SHALABH – HELGE TOUTENBURG – CHRISTIAN HEUMANN

Mean squared error matrix comparison of least squares and Stein-rule estimators for regression coefficients under non-normal disturbances

Summary - Choosing the performance criterion to be mean squared error matrix, we have compared the least squares and Stein-rule estimators for regression coefficients in a linear regression model when the disturbances are not necessarily normally distributed. It is shown that none of the two estimators dominates the other, except in the trivial case of merely one regression coefficient where least squares estimator is found to be superior in comparison to Stein-rule estimator.

Key Words - Linear regression model; Stein-rule estimator; Ordinary least squares estimator; Mean squared error matrix.

1. INTRODUCTION

Performance properties of the least squares and Stein-rule estimators for the coefficients in a linear regression model have been largely compared by considering various criteria like bias vector, scalar risk function under weighted and unweighted squared error loss functions, Pitman measures and concentration probability for normal as well as non-normal distributions such as multivariate $t$, Edgeworth-type, elliptically symmetric and spherically symmetric distributions. We do not propose to present a review of such a vast literature but we simply wish to note that the criterion of mean squared errors matrix for the comparison of estimators has received far less attention despite the fact that it is a strong criterion, see, e.g., Judge et al. (1985) and Rao et al. (2008). This paper reports our modest attempt to fill up this gap without assuming any specific distribution of the disturbances; all that we assume is the finiteness of moments up to order four.

The Stein-rule and least squares estimators have been studied and compared by Ullah et al. (1983) under the criterion of scalar risk function under unweighted squared error loss function, i.e., trace of mean squared error matrix which is a weaker criterion than mean squared error matrix. The results

Received April 2008 and revised July 2008.
derived by Ullah et al. (1983) can be derived as a particular case of the results presented in this paper. Moreover, the conclusions of Ullah et al. (1983) may not remain valid under the criterion of mean squared error matrix. Just for the sake of illustration, Ullah et al. (1983) derived the condition for the dominance of Stein-rule estimator over least squares estimator under symmetrical and mesokurtic distribution of errors which does not remain true under mean squared error matrix criterion. Our investigations have revealed that least squares and Stein-rule estimators may dominate each other, according to mean squared error matrix criterion, for asymmetrical disturbances but none dominates the other for symmetrical distributions with an exception to the trivial case of only one regression coefficient in the model where least squares estimator is found to be superior in comparison to Stein-rule estimator. Such differences in the statistical inferences provide the motivation to fill up such gaps and which is attempted in this paper. It may also be noted that Judge and Bock (1978) derived the bias vector and covariance matrix of Stein-rule estimator under the normally distributed disturbances. The assumption of normality may not remain valid in a variety of situations, see, e.g., Gnanadesikan (1977) and needs attention that what happens when assumption of normality is violated. Some work in this direction has been reported by Brandwein (1979) and Brandwein and Strawderman (1978, 1980) who considered a class of spherically symmetric distributions of disturbances and obtained the conditions for minimaxity of an improved family of estimators but they depend on the specification of distribution of disturbances. Also, the exact form of bias vector and covariance matrix of Stein-rule estimator is quite intricate to draw any clear conclusion and so most of the authors have studied the approximate bias vector and covariance matrix, e.g., by employing asymptotic approximation theory.

The organization of our presentation is as follows. Section 2 describes the model and presents an asymptotic approximation for the difference between the mean squared error matrices of least squares and Stein-rule estimators when disturbances are small, without assuming any functional form of the underlying distribution cannot be derived. Note that the covariance matrix and mean squared error matrix of least squares estimator are same due to its unbiasedness. The thus obtained expression is utilized for comparing the estimators for asymmetrical distributions in Section 3 and for symmetrical distributions in Section 4. A Monte-Carlo simulation experiment is conducted to study the finite sample properties of the estimators and its findings are reported in Section 5.

2. MEAN SQUARED ERROR MATRIX DIFFERENCE

Consider the following linear regression model

\[ y = X\beta + \sigma U \]  

(1)
Mean squared error matrix comparison

where $y$ is a $n \times 1$ vector of $n$ observations on the study variable, $X$ is a $n \times p$ full column rank matrix of $n$ observations on $p$ explanatory variables, $\beta$ is a $p \times 1$ vector of regression coefficients, $\sigma$ is an unknown positive scalar and $U$ is a $n \times 1$ vector of disturbances.

