Communications in Statistics - Theory and Methods

Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/lsta20

MSE Performance and Minimax Regret Significance Points for a HPT Estimator when each Individual Regression Coefficient is Estimated

Haifeng Xu

Department of Statistics, School of Economics, Xiamen University and Wang Yanan Institute for Studies in Economics (WISE), Xiamen University, Xiamen, China

To cite this article: Haifeng Xu (2013): MSE Performance and Minimax Regret Significance Points for a HPT Estimator when each Individual Regression Coefficient is Estimated, Communications in Statistics - Theory and Methods, 42:12, 2152-2164

To link to this article: http://dx.doi.org/10.1080/03610926.2011.605239

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: http://www.tandfonline.com/page/terms-and-conditions

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae, and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand, or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.
MSE Performance and Minimax Regret Significance Points for a HPT Estimator when each Individual Regression Coefficient is Estimated

HAIFENG XU
Department of Statistics, School of Economics, Xiamen University and Wang Yanan Institute for Studies in Economics (WISE), Xiamen University, Xiamen, China

In this article, we consider a heterogeneous preliminary test (HPT) estimator whose components are the OLS and feasible ridge regression (FRR) estimators, and derive the exact formulae for the moments of the HPT estimator using mathematical method. Since we cannot examine the MSE of the HPT estimator analytically, we execute the numerical evaluation to investigate the MSE performance of the HPT estimator, and compare the MSE performance of the HPT estimator with those of the FRR estimator and the usual OLS estimator. Furthermore, using the minimax regret criterion proposed by Sawa and Hiromatsu (1973), we derive the optimal critical points of the preliminary \( F \) test. Our results show that the optimal significance points are greater than 19% and the optimal significance points decrease as the denominator degrees of freedom of the preliminary \( F \) test statistic increases.

Keywords Feasible ridge regression estimator; Heterogeneous preliminary test estimator; Minimax regret criterion; Optimal critical points.

Mathematics Subject Classification 62J05; 62J07.

1. Introduction
Ridge regression estimators were proposed by Hoerl and Kennard (1970) to deal with the problem of multicollinearity. If we accept some bias in trade for a reduction in variance, the ridge regression estimator can have smaller mean squared error (MSE) than the ordinary least squares (OLS) estimator in a certain region of parameter space. For this reason, no matter whether there is a problem of multicollinearity or not, the ridge regression estimator may be preferred to the OLS estimator in some cases.

In applied regression analysis, we may be interested in a specific regression coefficient rather than all of them. For instance, Ohtani (1997) showed the following
MSE Performance and Optimal Significance Points

example. The import demand equation can be express as $y_t = \beta_0 + x_1t + x_2\beta_2 + \epsilon_t$, where $y_t$ indicates the imports of a country at time $t$, $x_1t$ indicates the real income (real GNP), and $x_2\beta_2$ indicates the ratio of the import price to the domestic price (relative price). If our concern is to predict the impact of change in the real income on trade account (ceteris paribus), then the estimated coefficient $\beta_1$ (income elasticity) is desired to be as accurate as possible. From this viewpoint, Ohtani and Kozumi (1996) and Ohtani (1997, 2000) examined the MSE performance of the Stein-rule (SR), positive-part Stein-rule (PSR) and minimum mean squared error (MMSE) estimators\(^1\) for each individual regression coefficient. Also, Srivastava and Giles (1984) investigated the finite sample properties of a pre-test generalized ridge regression (PTGRR) estimator for each individual regression coefficient in a canonical form. Huang (1999) showed the other situations that the estimation of each individual regression coefficient is important, and proposed a feasible ridge regression (FRR) estimator to estimate a specific regression coefficient. Also, several recent studies have focused on the FRR estimator, such as Ohtani (2007, 2008). Namba and Ohtani (2009) considered a linear regression model and incorporated a preliminary test for an inequality constraint on a regression coefficient into the ridge regression estimator proposed by Huang (1999). They also derived the exact formula for the risk of the estimator under the asymmetric LINEX loss function and they compared the risk performance of the proposed ridge regression estimator with that of the OLS estimator using numerical evaluations.

In the context of a linear regression model, a preliminary test for linear restrictions on regression coefficients can be used to choose the restricted or unrestricted estimators. Traditionally, the 1% or 5% level has been used as a level of a test. However, Sawa and Hiromatsu (1973) proposed a minimax regret criterion to choose the optimal critical point of the preliminary test. Their results showed that the optimal critical point of the preliminary $t$ test decreases slightly as the degrees of freedom increase, but it is nearly constant. In the context of estimating regression error variance, Toyoda and Wallace (1975) presented that optimal significance levels of a preliminary $F$ test vary somewhat but are close to 1/2 for most degrees of freedom and equal to 1/2 when the numerator and denominator of the degrees of freedom are equal. Also, Toyoda and Wallace, (1976) derived optimal critical points for a preliminary test in regression using a minimum average relative risk criterion. Further, Brook (1976) examined optimal critical points of the preliminary $F$ test and their corresponding significance levels using the quadratic risk function and the regret function proposed by Sawa and Hiromatsu (1973) for orthogonal and non-orthogonal cases, respectively. His results suggested that if the $F$-statistic is more than two, the unrestricted estimator should be used.

