MSE dominance of the PT-2SHI estimator over the positive-part Stein-rule estimator in regression

Namba, Akio

(Citation)
Journal of Statistical Planning and Inference,89(1-2):175-185

(Issue Date)
2000-08

(Resource Type)
journal article

(Version)
Accepted Manuscript

(URL)
https://hdl.handle.net/20.500.14094/90000345

MSE DOMINANCE OF THE PT-2SHI ESTIMATOR OVER THE POSITIVE-PART STEIN-RULE ESTIMATOR IN REGRESSION

Akio Namba

Graduate School of Economics
Kobe University
Rokko, Nada-ku, Kobe 657-8501
Japan

Abstract: In this paper, we consider a heterogeneous pre-test estimator which consists of the two-stage hierarchial information (2SHI) estimator and the Stein-rule (SR) estimator. This estimator is called the pre-test 2SHI (PT-2SHI) estimator. It is shown analytically that the PT-2SHI estimator dominates the SR estimator in terms of mean squared error (MSE) if the parameter values in the PT-2SHI estimator are chosen appropriately. Moreover, our numerical results show that the appropriate PT-2SHI estimator dominates the positive-part Stein-rule (PSR) estimator.

Keywords: Stein-rule estimator; Positive-part Stein-rule estimator; 2SHI estimator; Pre-test 2SHI estimator; Mean squared error; Dominance.

Short running title: MSE dominance of the PT-2SHI estimator

AMS subject classification: 62J07, 62C15.

1 Introduction

In the context of linear regression with k regression coefficients, the Stein-rule (SR) estimator proposed by Stein (1956) and James and Stein (1961) dominates the ordinary least squares (OLS) estimator in terms of mean squared error (MSE) for $k \geq 3$ if the predictive squared error loss function is used. Further, as is shown in Baranchik (1970), the SR estimator is dominated by the positive-part Stein-rule (PSR) estimator.

Recently, Shao and Strawderman (1994) proposed a family of estimators which dominate the PSR estimator. However, Shao and Strawderman (1994) assumed that the disturbance variance is known. Sugiura and Takagi (1996) extended their estimator to the case of unknown disturbance variance. However, since the dominance is valid when the condition $n \geq \max[9k + 10,\ 13k - 7]$ holds where n is sample size, Sugiura and Takagi's (1996) estimator needs relatively large sample size (e.g., when $k = 3$, $n \geq 37$).

As an improved estimator, Theil (1971) proposed the minimum mean squared error (MMSE) estimator. However, Theil's (1971) MMSE estimator includes unknown parameters, Farebrother (1975) suggested an operational variant of the MMSE estimator which is obtained by replacing the unknown parameters by the OLS estimators. As an extension of the MMSE estimator, Ohtani (1996) considered the adjusted minimum mean squared error (AMMSE) estimator which is obtained by adjusting the degrees of freedom of the operational variant of the MMSE estimator. He showed by numerical evaluations that if $k \leq 5$, the AMMSE estimator has smaller MSE than the SR and PSR estimators in a wide region of the noncentrality parameter.

In particular, when $k = 3$, the MSE of the AMMSE estimator can be much smaller than that of the PSR estimator for small values of the noncentrality parameter. Ohtani (1999) considered the heterogeneous pre-test estimator such that the AMMSE estimator is used if the null hypothesis that all the regression coefficients are zeroes (in other words, the value of the noncentrality parameter is zero) is accepted, and the SR estimator is used if the null hypothesis is rejected. Although the results were obtained by numerical evaluations, he showed that a heterogeneous pre-test estimator dominates the PSR estimator when $k = 3$ and the critical value of the pre-test is chosen appropriately. This indicates that the AMMSE estimator works effectively when the value of the noncentrality parameter is close to zero. However, when $k \geq 6$, the AMMSE estimator has larger MSE than the PSR estimator for small values of the noncentrality parameter. This indicates that if $k \geq 6$, there is no incentive to use the heterogeneous pre-test estimator since the AMMSE estimator never works effectively when the value of the noncentrality parameter is close to zero.

