MSE performance of the weighted average estimators consisting of shrinkage estimators

Namba, Akio

Ohtani, Kazuhiro

(Citation)
Communications in Statistics-Theory and Methods,47(5):1204-1214

(Issue Date)
2018

(Resource Type)
journal article

(Version)
Accepted Manuscript

MSE Performance of the Weighted Average Estimators Consisting of Shrinkage Estimators

Akio Namba* **and Kazuhiro Ohtani**†

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

Abstract

In this paper we consider a regression model and propose estimators which are the weighted averages of two estimators among three estimators; the Stein-rule (SR), the minimum mean squared error (MMSE) and the adjusted minimum mean squared error (AMMSE) estimators. It is shown that one of the proposed estimators has smaller mean squared error (MSE) than the positive-part Stein-rule (PSR) estimator over a moderate region of parameter space when the number of the regression coefficients is small (i.e., 3), and its MSE performance is comparable to the PSR estimator even when the number of the regression coefficients is not so small.

Key Words: Mean squared error, Stein-rule estimator, Minimum mean squared error estimator, Adjusted Minimum mean squared error estimator, weighted average estimator

*Akio Namba
Graduate School of Economics, Kobe University
Rokko, Nada-ku, Kobe, 657-8501, Japan
e-mail: namba@econ.kobe-u.ac.jp
†Kazuhiro Ohtani
Graduate School of Economics, Kobe University
Rokko, Nada-ku, Kobe, 657-8501, Japan
e-mail: kotani@econ.kobe-u.ac.jp

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 mean squared error (MSE) 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.

Also, as one of improved estimators, 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 (1996b) 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 the AMMSE estimator has smaller MSE than the SR and the PSR estimators in a wide region of the noncentrality parameter when $k \leq 5$, where k is the number of regression coefficients. 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 noncentrality parameter. Therefore, Ohtani (1999) considered a heterogeneous pre-test estimator such that the AMMSE estimator is used if the null hypothesis that all the regression coefficients are zeros (in other words, the value of 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. Extending the result of Ohtani (1999), Namba (2000) proposed another heterogeneous pre-test estimator and numerically showed that the proposed estimator has smaller MSE than the PSR estimator even when $k \geq 4$. However, since the estimators considered by Ohtani (1999) and Namba (2000) connect two different estimators via a pre-test based on the F-statistic, they are not smooth. This is a drawback of their estimators common to the PSR estimator because the smoothness is required for the estimator to be admissible.

In this paper, we propose estimators which are weighted averages of different kinds of estimators. The proposed estimators are smooth since they do not incorporate the pre-test. In the next section, we introduce the estimators, and derive an explicit formula for MSEs of the estimators in section 3. Using this formula, we examine the MSE performance of the estimators by numerical evaluations in section 4. Our numerical results show that one of our estimators has smaller MSE than the PSR estimator over a moderate region of parameter space. Finally, some concluding remarks are given in section 5.

2 Model and 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 $E[\epsilon\epsilon'] = \sigma^2 I_n$.

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

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

where $e = y - Xb$, and a is a constant such that $0 \le a \le 2(k-2)/(\nu+2)$, 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 $k \ge 3$. 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

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

As one of improved estimators, Theil (1971) proposed the minimum mean squared error (MMSE) estimator. However, since Theil's (1971) MMSE estimator includes unknown parameters, Farebrother (1975) suggested the following operational variant of the MMSE estimator

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

Hereafter, we call the operational variant of the MMSE estimator the MMSE estimator simply. There are several studies about the MMSE estimator and its variants. Some examples are Vinod (1976), Dwivedi and Srivastava (1978), Tracy and Srivastava (1994) and Ohtani (1996a, 1996b, 1997). Ohtani (1996a) derived the exact MSE of the MMSE estimator and a sufficient condition for the MMSE estimator to dominate the OLS estimator.

Furthermore, as an extension of the MMSE estimator, Ohtani (1996b) considered the following estimator which is obtained by adjusting the degrees of freedom of $b'Sb$ (i.e., k),

$$b_{\mathrm{AM}} = \left(\frac{b'Sb/k}{b'Sb/k + e'e/\nu} \right) b. \tag{7}$$

We call this estimator the adjusted MMSE (AMMSE) estimator. Ohtani (1996b) showed by numerical evaluations that if $k \leq 5$ the AMMSE estimator has smaller MSE than the SR, PSR and MMSE estimators in a wide region of noncentrality parameter defined as $\lambda = \beta' S \beta / \sigma^2$. Thus, Ohtani (1999) considered the following heterogeneous pre-test estimator:

$$\widehat{\beta}_{\mathrm{PT}}(\tau) = I(F \leq \tau) b_{\mathrm{AM}} + I(F > \tau) b_{\mathrm{SR}}, \tag{8}$$

where $F = (b'Sb/k)/(e'e/\nu)$ is the test statistic for $\mathrm{H}_0 : \beta = 0$ against $\mathrm{H}_1 : \beta \neq 0$, 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 values of the pre-test is chosen appropriately. Also, extending the result of Ohtani (1999), Namba (2000) proposed another pre-test estimator obtained by replacing the AMMSE estimator with the 2SHI estimator proposed by Tran Van Hoa and Chaturvedi (1990), and showed the dominance of the proposed estimator over the PSR estimator by numerical evaluations. Though the estimators proposed by Ohtani (1999) and Namba (2000) dominates the PSR estimator, they are not smooth because of the pre-test. This is a drawback of their estimators, since the smoothness is a necessary condition for the admissibility of an estimator.

