SMALL SAMPLE PROPERTIES OF A PRE-TEST STEIN-RULE ESTIMATOR FOR EACH INDIVIDUAL REGRESSION COEFFICIENT UNDER AN ALTERNATIVE NULL HYPOTHESIS IN THE PRE-TEST*

By AKIO NAMBA and KAZUHIRO OHTANI

In this paper we consider a pre-test Stein-rule (SR) estimator for each individual regression coefficient when the null hypothesis in the pre-test is that the regression coefficient to be estimated is a zero. We derive the explicit formula for the MSE of the pre-test SR estimator, and examine the MSE performance of the pre-test SR estimator by numerical evaluations. Our numerical results show that using the null hypothesis that all the regression coefficients are zeros yields a better MSE performance than using the null hypothesis that the regression coefficient to be estimated is a zero.

1. Introduction

In the context of a linear regression, the Stein-rule (SR) estimator proposed by Stein (1956) and James and Stein (1961) dominates the ordinary least squares (OLS) estimator in terms of predictive mean squared error (PMSE) if the number of the regression coefficients is larger than or equal to three. Though the SR estimator dominates the OLS estimator, Baranchik (1970) showed that the SR estimator is further dominated by the positive-part Stein-rule (PSR) estimator. As is shown in Judge and Bock (1978), the PSR estimator is considered as a pre-test estimator after a pre-test for the null hypothesis that all the regression coefficients are zeros. If the SR estimator is used when the null hypothesis is rejected in the pre-test with an appropriate critical value, and if the regression coefficients are estimated as zeros when it is accepted, then the pre-test estimator is the PSR estimator.

Though the SR and PSR estimators dominate the OLS estimator when all the regression coefficients are estimated simultaneously, the dominance does not necessarily hold when each individual regression coefficient is estimated separately (e.g., Efron and Morris (1972) and Rao and Shinozaki (1978)). Ohtani and Koizumi (1996) examined the MSE performance of the SR and PSR estimators when our concern is to estimate each individual regression coefficient, and showed that the SR and PSR estimators do not necessarily dominate the OLS estimator, while the MSE dominance of the PSR estimator over the SR estimator still holds. Since they use the PSR estimator to estimate a particular regression coefficient, their null hypothesis in the pre-test is that all the regression coefficients are zeros. If our concern is to estimate a particular regression coefficient, then we should use the null hypothesis that the regression coefficient to be estimated is a zero.

In this paper, our goal is to estimate each individual regression coefficient separately, and we consider a pre-test SR estimator when the null hypothesis in the pre-test is that the regression

* This work was supported by JSPS KAKENHI Grant Numbers 22530208, 232430308.
coefficient to be estimated is a zero. In section 2 we introduce the model and the estimators, and in section 3 we derive the explicit formula for the moments of the estimator. In section 4 we examine the small sample properties of the estimators by numerical evaluations, using the explicit formula for the bias and MSE. Our numerical results show that even when our goal is to estimate each individual regression coefficient separately, using the null hypothesis that all the regression coefficients are zeros rather yields a better MSE performance than using the null hypothesis that the regression coefficient to be estimated is a zero.

2. Model and the estimators

Consider a linear regression model,
\[ y = X\beta + \varepsilon, \quad \varepsilon \sim N(0, \sigma^2 I_n), \]  
(1)
where \( y \) is an \( n \times 1 \) vector of observations on a dependent variable, \( X \) is an \( n \times k \) matrix of full column rank of observations on nonstochastic independent variables, \( \beta \) is a \( k \times 1 \) vector of regression coefficients, and \( \varepsilon \) is an \( n \times 1 \) vector of normal error terms.