As is well known, the ridge regression estimator has smaller MSE than the OLS estimator over a certain region of parameter space. In particular, the MSE dominance of the FRR estimator over the OLS estimator holds when the individual regression coefficient is close to zero (see, for example, Huang, 1999; Ohtani, 2008). From this viewpoint, in this article, we consider the heterogeneous preliminary test (HPT) estimator whose components are the OLS and FRR estimators, and investigate the risk performance of the HPT estimator. We derive the exact general formula for the risk of the HPT estimator under the quadratic loss function.

\(^1\)The SR, PSR, and MMSE estimator are proposed by Stein (1956), Baranchik (1970), and Farebrother (1975), respectively.
Consider a linear regression model

\[ Y = X\beta + \varepsilon = x_1\beta_1 + X_2\beta_2 + \varepsilon, \quad \varepsilon \sim N(0, \sigma^2 I_n), \] (1)

where \( Y \) is an \( n \times 1 \) vector of observations on the explained variable, \( X = [x_1, X_2] \) is an \( n \times k \) matrix, \( x_1 \) is an \( n \times 1 \) vector of observations on an explanatory variable, and \( X_2 \) is an \( n \times (k-1) \) matrix of observations on other explanatory variables. \( \beta = [\beta_1, \beta_2]^T \) is a \( k \times 1 \) vector of regression coefficients for \( X \), \( \beta_1 \) is a scalar coefficient for \( x_1 \) which we have a special interest, \( \beta_2 \) is a \((k-1) \times 1\) vector of coefficients for \( X_2 \), and \( \varepsilon \) is an \( n \times 1 \) vector of error terms. As to the error terms, we assume that \( \varepsilon \) follows a normal distribution with location parameter 0 and scale parameter \( \sigma^2 \).

Since \( X = [x_1, X_2] \) and \( \beta = [\beta_1, \beta_2]^T \), the OLS estimator of \( \beta \) is \( b = (X'X)^{-1}X'Y \), and the OLS estimator of \( \beta_1 \) is \( b_1 = (x_1'M_2x_1)^{-1}x_1'M_2Y \), where \( M_2 = I_n - X_2(X_2X_2)^{-1}X_2' \). The distribution of \( b_1 \) can be written as

\[ b_1 \sim N(\beta_1, \sigma^2(x_1'M_2x_1)^{-1}). \] (2)

The FRR estimator of \( \beta_1 \) proposed by Huang (1999) is given by

\[
\hat{\beta}_1 = \left( x_1'M_2x_1 + \frac{s^2}{b_1^2} \right)^{-1} x_1'M_2y \\
= \left( \frac{(x_1'M_2x_1)b_1^2}{(x_1'M_2x_1)b_1^2 + s^2} \right) b_1,
\] (3)

where \( s^2 = (Y - Xb)'(Y - Xb)/v \) and \( v = n - k \).

Since the MSE dominance of the FRR estimator over the OLS estimator holds when the individual regression coefficient is close to zero, we consider the following HPT estimator whose components are the FRR and OLS estimators:

\[
\hat{\beta}_1(c) = I(F < c)\hat{\beta}_1 + I(F \geq c)b_1 \\
= I(F < c) \left( \frac{(x_1'M_2x_1)b_1^2}{(x_1'M_2x_1)b_1^2 + s^2} \right) b_1 + I(F \geq c)b_1,
\] (4)
where \( I(A) \) is an indicator function, \( c \) is a critical value of the pre-test, and \( F \) is the pre-test statistic for the null hypothesis \( H_0: \beta_1 = 0 \) against the alternative \( H_1: \beta_1 \neq 0 \):

\[
F = \frac{(x'_I M_2 x_I) b_I^2}{s^2}. \tag{5}
\]

Thus, \( \hat{\beta}_1(c) = \beta_1 \) if \( H_0 \) is accepted and \( \hat{\beta}_1(c) = b_1 \) otherwise.

### 3. Moments and Risk of the HPT Estimator

In this section, we derive the exact formulae for the moments of the MSE of the HPT estimator. As shown in the Appendix, the exact formulae for even and odd moments of \( \hat{\beta}_1(c) \) are

\[
E[\hat{\beta}_1(c)^{2r}] = E[I(F < c)(\hat{\beta}_1)^{2r} + I(F \geq c)(b_1)^{2r}]
= E\left[ I(F < c) \left( \frac{(x'_I M_2 x_I) b_I^2}{(x'_I M_2 x_I) b_I^2 + s^2} \right)^{2r} (b_1)^{2r} \right] + E[I(F \geq c)(b_1)^{2r}]
= H_1(2r, r; c) + H_2(0, r; c), \tag{6}
\]

and

\[
E[\hat{\beta}_1(c)^{2r+1}] = E[I(F < c)(\hat{\beta}_1)^{2r+1} + I(F \geq c)(b_1)^{2r+1}]
= E\left[ I(F < c) \left( \frac{(x'_I M_2 x_I) b_I^2}{(x'_I M_2 x_I) b_I^2 + s^2} \right)^{2r+1} (b_1)^{2r+1} \right] + E[I(F \geq c)(b_1)^{2r+1}]
= J_1(2r + 1, r; c) + J_2(0, r; c), \tag{7}
\]