Thus, in this paper, we propose the following two estimators which are weighted averages of the estimators introduced above:

$$\widehat{\beta}_{\mathrm{WA1}} = \left(\frac{F}{1+F} \right) b_{\mathrm{M}} + \left(1 - \frac{F}{1+F} \right) b_{\mathrm{AM}}, \tag{9}$$

$$\widehat{\beta}_{\mathrm{WA2}} = \left(\frac{F}{1+F} \right) b_{\mathrm{SR}} + \left(1 - \frac{F}{1+F} \right) b_{\mathrm{AM}}. \tag{10}$$

These estimators are smooth since they do not incorporate any pre-test. Similar to $\widehat{\beta}_{\mathrm{PT}}(\tau)$ given in (8), b_{AM} plays an important role in both $\widehat{\beta}_{\mathrm{WA1}}$ and $\widehat{\beta}_{\mathrm{WA2}}$ when the value of F is small.

In the next section, we derive the explicit formula for the MSEs of $\widehat{\beta}_{\mathrm{WA1}}$ and $\widehat{\beta}_{\mathrm{WA2}}$.

3 MSE of the estimators

In this section, we derive the formula for the MSEs of $\widehat{\beta}_{\mathrm{WA1}}$ and $\widehat{\beta}_{\mathrm{WA2}}$.

Noting that $b'Sb/e'e = kF/\nu$, substituting (3), (6) and (7) into (9) and (10), and conducting some calculations, we have

$$\widehat{\beta}_{\mathrm{WA1}} = \left[\frac{F^2}{(F+1)(F+1/k)} + \frac{F}{(1+F)^2}\right] b \tag{11}$$

and

$$\widehat{\beta}_{\mathrm{WA2}} = \left[\frac{F - a\nu/k}{F+1} + \frac{F}{(1+F)^2}\right] b. \tag{12}$$

Thus, the MSE of $\widehat{\beta}_{\mathrm{WA1}}$ is given by

$$\begin{aligned}\mathrm{MSE}[\widehat{\beta}_{\mathrm{WA1}}] &= \mathrm{E}[(\widehat{\beta}_{\mathrm{WA1}} - \beta)' S (\widehat{\beta}_{\mathrm{WA1}} - \beta)] \\ &= \mathrm{E}\left[\left\{\frac{F^4}{(F+1)^2(F+1/k)^2} + 2\frac{F^3}{(F+1)^3(F+1/k)} + \frac{F^2}{(F+1)^4}\right\} b'Sb\right] \\ &\quad -2\mathrm{E}\left[\left\{\frac{F^2}{(F+1)(F+1/k)} + \frac{F}{(F+1)^2}\right\}\beta' Sb\right] + \beta' S\beta. \end{aligned} \tag{13}$$

Similarly,

$$\begin{aligned}\mathrm{MSE}[\widehat{\beta}_{\mathrm{WA2}}] &= \mathrm{E}\left[\left\{\frac{1}{(F+1)^2(F-a\nu/k)^{-2}} + 2\frac{F}{(F+1)^3(F-a\nu/k)^{-1}} + \frac{F^2}{(F+1)^4}\right\} b'Sb\right] \\ &\quad -2\mathrm{E}\left[\left\{\frac{F^2}{(F+1)(F-a\nu/k)^{-1}} + \frac{F}{(F+1)^2}\right\}\beta' Sb\right] + \beta' S\beta. \end{aligned} \tag{14}$$

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

$$H(p,q,r,m;\alpha) = \mathrm{E}\left[\frac{F^r}{(F+1)^p(F+\alpha)^q}(b'Sb)^m\right] \tag{15}$$

and

$$J(p,q,r,m;\alpha) = \mathrm{E}\left[\frac{F^r}{(F+1)^p(F+\alpha)^q}(b'Sb)^m \beta' Sb\right]. \tag{16}$$

Then, we can express the MSEs of $\widehat{\beta}_{\mathrm{WA1}}$ and $\widehat{\beta}_{\mathrm{WA2}}$ as

$$\begin{aligned}\mathrm{MSE}[\widehat{\beta}_{\mathrm{WA1}}] &= H(2,2,4,1;1/k) + 2H(3,1,3,1;1/k) + H(4,0,2,1;1/k) \\ &\quad -2J(1,1,2,0;1/k) - 2J(2,0,1,0;1/k) + \beta' S\beta \end{aligned} \tag{17}$$

and

$$\begin{aligned}\mathrm{MSE}[\widehat{\beta}_{\mathrm{WA1}}] &= H(2,-2,0,1;-a\nu/k) + 2H(3,-1,1,1;-a\nu/k) + H(4,0,2,1;-a\nu/k) \\ &\quad -2J(1,-1,0,0;-a\nu/k) - 2J(2,0,1,0;-a\nu/k) + \beta' S\beta. \end{aligned} \tag{18}$$

