SMALL SAMPLE PROPERTIES OF A RIDGE REGRESSION ESTIMATOR WITH AN INEQUALITY CONSTRAINT

By KAZUHIRO OHTANI*

In this paper, we examine the small sample properties of a ridge regression estimator proposed by Huang (1999) to estimate a particular regression coefficient in the presence of an inequality constraint. We derive the exact formulas for the bias and MSE of the inequality constrained ridge regression (ICRR) estimator. Since the exact formulas for the bias and MSE of the ICRR estimator are very complicated, we compare the bias and MSE performances of the inequality constrained least squares (ICLS) estimator and the ICRR estimator based on numerical evaluations. Our numerical results show that the bias of the ICRR estimator is larger than that of the ICLS estimator. Although the MSE of the ICRR estimator is smaller than that of the ICLS estimator when a nuisance parameter is negative or around the origin, the result is reversed when the nuisance parameter exceeds a certain positive value.

1. Introduction

In applied regression analysis, we may have special interest in estimating a particular regression coefficient as accurately as possible. Huang (1999) showed some examples of such situations, proposed a ridge regression estimator to estimate a particular regression coefficient, and examined the large and small sample properties of the proposed ridge regression estimator. Also, Ohtani (1997) examined the small sample properties of the Stein-rule (SR) estimator and the minimum mean squared error (MMSE) estimator to estimate a particular regression coefficient.

In estimating a linear regression model, economic theory sometimes suggests an inequality constraint on a regression coefficient. For example, in estimating a demand equation, it is expected from economic theory that price elasticity is negative and income elasticity is positive. Lovell and Prescott (1970), Thomson and Schmidt (1982), and Judge and Yancey (1986) established the sampling properties of the inequality constrained least squares (ICLS) estimator of regression coefficients within the standard regression framework. In the context of shrinkage estimators, Judge et al. (1984) proposed an inequality constrained Stein-rule (ICSR) estimator and showed that the ICSR estimator dominates the ICLS estimator [see also Ohtani and Wan (1998)]. Further, Wan and Ohtani (2000) examined the small sample properties of inequality constrained MMSE (ICMMSE) estimators and showed by numerical evaluations that over much of the parameter space, the ICMMSE estimators are superior to the ICLS estimator in terms of relative risk based on mean squared error (MSE). However, the small sample properties of ridge regression estimators have not been examined in the presence of an inequality constraint.

In this paper, we examine the small sample properties of the ridge regression estimator

* The author is grateful for JSPS (Japan Society for the Promotion of Science) for partial financial support (Grant-in-Aid for Scientific Research).
proposed by Huang (1999) to estimate a particular regression coefficient in the presence of an inequality constraint. In Section 2 a model and estimators are presented, and in section 3 the exact formulas for the bias and MSE of the inequality constrained ridge regression (ICRR) estimator are derived. Since the exact formulas for the bias and MSE of the ICRR estimator are very complicated, we compare the bias and MSE performances of the ICLS and ICRR estimators based on numerical evaluations in section 4. Our numerical results show that the bias of the ICRR estimator is larger than that of the ICLS estimator. Although the MSE of the ICRR estimator is smaller than that of the ICLS estimator when a nuisance parameter, defined as $\theta_1$ later, is negative or around the origin, the result is reversed when a nuisance parameter exceeds a certain positive value.

2. Model and estimator

Consider a linear regression model,

$$y = x_1 \beta_1 + X_2 \beta_2 + \varepsilon,$$

(1)

where $y$ is an $n \times 1$ vector of observations on a dependent variable, $x_1$ is an $n \times 1$ vector of observations on an explanatory variable, $X_2$ is an $n \times (k-1)$ matrix of observations on other explanatory variables, $\beta_1$ is a scalar coefficient for $x_1$, $\beta_2$ is a $k-1$ vector of coefficients for $X_2$, and $\varepsilon$ is an $n \times 1$ vector of error terms. We assume that $x_1$ and $X_2$ are nonstochastic, the $n \times k$ matrix $[x_1, X_2]$ is of full column rank, and $X_2$ is distributed as $N(0, \sigma^2 I_n)$, where $I_n$ is an $n \times n$ identity matrix. Without loss of generality, we can assume that $\beta_1$ is a particular regression coefficient which we want to estimate as accurately as possible.