To improve the SR estimator, Tran Van Hoa and Chaturvedi (1990) suggested the family of two-stage hierarchial information (2SHI) estimators. Namba (1998) derived the exact MSE of the 2SHI estimator and showed that the 2SHI estimator with appropriate parameter values can have much smaller MSE than the PSR estimator for small values of the noncentrality parameter even when $k = 8$. Thus, in this paper, we consider a heterogeneous pre-test estimator which consists of the 2SHI estimator and the SR estimator (hereafter, the PT-2SHI estimator), and examine its MSE performance.

In section 2 of this paper we present the estimators, and in section 3 we derive the exact formula of the MSE of the PT-2SHI estimator. It is shown analytically that the PT-2SHI estimator dominates the SR estimator when the parameter values in the 2SHI estimator and the critical value of the pre-test are chosen appropriately. In section 4, we compare the MSE performance of the PT-2SHI estimator with those of the SR estimator and the PSR estimator by numerical evaluations. It is shown by numerical evaluations that the PT-2SHI estimator dominates the PSR estimator even for relatively large k if the parameter values in the 2SHI estimator and the critical value of the pre-test are chosen appropriately. Finally, some concluding remarks are given in section 5.

2 Model and the estimators

Consider a linear regression model,

$$y = X\beta + \epsilon, \quad \epsilon \sim N(0, \sigma^2 I_n). \tag{1}$$

where y is an $n \times 1$ vector of observations on the dependent variable, X is an $n \times k$ matrix of full column rank of observations on nonstochastic independent variables, β is a $k \times 1$ vector of coefficients, and ϵ is an $n \times 1$ vector of normal error terms with $E[\epsilon] = 0$ and $E[\epsilon\epsilon'] = \sigma^2 I_n$, where σ is constant but unknown.

The ordinary least squares (OLS) estimator of β is

$$b = S^{-1}X'y, \tag{2}$$

where $S = X'X$. In the context of linear regression, the Stein-rule (SR) estimator proposed by Stein (1956) is defined as

$$b_{SR} = \left(1 - \frac{ae'e}{b'Sb}\right) b, \tag{3}$$

where $e = y - Xb$, and a is a constant such that $0 \leq a \leq 2(k-2)/(\nu+2)$, where $\nu = n - k$. If we use the loss function

$$L(\bar{\beta}; \beta) = (\bar{\beta} - \beta)'S(\bar{\beta} - \beta), \tag{4}$$

where $\bar{\beta}$ is any estimator of β, the SR estimator dominates the OLS estimator in terms of mean squared error (MSE) for $k \geq 3$. As is shown in James and Stein (1961), the MSE of the SR estimator is minimized when $a = (k-2)/(\nu+2)$. Thus we use this value of a 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

$$b_{PSR} = \max\left[0, 1 - \frac{ae'e}{b'Sb}\right] b. \tag{5}$$

As an improved estimator, Ohtani (1996) proposed the adjusted minimum mean squared error (AMMSE) estimator:

$$b_{AM} = \left(\frac{b'Sb/k}{b'Sb/k + e'e/\nu}\right) b. \tag{6}$$

He showed by numerical evaluations that if $k \leq 5$ the AMMSE estimator has smaller MSE than the PSR estimators for small values of the noncentrality parameter defined as $\lambda = \beta' S \beta / \sigma^2$. Thus, Ohtani (1999) considered the following heterogeneous pre-test estimator:

$$\widehat{\beta}_{AM}(\tau) = I(F \leq \tau) b_{AM} + I(F > \tau) b_{SR}, \tag{7}$$

where $F = (b'Sb/k)/(e'e/\nu)$ is the test statistic for $H_0 : \beta = 0$, τ is the critical value of the pre-test, and $I(A)$ is an indicator function such that $I(A) = 1$ if an event A occurs and $I(A) = 0$ otherwise. He showed by numerical evaluations that the heterogeneous pre-test estimator dominates the PSR estimator for $k = 3$ if the critical value of the pre-test is chosen appropriately.

