Chapter 2

Developing biased Estimators as Alternative to Least Squares Estimator

As Chapter 1 explains the ways for dealing with MC, one of the strategies is using biased estimation methods. We believe that biased estimation methods are useful techniques that analysts should consider when dealing with MC. Marquardt and Snee (1975) observed, "it is often better to use some of the information in all of the regressors as biased estimation does, than to use all of the information in some regressors and none of the information in others, as variable selection does" (see Wadsworth, 1998). One of these estimation techniques is ordinary ridge regression (ORR), originally proposed by Hoerl and Kennard (1970).

ORR attacks the problem of MC by adding a small constant to the diagonal of \(X'X\) to improve its condition number. ORR can be obtained by augmenting the equation \(0 = k^{1/2}\beta + \epsilon'\) to the linear model in (1.1) where \(\epsilon'\) is a random vector of disturbances with \(\text{E}(\epsilon') = 0, \text{Var}(\epsilon') = \sigma^2I\) and \(\text{E}(\epsilon'\epsilon'') = 0\), then using the least squares method. As \(k\) becomes larger, the distance between \(k^{1/2}\beta\) and 0 increases. Therefore, augmenting \(0 = k^{1/2}\beta + \epsilon'\) to (1.1) introduces more bias to ORR. Therefore, it is desirable to select a small \(k\) and then \(X'X + kI\) may still
be ill-conditioned. Note that the condition number of $X'X + kI$ is a decreasing function of $k$. Therefore, if we want the condition number of $X'X + kI$ to be small, the value of $k$ should be large. Due to this reason, a small value of $k$ may not be enough to overcome the problem of MC, especially when the design matrix $X'X$ suffers strongly from singularity. In this case, $X'X + kI$ is still ill-conditioned and the resulting ORR may still be unstable.

In this chapter, we consider alternative biased estimators to reduce the effect of MC. For this, we first develop the $Al$ estimator. Then we propose new estimates of biasing parameters of shrinkage estimators. Two shrinkage type estimators are also introduced.

### 2.1 $Al$ estimator

Liu (1993) introduced a biased estimator as an alternative to the OLS when MC occurs. When two different estimators are available for a parameter, it is hoped that a combination of these two would inherit the advantages of both. With this view, Liu combined the ORR with the Stein estimator $\hat{\beta}_s = c\hat{\beta}$, where $0 < c < 1$ is a parameter (Stein 1956, James and Stein 1961), using the following argument.

ORR has the disadvantage of being unstable and it is a complicated function of $k$; but the advantage of ORR is its effectiveness in practice. Also, Stein estimator has an advantage that it is a linear function of $c$ while its disadvantage is that shrinkage effect is same for each component of OLS estimator. To overcome disadvantages of these two estimators, Liu defined the following estimators.

\[
\hat{\beta}_d = (X'X + I)^{-1}(X'Y + d\hat{\beta}), \quad 0 < d < 1. \tag{2.1}
\]

and

\[
\hat{\beta}_D = (X'X + I)^{-1}(X'Y + D\hat{\beta}), \tag{2.2}
\]

\[
\hat{\beta}_d = (X'X + I)^{-1}(X'Y + d\hat{\beta}), \quad 0 < d < 1. \tag{2.1}
\]

and

\[
\hat{\beta}_D = (X'X + I)^{-1}(X'Y + D\hat{\beta}), \tag{2.2}
\]
where \( D = \text{diag}\{d_1,d_2,...,d_p\} \) is the diagonal matrix of the biasing parameter \( d_i \) of Liu estimator. \( \hat{\beta}_d \) which is called Liu estimator (LE) and \( \hat{\beta}_D \) is called generalized Liu estimator (GLE). The advantage of \( \hat{\beta}_d \) over \( \hat{\beta}_k \) is that \( \hat{\beta}_d \) is a linear function of \( d \). However, \( \hat{\beta}_d \) still has the following problem. Since we want to find the LE, we have to use \( \hat{\beta} \) which is already unstable and often gives confusing results in the presence of MC.

The \( Al \) estimator is motivated by the following fact: The ordinary ridge estimator \( \hat{\beta}_k \) can be obtained by augmenting \( 0 = k^{1/2}\beta + \epsilon \) to (1.1) and then using the least squares method. Also, we can get the Liu estimator \( \hat{\beta}_d \) by augmenting the equation \( d\hat{\beta} = \beta + \epsilon \) to (1.1) and then using the least squares method. If we augment \( m(X'X)\hat{\beta} = \beta + \epsilon \) to (1.1) and application of the least squares method will give

\[
\hat{\beta}_m = (X'X + I)^{-1}(X'Y + m(X'X)\hat{\beta}), \quad 0 < m < \infty \tag{2.3}
\]

where \( m \) is a biasing parameter. The similarities among the \( Al \) estimator \( \hat{\beta}_m \), ORR estimator and GLE estimator are as follows:

\[
\begin{align*}
\hat{\beta}_{m=0} &= \hat{\beta}_{k=1}, \\
\hat{\beta}_m &= \hat{\beta}_{(D=mX'X)}.
\end{align*}
\]

It is well known that a linear regression model can be transformed to a canonical form by orthogonal transformation. The matrix \( X'X \) is a \( p \times p \) symmetric, non-singular matrix and can be reduced to a diagonal matrix by spectral decomposition \( V'X'XV = \Lambda \) where \( V \) is an orthogonal matrix and \( \Lambda \) is a \( p \times p \) diagonal matrix of the eigenvalues of \( X'X \), \( \lambda_1 > ... > \lambda_p \). The columns of \( V \) are the eigenvectors of \( X'X \). So, we can rewrite model (1.1) in the canonical form:

\[
Y = Z\alpha + \epsilon, \tag{2.4}
\]

where \( Z = XV \) and \( \alpha = V'\beta \). The advantage of the canonical form is that \( \alpha \) is
unit vector. For model (2.4), OLS estimator, ORR estimator, LE estimator and AI estimator are written as:

\[ \hat{\alpha} = \Lambda^{-1}Z'Y = A_1Y, \]  
\[ \hat{\alpha}_k = (\Lambda + kI)^{-1}Z'Y = A_2Y, \]  
\[ \hat{\alpha}_d = (\Lambda + I)^{-1}(I + d\Lambda^{-1})Z'Y = A_3Y, \]  
\[ \hat{\alpha}_m = (\Lambda + I)^{-1}A(I + mI)Z'Y = A_4Y, \]

respectively.

### 2.1.1 Criteria used in the comparison of biased estimation

In biased estimation, some biases are introduced while reducing inflation in the variance. Therefore, the objective is to strike a good balance between the bias and the variance. The two loss functions that can achieve this balance are

\[ \text{gmse}(\hat{\beta}) = E \left[ (\hat{\beta} - \beta)'B(\hat{\beta} - \beta) \right], \]
\[ \text{mspe}(\hat{\beta}) = E \left[ (\hat{\beta} - \beta)'X'X(\hat{\beta} - \beta) \right], \]

where \( B \) is a non-negative definite (n. n. d.) matrix. gmse is the generalized mean squared error of \( \hat{\beta} \), and mspe is the mean squared prediction error. The comparison of biased estimators with OLS has to be made in terms of a given \( B \). We assume \( B = I \) in this thesis. Therefore, the measure of goodness of fit will be

\[ \text{mse}(\hat{\beta}) = E \left[ (\hat{\beta} - \beta)'(\hat{\beta} - \beta) \right], \]
which is the scalar mean squared error (mse).

The matrix mean squared error (MMSE) of any two estimators $\hat{\theta}_1$ of $\hat{\theta}_2$ are given by

$$\text{MMSE}(\hat{\theta}_j) = \mathbb{E} \left[ \left( \hat{\theta}_j - \theta \right) \left( \hat{\theta}_j - \theta \right)' \right], j = 1, 2.$$

Theobald (1974) proved that $\text{gmse}(\hat{\theta}_1) > \text{gmse}(\hat{\theta}_2)$ for all n.n.d. matrices $B$ if and only if $\text{MMSE}(\hat{\theta}_1) - \text{MMSE}(\hat{\theta}_2)$ is an n.n.d. matrix. Thus, the superiority of $\hat{\theta}_2$ over $\hat{\theta}_1$ with respect to the MMSE criterion are examined by comparing the mean squared error matrices.

The MMSE of an estimator $\tilde{\theta}$ of $\theta$ can be written as follows:

$$\text{MMSE}(\tilde{\theta}) = \text{Cov}(\tilde{\theta}) + \left( \text{bias}(\tilde{\theta}) \right) \left( \text{bias}(\tilde{\theta}) \right)'$$

where $\text{Cov}(\tilde{\theta}) = \mathbb{E} \left[ \left( \tilde{\theta} - \mathbb{E}(\tilde{\theta}) \right) \left( \tilde{\theta} - \mathbb{E}(\tilde{\theta}) \right)' \right]$ is the variance-covariance matrix of $\tilde{\theta}$ and $\text{bias}(\tilde{\theta}) = \mathbb{E}(\tilde{\theta}) - \theta$ is the bias of $\tilde{\theta}$. Also, the mse can be written as follows:

$$\text{mse}(\tilde{\theta}) = \text{tr} \text{MMSE}(\tilde{\theta}) = \text{tr} \text{Cov}(\tilde{\theta}) + \left\| \mathbb{E}(\tilde{\theta}) - \theta \right\|^2,$$

where tr denotes trace.