where

\[
H_1(p, q; c) + H_2(p, q; c) = E\left[ I(F < c) \left( \frac{(x'_I M_2 x_I) b_I^2}{(x'_I M_2 x_I) b_I^2 + s^2} \right)^p (b_1)^q \right]
+ E\left[ I(F \geq c) \left( \frac{(x'_I M_2 x_I) b_I^2}{(x'_I M_2 x_I) b_I^2 + s^2} \right)^p (b_1)^q \right]
= \left( \frac{\sigma^2}{x'_I M_2 x_I} \right)^q \left[ P_1(p, q; c) + P_2(p, q; c) \right], \tag{8}
\]

\[
J_1(p, q; c) + J_2(p, q; c) = E\left[ I(F < c) \left( \frac{(x'_I M_2 x_I) b_I^2}{(x'_I M_2 x_I) b_I^2 + s^2} \right)^p (b_1)^q b_1 \right]
+ E\left[ I(F \geq c) \left( \frac{(x'_I M_2 x_I) b_I^2}{(x'_I M_2 x_I) b_I^2 + s^2} \right)^p (b_1)^q b_1 \right]
= \beta_1 \left( \frac{\sigma^2}{x'_I M_2 x_I} \right)^q \left[ T_1(p, q; c) + T_2(p, q; c) \right], \tag{9}
\]
From Eq. (15), the MSE performance of the expression:

\[ P_1(p, q; c) = 2^q v^p \sum_{i=0}^{\infty} w_i(\theta_i) \frac{\Gamma((v+1)/2+q+i)}{\Gamma(1/2+i)\Gamma(v/2)} \times \int_0^{c/(v+c)} t^{1/2+p+q+i-1}(1-t)^{v/2-1}[1+(v-1)t]^{-p} dt, \]

\[ T_1(p, q; c) = 2^q v^p \sum_{i=0}^{\infty} w_i(\theta_i) \frac{\Gamma((v+3)/2+q+i)}{\Gamma(3/2+i)\Gamma(v/2)} \times \int_0^{c/(v+c)} t^{1/2+p+q+i-1}(1-t)^{v/2-1}[1+(v-1)t]^{-p} dt, \]

\[ P_2(p, q; c) = 2^q \sum_{i=0}^{\infty} w_i(\theta_i) \frac{\Gamma((v+1)/2+q+i)}{\Gamma(1/2+i)\Gamma(v/2)} \int_0^{1} t^{1/2+p+q+i-1}(1-t)^{v/2-1} dt, \]

\[ T_2(p, q; c) = 2^q \sum_{i=0}^{\infty} w_i(\theta_i) \frac{\Gamma((v+3)/2+q+i)}{\Gamma(3/2+i)\Gamma(v/2)} \int_0^{1} t^{1/2+p+q+i-1}(1-t)^{v/2-1} dt. \]

where \( v = n - k, \theta_i = \beta_i^2 x_i^2 / \sigma^2, w_i(\theta_i) = e^{-\theta_i/2}(\theta_i/2)^{\theta_i}/\Gamma(\theta_i). \)

Using Eqs. (10)–(13), the MSE of the \( \hat{\beta}_1(c) \) can be expressed as follows:

\[ \text{MSE}(\hat{\beta}_1(c)) = E[(\hat{\beta}_1(c) - \beta_1)^2] \]

\[ = E((\hat{\beta}_1(c))^2) - 2\beta_1 E(\hat{\beta}_1(c)) + \beta_1^2 \]

\[ = \frac{\beta_1^2}{\theta_1} [P_1(2, 1; c) + P_2(0, 1; c)] - 2\beta_1^2 [T_1(1, 0; c) + T_2(0, 0; c)] + \beta_1^2. \]

(14)

Substituting Eqs. (10)–(13) into (14), and differentiating with respect to \( c \), we obtain:

\[ \frac{\partial \text{MSE}(\hat{\beta}_1(c))}{\partial c} = \frac{\beta_1^2}{\theta_1} \left[ \frac{\partial P_1(2, 1; c)}{c} + \frac{\partial P_2(0, 1; c)}{c} \right] - 2\beta_1^2 \left[ \frac{\partial T_1(1, 0; c)}{c} + \frac{\partial T_2(0, 0; c)}{c} \right] \]

\[ = -2\beta_1^2 \sum_{i=0}^{\infty} w_i(\theta_i) \frac{\Gamma((v+1)/2+i+1)}{\Gamma(1/2+i)\Gamma(v/2)} \frac{v^{1/2}c^{1/2}}{(v+c)^{1/2+i}} \]

\[ \times \frac{1}{\theta_1} \left[ \frac{(1+2c)}{(1+c)^2} - \frac{\theta_1}{(1+c)(1/2+i)} \right]. \]

(15)

From Eq. (15), the MSE performance of \( \hat{\beta}_1(c) \) depends on the sign of the following expression:

\[ \left[ \frac{(1+2c)}{(1+c)^2} - \frac{\theta_1}{(1+c)(1/2+i)} \right]. \]

(16)