Setting $X = [x_1, X_2]$ and $\beta = [\beta_1, \beta_2]'$, the least squares (LS) estimator of $\beta$ is

$$b = (X'X)^{-1}X'y,$$

(2)

and the LS estimator of $\beta_1$ is

$$b_1 = (x_1' M_2 x_1)^{-1}x_1' M_2 y,$$

(3)

where

$$M_2 = I_n - X_2(X_2'X_2)^{-1}X_2'. $$

(4)

The distribution of $b_1$ is a normal distribution with mean $\beta_1$ and variance $\sigma^2(x_1' M_2 x_1)^{-1}$:

$$b_1 \sim N(\beta_1, \sigma^2(x_1' M_2 x_1)^{-1}).$$

(5)

Following Huang (1999), the feasible ridge regression estimator of $\beta_1$ is given by
\[ \hat{\beta}_1 = \left( x_1'M_2x_1 + \frac{s^2}{b_1^2} \right)^{-1} x_1'M_2y \]

\[ = \left( \frac{(x_1'M_2x_1) b_2^2}{(x_1'M_2x_1) b_1^2 + s^2} \right) b_1, \tag{6} \]

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

In addition to the sample information, there may be prior information in the form of \( \beta_1 > 0 \). Then, the inequality constrained LS (ICLS) estimator of \( \beta_1 \) is

\[ \hat{b}_1 = I(b_1 > 0) b_1 + I(b_1 \leq 0) 0 \]

\[ = I(b_1 > 0) b_1, \tag{7} \]

where \( I(A) \) is an indicator function such that \( I(A) = 1 \) if an event \( A \) occurs and \( I(A) = 0 \) otherwise. Also, the inequality constrained ridge regression (ICRR) estimator of \( \beta_1 \) is

\[ \hat{\beta}_1^* = I(b_1 > 0) \hat{\beta}_1 + I(b_1 \leq 0) 0 \]

\[ = I(b_1 > 0) \hat{\beta}_1. \tag{8} \]

### 3. Bias and MSE of the ICRR Estimator

As is shown in the Appendix, the exact formula for the \( m \)-th moment of the ICRR estimator is written as

\[ E[(\hat{\beta}_1^*)^m] = \left( \frac{\sigma}{(x_1'M_2x_1)^{1/2}} \right)^m G(m), \tag{9} \]

where

\[ G(m) = \sum_{i=0}^{\infty} \frac{2^{(m+i)/2-1} \Gamma((m+i+1)/2) \theta_1^{i+1/2}}{\pi^{1/2} \Gamma(v/2)} i! e^{-\theta_1 i/2} \]

\[ \times \int_0^1 \frac{w^{(3m+i)/2} (1-w)^{v/2-1}}{[1+(v-1)w]^m} \, dw, \tag{10} \]

and

\[ \theta_1 = \frac{(x_1'M_2x_1)^{1/2} \beta_1}{\sigma}. \tag{11} \]

Using formula (9) and noting that \( \theta_1 = (x_1'M_2x_1)^{1/2} \beta_1 / \sigma \), the bias of the ICRR estimator is written as
Bias \( \hat{\beta}_1^* = E[\hat{\beta}_1^* - \beta_1] \)
\[ = \frac{\sigma}{(x_i'M_2x_i)^{1/2}} \left[ G(1) - \frac{(x_i'M_2x_i)^{1/2}}{\sigma} \beta_1 \right] \]
\[ = \frac{\sigma}{(x_i'M_2x_i)^{1/2}} [G(1) - \theta_1] . \] (12)

Also, the mean squared error (MSE) of the ICRR estimator is written as

\[
\text{MSE} \ (\hat{\beta}_1^*) = E[(\hat{\beta}_1^* - \beta_1)^2] \]
\[ = E[(\hat{\beta}_1^*)^2] - 2 \beta_1 E[\hat{\beta}_1^*] + \beta_1^2 \]
\[ = \frac{\sigma^2}{(x_i'M_2x_i)} \left[ (x_i'M_2x_i)^{1/2} - 2 \beta_1 \frac{(x_i'M_2x_i)^{1/2}}{\sigma} \right] G(1) + \beta_1^2 \]
\[ = \frac{\sigma^2}{(x_i'M_2x_i)} \left[ G(2) - 2 \theta_1 G(1) + \theta_1^2 \right] . \] (13)

In a similar way, the bias and MSE of the ICLS estimator are written as

\[
\text{Bias} \ (\hat{\theta}_1) = \frac{\sigma}{(x_i'M_2x_i)^{1/2}} [J(1) - \theta_1] , \] (14)

and

\[
\text{MSE} \ (\hat{\theta}_1) = \frac{\sigma^2}{(x_i'M_2x_i)} \left[ J(2) - 2 \theta_1 J(1) + \theta_1^2 \right] , \] (15)

where

\[
J(m) = \sum_{i=0}^{\infty} \frac{2^{(m+i)/2 - 1} \Gamma((m+i+1)/2)}{\pi^{1/2}} \frac{\theta_1^i}{i!} e^{-\theta_1^2} . \] (16)