It is assumed that the elements of $U$ are independently and identically distributed following a distribution with first four moments as $0$, $1$, $\gamma_1$ and $(\gamma_2 + 3)$. Here $\gamma_1$ and $\gamma_2$ are the Pearson’s measures of skewness and kurtosis, respectively.

The least squares estimator of $\beta$ is given by

$$ b = (X'X)^{-1}X'y $$

which is the best estimator in the class of all linear and unbiased estimators.

If $H$ denotes the prediction matrix $X(X'X)^{-1}X'$ and $ar{H} = (I - H)$, the Stein-rule estimator of $\beta$ is given by

$$ \hat{\beta} = \left[ 1 - \frac{k}{n-p+2} \cdot \frac{y'\bar{H}y}{y'Hy} \right] b $$(3)

which essentially defines a class of non-linear and biased estimators characterized by a positive scalar $k$, see, e.g., Judge and Bock (1978) and Saleh (2006).

For comparing the estimators $b$ and $\hat{\beta}$ with respect to the criterion of mean squared error matrix, we do not assume any specific distribution of disturbances. All that we assume is that the first four moments of distribution of disturbances are finite.

It is easy to see that the difference in the mean squared error matrices of $b$ and $\hat{\beta}$ is

$$ \Delta = E(b - \beta)(b - \beta)' - E(\hat{\beta} - \beta)(\hat{\beta} - \beta)' $$

$$ = \left( \frac{k}{n-p+2} \right) E \left[ \frac{y'\bar{H}y}{y'Hy} \left\{ b(b - \beta)' + (b - \beta)b' \right\} \right] $$

$$ - \left( \frac{k}{n-p+2} \right)^2 E \left[ \left( \frac{y'\bar{H}y}{y'Hy} \right)^2 bb' \right]. $$

(4)

Under the specification of first four moments of the disturbances, it is not possible to evaluate the expectations in (4). We therefore derive an asymptotic approximation for $\Delta$ when disturbances are small, i.e., when $\sigma$ is small and tends towards zero; see, e.g., Ullah et al. (1983).
Using (1), we can express

\[
\frac{y' \tilde{H} y}{y' H y} = \sigma^2 \frac{U' \tilde{H} U}{\beta' X' \beta} \left[ 1 + 2 \sigma \frac{\beta' X' U}{\beta' X' \beta} + \sigma^2 \frac{U' H U}{\beta' X' \beta} \right]^{-1}
\]

\[
= \sigma^2 \frac{U' H U}{\beta' X' \beta} - 2 \sigma^3 \frac{U' H U \cdot \beta' X' U}{(\beta' X' \beta)^2} + O_p(\sigma^4).
\]

Hence we get

\[
\frac{y' \tilde{H} y}{y' H y} b (b - \beta)' = \sigma^3 \frac{U' \tilde{H} U}{\beta' X' \beta} \beta U' X (X' X)^{-1}
\]

\[
+ \sigma^4 \frac{U' H U}{\beta' X' \beta} \left[ (X' X)^{-1} - \frac{2 \beta' X' \beta}{\beta' X' \beta} \right] X' U \cdot U' X (X' X)^{-1}
\]

\[
+ O_p(\sigma^5)
\]

(5)

\[
\left( \frac{y' \tilde{H} y}{y' H y} \right)^2 b b' = \sigma^4 \left( \frac{U' \tilde{H} U}{\beta' X' \beta} \right)^2 \beta \beta' + O_p(\sigma^5).
\]

(6)

Next, we observe that

\[
E(U' \tilde{H} U \cdot U) = \gamma_1 d
\]

\[
E(U' \tilde{H} U \cdot U U') = \gamma_2 D + (n - p) I + 2 \tilde{H}
\]

where \(d\) denotes the column vector formed by the diagonal element of \(\tilde{H}\) and \(D\) is a diagonal matrix with diagonal elements same as that of \(\tilde{H}\).