As is shown in Appendix A, explicit formulae for $H(p,q,r,m;\alpha)$ and $J(p,q,r,m;\alpha)$ are given by

$$H(p,q,r,m;\alpha) = (2\sigma^2)^m \left(\frac{k}{\nu}\right)^{p+q-r} \sum_{i=0}^{\infty} w_i(\lambda) G_i(p,q,r,m;\alpha), \tag{19}$$

$$J(p,q,r,m;\alpha) = \beta' S\beta (2\sigma^2)^m \left(\frac{k}{\nu}\right)^{p+q-r} \sum_{i=0}^{\infty} w_i(\lambda) G_{i+1}(p,q,r,m;\alpha), \tag{20}$$

where $w_i(\lambda) = \exp(-\lambda/2)(\lambda/2)/i!$ and

$$G_i(p,q,r,m;\alpha) = \frac{\Gamma((\nu+k)/2+m+i)}{\Gamma(\nu/2)\Gamma(k/2+i)} \int_0^1 \frac{t^{k/2+m+r+i-1}(1-t)^{\nu/2+p+q-r-1}}{[t+k(1-t)/\nu]^p[t+\alpha k(1-t)/\nu]^q} dt. \tag{21}$$

Substituting (19) and (20) into (17) and (18), we obtain explicit formulae for the MSEs of $\widehat{\beta}_{\mathrm{WA1}}$ and $\widehat{\beta}_{\mathrm{WA2}}$.

Since further theoretical analysis of the MSEs of $\widehat{\beta}_{\mathrm{WA1}}$ and $\widehat{\beta}_{\mathrm{WA2}}$ is difficult, we execute numerical evaluation in the next section.

4 Numerical results

In this section, we compare the MSEs of estimators by numerical evaluations[1].

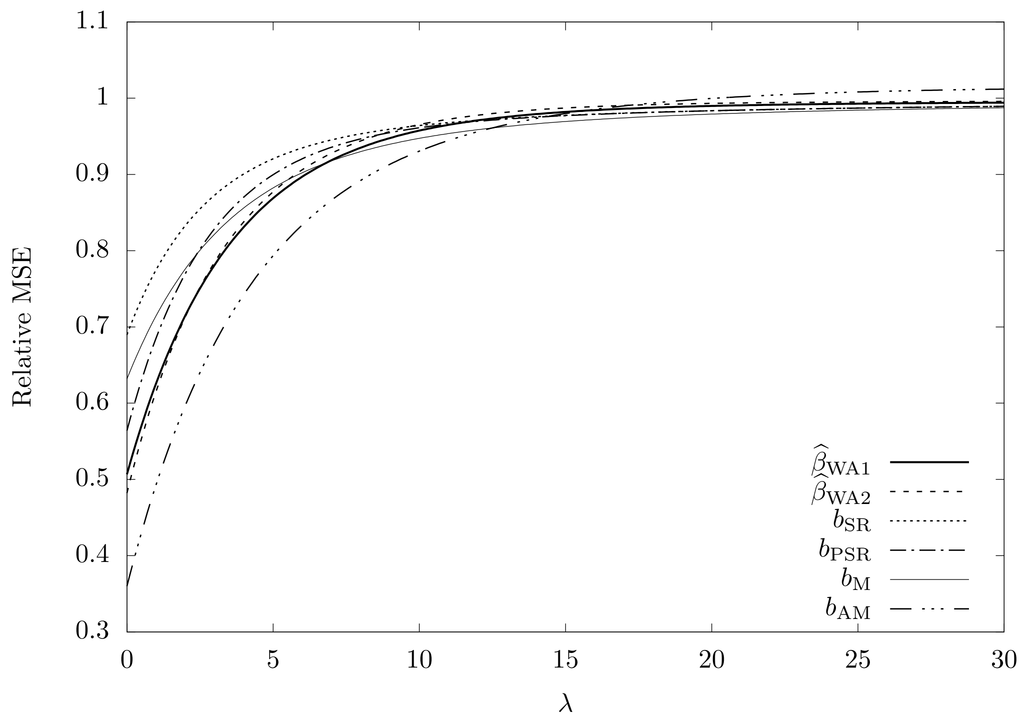
The parameter values used in the numerical evaluations were k = 3, 4, 5, 8, 10, n = 20, 30, 50, and λ = various values. The numerical evaluations were executed on a personal computer, using the FORTRAN code. In evaluating the integral in $G_i(p,q,r,m;\alpha)$ given in (21), we used Simpson's rule with 500 equal subdivisions. The infinite series in $H(p,q,r,m;\alpha)$ and $J(p,q,r,m;\alpha)$ were judged to converge when the increment of the series gets smaller than 10^{-12}. In order to compare the MSE performances of the estimators, we evaluated the values of relative MSE defined as $\mathrm{MSE}(\bar{\beta})/\mathrm{MSE}(b)$, 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. Since the results for $k = 3$, $k = 5$ and $n = 30$ are qualitatively typical, we show the results only for these cases.