Following Judge and Yancey (1986, p. 11.), we reparameterize the model (1) and work with the following orthonormal counterpart:
\[ y = Z\gamma + \epsilon, \]  
(2)
where \( Z = XS^{-1/2} \), \( \gamma = S^{1/2} \beta \), and \( S^{1/2} \) is the symmetric matrix such that \( S^{1/2} S^{-1/2} = Z'Z = I_k \), where \( S = X'X \). Then the ordinary least squares (OLS) estimator of \( \gamma \) is
\[ c = Z'y. \]  
(3)

In the context of reparameterized model, the Stein-rule (SR) estimator proposed by Stein (1956) is defined as
\[ c_{SR} = \left( 1 - \frac{\alpha \epsilon' \epsilon}{\epsilon' \epsilon} \right) c, \]  
(4)
where \( \epsilon = y - Zc \) and \( \alpha \) is a constant such that \( 0 \leq \alpha \leq 2(k-2)/(\nu + 2) \), where \( \nu = n - k \). As shown by Stein (1956), the SR estimator dominates the OLS estimator in terms of predictive mean squared error (PMSE) when \( k \geq 3 \). Also, James and Stein (1961) showed that the PMSE of the SR estimator is minimized when \( \alpha = (k - 2)/(\nu + 2) \). Thus, we use this value of \( \alpha \) hereafter.

Although the SR estimator dominates the OLS estimator, Baranchik (1970) showed that the SR estimator is further dominated by the positive-part Stein-rule (PSR) estimator defined as
\[ c_{\text{FSR}} = \max \left[ 0, \ 1 - \frac{a c' \epsilon}{c' c} \right] c. \] \tag{5}

The PSR estimator is considered as a pre-test estimator:

\[ c_{\text{FSR}} = I(F \leq \tau_{\text{FSR}}) \left( 1 - \frac{a c' \epsilon}{c' c} \right) c, \] \tag{6}

where \( I(A) \) is an indicator function such that \( I(A) = 1 \) if an event \( A \) occurs and \( I(A) = 0 \) otherwise, \( F = \frac{(c' \epsilon / k)}{(c' \epsilon / \nu)} \) is the test statistic for the null hypothesis \( H_0 : \gamma = 0 \) against the alternative \( H_1 : \gamma \neq 0 \), and \( \tau_{\text{FSR}} = a \nu / k \). Note that \( \tau_{\text{FSR}} \) is obtained by dividing \( a \nu \) by the degrees of freedom of the numerator of the test statistic \( F \) (i.e., \( k \)). Regarding the PSR estimator, the pre-test is conducted simultaneously for \( H_0 : \gamma = 0 \). Ohtani and Kozumi (1996) showed that the PSR estimator for each individual regression coefficient dominates the SR estimator in terms of MSE. However, the PSR estimator pre-tests simultaneously for \( H_0 : \gamma = 0 \) even when our goal is to estimate the \( i \) th regression coefficient. Thus, we propose an alternative pre-test SR estimator such that the pre-test is conducted individually. Let \( h \) be a \( k \times 1 \) vector with known elements. If \( h_i \) is the \( i \) th row vector of \( S^{-1/2} \), the estimator \( h' c_{\text{FSR}} \) is the \( i \) th element of the SR estimator for \( \beta \). Similarly, we consider the following pre-test estimator for the \( i \) th element of \( \beta \):

\[ h_i' \gamma_{\tau_i} = I(F_i \geq \tau_i) \left( 1 - \frac{a c' \epsilon}{c' c} \right) h_i' c, \] \tag{7}

where \( F_i = \frac{(h_i' \epsilon)^2}{(c' \epsilon / \nu)} \) and \( \tau_i \) is a critical value of the pre-test. We call this estimator the pre-test SR (PTSR) estimator. Regarding the PTSR estimator, the pre-test is conducted for the null hypothesis \( H_0 : \beta_i = 0 \), where \( \beta_i \) is the \( i \) th element of \( \beta \). We derive the explicit formula for the MSE of the PTSR estimator in the next section.