Using these results along with (5) and (6) in (4), we obtain the following expression for \(\Delta\) up to order \(O(\sigma^4)\):

\[
\Delta = \sigma^3 \left( \frac{\gamma_1 k}{n - p + 2} \right) (\beta h' + h \beta') + \frac{\sigma^4 (n - p) k W}{(n - p + 2)(\beta' X' \beta)^2}
\]

(7)

where

\[
h = (X' X)^{-1} X' d
\]

\[
W = 2(\beta' X' \beta) \left[ (X' X)^{-1} + \left( \frac{\gamma_2}{n - p} \right) (X' X)^{-1} X' D X (X' X)^{-1} \right]
\]

\[
- 2 \left( \frac{\gamma_2}{n - p} \right) \left[ \beta' \beta X' D X (X' X)^{-1} + (X' X)^{-1} X' D X \beta \beta' \right]
\]

\[
- [4 + k (1 + g \gamma_2)] \beta \beta'
\]

(9)

with

\[
g = \frac{\text{tr} D \tilde{H}}{(n - p)(n - p + 2)}.
\]
3. COMPARISONS FOR ASYMMETRICAL DISTURBANCES

If the distribution of disturbances is not symmetrical so that $\gamma_1$ is different from 0, the leading term in (7) helps in comparing the asymptotic efficiency. We thus have

$$\Delta = \sigma^3 \left( \frac{\gamma_1 k}{n - p + 2} \right) (\beta' h' + h\beta')$$

up to order $O(\sigma^3)$ only.

It may be observed that the quantity $h$ is a column vector of covariances between the elements of least squares estimator $b$ and the residual sum of squares $(y - Xb)'(y - Xb)$. Further, the matrix $\beta h'$ has at most rank 1 with non-zero characteristic root as $\beta'h$.

The expression on the right hand side of (10) is positive semi-definite when $\gamma_1$ and $\beta'h(\neq 0)$ have same signs. This implies that Stein-rule estimator is not inferior to least squares estimator according to the strong criterion of mean squared error matrix, to the order of our approximation, when $\beta'h$ is positive for positively skewed distributions and is negative for negatively skewed distributions of disturbances. In other words, the Stein-rule estimator for all positive values of $k$ will asymptotically dominate the least squares estimator for those asymmetrical distributions for which skewness measure $\gamma_1$ and $\beta'h$ have identical signs.

The reverse is true, i.e., the Stein-rule estimator for all $k$ fails to dominate the least squares estimator at least asymptotically when $\gamma_1$ and $\beta'h$ have opposite signs and the underlying distribution of disturbances is asymmetrical.

4. COMPARISON OF SYMMETRICAL DISTRIBUTIONS

When the distribution is symmetric ($\gamma_1 = 0$) and for $\beta'h = 0$, it is observed from (10) that both the least squares and the Stein-rule estimators are equally good at least up to the order of our approximation. This means that we need to consider higher order approximations for $\Delta$ in order to make a choice between least squares and Stein-rule estimators.

We thus observe from (7) that the expression $\Delta$ under symmetry of the distribution of disturbances becomes

$$\Delta = \sigma^4 \frac{(n - p)k}{(n - p + 2)(\beta'X'X\beta)^2} W.$$  \hspace{1cm} (11)

Now we present some results which will be utilized for examining the definiteness of the matrix $W$.

**Lemma 1.** The matrices $X'DX(X'X)^{-1}$ and $[I - X'DX(X'X)^{-1}]$ are at least positive semi definite.
Proof. Observing from Chatterjee and Hadi (1988, Chap. 2) that the diagonal elements of any idempotent matrix lies between 0 and 1, it is obvious that the diagonal matrices $D$ and $(I - D)$ will be at least positive semi-definite. Consequently, the matrices $X'DX(X'X)^{-1}$ and $[I - X'DX(X'X)^{-1}]$ will also be at least positive semi-definite.

**Lemma 2.** The quantity $(1 + g\gamma_2)$ is positive.

Proof. As the diagonal elements of $D$ lie between 0 and 1, we have

$$0 \leq \text{tr}D\tilde{H} \leq \text{tr}\tilde{H} = (n - p)$$

whence we obtain

$$0 \leq g = \frac{\text{tr}D\tilde{H}}{(n - p)(n - p + 2)} \leq \frac{1}{n - p + 2}.$$  

Using it and noting that $(2 + \gamma_2)$ is always non-negative, we find $(1 + g\gamma_2)$ to be positive.

