IMPROVING THE ‘HKB’ ORDINARY TYPE RIDGE ESTIMATOR

Takakatsu Inoue *

The HKB estimator of Hoerl, Kennard and Baldwin is known to be an ordinary ridge type shrinkage estimator and an adjustment of degrees of freedom in the ordinary ridge estimator of Farebrother. The HKB estimator has a smaller predictive mean squared error (MSE) than the positive-part Stein-rule (PP) estimator in the wide region of the noncentral parameter when the degrees of freedom $q = 3 \sim 6$, but does not satisfy the sufficient condition to dominate the OLS estimator of Baranchik or Efron and Morris when $q = 3$. In this paper, a sufficient condition to dominate the OLS estimator is derived, and the modified HKB estimator is constructed to dominate the OLS estimator and have a smaller MSE than the PP estimator in the wide region of the noncentral parameter.

Key words and phrases: ridge regression, MSE dominance condition, HKB estimator, positive-part Stein-rule estimator, predictive mean squared error

1. Introduction

Given sample $\{(y_s, x_s); s = 1, 2, \ldots, n\}$, consider the following linear regression model:

$$y = X\beta + \epsilon, \quad \epsilon \sim N(0, \sigma^2 I_n),$$

(1.1)

where $y = (y_1, \ldots, y_s, \ldots, y_n)'$ is an objective variable vector, and $X$ is a known $n \times q$ design matrix with full column rank. $\beta = (\beta_1, \ldots, \beta_q)'$ is an unknown regression coefficient vector and $\epsilon$ is an $n \times 1$ vector following the $n$-variates normal distribution with mean $0$ and dispersion matrix $\sigma^2 I_n$, i.e., $\epsilon \sim N(0, \sigma^2 I_n)$.

Under the model (1.1), the Ordinary Least Squares (OLS) estimator is $\hat{\beta} = (X'X)^{-1}X'y$ and the distribution of $\hat{\beta}$ is $\hat{\beta} \sim N(\beta, \sigma^2(X'X)^{-1})$.

Farebrother (1975) showed that the estimator

$$\beta_F = \frac{\hat{\beta}'X'X\hat{\beta}}{\hat{\beta}'X'X\hat{\beta} + \sigma^2 \hat{\beta}}$$

(1.2)

is the best linear homogeneous estimator of $\beta$, and the predictive mean squared error $MSE(\beta_F)$ is

$$MSE(\beta_F) = E\{(\beta_F - \beta)'X'X(\beta_F - \beta)\}$$

(1.3)

$$= (1 - \phi)^2 \beta'X'X\beta + \phi^2 q\sigma^2,$$

where $E$ is the expectation over the sample $y$, $\phi = \beta'X'X\beta/(\beta'X'X\beta + \sigma^2)$, and $\phi$ is considered to be the shrinkage parameter in population. However,
the estimator $\mathbf{\beta}_F$ includes the unknown parameters $\mathbf{\beta}$ and $\sigma^2$. Farebrother (1975) suggested an operational estimator

$$(1.4) \quad \hat{\mathbf{\beta}}_F = \frac{\hat{\mathbf{\beta}}' X' X \hat{\mathbf{\beta}}}{\hat{\mathbf{\beta}}' X' X \hat{\mathbf{\beta}} + \hat{\sigma}^2} = \hat{\phi}_F \hat{\mathbf{\beta}},$$

where $\hat{\sigma}^2 = (\mathbf{y} - \mathbf{X}\hat{\mathbf{\beta}})'(\mathbf{y} - \mathbf{X}\hat{\mathbf{\beta}})/(n - q)$ and $\hat{\phi}_F = \hat{\mathbf{\beta}}' X' X \hat{\mathbf{\beta}} / \{ \hat{\mathbf{\beta}}' X' X \hat{\mathbf{\beta}} + \hat{\sigma}^2 \}$ is the estimator of the shrinkage factor $\phi$. Denoting $F = \{ \hat{\mathbf{\beta}}' X' \hat{\mathbf{\beta}} / \hat{\sigma}^2 \}$, the estimator of the shrinkage factor is written as

$$(1.5) \quad \hat{\phi}_F = \frac{F}{F + 1/q},$$

where $F$ is the $F$ statistic with the degrees of freedom $(q, \nu = n - q)$ and the noncentral parameter $\lambda = \mathbf{\beta}' X' X \mathbf{\beta} / \hat{\sigma}^2$, i.e., $F \sim F(q, \nu, \lambda)$.

Similarly, a modified estimator of Hoerl, Kennard and Baldwin (1975) (see Vinod and Ullah (1981)) is written as

$$(1.6) \quad \hat{\mathbf{\beta}}_{HKB} = \hat{\phi}_{HKB} \hat{\mathbf{\beta}}, \quad \hat{\phi}_{HKB} = \frac{F}{F + 1/q}.$$  

This estimator was also taken up in Ohtani (1996) and was called 'the adjusted minimum mean squared error estimator'. We call it 'HKB' estimator.