### 3. Moments of the estimator

In this section we derive the explicit formula for the MSE of the PTSR estimator. Since the elements of \( h \) are known, we assume that \( h'h = 1 \) without loss of generality. Then, the bias and MSE of the PTSR estimator are

\[ \text{Bias}[h_i' \gamma_{\tau_i}] = E \left[ I(F_i \geq \tau_i) \left( 1 - \frac{a c' \epsilon}{c' c} \right) (h_i' c) \right] - h_i' \gamma_i, \] \tag{8}

and
\[
\text{MSE}[h'(r)] = \mathbb{E}\left[I(F_i \geq r_i) \left(1 - \frac{a(b+c)}{c} \right)^2 (h'(r))^2 \right] - 2h'\gamma \mathbb{E}\left[I(F_i \geq r_i) \left(1 - \frac{a(b+c)}{c} \right)^2 (h'(r))^2 \right] + (h'(r))^2, \tag{9}
\]

If we define the functions as

\[
H(p; q; r) = \mathbb{E}\left[I(F_i \geq r_i) \left(1 - \frac{a(b+c)}{c} \right)^p (h'(r))^2 \right], \tag{10}
\]

\[
J(p; q; r) = \mathbb{E}\left[I(F_i \geq r_i) \left(1 - \frac{a(b+c)}{c} \right)^p (h'(r))^2 (h'(r))^2 \right], \tag{11}
\]

then the bias and the MSE of the PTSR estimator are written as

\[
\text{Bias}[h'(r)] = J(1, 0; r) - h'(r), \tag{12}
\]

\[
\text{MSE}[h'(r)] = H(2, 1; r) - 2h'\gamma J(1, 0; r) + (h'(r))^2, \tag{13}
\]

respectively.

As shown in the appendix, the explicit formulae for \(H(p; q; r)\) and \(J(p; q; r)\) are

\[
H(p; q; r) = (2\sigma^2)^p \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} w_i(\lambda_1)w_j(\lambda_2)G_{i+j}(p; q; r), \tag{14}
\]

\[
J(p; q; r) = (h'(r))(2\sigma^2)^p \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} w_i(\lambda_1)w_j(\lambda_2)G_{i+j+1}(p; q; r), \tag{15}
\]

where \(\lambda_1 = \gamma/(h'\gamma)/\sigma^2, \lambda_2 = \gamma/k + h'\gamma)/\sigma^2, w_i(\lambda) = \exp(-\lambda/2)(\lambda/2)^i/i!, \) and

\[
G_{ij}(p; q; r) = \frac{\Gamma(1/2 + q + i)}{\Gamma(q + i + j)} \frac{\Gamma((\nu + k)/2 + q - i + j)}{\Gamma((\nu/2) + q - i + j)} \prod_{k=1}^{\nu + k} I_{i+j}(1/2 + q + i)
\]

\[
\times \int_{\tau_i/(\nu + k)}^{1} \left(1 - \frac{t}{t_i} \right)^p \left(1 - \frac{t}{1 - t} \right)^{p/2 - 1} dt. \tag{16}
\]

where \(\tau_i^\nu = [(\nu + k)t - \tau_i]/(\nu t), \) and \(I_{i+j}(\cdot, \cdot)\) is the incomplete beta function ratio. See, for example, Abramowitz and Stegun (1972) for the definitions and properties of the incomplete beta function ratio.

Using these formulae, we examine the sampling property of the PTSR estimator by numerical evaluations.
4. Numerical analysis

Using the formula derived in the previous section, we examine the performance of the PTSR estimator by numerical evaluations.