We consider the cases of \( \theta_1 = 0 \) and \( \theta_1 > 0 \) separately. When \( \theta_1 = 0 \), Eq. (16) is positive for all \( c \geq 0 \) and the MSE of \( \hat{\beta}_1(c) \) is monotonically decreasing function of \( c \). This result indicates that the MSE of \( \hat{\beta}_1(c) \) is minimized when \( c \) is infinity. In this case, the desirable estimator is the FRR estimator. Unfortunately, when \( \theta_1 > 0 \), there is the running parameter \( i \) from zero to \( \infty \) in Eq. (16), we cannot
determine whether Eq. (16) is positive or negative. Thus, we examine the MSE of HPT estimators corresponding to some significance levels with that of the FRR estimator using numerical evaluations in the next section, and examine optimal critical points of the preliminary F test using the minimax regret criterion proposed by Sawa and Hiromatsu (1973) in Sec. 5.

4. MSE Performance of the HPT Estimator

In this section, we execute the numerical evaluations to investigate the risk performance of the HPT estimator \( \hat{\beta}_1(c) \), and compare with those of the FRR estimator \( \hat{\beta}_1 \) and the usual OLS estimator \( b_1 \). In numerical evaluations, since

\[
\text{MSE}(b_1) = \sigma^2/x_1'M_2x_1,
\]

we use the relative MSE defined as

\[
\frac{\text{MSE}(\hat{\beta}_1(c))}{\text{MSE}(b_1)} = P_1(2, 1; c) + P_2(0, 1; c) - 2\theta_1 [T_1(1, 0; c) + T_2(0, 0; c)] + \theta_1. \tag{17}
\]

The parameter values used in numerical evaluations are as follows: \( v(=n-k) = 4, 8, 16, 24, 60, 120; \theta_1 = \) various values. Also, \( c \) is chosen so that the pretest is conducted at 1%, 5%, 10%, 20%, and 50% significance levels. The numerical evaluations were executed on a personal computer, using a FORTRAN code. Whenever the increment of infinite series in the formulae of \( P_1(p, q; c) \), \( T_1(p, q; c) \), \( P_2(p, q; c) \), and \( T_2(p, q; c) \) given in Eqs. (10), (11), (12) and (13) becomes smaller than \( 10^{-12} \), the infinite series are judged to converge. Further, the integral in Eqs. (10)–(13) are calculated using the Simpson’s 3/8 rule with 1,000 equal subdivisions. Since we calculate the relative MSE defined in (17), \( \hat{\beta}_1(c) \) has smaller MSE than \( b_1 \) when the relative MSE is smaller than unity. Since the results for \( v = 4, 24, 120 \) is typical, we show the results for these cases and discuss in detail. Since \( \hat{\beta}_1(c) = \hat{\beta}_1 \) when \( x = 0.00 \), the HPT estimator becomes the FRR estimator in this case.

Tables 1–3 show the relative MSE of the HPT and FRR estimators for \( v = 4, 24, 120 \). As shown in the tables, for any given \( v \) and \( x \), as \( \theta_1 \) increases, the values of the relative MSE of both \( \hat{\beta}_1(c) \) and \( \hat{\beta}_1 \) climb up, decline and finally converge to unity from above. Also, as is clear in the tables, both \( \hat{\beta}_1(c) \) and \( \hat{\beta}_1 \) have smaller MSE than \( b_1 \) when \( \theta_1 \) (or \( \beta_1 \)) is close to zero. It means that the HPT and FRR estimators are superior to the usual OLS estimator when the value of \( \theta_1 \) is not large. We also find out that the range of dominance of the HPT and FRR estimators over the usual OLS estimator get narrow as \( v \) increases.

Our numerical results also show that the relative MSE of \( \hat{\beta}_1(c) \) is smaller than that of \( \hat{\beta}_1 \) in some cases of \( \theta_1 \). For example, when \( v = 4 \) and \( \theta_1 \geq 100.0 \), the HPT estimators for all kinds of \( x \) are smaller than the FRR estimator. Comparing Tables 1 and 2, we can find that the range of dominance of the HPT estimator over the FRR estimator gets wide as \( v \) increases. Further, as \( x \) gets large, the range where the relative risk of the HPT estimator is smaller than that of the FRR estimator gets wide. However, when \( \theta_1 \) is close to zero, the relative risk of the HPT estimator increases as \( x \) increases. It is clear that there exists a trade-off between them. Therefore, it seems that a suitable \( x \) level should be chosen to obtain the desirable HPT estimator in some sense.
Relative MSE of $\hat{\beta}_1$ for $v = n - k = 4$