To improve the SR estimator, Tran Van Hoa and Chaturvedi (1990) suggested the following family of two-stage hierarchial information (2SHI) estimators:

$$b_H = \left(1 - c\omega \frac{e'e}{b'Sb} - c(1-\omega) \frac{e'e}{b'Sb + c^* e'e}\right) b, \tag{8}$$

where c and c^* are two non-negative characterizing scalars, and ω is a constant such that $0 \leq \omega \leq 1$. If $c \leq 2(k-2)/(\nu+2)$, the 2SHI estimator is minimax since it satisfies Baranchik's (1970) condition. When $c^* = 0$ or when $\omega = 1$, the 2SHI estimator reduces to the SR estimator. Also, the 2SHI estimator reduces to the AMMSE estimator when $\omega = 0$ and $c = c^* = k/\nu$. Namba (1998) derived the explicit formula of the MSE of the 2SHI estimator and showed by numerical evaluations that the 2SHI estimator has much smaller MSE than the PSR estimator in the neighborhood of $\lambda = 0$ even when $k = 8$ if the values of c, c^* and ω are chosen appropriately. Thus, in this paper, we consider the following heterogeneous pre-test estimator which consists of the 2SHI estimator and the SR estimator:

$$\widehat{\beta}_H(\tau) = I(F \leq \tau) b_H + I(F > \tau) b_{SR}. \tag{9}$$

Hereafter, we call this estimator the pre-test 2SHI (PT-2SHI) estimator. The PT-2SHI estimator is inadmissible since it is not smooth enough.

3 MSE of the PT-2SHI estimator

The MSE of the PT-2SHI estimator is

$$
\begin{aligned}
MSE[\widehat{\beta}_H(\tau)] &= E[(\widehat{\beta}_H(\tau)-\beta)'S(\widehat{\beta}_H(\tau)-\beta)] \\
&= E[\widehat{\beta}'_H(\tau)S\widehat{\beta}_H(\tau)] - 2E[\beta' S\widehat{\beta}_H(\tau)] + \beta' S\beta \\
&= E[I(F\le\tau)b'_H Sb_H + I(F>\tau)b'_{SR}Sb_{SR}] \\
&\quad -2E[I(F\le\tau)\beta' Sb_H + I(F>\tau)\beta' Sb_{SR}] + \beta' S\beta. \qquad (10)
\end{aligned}
$$

Here, we define the functions, $H(p,q,r,\alpha;\tau)$ and $J(p,q,r,\alpha;\tau)$, as

$$
H(p,q,r,\alpha;\tau) = E\left[I(F\le\tau)\frac{(b'Sb)^p(e'e)^q}{(b'Sb+\alpha e'e)^r}\right], \qquad (11)
$$

and

$$
J(p,q,r,\alpha;\tau) = E\left[I(F\le\tau)\frac{(b'Sb)^p(e'e)^q}{(b'Sb+\alpha e'e)^r}\beta' Sb\right], \qquad (12)
$$

and let $a_1 = c\omega$, $a_2 = c(1-\omega)$ and $a_3 = c^*$. Performing some manipulations, we have:

$$
\begin{aligned}
&MSE[\widehat{\beta}_H(\tau)] \\
&= (a_1^2 - a^2)H(0,2,1,0;\tau) + a_2^2 H(1,2,2,a_3;\tau) - 2(a_1-a)H(0,1,0,0;\tau) \\
&\quad -2a_2 H(1,1,1,a_3;\tau) + 2a_1a_2 H(0,2,1,a_3;\tau) + H(1,0,0,0;\infty) \\
&\quad -2aH(0,1,0,0;\infty) + a^2 H(0,2,1,0;\infty) - 2(a-a_1)J(0,1,1,0;\tau) \\
&\quad +2a_2 J(0,1,1,a_3;\tau) - 2J(0,0,0,0;\infty) + 2aJ(0,1,1,0;\infty) + \beta' S\beta. \qquad (13)
\end{aligned}
$$