Figure 1 shows the results for $k = 3$ and $n = 30$. We can see that all estimators considered here except for b_{AM} dominate the OLS estimator b in terms of MSE, and $\widehat{\beta}_{\mathrm{WA2}}$ has smaller MSE than the other estimators considered here except for b_{AM} when λ is small (i.e., $\lambda \leq 2.0$). Moreover, $\widehat{\beta}_{\mathrm{WA2}}$ has smaller MSE than b_{PSR} when λ is small and moderate (i.e., $\lambda \leq 8.0$). Though b_{AM} has smallest MSE among all the estimators when λ is small and moderate (i.e., $\lambda \leq 12.0$), it has slightly larger MSE than the OLS estimator for large values of λ (i.e., $\lambda \geq 22.0$). Also, though $\widehat{\beta}_{\mathrm{WA2}}$ has smaller MSE than $\widehat{\beta}_{\mathrm{WA1}}$ around $\lambda = 0$, the region where $\widehat{\beta}_{\mathrm{WA1}}$ has smaller MSE than b_{PSR} (i.e., $\lambda \leq 10.0$) is slightly wider than that of $\widehat{\beta}_{\mathrm{WA2}}$ (i.e., $\lambda \leq 8.0$). Comparing $\widehat{\beta}_{\mathrm{WA1}}$ and $\widehat{\beta}_{\mathrm{WA2}}$, the MSE of $\widehat{\beta}_{\mathrm{WA2}}$ around $\lambda = 0$ is smaller than that of $\widehat{\beta}_{\mathrm{WA1}}$ while the MSE of $\widehat{\beta}_{\mathrm{WA2}}$ is slightly larger than that of $\widehat{\beta}_{\mathrm{WA1}}$ for $\lambda \geq 3.0$. Comparing the gain and the loss of MSE, $\widehat{\beta}_{\mathrm{WA2}}$ may be preferred to b_{PSR} and $\widehat{\beta}_{\mathrm{WA1}}$.

Figure 2 shows the results for $k = 5$ and $n = 30$. Comparing Figures 1 and 2, we see that when k increases from 3 to 5, the MSE performance changes largely. Though all estimators considered here

[1] As is suggested by an anonymous referee, we also calculated the squared lengths of bias of the estimators under the simplified assumptions. See Appendix B for details.

Figure 1: MSEs for $k = 3$ and $n = 30$.



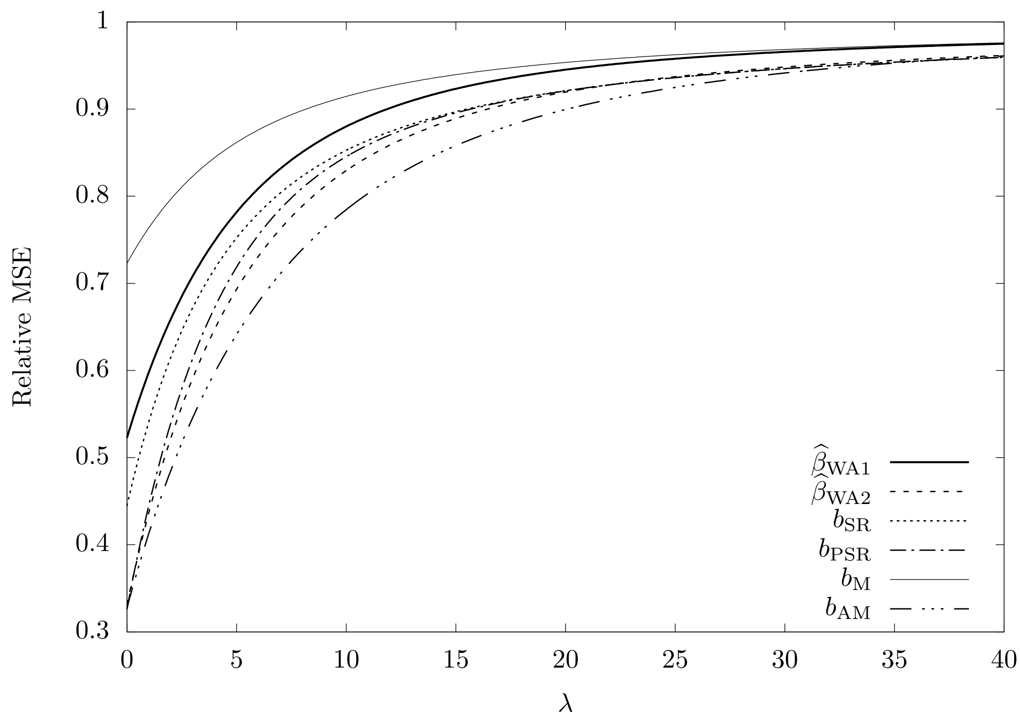
dominate the OLS estimator, $\widehat{\beta}_{\mathrm{WA1}}$ is dominated by b_{PSR}, and both $\widehat{\beta}_{\mathrm{WA1}}$ and $\widehat{\beta}_{\mathrm{WA2}}$ no longer have smaller MSE than b_{PSR} around $\lambda = 0$. Also, $\widehat{\beta}_{\mathrm{WA2}}$ dominates $\widehat{\beta}_{\mathrm{WA1}}$ and b_{M}, and $\widehat{\beta}_{\mathrm{WA2}}$ has smaller MSE than b_{PSR} for $0.4 \leq \lambda \leq 22.0$. This indicates that the MSE performances of two estimators, $\widehat{\beta}_{\mathrm{WA2}}$ and b_{PSR}, are comparable.