The mean squared errors of $\hat{\theta}_b$ and $\hat{\alpha}_b$ are the same, where $\hat{\theta}_b$ can be any estimator and $\hat{\alpha}_b$ is its canonical form, since

$$\text{mse}(\hat{\theta}_b) = \mathbb{E} \left( \beta - \hat{\beta}_d \right)' \left( \beta - \hat{\beta}_d \right)$$

$$= \mathbb{E} \left( V\alpha - V\hat{\alpha}_d \right)' \left( V\alpha - V\hat{\alpha}_d \right)$$

$$= \mathbb{E} \left( \alpha - \hat{\alpha}_d \right)' V'^{V} \left( \alpha - \hat{\alpha}_d \right)$$

$$= \mathbb{E} \left( \alpha - \hat{\alpha}_d \right)' \left( \alpha - \hat{\alpha}_d \right)$$

$$= \text{mse}(\hat{\alpha}_d).$$
For any two estimators $\tilde{\beta}_j = A_j Y$, $j = 1, 2$, we have

$$\Delta = \text{MMSE}(\tilde{\beta}_1) - \text{MMSE}(\tilde{\beta}_2)$$
$$= \text{Cov}(\tilde{\beta}_1) - \text{Cov}(\tilde{\beta}_2) + (\text{bias}(\tilde{\beta}_1))(\text{bias}(\tilde{\beta}_1))'$$
$$- (\text{bias}(\tilde{\beta}_2))(\text{bias}(\tilde{\beta}_2))'$$
$$= \sigma^2(A_1A_1' - A_2A_2') + (\text{bias}(\tilde{\beta}_1))(\text{bias}(\tilde{\beta}_1))'$$
$$- (\text{bias}(\tilde{\beta}_2))(\text{bias}(\tilde{\beta}_2))'.$$

In order to inspect whether $\Delta$ is positive definite (p.d.), we may confine ourselves to the following case:

$$G = \text{Cov}(\tilde{\beta}_1) - \text{Cov}(\tilde{\beta}_2) > 0 \quad (p.d.).$$

As $(\text{bias}(\tilde{\beta}_1))(\text{bias}(\tilde{\beta}_1))' > 0$, it is easy to see that $G > 0$ implies $G + (\text{bias}(\tilde{\beta}_1))(\text{bias}(\tilde{\beta}_1))' > 0$ (see Rao et al., 2008). Hence the problem of deciding whether $\Delta > 0$ reduces to that of deciding whether a matrix of type $A - cc'$ is p.d. when $A$ is p.d. Then we have the following result.

**Lemma 2.1** (Farebrother, 1976): Let $A$ be a positive definite matrix, $c$ be a nonzero vector and $\theta$ be a positive scalar. Then $A - cc'$ is p.d. iff $c'A^{-1}c < \theta$.

For this, we have the following lemma.

**Lemma 2.2** (Trenkler, 1980): Let $\tilde{\beta}_j = A_j Y$, $j = 1, 2$, be two homogeneous linear estimators of $\beta$ such that $(A_1A_1' - A_2A_2')$ is p.d. If

$$\beta'(A_2X - I)'(A_1A_1' - A_2A_2')^{-1}(A_2X - I)\beta < \sigma^2,$$

then $\Delta$ is p. d.

### 2.1.2 Comparison of the $A_1$ estimator with other estimators

In this section, we compare $A_1$ estimator with OLS estimator, ORR estimator and LE estimator in terms of MMSE.
Comparison between the Al estimator and the OLS estimator

The variance-covariance matrix and bias of $\hat{a}(m)$ are given by

$$\text{Cov}(\hat{a}(m)) = \sigma^2 (A + I)^{-1} (I + mI)(A + I)(A + I)^{-1} \quad (2.9)$$

$$\text{bias}(\hat{a}(m)) = (A + I)^{-1} (A - mI)\alpha \alpha'(A - mI)(A + I)^{-1} \quad (2.10)$$

The variance-covariance matrix of $\hat{a}$ is given by

$$\text{Cov}(\hat{a}) = \sigma^2 A^{-1} \quad (2.11)$$

By combining (2.9) and (2.10), we obtain the MMSE of Al estimator

$$\text{MMSE}(\hat{a}(m)) = \sigma^2 (A + I)^{-1} (I + mI)(A + I)(A + I)^{-1} + (A + I)^{-1} (A - mI)\alpha \alpha'(A - mI)(A + I)^{-1} \quad (2.12)$$

Also,

$$\text{MMSE}(\hat{a}) = \sigma^2 A^{-1}. \quad (2.13)$$

Now, we have the following theorem.

**Theorem 2.1** If $\text{bias}(\hat{a}(m))'(A_1 A_1' - A_4 A_4')^{-1} \text{bias}(\hat{a}(m)) < \sigma^2$, then $\text{MMSE}(\hat{a}) - \text{MMSE}(\hat{a}(m))$ is p.d. for $0 < m < 1/\lambda_i$, $i = 1, 2, ..., p.$

**Proof:**

Using the estimators $\hat{a}$ and $\hat{a}(m)$ in (2.5) and (2.8) respectively, we obtain

$$\text{Cov}(\hat{a}) - \text{Cov}(\hat{a}(m)) = \sigma^2 (A_1 A_1' - A_4 A_4')$$

$$= \sigma^2 \left[ (A^{-1} - (A + I)^{-1}(I + mI)\Lambda (I + mI)(A + I)^{-1} \right]$$
\[ \sigma^2 \left[ \text{diag} \left\{ \frac{1}{\lambda_i} - \frac{(1 + m)^2 \lambda_i}{(\lambda_i + 1)^2} \right\}_{i=1}^p \right] \]

\( (A_1 A_1' - A_4 A_4') \) will be p. d. if and only if \( \left[ \text{diag} \left\{ \frac{1}{\lambda_i} - \frac{(1 + m)^2 \lambda_i}{(\lambda_i + 1)^2} \right\}_{i=1}^p \right] \) is p. d.

\( \Leftrightarrow \lambda_i(1 + m) - (\lambda_i + 1) < 0. \)

This inequality requires that \( (A_1 A_1' - A_4 A_4') \) is p.d. for \( 0 < m < 1/\lambda_i \), \( i = 1, 2, \ldots, p. \) By lemma 2.1, the proof of Theorem (2.1) is complete. \( \square \)

**Comparison between the Al estimator and ORR estimator**

The variance-covariance matrix and bias of \( \hat{\alpha}_k \) are given by

\[
\begin{align*}
\text{Cov}(\hat{\alpha}_k) &= \sigma^2(\Lambda + kI)^{-1}\Lambda(\Lambda + kI)^{-1} \\
\text{bias}(\hat{\alpha}_k) &= k^2(\Lambda + kI)^{-1}\alpha\alpha'(\Lambda + kI)^{-1} \\
\text{MMSE}(\hat{\alpha}_k) &= \sigma^2(\Lambda + kI)^{-1}\Lambda(\Lambda + kI)^{-1} + k^2 \\
&\quad \times (\Lambda + kI)^{-1}\alpha\alpha'(\Lambda + kI)^{-1}. \quad (2.14)
\end{align*}
\]

Let us now fix \( m \) such that \( m\lambda_i < 1 \) for all \( i = 1, 2, \ldots, p. \) So, we state the following theorem:

**Theorem 2.2**

\( a) \) If \( \text{bias}(\hat{\alpha}_m)(A_2 A_2' - A_4 A_4')^{-1}\text{bias}(\hat{\alpha}_m) < \sigma^2 \), then \( \text{MMSE}(\hat{\alpha}_k) - \text{MMSE}(\hat{\alpha}_m) \) is p. d. for \( 0 < k^* < k \).

\( b) \) If \( \text{bias}(\hat{\alpha}_k)(A_2 A_2' - A_4 A_4')^{-1}\text{bias}(\hat{\alpha}_k) < \sigma^2 \), then \( \text{MMSE}(\hat{\alpha}_m) - \text{MMSE}(\hat{\alpha}_k) \) is p.d. for \( 0 < k^* < k \).

**Proof:**

Using the estimators in (2.6) and (2.8) we obtain

\[
\begin{align*}
\text{Cov}(\hat{\alpha}_k) - \text{Cov}(\hat{\alpha}_m) &= \sigma^2(A_2 A_2' - A_4 A_4') \\
&= \sigma^2 \left[ (\Lambda + kI)^{-1}\Lambda(\Lambda + kI)^{-1} - (\Lambda + kI)^{-1}(I + mI)\Lambda(I + mI)(\Lambda + I)^{-1} \right] \\
&= \sigma^2 \left\{ \text{diag} \left\{ \frac{\lambda_j}{(\lambda_j + k)^2} - \frac{(1 + m)^2 \lambda_i}{(\lambda_i + 1)^2} \right\}_{j=1}^p \right\}.
\end{align*}
\]
Clearly, \((A_2A'_2 - A_4A'_4)\) is p. d. if and only if \((\lambda_j + 1)(\lambda_j + k)(1 + m) < 0\).

