in case of all distributions under consideration, viz., normal, $t$, and gamma, the variability of all the estimators $b^{(m)}$, $b^{(2)}$, $b^{(3)}$ and $b^{(4)}$ decreases as the sample size increases. This
Table 5
Empirical mean squared error matrices (EM) of the estimators when $e_i \sim t_{12}$, $\phi_{ij} \sim t_{12}$ and $\delta_{ij} \sim t_{12}$

<table>
<thead>
<tr>
<th>$b^{(m)}_\delta$</th>
<th>$b^{(1)}_\delta$</th>
<th>$b^{(2)}_\delta$</th>
<th>$b^{(3)}_\delta$</th>
<th>$b^{(4)}_\delta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma^2 = 0.4$</td>
<td>$\sigma^2 = 0.4$</td>
<td>$\sigma^2 = 0.4$</td>
<td>$\sigma^2 = 0.4$</td>
<td>$\sigma^2 = 0.4$</td>
</tr>
<tr>
<td>0.0012</td>
<td>-0.0111</td>
<td>-0.0030</td>
<td>-0.0046</td>
<td>0.0149</td>
</tr>
<tr>
<td>0.0111</td>
<td>0.1015</td>
<td>0.0277</td>
<td>0.0424</td>
<td>-0.1370</td>
</tr>
<tr>
<td>-0.0030</td>
<td>0.0277</td>
<td>0.0075</td>
<td>0.0116</td>
<td>-0.0374</td>
</tr>
<tr>
<td>-0.0046</td>
<td>0.0424</td>
<td>0.0116</td>
<td>0.0177</td>
<td>-0.0573</td>
</tr>
<tr>
<td>0.0149</td>
<td>-0.1370</td>
<td>-0.0374</td>
<td>-0.0573</td>
<td>0.1850</td>
</tr>
<tr>
<td>$\sigma^2 = 0.4$</td>
<td>$\sigma^2 = 0.4$</td>
<td>$\sigma^2 = 0.4$</td>
<td>$\sigma^2 = 0.4$</td>
<td>$\sigma^2 = 0.4$</td>
</tr>
<tr>
<td>0.0013</td>
<td>-0.0115</td>
<td>-0.0031</td>
<td>-0.0048</td>
<td>0.0156</td>
</tr>
<tr>
<td>-0.0115</td>
<td>0.1057</td>
<td>0.0288</td>
<td>0.0442</td>
<td>-0.1427</td>
</tr>
<tr>
<td>-0.0031</td>
<td>0.0288</td>
<td>0.0079</td>
<td>0.0121</td>
<td>-0.0389</td>
</tr>
<tr>
<td>-0.0048</td>
<td>0.0442</td>
<td>0.0121</td>
<td>0.0185</td>
<td>-0.0597</td>
</tr>
<tr>
<td>0.0156</td>
<td>-0.1427</td>
<td>-0.0389</td>
<td>-0.0597</td>
<td>0.1927</td>
</tr>
</tbody>
</table>

establishes that the estimators are consistent and confirms the theoretical findings also. Among these estimators, $b^{(4)}_\delta$ has the largest variability, while $b^{(m)}_\delta$ and $b^{(3)}_\delta$ have the smaller variability than $b^{(2)}_\delta$ and $b^{(4)}_\delta$. There is no significant difference between the variabilities of the estimators $b^{(m)}_\delta$ and $b^{(3)}_\delta$. The eigenvalues of the difference in EMs of the
Table 6
Empirical mean squared error matrices (EM) of the estimators when \( e_i \sim \text{Gamma}(5, \sqrt{\sigma^2_i}/5) \), \( \phi_{ij} \sim \text{Gamma}(5, \sqrt{\sigma^2_{ij}}/5) \) and \( \delta_{ij} \sim \text{Gamma}(5, \sqrt{\sigma^2_{ij}}/5) \)