The estimator of James and Stein (1961) is written as

$$(1.7) \quad \hat{\mathbf{\beta}}_{JS} = \hat{\phi}_{JS} \hat{\mathbf{\beta}}, \quad \hat{\phi}_{JS} = 1 - \frac{\nu}{\nu - 2} \cdot \frac{q - 2}{q} \cdot \frac{1}{F} \quad (q \geq 3),$$

and the positive part Stein-rule estimator is written as

$$(1.8) \quad \hat{\mathbf{\beta}}_{PP} = \hat{\phi}_{PP} \hat{\mathbf{\beta}}, \quad \hat{\phi}_{PP} = \begin{cases} \hat{\phi}_{JS} & (F > \frac{\nu}{\nu - 2} \cdot \frac{q - 2}{q} \quad \text{and} \quad q \geq 3) \\ 0 & (F \leq \frac{\nu}{\nu - 2} \cdot \frac{q - 2}{q} \quad \text{and} \quad q \geq 3) \end{cases}. $$

Hereafter, we call the estimator $\hat{\mathbf{\beta}}_{JS}$ of (1.7) the 'JS' estimator and call the estimator $\hat{\mathbf{\beta}}_{PP}$ of (1.8) the 'PP' estimator. James and Stein (1961) showed that the JS estimator dominates the OLS estimator in terms of predictive mean squared error, while, Baranchik (1970) showed that the PP estimator dominates the JS estimator. Furthermore, some modified PP estimators have been proposed and investigated (Judge and Bock (1978), Shao and Strawderman (1994,1996), Gruber (1998)).

Ohtani (1996b) derived the exact MSE of the HKB estimator, and showed that the HKB estimator has a smaller predictive mean squared error than the ordinary ridge estimator of Farebrother (1975) or the PP estimator in the wide region of the noncentral parameter (for $q = 2, 4,$ and 6).

Inoue (2000) showed that the HKB estimator has a smaller MSE than the PP estimator in a wide region of $\lambda$ when $q = 3 \sim 6$, but a larger MSE.
than the PP estimator when \( q > 6 \). Furthermore the modified HKB (we call it the ‘M’ estimator)
\[
\hat{\beta}_M = \frac{F^b}{a + F^b}
\]
was investigated, and showed that the M estimator has a smaller MSE than the HKB estimator or the PP estimator in a wide region of \( \lambda \). However, it was shown that the M estimator does not satisfy the sufficient condition of Baranchik (1970) or Efron and Morris (1976) to dominate the OLS estimator when \( q = 3 \sim 6 \).

In the shrinkage estimators taken the above, the corresponding shrinkage factor estimator \( \hat{\phi} \) is expressed as a function of \( F \) statistic. We then consider the shrinkage estimator
\[
\tilde{\beta} = \tilde{\phi}(F)\hat{\beta},
\]
(1.10)
where \( \tilde{\phi}(F) \) is an arbitrary function of \( F \) statistic.

In this paper, we show the exact predictive mean squared error \( MSE(\tilde{\beta}) \) in Section 2, and derive a sufficient condition for the estimator \( \tilde{\beta} = \tilde{\phi}(F)\hat{\beta} \) to dominate the OLS estimator \( \hat{\beta} \) in the terms of the predictive MSE by expanding that of Baranchik (1970) in Section 3. Based on the sufficient condition, we show the region of \( (a, b) \) in the M estimator to dominate the OLS estimator in Section 4, and evaluate the efficiency of the M estimator by numerical analysis in Section 5.

2. Predictive mean squared error

For a convenience of notation, we denote the eigen value decomposition of \( X'X = G\Lambda G' \), where \( G \) is the orthogonal matrix of order \( q \) and \( \Lambda = \text{diag}\{\lambda_j\} \) is the diagonal matrix with the \( j \)-th element \( \lambda_j > 0 \) \((j = 1, 2, \ldots, q)\). We also denote \( \alpha = \Lambda^{1/2}G'\beta, \hat{\alpha} = \Lambda^{1/2}G'\hat{\beta}, \tilde{\alpha} = \Lambda^{1/2}G'\tilde{\beta} \). Then the predictive mean squared error \( MSE(\tilde{\beta}) \) is expressed as
\[
MSE(\tilde{\beta}) = E\{(\tilde{\beta} - \beta)'X'X(\tilde{\beta} - \beta)\}
= E\{ (\tilde{\alpha} - \alpha)'(\tilde{\alpha} - \alpha) \}
= E\{ \tilde{\phi}^2\tilde{\alpha}'\tilde{\alpha} \} - 2\alpha' E\{ \tilde{\phi}\tilde{\alpha} \} + \alpha'\alpha.
\]
(2.1)

We define \( U_0 \) and \( U \) as \( U_0 = \nu\hat{\sigma}^2/\sigma^2 \) and \( U = \hat{\alpha}'\hat{\alpha}/\sigma^2 \), where \( \nu = n - q \). Then, \( U_0 \) is distributed as the chi-square distribution with \( \nu \) degrees of freedom, i.e., \( U_0 \sim \chi^2(\nu) \), and \( U \) is independent of \( U_0 \) and distributed as the noncentral chi-square distribution with \( q \) degrees of freedom and noncentral parameter \( \lambda = \hat{\alpha}'\hat{\alpha}/\sigma^2 \), i.e., \( U \sim \chi^2(q, \lambda) \). From \( F = (\nu/q)U/U_0, \tilde{\phi} \) is a function of \( (U, U_0) \), i.e., \( \tilde{\phi} = \tilde{\phi}(F) = \tilde{\phi}(U, U_0) \). To evaluate \( MSE(\tilde{\beta}) \), if we define the following two functions:
\[
H(l, m) = E\left\{ \tilde{\phi}^l\left(\frac{\tilde{\alpha}'\tilde{\alpha}}{\sigma^2}\right)^m \right\} = E\{ \tilde{\phi}^lU^m \},
\]
(2.2)
\[
J(l, m) = \frac{1}{\sigma^2}\alpha'E\{ \tilde{\phi}^lU^m\tilde{\alpha} \},
\]
(2.3)
then we have