<table>
<thead>
<tr>
<th>$\theta_1$</th>
<th>0.00</th>
<th>0.01</th>
<th>0.05</th>
<th>0.10</th>
<th>0.20</th>
<th>0.50</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>0.519</td>
<td>0.522</td>
<td>0.545</td>
<td>0.585</td>
<td>0.677</td>
<td>0.908</td>
</tr>
<tr>
<td>0.1</td>
<td>0.554</td>
<td>0.557</td>
<td>0.582</td>
<td>0.624</td>
<td>0.716</td>
<td>0.925</td>
</tr>
<tr>
<td>0.5</td>
<td>0.680</td>
<td>0.684</td>
<td>0.718</td>
<td>0.766</td>
<td>0.852</td>
<td>0.981</td>
</tr>
<tr>
<td>1.0</td>
<td>0.808</td>
<td>0.815</td>
<td>0.858</td>
<td>0.911</td>
<td>0.985</td>
<td>1.029</td>
</tr>
<tr>
<td>1.5</td>
<td>0.911</td>
<td>0.919</td>
<td>0.971</td>
<td>1.025</td>
<td>1.084</td>
<td>1.059</td>
</tr>
<tr>
<td>2.0</td>
<td>0.992</td>
<td>1.003</td>
<td>1.061</td>
<td>1.115</td>
<td>1.156</td>
<td>1.077</td>
</tr>
<tr>
<td>3.0</td>
<td>1.106</td>
<td>1.122</td>
<td>1.191</td>
<td>1.236</td>
<td>1.242</td>
<td>1.089</td>
</tr>
<tr>
<td>5.0</td>
<td>1.215</td>
<td>1.242</td>
<td>1.315</td>
<td>1.332</td>
<td>1.275</td>
<td>1.071</td>
</tr>
<tr>
<td>10.0</td>
<td>1.236</td>
<td>1.285</td>
<td>1.318</td>
<td>1.261</td>
<td>1.145</td>
<td>1.019</td>
</tr>
<tr>
<td>20.0</td>
<td>1.156</td>
<td>1.226</td>
<td>1.142</td>
<td>1.066</td>
<td>1.017</td>
<td>1.001</td>
</tr>
<tr>
<td>50.0</td>
<td>1.068</td>
<td>1.086</td>
<td>1.004</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>100.0</td>
<td>1.035</td>
<td>1.007</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>200.0</td>
<td>1.018</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
</tbody>
</table>

Note: Since $\hat{\beta}_1(c) = \hat{\beta}_1$, when $\alpha = 0.00$, the HPT estimator becomes the FRR estimator.

5. Optimal Significance Points Using the Minimax Regret Criterion

The relative risk performances of the HPT and FRR estimators are presented in Sec. 4. However, we cannot decide the optimal critical point of the preliminary $F$

Relative MSE of $\hat{\beta}_1$ for $v = n - k = 24$

<table>
<thead>
<tr>
<th>$\theta_1$</th>
<th>0.00</th>
<th>0.01</th>
<th>0.05</th>
<th>0.10</th>
<th>0.20</th>
<th>0.50</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>0.476</td>
<td>0.489</td>
<td>0.541</td>
<td>0.603</td>
<td>0.713</td>
<td>0.927</td>
</tr>
<tr>
<td>0.1</td>
<td>0.514</td>
<td>0.530</td>
<td>0.587</td>
<td>0.649</td>
<td>0.754</td>
<td>0.941</td>
</tr>
<tr>
<td>0.5</td>
<td>0.650</td>
<td>0.678</td>
<td>0.751</td>
<td>0.813</td>
<td>0.893</td>
<td>0.988</td>
</tr>
<tr>
<td>1.0</td>
<td>0.789</td>
<td>0.834</td>
<td>0.919</td>
<td>0.974</td>
<td>1.023</td>
<td>1.027</td>
</tr>
<tr>
<td>1.5</td>
<td>0.899</td>
<td>0.961</td>
<td>1.051</td>
<td>1.094</td>
<td>1.114</td>
<td>1.051</td>
</tr>
<tr>
<td>2.0</td>
<td>0.986</td>
<td>1.064</td>
<td>1.153</td>
<td>1.182</td>
<td>1.174</td>
<td>1.064</td>
</tr>
<tr>
<td>3.0</td>
<td>1.108</td>
<td>1.215</td>
<td>1.287</td>
<td>1.286</td>
<td>1.233</td>
<td>1.071</td>
</tr>
<tr>
<td>5.0</td>
<td>1.222</td>
<td>1.366</td>
<td>1.373</td>
<td>1.320</td>
<td>1.220</td>
<td>1.053</td>
</tr>
<tr>
<td>10.0</td>
<td>1.236</td>
<td>1.353</td>
<td>1.229</td>
<td>1.152</td>
<td>1.077</td>
<td>1.012</td>
</tr>
<tr>
<td>20.0</td>
<td>1.148</td>
<td>1.098</td>
<td>1.027</td>
<td>1.012</td>
<td>1.004</td>
<td>1.000</td>
</tr>
<tr>
<td>50.0</td>
<td>1.061</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>100.0</td>
<td>1.031</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>200.0</td>
<td>1.015</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
</tbody>
</table>

Note: Since $\hat{\beta}_1(c) = \hat{\beta}_1$, when $\alpha = 0.00$, the HPT estimator becomes the FRR estimator.
MSE Performance and Optimal Significance Points