The numerical evaluations were executed using the FORTRAN code. In calculating the integral in $G_{ij}(p; q; \tau_i)$ given in (16), we used the Simpson’s 3/8 rule with 200 equal subdivisions. The double infinite series in $H(p, q; \tau_i)$ and $J(p, q; \tau_i)$ are judged to converge when the increment of series gets smaller than $10^{-12}$. To compare the performances of the estimators, we evaluated the values of relative bias and relative MSE, defined as $\text{Bias}[h^\gamma]/\sigma$ and $\text{MSE}[h^\gamma]/\text{MSE}[h']$, where $\gamma$ is any estimator of $\gamma$. Thus, the estimator $h^\gamma$ has smaller MSE than the OLS estimator when the value of the relative MSE is smaller than unity. From (13), (14) and (15), we can easily show that the relative MSE of the PTSR estimator depends on the values of $k, n, \lambda_1$, and $\lambda_2$. Also, the relative bias of the PTSR estimator depends on the values of $k, n, h^\gamma/\sigma$ and $\lambda_2$. Thus, we use the following parameter values: $k = 3, 5, 8, n = 20, 30, 40, \lambda_1 = u(\lambda_1 + \lambda_2)$, where $u$ is a constant between 0 and 1, and $\lambda_1 + \lambda_2 = \text{various values}$. In calculating the biases of the estimators, we assume that $h^\gamma/\sigma > 0$ without loss of generality. For the purpose of comparison, we evaluated the relative MSE of the SR and PSR estimator using the formula derived in Ohtani and Kozumi (1996).

Since there is no theoretical information about the choice of the critical value of the pre-test of the PTSR estimator (i.e., $\tau_i$), we tried several values on $\tau_i$, and found that PTSR estimator has small MSE over a wide region of parameter space when $\tau_i = \tau_i^*$. Similar to the case of the SR estimator, $\tau_i^*$ is obtained by dividing $\alpha$, by the degrees of freedom of the numerator of the test statistic $F_i$ (i.e., 1). Thus, we showed the results for $\tau_i = \tau_i^*$. Also, since the results for $k = 5$ and $n = 30$ are qualitatively typical, we do not show the results for the other cases.

Tables 1 and 2 show the relative biases and relative MSE’s for $k = 5$ and $n = 30$. In Tables 1 and 2 ‘SR’ indicates the usual SR estimator for the $i$ th regression coefficient without any pre-test, and ‘PSR’ the pre-test estimator after the pre-test for $H_0 : \beta = 0$ (all coefficients are zeros). On the other hand, ‘PTSR’ is the pre-test estimator after the pre-test for $H_1 : \beta_i = 0$ (only the $i$ th coefficient is a zero). Although both the PSR estimator and the PTSR estimator are pre-test estimators based on the SR estimator, the null hypotheses in the pre-tests are different between the PTSR and PSR estimator.

From Table 1, we see that the PSR estimator has smaller absolute bias than the SR estimator. However, the SR estimator has smaller absolute bias than the PTSR estimator.

We see from Table 2 that PTSR estimator has smaller MSE than the PSR estimator when both $\lambda_1 + \lambda_2$ and $w$ are small. However, the PSR estimator has smaller MSE than PTSR over a wide region of the parameter space. In particular, when $\lambda_1 + \lambda_2 = 3.0$ and $w \geq 0.7$, the MSE of the PTSR estimator can be much larger than that of the PSR estimator. As a whole, the PSR estimator seems to have a better MSE performance than $h^\gamma/\tau_i^*$.

This indicates that even if our concern is to estimate the $i$ th regression coefficient as accurate as possible, the null hypothesis $H_0 : \beta = 0$ is preferable to the null hypothesis $H_0 : \beta_i = 0$. 

\textit{Note:} The dou...
Appendix

We derive the formulae for $H(p; q; \tau_i)$ and $J(p; q; \tau_i)$ in the appendix. First, we derive the formula for $H(p; q; \tau_i)$. Let $u_1 = (h'c)^2/\sigma^2$, $u_2 = c'[h_k - hh'/\gamma]/\sigma^2$ and $u_3 = c'c/\sigma^2$. Then $u_1 \sim \chi^2_1(\lambda_1)$, $u_2 \sim \chi^2_2(\lambda_2)$, where $\lambda_1 = \gamma''/\gamma/\sigma^2$, $\lambda_2 = \gamma''/\gamma/\sigma^2$, and $\chi^2_2(\lambda)$ is the noncentral chi-square distribution with $f$ degrees of freedom and noncentrality parameter $\lambda$. Also, $u_3$ is distributed as the chi-square distribution with $l$ degrees of freedom, and $u_1$, $u_2$ and $u_3$ are mutually independent.