\begin{equation}
MSE(\hat{\beta})/\sigma^2 = H(2, 1) - 2J(1, 0) + \lambda.
\end{equation}

First, we evaluate \(H(l, m)\). From \(U_0 \sim \chi^2(\nu)\) and \(U \sim \chi^2(q, \lambda)\), the corresponding density functions are

\begin{equation}
f_0(u_0, \nu) = \frac{1}{2^{\nu/2}\Gamma(\nu/2)} u_{0}^{\nu/2 - 1} \exp \left\{-\frac{u_0}{2} \right\},
\end{equation}

\begin{equation}
f(u) = \sum_{i=0}^{\infty} \omega_{i}(\lambda)f_0(u, q + 2i) \quad \text{where} \quad \omega_{i}(\lambda) = \frac{(\lambda/2)^i}{i!} \exp \{-\lambda/2\}.
\end{equation}

Then we have

\begin{equation}
H(l, m) = \int_{u=0}^{\infty} \int_{u_0=0}^{\infty} (\tilde{\phi}^l u^m) f_0(u_0, \nu) dudu_0
= \sum_{i=0}^{\infty} \omega_{i}(\lambda) \int_{u=0}^{\infty} \int_{u_0=0}^{\infty} (\tilde{\phi}^l u^m) f_0(u, q + 2i) f_0(u_0, \nu) dudu_0
= \sum_{i=0}^{\infty} \omega_{i}(\lambda) G_{i}(l, m),
\end{equation}

where

\begin{equation}
G_{i}(l, m) = \int_{u=0}^{\infty} \int_{u_0=0}^{\infty} (\tilde{\phi}^l u^m) f_0(u, q + 2i) f_0(u_0, \nu) dudu_0
= E_{U_0}\left\{E_U\{(\tilde{\phi}^l U^m) \mid U_0 = u_0\}\right\},
\end{equation}

where \(E_U\{\cdot \mid U_0 = u_0\}\) means the conditional expectation with respect to \(U \sim \chi^2(q + 2i)\) on the condition \(U_0 = u_0\), \(E_{U_0}\) means the expectation with respect to \(U_0 \sim \chi^2(\nu)\). Next, we evaluate \(J(l, m) = \alpha' E\{\tilde{\phi}^l U^m \hat{\alpha}/\sigma^2\}\). Noting \(\lambda = \alpha' \alpha/\sigma^2\), \(\partial \lambda / \partial \alpha = 2\alpha/\sigma^2\), we have

\begin{equation}
\frac{\partial \omega_{i}(\lambda)}{\partial \alpha} = \frac{\alpha}{\sigma^2} \omega_{i}(\lambda)\{i(\lambda/2)^{-1} - 1\} \quad (i = 0, 1, \ldots).
\end{equation}

Then we have

\begin{equation}
\frac{\partial}{\partial \alpha} H(l, m) = \frac{\alpha}{\sigma^2} \sum_{i=0}^{\infty} \omega_{i}(\lambda)\{i(\lambda/2)^{-1} - 1\} G_{i}(l, m).
\end{equation}

Since \(\tilde{\phi} = \tilde{\phi}(U, U_0) = \phi(\hat{\alpha}' \hat{\alpha}, U_0), U = \hat{\alpha}' \hat{\alpha}/\sigma^2\), and \(\hat{\alpha} \sim N(\alpha, \sigma^2 I_q), U_0 = \nu \sigma^2/\sigma^2 \sim \chi^2(\nu)\), an alternate expression of \(H(l, m)\) based on the distribution of \(\hat{\alpha}\) is

\begin{equation}
H(l, m) = \int_{\hat{\alpha}} \int_{u_0=0}^{\infty} \left[ \tilde{\phi}^l \left( \frac{1}{\sigma^2} \hat{\alpha}' \hat{\alpha} \right)^m \right] f_N(\hat{\alpha}) f_0(u_0) d\hat{\alpha} du_0.
\end{equation}

Noting \(\partial f_N(\hat{\alpha})/\partial \alpha = (\hat{\alpha} - \alpha) f_N(\hat{\alpha})/\sigma^2\) and differentiating (2.9) with respect to \(\alpha\), we have

\begin{equation}
\frac{\partial}{\partial \alpha} H(l, m) = \int_{\hat{\alpha}} \int_{u_0=0}^{\infty} \left[ \tilde{\phi}^l \left( \frac{1}{\sigma^2} \hat{\alpha}' \hat{\alpha} \right)^m \right]
\end{equation}
From (2.5) and (2.11),

\[ \phi(l, m) = \sum_{i=0}^{\infty} \omega_i(\lambda) \{ G_i(2, 1) - 4iG_i(1, 0) + 2iG_i(0, 0) \} \]

\[ = \sum_{i=0}^{\infty} \omega_i(\lambda) E_U \{ (\tilde{\phi}^2 U - 4i\tilde{\phi} + 2i) \mid U_0 = u_0 \}. \]