We solve this inequality to get

\[
k_j^* = \frac{1 - m\lambda_j}{1 + m}, \quad j = 1, 2, ..., p.
\]

So, \((A_2A'_2 - A_4A'_4)\) is p.d. for \(0 < k < k_j^*\). Similarly, \((A_4A'_4 - A_2A'_2)\) will be p.d. if and only if \((\lambda_j + k)(m + 1) - (\lambda_j + 1) > 0\). This inequality requires that \((A_4A'_4 - A_2A'_2)\) is p.d. for \(0 < k_j^* < k\). The proof is complete by lemma (2.2).

Let us now fix \(k\) such that \(0 < k < 1\). Then, we have the following theorem.

**Theorem 2.3**  
a) If \(bias(\hat{\alpha}_k)'(A_4A'_4 - A_2A'_2)^{-1}bias(\hat{\alpha}_k) < \sigma^2\), then \(MMSE(\hat{\alpha}_m) - MMSE(\hat{\alpha}_k)\) is p.d. for \(0 < m < m_j^*\).

b) If \(bias(\hat{\alpha}_m)'(A_2A'_2 - A_4A'_4)^{-1}bias(\hat{\alpha}_m) < \sigma^2\), then \(MMSE(\hat{\alpha}_k) - MMSE(\hat{\alpha}_m)\) is p.d. for \(0 < m < m_j^*\).

Here,

\[
m_j^* = \frac{1 - k}{\lambda_j + k}, \quad j = 1, 2, ..., p.
\]

**Proof:**

Use the estimators \(\hat{\alpha}_k\) and \(\hat{\alpha}_m\) in (2.6) and (2.8) to obtain

\[
\begin{align*}
\text{Cov}(\hat{\alpha}_m) - \text{Cov}(\hat{\alpha}_k) &= \sigma^2(A_4A'_4 - A_2A'_2) \\
&= \sigma^2 \left[ (\Lambda + I)^{-1}(I + mI)(\Lambda + I)^{-1} \right] \\
&= \sigma^2 \left[ \text{diag} \left\{ \frac{(1 + m)^2 \lambda_j}{(\lambda_j + 1)^2} - \frac{\lambda_j}{(\lambda_j + k)^2} \right\}_{j=1}^p \right].
\end{align*}
\]

Evidently, \((A_4A'_4 - A_2A'_2)\) will be p.d. if and only if \((\lambda_j + k)(1 + m) - (\lambda_j + 1) > 0\). By solving this inequality we get

\[
m_j^* = \frac{1 - k}{\lambda_j + k}, \quad j = 1, 2, ..., p.
\]
So, \((A_4A'_4 - A_2A'_2)\) is p.d. for \(0 < m'_* < m\). Similarly, \((A_2A'_2 - A_4A'_4)\) will be p.d. if and only if \((\lambda_j + k)(1 + m) - (\lambda_j + 1) < 0\) This inequality requires that \((A_2A'_2 - A_4A'_4)\) is p.d. for \(0 < m < m'_*\). The proof is complete by lemma (2.2).

\[\Box\]

**Comparison between the \(A_l\) estimator and the LE estimator**

The variance-covariance matrix and bias of \(\hat{\alpha}_d\) are given respectively by

\[
\text{Cov}(\hat{\alpha}_d) = \sigma^2(\Lambda + I)^{-1}(\Lambda + dI)\Lambda - 1(\Lambda + dI)(\Lambda + I)^{-1}
\]

\[
\text{bias}(\hat{\alpha}_d) = (\Lambda + I)^{-1}(I - dI)\alpha \alpha' (I - dI)(\Lambda + I)^{-1}
\]

\[
\text{MMSE}(\hat{\alpha}_d) = \sigma^2(\Lambda + I)^{-1}(\Lambda + dI)\Lambda - 1(\Lambda + dI)(\Lambda + I)^{-1} + \\
(\Lambda + I)^{-1}(I - dI)\alpha \alpha' (I - dI)(\Lambda + I)^{-1}.
\]

Let us fix \(d\). We state the following theorem:

**Theorem 2.4**

\(a\) If \(\text{bias}(\hat{\alpha}_d)'(A_4A'_4 - A_3A'_3)^{-1}\text{bias}(\hat{\alpha}_d) < \sigma^2\), then \(\text{MMSE}(\hat{\alpha}_m) - \text{MMSE}(\hat{\alpha}_d)\) is p. d. for \(0 < m'_* < m\).

\(b\) If \(\text{bias}(\hat{\alpha}_m)'(A_3A'_3 - A_4A'_4)^{-1}\text{bias}(\hat{\alpha}_m) < \sigma^2\), then \(\text{MMSE}(\hat{\alpha}_d) - \text{MMSE}(\hat{\alpha}_m)\) is p. d. for \(0 < m < m'_*\).

Here,

\[ m'_* = \frac{d}{\lambda_j}, \quad j = 1, 2, \ldots, p. \]

**Proof:**

Use the estimators \(\hat{\alpha}_d\) and \(\hat{\alpha}_m\) in (2.7) and (2.8) to obtain

\[
\text{Cov}(\hat{\alpha}_m) - \text{Cov}(\hat{\alpha}_d) = \sigma^2(A_4A'_4 - A_3A'_3)
\]

\[
= \sigma^2[(\Lambda + I)^{-1}(I + mI)\Lambda(I + mI)(\Lambda + I)^{-1} - \\
(\Lambda + I)^{-1}(\Lambda + dI)\Lambda^{-1}(\Lambda + dI)(\Lambda + I)^{-1}]
\]

\[
= \sigma^2 \left[ \text{diag} \left\{ \frac{(1 + m)^2 \lambda_j}{(\lambda_j + 1)^2} - \frac{(\lambda_j + d)^2}{\lambda_j(\lambda_j + 1)^2} \right\}^p \right].
\]
\[(A_4A_3' - A_3A_4')\] is p.d. if and only if \(\lambda_j(1 + m) - (\lambda_j + d) > 0\). This inequality requires that \((A_4A_3' - A_3A_4')\) is p.d. for \(0 < m^*_j < m\). Similarly, \((A_3A_4' - A_4A_3')\) is p.d. if and only if \(\lambda_j(1 + m) - (\lambda_j + d) < 0\). This inequality requires that \((A_3A_4' - A_4A_3')\) is p.d. for \(0 < m < m^*_j\). The proof is complete by lemma (2.2). □

Let us fix \(m\) such that \(m\lambda_j < 1\) and let \(0 < d < d^*_j\). Then, we have the following theorem.

**Theorem 2.5**

a) If \(\text{bias}(\hat{\alpha}_d)'(A_4A_3' - A_3A_4')^{-1}\text{bias}(\hat{\alpha}_d) < \sigma^2\), then \(\text{MMSE}(\hat{\alpha}_m) - \text{MMSE}(\hat{\alpha}_d)\) is p. d. for \(0 < d^*_j < d\).

b) If \(\text{bias}(\hat{\alpha}_m)'(A_3A_4' - A_4A_3')^{-1}\text{bias}(\hat{\alpha}_m) < \sigma^2\), then \(\text{MMSE}(\hat{\alpha}_d) - \text{MMSE}(\hat{\alpha}_m)\) is p. d. for \(0 < d < d^*_j\).

Here, \(d^*_j = m\lambda_j\).

**Proof:**

Use the estimators \(\hat{\alpha}_d\) and \(\hat{\alpha}_m\) in (2.7) and (2.8) to obtain

\[
\text{Cov}(\hat{\alpha}_d) - \text{Cov}(\hat{\alpha}_m) = \sigma^2(A_3A_4' - A_4A_3')
\]

\[
= \sigma^2[(\Lambda + I)^{-1}(\Lambda + dI)\Lambda^{-1}(\Lambda + dI)(\Lambda + I)^{-1} -
(\Lambda + I)^{-1}(I + mI)\Lambda(I + mI)(\Lambda + I)^{-1}]
\]

\[
= \sigma^2 \left[ \text{diag} \left\{ \frac{(\lambda_j + d)^2}{\lambda_j(\lambda_j + 1)^2} - \frac{(1 + m)^2\lambda_j}{(\lambda_j + 1)^2} \right\} \right]_{j=1}^p.
\]

Clearly, \(A_3A_4' - A_4A_3'\) is p. d. if and only if \((\lambda_j + d) - (1 + m)\lambda_j > 0\). Thus, \(A_3A_4' - A_4A_3'\) is p.d. for \(0 < d^*_j < d\). Similarly, \(A_4A_3' - A_3A_4'\) is p.d. if and only if \((\lambda_j + d) - (1 + m)\lambda_j < 0\). This inequality requires that \(A_4A_3' - A_3A_4'\) is p.d. for \(0 < d < d^*_j\). The proof is complete by lemma (2.2). □