When it is difficult to derive an exact distribution of an estimator, the bootstrap methods proposed by Efron (1979) have often been applied. (Some examples are Chi and Judge (1985), Brownstone (1990) and Yi (1991).) However, as is shown in Zaman (1996), the bootstrap methods are not so valid when an estimator is not smooth. Since the MSE performances of $\widehat{\beta}_{\mathrm{WA2}}$ and b_{PSR} are comparable and $\widehat{\beta}_{\mathrm{WA2}}$ is smooth while b_{PSR} is not smooth, we can say that $\widehat{\beta}_{\mathrm{WA2}}$ can be a good smooth alternative to the PSR estimator, in particular when we apply the bootstrap methods.

5 Concluding remarks

In this paper, we proposed estimators for regression coefficients which are weighted averages of two shrinkage estimators. Our numerical results show that the estimator $\widehat{\beta}_{\mathrm{WA2}}$ which is a weighted average of the SR and the AMMSE estimators has smaller MSE than the PSR estimator over a moderate region

Figure 2: MSEs for $k = 5$ and $n = 30$.



of parameter space when $k = 3$. Even when $k > 3$, $\widehat{\beta}_{\mathrm{WA2}}$ has comparable MSE performance to the PSR estimator b_{PSR}. Also, the proposed estimators $\widehat{\beta}_{\mathrm{WA1}}$ and $\widehat{\beta}_{\mathrm{WA2}}$ have smaller MSE than the OLS estimator for all parameter values considered in this paper. Moreover, because of their structure, the proposed estimators are smooth. Compared with the PSR estimator, this is the virtue of proposed estimators.

Considering this result, we may able to construct smooth estimators which have more preferable performance by taking weighted averages of possibly some other estimators. However, seeking for such estimators is beyond the scope of this research and a remaining problem for future research.

Appendix A

First, we derive the formula for $H(p, q, r, m; \alpha)$. 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 $\lambda = \beta'S\beta/\sigma^2$, and u_2 is distributed as the chi-square distribution with $\nu = n - k$ degrees of freedom.

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

$$H(p,q,r,m;\alpha) = \mathrm{E}\left[\frac{(\frac{\nu}{k}\frac{u_1}{u_2})^r}{(\frac{\nu}{k}\frac{u_1}{u_2}+1)^p(\frac{\nu}{k}\frac{u_1}{u_2}+\alpha)^q}(\sigma^2 u_1)^m\right]$$

$$= \nu^r k^{p+q-r} (\sigma^2)^m \sum_{i=0}^{\infty} K_i$$

$$\times \int_0^\infty \int_0^\infty \frac{u_1^{k/2+m+r+i-1} u_2^{\nu/2+p+q-r-1}}{(\nu u_1 + k u_2)^p (\nu u_1 + \alpha k u_2)^q} \exp[-(u_1+u_2)/2] du_1\, du_2, \tag{22}$$

where

$$K_i = \frac{w_i(\lambda)}{2^{(\nu+k)/2+i}\Gamma(\nu/2)\Gamma(k/2+i)}, \tag{23}$$

and $w_i(\lambda) = \exp(-\lambda/2)(\lambda/2)^i/i\,!$.

Making use of the change of variables, $v_1 = u_1/u_2$ and $v_2 = u_2$, the integral in (22) reduces to

$$\int_0^\infty \int_0^\infty \frac{v_1^{k/2+m+r+i-1} v_2^{(\nu+k)/2+m+i-1}}{(\nu v_1 + k)^p (\nu v_1 + \alpha k)^q} \exp[-(v_1+1)/2] dv_1\, dv_2. \tag{24}$$

Again, making use of the change of variable, $z = (1+v_1)v_2/2$, (24) reduces to

$$2^{(\nu+k)/2+m+1}\Gamma((\nu+k)/2+m+i) \int_0^\infty \frac{v_1^{k/2+m+r+i-1}}{(\nu v_1 + k)^p (\nu v_1 + \alpha k)^q (v_1+1)^{(\nu+k)/2+m+1}} dv_1. \tag{25}$$

Further, making use of the change of variable, $t = v_1/(1+v_1)$, and substituting (25) in (22) and performing some manipulations, (22) reduces to (19) in the text.