<table>
<thead>
<tr>
<th>Estimator</th>
<th>EM ( n = 25 )</th>
<th>EM ( n = 45 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \sigma^2_{b_{(m)}^0} = 0.4 )</td>
<td>( \sigma^2_{b_{(m)}^0} = 0.4 ), ( \sigma^2_{\delta} = 0.4 )</td>
<td>( \sigma^2_{b_{(m)}^0} = 0.4 ), ( \sigma^2_{\delta} = 0.4 )</td>
</tr>
<tr>
<td>( b_{(m)}^0 )</td>
<td>0.0006</td>
<td>0.0047</td>
</tr>
<tr>
<td></td>
<td>0.0054</td>
<td>0.0047</td>
</tr>
<tr>
<td></td>
<td>0.0155</td>
<td>0.0065</td>
</tr>
<tr>
<td></td>
<td>0.0022</td>
<td>0.0065</td>
</tr>
<tr>
<td></td>
<td>0.0073</td>
<td>0.0065</td>
</tr>
<tr>
<td>( b_{(m)}^{(2)} )</td>
<td>0.0012</td>
<td>0.0012</td>
</tr>
<tr>
<td></td>
<td>0.0112</td>
<td>0.0112</td>
</tr>
<tr>
<td></td>
<td>0.0031</td>
<td>0.0031</td>
</tr>
<tr>
<td></td>
<td>0.0047</td>
<td>0.0047</td>
</tr>
<tr>
<td></td>
<td>0.0151</td>
<td>0.0151</td>
</tr>
<tr>
<td>( b_{(m)}^{(3)} )</td>
<td>0.0012</td>
<td>0.0012</td>
</tr>
<tr>
<td></td>
<td>0.0111</td>
<td>0.0111</td>
</tr>
<tr>
<td></td>
<td>0.0030</td>
<td>0.0030</td>
</tr>
<tr>
<td></td>
<td>0.0046</td>
<td>0.0046</td>
</tr>
<tr>
<td></td>
<td>0.0149</td>
<td>0.0149</td>
</tr>
<tr>
<td>( b_{(m)}^{(4)} )</td>
<td>0.0015</td>
<td>0.0015</td>
</tr>
<tr>
<td></td>
<td>0.0142</td>
<td>0.0142</td>
</tr>
<tr>
<td></td>
<td>0.0039</td>
<td>0.0039</td>
</tr>
<tr>
<td></td>
<td>0.0059</td>
<td>0.0059</td>
</tr>
<tr>
<td></td>
<td>0.0191</td>
<td>0.0191</td>
</tr>
</tbody>
</table>

EM \( \sigma^2_{b_{(m)}^0} = 1.0476, \ \sigma^2_{\delta} = 1.0476, \ \sigma^2_{\delta} = 1.0476 \)

| \( b_{(m)}^{(2)} \) | 0.0025 | 0.0025 |
| | 0.0031 | 0.0031 |
| | 0.0063 | 0.0063 |
| | 0.0097 | 0.0097 |
| | 0.0312 | 0.0312 |
| \( b_{(m)}^{(3)} \) | 0.0076 | 0.0076 |
| | 0.0699 | 0.0699 |
| | 0.0191 | 0.0191 |
| | 0.0029 | 0.0029 |
| | 0.0044 | 0.0044 |
| \( b_{(m)}^{(4)} \) | 0.0034 | 0.0034 |
| | 0.0311 | 0.0311 |
| | 0.0885 | 0.0885 |
| | 0.0149 | 0.0149 |
| | 0.0097 | 0.0097 |
| \( b_{(m)}^{(5)} \) | 0.0338 | 0.0338 |
| | 0.0395 | 0.0395 |
| | 0.0444 | 0.0444 |
| | 0.1294 | 0.1294 |
| | 0.4179 | 0.4179 |

estimators \( b_{(m)}^{(0)}, b_{(m)}^{(2)}, b_{(m)}^{(3)} \) and \( b_{(m)}^{(4)} \) under all the distributional assumptions are calculated and it is found that