**Lemma 3.** For any $m \times 1$ vector $\phi$, a necessary and sufficient condition for matrix $(I - \phi\phi')$ to be positive definite is that $\phi\phi'$ is less than 1.

Proof. See Yancy, Judge and Bock (1974).

**Lemma 4.** For any $m \times 1$ vector $\phi$ and any $m \times m$ positive definite matrix $\Phi$, the matrix $(\phi\phi' - \Phi)$ cannot be non-negative definite for $m$ exceeding 1.


When the distribution of disturbances is mesokurtic ($\gamma_2 = 0$), the expression (11) reduces to the following:

$$W = 2(\beta'X'X\beta)(X'X)^{-1} - (4 + k)\beta\beta'$$

$$= 2(\beta'X'X\beta)(X'X)^{-1/2} \left[ I - \left( \frac{4 + k}{2\beta'X'X\beta} \right) (X'X)^{1/2} \beta\beta'(X'X)^{1/2} \right] (X'X)^{-1/2}$$  \hspace{1cm} (12)

which, applying Lemma 3, will be positive if and only if

$$\left( \frac{4 + k}{2} \right) < 1$$  \hspace{1cm} (13)

but it holds in no case for positive values of $k$. This implies that no Stein-rule estimator dominates the least squares estimator with respect to the criterion of mean squared error matrix, to the order of our approximation, for mesokurtic distributions of disturbances.
Similarly, let us examine the dominance of least squares estimator over the Stein-rule estimator, that is we need to examine the positive definiteness of the matrix \((-W)\). Now applying Lemma 4, it is easy to see that \((-W)\) is positive definite only in the trivial case of \(p = 1\). Thus for \(p > 1\), the least squares estimator does not dominates the Stein-rule estimator with respect to the mean squared error matrix criterion for mesokurtic distributions.

The above findings match the well-known result regarding the failure of least squares and Stein-rule estimators over each other under normality of disturbances when the performance criterion is mean squared error matrix rather than, for example, its trace or any other weak mean squared error criterion as in Ullah et al. (1983).

Next, let us investigate the nature of the matrix \(W\) for a leptokurtic distribution \((\gamma_2 > 0)\). Using Lemma 1, we observe that the matrix \(W\) is positive definite as long as the matrix expression

\[
2(\beta'X'X\beta) \left[ (X'X)^{-1} + \left( \frac{\gamma_2}{n-p} \right) (X'X)^{-1}X'DX(X'X)^{-1} \right]
\]

is positive definite.

Applying Lemma 3, it is seen that (14) is positive definite if and only if

\[
\frac{\beta' \left[ (X'X)^{-1} + \left( \frac{\gamma_2}{n-p} \right) (X'X)^{-1}X'DX(X'X)^{-1} \right]^{-1} \beta}{2\beta'X'X\beta} < 1.
\]

As the diagonal elements of diagonal matrix lie between 0 and 1, we have

\[
1 \leq \frac{\beta' \left[ (X'X)^{-1} + \left( \frac{\gamma_2}{n-p} \right) (X'X)^{-1}X'DX(X'X)^{-1} \right]^{-1} \beta}{\beta'X'X\beta} \leq \left( 1 + \frac{\gamma_2}{n-p} \right)
\]

which, when used in (15), clearly reveals that the inequality (15) cannot hold true for positive \(k\). This implies that Stein-rule estimator cannot be superior to least squares estimator according to mean squared error matrix criterion when the distribution of disturbances is leptokurtic.

Similarly, for the dominance of least squares estimator over Stein-rule estimator, the matrix \((-W)\) should be positive definite. This will be the case
so long as the matrix
\[ [4 + k(1 + g\gamma_2)]\beta\beta' - 2(\beta'X'X\beta)(X'X)^{-1} + \left(\frac{\gamma_2}{n-p}\right)(X'X)^{-1}X'DX(X'X)^{-1} \]
\[(17)\]
is positive definite.