Table 3
Relative MSE of $\hat{\beta}_1(c)$ for $\nu = n - k = 120$

<table>
<thead>
<tr>
<th>$\theta_1$</th>
<th>0.00</th>
<th>0.01</th>
<th>0.05</th>
<th>0.10</th>
<th>0.20</th>
<th>0.50</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>0.469</td>
<td>0.483</td>
<td>0.539</td>
<td>0.603</td>
<td>0.713</td>
<td>0.923</td>
</tr>
<tr>
<td>0.1</td>
<td>0.507</td>
<td>0.526</td>
<td>0.587</td>
<td>0.651</td>
<td>0.755</td>
<td>0.938</td>
</tr>
<tr>
<td>0.5</td>
<td>0.645</td>
<td>0.681</td>
<td>0.758</td>
<td>0.820</td>
<td>0.899</td>
<td>0.988</td>
</tr>
<tr>
<td>1.0</td>
<td>0.785</td>
<td>0.843</td>
<td>0.933</td>
<td>0.986</td>
<td>1.031</td>
<td>1.030</td>
</tr>
<tr>
<td>1.5</td>
<td>0.897</td>
<td>0.977</td>
<td>1.069</td>
<td>1.109</td>
<td>1.123</td>
<td>1.056</td>
</tr>
<tr>
<td>2.0</td>
<td>0.985</td>
<td>1.086</td>
<td>1.174</td>
<td>1.183</td>
<td>1.178</td>
<td>1.070</td>
</tr>
<tr>
<td>3.0</td>
<td>1.108</td>
<td>1.247</td>
<td>1.309</td>
<td>1.300</td>
<td>1.240</td>
<td>1.077</td>
</tr>
<tr>
<td>5.0</td>
<td>1.223</td>
<td>1.403</td>
<td>1.386</td>
<td>1.325</td>
<td>1.222</td>
<td>1.057</td>
</tr>
<tr>
<td>10.0</td>
<td>1.236</td>
<td>1.355</td>
<td>1.217</td>
<td>1.143</td>
<td>1.074</td>
<td>1.012</td>
</tr>
<tr>
<td>20.0</td>
<td>1.147</td>
<td>1.072</td>
<td>1.021</td>
<td>1.009</td>
<td>1.003</td>
<td>1.000</td>
</tr>
<tr>
<td>50.0</td>
<td>1.060</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>100.0</td>
<td>1.030</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>200.0</td>
<td>1.015</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
</tbody>
</table>

Note: Since $\hat{\beta}_1(c) = \hat{\beta}_1$, when $\alpha = 0.00$, the HPT estimator becomes the FRR estimator.

Table 4
Optimal critical values, their percentage levels of the pre-test, and least favorable values of $\theta_1$ for $\nu = 4, 8, 16, 24, 60, 120$

<table>
<thead>
<tr>
<th>$\nu$</th>
<th>$\theta_1^*$</th>
<th>$c^*$</th>
<th>$x^*$</th>
<th>$\theta_1^L$</th>
<th>$\theta_1^U$</th>
<th>$\delta_1^L(= \delta_1^U)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>2.1</td>
<td>1.60</td>
<td>27.46%</td>
<td>0.0</td>
<td>4.0</td>
<td>0.225</td>
</tr>
<tr>
<td>8</td>
<td>2.1</td>
<td>1.69</td>
<td>22.98%</td>
<td>0.2</td>
<td>3.6</td>
<td>0.232</td>
</tr>
<tr>
<td>16</td>
<td>2.1</td>
<td>1.74</td>
<td>20.57%</td>
<td>0.4</td>
<td>3.6</td>
<td>0.240</td>
</tr>
<tr>
<td>24</td>
<td>2.1</td>
<td>1.75</td>
<td>19.83%</td>
<td>0.4</td>
<td>3.6</td>
<td>0.242</td>
</tr>
<tr>
<td>60</td>
<td>2.1</td>
<td>1.74</td>
<td>19.22%</td>
<td>0.5</td>
<td>3.6</td>
<td>0.246</td>
</tr>
<tr>
<td>120</td>
<td>2.1</td>
<td>1.70</td>
<td>19.48%</td>
<td>0.5</td>
<td>3.6</td>
<td>0.251</td>
</tr>
</tbody>
</table>

Note: $n =$ number of points in the sample space; $k =$ number of parameters; $\theta_1^*$ indicates that $R_{\theta_1}(\theta_1^*, c) = R_{\theta_1}(\theta_1^*, c)$; $c^*$ = critical value of the pre-test of estimation; $x^*$ = percentage alpha level for the central $F$ distribution, and degrees of freedom is 1 and $n - k$; $\theta_1^L$, $\theta_1^U =$ optimal lower and upper $\theta_1$ values; $\delta_1^L(\delta_1^U) =$ the minimized regret function values of $\hat{\beta}_1(b_1)$.

test by the results of the risk performance. In this section, using the minimax regret criterion proposed by Sawa and Hiromatsu (1973), an optimal value of $c$ will be found by minimizing the maximum regret over all possible values of $\theta_1$.