\[
\text{EM}(b_{(m)}^{(0)}) \approx \text{EM}(b_{(m)}^{(2)}) \leq \text{EM}(b_{(m)}^{(3)}) \leq \text{EM}(b_{(m)}^{(4)}),
\]
where \( \leq \) \_L implies Loëwner ordering. For the two square matrices \( G_1 \) and \( G_2 \), \( G_1 - G_2 \leq \) \_L \( 0 \) implies that \( G_1 - G_2 \) is negative semi-definite. Thus from (25), it is clear that there is no significant difference between the variability of the estimators \( b(m) \) and \( b(3) \) and both these estimators are not worse than \( b(2) \) and \( b(4) \) under the criterion of Loëwner ordering of empirical mean squared matrices, while \( b(2) \) is not worse than \( b(4) \) under the same criterion.

Under the normal distribution of measurement errors, when \( \sigma^2_1 \) and \( \sigma^2_\phi \) are kept fixed and only \( \sigma^2_{\varepsilon} \) is increased, then the variabilities of all the estimators increases. When \( \sigma^2_1 \) and \( \sigma^2_\phi \) are increased independently keeping the other two fixed, then there is no significant change in the variability of the estimators. It is also noticed that when any two variances out of the \( \sigma^2_1 \), \( \sigma^2_\varepsilon \) and \( \sigma^2_\phi \) are increased simultaneously keeping the remaining one fixed, then the variability of the estimators is increased only when \( \sigma^2_\phi \) is increased. The effect of other two measurement error variances seems to be not as high as of \( \sigma^2_\phi \). When all the three variances are increased simultaneously, the increments in the variability of the estimators are almost the same as if only \( \sigma^2_\phi \) is increased. So \( \sigma^2_\phi \) plays a dominant role in the EMs of the estimators. The same is evident from the asymptotic theory also.

Comparing the EMs of the estimators \( b(m) \), \( b(2) \), \( b(3) \) and \( b(4) \) under different choices of the distributions, it is observed that the difference among the variabilities is not significant when \( \sigma^2_1 \), \( \sigma^2_\varepsilon \) and \( \sigma^2_\phi \) are small. When \( \sigma^2_\phi \) increases, then the variability of the estimators under \( t \) distributed measurement errors increases whereas it decreases in case of gamma distributed measurement errors. On the other hand, when \( \sigma^2_\varepsilon \) increases, then the variability of the estimators decreases under the \( t \) as well as gamma distributed measurement errors. The rate of such decrement is higher under the \( t \) distributed measurement errors than under the corresponding gamma distributed measurement errors. This clearly shows the effect of the coefficients of skewness and kurtosis on the variability of these estimators.

When the EMs of the four estimators in normal distribution case are compared under the same sample size and same variance with the corresponding EMs in the gamma distribution, we find that the variability of \( b(4) \) decreases significantly when sample size is 25. This change is negligible when the sample size is large. The change in the variabilities of \( b(m) \), \( b(2) \) and \( b(3) \) is not significant.

Now we compare the EMs under the \( t \) and gamma distributed measurement errors. We observe that the difference between the variabilities of \( b(2) \) and \( b(4) \) in case of the \( t \) distribution is higher than under the gamma distribution of measurement errors. The difference in the variabilities of \( b(m) \) and \( b(3) \) under \( t \) and gamma distributed measurement errors is not significant. In large samples, the variability of \( b(4) \) is almost same under the \( t \) and gamma distributed measurement errors.

The difference in the EM of different estimators under the \( t \) and gamma distributed measurement errors with their corresponding values under normally distributed measurement errors is significant. The magnitude of such difference essentially depends on the direction and magnitude of the departure from symmetry and peakedness of the distribution of measurement errors. Such changes are also explained by the asymptotic theory. The effect of departure from normality is seen to be more prominent in small samples than in large samples.

