CHAPTER III

ESTIMATION OF MULTICOLLINEARITY RIDDEN

CLASSICAL LINEAR REGRESSION MODEL

1. INTRODUCTION

A crucial condition for the application of Ordinary Least Squares (OLS) while estimating coefficients in a linear regression model, is that the explanatory variables are not perfectly correlated. If they are so, the parameters are not estimable in the sense that it is impossible to obtain numerical values for each parameter separately and consequently the OLS method breaks down. This presence of association among explanatory variables is often termed as multicollinearity. It is a phenomenon inherent in many economic relationships due to the very nature of variables. For instance, rise in income, consumption, savings, prices, employment, investments, etc. in periods of economic boom and their decrease in periods of recession contribute to multicollinearity. So does the use of lagged variables in the relationship. Though multicollinearity is quite frequent in cross-sectional data, it is more common and serious in time series data. Srivastava (1975) has presented a brief account of these.

When multicollinearity is present, one may obtain estimates with large standard error which is obviously an
undesirable feature from statistical inference point of view. Besides it, the estimator may show considerable instability, i.e. small variations in the data matrix may lead to large variations in estimates as well as standard errors. Quite frequently the estimates may have wrong signs; see Gunst and Mason (1980) for an interesting exposition.

Various solutions to cope up with the vexing problem of multicollinearity have been presented in literature. Among them, an interesting proposal was forwarded by Feldstein (1973) who conceived two estimators - Conditionally Omitted Variable (COV) and the Weighted Average (WTD) estimators and studied their properties. He also compared COV and WTD estimators with the OLS estimator using simulation experiments. The limitations of such experiments are well known, as such it may well be risky to generalise his findings.

In the present chapter, we study exact finite sample properties of both, the COV and the WTD, estimators through analytical methods and compare them with those of the OLS estimator.

2. THE MODEL AND THE ESTIMATORS

Feldstein (1973) considered a two-regressor model

\[ Y_t = \beta x_t + \gamma z_t + u_t; \quad t = 1, 2, \ldots, T \]

\[ \mathbb{E}(u_t) = 0, \quad \mathbb{E}(u_t^2) = \sigma^2 \quad \text{and} \quad \mathbb{E}(u_t u_{t-g}) = 0 \]

where \( x_t \) and \( z_t \) are non-stochastic, each variable having
mean zero and \( u_t \) is a serially independent random disturbance (scalar) term with mean zero and constant variance \( \sigma^2 \). It is further assumed that explanatory variables \( x \) and \( z \) are highly correlated. If \( \beta \) is the only parameter of interest in (1), he relaxed the condition of unbiasedness on the estimators and took the mean-squared error (MSE) as measure of their goodness.

2.1 COV ESTIMATOR

If \( \hat{\beta} \) is the OLS estimator and \( \hat{b} \) is the Omitted Variable (OV) estimator of \( \beta \) in (1), then \( \hat{b} \) is, in fact, the OLS estimator of \( b \) in the following omitted variable equation:

\[
y_t = bx_t + v_t \quad ; \quad t = 1, 2, \ldots, T
\]

where \( v_t \) is the usual random disturbance term with zero mean and constant variance.

In fact Bancroft (1944) first studied the COV-type estimator adopting a procedure of preliminary test whether \( 7 \) equals zero; if so, \( z \) could obviously be omitted without introducing any bias into the estimate of \( \beta \). He concentrated on deriving estimates of the bias when \( 7 \) does not equal zero but he did not calculate the MSE of the estimator of \( \beta \). Further studies were conducted, among others, by Toro-Vizcorondo and Wallace (1968, 1969) and Wallace (1972).