As is shown in appendix, the explicit formulae of $H(p,q,r,\alpha;\tau)$ and $J(p,q,r,\alpha;\tau)$ are

$$
H(p,q,r,\alpha;\tau) = (2\sigma^2)^{p+q-r}\sum_{i=0}^{\infty} w_i(\lambda)G_i(p,q,r,\alpha;\tau), \qquad (14)
$$

and

$$
J(p,q,r,\alpha;\tau) = \beta' S\beta(2\sigma^2)^{p+q-r}\sum_{i=0}^{\infty} w_i(\lambda)G_{i+1}(p,q,r,\alpha;\tau), \qquad (15)
$$

where

$$G_i(p,q,r,\alpha;\tau) = \frac{\Gamma((\nu+k)/2+p+q-r+i)}{\Gamma(k/2+i)\Gamma(\nu/2)} \int_0^{\tau^*} \frac{t^{k/2+p+i-1}(1-t)^{\nu/2+q-1}}{[\alpha+t(1-\alpha)]^r} dt, \tag{16}$$

$w_i(\lambda) = \exp(-\lambda/2)(\lambda/2)^i/i\,!$, $\lambda = \beta' S\beta/\sigma^2$ and $\tau^* = k\tau/(\nu+k\tau)$.

Substituting (14) and (15) in (13), we obtain the explicit formula of the MSE of the PT-2SHI estimator.

Differentiating $G_i(p,q,r,\alpha;\tau)$ with respect to τ, we have:

$$\begin{aligned} &\frac{\partial G_i(p,q,r,\alpha;\tau)}{\partial\tau} \\ &\quad = \frac{\Gamma((\nu+k)/2+p+q-r+i)}{\Gamma(k/2+i)\Gamma(\nu/2)} \frac{k^{k/2+p+i}\nu^{\nu/2+q}\tau^{k/2+p+i-1}}{(k\tau+\alpha\nu)^r(\nu+k\tau)^{(\nu+k)/2+p+q-r+i}}. \end{aligned} \tag{17}$$

Utilizing (17), and performing some manipulations, we obtain:

$$\begin{aligned} &\frac{\partial MSE[\widehat{\beta}_H(\tau)]/2\sigma^2}{\partial\tau} \\ &\quad = \sum_{i=0}^{\infty} w_i(\lambda) \frac{\Gamma((\nu+k)/2+i+1)}{\Gamma(k/2+i)\Gamma(\nu/2)} \frac{k^{k/2+i}\nu^{\nu/2}\tau^{k/2+i-1}}{(\nu+k\tau)^{(\nu+k)/2+i+1}} \\ &\qquad \times \Bigg[(a_1^2-a^2)\frac{\nu^2}{k\tau} + a_2^2\frac{k\nu^2\tau}{(k\tau+a_3\nu)^2} - 2(a_1-a)\nu - 2a_2\frac{k\nu\tau}{(k\tau+a_3\nu)} \\ &\qquad\qquad +2a_1a_2\frac{\nu^2}{(k\tau+a_3\nu)} - \lambda(a-a_1)\frac{\nu}{k/2+i} + \lambda a_2\frac{k\nu\tau}{(k/2+i)(k\tau+a_3\nu)}\Bigg] \\ &\quad = \sum_{i=0}^{\infty} w_i(\lambda) \frac{\Gamma((\nu+k)/2+i+1)}{\Gamma(k/2+i)\Gamma(\nu/2)} \frac{k^{k/2+i}\nu^{\nu/2+1}\tau^{k/2+i-1}}{(\nu+k\tau)^{(\nu+k)/2+i+1}} D_1(\tau)D_2(\tau), \end{aligned} \tag{18}$$

where

$$D_1(\tau) = a - a_1 - a_2\frac{k\tau}{k\tau+a_3\nu}, \tag{19}$$

and

$$D_2(\tau) = -a\frac{\nu}{k\tau} - a_1\frac{\nu}{k\tau} - a_2\frac{\nu}{k\tau+a_3\nu} + 2 - \frac{\lambda}{k/2+i}. \tag{20}$$

From (18), a necessary condition of $MSE[\widehat{\beta}_H(\tau)]$ to be monotone decreasing on $\tau \in (0,\tau_0)$ is:

$$D_1(\tau)D_2(\tau) \le 0, \tag{21}$$

and a sufficient condition is:

$$D_1(\tau)D_2(\tau) \leq 0 \text{ and } D_1(\tau) \geq 0. \tag{22}$$

If $a - a_1 - a_2 > 0$, $D_1(\tau) > 0$. Also, if $a - a_1 - a_2 < 0$ and $a - a_1 > 0$, $D_1(\tau) \geq 0$ for $0 \leq \tau \leq \tau_1$, where τ_1 is the positive solution of $D_1(\tau) = 0$.