Applying Lemma 4, we observe that the matrix expression (17) cannot be positive definite for \( p > 1 \). This means that the least squares estimator is not superior to Stein-rule estimator, except the trivial case of \( p = 1 \), when the distribution of disturbances is leptokurtic.

Finally, consider the case of platykurtic distributions (\( \gamma_2 < 0 \)). In this case, \( W \) is positive definite if the matrix
\[ 2(\beta'X'X\beta) \left(1 + \frac{\gamma_2}{n-p}\right)(X'X)^{-1} - [4 + k(1 + g\gamma_2)]\beta\beta' \]
\[(18)\]
is positive definite.

Notice that the matrix expression (18) cannot be positive definite if \((n - p + \gamma_2)\) is negative or zero. On the other hand, if \((n - p + \gamma_2)\) is positive, an application of Lemma 2 and Lemma 3 suggests that the expression (18) is positive definite if and only if
\[ \left[ \frac{4 + k(1 + g\gamma_2)}{2 \left(1 + \frac{\gamma_2}{n-p}\right)} \right] < 1 \]
\[(19)\]
but it can never hold true for positive values of \( k \) as \( \gamma_2 \) is negative.

It is thus found that no Stein-rule estimator is superior to least squares estimator for platykurtic distributions of disturbances.

Finally, let us check whether the least squares estimator dominates the Stein-rule estimator in case of a platykurtic distribution. For this purpose, it suffices to examine the nature of following matrix expression.
\[ [4 + k(1 + g\gamma_2) + 2 \left(\frac{\gamma_2}{n-p}\right)]\beta\beta' - 2(\beta'X'X\beta)(X'X)^{-1} \]
\[(20)\]
as \( \gamma_2 \) is negative.

Using Lemma 4 and observing that \((2 + \gamma_2)\) is always positive, it follows that (20) cannot be positive definite except when \( p = 1 \). It means that the least squares estimator cannot be superior to the Stein-rule estimator provided the number of unknown coefficients in the model is more than one.
5. Monte-Carlo simulation

We conducted a Monte-Carlo simulation experiment to study the behavior of least squares estimator $b$ and Stein-rule estimator $\hat{\beta}$ in finite samples under the mean squared error matrix criterion. We considered two different sample sizes ($n = 40$ and $n = 100$) and following three different distributions of disturbances:

1. normal distribution (having no skewness and no kurtosis) with mean 0 and variance 1,
2. $t$ distribution (having no skewness but non-zero kurtosis) with 5 degrees of freedom ($t(5)$) and
3. beta distribution $Beta(1, 2)$ (having non-zero skewness and non-zero kurtosis).

All the random observations from $t$ and beta distributions are suitably scaled to have zero mean and unity variance. The difference in the results under normal, $t$ and beta distributions may be considered as arising due to the skewness and kurtosis of the distributions. Other parameter settings are $p = 6$; 21 different settings for $\beta$ as $\beta_1 = \beta_2 = \ldots = \beta_6 = c$ starting from $c = 0$ to $c = 0.1$ with an increment of 0.005; each row of $X$ was generated from a multivariate normal distribution with mean vector $(1, 1, 1, -1, -1, -1)'$ and identity covariance matrix $I$ (no collinearity in the columns of $X$) and $\epsilon$’s were generated under three different distributions of disturbances. The empirical mean squared error matrix of both the estimators was estimated based on these 5000 runs. After completion of 5000 runs, we calculated the eigenvalues of the difference of empirical mean squared error matrices of $b$ and $\hat{\beta}$. If all eigenvalues were positive, than we concluded that Stein-rule estimator was better than least squares estimator under the mean squared error matrix criterion in that setting.