Using $u_1$, $u_2$ and $u_3$, $H(p; q; \tau_i)$ is expressed as

$$H(p; q; \tau_i) = \mathbb{E} \left[ I(F_1 \geq \tau_i) \left( \frac{u_1 + u_2 - au_3}{u_1 + u_2} \right)^{\gamma} \sigma^{-2} u_1 \gamma \right]$$

$$= \sigma^{2\gamma} \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} K_{ij} \int \int \frac{u_1 + u_2 - au_3}{u_1 + u_2} u_1^{1/2+q+i-1} u_2^{(k-1)/2+j-1} u_3^{\gamma-1} \times \exp \left[ -\frac{u_1 + u_2 + u_3}{2} \right] du_3 du_2 du_1,$$  \(17\)

where
Table 2. Relative MSE’s for $k = 5$ and $n_1 = 30$.

<table>
<thead>
<tr>
<th>Estimator</th>
<th>$v$</th>
<th>$\lambda_1 + \lambda_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.5</td>
<td>3.0</td>
</tr>
<tr>
<td>SR</td>
<td>0.1</td>
<td>0.471</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>0.521</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.571</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>0.621</td>
</tr>
<tr>
<td></td>
<td>0.9</td>
<td>0.670</td>
</tr>
<tr>
<td></td>
<td>1.0</td>
<td>0.695</td>
</tr>
<tr>
<td>PSR</td>
<td>0.1</td>
<td>0.370</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>0.408</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.447</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>0.485</td>
</tr>
<tr>
<td></td>
<td>0.9</td>
<td>0.523</td>
</tr>
<tr>
<td></td>
<td>1.0</td>
<td>0.543</td>
</tr>
<tr>
<td>PTSR</td>
<td>0.1</td>
<td>0.260</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>0.340</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.418</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>0.492</td>
</tr>
<tr>
<td></td>
<td>0.9</td>
<td>0.564</td>
</tr>
<tr>
<td></td>
<td>1.0</td>
<td>0.599</td>
</tr>
</tbody>
</table>

\[ K_{ij} = \frac{w_i(\lambda_1)w_j(\lambda_2)}{2^{(\nu+k)/2+i+j}\Gamma((\nu/2)\Gamma((k-1)/2+j}}, \quad (18) \]

and \( w_\nu(\lambda) = \exp(-\nu\lambda/2)(\lambda/2)^i/\nu! \).

Making use of the change of variables, \( v_1 = u_1/u_3 \) and \( v_2 = u_2/u_3 \), the integral part in (17) reduces to

\[
\int_{\tau_i/\nu}^{\infty} \int_{\tau_0}^{\infty} \left( \frac{v_1 + v_2 - a}{v_1 + v_2} \right)^\nu F_{1/2+q+i-1, (k-1)/2+j-1, (\nu+k)/2+q+i+j-1} dU_3 dU_2 dU_1.
\]

(19)

Again, making use of the change of variable, \( z = u_3(v_1 + v_2 + 1)/2 \), (19) reduces to

\[
2^{(\nu+k)/2+q+i+j}\Gamma((\nu+k)/2+q+i+j) \times \int_{\tau_i/\nu}^{\infty} \int_{\tau_0}^{\infty} \left( \frac{v_1 + v_2 - a}{v_1 + v_2} \right)^\nu F_{1/2+q+i-1, (k-1)/2+j-1, (\nu+k)/2+q+i+j-1} dU_3 dU_2 dU_1,
\]