Writing \( x = (x_1, x_2, \ldots, x_T)' \), \( y = (y_1, y_2, \ldots, y_T)' \)
and \( z = (z_1, z_2, \ldots, z_T)' \), let us define the ratio as below.
(4) \[ t_\gamma = \gamma / \{ \sigma^2 (x'x) D^{-1} \}^{1/2} \]

where

(5) \[ D = \left| (x'x)(z'z) - (x'z)^2 \right| \]

Clearly, \( t_\gamma \) is unknown as it involves unknown parameters \( \gamma \) and \( \sigma^2 \). An operational estimator of \( t_\gamma \) is given by

(6) \[ \hat{t}_\gamma = \hat{\gamma} / \{ \hat{\sigma}^2 (x'x) D^{-1} \}^{1/2} \]

where \( \hat{\sigma}^2 \) is the unbiased estimator of \( \sigma^2 \) on \((\hat{\beta}, \hat{\gamma})\).

Feldstein (1973) showed that the MSE of \( \hat{\beta} \) is smaller (larger) than that of \( \hat{\beta} \) according as \( t_\gamma^2 < 1 \) ( \( t_\gamma^2 > 1 \)). The COV estimator forwarded by him is a probabilistic mixture of \( \hat{\beta} \) and \( \hat{\beta} \). Its relative MSE, therefore, depends upon the probability of selecting the OLS and COV estimators. Relative to a parameter \( \xi \), the COV estimator, which is operational as well, is thus defined by

(7) \[ \beta[\text{COV } \xi] = \begin{cases} \hat{\beta} & \text{if } \hat{t}_\gamma \geq \xi \\ \hat{\beta} & \text{if } \hat{t}_\gamma < \xi \end{cases} \]

He suggested, through MSE's estimated from simulation experiments, that COV estimator should not be substituted for OLS estimator unless there is strong prior belief that \( t_\gamma \) is less than one. However, one may not be inclined to generalise from the above findings.

2.2 WTD ESTIMATOR

The COV estimator is defined by essentially arbitrary test statistic for selecting, in any sample, between
the OLS and the OV estimators. As a generalisation of above procedure, a shrinkage-type estimator is defined, which is a weighted average of the OLS and the OV estimators, as below

\[ \hat{\beta}^* = \lambda \hat{\beta} + (1-\lambda) \hat{\delta} \]

where \( \lambda \) is the characterising scalar. The optimal value of \( \lambda \), which minimises the MSE of \( \beta^* \) is provided by

\[ \lambda^* = \frac{\gamma^2_D}{(x'x)\sigma^2 + \gamma^2_D} \]

employing \( t_\gamma \) defined by (5), \( \lambda^* \) is given by

\[ \lambda^* = \frac{t_\gamma^2}{1 + t_\gamma^2} \]

However, the true value of \( t_\gamma \) being unknown, an operational value of \( \lambda^* \) is obtained by replacing \( t_\gamma \) with \( \hat{t}_\gamma \) defined at (6). It is used to define operational WTD estimator.

Feldstein (1973) used the simulation experimental average data for analysing the MSE of an operational weighted estimator defined by

\[ \tilde{\beta} = \hat{\lambda}^* \beta + (1-\hat{\lambda}^*) \hat{\delta} \quad \text{with} \]

\[ \hat{\lambda}^* = \hat{t}_\gamma^2 / (1 + \hat{t}_\gamma^2) \]

On the basis of certain arbitrary values of \( t_\gamma \), he established that \( \tilde{\beta} \) is always better than the OLS provided the true value of \( t_\gamma \) is known. However, he opined that relative gain decreases as \( t_\gamma \) increases and, in the process,
this gain will be outweighed by the error in estimating $t_\gamma$, if it has to be.

Therefore, Feldstein (1973) cautioned against arbitrary rejection of OLSE in favour of the COV or even the WTD estimator. He suggested that even when a WTD estimator is used in an attempt to reduce the MSE, the OLSE along with the associated covariance matrix should also be presented, as a measure of the effect of collinearity.