As for the \( G_i(l, m) \) of (2.6), we make use of the change of variables, \( v = u/u_0 = (q/\nu)F \), \( v_0 = u_0 \), (noting \( \partial(u, u_0)/\partial(v, v_0) = v_0 \), and \( \tilde{\phi} = \tilde{\phi}(u, u_0) = \tilde{\phi}(v) \)) and make use of the change of variable \( t = (v + 1)v_0/2 \) (noting \( dv_0 = 2/(v + 1)dt \)), and make use of the change of variable, \( z = v/(1 + v) \) (noting \( dv = (1 - z)^{-2}dz \), and \( \tilde{\phi} = \tilde{\phi}(v) = \tilde{\phi}(z) \)), then we have

\[ G_i(l, m) = E_U \{ (\tilde{\phi}^l U^m) \mid U_0 = u_0 \} \]

\[ = \int_{u=0}^{\infty} \int_{v=0}^{\infty} (\tilde{\phi}^l u^m) f_0(u, q + 2i) f_0(u_0, \nu) du dv \]

\[ = c_0 \times \int_{u=0}^{\infty} \int_{v=0}^{\infty} (\tilde{\phi}^l u^m) u^{(q+2i)/2-1} u_0^{(m+(q+\nu)/2+i)-1} \]

\[ \times \exp \left\{ -\frac{(u + u_0)}{2} \right\} dudv_0 \]

\[ = c_0 \times \int_{v=0}^{\infty} \int_{v_0=0}^{\infty} \tilde{\phi}^l v^{(m+q/2+i)-1} v_0^{(m+(q+\nu)/2+i)-1} \]

\[ \times \exp \left\{ -\frac{v_0(v + 1)}{2} \right\} dv dv_0 \]

\[ = c_0 \times \int_{v=0}^{\infty} \tilde{\phi}^l \frac{2^{m+(q+\nu)/2+i} v^{(m+q/2+i)-1}}{(1 + v)^{m+(q+\nu)/2+i}} \]

\[ \times \left\{ \int_{t=0}^{\infty} t^{(m+(q+\nu)/2+i)-1} e^{-t} dt \right\} dv \]

\[ = 2^m \frac{\Gamma \left( \frac{m + q + \nu}{2} + i \right)}{\Gamma(q/2 + i)\Gamma(\nu/2)} \int_{z=0}^{1} \tilde{\phi}^l z^{(m+q/2+i)-1} (1 - z)^{\nu/2-1} dz, \]
where $c_0 = \{2^{q/2+i+\nu/2}\Gamma(q/2+i)\Gamma(\nu/2)\}^{-1}$, and $\Gamma()$ is the Gamma function. Using (2.13), we obtain the alternative expression of $MSE(\hat{\beta})/\sigma^2$,

$$
MSE(\hat{\beta})/\sigma^2 = \sum_{i=0}^{\infty} \omega_i(\lambda) \left\{ G_i(2, 1) - 4iG_i(1, 0) + 2iG_i(0, 0) \right\}
$$

$$
= \sum_{i=0}^{\infty} \omega_i(\lambda) \frac{1}{B(q/2+i, \nu/2)} \times \int_{z=0}^{1} \left\{ 2 \left( \frac{q + \nu}{2} + i \right) \tilde{\phi}^2 z - 4i\tilde{\phi} + 2i \right\} Z^{(q/2+i)-1} (1 - z)^{\nu/2-1} dz
$$

$$
= \sum_{i=0}^{\infty} \omega_i(\lambda) E_Z \left\{ (q + \nu + 2i) \tilde{\phi}^2 Z - 4i\tilde{\phi} + 2i \right\},
$$

where $B(i, j) = (\Gamma(i) \cdot \Gamma(j))/\Gamma(i+j)$ is the beta function, and $E_Z$ means expectation with respect to the Beta distribution $Z \sim Beta(q/2+i, \nu/2)$. When the shrinkage factor estimator $\tilde{\phi}(F)$ is given as a function of $F$ statistic, the $\tilde{\phi}(F)$ is expressed as a function of $Z$ by using relation

$$
Z = \frac{qF}{\nu + qF}, \quad \left( F = \frac{\nu}{q} \cdot \frac{Z}{1-Z} \right).
$$

Therefore, the quantity $MSE(\tilde{\alpha})/\sigma^2$ can be evaluated by (2.14) and (2.15).

3. **Sufficient condition for MSE dominance**

It is known that the PP estimator dominates the OLS estimator in the terms of the predictive $MSE$ when $q \geq 3$ (see Baranchik (1970)). Also a sufficient condition for the estimator $\hat{\beta} = \tilde{\phi}(F)\hat{\beta}$ $(q \geq 3)$ to dominate the OLS estimator $\hat{\beta}$ is given by Baranchik (1970) or Efron and Morris (1976). The sufficient condition of Baranchik is expressed by using

$$
r(G) = \{1 - \tilde{\phi}(G)\}G, \quad G = \frac{q}{\nu}F,
$$

(b-1) $r(G)$ is monotone, nondecreasing, and

(b-2) $0 \leq r(G) \leq 2(q - 2)/(\nu + 2)$.

