Shalabh, Toutenburg, Heumann:

Mean Squared Error Matrix comparison of Least Squares and Stein-Rule Estimators for Regression Coefficients under Non-normal Disturbances

Sonderforschungsbereich 386, Paper 496 (2006)

Mean Squared Error Matrix comparison of Least Squares and Stein-Rule Estimators for Regression Coefficients under Non-normal Disturbances

Shalabh [1] H. Toutenburg [2] C. Heumann [3]

Abstract

Choosing the performance criterion to be mean squared error matrix, we have compared the least squares and Stein-rule estimators for 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 is found to be superior in comparisons to Stein-rule estimators.

Key Words: Linear regression model, Stein rule estimator, ordinary least squares estimator, mean squared error matrix.

[1]Corresponding Author,

Department of Mathematics and Statistics, Indian Institute of Technology, Kanpur 208016, India. E-mail: shalab@iitk.ac.in, shalabh1@yahoo.com

Acknowledgement: The first author gratefully acknowledges the support from Alexander von Humboldt Foundation, Germany for Humboldt Fellowship.

[2]Department of Statistics, Ludwig–Maximilians–University Munich, Ludwigstr. 33, 80539 Munich, Germany. E-mail: helge.toutenburg@stat.uni-muenchen.de

[3]Department of Statistics, Ludwig–Maximilians–University Munich, Ludwigstr. 33, 80539 Munich, Germany. E-mail: chris@stat.uni-muenchen.de

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 unweighed 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 and Toutenburg (1999). 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 under four.

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. 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 and its findings are reported in Section 5.

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 square estimators is found to be superior in comparison to Stein-rule estimators.

2 Mean Squared Error Matrix Difference

Consider the following linear model

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

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, β is a $p \times 1$ vector of regression coefficients, σ 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 $0, 1, \gamma_1 \text{and} (\gamma_2 + 3)$. Here γ_1 and γ_2 are the Pearsons measures of skewness and kurtosis, respectively.

The least squares estimator of β is given by

$$b = (X'X)^{-1}X'y \tag{2.2}$$

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 $\bar{H} = (I - H)$, the Stein-rule estimators of β are given by

$$\hat{\beta} = \left[1 - \frac{k}{n-p+2} \cdot \frac{y'\bar{H}y}{y'Hy}\right] b \tag{2.3}$$

which essentially defines a class of non-linear and biased estimators characterized by a positive scalar k, see, e.g., Judge and Bock (1978), 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 first four moments are finite.

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

$$\begin{aligned}
\Delta &= \mathrm{E}(b-\beta)(b-\beta)' - \mathrm{E}(\hat{\beta}-\beta)(\hat{\beta}-\beta)' \\
&= \Big(\frac{k}{n-p+2}\Big)\,\mathrm{E}\Big[\frac{y'\bar{H}y}{y'Hy}\{b(b-\beta)' + (b-\beta)b'\}\Big] \\
&\quad -\Big(\frac{k}{n-p+2}\Big)^2\,\mathrm{E}\Big[\Big(\frac{y'\bar{H}y}{y'Hy}\Big)^2 bb'\Big]. \qquad (2.4)
\end{aligned}$$

Under the specification of first four moments of the disturbances, it is not possible to evaluate the expectations in (2.4). We therefore derive an asymptotic approximation for Δ when disturbances are small, i.e., when σ is small and tends towards zero; see, e.g., Ullah, Srivastava and Chandra (1983).
Using (2.1), we can express

$$\begin{aligned}
\frac{y'\bar{H}y}{y'Hy} &= \sigma^2\frac{U'\bar{H}U}{\beta X'X\beta}\Big[1 + 2\sigma\frac{\beta'X'U}{\beta'X'X\beta} + \sigma^2\frac{U'HU}{\beta'X'X\beta}\Big]^{-1} \\
&= \sigma^2\frac{U'\bar{H}U}{\beta'X'X\beta} - 2\sigma^3\frac{U'\bar{H}U\cdot\beta'X'U}{(\beta'X'X\beta)^2} + O_p(\sigma^4)
\end{aligned}$$