We also compared the bias and variability of \( b(1) \) with \( b(m) \), \( b(2) \), \( b(3) \) and \( b(4) \) to explore whether the incorporation of additional information in the estimation procedure leads to improvement of estimators or not. The nature of bias of \( b(1) \) is similar to other estimators and there is not much difference in the magnitude of bias of \( b(1) \) relative to other estimators. It was observed in all parameter settings that \( b(m) \), \( b(2) \), \( b(3) \) and \( b(4) \) have smaller mean squared error than \( b(1) \). This clearly indicates that the use of prior information improves the consistent estimator \( b(1) \) in terms of its variability. In most of the parametric settings, the estimator \( b(m) \) and \( b(3) \) are found to have better performance than \( b(2) \) and \( b(4) \). It may be noted that the analytic dominance conditions depend on the unknown parameters, skewness and kurtosis of the distribution of measurement errors and random error component. So our simulation results and guidelines are also affected in a similar way.

Acknowledgment

The authors would like to thank the Guest Editor and the referees for their constructive comments to improve the paper.
Appendix

For a matrix $B$, let $(B)_{ij}$ denote the $(i, j)$th element of the matrix $B$. We use the following definition:

**Definition.** Let $\{A_n : n = 1, 2, \ldots\}$ be a sequence of random matrices and let $\{b_n : n = 1, 2, \ldots\}$ be a sequence of real numbers. As $n \to \infty$, we say that (i) $A_n = \text{Op}(b_n)$ ($A_n = \text{op}(b_n)$) if every element of the random matrix $A_n$ is $\text{Op}(b_n)$ ($\text{op}(b_n)$) and (ii) plim $A_n = A$ if plim $(A_n)_{ij} = (A)_{ij} \forall i, j$.

We present some results in Lemma 1 which are used in this paper. The proof of the lemma being obvious is omitted.

**Lemma 1.** As $n \to \infty$,

(i) \( (1/\sqrt{n})M'\Phi = \text{Op}(1), \quad (1/\sqrt{n})M'\Delta = \text{Op}(1), \quad (1/\sqrt{n})M'\varepsilon = \text{Op}(1), \)

(ii) \( (1/\sqrt{n})\Phi' = \text{Op}(1), \quad (1/\sqrt{n})\Delta' = \text{Op}(1), \quad (1/\sqrt{n})\varepsilon' = \text{Op}(1), \)

(iii) \( (1/\sqrt{n})\Delta' \cdot \sqrt{n}\sigma^2_{\phi} I_p = \text{Op}(1), \quad (1/\sqrt{n})\Phi' \cdot \sqrt{n}\sigma^2_{\phi} I_p = \text{Op}(1), \)

(iv) \( \text{plim} \left( (1/n)\Delta' \right) = \sigma^2_{\phi} I_p, \quad \text{plim} \left( (1/n)\Phi' \right) = \sigma^2_{\phi} I_p, \)

(v) \( \text{plim} \left( (1/n)\Delta' \right) = \text{plim} \left( (1/n)\Phi' \right) = \text{plim} \left( (1/n)\varepsilon' \right) = \text{plim} \left( (1/n)\varepsilon' \right) = 0, \)

(vi) \( \text{plim} \left( (1/n)S \right) = \Sigma, \quad \text{plim} \left( (1/n)X'y \right) = (\Sigma - \sigma^2_{\phi} I_p)\beta, \) where $S = X'X, \Sigma = \Sigma + \sigma^2_{\phi} I_p + \sigma^2_{\phi} I_p$.

Using (1)–(4), it can be verified that

\[
S = n\Sigma_X + \sqrt{n}H,
\]

\[
X'(\varepsilon - \Delta\beta) = \sqrt{n}h - n\sigma^2_{\phi} \beta,
\]

\[
S - n\sigma^2_{\phi} I_p = n\Sigma_T + \sqrt{n}H
\]

and

\[
X'y = (n\Sigma_T + \sqrt{n}H)\beta + \sqrt{n}h,
\]

where

\[
\Sigma_X := \frac{1}{n}M'M + \sigma^2_{\phi} I_p + \sigma^2_{\phi} I_p,
\]

\[
\Sigma_T := \frac{1}{n}M'M + \sigma^2_{\phi} I_p,
\]