The sufficient condition of Efron and Morris is expressed by using

$$
\tau(H) = \{1 - \tilde{\phi}(H)\}H/(q - 2), \quad H = \frac{(\nu + 2)q}{\nu}F,
$$

$$
\psi(H) = H^{(q-2)/2}\tau(H)/(2 - \tau(H))^{1+2c}, \quad c = \frac{q - 2}{\nu + 2},
$$

(e-1) $0 \leq \tau(H) \leq 2$, and

(e-2) $\psi(H)$ is nondecreasing for all $H$ with $\tau(H) < 2$, and
if $H_1$ exists such that $\tau(H_1) = 2$, then $\tau(H) = 2$ for all $H \geq H_1$.

The relations $H = (\nu + 2)G$ and $\tau(H) = r(G)/c$ hold.

For the M estimator, we have

$$r(G) = \begin{cases} 1 - \frac{F^b}{a + F^b} & G = \frac{a_g G}{a_g + G^b}, \\ (a_g = (q/\nu)^b a), \end{cases}$$

$$r'(G) = \frac{a_g}{(a_g + G^b)} \{a_g - (b - 1)G^b\}.$$
\[ (\phi U - 4i)\tilde{\phi} + 2i - q \]
\[ = (\phi U - 4i)(1 - s(G)G^{-d}) + 2i - q \]
\[ = \tilde{\phi}U - 4i - (\phi U - 4i)s(G)G^{-d} + 2i - q \]
\[ = (1 - s(G)G^{-d})U - 4i - (\phi U - 4i)s(G)G^{-d} + 2i - q \]
\[ = U - 2i - q + s(G)G^{-d}U \left( -1 - \tilde{\phi} + \frac{4i}{U} \right), \]

and the condition (c-2)
\[ \tilde{\phi}(G) \geq 1 - c_1 G^{-1}, \]
we have the inequality

\[ (3.9) \quad R_1 = E_U \{(\tilde{\phi}^2 U - 4i\tilde{\phi} + 2i - q) \mid U_0 = u_0\} \]
\[ = E_U \left\{ s(G)G^{-d}U \left( -1 - \tilde{\phi}(G) + \frac{4i}{U} \right) \mid U_0 = u_0 \right\} \]
\[ \leq E_U \left\{ s(G)G^{-d}U \left( -2 + c_1 G^{-1} + \frac{4i}{U} \right) \mid U_0 = u_0 \right\} \]
\[ = E_U \left\{ s \left( \frac{U}{U_0} \right) U_0^d U^{1-d} \left( -2 + \frac{c_1 U_0 + 4i}{U} \right) \mid U_0 = u_0 \right\}. \]

Denoting
\[ W_1(U, U_0) = U^{1-d} \left( -2 + \frac{c_1 U_0 + 4i}{U} \right), \]

\[ W_1(U, U_0) \text{ for any fixing } U_0 = u_0 \text{ has} \]
\[ W_1(U, U_0) \geq 0 \quad (0 \leq U \leq w_1), \]
\[ W_1(U, U_0) < 0 \quad (U > w_1), \]

where \( w_1 = (c_1 U_0 + 4i)/2 \). Then, noting the condition (c-1), we have the inequality

\[ (3.10) \quad R_1 = E_U \left\{ s \left( \frac{U}{U_0} \right) U_0^d \cdot W_1(U, U_0) \mid U_0 = u_0 \right\} \]
\[ \leq E_U \left\{ s \left( \frac{w_1}{U_0} \right) U_0^d \cdot W_1(U, U_0) \mid U_0 = u_0, U \leq w_1 \right\} P\{U \leq w_1\} \]
\[ + E_U \left\{ s \left( \frac{w_1}{U_0} \right) U_0^d \cdot W_1(U, U_0) \mid U_0 = u_0, U > w_1 \right\} P\{U > w_1\} \]
\[ = E_U \left\{ s \left( \frac{w_1}{U_0} \right) U_0^d \cdot W_1(U, U_0) \mid U_0 = u_0 \right\} \]

From \( U \sim \chi^2(q + 2i) \), and \( E \{U^\alpha\} = 2^\alpha\Gamma((q + 2i)/2 + \alpha)/\Gamma((q + 2i)/2) \), we have

\[ (3.11) \quad E_U \{W_1(U, U_0) \mid U_0 = u_0\} \]
\[ = E_U \{-2U^{1-d} + (c_1 U_0 + 4i) U^d \mid U_0 = u_0\} \]
\[ R_1 \leq s \left( \frac{w_1}{U_0} \right) U_0^d \cdot \frac{2^{-d} \Gamma((q + 2i)/2 - d)}{\Gamma(q + 2i)} \times \{ -2(q - 2d) + c_1 U_0 \} \]  