Next, we derive the formula for $J(p,q,r,m;\alpha)$. Differentiating $H(p,q,r,\alpha)$ given in (19) with respect to β, we obtain

$$\begin{aligned}
\frac{\partial H(p,q,r,\alpha)}{\partial \beta} &= (2\sigma^2)^m \left(\frac{k}{\nu}\right)^{p+q-r} \sum_{i=0}^{\infty} \left[\frac{\partial w_i(\lambda)}{\partial \beta}\right] G_i(p,q,r,m;\alpha) \\
&= (2\sigma^2)^m \left(\frac{k}{\nu}\right)^{p+q-r} \sum_{i=0}^{\infty} \left[-\frac{S\beta}{\sigma^2} w_i(\lambda) + \frac{S\beta}{\sigma^2} w_{i-1}(\lambda)\right] G_i(p,q,r,m;\alpha) \\
&= -\frac{S\beta}{\sigma^2} H(p,q,r,m;\alpha) + \frac{S\beta}{\sigma^2}(2\sigma^2)^m \left(\frac{k}{\nu}\right)^{p+q-r} \sum_{i=0}^{\infty} w_i(\lambda) G_{i+1}(p,q,r,m;\alpha),
\end{aligned} \tag{26}$$

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

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

$$\begin{aligned}
H(p,q,r,m;\alpha) &= \mathrm{E}\left[\frac{(\frac{\nu}{k}\frac{b'Sb}{e'e})^r}{(\frac{\nu}{k}\frac{b'Sb}{e'e}+1)^p(\frac{\nu}{k}\frac{b'Sb}{e'e}+\alpha)^q}(b'Sb)^m\right] \\
&= \int_0^\infty \int_{-\infty}^\infty \frac{(\frac{\nu}{k}\frac{b'Sb}{e'e})^r}{(\frac{\nu}{k}\frac{b'Sb}{e'e}+1)^p(\frac{\nu}{k}\frac{b'Sb}{e'e}+\alpha)^q}(b'Sb)^m f_N(b) f_e(e'e) db\, d(e'e),
\end{aligned} \tag{27}$$

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

$$f_N(b) = \frac{1}{(2\pi)^{k/2}|\sigma^2 S^{-1}|^{1/2}} \exp\left[-\frac{(b-\beta)'S(b-\beta)}{2\sigma^2}\right]. \tag{28}$$

Differentiating $H(p,q,r,m;\alpha)$ given in (27) with respect to β, we obtain:

$$\frac{\partial H(p,q,r,m;\alpha)}{\partial \beta} = \frac{1}{\sigma^2}\mathrm{E}\left[\frac{(\frac{\nu}{k}\frac{b'Sb}{e'e})^r}{(\frac{\nu}{k}\frac{b'Sb}{e'e}+1)^p(\frac{\nu}{k}\frac{b'Sb}{e'e}+\alpha)^q}(b'Sb)^m Sb\right] - \frac{S\beta}{\sigma^2} H(p,q,r,m;\alpha). \tag{29}$$

Equating (26) and (29), and multiplying β' from left, we obtain (20) in the text.

Appendix B

In this Appendix, we explain the way to calculate the squared lengths of bias of the estimators under some simplified conditions. In order to make the calculation tractable, we assume $S = X'X = I$, $\beta = \overline{\beta}[1, 1, \dots, 1]' = \overline{\beta}\iota$, where $\overline{\beta}$ is a scalar and ι is a $k \times 1$ column vector whose elements are 1s. In addition, we assume $\sigma^2 = 1$ for simplicity. These conditions are assumed only when we calculate the the squared lengths of bias of the estimators. Under these assumptions, we have

$$\lambda = \beta' S \beta / \sigma^2 = \overline{\beta}^2 \iota'\iota = k\overline{\beta}^2.$$

Also, equation (16) can be transformed as follows

$$\begin{aligned} J(p,q,r,0;\alpha) &= \mathrm{E}\left[\frac{F^r}{(F+1)^p(F+\alpha)^q}\beta' S b\right] \\ &= \mathrm{E}\left[\frac{F^r}{(F+1)^p(F+\alpha)^q}\beta' b\right] \\ &= \overline{\beta}\iota'\mathrm{E}\left[\frac{F^r}{(F+1)^p(F+\alpha)^q} b\right] \end{aligned}$$

under the above mentioned conditions. Moreover, under the same conditions, all the k elements of $\frac{F^r}{(F+1)^p(F+\alpha)^q} b$ are distributed as the same distribution, and accordingly, all the elements of $\mathrm{E}\left[\frac{F^r}{(F+1)^p(F+\alpha)^q} b\right]$ are the same. Thus, if we denote the elements of $\mathrm{E}\left[\frac{F^r}{(F+1)^p(F+\alpha)^q} b\right]$ by ξ, we have:

$$J(p,q,r,0;\alpha) = k\overline{\beta}\xi$$

Similarly, if we denote the elements of $\mathrm{E}\left[\frac{F^{r'}}{(F+1)^{p'}(F+\alpha')^{q'}} b\right]$ as ζ, we have

$$J(p',q',r',0;\alpha') = k\overline{\beta}\zeta$$

Thus, we obtain

$$\begin{aligned} \left\{\mathrm{E}\left[\frac{F^r}{(F+1)^p(F+\alpha)^q} b\right]\right\}' \mathrm{E}\left[\frac{F^{r'}}{(F+1)^{p'}(F+\alpha')^{q'}} b\right] &= k\xi\zeta \\ &= \frac{J(p,q,r,0;\alpha)J(p',q',r',0;\alpha')}{k\overline{\beta}^2}. \end{aligned}$$