3. THE EXACT FINITE SAMPLE PROPERTIES

With a view to discuss exact properties of COV and WTD estimators, we assume that the disturbance $u_\tau$ follows a normal distribution. This implies that the estimators $\hat{\beta}$ and $\hat{\gamma}$ follow jointly a bivariate normal distribution with the mean vector

$$E(\hat{\beta}, \hat{\gamma})^* = (\beta, \gamma)^*$$

and the covariance matrix given by

$$V(\beta, \gamma)^* = \sigma^2 \begin{bmatrix} x'x & x'z \\ z'x & z'z \end{bmatrix}^{-1}$$

Moreover, the OV estimator $b$ may be, it is recalled (Feldstein (1973) p 340), written as under

$$\hat{b} = \hat{\beta} + (x'x)^{-1}(x'z) \hat{\gamma}$$

writing

$$\sigma^2 = \sigma^2(x'x) \Sigma^{-1}, \quad \hat{\sigma}^2 = \frac{\sigma^2}{(x'x)D^{-1}}$$
the marginal distribution of \( \hat{\gamma} \) is \( N(\gamma, \sigma_\gamma^2) \). Since \( (n\hat{\sigma}_\gamma^2/\sigma_\gamma^2) \) will thus be distributed as \( \chi^2 \) with \( n (= T-2) \) degrees of freedom, it is easily verified from (6) that

\[
(17) \quad \hat{\gamma}^2 = \frac{\hat{\gamma}^2}{\sigma_\gamma^2}
\]

and \( \hat{\gamma} \) follows a non-central Student's t-distribution with \( n \) degrees of freedom and the non-centrality parameter \( (\gamma/\sigma_\gamma) \).

Define the two variables as under

\[
(18) \quad v = \frac{n\hat{\sigma}_\gamma^2}{\sigma_\gamma^2} \\
(19) \quad w = \frac{\hat{\gamma}^2}{\sigma_\gamma^2}
\]

It is noted that \( w^2 \) is a non-central \( \chi^2 \)-variate parameter with one degree of freedom and the non-centrality parameter is given by

\[
(20) \quad \phi = \frac{\gamma^2}{\sigma_\gamma^2}
\]

Employing (18) and (19) in (17) reduces it to

\[
(21) \quad \hat{\gamma}^2 = \frac{n\hat{\sigma}_\gamma^2}{v}
\]

which, in turn, reduces (12) to the following form.

\[
(22) \quad \hat{\Lambda}^* = \frac{w^2}{w^2 + v}
\]

It may be re-written as below

\[
(23) \quad \hat{\Lambda}^* = \frac{w^2}{w^2 + v} \left( 1 - \left( \frac{n}{n-1} \right) \left( \frac{v}{w^2 + v} \right) \right)^{-1}
\]
Expanding the right hand side of (23) by Taylor's series, we may write

(24) \[ \hat{\lambda}^* = \sum_{\alpha=0}^{\infty} \frac{|n-1|^\alpha}{(n)!} \frac{w'^\alpha}{(w^{2}+v)^{\alpha+1}} \]

**THEOREM I**

The bias and the mean-squared error of the COV estimator \( \beta(\text{COV} \xi) \) are given by

(25) \[ \text{Bias } \beta(\text{COV} \xi) = -\left( \frac{x'^2}{x'^x} \right) \frac{\sigma_\gamma e^{-\theta/2}}{\Gamma(n/2)\sqrt{2\pi}} \]

\[ \sum_{j=1}^{\infty} \frac{(\theta/2)^{j/2}}{j!} 2^j \left[ \frac{1}{\frac{j+n+2}{2}} \right] B_{\xi^*} \left( \frac{n}{2} + \frac{j+2}{2} \right) \]