Then we have

\[ R_0 = E_{U_0} \{ R_1 \} \]

\[ \leq E_{U_0} \left\{ s \left( \frac{w_1}{U_0} \right) U_0^d \frac{2^{-d} \Gamma((q + 2i)/2 - d)}{\Gamma(q + 2i)} \times \{ -2(q - 2d) + c_1 U_0 \} \right\} \]

\[ = \frac{2^{-d} \Gamma((q + 2i)/2 - d)}{\Gamma(q + 2i)} \times E_{U_0} \left\{ s \left( \frac{c_1}{2} + \frac{2q}{U_0} \right) U_0^d \{ -2(q - 2d) + c_1 U_0 \} \right\} . \]

Denoting

\[ W_2(U_0) = U_0^d \{ -2(q - 2d) + c_1 U_0 \} , \]

\[ W_2(U_0) \text{ has} \]

\[ W_2(U_0) < 0 \quad (0 \leq U_0 \leq w_2) , \]

\[ W_2(U_0) \geq 0 \quad (U_0 > w_2) , \]

where \( w_2 = 2(q - 2d)/c_1 \). Then we have the inequality

\[ E_{U_0} \left\{ s \left( \frac{c_1}{2} + \frac{2q}{w_2} \right) W_2(U_0) \right\} \]

\[ \leq s \left( \frac{c_1}{2} + \frac{2q}{w_2} \right) E_{U_0} \{ W_2(U_0) \mid U_0 \leq w_2 \} \{ U_0 \leq w_2 \} \]

\[ + s \left( \frac{c_1}{2} + \frac{2q}{w_2} \right) E_{U_0} \{ W_2(U_0) \mid U_0 > w_2 \} \{ U_0 > w_2 \} \]

\[ = s \left( \frac{c_1}{2} + \frac{2q}{w_2} \right) E_{U_0} \{ W_2(U_0) \} \]

\[ = s \left( \frac{c_1}{2} + \frac{2q}{w_2} \right) \left[ -2(q - 2d) \frac{2^d \Gamma(\nu/2 + d)}{\Gamma(\nu/2)} \right. \]

\[ + c_1 \frac{2^{d+1} \Gamma(\nu/2 + d + 1)}{\Gamma(\nu/2)} \]
and
\[ R_0 \leq \frac{2^{-d}\Gamma((q+2i)/2 - d)}{\Gamma(q+2i)} \times s \left( \frac{c_1}{2} + \frac{2q}{w_2} \right)^2 \frac{2^d\Gamma(\nu/2 + d)}{\Gamma(\nu/2)} \left[-2(q - 2d) + c_1 (\nu + 2d)\right]. \]

Then we see that

\[ (3.16) \quad \text{if} \quad c_1 \leq \frac{2(q - 2d)}{\nu + 2d}, \quad \text{then} \quad R_0 \leq 0 \quad \text{and then} \quad R \leq 0. \]

We therefore obtain the theorem. Note that \( d \) is any value such that \( 0 < d < q/2 \). When \( d = 1 \), the sufficient condition of the Theorem is equivalent to that of Baranchik, and from \( 0 < d(= 1) < q/2 \), we have \( q > 2 \). If \( \tilde{\phi}(F) \) exists such that \( s(G) \) satisfies the conditions (c-1) and (c-2) in \( 0 < d < 1/2 \), the estimator \( \tilde{\beta} = \tilde{\phi}(F)\hat{\beta} \) \((q > 2d = 1)\) dominates the OLS estimator. Then we may apply this sufficient condition of the theorem not only for \( q \geq 3 \) but also for \( q = 1, 2 \).

4. Sufficient condition for the M estimator

In this section, we derive the sufficient condition for the M estimator to dominate the OLS estimator, and show the region of \((a,b)\) in \( \tilde{\phi}_M = F^b/(a + F^b) \) which satisfy the sufficient condition.

From \( \tilde{\phi}(F) = F^b/(a + F^b), \ G = (q/\nu)F, \)

\[ (4.1) \quad \tilde{\phi}(G) = \frac{G^b}{a_g + G^b}, \quad \text{where} \quad a_g = \left(\frac{q}{\nu}\right)^b a. \]

The function \( s(G) \) and its derivative \( s'(G) \) are

\[ s(G) = \{1 - \tilde{\phi}(G)\}G^d = \frac{a_gG^d}{a_g + G^b}, \]
\[ (4.3) \quad s'(G) = \frac{a_gG^{d-1}}{(a_g + G^b)^2}\{a_gd + (d-b)G^b\}. \]

Then we can see that if \( d \geq b \) then \( s'(G) \geq 0 \) and then the condition (c-1) is satisfied. Next, we evaluate the condition (c-2). Setting \( t(G) = s(G)G^{1-d}(\nu + 2d)/(q - 2d) \), the condition (c-2) becomes \( t(G) \leq 2 \). The function \( t(G) \) and its derivative \( t'(G) \) are

\[ t(G) = \frac{a_gG^{1-d}}{a_g + G^b}\{a_g(1 - 1/b)G^b\} \times \frac{\nu + 2d}{q - 2d} \quad (d < q/2), \]
\[ t'(G) = \frac{a_g}{(a_g + G^b)^2}\{a_g + (1 - b)G^b\} \times \frac{\nu + 2d}{q - 2d}. \]

Then we can see that:
(1) if $0 \leq b < 1$, then $t'(G) > 0$ and then $\text{Max } t(G) = t(\infty) = \infty$. Then the condition (c-2) is not satisfied. Adding the condition (c-1): $b \leq d < q/2$, the sufficient condition for the M estimator is not satisfied when $q = 1, 2$.