Hence we get

$$\begin{aligned}
\frac{y'\bar{H}y}{y'Hy}b(b-\beta) &= \sigma^3\frac{U'\bar{H}U}{\beta'X'X\beta}\beta U'X(X'X)^{-1} \\
&\quad +\gamma^4\frac{U'HU}{\beta'X'X\beta}\Big[(X'X)^{-1} - \frac{2}{\beta'X'X\beta}\beta\beta'\Big] \\
&\quad X'U\cdot U'X(X'X)^{-1} + O_p(\sigma^5) \qquad (2.5) \\
\Big(\frac{y'\bar{H}y}{y'Hy}\Big)^2 bb' &= \sigma^4\Big(\frac{U'\bar{H}U}{\beta'X'X\beta}\Big)^2\beta\beta' + O_p(\sigma^5). \qquad (2.6)
\end{aligned}$$

Next, we observe that

$$\begin{aligned}
\mathrm{E}(U'\bar{H}U\cdot U) &= \gamma_1 d \\
\mathrm{E}(U'\bar{H}U\cdot UU') &= \gamma_2 D + (n-p)I + 2\bar{H}
\end{aligned}$$

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

Using these results along with (2.5) and (2.5) in (2.4), we obtain the following expression for Δ up to order $O(\sigma^4)$:

$$\Delta = \sigma^3\Big(\frac{\gamma_1 k}{n-p+2}\Big)(\beta h' + h\beta') + \frac{\sigma^4(n-p)kW}{(n-p+2)(\beta' X' X\beta)^2} \tag{2.7}$$

where

$$\begin{aligned} h &= (X'X)^{-1}X'd & (2.8)\\ W &= 2(\beta' X' X\beta)\Big[(X'X)^{-1} + \Big(\frac{\gamma_2}{n-p}\Big)(X'X)^{-1}X'DX(X'X)^{-1}\Big] \\ &\quad -2\Big(\frac{\gamma_2}{n-p}\Big)\Big[\beta'\beta X'DX(X'X)^{-1} + (X'X)^{-1}X'DX\beta\beta'\Big] \\ &\quad -[4 + k(1 + g\gamma_2)]\beta\beta' & (2.9) \end{aligned}$$

with

$$g = \frac{\text{tr} D\bar{H}}{(n-p)(n-p+2)}.$$

3 Comparisons for Asymmetrical Disturbances

If the distribution of disturbances is not symmetrical so that γ_1 is different from 0, the leading term in (2.7) helps in comparing the asymptotic efficiency. We thus have

$$\Delta = \sigma^3\Big(\frac{\gamma_1 k}{n-p+2}\Big)(\beta h' + h\beta') \tag{3.1}$$

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 (3.1) is positive semi-definite when γ_1 and $\beta' h (\neq 0)$ have same signs. This implies that Stein-rule estimators are 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 estimators for all positive values of k will asymptotically dominate the least squares estimators for those asymmetrical distributions for which skewness measure γ_1 and $\beta' h$ have identical signs.

The reverse is true, i.e., the Stein-rule estimators for all k fail to dominate the least squares estimator at least asymptotically when γ_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 (3.1) 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 Δ in order to make a choice between least squares and Stein-rule estimators.

We thus observe from (2.7) that the expression Δ under symmetry of the distribution of disturbances becomes

$$\Delta = \sigma^4 \frac{(n-p)k}{(n-p+2)(\beta' X' X \beta)^2} W \tag{4.1}$$

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

<u>*Lemma 1*</u>: The matrices $X' D X (X' X)^{-1}$ and $[I - X' D X (X' X)^{-1}]$ are at least positive semi definite.

<u>*Proof*</u>: 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

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

whence we obtain

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

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 ϕ, 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 ϕ and any $m \times m$ positive definite matrix Φ, the matrix $(\phi\phi' - \Phi)$ cannot be non-negative definite for m exceeding 1.

Proof: See Guilky and Price (1981).