Also, if we define

$$\begin{aligned} D_2^*(\tau) &= -a\frac{\nu}{k\tau} - a_1\frac{\nu}{k\tau} - a_2\frac{\nu}{k\tau + a_3\nu} + 2 \\ &= \frac{2k^2\tau^2 + \nu k(2a_3 - a - a_1 - a_2)\tau - \nu^2 a_3(a + a_1)}{k\tau(k\tau + a_3\nu)}, \end{aligned} \tag{23}$$

then $D_2^*(\tau) \geq D_2(\tau)$. Since a quadratic equation

$$2k^2\tau^2 + \nu k(2a_3 - a - a_1 - a_2)\tau - \nu^2 a_3(a + a_1) = 0 \tag{24}$$

has one positive solution and one negative solution, $D_2^*(\tau) \leq 0$ for $0 < \tau \leq \tau_2$, where τ_2 is the positive solution of (24).

Thus, the MSE of the PT-2SHI estimator is a monotone decreasing function of τ for $0 \leq \tau \leq \tau_2$ if $a - a_1 - a_2 > 0$. Also, if $a - a_1 > 0$ and $a - a_1 - a_2 < 0$, the MSE of the PT-2SHI estimator is a monotone decreasing function for $0 \leq \tau \leq \min[\tau_1, \tau_2]$. Since the PT-2SHI estimator reduces to the SR estimator when $\tau = 0$, we obtain the following two theorems.

Theorem 1 *If $a - a_1 - a_2 > 0$, the PT-2SHI estimator dominates the SR estimator for $0 < \tau \leq \tau_2$. Particularly, the PT-2SHI estimator with $\tau = \tau_2$ has the smallest MSE among the class of the PT-2SHI estimators with $0 < \tau \leq \tau_2$.*

Theorem 2 *If $a - a_1 > 0$ and $a - a_1 - a_2 < 0$, the PT-2SHI estimator dominates the SR estimator for $0 < \tau \leq \tau_3$, where $\tau_3 = \min[\tau_1, \tau_2]$. Particularly, the PT-2SHI estimator with $\tau = \tau_3$ has the smallest MSE among the class of the PT-2SHI estimators with $0 < \tau \leq \tau_3$.*

Since further theoretical analysis of the MSE of the PT-2SHI estimator seems to be difficult, we compare the MSE's of the PSR estimator and the PT-2SHI estimator by numerical evaluations in the next section.

Table 1: Values of τ_4 for $\omega = 0.0$ and $c_H = 30.0$

	k							
n	3	4	5	6	8	10	12	15
20	.3310	.4923	.5870	.6463	.7125	.7392	.7396	.6880
25	.3390	.5060	.6042	.6681	.7433	.7820	.8003	.8010
30	.3440	.5142	.6153	.6815	.7613	.8053	.8305	.8471
35	.3480	.5202	.6230	.6905	.7730	.8200	.8486	.8722
40	.3501	.5250	.6285	.6971	.7813	.8300	.8606	.8880

4 Numerical analysis

In the numerical evaluations, we used the values of $c = c^* = c_H(k-2)/(\nu+2)$, where c_H is a positive constant. Also, the numerical evaluations were executed for $k = 3, 4, 5, \ldots, 15$, $n = 20$, 25, 30, 35, 40 and $\lambda =$ various values. The numerical evaluations were executed on a personal computer, using the FORTRAN code. In evaluating the integral in $G_i(p,q,r,\alpha;\tau)$ given in (16), we used Simpson's rule with 500 equal subdivisions. The infinite series in $H(p,q,r,\alpha;\tau)$ and $J(p,q,r,\alpha;\tau)$ were judged to converge when the increment of the series got smaller than 10^{-12}.

To compare the MSE performances of the estimators, we evaluated the values of relative MSE defined as $MSE(\bar{\beta})/MSE(b)$, where $\bar{\beta}$ is any estimator of β. Thus, the estimator $\bar{\beta}$ has smaller MSE than the OLS estimator when the value of relative MSE is smaller than unity.