(2) if $b = 1$, then $t(G) = \frac{a_g G}{a_g + G} \times \frac{\nu + 2d}{q - 2d}$ and $\text{Max } t(G) = t(\infty) = a_g \times \frac{\nu + 2d}{q - 2d}$.

Then the condition (c-2) is satisfied if $a_g \leq 2 \frac{q - 2d}{\nu + 2d} \times \frac{\nu}{q}$.

(3) if $b > 1$, then $\text{Max } t(G) = t(G_0)$ where $G_0 = \{a_g/(b - 1)\}^{1/b}$. From

$$t(G_0) = \frac{b - 1}{b} \times \left(\frac{a_g}{b - 1}\right)^{1/b} \cdot \frac{\nu + 2d}{q - 2d},$$

the condition (c-2): $t(G) \leq 2$ is satisfied if

$$a_g \leq (b - 1) \left\{ \frac{2b}{b - 1} \times \frac{q - 2d}{\nu + 2d} \right\}^b.$$
In Table 1, the values of $a(q) = a_{\text{peak}}(q), b(q) = b_{\text{peak}}(q)$ and $a_1(q) = \lim_{b \to 1^+} a(b, q)$ are presented.

<table>
<thead>
<tr>
<th>$\nu$</th>
<th>$\nu = 20$</th>
<th>$\nu = 40$</th>
<th>$\nu = 200$</th>
<th>$\nu = 1000$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$q$</td>
<td>$a(q)$</td>
<td>$b(q)$</td>
<td>$a_1(q)$</td>
<td>$a(q)$</td>
</tr>
<tr>
<td>3</td>
<td>0.606</td>
<td>1.000</td>
<td>0.606</td>
<td>1.000</td>
</tr>
<tr>
<td>4</td>
<td>1.148</td>
<td>1.183</td>
<td>0.909</td>
<td>1.225</td>
</tr>
<tr>
<td>5</td>
<td>1.635</td>
<td>1.320</td>
<td>1.091</td>
<td>1.487</td>
</tr>
<tr>
<td>6</td>
<td>2.116</td>
<td>1.576</td>
<td>1.299</td>
<td>1.867</td>
</tr>
<tr>
<td>7</td>
<td>2.595</td>
<td>1.635</td>
<td>1.414</td>
<td>2.265</td>
</tr>
<tr>
<td>8</td>
<td>3.072</td>
<td>1.693</td>
<td>1.538</td>
<td>2.675</td>
</tr>
<tr>
<td>9</td>
<td>3.549</td>
<td>1.751</td>
<td>1.664</td>
<td>3.105</td>
</tr>
<tr>
<td>10</td>
<td>4.025</td>
<td>1.809</td>
<td>1.792</td>
<td>3.545</td>
</tr>
<tr>
<td>11</td>
<td>4.500</td>
<td>1.867</td>
<td>1.920</td>
<td>4.005</td>
</tr>
<tr>
<td>12</td>
<td>4.972</td>
<td>1.925</td>
<td>2.050</td>
<td>4.480</td>
</tr>
<tr>
<td>13</td>
<td>5.443</td>
<td>1.983</td>
<td>2.180</td>
<td>4.965</td>
</tr>
<tr>
<td>14</td>
<td>5.910</td>
<td>2.041</td>
<td>2.310</td>
<td>5.460</td>
</tr>
<tr>
<td>15</td>
<td>6.375</td>
<td>2.100</td>
<td>2.440</td>
<td>5.965</td>
</tr>
<tr>
<td>16</td>
<td>6.836</td>
<td>2.159</td>
<td>2.570</td>
<td>6.480</td>
</tr>
<tr>
<td>17</td>
<td>7.293</td>
<td>2.219</td>
<td>2.700</td>
<td>6.980</td>
</tr>
<tr>
<td>18</td>
<td>7.746</td>
<td>2.279</td>
<td>2.830</td>
<td>7.480</td>
</tr>
<tr>
<td>19</td>
<td>8.196</td>
<td>2.339</td>
<td>2.960</td>
<td>7.980</td>
</tr>
<tr>
<td>20</td>
<td>8.640</td>
<td>2.399</td>
<td>3.090</td>
<td>8.480</td>
</tr>
</tbody>
</table>

Noting $a = a_g(\nu/q)^b$, and we denote the function $a(b, q, d)$ for $b > 1$ by

\[(4.6) \quad a(b, q, d) = a_g \left(\frac{\nu}{q}\right)^b = (b - 1) \left\{ \frac{2b}{b - 1} \times q - 2d \right\}^b \times \left(\frac{\nu}{q}\right)^b,
\]

and also denote that for $b = 1$ by

\[(4.7) \quad a(1, q, d) = \lim_{b \to 1^+} a(b, q, d) = \frac{2\nu(q - 2d)}{(\nu + 2d)q}.
\]

Then binding the conditions (c-2) for $b = 1$ and that for $b > 1$, we can see that if $a \leq a(b, q, d)$ and $b \geq 1$, then the condition (c-2) is satisfied. Adding the condition (c-1): $b \leq d < q/2$, the sufficient condition for the M estimator to dominate the OLS estimator is express as

\[a \leq a(b, q, d) \quad (1 \leq b \leq d < q/2).\]