When the distribution of disturbances is mesokurtic ($\gamma_2 = 0$), the expression (4.1) reduces to the following:

$$\begin{aligned} W &= 2(\beta'X'X\beta)(X'X)^{-1} - (4+k)\beta\beta' \\ &= 2(\beta'X'X\beta)(X'X)^{-\frac{1}{2}}\left[I - \left(\frac{4+k}{2\beta'X'X\beta}\right)(X'X)^{\frac{1}{2}}\beta\beta'(X'X)^{\frac{1}{2}}\right] \\ &\quad \cdot (X'X)^{-\frac{1}{2}} \end{aligned} \qquad (4.2)$$

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

$$\left(\frac{4+k}{2}\right) < 1 \qquad (4.3)$$

but it holds in no case for positive values of k. This implies that no Stein-rule estimator dominates the least squares dominator 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 estimators, 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 dominate the Stein-rule estimators 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 criteria.

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

$$2(\beta' X' X \beta)\Big[(X'X)^{-1} + \Big(\frac{\gamma_2}{n-p}\Big)(X'X)^{-1}X'DX(X'X)^{-1}\Big] - \Big[2\Big(\frac{\gamma_2}{n-p}\Big) + 4 + k(1 + g\gamma_2)\Big]\beta\beta' \tag{4.4}$$

is positive definite.

Applying Lemma 3, it is seen that (4.4) is positive definite if and only if

$$\Big[2\Big(\frac{\gamma_2}{n-p}\Big) + 4 + k(1+g\gamma_2)\Big]\frac{\beta'\Big[(X'X)^{-1} + \Big(\frac{\gamma_2}{n-p}\Big)(X'X)^{-1}X'DX(X'X)^{-1}\Big]^{-1}\beta}{2\beta'X'X\beta} < 1 \tag{4.5}$$

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

$$1 \leq \frac{\beta'\Big[(X'X)^{-1} + \Big(\frac{\gamma_2}{n-p}\Big)(X'X)^{-1}X'DX(X'X)^{-1}\Big]^{-1}\beta}{\beta'X'X\beta} \leq \Big(1 + \frac{\gamma_2}{n-p}\Big) \qquad (4.6)$$

which, when used in (4.5), clearly reveals that the inequality (4.5) cannot hold true for positive k. This implies that Stein-rule estimators 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 estimators, the matrix $(-W)$ should be positive definite. This will be the case so long as the matrix

$$[4 + k(1 + g\gamma_2)]\beta\beta' - 2(\beta'X'X\beta)\Big[(X'X)^{-1} + \Big(\frac{\gamma_2}{n-p}\Big)(X'X)^{-1}X'DX(X'X)^{-1}\Big] \qquad (4.7)$$

is positive definite.

Applying Lemma 4, we observe that the matrix expression (4.7) 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

$$2(\beta'X'X\beta)\Big(1 + \frac{\gamma_2}{n-p}\Big)(X'X)^{-1} - [4 + k(1 + g\gamma_2)]\beta\beta' \qquad (4.8)$$

is positive definite.

Notice that the matrix expression (4.8) 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 (4.8) is positive definite if and only if

$$\Bigg[\frac{4 + k(1 + g\gamma_2)}{2\Big(1 + \frac{\gamma_2}{n-p}\Big)}\Bigg] < 1 \qquad (4.9)$$

but it can never hold true for positive values of k as γ_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 weather the least squares estimator dominates the Stein-rule estimators in case of a platykurtic distribution. For this purpose, it suffices to examine the nature of following matrix expression.

$$\Big[4 + k(1 + g\gamma_2) + 2\Big(\frac{\gamma_2}{n-p}\Big)\Big]\beta\beta' - 2(\beta'X'X\beta)(X'X)^{-1} \tag{4.10}$$

as γ_2 is negative.