Since the bias of $\widehat{\beta}_{\mathrm{WA1}}$ is

$$\mathrm{Bias}[\widehat{\beta}_{\mathrm{WA1}}] = \mathrm{E}\left[\frac{F^2}{(F+1)(F+1/k)} b\right] + \mathrm{E}\left[\frac{F}{(F+1)^2} b\right] - \beta,$$

the squared length of the bias is given by

$$\begin{aligned} \mathrm{Bias}[\widehat{\beta}_{\mathrm{WA1}}]'\mathrm{Bias}[\widehat{\beta}_{\mathrm{WA1}}] &= \left\{\mathrm{E}\left[\frac{F^2}{(F+1)(F+1/k)} b\right]\right\}' \mathrm{E}\left[\frac{F^2}{(F+1)(F+1/k)} b\right] \\ &\quad + \left\{\mathrm{E}\left[\frac{F}{(F+1)^2} b\right]\right\}' \mathrm{E}\left[\frac{F}{(F+1)^2} b\right] + \beta'\beta \end{aligned}$$

$$
\begin{aligned}
&+2\left\{\mathrm{E}\left[\frac{F^2}{(F+1)(F+1/k)}b\right]\right\}'\mathrm{E}\left[\frac{F}{(F+1)^2}b\right] \\
&-2\beta'\mathrm{E}\left[\frac{F^2}{(F+1)(F+1/k)}b\right] \\
&-2\beta'\mathrm{E}\left[\frac{F}{(F+1)^2}b\right] \\
=\ & J(1,1,2,0;1/k)J(1,1,2,0;1/k)/k\overline{\beta}^2 \\
&+J(2,0,1,0;1/k)J(2,0,1,0;1/k)/k\overline{\beta}^2+k\overline{\beta}^2 \\
&+2J(1,1,2,0;1/k)J(2,0,1,0;1/k)/k\overline{\beta}^2 \\
&-2J(1,1,2,0;1/k)-2J(2,0,1,0;1/k).
\end{aligned}
$$

The squared lengths of the bias vectors of the other estimators are obtained in a similar way.

Using the formula obtained above, we calculated the squared bias lengths of the estimators defined as

$$[\mathrm{Bias}[\tilde{\beta}]'\mathrm{Bias}[\tilde{\beta}]/\mathrm{MSE}[b],$$

where, $\tilde{\beta}$ is any estimator for β. The results for $k=3$ and 5, $n=30$, and various values of $\lambda=k\overline{\beta}^2$ are given in Figures 3 and 4. From the Figures, we can see:

1. The bias of the MMSE estimator is smaller than that of the AMMSE estimator. As a whole, the bias of the MMSE estimator is small compared with those of the other estimators.

2. The SR estimator has smaller bias than the AMMSE estimator for $k=3$, however, the former has larger bias than the latter for $k=5$.

3. Generally, the bias of $\widehat{\beta}_{\mathrm{WA1}}$ exists between those of the MMSE and the AMMSE estimators, and the bias of $\widehat{\beta}_{\mathrm{WA1}}$ exists between those of the AMMSE and the SR estimators.

The bias of an estimator is mostly determined by the shrink factor which is a functions of F. Also, F will be large when λ is large. Further, $\widehat{\beta}_{\mathrm{WA1}}$ is an weighted average of b_{M} and b_{AM}, and $\widehat{\beta}_{\mathrm{WA2}}$ is an weighted average of b_{SR} and b_{AM}. From these facts, we can consider most of the above results on bias vectors are as expected.

Acknowledgment

The authors are grateful for anonymous referees for helpful comments and suggestions. This work was supported by JSPS KAKENHI Grant Number 23243038, 26780136.

Figure 3: Squared lengths of bias of the estimators for $k = 3$ and $n = 30$.