4. Numerical Analysis

Since the exact formulas for the bias and MSE of the ICRR and ICLS estimators given in (12) to (15) are very complicated, their theoretical analysis is very difficult. Thus, we examine the bias and MSE performances of the ICRR and ICLS estimators by numerical evaluations. In the numerical evaluations, we numerically evaluate the relative bias and MSE of the ICRR estimator defined by
\[
\frac{\text{Bias}(\hat{\beta}_1)}{\sigma/(x_1'M_2x_1)^{1/2}} = G(1) - \theta_1, \quad (17)
\]

and

\[
\frac{\text{MSE}(\hat{\beta}_1)}{\sigma^2/(x_1'M_2x_1)} = G(2) - 2\theta_1 G(1) + \theta_1^2. \quad (18)
\]

In a similar way, the relative bias and MSE of the ICLS estimator are defined.

Since the MSE of the LS estimator is \(\sigma^2/(x_1'M_2x_1)\), the relative MSE's of the ICRR and ICLS estimators are relative MSE's to the LS estimator. Thus, if the relative MSE's of the ICRR and ICLS estimators are smaller than unity, the ICRR and ICLS estimators are more efficient than the LS estimator in terms of MSE.

The parameter values used in the numerical evaluations are as follows: \(k = 2, 3, 4, 8; n = 20, 30, 40; \theta_1 = \text{various values}\). The infinite series in the formulas of \(G(m)\) and \(J(m)\) given in (10) and (16) are judged to converge when the increment becomes smaller than \(10^{-12}\). Also, the integral in \(G(m)\) is numerically calculated based on the Simpson's 3/8 rule with 200 equal subdivisions.

Table 1 and Figures 1 and 2 show the bias and MSE performances of the ICLS and ICRR estimators for \(k = 3\) and \(n = 20\). Since the results for \(k = 3\) and \(n = 20\) are typical, we discuss the results based on Table 1 and Figures 1 and 2.

TABLE 1. Bias and MSE of the ICLS and ICRR estimators for \(k = 3\) and \(n = 20\)