\[
H := \frac{1}{\sqrt{n}}(M'\Phi + M'\Delta + \Phi' M + \Phi' + \Delta' M + \Delta' \Phi) + \left( \frac{1}{\sqrt{n}}\Phi' - \sqrt{n}\sigma^2_{\phi} \right) + \left( \frac{1}{\sqrt{n}}\Delta' \Delta - \sqrt{n}\sigma^2_{\phi} \right)
\]

and

\[
h := \frac{1}{\sqrt{n}}(M'\varepsilon + \Phi' \varepsilon + \Delta' \varepsilon) - \frac{1}{\sqrt{n}}(M'\Delta + \Phi' \Delta)\beta - \frac{1}{\sqrt{n}}(\Delta' \Delta - n\sigma^2_{\phi} I_p)\beta.
\]

Using Lemma 1, it follows that $\Sigma_X = \text{O}(1), \Sigma_T = \text{O}(1), H = \text{Op}(1)$ and $h = \text{Op}(1)$.

Following lemma is useful in proving various results of Section 3. For the proof, see Rao and Rao (1998).

**Lemma 2.** Let $C = (c_{ij})$ be a $m \times m$ matrix and let $\|C\|_1 = \max_{1 \leq i \leq m} \sum_{j=1}^{m} |c_{ij}|$ and $\|C\|_2 = \max_{1 \leq j \leq m} \sum_{i=1}^{m} |c_{ij}|$ be the maximum column sum and maximum row sum matrix norms, respectively. If $\|C\|_1 < 1$ and/or $\|C\|_2 < 1$, ...
then \((I_m - C)\) is invertible and
\[(I_m - C)^{-1} = \sum_{i=1}^{\infty} C^i,\]
where \(C^0 = I_m\).

**Proof of (13).** First consider from (8) and using Lemma 2, the estimation error of \(b_{\delta}^{(1)}\) is
\[
b_{\delta}^{(1)} - \beta = (S - n\sigma_\delta^2)^{-1} X'y - \beta
\]
\[
= n^{-1/2}(I_p + n^{-1/2}\Sigma_{T}^{-1} H)^{-1}\Sigma_{T}^{-1} h
\]
\[
= n^{-1/2}(I_p - n^{-1/2}\Sigma_{T}^{-1} H)\Sigma_{T}^{-1} h + O_p(n^{-3/2}).
\]
(26)

Now the estimation error of the estimator of \(b_{\delta}^{(m)}\) can be expressed as
\[
b_{\delta}^{(m)} - \beta = (b_{\delta}^{(1)} - \beta) + (S - n\sigma_\delta^2 I_p)^{-1} R'(R(S - n\sigma_\delta^2 I_p)^{-1} R')^{-1}(r - Rb_{\delta})\]
From the definition of \(H\) and since \(\text{plim} (n^{-1/2}\Sigma_{T}^{-1} H) = 0\), using Lemma 2, we have
\[
(S - n\sigma_\delta^2 I_p)^{-1} = (n\Sigma_T + \sqrt{n}H)^{-1}
\]
\[
= \frac{1}{n}(I_p - n^{-1/2}\Sigma_{T}^{-1} H)\Sigma_{T}^{-1} + O_p(n^{-2})
\]
and
\[
\frac{1}{n}[R(S - n\sigma_\delta^2 I_p)^{-1} R']^{-1} = [R((I_p - n^{-1/2}\Sigma_{T}^{-1} H)\Sigma_{T}^{-1} + O_p(n^{-1}))R']^{-1}
\]
\[
= (I_p + n^{-1/2}R_{T}^{-1} R_{T}^{-1} H\Sigma_{T}^{-1} R')R_{T}^{-1} + O_p(n^{-1}),
\]
(29)

where \(R_{T} = R\Sigma_{T}^{-1} R'\). Thus using (27)–(29), we have
\[
\sqrt{n}(b_{\delta}^{(m)} - \beta) = A_m h + \frac{1}{\sqrt{n}}Z_m,
\]
(30)
where \(A_m := (I_p - Q_T\Sigma_T)\Sigma_{T}^{-1}, \ Q_T := \Sigma_{T}^{-1} R'(R\Sigma_{T}^{-1} R')^{-1} R\Sigma_{T}^{-1}\) and \(Z_m = O_p(1)\).