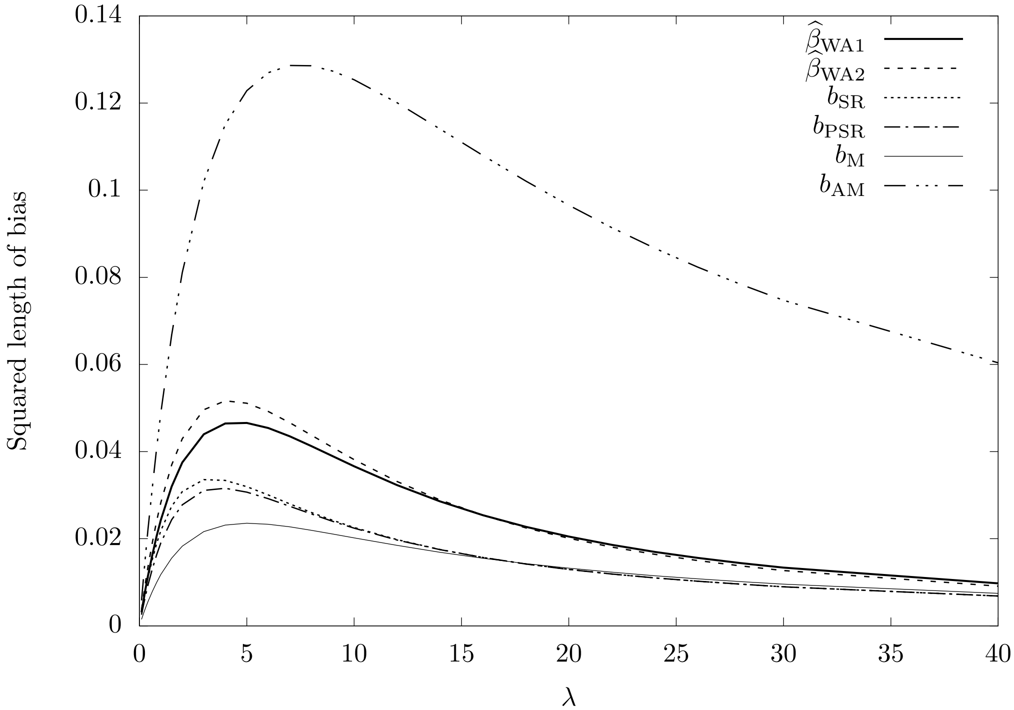
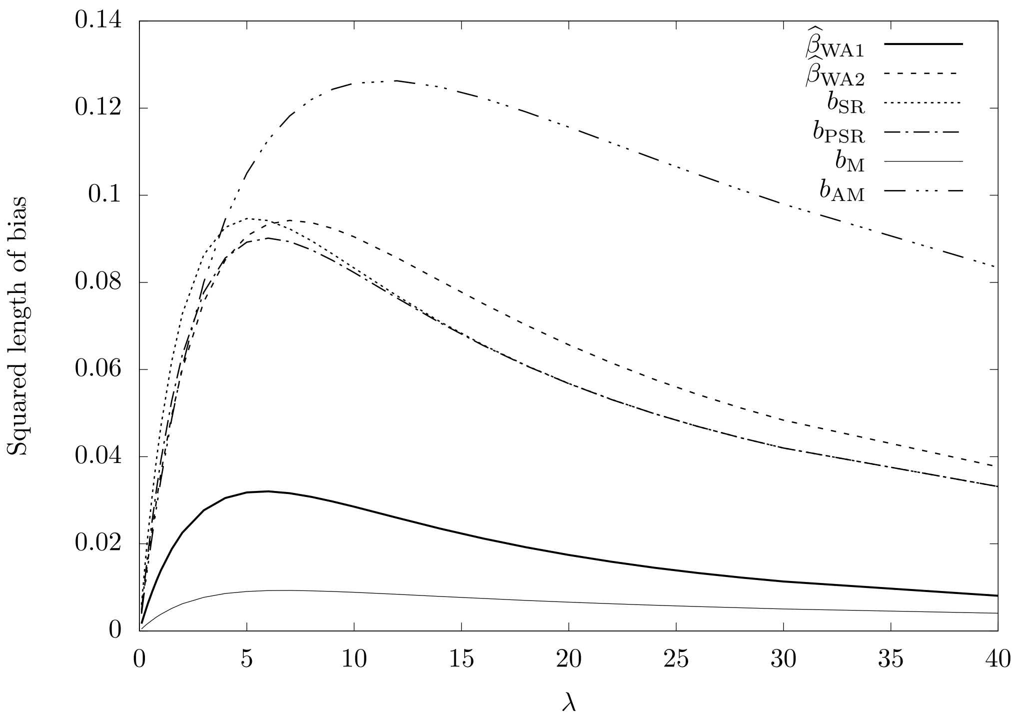


Figure 4: Squared lengths of bias of the estimators for $k = 5$ and $n = 30$.



References

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

Brownstone, D. (1990) Bootstrapping improved estimators for linear regression models. *Journal of Econometrics* 44, 171–187.

Chi, X. W. & G. Judge (1985) On assessing the precision of Stein's estimator. *Economics Letters* 18, 143–148.

Dwivedi, T. & V. Srivastava (1978) On the minimum mean squared error estimators in a regression model. *Communications in Statistics-Theory and Methods* 7, 487–494.

Efron, B. (1979) Bootstrap methods: Another look at the jackknife. *Annals of Statistics* 7, 1–26.

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

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

Namba, A. (2000) MSE dominance of the PT-2SHI estimator over the positive-part Stein-rule estimator in regression. *Journal of Statistical Planning and Inference* 89, 175–185.

Ohtani, K. (1996a) Exact small sample properties of an operational variant of the minimum mean squared error estimator. *Communications in statistics-Theory and methods* 25, 1223–1231.

Ohtani, K. (1996b) 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. (1997) Minimum mean squared error estimation of each individual coefficient in a linear regression model. *Journal of Statistical Planning and Inference* 62, 301–316.

Ohtani, K. (1999) MSE performance of a heterogeneous pre-test estimator. *Statistics & probability letters* 41, 65–71.

Stein, C. (1956) Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, Volume 1, pp. 197–206. University of California Press, Berkeley.

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

Tracy, D. S. & A. K. Srivastava (1994) Comparison of operational variants of best homogeneous and heterogeneous estimators in linear regression. *Communications in statistics-Theory and Methods* 23, 2313–2322.

Van Hoa, T. & A. Chatruvedi (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.

Vinod, H. (1976) Simulation and extension of a minimum mean squared error estimator in comparison with Stein's. *Technometrics* 18, 491–496.

Yi, G. (1991) Estimating the variability of the Stein estimator by bootstrap. *Economics Letters* 37, 293–298.

Zaman, A. (1996) *Statistical Foundations for Econometric Techniques*, Academic Press.