<table>
<thead>
<tr>
<th>(\theta_1)</th>
<th>ICLS</th>
<th>ICLS</th>
<th>ICRR</th>
<th>ICRR</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bias</td>
<td>MSE</td>
<td>Bias</td>
<td>MSE</td>
</tr>
<tr>
<td>(-1.5)</td>
<td>1.529</td>
<td>2.361</td>
<td>1.511</td>
<td>2.288</td>
</tr>
<tr>
<td>(-1.4)</td>
<td>1.437</td>
<td>2.092</td>
<td>1.414</td>
<td>2.007</td>
</tr>
<tr>
<td>(-1.3)</td>
<td>1.346</td>
<td>1.846</td>
<td>1.317</td>
<td>1.747</td>
</tr>
<tr>
<td>(-1.2)</td>
<td>1.256</td>
<td>1.622</td>
<td>1.222</td>
<td>1.508</td>
</tr>
<tr>
<td>(-1.1)</td>
<td>1.169</td>
<td>1.421</td>
<td>1.128</td>
<td>1.291</td>
</tr>
<tr>
<td>(-1.0)</td>
<td>1.083</td>
<td>1.242</td>
<td>1.035</td>
<td>1.096</td>
</tr>
<tr>
<td>(-0.9)</td>
<td>1.000</td>
<td>1.084</td>
<td>0.943</td>
<td>0.921</td>
</tr>
<tr>
<td>(-0.8)</td>
<td>0.920</td>
<td>0.948</td>
<td>0.853</td>
<td>0.768</td>
</tr>
<tr>
<td>(-0.7)</td>
<td>0.843</td>
<td>0.832</td>
<td>0.765</td>
<td>0.635</td>
</tr>
<tr>
<td>(-0.6)</td>
<td>0.769</td>
<td>0.735</td>
<td>0.679</td>
<td>0.523</td>
</tr>
<tr>
<td>(-0.5)</td>
<td>0.698</td>
<td>0.657</td>
<td>0.595</td>
<td>0.431</td>
</tr>
<tr>
<td>(-0.4)</td>
<td>0.630</td>
<td>0.597</td>
<td>0.513</td>
<td>0.358</td>
</tr>
<tr>
<td>(-0.3)</td>
<td>0.567</td>
<td>0.552</td>
<td>0.435</td>
<td>0.303</td>
</tr>
<tr>
<td>(-0.2)</td>
<td>0.507</td>
<td>0.522</td>
<td>0.360</td>
<td>0.266</td>
</tr>
<tr>
<td>(-0.1)</td>
<td>0.451</td>
<td>0.505</td>
<td>0.287</td>
<td>0.246</td>
</tr>
<tr>
<td>(-0.0)</td>
<td>0.399</td>
<td>0.500</td>
<td>0.219</td>
<td>0.240</td>
</tr>
<tr>
<td>(0.1)</td>
<td>0.351</td>
<td>0.505</td>
<td>0.154</td>
<td>0.248</td>
</tr>
<tr>
<td>(0.2)</td>
<td>0.307</td>
<td>0.518</td>
<td>0.093</td>
<td>0.269</td>
</tr>
<tr>
<td>(0.3)</td>
<td>0.267</td>
<td>0.538</td>
<td>0.036</td>
<td>0.301</td>
</tr>
<tr>
<td>(0.4)</td>
<td>0.230</td>
<td>0.563</td>
<td>-0.017</td>
<td>0.342</td>
</tr>
<tr>
<td>(0.5)</td>
<td>0.198</td>
<td>0.593</td>
<td>-0.066</td>
<td>0.390</td>
</tr>
<tr>
<td>(0.6)</td>
<td>0.169</td>
<td>0.625</td>
<td>-0.111</td>
<td>0.445</td>
</tr>
<tr>
<td>(0.7)</td>
<td>0.143</td>
<td>0.658</td>
<td>-0.151</td>
<td>0.504</td>
</tr>
</tbody>
</table>
We see from Table 1 and Figure 1 that the ICLS estimator has positive bias when the inequality constraint is not true (i.e., $\theta_1 = [(\alpha_1' M \alpha_1)^{1/2}/\sigma] \beta_1 \leq 0$ and thus $\beta_1 \leq 0$). When
the inequality constraint is true (\( \theta_1 > 0 \)) and the value of \( \theta_1 \) is smaller than 2.0, the ICLS estimator has positive bias. However, as the value of \( \theta_1 \) exceeds 2.0, the bias almost disappears. Although the ICRR estimator has positive bias when the inequality constraint is not true (i.e., \( \theta_1 \leq 0 \)) or the value of \( \theta_1 \) is less than 0.4, it has negative bias when the value of \( \theta_1 \) exceeds 0.4 and the bias hardly disappears. As a whole, the ICLS estimator has smaller bias than the ICRR estimator.

We see from Table 1 and Figure 2 that the relative MSE of the ICLS estimator decreases as the value of \( \theta_1 \) increases from a negative value to zero, attains a minimum, and then increases and approaches unity from below as the value of \( \theta_1 \) increases from zero. Also, the relative MSE of the ICRR estimator decreases as the value of \( \theta_1 \) increases from a negative value to zero, and attains a minimum. However, as the value of \( \theta_1 \) increases from zero, the MSE of the ICRR estimator increases, becomes larger than unity, attains a maximum, and then approaches unity from above. Since the relative MSE of the LS estimator is unity, both the ICLS and ICRR estimators have a much smaller MSE than the LS estimator around \( \theta_1 = 0 \).

Although the ICRR estimator has a smaller MSE than the ICLS estimator when the value of \( \theta_1 \) is smaller than 1.1, the ICRR estimator has a larger MSE than the ICLS estimator when the value of \( \theta_1 \) exceeds 1.2. In addition, the ICRR estimator has a larger MSE than the LS estimator when the value of \( \theta_1 \) is larger than 1.5, while the ICLS estimator always has a smaller MSE than the LS estimator when the inequality constraint is true (i.e., \( \theta_1 > 0 \)). Comparing the maximum gain of the MSE of the ICRR estimator at \( \theta_1 = 0 \) and the maximum loss around \( \theta_1 = 2.7 \), the gain is larger than the loss. These results show that the ICRR estimator is preferable to the ICLS estimator in terms of MSE if we hope to obtain the maximum gain in MSE. However, the ICLS estimator is preferable to the ICRR estimator in terms of MSE if we hope to avoid the maximum loss in MSE.