Noting that $d$ is any constant value such that $d \geq b$, and the function $a(b, q, d)$ has the maximum value when $d = b$. We therefore obtain the following remark:

**Remark** A sufficient condition for the M estimator $\tilde{\beta} = \tilde{\phi}(F)\hat{\beta}$ ($q \geq 3$) to dominate the OLS estimator $\hat{\beta}$ in the terms of the predictive MSE is expressed as

\[(4.8) a \leq a(b, q) = (b - 1) \left\{ \frac{2b}{b - 1} \times q - 2d \right\}^b \times \left(\frac{\nu}{q}\right)^b \quad (1 \leq b < q/2).
\]
5. Numerical evaluation

We obtain the sufficient condition for the M estimator in (4.8). From (4.8), the coefficient $a$ depends on the value of $b$ ($1 \leq b < q/2$). For the coefficient $a$, we may expect that the coefficient $a$ is taken with possible
range greatly, because the HKB estimator (with $a = 1$, $b = 1$) has a smaller MSE in the wide region of $\lambda$ than the estimator of Farebrother (with $a = 1/q$, $b = 1$). So drawing the graph of the functions $\{a(b, q) : q = 3(1)12\}$, we have the first step graphs (the left graph for $\nu = 20$ and the right one for $\nu = 200$) shown in Figure 1.
Figure 4. The $MSE(\mid \lambda)$ and $MSE(\mid \phi)$ with $\nu = 200$, $w = 0.85$.

For each $q$, finding the peak point $(a_{\text{peak}}(q), b_{\text{peak}}(q))$ such that $a(b, q)$ has the maximum value $a_{\text{peak}}(q)$ at $b = b_{\text{peak}}(q)$, and drawing the functions $a_{\text{peak}}(q)$, and $b_{\text{peak}}(q)$, we have the second step graphs of $a_{\text{peak}}(q)$ and the third step graphs for $b_{\text{peak}}(q)$. The graphs of $a(1, q) = \lim_{b\to+1+} a(b, q)$ are
shown in the last step graphs in Figure 1. Also the values of \( a_{\text{peak}}(q) \), \( b_{\text{peak}}(q) \) and \( a(1,q) \) are shown in Table 1. For the coefficient \( b \), we take up the range \( 1 \leq b(q) \leq b_{\text{peak}}(q) \), i.e.

\[
b(q) = w \cdot 1 + (1 - w) \cdot b_{\text{peak}}(q) \quad (0 \leq w \leq 1).
\]

For a case of \( w = 0.85 \), which is chosen in some other cases by evaluation of relative efficiency, we show the results: the estimator of shrinkage factor \( \hat{\phi}(F) \) and \( \hat{\phi}(Z) \) for \( q = 3, 4, 6, 12, 20 \) with \( \nu = 20 \) are shown in Figure 2. The relative efficiency \( \text{MSE}(\hat{\beta})/\text{MSE}(\hat{\beta}) \) for \( q = 3, 4, 6, 12 \) and \( 20 \) are shown in Figure 3 (\( \nu = 20 \)), in Figure 4 (\( \nu = 200 \)), where the horizontal axis of each left graph means \( \lambda \) and that of right graph means shrinkage parameter in population \( \phi = \lambda/(\lambda+1) \), and the value of relative efficiency was computed in the range \( 0 \leq \lambda \leq 999 \) (\( 0 \leq \phi \leq 0.999 \)). From these results, we can see:

1. For the case of \( q = 3 \) with \( \nu = 20 \), the HKB estimator does not dominate the OLS estimator, and the M estimator or the PP estimator dominates the OLS estimator, and the M estimator has smaller MSE than the PP estimator in almost all the region of \( \lambda \) (or \( \phi \)).
2. For the case of \( q = 4 \) with \( \nu = 20 \), the performance of the M estimator and the HKB estimator are comparable, and the M estimator or the HKB estimator has smaller MSE than the PP estimator in almost all the region of \( \lambda \) (or \( \phi \)).
3. For the case of \( q = 6 \) with \( \nu = 20 \), although the M estimator has slightly larger MSE than the PP estimator when \( \phi \) is close to zero, the M estimator has smaller MSE than the PP estimator in the wide region of \( \lambda \) (or \( \phi \)). The HKB estimator has larger MSE than the M estimator when \( \lambda \) is close to zero.
4. For the cases of \( q = 12, 20 \) with \( \nu = 20 \), based on each left graph, the M estimator has smaller MSE than the HKB estimator or the PP estimator in the wide region of \( \lambda \). Also based on each right graph, the M estimator or the PP estimator has smaller MSE than the HKB estimator, and the MSE performance of the M estimator and the PP estimator are comparable.
5. Comparing the case of \( \nu = 20 \) and that of \( \nu = 200 \) in each \( q \), we see that the HKB estimator or the PP estimator has similar efficiency for the variation of \( \nu \), and the efficiency of the M estimator depends on the value \( \nu \). Furthermore, the efficiency of the M estimator is slightly improved when \( \nu \) increases.

Acknowledgements
The author wishes to thank the referees for their helpful comments.

References