In a similar way to Ohtani (1999), we found the critical value of the pre-test, τ_4, such that $MSE[\widehat{\beta}_H(\tau_4)] = MSE[b_{PSR}]$ down to five decimal places at $\lambda = 0$.

We found by trial and error that the relatively small value of ω (e.g., $\omega \leq 0.1$) and the relatively large value of c_H (e.g., $c_H \geq 8.0$) work effectively. Thus, we use the values $\omega = 0.0$ and $c_H = 30.0$ in the following numerical evaluations. The values of τ_4 for $\omega = 0.0$, $c_H = 30.0$, k =3, 4, 5, 6, 8, 10, 12, 15 and $n = 20$, 25, 30, 35, 40 are shown in Table 1.

The results for $k = 5$ and $n = 20$ are shown in Table 2. Since these results are qualitatively typical, we do not show the results for the other cases. Though the PT-2SHI estimator with $\tau = \tau_2$ dominates the SR estimator from Theorem 2, it has slightly larger MSE than the PSR

Table 2
Relative MSE's of the SR, PSR, and the PT-2SHI estimators
for $k = 5$ and $n = 20$

λ	SR	PSR	PT − 2SHI		
			$\tau = \tau_1$ (0.5477)	$\tau = \tau_2$ (0.5133)	$\tau = \tau_4$ (0.5870)
.0	.47059	.35330	.35336	.35335	.35330
.1	.48103	.36608	.36605	.36605	.36600
.2	.49117	.37855	.37843	.37845	.37841
.3	.50104	.39073	.39052	.39056	.39052
.4	.51062	.40261	.40233	.40238	.40235
.6	.52902	.42553	.42511	.42519	.42517
.8	.54642	.44737	.44682	.44693	.44692
1.0	.56288	.46818	.46753	.46765	.46765
1.5	.60036	.51603	.51516	.51533	.51534
2.0	.63318	.55845	.55745	.55765	.55768
3.0	.68747	.62951	.62841	.62864	.62869
4.0	.73000	.68565	.68458	.68481	.68487
5.0	.76376	.73020	.72924	.72945	.72951
6.0	.79094	.76577	.76493	.76512	.76519
7.0	.81309	.79437	.79366	.79383	.79388
8.0	.83138	.81754	.81696	.81710	.81715
9.0	.84666	.83648	.83602	.83613	.83617
10.0	.85955	.85211	.85174	.85183	.85187
12.0	.88003	.87609	.87587	.87593	.87595
14.0	.89545	.89341	.89328	.89331	.89333
16.0	.90745	.90639	.90632	.90634	.90635
18.0	.91701	.91647	.91643	.91644	.91645
20.0	.92480	.92453	.92451	.92451	.92452
22.0	.93127	.93113	.93112	.93112	.93112
24.0	.93671	.93665	.93664	.93664	.93664
26.0	.94137	.94133	.94133	.94133	.94133
28.0	.94539	.94537	.94537	.94537	.94537
30.0	.94889	.94888	.94888	.94888	.94888
40.0	.96131	.96131	.96131	.96131	.96131
50.0	.96888	.96888	.96888	.96888	.96888
100.0	.98428	.98428	.98428	.98428	.98428
150.0	.98948	.98948	.98948	.98948	.98948

estimator in the neighborhood of $\lambda = 0$. Also, we see from Table 2 that the PT-2SHI estimator with $\tau = \tau_1$ has smaller MSE than the PT-2SHI estimator with $\tau = \tau_2$ over a wide region of λ though the MSE of the PT-2SHI estimator with $\tau = \tau_1$ is slightly larger than that of the PT-2SHI estimator with $\tau = \tau_2$ around $\lambda = 0$. Furthermore, the PT-2SHI estimator with $\tau = \tau_4$ dominates the PSR estimator. Though we do not show the results for the other cases, the PT-2SHI estimator with $\tau = \tau_4$ dominates the PSR estimator for the other values of k and n considered here.

5 Concluding remarks

Shao and Strawderman (1994) proposed a family of estimators which dominate the PSR estimator though they assumed that the disturbance variance is known. Sugiura and Takagi (1996) extended Shao and Strawderman's (1994) estimator to the case of unknown disturbance variance. However, relatively large sample size is needed for Sugiura and Takagi's estimator to dominate the PSR estimator since the dominance is valid when the condition $n \geq \max[9k + 10, 13k - 7]$ holds.