The best estimator depends on the unknown parameters \(\beta\) and \(\sigma^2\) and also on the choice of the biasing parameter. This makes it difficult to apply these results in practice. For practical purposes, we have to replace these unknown parameters by some suitable estimates. Liu (1993) gave estimates of \(d\) by analogy with the
estimate of $k$ in ridge regression. Two of these estimates are:

\[
\hat{d}_{mn} = 1 - \hat{\sigma}^2 \left[ \sum_{i=1}^{p} \frac{1}{(1 + \lambda_i)} / \sum_{i=1}^{p} \frac{\hat{\alpha}_i^2}{(1 + \lambda_i)^2} \right], \tag{2.16}
\]

\[
\hat{d}_{CL} = 1 - \hat{\sigma}^2 \left[ \sum_{i=1}^{p} \frac{1}{(1 + \lambda_i)} / \sum_{i=1}^{p} \frac{\lambda_i \hat{\alpha}_i^2}{(1 + \lambda_i)^2} \right], \tag{2.17}
\]

where $\hat{\alpha}$ and $\hat{\sigma}^2$ are the OLS estimates of $\alpha$ and $\sigma^2$ respectively. Hoerl and Kennard (1970 a,b), Hoerl and Kennard and Baldwin (1975) and Lawless and Wang (1976) suggested the following operational ridge parameters, respectively:

\[
\hat{k}_{HK} = \frac{\hat{\sigma}^2}{\sum_{i=1}^{p} \hat{\alpha}_i^2}, \tag{2.18}
\]

\[
\hat{k}_{HKB} = \frac{p\hat{\sigma}^2}{\sum_{i=1}^{p} \hat{\alpha}_i^2}, \tag{2.19}
\]

\[
\hat{k}_{LW} = \frac{p\hat{\sigma}^2}{\sum_{i=1}^{p} \lambda_i \hat{\alpha}_i^2}. \tag{2.20}
\]

Other methods of estimating $k$ are available, but we restrict our attention to the estimators in (2.18)-(2.20). See Clark and Troskie (2006) and Muniz and Kibria (2009) for more information.

Now let us estimate $m$. The scalar mean squared error of $\hat{\alpha}'$ estimator is

\[
\text{mse}(\hat{\alpha}_m) = \sigma^2 \sum_{i=1}^{p} \frac{(1 + m)^2 \lambda_i}{(1 + \lambda_i)^2} + \sum_{i=1}^{p} \frac{(m \lambda_i - 1)^2 \hat{\alpha}_i^2}{(1 + \lambda_i)^2}. \tag{2.21}
\]

Since (2.21) involves a quadratic form, it can be minimized with respect to $m$ to
obtain

\[ m = \left[ \frac{\sum_{i=1}^{p} \lambda_i (\alpha_i^2 - \sigma^2)}{(1 + \lambda_i)^2} \right] \left/ \left[ \frac{\sum_{i=1}^{p} \lambda_i (\alpha_i^2 + \sigma^2)}{(1 + \lambda_i)^2} \right] \right. \]. \tag{2.22}

Since \( \alpha_i^2 \) and \( \sigma^2 \) are unknown, they are substituted by their unbiased estimators \( \hat{\alpha}_i^2 - \frac{\hat{\sigma}^2}{\lambda_i} \) and \( \hat{\sigma}^2 \), respectively, to obtain the estimate of \( m \).

\[ \hat{m} = \left[ \frac{\sum_{i=1}^{p} \lambda_i \hat{\alpha}_i^2}{\sum_{i=1}^{p} (1 + \lambda_i)^2} \right] - \hat{\sigma}^2 \left[ \frac{\sum_{i=1}^{p} \lambda_i^2 \sigma_i^2}{\sum_{i=1}^{p} (1 + \lambda_i)^2} \right]. \tag{2.23} \]

### 2.1.3 Simulation results

We carry out a simulation to study the behavior of the Al estimator under different degrees of collinearity and for different values of \( \sigma^2 \). For this simulation, we take \( p = 3 \) and \( n = 50, 100, \) and \( 150 \). The explanatory variables are generated by

\[ x_{ij} = (1 - \rho^2)^{1/2} z_{ij} + \rho z_{ij} \quad i = 1, \ldots, n; j = 1, \ldots, 3 \]

where \( z_{ij} \) are standard normal pseudorandom numbers and \( \rho \) is the correlation between any two explanatory variables. Seven values of correlation are considered, \( \rho = 0.5, 0.6, 0.7, 0.8, 0.9, 0.99, 0.999 \) and \( 0.9999 \).

For every simulation, let \( \beta_L \) denote the eigenvector corresponding to the largest eigenvalue of \( X'X \) and \( \beta_S \) denote the eigenvector corresponding to the smallest eigenvalue of \( X'X \). Observations on the dependent variable are generated by the following equation.

\[ y_{ig} = \beta_{1g} x_{i1} + \beta_{2g} x_{i2} + \beta_{3g} x_{i3} + \epsilon_i, \quad i = 1, 2, \ldots, n \quad g = L, S, \]

where the \( \epsilon_i \sim N(0, \sigma^2) \). We use three values of \( \sigma: 0.1, 1 \) and \( 10 \). The following seven estimators are compared:

**OLS:** Least squares estimator
RRHK: Ridge regression with estimate $k$ by formula (2.18)
RRHKB: Ridge regression with estimate $k$ by formula (2.19)
RRLW: Ridge regression with estimate $k$ by formula (2.20)
LIU1: Liu estimator with estimate $d$ by formula (2.16)
LIU2: Liu estimator with estimate $d$ by formula (2.17)
AL: Al estimator with estimate $m$ by formula (2.23)

For each choice of $\rho$ and $\sigma$, the experiment is replicated 1000 times. After 1000 samples are generated, the estimated mean squared error (EMSE) is computed for each of the above four estimators. EMSE is defined by

$$EMSE(\hat{\beta}) = \frac{1}{1000} \sum_{j=1}^{1000} \sum_{i=1}^{3} (\hat{\beta}_{ij} - \beta_i)^2,$$

where $\hat{\beta}_{ij}$ denotes the estimate of the $i^{th}$ parameter in $j^{th}$ replication and $\beta_i$ denotes the value of $i^{th}$ parameter. The results are presented in Tables A1-A4. We notice the following points in these tables.

1. The Al estimator is sensitive to change in $\sigma$. This is clear from its performance when $\sigma=0.1$ and when $\sigma=1$ or $\sigma=10$.

2. Preference of Al estimator over Liu estimator (LIU1 and LIU2) improves with increasing values of $\sigma$ and $\rho$, regardless of the sample size.

3. The Al estimator is better than ORR estimator when $k$ is estimated by (2.18) and (2.20). On the other hand, ORR estimator is sometimes better than Al estimator when $k$ is estimated by (2.19).

4. Liu estimator (LIU1 and LIU2) is not better than any other estimators when $n$, $\sigma$ and $\rho$ are large.
2.1.4 Numerical example

In this section, we consider a dataset on Portland cement, originally due to Wood et al. (1932), to illustrate the results of the previous section. This dataset has been widely analysed in Hald (1952), Daniel and Wood (1980), Nomura (1988), Kaçiranlar et al. (1999) and Liu (2003) etc. This dataset comes from an experimental investigation of the heat evolved during the setting and hardening of Portland cements of varied composition and the dependence of this heat on the percentages of four compounds in the clinkers from which the cement was produced. In this example, the dependent variable $Y$ is defined as heat evolved in calories per gram of cement. The independent variables are amounts of the following compounds: tricalcium aluminate ($X_1$), tricalcium silicate ($X_2$), tetracalcium aluminoferrite ($X_3$), and dicalcium silicate ($X_4$). The model includes the intercept term. The matrix $X'X$ has eigenvalues $\lambda_1 = 211.367$, $\lambda_2 = 77.236$, $\lambda_3 = 28.459$, $\lambda_4 = 10.267$ and $\lambda_5 = 0.0349$. The condition number of $X$ is $CN = 6056.37$ and so $X$ may be considered as being quite "ill-Conditioned". OLS estimator is:

$$\hat{\beta} = (X'X)^{-1}X'Y = \begin{bmatrix} \hat{\beta}_0 & \hat{\beta}_1 & \hat{\beta}_2 & \hat{\beta}_3 & \hat{\beta}_4 \end{bmatrix}'$$

$$= [62.4052 \ 1.5511 \ 0.5102 \ 0.1019 \ -0.1441]' .$$

Most authors recommend standardizing the data so that $X'X$ is in the form of a correlation matrix. An advantage of standardization of the data is that the regression coefficients can then be expressed in comparable numerical units. The standardization is accomplished by transforming the linear model $Y = X\beta + \epsilon$ to $Y_s = X_s\hat{\beta}_s + \epsilon$. Another advantage of standardizing the data is that it can show highly correlated variables. The corresponding OLS estimator is:

$$\hat{\beta}_s = (X'_sX_s)^{-1}X'_sY_s = [0.6065 \ 0.5277 \ 0.0434 \ -0.1603]' .$$
Since there are thirteen observations and four parameters in the standardized data, we obtain the following

\[ \sigma^2 \frac{\beta - X_s\beta_s}{n - p} = 0.00196 \]

The eigenvalues of \( X_s^tX_s \) are 2.2357, 1.5760, 0.1866, 0.00162. The 4 x 4 matrix \( V \) is the matrix of normalized eigenvectors, \( \Lambda \) is a 4 x 4 diagonal matrix of eigenvalues of \( X_s^tX_s \) such that \( X_s^tX_s = VAV^t \). Then \( Z = X_sV \) and \( \alpha = V^t\beta_s \) so that \( Y_s = X_s\beta_s + \epsilon = X_sVV^t\beta_s + \epsilon = Z\alpha + \epsilon \). In orthogonal coordinates, the OLS estimator is:

\[ \hat{\alpha} = \Lambda^{-1}Z'Y_s = [0.65696, -0.00831, 0.3028, 0.388]^t. \]

From Eq.(2.18), we have \( k_{HK} = 0.0029 \). Using Eq.(2.19) we get \( k_{HK,B} = 0.01163 \) and using Eq.(2.20) we get \( k_{LW} = 0.00798 \). Also, (2.16) and (2.17) give the estimators of \( d \) for use in Liu estimator. From Eq.(2.16) we have \( \hat{d}_{nm} = 0.9806 \) and using Eq.(2.17) we get \( \hat{d}_{CL} = 0.9524 \). Using Eq.(2.23), the estimator of \( m \) can be obtained where \( \hat{m} = 0.4461 \).