We have from (10)
\[
b_{\delta}^{(2)} - \beta = (b_{\delta}^{(1)} - \beta) + S^{-1} R'(RS^{-1} R')^{-1}(r - Rb_{\delta}^{(1)}).
\]
(31)

Since \(\Sigma_{X}^{-1} = O(1)\) and \(\text{plim} (n^{-1/2}\Sigma_{X}^{-1} H) = 0\), using Lemma 2, we have
\[
S^{-1} = (n\Sigma_X + \sqrt{n}H)^{-1}
\]
\[
= \frac{1}{n}(I_p - n^{-1/2}\Sigma_{X}^{-1} H)\Sigma_{X}^{-1} + O_p(n^{-2}),
\]
(32)

for sufficiently large \(n\). Let \(R_{X} = R\Sigma_{X}^{-1} R'\) so that \(R_{X}^{-1} = O(1)\). Moreover, since \(\text{plim}(n^{-1/2}R_{X}^{-1} R\Sigma_{X}^{-1} H\Sigma_{X}^{-1} R') = 0\), so we have from (32),
\[
\frac{1}{n}[R(S^{-1} R')^{-1} = [R((I_p - n^{-1/2}\Sigma_{X}^{-1} H)\Sigma_{X}^{-1} + O_p(n^{-1}))R']^{-1}
\]
\[
= (I_p + n^{-1/2}R_{X}^{-1} R\Sigma_{X}^{-1} H\Sigma_{X}^{-1} R')R_{X}^{-1} + O_p(n^{-1}).
\]
(33)
Using (31)–(33), we have
\[ b_\delta^{(2)} - \beta = n^{-1/2}(I_p - Q\Sigma_X)(I_p - n^{-1/2}(\Sigma_X^{-1} H - \Sigma_X^{-1} Q\Sigma_X))\Sigma_X^{-1} h + O_p(n^{-3/2}). \]

Thus
\[ \sqrt{n}(b_\delta^{(2)} - \beta) = A_2 h + \frac{1}{\sqrt{n}} Z_2, \]

(34)

where \( A_2 := (I_p - Q\Sigma_X)\Sigma_X^{-1} \), \( Q := \Sigma_X^{-1} R'(R^{-1} R')^{-1} R\Sigma_X^{-1} \) and \( Z_2 = O_p(1) \).

From (11), we have
\[ b_\delta^{(3)} = [I_p - n\sigma_\delta^2 A\Sigma^{-1}]^{-1} b_R, \]

(35)

where \( A := I_p - S^{-1} R' R_S^{-1} R \). Also,
\[ r - Rb = -RS^{-1}X' (\epsilon - \Delta \beta), \]
\[ b = S^{-1}X'(\epsilon - \Delta \beta) + \beta, \]

and therefore
\[ b_R = \beta + AS^{-1}X'(\epsilon - \Delta \beta) \]
\[ = (I_p - n\sigma_\delta^2 A\Sigma^{-1}) \beta + \sqrt{n} AS^{-1} h. \]

(36)

From (35) and (36), we have
\[ b_\delta^{(3)} = [I_p - n\sigma_\delta^2 A\Sigma^{-1}]^{-1}[(I_p - n\sigma_\delta^2 A\Sigma^{-1}) \beta + \sqrt{n} AS^{-1} h], \]
\[ b_\delta^{(3)} - \beta = \frac{1}{\sqrt{n}} [I_p - \sigma_\delta^2 A(nS^{-1})^{-1} A(nS^{-1})h]. \]

(37)

Using (32) and (33), we have
\[ A = I_p - [(I_p - n^{-1/2}\Sigma_X^{-1} H)\Sigma_X^{-1} + O_p(n^{-1})]R' \]
\[ \cdot [(I_p + n^{-1/2}\Sigma_X^{-1} H\Sigma_X^{-1} R')R_X^{-1} + O_p(n^{-1})]R \]
\[ = (I_p - Q\Sigma_X) \left[ I_p + \frac{1}{\sqrt{n}} \Sigma_X^{-1} H Q\Sigma_X \right] + O_p(n^{-1}). \]