Using Lemma 4 and observing that $(2+\gamma_2)$ is always positive, it follows that (4.10) 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 behaviour of the two estimators b (OLSE) and $\hat{\beta}$ (Stein rule estimator (SRE)) in finite samples under the MSE-matrix criterion. We used the R programming environment (`http://www.r-project.org`) to conduct the simulation study. We considered six basic designs for the simulation study. The two different sample sizes ($n = 40$ and $n = 100$) are combined with each of the three different following error distributions:

(i) normal distribution (having no skewness and no kurtosis) with mean 0 and variance 1,

(ii) t-distribution (having no skewness but non-zero kurtosis) with 5 degrees of freedom and

(iii) beta distribution $Beta(1, 2)$ (having non-zero skewness and non-zero kurtosis). All the random observations from t and beta distributions are suitable scaled to have zero mean and unity variance. The difference in the results under normal, t and beta distribution may be considered as arising due to the skewness and kurtosis of the distributions. For each of the six combinations of sample sizes and error distributions, we programmed three loops which we call as β-loop, X-loop and ϵ-loop. The number of explanatory variables was fixed to $p = 6$. The β-loop was the outermost loop, followed by X-loop and ϵ-loop was the innermost loop.

The β-loop was constructed such that each of the $p = 6$ parameters were set to a fixed value (such that $\beta_1 = \beta_2 = \ldots = \beta_6 = c$) starting from $c = 0$ to $c = 0.1$ with an increment of 0.005. This resulted in a loop length of 21 different parameter settings for the vector β.

The X-loop was set to be of length 10000 and we generated a new X every time. Each row of X was generated from a multivariate normal distribution with mean $(1, 1, 1, -1, -1, -1)$ and identity covariance matrix I (no collinearity in the columns of X). Then we calculated the true response vector (without error) as $\tilde{y} = X\beta$.

The innermost ϵ-loop was set as of length 5000 and a new vector y was generated every time as $y = \tilde{y} + \epsilon$ under the three different error distributions. Note that since this is an inner loop of X and β-loops, so $\tilde{y}$ was always the same for each of the 5000 runs and only errors were generated in every run. The empirical MSE-matrix of both estimators was estimated based on these 5000 runs. After completion of 5000 runs, we calculated the eigenvalues of the difference of empirical MSE-matrices of b and $\hat{\beta}$. If all eigenvalues were positive, than we concluded that SRE was better than OLSE under the MSE-matrix criterion in that setting.

We considered all together 6.3×10^9 different data sets in this set up. In fact, we got 10000 different comparisons of the estimated empirical MSE-matrices of b and

$\hat{\beta}$ for each of the 21 different parameter settings of β and for each of the six basic designs. We counted the number of cases where the SRE was better than OLSE to comprehend the simulation output. The results are presented in Figure 1. We plotted a curve between the number of cases where SRE is better than OLSE and value of β under each sample size. The plots corresponding to the sample sizes 40 and 100 differ with respect to the distribution of errors. It is seen that the performance of SRE compared to the OLS estimator depends on the sample size and β. The SRE appears to be better than OLSE only for small (absolute) values of β. When the sample size is low, the probability is higher that SRE is better than OLSE for larger values of β in comparison to the case when sample size is large. We also note that the difference in the results among the different error distribution (when sample size is fixed) seems to be small but the number of cases of superiority of SRE over OLSE are different. For example, when $\beta = 0.1$ and $n = 40$, then SRE is better than OLSE 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 5.1 shows some selected outcomes for illustration.

We observe from Table 5.1 that the t-distribution has maximum and beta distribution has minimum number of cases for the superiority of SRE over OLSE. The rate of decrement of the number of cases of superiority of SRE over OLSE heavily depends on the sample size and the value of parameter vector β. When β is small, say, ≤ 0.03, then mostly SRE is better than OLSE, irrespective of the sample size. The number of case of superiority of SRE over OLSE are higher in smaller sample size ($n = 40$) than larger sample size ($n = 100$).

These findings indicate that there is no uniform dominance of SRE over OLSE and vice versa. Under the same experimental settings, SRE is not superior to OLSE 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 σ, then the number of cases of superiority of SRE over OLSE will change from the present values.