Our objective is to compare the trace of the estimated mean squared error matrix of \( \hat{\alpha}_m \) with the traces of the estimated mean squared error matrices of OLS estimator, ORR estimator, and Liu estimator. The trace of the mean squared error matrix of \( \hat{\alpha}_m \) is given by Eq.(2.21). The trace of the mean square error matrix of \( \hat{\alpha} \) is given by

\[ \text{mse}(\hat{\alpha}) = \sigma^2 \sum_{i=1}^{P} \frac{1}{\lambda_i}. \]

The trace of the mean squared error matrix of \( \hat{\alpha}_k \) is given by

\[ \text{mse}(\hat{\alpha}_k) = \sum_{i=1}^{P} \frac{\lambda_i \sigma^2 + k^2 \alpha_i^2}{(\lambda_i + k)^2}. \]
Also, the trace of the mean squared error matrix of $\hat{a}_d$ is given by

$$\text{mse}(\hat{a}_d) = \sum_{i=1}^{p} \frac{\lambda_i + d + 1}{\lambda_i + 1} \sigma^2 + (1 - d)^2 \lambda_i \sigma^2 \frac{\lambda_i \sigma^2}{(\lambda_i + 1)^2}.$$ 

By substituting $\hat{a}$ and $\hat{\sigma}^2$ for $\alpha$ and $\sigma^2$, we get the estimates for Eqs.(2.6)-(2.8) and their estimated mean squared errors. Various values of $m$, $k$ and $d$ and their mse and estimators are given in Tables 2.1-2.3.
Table 2.3: Values of $\hat{\alpha}_d$ and $\text{mse}(\hat{\alpha}_d)$ for various values of $d$

<table>
<thead>
<tr>
<th>$d$</th>
<th>$\hat{\alpha}_d$</th>
<th>$\text{mse}(\hat{\alpha}_d)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.453924 -0.005083</td>
<td>0.047614 0.000629 0.257569</td>
</tr>
<tr>
<td>0.00014</td>
<td>0.453952 -0.005084</td>
<td>0.047649 0.000683 0.257498</td>
</tr>
<tr>
<td>0.21</td>
<td>0.496561 -0.005761</td>
<td>0.101197 0.081985 0.216189</td>
</tr>
<tr>
<td>0.9</td>
<td>0.636655 -0.007986</td>
<td>0.277255 0.349295 0.991436</td>
</tr>
<tr>
<td>0.95</td>
<td>0.646806 -0.008147</td>
<td>0.290012 0.368866 1.10190</td>
</tr>
<tr>
<td>0.98</td>
<td>0.652897 -0.008244</td>
<td>0.297677 0.380288 1.17171</td>
</tr>
<tr>
<td>1</td>
<td>0.656958 -0.008309</td>
<td>0.302770 0.388036 1.21971</td>
</tr>
</tbody>
</table>

Explanation of the results:

1. Using $m_i^* = 1/\lambda_i$, $i=1,2,3,4$ in Theorem 2.1, we may find the following values of $m_i^*$: 0.447; 5.359; 615.860. Comparing $\text{mse}(\hat{\alpha})$ with $\text{mse}(\hat{\alpha}_m(0.446))$ for $m = 0.446 < \min(m_i^*) = 0.447$ we see that $\text{mse}(\hat{\alpha}_m(0.446)) = 0.206985$ is smaller than $\text{mse}(\hat{\alpha}) = 1.21971$ as stated in Theorem 2.1.

2. Let us make $m = 0.44$, which satisfies $m \lambda_i < 1$ fixed. Using $k_i^* = (1 - m \lambda_i)/(1 + m)$ in Theorem 2.2, we may find the following values of $k_i^*$: 0.011313; 0.2128; 0.6374; 0.6939. Comparing $\text{mse}(\hat{\alpha}_m(0.44))$ with $\text{mse}(\hat{\alpha}_k(0.7))$ for $k = 0.7 > \max(k_i^*) = 0.6939$ we see that $\text{mse}(\hat{\alpha}_m(0.44)) = 0.2071$ is smaller than $\text{mse}(\hat{\alpha}_k(0.7)) = 0.2331$ as stated in Theorem 2.2(a). Also for $\text{mse}(\hat{\alpha}_m(0.44))$ with $\text{mse}(\hat{\alpha}_k(0.11))$ for $k = 0.11 < \min(k_i^*) = 0.113$ we have the following inequality from Theorem 2.2(b): $\text{mse}(\hat{\alpha}_m(0.44)) = 0.2071 > \text{mse}(\hat{\alpha}_k(0.11)) = 0.146$.

The plot in Figure 2.1 shows the mean squared errors of OLS, Liu, ORR and $Al$ estimators, when the shrinkage parameters $d$, $k$ and $m$ are in $[0,1]$, and it indicates that $Al$ estimator at some points has smaller mean squared error than other estimators.
3. Let us fix \( k = 0.002 \). Using \( m_i^* = (1 - k) / (\lambda_i + k) \), \( i = 1, 2, 3, 4 \) in Theorem 2.3 we find the following values of \( m_i^* \): 0.446; 0.632; 5.291; 275.406. In this case \( \text{mse}(\hat{\alpha}_m(275.5)) = 15843.3 > \text{mse}(\hat{\alpha}_k(0.002)) = 0.3006 \) for \( m = 275.5 > \max(m_i^*) = 275.406 \) as stated in Theorem 2.3(a). Also \( \text{mse}(\hat{\alpha}_m(0.445)) = 0.207 < \text{mse}(\hat{\alpha}_k(0.002)) = 0.3006 \) for \( m = 0.446 < \min(m_i^*) = 0.446 \) as stated in Theorem 2.3(b).

The plot in Figure 2.2 shows the mse of \( Al \) estimator compared with Liu and ORR estimators when \( m \) and \( d \) are in \([0,0.5]\) and \( m \) and \( k \) are in \([0,0.01]\). It indicates that as \( m \) and \( d \) increase, mse of \( Al \) estimator slightly increases compared to mse of Liu estimator, while when \( m \) and \( k \) is close to 0, mse of \( Al \) estimator is smaller than that of ORR.

4. Let \( d = 0.9 \) be fixed. We find \( m_i^* = d / \lambda_i : 0.403; 0.571; 4.823; 554.274 \) using Theorem 2.4. Comparing \( \text{mse}(\hat{\alpha}_d(0.9)) \) with \( \text{mse}(\hat{\alpha}_m(554.3)) \) for \( m = 554.3 > \max(m_i^*) = 554.27 \), we see that \( \text{mse}(\hat{\alpha}_d(0.9)) = 0.9914 \) is smaller than
Figure 2.2: The estimated values of mse for ORR, LE and Al estimators for different values of the shrinkage parameters.

$mse(\hat{\alpha}_m(554.3)) = 64250.3$ as stated in Theorem 2.4(a). As for $mse(\hat{\alpha}_d(0.9))$ with $mse(\hat{\alpha}_m(0.4))$ for $m=0.4 < min(m_l) = 0.403$, we have the following inequality from Theorem 2.4(b):

$mse(\hat{\alpha}_d(0.9)) = 0.9914 > mse(\hat{\alpha}_m(0.4)) = 0.208.$

5. Let $m=0.09$. It satisfies $m \lambda_i < 1$. Using $d^*_i = m \lambda_i$ in Theorem 2.5, we find the following values of $d^*_i$: 0.2012; 0.1418; 0.0167; 0.00014. Comparing $mse(\hat{\alpha}_m(0.09))$ with $mse(\hat{\alpha}_d(0.21))$ for $d=0.21 > max(d^*_i) = 0.2012$, we see that $mse(\hat{\alpha}_m(0.09)) = 0.241 > mse(\hat{\alpha}_d(0.21)) = 0.216$ as stated in Theorem 2.5(a). As for $mse(\hat{\alpha}_m(0.09))$ with $mse(\hat{\alpha}_d(0.00013))$ for $d=0.0013 < min(d^*_i) = 0.00014$ we have the following inequality from Theorem 2.5(b) $mse(\hat{\alpha}_m(0.091)) = 0.241 < mse(\hat{\alpha}_d(0.00013)) = 0.2575.$
2.2 New types of shrinkage estimator

If $\rho$ is considered as a known constant, $\hat{\beta}(\rho)$ is a homogeneous linear estimator satisfying

\[
\begin{align*}
E(\hat{\beta}(\rho)) &= (1 + \rho)^{-1}\beta, \\
\text{bias}(\hat{\beta}(\rho)) &= -\rho(1 + \rho)^{-1}\beta, \\
\text{Var}(\hat{\beta}(\rho)) &= \sigma^2(1 + \rho)^{-2}(X'X)^{-1}.
\end{align*}
\]

We investigate the condition for MMSE of SE to be smaller than MMSE of OLS.