We considered all together $6.3 \times 10^9$ different data sets in this set up. In fact, we got 10000 different comparisons of the estimated empirical mean squared error matrices of $b$ and $\hat{\beta}$ for each of the 21 different parameter settings of $\beta$ and for each of the six basic designs. We counted the number of cases where Stein-rule estimator was better than least squares estimator to comprehend the simulation output. The results are presented in Figure 1. We plotted a curve between the number of cases where Stein-rule estimator is better than least squares estimator and value of $\beta$ under each sample size. The difference in the plots corresponding to different sample sizes under different distributions of disturbances is not clearly depicted. So we have presented only one case under beta distribution of disturbances for illustration. Other plots under normal and $t$ distributed disturbances are similar to the plots as under beta distributed disturbances. It is seen that the performance of Stein-rule estimator compared
Error dist.: Beta (1,2), centered and scaled to variance 1

Error dist.: Beta (1,2), centered and scaled to variance 1

Figure 1. Behaviour of number of cases in which Stein-rule estimator is better than least squares estimator with respect to $\beta$ under beta distribution of disturbances and sample sizes.

to the least squares estimator depends on the sample size and $\beta$. The Stein-rule estimator appears to be better than least squares estimator only for small (absolute) values of $\beta$. When the sample size is low, the probability is higher that Stein-rule estimator is better than least squares estimator for larger values of $\beta$ in comparison to the case when sample size is large. We also note that the difference in the results among the different distributions of disturbances (when sample size is fixed) seems to be small but the number of cases of superiority of Stein-rule estimator over least squares estimator are different. For example, when $\beta = 0.1$ and $n = 40$, then Stein-rule estimator is better than least squares estimator in

- 59 cases for the normal distribution,
- 50 cases for the beta distribution but
- 91 cases for the $t(5)$ distribution,

out of 10000 simulations. Table 1 shows some selected outcomes for illustration.

We observe from Table 1 that the $t$ distribution has maximum and beta distribution has minimum number of cases for the superiority of Stein-rule estimator over least squares estimator. The rate of decrement of the number of cases of superiority of Stein-rule estimator over least squares estimator heavily depends on the sample size and the value of parameter vector $\beta$. When $\beta$ is small, say, $\leq 0.03$, then mostly Stein-rule estimator is found to be better than least squares estimator, irrespective of the sample size. The number of cases of superiority of Stein-rule estimator over least squares estimator are higher in smaller sample size ($n = 40$) than larger sample size ($n = 100$).
These findings indicate that there is no uniform dominance of Stein-rule estimator over least squares estimator and vice versa. Under the same experimental settings, Stein-rule estimator is not superior to least squares estimator under all the cases. This finding goes along with the results reported in Sections 3 and 4. Though our findings in Section 3 and 4 are based on small sigma approach but it is clear from the simulated results that they hold true in finite samples and for $\sigma = 1$. If we take any other choice of $\sigma$, then the number of cases of superiority of Stein-rule estimator over least squares estimator will change from the present values.

We are presenting below the mean squared error matrix of Stein-rule estimator and covariance matrix of least squares estimator for illustration when $\beta = 0.07$ and $n = 40$ under all the three distributions of disturbances to get an idea about the non-normality effect. If there was no effect of coefficients of skewness and kurtosis, then all the matrices are expected to be nearly same.

- **Normal distribution**

  Stein-rule estimator:

  $\begin{pmatrix}
  0.01969 & -0.005705 & 0.0003417 & 0.003382 & 0.004405 & 0.005621 \\
  -0.005705 & 0.01747 & 0.001881 & 0.004259 & 0.00371 & 0.004046 \\
  0.0003417 & 0.001881 & 0.01095 & 0.005669 & 0.003347 & 0.004802 \\
  0.003382 & 0.004259 & 0.005669 & 0.01263 & 0.003933 & 0.002561 \\
  0.004405 & 0.00371 & 0.003347 & 0.005669 & 0.003933 & 0.00104 \\
  0.005621 & 0.004046 & 0.004802 & 0.002561 & 0.00104 & 0.0121 
\end{pmatrix}$
Least squares estimator:

\[
\begin{bmatrix}
0.04615 & -0.02202 & -0.006222 & 0.001424 & 0.004545 & 0.00936 \\
-0.02202 & 0.03941 & -0.002366 & 0.004601 & 0.002678 & 0.003218 \\
0.006222 & -0.002366 & 0.02204 & 0.008096 & 0.001902 & 0.005629 \\
0.001424 & 0.004601 & 0.008096 & 0.02576 & 0.003422 & -0.0004868 \\
0.004545 & 0.002678 & 0.001902 & 0.003422 & 0.01965 & -0.004275 \\
0.00936 & 0.003218 & 0.005629 & -0.0004868 & -0.004275 & 0.02568
\end{bmatrix}
\]