	$n = 40$			$n = 100$		
β	$N(0,1)$	$t(5)$	$Beta(1,2)$	$N(0,1)$	$t(5)$	$Beta(1,2)$
0	10000	10000	9999	10000	10000	10000
0.01	9999	9999	10000	10000	10000	10000
0.02	10000	10000	10000	10000	10000	10000
0.03	10000	10000	10000	10000	10000	10000
0.04	10000	10000	9998	9690	9751	9704
0.045	10000	10000	10000	6035	6135	6051
0.05	9999	9997	10000	1294	1408	1216
0.055	9958	9964	9956	97	115	85
0.06	9611	9680	9602	3	4	2
0.065	8449	8710	8408	1	0	1
0.07	6364	6745	6225	0	0	0
0.075	3984	4359	4040	0	0	0
0.08	2128	2432	2102	0	0	0
0.085	1021	1131	983	0	0	0
0.09	425	522	428	0	0	0
0.095	159	211	135	0	0	0
0.1	59	91	50	0	0	0

Table 5.1: Number of times where the Stein-rule estimator is better than OLS in 10000 simulations under different error distributions.

We are presenting below the MSE matrix of SRE and covariance matrix of OLSE for illustration when $\beta = 0.07$ and $n = 40$ under all three error distributions to get an idea about the non-normality effect. If there was no effect of coefficients of skewness and kurtosis, then all the matrices should be nearly same.

- **Normal distribution**

SRE:

$$\begin{bmatrix} 0.01969 & -0.005705 & 0.0003417 & 0.003382 & 0.004405 & 0.005621 \\ -0.005705 & 0.01747 & 0.001881 & 0.004259 & 0.00371 & 0.004046 \\ 0.0003417 & 0.001881 & 0.01095 & 0.005669 & 0.003347 & 0.004802 \\ 0.003382 & 0.004259 & 0.005669 & 0.01263 & 0.003933 & 0.002561 \\ 0.004405 & 0.00371 & 0.003347 & 0.003933 & 0.01016 & 0.00104 \\ 0.005621 & 0.004046 & 0.004802 & 0.002561 & 0.00104 & 0.0121 \end{bmatrix}$$

OLSE:

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

- ***t*-distribution**

SRE:

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

OLSE:

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

- **Beta distribution**

SRE:

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

OLSE:

$$\begin{bmatrix} 0.02101 & -0.0004467 & -0.002724 & 0.003978 & 0.003177 & 0.004978 \\ -0.0004467 & 0.02111 & -0.004928 & 0.0007727 & 0.00471 & 0.00227 \\ -0.002724 & -0.004928 & 0.02891 & 0.004476 & 0.005579 & 0.009623 \\ 0.003978 & 0.0007727 & 0.004476 & 0.01921 & -0.004037 & 0.00003745 \\ 0.003177 & 0.00471 & 0.005579 & -0.004037 & 0.01545 & 0.0003523 \\ 0.004978 & 0.00227 & 0.009623 & 0.00003745 & 0.0003523 & 0.01947 \end{bmatrix}$$

It is clear from these values that the covariance matrix of OLSE and MSE matrix of SRE are not same under different distribution 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 5.1 clearly indicates the effect of non-normality of error distributions on the variability of these estimators and the superiority of SRE and OLSE 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 MSE and covariance matrices of SRE and OLSE 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.

References

Chatterjee, S., and Hadi, A.S. (1988): **Sensitivity Analysis in Linear Regression**, John Wiley.

Guilky, D.K. and Price, J.M. (1981): "On comparing restricted least squares estimators", *Journal of Econometrics*, **15**, 397-404.

Judge, G.G. and M.E. Bock (1978): **The Statistical Implication of Pre-Test and Stein-rule Estimators**, North Holland.