**Theorem 2.6** Assume that $\rho$ is known and $\rho > 0$. Then the difference $\Delta = \text{MMSE}(\hat{\beta}) - \text{MMSE}(\hat{\beta}(\rho))$ is n.n.d. if

\[
0 < \rho < \frac{2\sigma^2}{\beta'X'X\beta}, \quad \beta \neq 0.
\]

**Proof:**

The MMSE of SE is given by

\[
\text{MMSE}(\hat{\beta}(\rho)) = \sigma^2(1 + \rho)^{-2}(X'X)^{-1} + \rho^2(1 + \rho)^{-2}\beta'\beta'.
\]

(2.24)

Hence,

\[
\text{MMSE}(\hat{\beta}) - \text{MMSE}(\hat{\beta}(\rho)) = \sigma^2 A(X'X)^{-1} - \beta'\beta',
\]

where $A = \frac{2\rho}{\rho}$. Since $(X'X)^{-1}$ is p. d. and $\sigma^2 A$ is a positive scaler, we apply lemma (2.1). Then, $\Delta$ is n. n. d. if and only if

\[
\beta'X'X\beta < \frac{2 + \rho}{\rho} \sigma^2.
\]
For any $\beta \neq 0$, positive definiteness of $X'X$ implies the identity $\beta'X'X\beta \neq 0$. If the parameters $\beta \neq 0$ and $\sigma^2$ satisfy $\beta'X'X\beta \leq \sigma^2$, then $\beta'X'X\beta \leq \frac{\rho^2 + 2\sigma^2}{\rho}$ for any $\rho > 0$ and hence for any $0 < \rho < \frac{2\sigma^2}{\beta'X'X\beta}$. If $\beta \neq 0$ and $\sigma^2$ satisfy $\beta'X'X\beta > \sigma^2$, then $\beta'X'X\beta \leq \frac{\rho^2 + 2\sigma^2}{\rho}$ can be written as

$$\rho \leq \frac{2\sigma^2}{\beta'X'X\beta - \sigma^2}.$$ 

But, this inequality is satisfied for any $0 < \rho < 2\sigma^2/\beta'X'X\beta$ whenever $\beta'X'X\beta \geq \sigma^2$. 

We have seen that, in presence of MC, the squared length of $\beta$ is often overestimated by the OLS estimator $\hat{\beta}$. The SE estimator reduces the squared length of $\hat{\beta}$. 

$$\|\hat{\beta}(\rho)\|^2 < \|\hat{\beta}\|^2.$$ 

Unfortunately, the SE has a negative aspect. The curve of SE over the parameter space goes from $\hat{\beta}$ to 0 as $\rho$ goes from 0 to $\infty$. Therefore, the curve of SE path through the parameter space from $\hat{\beta}$ to 0. To solve this problem, we must find an estimator whose length is closer to $\beta$ than $\hat{\beta}$. 

Since ORR and LE have a smaller length than OLS estimator, it might be preferable to replace the OLS estimator by ORR or the LE estimators. The new types of the SE are given as follows:

$$\hat{\beta}_k(\rho) = \left( \frac{1}{1 + \rho} \right) \hat{\beta}_k. \quad (2.25)$$ 

$$\hat{\beta}_d(\rho) = \left( \frac{1}{1 + \rho} \right) \hat{\beta}_d. \quad (2.26)$$ 

which shrinks each component of ORR and LE estimators by $(1 + \rho)^{-1}$. We call these proposed estimators, the shrinkage ridge regression (SRR) and the shrinkage Liu (SL) estimators respectively. Also we can see that $(1 + \rho)^{-1}$ will be between
1 and 0 where

\[ \lim_{\rho \to 0} (1 + \rho)^{-1} = 1. \]

and

\[ \lim_{\rho \to \infty} (1 + \rho)^{-1} = 0. \]

2.2.1 Statistical properties of the proposed shrinkage estimators

By using model (2.4), the SE, the SRR and the SL estimators are given as follows:

\[ \hat{\alpha}(\rho) = \frac{1}{1 + \rho} \hat{\alpha} = A_2 Y, \]

\[ \hat{\alpha}_k(\rho) = \frac{1}{1 + \rho} \hat{\alpha}_k = A_6 Y, \]

\[ \hat{\alpha}_d(\rho) = \frac{1}{1 + \rho} \hat{\alpha}_d = A_7 Y, \tag{2.27} \]

respectively.

Now, we find some useful properties of proposed estimators.

\[ \mathbb{E}(\hat{\alpha}_k(\rho)) = (1 + \rho)^{-1}(\Lambda + kI)^{-1}\Lambda \alpha. \]

\[ \text{bias}(\hat{\alpha}_k(\rho)) = ((1 + \rho)^{-1}(\Lambda + kI)^{-1}\Lambda - I)\alpha \]

\[ \text{Var}(\hat{\alpha}_k(\rho)) = \sigma^2(1 + \rho)^{-2}(\Lambda + kI)^{-1}\Lambda(\Lambda + kI)^{-1}. \tag{2.28} \]

\[ \text{MMSE}(\alpha k(\rho)) = \sigma^2(1 + \rho)^{-2}(\Lambda + kI)^{-1}\Lambda(\Lambda + kI)^{-1} + \\
(1 + \rho)^{-1}(\Lambda + kI)^{-1}\Lambda - I)\alpha\alpha' + \\
((1 + \rho)^{-1}(\Lambda + kI)^{-1}\Lambda - I). \tag{2.29} \]
\[
E(\hat{\alpha}_d(\rho)) = (1 + \rho)^{-1}(\Lambda + I)^{-1}(\Lambda + dI)\alpha \\
\text{bias}(\hat{\alpha}_d(\rho)) = ((1 + \rho)^{-1}(\Lambda + I)^{-1}(\Lambda + dI) - I)\alpha.
\]

\[
\text{Var}(\hat{\alpha}_d(\rho)) = \sigma^2(1 + \rho)^{-2}(\Lambda + I)^{-1}(\Lambda + dI)(\Lambda + I)^{-1}. \\
\text{MMSE}(\hat{\alpha}_d(\rho)) = \sigma^2(1 + \rho)^{-2}(\Lambda + I)^{-1}(\Lambda + dI)\Lambda^{-1}(\Lambda + dI)(\Lambda + I)^{-1} \\
+ ((1 + \rho)^{-1}(\Lambda + I)^{-1}(\Lambda + dI) - I)\alpha \\
\alpha'((1 + \rho)^{-1}(\Lambda + I)^{-1}(\Lambda + dI) - I) \\
\text{(2.30)}
\]

Now, we state the following theorems:

**Theorem 2.7** Assume that \( \rho \) is known and \( \rho > 0 \) and \( 0 < k < 1 \). Then the variance of \( \hat{\alpha}_k(\rho) \) is smaller than the variance of \( \hat{\alpha}(\rho) \).

**Proof:**

\[
\text{Var}(\hat{\alpha}(\rho)) - \text{Var}(\hat{\alpha}_k(\rho)) = \sigma^2(1 + \rho)^{-2} \left[ \Lambda^{-1} - (\Lambda + kI)^{-1}\Lambda(\Lambda + kI)^{-1} \right] \\
= \sigma^2(1 + \rho)^{-2} \left[ \text{diag} \left\{ \frac{1}{\lambda_i} - \frac{\lambda_i}{(\lambda_i + k)^2} \right\} \right]_{i=1}^p.
\]

Since \( 0 < k < 1 \), the proof is complete.

**Theorem 2.8** Assume that \( \rho \) is known and \( \rho > 0 \) and \( 0 < k < 1 \). Then the variance of \( \hat{\alpha}_k(\rho) \) is smaller than the variance of \( \hat{\alpha}_k \).

**Proof:**

\[
\text{Var}(\hat{\alpha}_k) - \text{Var}(\hat{\alpha}_k(\rho)) = \sigma^2 \left[ (\Lambda + I)^{-2}\Lambda - (1 + \rho)^{-2}(\Lambda + I)^{-2}\Lambda \right] \\
= \sigma^2 \left[ \text{diag} \left\{ \frac{\lambda_i}{(\lambda_i + k)^2} - (1 + \rho)^{-2}\frac{\lambda_i}{(\lambda_i + k)^2} \right\} \right]_{i=1}^p.
\]

46
Since \((1 + \rho)^{-1} < 1\), the proof is complete.

**Theorem 2.9** Assume that \(\rho\) is known and \(\rho > 0\) and \(0 < d < 1\). Then the variance of \(\hat{\alpha}_d(\rho)\) is smaller than the variance of \(\hat{\alpha}(\rho)\).