Although we cannot show theoretically, our numerical results show that the MSE of the HPT estimator has a minimum value at $c = +\infty$ for $\theta_1 < \theta_1^*$ and at $c = 0$ for $\theta_1 \geq \theta_1^*$, where $\theta_1^*$ is the value of $\theta_1$ such that $R_{\theta_1}(\theta_1^*) = R_{\theta_1}(\theta_1^*)^2$. The values

$2 R_{\theta_1}(\theta_1^*)$ and $R_{\theta_1}(\theta_1^*)$ are quadratic risk functions (MSE) for the FRR ($\hat{\beta}_1$) and the OLS estimator ($b_1$), respectively.
of $\theta^*_1$ are shown in Table 4. The distances $\delta_L = \delta(\theta^*_1, c)$ and $\delta_U = \delta(\theta^*_1, c)$ indicate the maximum additional risk of $\hat{\beta}_1(c)$ above those of $\hat{\beta}_1$ and $b_1$. This means that $\delta_L$ and $\delta_U$ are the regret for using $\hat{\beta}_1(c)$ instead of $\hat{\beta}_1$ or $b_1$. Figure 1 shows the typical quadratic risk functions for different critical values. As is shown in Brook (1976) and illustrated by Fig. 1 that as $c$ increases, $\delta_L$ decreases and $\delta_U$ increases. Our purpose is to control both $\delta_L$ and $\delta_U$, and try to find an optimum value $c = c^*$ which minimizes the maximum regret over all values of $\theta_1$.

Optimal values of $c$ (says, $c^*$) and percentage levels of the preliminary test (says, $\alpha^*$) are presented in Table 4.

From Table 4, it is clear that the outstanding property of the optimal critical value, $c^*$, is that it does not vary much. $c^*$ only increases by about 0.1 as the denominator degrees of freedom of the preliminary test $F$ distribution ($v = n - k$) increases from 4 to 120. According to the results, $c^*$ increases as $v$. However, $\alpha^*$ decreases as $v$ increases.

Traditionally, the 1%, 5%, and 10% significance levels have been used in a testing procedure. If we use the 5% significance level in a pre-test in our context, it may lead to the possible large regret. However, if we use the optimal critical value in the sense of minimax regret, we can avoid obtaining such a large regret. This indicates that when at least a test is conducted to choose an estimator as a pre-test, the indiscriminate use of traditional significance levels should be avoided in the sense of minimax regret.

Table 4 also gives least favorable values, $\theta^*_L$, $\theta^*_U$, of the non centrality parameter of a noncentral $F$ distribution. As $v$ increases, $\theta^*_L$ increases and $\theta^*_U$ slightly decreases. As $v$ increases, $\theta^*_1$ is steady at 2.1. Further, $\delta^*_L (= \delta^*_U)$ increases as $v$ increases. It indicates that the minimum value of the maximum regret function increases as the denominator degrees of freedom of the preliminary test $F$ distribution ($v$) increases.
6. Conclusion

In this article, we considered the HPT estimator whose components are the OLS and FRR estimators, and derived the exact formulae for the moments of the HPT estimator using mathematical method. Since we cannot examine the MSE performances analytically, we executed the numerical evaluation to investigate the MSE performance of the HPT estimator, and compared the MSE performance of the HPT estimator with those of the FRR estimator and the usual OLS estimator. Furthermore using the minimax regret criterion proposed by Sawa and Hiromatsu (1973), we derived the optimal critical points of the preliminary F test. Our results show that the optimal significance points are greater than 19%. As suggested by a referee, future research can be done to examine the sampling properties of a similar weighted average least squares (WALS) estimator combining the OLS and FRR estimators, which would probably exhibit superior properties to the pre-test estimator considered in this article.

Appendix

In this appendix, we derive the exact formulae for the case of the even and odd moments of the HPT estimator. First, we derive the first part of the even moments of \( \hat{\beta}_1(c) \) in Eq. (6). Let \( u_1 = x_i^1M_1x_1b_1^2/\sigma^2 \) and \( u_2 = e_1^2/\sigma^2 = vs^2/\sigma^2 \), where \( v = n - k \). Then, \( u_1 \sim \chi_{1}^{2}(\theta_1) \) and \( u_2 \sim \chi_{v}^{2}(\theta_2) \), where \( \chi_{1}^{2}(\theta_1) \) is the non central chi-square distribution with \( f \) degrees of freedom and non centrality parameter \( \theta_1 \), and \( \theta_1 = \beta_1^2x_i^1M_1x_1/\sigma^2 \). Using \( u_1 \) and \( u_2 \), \( H_1(p, q, c) \) in Eq. (6) can be written as

\[
H_1(p, q, c) = E \left[ I \left( \frac{u_1}{u_2/v} < c \right) \left( \frac{\sigma^2 u_1}{\sigma^2 u_1 + \sigma^2 u_2/v} \right)^p \left( \frac{\sigma^2}{x_i^1M_2x_1} \right)^q u_1^{1/2} \right]
\]

where

\[
K_i = \frac{w_i(\theta_i)}{2^{(v+1)/2\nu} \Gamma(1/2 + i) \Gamma(v/2)}.
\]

Making use of the change of variables, \( r_1 = u_1 + u_2 \) and \( r_2 = u_1/2 \), Eq. (18) can be written as

\[
a^q \sum_{i=0}^{\infty} K_i \int_{0}^{\infty} r_i^{(v+1)/2 + q + i - 1} e^{-r_1/2} dr_1 \int_{0}^{\infty} r_2^{1/2 + q + i - 1} (1 + r_2)^{v/2 + q + i} dr_2.
\]

Also, making use of the change of a variable, \( z = r_1/2 \), the first integral in Eq. (20) reduces to

\[
2^{(v+1)/2 + p + q + i} \Gamma((v + 1)/2 + q + i).
\]
Further, making use of the change of a variable, \( t = r_2/(1 + r_2) \), the second integral in Eq. (20) reduces to

\[
y^q \int_0^{c/(y+c)} t^{1/2+p+q+i-1} (1 - t)^{y/2-1} [1 + (y-1)t]^{-p} dt.
\] (22)

Substituting Eqs. (21) and (22) into Eq. (20), and performing some manipulations, we obtain \( H_1(p, q; c) \) in Eq. (6).