Our results do not support a definite choice between the ICRR and ICLS estimators. Wan (1994) considered the pre-test ICLS estimator such that the ICLS estimator is used if the inequality constraint is accepted in the pre-test and the LS estimator is used if the inequality constraint is rejected. In a similar way, the pre-test ICRR estimator may be considered. A comparison of the risk performances of the pre-test ICLS and ICRR estimators might yield more unambiguous results. However, since consideration of such a pre-test estimator is beyond the scope of this paper, this problem remains.

Appendix

In this appendix, we derive the exact formula for the \( m \)-th moment of the ICRR estimator.

Denoting \( u_1 = (x_1' M_2 x_1)^{1/2} b_1 / \sigma \) and \( u_2 = e' e / \sigma^2 \), \( u_1 \) is distributed as \( u_1 \sim N(\theta_1, 1) \), \( \theta_1 = (x_1' M_2 x_1)^{1/2} \beta_1 / \sigma \), and \( u_2 \) as a chi-square distribution with \( v = n - k \) degrees of freedom. Note that \( u_1 \) and \( u_2 \) are mutually independent. Since \( (x_1' M_2 x_1)^{1/2} / \sigma \) is positive, \( b_1 > 0 \) and \( u_1 > 0 \) are equivalent. Thus, using \( u_1 \) and \( u_2 \), the ICRR estimator is written as
\[ \hat{\beta}_1 = I(b_1 > 0) \beta_1 \]

Then, the \( m \)-th moment of the ICRR estimator is written as

\[ E[(\hat{\beta}_1)^m] = \int_0^{\infty} \int_0^{\infty} \left( \frac{\sigma}{(x_1^2 M_2 x_1)^{1/2}} \right)^m u_1^m \frac{1}{(2\pi)^{1/2}} \exp\left[-(u_1 - \theta_1)^2/2\right] \times \frac{1}{2^{v/2} \Gamma(v/2)} u_2^{v'/2-1} \exp(-u_2/2) du_1 du_2. \]  

Using the formula \( \exp(x) = \sum_{i=0}^{\infty} x^i/i! \), we obtain

\[ \exp[-(u_1 - \theta_1)^2/2] = \exp(-u_1^2/2) \exp(\theta_1 u_1) \exp(-\theta_1^2/2) = \exp(-u_1^2/2) \left( \sum_{i=0}^{\infty} \frac{\theta_1^i u_1^i}{i!} \right) \exp(-\theta_1^2/2). \]

Thus, (20) reduces to

\[ \sum_{i=0}^{\infty} K_i \int_0^{\infty} \int_0^{\infty} \left( \frac{u_1^2}{u_1^2 + u_2^2} \right)^m u_1^{m+i} \exp(-u_1^2/2) u_2^{v'/2-1} \exp(-u_2/2) du_1 du_2, \]

where

\[ K_i = \left( \frac{\beta_1}{\theta_1} \right)^m \frac{1}{\pi^{1/2} 2^{(i+1)/2} \Gamma(v_i/2)} \frac{\theta_1^i}{i!} \exp(-\theta_1^2/2). \]

Making use of the change of variable, \( v_1 = u_1^2 \), (22) reduces to

\[ \sum_{i=0}^{\infty} \frac{K_i}{2} \int_0^{\infty} \int_0^{\infty} \left( \frac{v_1}{v_1 + u_2^2} \right)^m v_1^{(m+i+1)/2} u_2^{v'/2-1} \exp[-(v_1 + u_2^2)/2] dv_1 du_2. \]

Again, making use of the change of variables, \( t_1 = v_1 + u_2 \) and \( t_2 = v_1/u_2 \), (24) reduces to

\[ \sum_{i=0}^{\infty} \frac{K_i}{2} \int_0^{\infty} t_1^{(m+i+1)/2} - t_1/2 dt_1 \int_0^{\infty} \frac{t_2^{(3m+i+1)/2}}{(1 + t_2^{(m+i+1)/2})^{m+1/2}} dt_2. \]
\[
\sum_{i=0}^{\infty} \frac{K_i^2 2^{(m+v+i+1)/2}}{\Gamma((m + v + i +1)/2)} \int_0^{\infty} t_2^{(3m+i-1)/2} \left(1 + t_2^{m+v+i+1}/2\right) \frac{t_2^{(m+v+i+1)/2} (t_2 + 1/V)^m}{(t_2 + 1/V)^m} \, dt_2.
\]

Finally, making use of the change of variable, \( w = t_2/(1 + t_2) \), and performing some manipulations, we obtain (9) in the text.

Professor, Graduate School of Economics, Kobe University

REFERENCES