**Proof**

\[
\text{Var}(\hat{\alpha}(\rho) - \text{Var}(\hat{\alpha}_d(\rho)) = \sigma^2 (1 + \rho)^{-2} [\Lambda^{-1} - (\Lambda + I)^{-2}(\Lambda + dI)^2 \Lambda^{-1} ]
\]

\[
= \sigma^2 (1 + \rho)^{-2} \left[ \text{diag} \left\{ \frac{1}{\lambda_i} - \frac{(\lambda_i + d)^2}{\lambda_i(\lambda_i + 1)^2} \right\}_{i=1}^p \right]
\]

Since \(0 < d < 1\), the proof is complete.

**Theorem 2.10** Assume that \(\rho\) is known and \(\rho > 0\) and \(0 < d < 1\). Then the variance of \(\hat{\alpha}_d(\rho)\) is smaller than the variance of \(\hat{\alpha}_d\).

**Proof**

\[
\text{Var}(\hat{\alpha}_d) - \text{Var}(\hat{\alpha}_d(\rho)) = \sigma^2 (\Lambda + I)^{-2}(\Lambda + dI)^2 \Lambda^{-1} - (1 + \rho)^{-2}(\Lambda + I)^{-2}
\]

\[
= \sigma^2 \left[ \text{diag} \left\{ \frac{(\lambda_i + d)^2}{\lambda_i(\lambda_i + 1)^2} - (1 + \rho)^{-2} \frac{(\lambda_i + d)^2}{\lambda_i(\lambda_i + 1)^2} \right\}_{i=1}^p \right]
\]

Since \((1 + \rho)^{-1} < 1\), the proof is completed.

### 2.2.2 Comparisons among the shrinkage estimators

By using the MMSE of the shrinkage estimator, we start to make a comparison among them to see the performance of the new shrinkage estimators with others. Thus, we have the following theorems:
Theorem 2.11 If \( \beta(\hat{\alpha}_k(\rho))' (A_5 A_3' - A_6 A_6')^{-1} \beta(\hat{\alpha}_k(\rho)) < (1+\rho)^{-2}\sigma^2 \), then \( \Delta = \text{MMSE}(\hat{\alpha}(\rho)) - \text{MMSE}(\hat{\alpha}_k(\rho)) \) is p. d.

Proof:

By Theorem 2.8 and Lemma 2.1, the proof is complete. \( \square \)

Theorem 2.12 If \( \beta(\hat{\alpha}_d(\rho))' (A_2 A_2' - A_7 A_7')^{-1} \beta(\hat{\alpha}_d(\rho)) < (1+\rho)^{-2}\sigma^2 \), then \( \Delta = \text{MMSE}(\hat{\alpha}(\rho)) - \text{MMSE}(\hat{\alpha}_d(\rho)) \) is p. d.

Proof

By Theorem 2.9 and Lemma 2.2, the proof is complete. \( \square \)

Theorem 2.13 If \( \beta(\hat{\alpha}_d(\rho))' (A_3 A_3' - A_7 A_7')^{-1} \beta(\hat{\alpha}_d(\rho)) < (1+\rho)^{-2}\sigma^2 \), then \( \Delta = \text{MMSE}(\hat{\alpha}(\rho)) - \text{MMSE}(\hat{\alpha}_d(\rho)) \) is p. d.

Proof

By Theorem 2.10 and Lemma 2.1, the proof is complete. \( \square \)

Theorem 2.14 If \( \beta(\hat{\alpha}_d(\rho))' (A_3 A_3' - A_7 A_7')^{-1} \beta(\hat{\alpha}_d(\rho)) < (1+\rho)^{-2}\sigma^2 \), then \( \Delta = \text{MMSE}(\hat{\alpha}(\rho)) - \text{MMSE}(\hat{\alpha}_d(\rho)) \) is p. d.

Proof

The proof follows from Theorem 2.10 and Lemma 2.2. \( \square \)

2.2.3 The generalized shrinkage estimator (GSE)

Let us consider \( \rho_i > 0 \; \forall i = 1, ..., p \). Then, the generalized shrinkage estimator (GSE) is defined as follows:

\[
\hat{\beta}_i(\rho) = \frac{1}{1 + \rho_i} \hat{\beta}_i. \tag{2.32}
\]
This estimator solves the shrinkage estimator problem as we explained in the previous section. The scalar mean squared error for the model (1.1) of \( \hat{\alpha}_i(\rho_i) \) is given by

\[
mse(\hat{\alpha}_i(\rho_i)) = \sigma^2 \left( \frac{1}{1 + \rho_i} \right)^2 \frac{1}{\lambda_i} + \left( \frac{\rho_i}{1 + \rho_i} \right)^2 \alpha_i^2.
\] (2.33)

To find the optimal value of \( \rho_i \), we minimize the mse of GSE with respect to the shrinkage estimator \( \rho_i \). After some simplification, we obtain

\[
\rho_i \text{ opt} = \frac{\sigma^2}{\lambda_i \alpha_i^2}.
\] (2.34)

Since \( \rho_i \text{ opt} \) depends on the unknown parameters \( \sigma^2 \) and \( \alpha_i \), we replace them by their unbiased estimators. So,

\[
\hat{\rho}_i \text{ opt} = \frac{\hat{\sigma}^2}{\hat{\lambda}_i \hat{\alpha}_i^2}.
\] (2.35)

### 2.2.4 The choice of the shrinkage parameter

For every \( (\beta \neq 0, \sigma^2) \) there is an optimal value \( \rho_{opt} \) such that

\[
mse(\hat{\beta}(\rho_{opt})) \leq mse(\hat{\beta}(\rho)) \text{ for any } \rho \geq 0.
\]

mse of SE is given as

\[
mse(\hat{\beta}(\rho)) = tr[MMSE(\hat{\beta}(\rho))] = \sigma^2 tr[(X'X)^{-1}] + \frac{\rho^2}{(1 + \rho)^2} \beta' \beta.
\]

If we consider the mse of SE as a function of \( \rho \) and determine its derivative with respect to \( \rho \) and put it equal to zero and solve it for \( \rho \), we obtain

\[
\rho_{i \text{ opt}} = \frac{\sigma^2 tr[(X'X)^{-1}]}{\beta' \beta}.
\] (2.36)
Now we can state the following theorem

**Theorem 2.15** The inequality $\text{mse}(\hat{\beta}(\rho_{opt})) \leq \text{mse}(\tilde{\beta}(\rho))$ for any $\rho \geq 0$ holds true, where

$$\rho_{opt} = \frac{\sigma^2 \text{tr}[(X'X)^{-1}]}{\beta'\beta}.$$

In the literature, there are other shrinkage estimators for $\beta$. We will present some of them and propose some new methods, by introducing the corresponding estimator for the shrinkage parameter $\rho$.

**The minimum mean squared error estimator (MSE)**

Farebrother (1975) suggested the minimum mean squared error estimator (MSE). Let $b = \beta^* + AY$ be the linear estimator where

$$\beta^* = \beta\beta'X'X\beta + \sigma^2 I^{-1}Y.$$

(2.37)

and $A$ is an arbitrary matrix. Then MSE of $b$ is defined as:

$$\text{MSE}(b) = [1 - \beta'X'(X\beta\beta'X' + \sigma^2 I)^{-1}X\beta] \beta'\beta'X' + A(X\beta\beta'X' + \sigma^2 I)A'.

(2.38)

**Lemma 2.3**

$$(B - cc')^{-1} = \left( B^{-1} + \frac{B^{-1}cc'B^{-1}}{1 - c'B^{-1}c} \right).$$

Using lemma 2.3, we have:

$$(X\beta\beta'X' + \sigma^2 I)^{-1} = \frac{1}{\sigma^2} (I - X\beta(\sigma^2 + \beta'X'X\beta)^{-1}\beta'X').

(2.39)$$
Substituting (2.39) with (2.37) we obtain

\[ \hat{\beta}^* = \frac{\beta'X'Y}{\sigma^2 + \beta'X'X\hat{\beta}}. \quad (2.40) \]

Since \( \beta^* \) in (2.40) depends on the unknown parameters \( \beta \) and \( \sigma^2 \), therefore it is not operational. Farebrother (1975) proposed to replace \( \beta \) and \( \sigma^2 \) by \( \hat{\beta} \) and \( \hat{\sigma}^2 \) respectively.

Hence (2.40) changes to

\[ \hat{\beta}_M = \frac{\hat{\beta}'X'Y}{\hat{\sigma}^2 + \hat{\beta}'X'X\hat{\beta}}. \quad (2.41) \]

**Remark:** The \( \hat{\beta}_M \) can explicitly be written as the SE if \( \hat{\rho} = \frac{\hat{\sigma}^2}{\hat{\beta}'X'X\hat{\beta}} \).

**Adjusted minimum mean squared error estimator**

Ohtani (1996) considered the adjusted minimum mean squared error (AMSE) estimator which is defined as:

\[ \hat{\beta}_{AM} = \hat{\beta}(\hat{\rho}_{AM}) = \frac{\hat{\beta}'X'Y/p}{\hat{\sigma}^2 + \hat{\beta}'X'X\hat{\beta}/p}. \quad (2.42) \]

where \( \hat{\rho}_{AM} = \frac{\hat{\rho}\sigma^2}{\hat{\beta}'X'X\hat{\beta}} \).