Thus,
\[ A(nS^{-1}) = (\Sigma_X^{-1} - Q) - n^{-1/2}(\Sigma_X^{-1} - Q)H(\Sigma_X^{-1} - Q) + O_p(n^{-1}). \]

(38)

Since \( Q = O(1) \) and \( \text{plim}[n^{-1/2}\sigma_\delta^2 (\Sigma_X^{-1} - Q)H(\Sigma_X^{-1} - Q)] = 0 \), we have
\[ [I_p - \sigma_\delta^2 A(nS^{-1})]^{-1} = [I_p + n^{-1/2}\sigma_\delta^2 (I_p - \sigma_\delta^2 (\Sigma_X^{-1} - Q))]^{-1} \]
\[ \cdot ((\Sigma_X^{-1} - Q)H(\Sigma_X^{-1} - Q) + O_p(n^{-1/2}))^{-1} \]
\[ \cdot (I_p - \sigma_\delta^2 (\Sigma_X^{-1} - Q))^{-1} \]
\[ = [I_p + O_p(n^{-1/2})](I_p - \sigma_\delta^2 (\Sigma_X^{-1} - Q))^{-1} \]
\[ = [I_p - \sigma_\delta^2 (\Sigma_X^{-1} - Q)]^{-1} + O_p(n^{-1/2}). \]

(39)
Using (37)–(39), we have

\[ b_\delta^{(3)} - \beta = \frac{1}{\sqrt{n}}[IP - \sigma_\delta^2(S_X^{-1} - Q)]^{-1} + O_p(n^{-1/2}) \]

\[ \cdot [(S_X^{-1} - Q) - n^{-1/2}(S_X^{-1} - Q)H(S_X^{-1} - Q) + O_p(n^{-1})]h \]

\[ = \frac{1}{\sqrt{n}}[IP - \sigma_\delta^2(S_X^{-1} - Q)]^{-1}(S_X^{-1} - Q)h + O_p(n^{-1}). \]

Thus

\[ \sqrt{n}(b_\delta^{(3)} - \beta) = A_3h + \frac{1}{\sqrt{n}}Z_3, \]

where \( A_3 := [IP - \sigma_\delta^2(S_X^{-1} - Q)]^{-1}(S_X^{-1} - Q) \) and \( Z_3 = O_p(1) \).

Next, the estimation error of \( b_\delta^{(4)} \) can be written as

\[ b_\delta^{(4)} - \beta = [IP - R'(RR')^{-1}R](b_\delta^{(1)} - \beta). \]

Using (26) we have

\[ b_\delta^{(4)} - \beta = \frac{1}{\sqrt{n}}[IP - R'(RR')^{-1}R](IP - n^{-1/2}S_T^{-1}H)S_T^{-1}h + O_p(n^{-3/2}). \]

Thus

\[ \sqrt{n}(b_\delta^{(4)} - \beta) = A_4h + \frac{1}{\sqrt{n}}Z_4, \]

where \( A_4 := [IP - R'(RR')^{-1}R]S_T^{-1} \) and \( Z_4 = O_p(1) \). □

**Proof of Theorem 1.** First consider,

\[ h = \frac{1}{\sqrt{n}}X'i - \frac{1}{\sqrt{n}}(X'A - n\sigma_\delta^2IP)\beta \]

\[ = (IP, -(IP \otimes \beta')) \sum_{i=1}^{n} \frac{1}{\sqrt{n}} \left( x_i'\epsilon_i \right) \]

\[ = \sum_{i=1}^{n} C_i \omega_i, \]

where \( \delta_i, x_i, m_i \) and \( \phi_i \) are \( i \)th \((1 \times p)\) row vectors of the matrices \( A, X, M \) and \( \Phi \), respectively \( (i = 1, 2, \ldots, n) \),