Then, we derive the first part of the odd moments of \( \hat{\beta}_1(c) \), \( J_1(p, q; c) \) in Eq. (7). Differentiating \( H_1(p, q; c) \) given in Eq. (18) with respect to \( \beta_1 \), we can write

\[
\frac{\partial H_1(p, q; c)}{\partial \beta_1} = 2^q \left( \frac{\sigma^2}{x_1' M_2 x_1} \right)^q y^q \sum_{i=0}^{\infty} \frac{\partial w_i(\theta_1)}{\partial \beta_1} G_i(p, q; c)
\]

\[
= 2^q \left( \frac{\sigma^2}{x_1' M_2 x_1} \right)^q y^q \sum_{i=0}^{\infty} \left[ -\left( \frac{x_1' M_2 x_1}{\sigma^2} \right) \beta_1 w_i(\theta_1) \right] G_i(p, q; c)
\]

\[
+ \left( \frac{x_1' M_2 x_1}{\sigma^2} \right) \beta_1 H_1(p, q; c)
\]

\[
= -\left( \frac{x_1' M_2 x_1}{\sigma^2} \right) \beta_1 H_1(p, q; c) + \beta_1 2^q \left( \frac{\sigma^2}{x_1' M_2 x_1} \right)^q \left( \frac{x_1' M_2 x_1}{\sigma^2} \right) y^q \sum_{i=0}^{\infty} w_i(\theta_1) G_{i+1}(p, q; c),
\] (23)

where we define \( w_{-1}(\theta_1) = 0 \), and

\[
G_i(p, q; c) = \frac{\Gamma(v + 1)/2 + q + i}{\Gamma(1/2 + i)\Gamma(v/2)} \int_0^{c/(y+c)} t^{1/2+p+q+i-1} (1 - t)^{y/2-1} [1 + (y-1)t]^{-p} dt.
\] (24)

On the other hand, using \( b_1 \) and \( s^2 \), \( H_1(p, q; c) \) can be expressed as

\[
H_1(p, q; c) = E \left[ I(F < c) \left( \frac{(x_1' M_2 x_1) b_1^2}{(x_1' M_2 x_1) b_1^2 + s^2} \right)^p (b_1^{-2})^q \right]
\]

\[
= \iiint_{F<\epsilon} \left( \frac{(x_1' M_2 x_1) b_1^2}{(x_1' M_2 x_1) b_1^2 + s^2} \right)^p (b_1^{-2})^q f_{b_1}(b_1) f_{s^2}(s^2) db_1 ds^2,
\] (25)

where

\[
f_{b_1}(b_1) = \frac{1}{\sqrt{2\pi\sigma}/\sqrt{x_1' M_2 x_1}} \exp \left[ -\frac{(b_1 - \beta_1)^2}{2\sigma^2/(x_1' M_2 x_1)} \right],
\] (26)

and \( f_{s^2}(s^2) \) is the probability density function of \( s^2 \). Then differentiating \( H_1(p, q; c) \) given in Eq. (25) with respect to \( \beta_1 \), we have

\[
\frac{\partial H_1(p, q; c)}{\partial \beta_1} = \iiint_{F<\epsilon} \left( \frac{x_1' M_2 x_1}{\sigma^2} \right) (b_1 - \beta_1)
\]

\[
= \iiint_{F<\epsilon} \left( \frac{x_1' M_2 x_1}{\sigma^2} \right) (b_1 - \beta_1)
\]
\[ \times \left( \frac{(x_i^TM_2x_i)b^2_1}{(x_i^TM_2x_i)b^2_1 + s^2} \right)^p (b^2_1)^q f_{b_1}(b_1)f_{\sigma^2}(s^2)db_1ds^2 \]
\[ = \left( \frac{x_i^TM_2x_i}{\sigma^2} \right) J_1(p, q; c) - \left( \frac{x_i^TM_2x_i}{\sigma^2} \right) \beta_1 H_1(p, q; c). \tag{27} \]

Equating Eqs. (23) and (27), we obtain \( J_1(p, q; c) \) in Eq. (7).

Similarly, using the same way, we can obtain \( H_2(p, q; c) \) and \( J_2(p, q; c) \) in Eqs. (6) and (7), respectively. Further, using \( H_1(p, q; c) \) and \( H_2(p, q; c) \), we have the even moments of \( \hat{\beta}_1(c) \) in Eq. (6). Similarly, using \( J_1(p, q; c) \) and \( J_2(p, q; c) \), we obtain the odd moments of \( \hat{\beta}_1(c) \) in Eq. (7).

Acknowledgments

I would like to thank Kazuhiro Ohtani for his guidance and suggestions. I would also like to thank Hisashi Tanizaki, Akio Namba, and an anonymous referee for their useful comments.

References