**t distribution**

Stein-rule estimator:

\[
\begin{bmatrix}
0.01574 & 0.001746 & 0.00007353 & 0.005134 & 0.00259 & 0.00615 \\
0.001746 & 0.009077 & -0.0002147 & 0.001766 & 0.004082 & 0.004697 \\
0.00007353 & -0.0002147 & 0.01047 & 0.003891 & 0.005382 & 0.0004504 \\
0.005134 & 0.001766 & 0.003891 & 0.01395 & -0.0003272 & -0.001497 \\
0.00259 & 0.004082 & 0.005382 & -0.0003272 & 0.01223 & 0.0006263 \\
0.00615 & 0.004697 & 0.0004504 & -0.001497 & 0.0006263 & 0.0146
\end{bmatrix}
\]

Least squares estimator:

\[
\begin{bmatrix}
0.03415 & -0.002385 & -0.006143 & 0.007739 & -0.0009113 & 0.008991 \\
-0.002385 & 0.01694 & -0.006772 & -0.002186 & 0.00433 & 0.005431 \\
-0.006143 & -0.006772 & 0.02144 & 0.004556 & 0.007315 & -0.00511 \\
0.007739 & -0.002186 & 0.004556 & 0.03072 & -0.007534 & -0.009714 \\
-0.0009113 & 0.00433 & 0.007315 & -0.007534 & 0.02447 & -0.005077 \\
0.008991 & 0.005431 & -0.00511 & -0.009714 & -0.005077 & 0.03068
\end{bmatrix}
\]

**Beta distribution**

Stein-rule estimator:

\[
\begin{bmatrix}
0.01049 & 0.002311 & 0.00166 & 0.004018 & 0.00388 & 0.004543 \\
0.002311 & 0.01069 & 0.0008187 & 0.002575 & 0.004466 & 0.003389 \\
0.00166 & 0.0008187 & 0.01356 & 0.00439 & 0.004765 & 0.006381 \\
0.004018 & 0.002575 & 0.00439 & 0.009821 & 0.0009748 & 0.002586 \\
0.00388 & 0.004466 & 0.004765 & 0.0009748 & 0.008549 & 0.002869 \\
0.004543 & 0.003389 & 0.006381 & 0.002586 & 0.002869 & 0.009958
\end{bmatrix}
\]

Least squares estimator:

\[
\begin{bmatrix}
0.02101 & -0.0004467 & -0.002724 & 0.003978 & 0.003177 & 0.004978 \\
-0.0004467 & 0.02111 & -0.004928 & 0.0007727 & 0.00471 & 0.00227 \\
-0.002724 & -0.004928 & 0.02891 & 0.004476 & 0.005579 & 0.009623 \\
0.003978 & 0.0007727 & 0.004476 & 0.01921 & -0.004037 & 0.0003745 \\
0.003177 & 0.00471 & 0.005579 & -0.004037 & 0.01545 & 0.0003523 \\
0.004978 & 0.00227 & 0.009623 & 0.0003745 & 0.0003523 & 0.01947
\end{bmatrix}
\]

It is clear from these values that the covariance matrix of least squares estimator and mean squared error matrix of Stein-rule estimator are not same under
different distributions when all other parameters are same except the coefficient of skewness and kurtosis. In some cases, such difference is rather high. The simulated results and the results compiled in Table 1 clearly indicates the effect of non-normality of distribution of disturbances on the variability of these estimators and the superiority of Stein-rule estimator and least squares estimator over each other. It is difficult to give any clear guidelines for the user and to explain the effect of coefficients of skewness and kurtosis from the mean squared error and covariance matrices of Stein-rule estimator and least squares estimator respectively based on simulated results. In most of the cases, it is found that such effect follows the conditions reported in Section 3 and 4.

ACKNOWLEDGMENTS

The authors are grateful to the referee for providing illuminating comments to improve the paper. The first author gratefully acknowledges the support from Alexander von Humboldt Foundation, Germany for Humboldt Fellowship.

REFERENCES