(26) \[ \text{MSE } \beta(\text{COV} \xi) = \sigma_\beta^2 + \left( \frac{x'^2}{x'^x} \right)^2 \frac{2\sigma_\gamma e^{-\theta/2}}{\Gamma(n/2)\sqrt{2\pi}} \]

\[ \sum_{j=0}^{\infty} \frac{(2\theta)^{j/2}}{j!} \left\{ \frac{2\theta}{\frac{j+n+2}{2}} \right. \left[ \frac{1}{\frac{j+n+2}{2}} \right] B_{\xi^*} \left( \frac{n}{2} + \frac{j+2}{2} \right) \]

\[ - \left[ \frac{1}{\frac{j+n+3}{2}} \right] B_{\xi^*} \left( \frac{n}{2} + \frac{j+3}{2} \right) \]

where

(27) \[ \sigma_\beta^2 = \sigma^2 (z'^z) D^{-1} \]

(28) \[ \xi^* = n / (n + \xi^2) \]

(29) \[ B_{\xi^*}(p,q) = \int_0^{\xi^*} h^{p-1} (1 - h)^{q-1} dh \]
(29) represents the well known Incomplete Beta function (see Johnson and Kotz (1970) p 38).

Similarly, under identical assumptions, the properties of the WTD estimator \( \hat{\beta} \) are presented in the following theorem.

**Theorem II**

The bias and the mean-squared error of the weighted average (WTD) estimator are given by

\[
\text{Bias (} \hat{\beta} \text{)} = \frac{\gamma e^{-\gamma/2}}{n} \left( \frac{x'z}{x'x} \right).
\]

\[
\sum_{\alpha=0}^{\infty} \sum_{j=0}^{\alpha} \frac{(n-1)^{\alpha} (e/2)^j}{n^j} \frac{B\left(\alpha + \frac{n+2}{2}, j + \frac{3}{2}\right)}{B\left(\frac{n}{2}, j + \frac{3}{2}\right)}
\]

\[
\text{MSE (} \hat{\beta} \text{)} = \sigma^2 + \frac{x'z^2}{x'x} \frac{2\sigma^2 e^{-\gamma/2}}{n}
\]

\[
\sum_{\alpha=0}^{\infty} \sum_{j=0}^{\alpha} \frac{(n-1)^{\alpha} (e/2)^j}{n^j} \frac{B\left(\alpha + \frac{n+1}{2}, j + \frac{3}{2}\right)}{B\left(\frac{n}{2}, j + \frac{3}{2}\right)}
\]

\[
\left\{ \frac{(\alpha+1)(\alpha+2)}{2n} - \frac{2j-e+1}{n(j+\alpha+2)} \right\} \frac{B\left(\frac{n+2}{2}, j + \frac{3}{2}\right)}{B\left(\frac{n}{2}, j + \frac{3}{2}\right)}
\]

where \( B(p, q) \) is the well-known Beta function.

From the expressions of bias and mean-squared error for both, the COV and the WTD estimators, it is clear that meaningful conclusions cannot be drawn from their compari-
son, we have therefore, evaluated them for a few selected values of the parameters.

4. **EFFICIENCY OF ESTIMATORS**

The results of the preceding section pose interesting questions of practical significance. How does the two estimators — the COV and the WTD, compare with the OLS estimator of $\beta$? What impact the degree of relationship between $x$ and $z$ variates may have on the goodness of the estimators? In order to answer these, we compute numerical results precisely for the same set of parameters as envisaged by Feldstein (1973).

Therefore, $r_{xz}$, the linear correlation coefficient between the explanatory variables $x$ and $z$ is assumed to be 0.98, 0.90 and 0.70 respectively for the three tables. Considering the regression sample of size 20 ($= T$), we take twelve equidistant values of $\bar{e}$ ranging from 0.0625 to 9.0000 (or $t_\gamma$ from 0.25 to 3.00). Other parameters are chosen as $(x'x) = (z'z) = 1$ and $\gamma = 1$. From these the remaining parameters viz. $\sigma^2$, $x'z$ are calculated. The values of $\delta$ used to define the COV estimators are also chosen, like Feldstein (1973), so as to correspond to particular probabilities of using the omitted variable estimator $\hat{\beta}^\Delta$ when $t_\gamma$ is actually less than one. For example, $\delta = 1.741$ corresponds to implication that $\hat{\beta}^\Delta$ is to be used in 75 percent cases when $t_\gamma$ equals to 1 and in more cases if $t_\gamma$ is less than 1. Correspondingly, $\delta$ are obtained from (28) and for convenience, we have included both,
and $\hat{\lambda}$ in the tables.