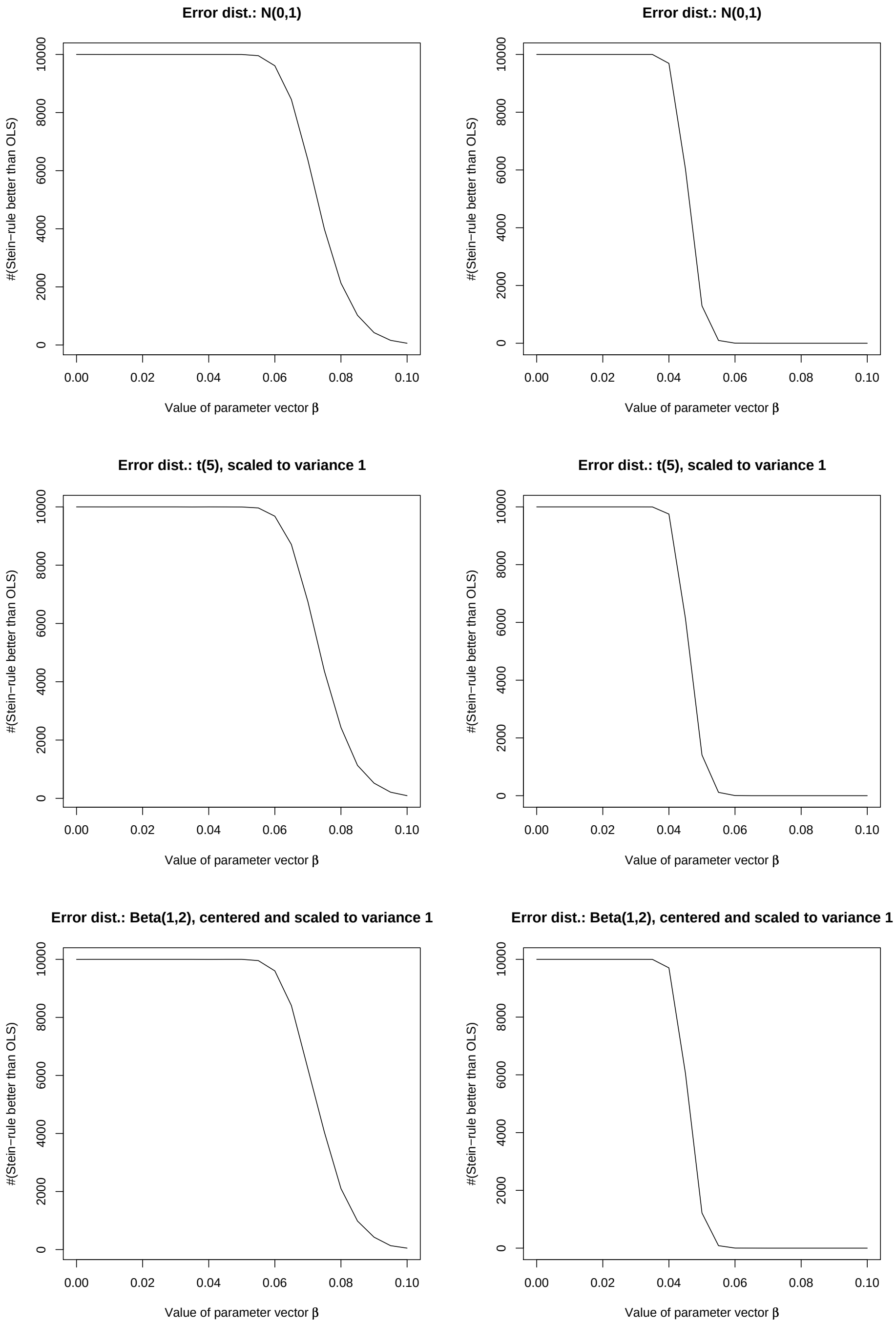
Judge, G.H., Griffiths, W.E., Hill, R.C., Lutekepohl, H. and Lee, T (1985): **The Theory and Practice of Econometrics**, John Wiley.

Rao, C.R. and H. Toutenburg (1999):**Linear Models: Least Squares and Alternatives**, Second Edition, Springer.

Saleh, A.K.Md.E. (2006): **Theory of Preliminary Test and Stein-type Estimation with Apllications**, John Wiley & Sons.

Ullah, A., Srivastava V.K. and Chandra, R. (1983): "Properties of shrinkage estimators in linear regression when disturbances are not normal", *Journal of Econometrics*, **21**, 389-402.

Yancy, T.A., Judge, G.G. and Bock, M.E. (1974): "A mean square error test when stochastic restriction are used in regression", *Communications in Statistics*, **3**, 755-768.



$n = 40$ $\qquad$ $n = 100$

Figure 5.1: Behaviour of number of cases in which SRE is better than OLSE with respect to β under different error distributions and sample sizes