Also, Ohtani (1999) showed that the heterogeneous pre-test (HPT) estimator which consists of the AMMSE and SR estimators dominates the PSR estimator when $k = 3$. However, the dominance is not hold when $k \geq 4$.

Our numerical results show that the PT-2SHI estimator with the appropriate parameter values dominates the PSR estimator even when $k = 15$ and $n = 20$. Though they were found by trial and error, the parameter values we used in numerical evaluations are restricted and somewhat tentative. Thus, there may be more appropriate values such that the MSE performance of the PT-2SHI estimator is improved over a wide region of the parameter space. However, since it is beyond the scope of this paper to seek such values, it is a remaining problem for future research.

Appendix

First, we derive the formula for $H(p,q,r,\alpha;\tau)$. Let $u_1 = b'Sb/\sigma^2$ and $u_2 = e'e/\sigma^2$. Then, u_1 is distributed as the noncentral chi-square distribution with k degrees of freedom and noncentrality parameter $\lambda = \beta'S\beta/\sigma^2$, u_2 is distributed as the chi-square distribution with $\nu = n-k$ degrees of freedom, and u_1 and u_2 are independent.

Using u_1 and u_2, $H(p,q,r,\alpha;\tau)$ is expressed as

$$\begin{aligned} &H(p,q,r,\alpha;\tau)\\ &\quad= E\left[I(F\le\tau)\frac{(\sigma^2u_1)^p(\sigma^2u_2)^q}{(\sigma^2u_1+\alpha\sigma^2u_2)^r}\right]\\ &\quad= (\sigma^2)^{p+q-r}\sum_{i=0}^{\infty}K_i\iint_R u_1^{k/2+i-1}u_2^{\nu/2-1}\exp[-(u_1+u_2)/2]\frac{u_1^pu_2^q}{(u_1+\alpha u_2)^r}du_1\,du_2, \end{aligned} \tag{25}$$

where

$$K_i = \frac{w_i(\lambda)}{2^{(\nu+k)/2+i}\Gamma(\nu/2)\Gamma(k/2+i)}, \tag{26}$$

$w_i(\lambda) = \exp(-\lambda/2)(\lambda/2)^i/i\,!$, and R is the region such that $\{(u_1,u_2)|u_1\ge 0, u_2\ge 0 \text{ and } u_1/u_2 \le k\tau/\nu\}$.

Making use of the change of variables, $v_1 = u_1/u_2$ and $v_2 = u_2$, (25) reduces to:

$$(\sigma^2)^{p+q-r}\sum_{i=0}^{\infty}K_i\int_0^{\tau^{**}}\int_0^{\infty}\frac{v_1^{k/2+p+i-1}v_2^{(\nu+k)/2+p+q-r+i-1}}{(v_1+\alpha)^r}\exp[-(1+v_1)v_2/2]dv_2\,dv_1, \tag{27}$$

where $\tau^{**} = k\tau/\nu$.

Again, making use of the change of variable, $z = (1+v_1)v_2/2$, (27) reduces to:

$$\begin{aligned} &(\sigma^2)^{p+q-r}\sum_{i=0}^{\infty}K_i2^{(\nu+k)/2+p+q-r+i}\Gamma((\nu+k)/2+p+q-r+i)\\ &\qquad\times\int_0^{\tau^{**}}\frac{v_1^{k/2+p+i-1}}{(1+v_1)^{(\nu+k)/2+p+q-r+i}(v_1+\alpha)^r}dv_1. \end{aligned} \tag{28}$$

Further, making use of the change of variable, $t = v_1/(1+v_1)$, and substituting (26) in (28), (28) reduces to (14) in the text.