(20)
Further, making use of the change of variables, \( w_1 = v_1 + v_2 \) and \( w_2 = v_2/(v_1 + v_2) \), the integral part in (20) reduces to

\[
B((k - 1)/2 + j, 1/2 + q + i) \\
\times \int_{\tau_1/\nu}^{\infty} I_{\nu} ((k - 1)/2 + j, 1/2 + q + i) \left( \frac{w_1 - a}{w_1} \right)^p \\
\times \frac{w_1^{k/2+q+i+j-1}}{(w_1 + 1)^{(n+k)/(2+q+i+j)}} \, dw_1,
\]

(21)

where \( I_\nu(\cdot, \cdot) \) is the incomplete beta function ratio. Finally, making use of the change of variable, \( t = w_1/(1 + w_1) \), and performing some manipulations, we obtain (14) in the text.

Next, we derive the formula for \( J(p, q; \tau_1) \). Differentiating \( H(p, q; \tau_1) \) given in (14) with respect to \( \gamma \), we obtain

\[
\frac{\partial H(p, q; \tau_1)}{\partial \gamma} = (2\sigma^2)^q \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \left[ \frac{\partial w_j(\lambda_1)}{\partial \gamma} w_j(\lambda_2) + w_j(\lambda_1) \frac{\partial w_j(\lambda_2)}{\partial \gamma} \right] G_{ij}(p, q; \tau_1),
\]

\[
= (2\sigma^2)^q \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \left[ - \frac{h'\gamma}{\sigma^2} w_i(\lambda_1) w_j(\lambda_2) + \frac{h'\gamma}{\sigma^2} w_{i-1}(\lambda_1) w_j(\lambda_2)
\right.
\]

\[
\left. - \frac{(I_k - h'\gamma)}{\sigma^2} w_i(\lambda_1) w_j(\lambda_2) + \frac{(I_k - h'\gamma)}{\sigma^2} w_{i-1}(\lambda_1) w_j(\lambda_2) \right]
\times G_{ij}(p, q; \tau_1),
\]

(22)

where we may define \( w_{-1}(\lambda) = 0 \). Since \( h'h' = 1 \) and \( h'(I_k - h'\gamma) = 0 \), we obtain

\[
h' \frac{\partial H(p, q; \tau_1)}{\partial \gamma} = \frac{h'\gamma}{\sigma^2} H(p, q; \tau_1) + \frac{h'\gamma}{\sigma^2} (2\sigma^2)^q
\]

\[
\times \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \frac{w_i(\lambda_1) w_j(\lambda_2)}{w_{i-1}(\lambda_1) w_j(\lambda_2)} G_{i+1,j}(p, q; \tau_1).
\]

(23)

Also, expressing \( H(p, q; \tau_1) \) by \( c'c \) and \( c'e \), we have

\[
H(p, q; \tau_1) = \int_{\Gamma_1 \geq \tau_1} \left( 1 - \frac{\alpha c'e}{c'c} \right)^p (H'c)^{\frac{\gamma}{2}} f_N(c) f_e(c'e) \, dc \, dc',
\]

(24)

where \( f_e(c'e) \) is the density function of \( c'E \) and

\[
f_N(c) = \frac{1}{(2\pi)^{k/2} \sigma^k} \exp \left[ -\frac{(c - \gamma)'(c - \gamma)}{2\sigma^2} \right],
\]

(25)

Differentiating (24) with respect to \( \gamma \) and multiplying \( h' \) from the left, we obtain
Equating (23) and (26), we obtain (15) in the text.

\[
\frac{h^T H(p; \theta; \tau_i)}{\gamma} = \frac{1}{\sigma^2} \mathbb{E}\left[ I(F_i \geq \tau_i) \left( 1 - \frac{\alpha^T c}{c^T c} \right)^F (h^T c)^{2q} H' c \right] - \frac{h^T \gamma}{\sigma^2} H(p; \theta; \tau_i). \tag{26}
\]  

REFERENCES