\[ C_i = \frac{1}{\sqrt{n}}((IP, -(IP \otimes \beta'))((IP + 1 \otimes m_i^i) \ IP_{p_2+p} \ IP_{p_2+p}) \]

are \( p \times (2p^2 + 3p + 1) \) non-stochastic matrices and

\[ \omega_i = (\epsilon_i \ \delta_i' \ \phi_i' \ \epsilon_i \ \delta_i \ \vec(\phi_i' \ \delta_i) \ \vec(\epsilon_i' \ \delta_i - \sigma_\delta^2IP))' \]

are \((2p^2 + 3p + 1) \times 1 \) independent and identically distributed random vectors, \( i = 1, 2, \ldots, n \). Since,

\[ E(h) = 0, \]

\[ E(hh') = \sigma_\delta^2S_X + \sigma_\delta^2(tr\beta\beta')S_X + \sigma_\delta^2\beta\beta' + \gamma_{1}^{\delta} \sigma_\delta^4 f(IP, \beta\beta') \]

\[ + \gamma_{1}^{\delta} \sigma_\delta^3 \left\{ f \left( \frac{1}{n}M'e_n' \rho, \beta\beta' \right) + \left( f \left( \frac{1}{n}M'e_n' \rho, \beta\beta' \right) \right)' \right\}, \]
\[
\lim_{n \to \infty} E(hh') = \sigma^2_e \Sigma + \sigma^2_e (tr \, \beta \beta') \Sigma + \sigma_0^2 \beta \beta' + \gamma_1 \sigma_0^2 \{ f(1_p, \beta \beta') \\
+ \gamma_2 \sigma_0^3 \{ f(\sigma_\mu e'_p, \beta \beta') + (f(\sigma_\mu e'_p, \beta \beta'))' \},
\]

therefore by central limit theorem, \( h \) has a limiting normal distribution with mean vector 0 and covariance matrix given by \( \Omega := \lim_{n \to \infty} E(hh') \).

From (30), (34), (40) and (42), since \( Z_l = O_p(1) \), it follows that \( (1/\sqrt{n}) Z_l \to^p 0, l = m, 2, 3, 4 \). Therefore, by Slutsky’s lemma asymptotic distribution of \( \sqrt{n}(b_\delta^{(l)} - \beta) \) is same as the asymptotic distribution of \( A_l h \). Since \( \lim_{n \to \infty} A_l = A_l^* \), where \( A_m^*, A_2^*, A_3^* \) and \( A_4^* \) are given in (15)–(18), asymptotic distribution of \( \sqrt{n}(b_\delta^{(l)} - \beta) \) is normal with mean vector 0 and covariance matrix \( A_l^* \Omega A_l'^*, l = m, 2, 3, 4 \).

**Proof of Theorem 2.** From Theorem 1, it is clear that \( \sqrt{n}(b_\delta^{(l)} - \beta) \) has an asymptotic distribution, which is same as the asymptotic distribution of \( A_l h, l = m, 2, 3, 4 \). Now, suppose \((A_l'A_l - A_k'A_k) \) \((l, k = 2, 3, 4, l \neq k)\) is a positive semi-definite matrix, therefore

\[
\|A_l h\| \geq \|A_k h\|,
\]

with probability one. Here \( \| \cdot \| \) is a vector norm defined as the square root of the sum of squares of elements of the vector. From (44),

\[
P(\|A_l h\| \leq z) \leq P(\|A_k h\| \leq z) \quad \text{for} \quad z \geq 0.
\]

Thus

\[
\lim_{n \to \infty} P(\|A_l h\| \leq z) \leq \lim_{n \to \infty} P(\|A_k h\| \leq z).
\]

It follows from (45),

\[
\lim_{n \to \infty} P(\|\sqrt{n}(b_\delta^{(l)} - \beta)\| \leq z) \leq \lim_{n \to \infty} P(\|\sqrt{n}(b_\delta^{(k)} - \beta)\| \leq z),
\]

which is the result stated in Theorem 2.

**References**