Next, we derive the formula for $J(p,q,r,\alpha;\tau)$. Differentiating $H(p,q,r,\alpha;\tau)$ given in (14) with respect to β, we obtain:

$$\begin{aligned}
&\frac{\partial H(p,q,r,\alpha;\tau)}{\partial\beta} \\
&\quad = (2\sigma^2)^{p+q-r}\sum_{i=0}^{\infty}\left[\frac{\partial w_i(\lambda)}{\partial\beta}\right]G_i(p,q,r,\alpha;\tau) \\
&\quad = (2\sigma^2)^{p+q-r}\sum_{i=0}^{\infty}\left[-\frac{S\beta}{\sigma^2}w_i(\lambda)+\frac{S\beta}{\sigma^2}w_{i-1}(\lambda)\right]G_i(p,q,r,\alpha;\tau) \\
&\quad = -(S\beta/\sigma^2)H(p,q,r,\alpha;\tau)+(S\beta/\sigma^2)(2\sigma^2)^{p+q-r}\sum_{i=0}^{\infty}w_i(\lambda)G_{i+1}(p,q,r,\alpha;\tau), \qquad (29)
\end{aligned}$$

where we define $w_{-1}(\lambda)=0$.

Expressing $H(p,q,r,\alpha;\tau)$ by $b'Sb$ and $e'e$, we have:

$$\begin{aligned}
H(p,q,r,\alpha;\tau) &= E\left[I(F\le\tau)\frac{(b'Sb)^p(e'e)^q}{(b'Sb+\alpha e'e)^r}\right] \\
&= \iint_{R'}\frac{(b'Sb)^p(e'e)^q}{(b'Sb+\alpha e'e)^r}f_N(b)f_e(e'e)db\,d(e'e), \qquad (30)
\end{aligned}$$

where $f_e(e'e)$ is the density function of $e'e$,

$$f_N(b)=\frac{1}{(2\pi)^{k/2}|\sigma^2S^{-1}|^{1/2}}\exp\left[-\frac{(b-\beta)'S(b-\beta)}{2\sigma^2}\right] \qquad (31)$$

is the density function of b, and R' is the region such that $\{(b,e)|(b'Sb/k)/(e'e/\nu)\le\tau\}$.

Differentiating $H(p,q,r,\alpha;\tau)$ given in (30) with respect to β, we obtain:

$$\frac{\partial H(p,q,r,\alpha;\tau)}{\partial\beta}=\frac{1}{\sigma^2}E\left[I(F\le\tau)\frac{(b'Sb)^p(e'e)^q}{(b'Sb+\alpha e'e)^r}Sb\right]-\frac{S\beta}{\sigma^2}H(p,q,r,\alpha;\tau). \qquad (32)$$

Equating (29) and (32), and multiplying β' from left, we obtain (15) in the text.

Acknowledgment

The author is grateful to Kazuhiro Ohtani for his very helpful guidance and to two anonymous referees and Michael D. Perlman for their very useful comments.

References

Baranchik, A.J. (1970), A family of minimax estimators of the mean of a multivariate normal distribution, *Annals of Mathematical Statistics*, 41, 642-645.

Farebrother, R.W. (1975), The minimum mean square error linear estimator and ridge regression, *Technometrics*, 17, 127-128.

James, W. and C. Stein (1961), Estimation with quadratic loss, *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability 1*, (Berkeley, University of California Press), 361-379.

Namba, A. (1998), Small sample properties of shrinkage estimators in econometrics, Master Thesis, Kobe University.

Ohtani, K. (1996), On an adjustment of degrees of freedom in the minimum mean squared error estimator, *Communications in Statistics-Theory and Methods*, 25, 3049-3058.

Ohtani, K. (1999), MSE performance of a heterogeneous pre-test estimator, *Statistics and Probability Letters*, 41, 65-71.

Shao, P.Y. and W.E. Strawderman (1994), Improving on the James-Stein positive-part estimator, *Annals of Statistics*, 22, 1517-1538.

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

Sugiura, N. and Y. Takagi (1996), Dominating James-Stein positive-part estimator for normal mean with unknown covariance matrix, *Communications in Statistics-Theory and Methods*, 25, 2875-2900.

Theil, H. (1971), *Principles of Econometrics*, (John Wiley, New York).

Tran Van Hoa and A. Chaturvedi (1990), Further results on the two-stage hierarchial information (2SHI) estimators in the linear regression models, *Communications in Statistics-Theory and Methods*, 19, 4697-4704.