**Remark:** The estimator of \( \rho \) in (2.42) is the same as the lawless and Wang (1976) estimator of the shrinkage parameter of ridge regression.

**The Stein estimator**

Stein (1956) and James Stein (1961) introduced another class of biased estimators. While the original thrust of ridge regression was to alleviate effects of MC, this type was aimed solely at reducing the MMSE (Groß, 2003).

The predictive mean squared error of SE is given as follows:

\[ \text{PMSE}(\hat{\beta}(\rho)) = \text{tr}[X'X\text{MMSE}(\hat{\beta}(\rho))] = \frac{\rho\sigma^2}{(1 + \rho)^2} + \frac{\rho^2}{(1 + \rho)^2} \beta'X'X\beta. \]
By minimizing PMSE of SE with respect to $\rho$, we obtain

$$\rho_{opt} = \frac{p\sigma^2}{\beta'X'X\beta},$$

We rewrite $\rho_{opt}$ as

$$\rho_{opt} = \frac{p\sigma^2}{\beta'X'X\beta + p\sigma^2 - p\sigma^2},$$

By replacing $\beta'X'X\beta + p\sigma^2$ by its unbiased estimator $\hat{\beta}X'X\hat{\beta}$ and $p\sigma^2$ by $c\hat{e}'\hat{e}$, where $\hat{e} = Y - X\hat{\beta}$ and $c > 0$ so that $\hat{\beta}X'X\hat{\beta} - c\hat{e}'\hat{e}$ is positive. Then

$$\hat{\rho}_{S} = \frac{c\hat{e}'\hat{e}}{\beta'X'X\beta - c\hat{e}'\hat{e}}.$$  \hspace{1cm} (2.43)

Therefore,

$$\hat{\beta}_S = \frac{1}{1 + \hat{\rho}_S \hat{\beta}} = \left[1 - \frac{c\hat{e}'\hat{e}}{\beta'X'X\beta}\right] \hat{\beta}. \hspace{1cm} (2.44)$$

If $p \geq 3$ and $c = \frac{p - 2}{n - p + 2}$, then $\hat{\beta}_S$ will be the Stein estimator.

**The arithmetic mean of the shrinkage parameter of GSE**

Here, we consider a new estimator of $\rho$ by using the arithmetic mean of $\hat{\rho}_{opt}$ in (2.36) which produces the following estimator

$$\hat{\rho}_A = \frac{1}{p} \sum_{i=1}^{p} \frac{\hat{\sigma}^2}{\lambda_i\hat{\beta}_i^2}. \hspace{1cm} (2.45)$$

**The geometric mean of the shrinkage parameter of GSE**

If we take the geometric mean of $\hat{\rho}_{opt}$ in (2.36), we may obtain a new estimator of $\rho$

$$\hat{\rho}_G = \frac{\hat{\sigma}^2}{(\Pi_{i=1}^{p} \lambda_i\hat{\beta}_i^2)^{1/p}}. \hspace{1cm} (2.46)$$
The median of the shrinkage parameter of GSE

By using the median of \( \hat{\rho}_{opt} \) in (2.36), we propose the following estimator of \( \rho \) for \( p \geq 3 \)

\[
\hat{\rho}_{ME} = \text{Median}\{\frac{\hat{\sigma}^2}{\hat{\lambda}_i \hat{\sigma}_i^2}\}.
\]  

(2.47)

Remark: Kibria (2003) introduced the estimators in (2.45-2.47) as new estimators of the shrinkage parameter of ridge regression.

2.2.5 Monte Carlo simulations

In order to assess the suitability of the new types of SE and the new estimate of the shrinkage parameter we undertake a simulation study. Five parameters of the CMLR model with 50, 100 and 150 observations is generated. The explanatory variables are generated by

\[
x_{ij} = \begin{cases} 
(1 - \gamma^2_{1,1})^{1/2} z_{ij} + \gamma_1 z_{i6} & \text{for } j=1,2,3, \ i=1,2,...,n \\
(1 - \gamma^2_{2,2})^{1/2} z_{ij} + \gamma_2 z_{i6} & \text{for } j=4,5, \ i=1,2,...,n,
\end{cases}
\]

where the six \( z \) variables are independent \( N(0,1) \) random variables and \( \gamma_1 \) and \( \gamma_2 \) are the correlations between any two explanatory variables. Five different combinations of \( (\gamma_1, \gamma_2) \) are considered: \( (0.999,0.999), (0.999,0.9), (0.99,0.1), (0.9,0.9) \) and \( (0.8,0.4) \). These combinations with the condition number of the correlation matrix of the \( X \) variables are presented in Table (5) in Appendix A. Tables 6-8 in the Appendix A display the eigenvalues and the vector \( \pi_{ii} = \pi_{11},...,5 \) associated with maximum eigenvalue of the correlation matrix of the design matrix for the different collinearity levels and sample sizes.

The \( Y \) vector for each simulation is generated by the following equation

\[
Y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_5 x_{i5} + \epsilon_i,
\]
where $\epsilon_i \sim N(0, \sigma^2)$. We consider six different $\sigma^2$ values: 0.01, 0.1, 0.25, 1, 5 and 25. In this simulation, the following nineteen estimators are considered:

1. **OLS**: Least squares estimator.

2. **SHOLSF**: Shrinkage least squares estimator (with Farebrother shrinkage parameter).

3. **SHOLSOH**: Shrinkage least squares estimator (with Ohtani shrinkage parameter).

4. **SHOLSST**: Shrinkage least squares estimator (with Stein shrinkage parameter).

5. **SHOLSAM**: Shrinkage least squares estimator (with the arithmetic mean of the shrinkage parameter).

6. **SHOLSGM**: Shrinkage least squares estimator (with the geometric mean of the shrinkage parameter).

7. **SHOLSMED**: Shrinkage least squares estimator (with the median mean of the shrinkage parameter).

8. **SHRIDGEOF**: Shrinkage Ridge regression estimator (with Farebrother shrinkage parameter).
9. SHRIDGEOH: Shrinkage Ridge regression estimator (with Ohtani shrinkage parameter).

10. SHRIDGEST: Shrinkage Ridge regression estimator (with Stein shrinkage parameter).

11. SHRIDGEAM: Shrinkage Ridge regression estimator (with the arithmetic mean of the shrinkage parameter).

12. SHRIDGEGM: Shrinkage Ridge regression estimator (with the geometric mean of the shrinkage parameter).

13. SHRIDGEMED: Shrinkage Ridge regression estimator (with the median mean of the shrinkage parameter).

14. SHLIUF: Shrinkage Liu estimator (with Farebrother shrinkage parameter).

15. SHLIUOH: Shrinkage Liu estimator (with Ohtani shrinkage parameter).

16. SHLIUST: Shrinkage Liu estimator (with Stein shrinkage parameter).

17. SHLIUAM: Shrinkage Liu estimator (with the arithmetic mean of the shrinkage parameter).
18. SHLIUGM: Shrinkage Liu estimator (with the geometric mean of the shrinkage parameter).

19. SHLIUMED: Shrinkage Liu estimator (with the median mean of the shrinkage parameter).

For each choice of $(\gamma_1, \gamma_2)$ and $\sigma^2$ for each sample size, the process is repeated 1000 times and the estimated mean squared error (EMSE) is calculated for each estimator of this simulation, where the EMSE is defined by

$$EMSE(\hat{\beta}) = \frac{1}{1000} \sum_{j=1}^{1000} \sum_{i=1}^{5} (\hat{\beta}_{ij} - \beta_i)^2,$$

where $\hat{\beta}_{ij}$ denotes the estimate of the $i$th parameter in $j$th replication and $\beta_i$ denotes the $i$th of the true parameter values.

The results of the simulation study are presented in Table 5 and Tables 9-11 in Appendix A. Based on Table 5, we note the following point:

The condition number has a positive relationship with the degree of correlation between the explanatory variables and at the same time it has a negative relationship with the number observations.

Based on Tables 9-11, we observe the following points:

- The EMSE of the estimators increase most the times as the sample size decreases and it is visible for $\sigma^2 > 1$.

- The performance of proposed estimators dominate OLS estimator in the sense of having smaller EMSE. This can be seen easily for moderate to high correlation and when $\sigma^2$ increases. Also, when the correlation and $\sigma^2$ increase and the sample size decreases, the performance of the proposed estimators will improve.
• The proposed shrinkage estimators have EMSE smaller than the shrinkage least squares estimator for all types of shrinkage parameter, except for Stein parameter in some cases.

• The proposed shrinkage parameters improve the precision of the shrinkage estimators compared with Farebrother and Ohtani and Stein shrinkage parameter estimators, where they obtain EMSE smaller than others except $\hat{\rho}_{GM}$. Also, the performance of these shrinkage parameters are affected by the degree of correlation, $\sigma^2$ and the sample size.

• Through the simulation results, the performance of the proposed shrinkage parameters among themselves shows that $\hat{\rho}_M$ is better than $\hat{\rho}_{GM}$ and $\hat{\rho}_{MED}$.

• The value of $\sigma^2$ affects the performance of OLS and other estimators regardless of the level of correlations between the explanatory variables.