All the tables present the ratio of the MSE of OLS estimator to those of every other estimator, which in each case has been treated as 'relative efficiency'.

Stein (1956, 1962) established that OLS estimator can not be dominated by any other estimator, even outside the class of unbiased estimators, when upto only two parameters are to be estimated. Feldstein (1973) observed that use of a COV estimator makes the MSE loss function less sensitive to the value of $t_\gamma$ than the OLS estimator. He further observed that none of the COV estimators dominates the OLSE over the entire range $0.0625 \leq \theta \leq 9.0000$ (or $0.25 \leq t_\gamma \leq 3.00$). Based on their exact finite sample properties, the results in Table 4.1, 4.2 and 4.3 also present a similar picture. It is evident from all tables that COV estimator is uniformly better for all values of $\hat{\lambda}$ for $\theta \leq 0.5625$ and the gain shows a decreasing trend as $\hat{\lambda}$ increases (or $\hat{\lambda}$ decreases). For $\theta \geq 1.00$ this estimator is dominated by the OLS estimator everywhere except for $\hat{\lambda} \leq 0.9430$ (or $\hat{\lambda} \geq 1.072$) when both estimators are approximately equivalent.

Moreover, we observe that higher is the degree of collinearity in the model, larger will be the efficiency of COV estimator wherever it dominates the OLSE. But, at the same time, it will lose heavily wherever the OLSE dominates. Thus, we suggest that COV estimator should be used only if there is a strong prior belief that $\theta$ is less than one.
TABLE 4.1

RELATIVE EFFICIENCIES FOR COV AND WTD ESTIMATORS

\( r_{xz} = 0.70 \)

<table>
<thead>
<tr>
<th>Est</th>
<th>COV</th>
<th>COV</th>
<th>COV</th>
<th>COV</th>
<th>COV</th>
<th>COV</th>
<th>WTD</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \xi )</td>
<td>0.6991</td>
<td>0.7648</td>
<td>0.8116</td>
<td>0.8624</td>
<td>0.9430</td>
<td>0.9500</td>
<td></td>
</tr>
<tr>
<td>( \xi )</td>
<td>2.8600</td>
<td>2.4170</td>
<td>2.1000</td>
<td>1.7410</td>
<td>1.0720</td>
<td>1.0000</td>
<td></td>
</tr>
</tbody>
</table>

\[ \xi^* = (\xi^2 + n)^{-1}, n \quad \text{and} \quad \varepsilon = t_y^2 \]

Note: If the true value \( t_y \) were known, Feldstein (1973) opined that WTD estimator \( \hat{\beta} \) would dominate the OLS estimator irrespective of the value of \( t_y \). However, he added that relative gain in efficiency of \( \hat{\beta} \) would decrease as the value of \( t_y \) increases. From the tables we observe that the WTD estimator is also not uniformly better than the OLS though it appears to be less sensitive to variations in values of \( t_y \), than the COV estimator. Moreover, there is clear evidence from all the tables that \( \hat{\beta} \) will dominate the OLS estimator \( \hat{\beta} \) whenever \( \varepsilon \leq 1.5625 \).
TABLE 4.2

RELATIVE EFFICIENCIES FOR COV AND WTD ESTIMATORS

( $r_{xz} = 0.90$ )

<table>
<thead>
<tr>
<th>Est</th>
<th>COV</th>
<th>COV</th>
<th>COV</th>
<th>COV</th>
<th>COV</th>
<th>COV</th>
<th>COV</th>
<th>WTD</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\xi$</td>
<td>0.6991</td>
<td>0.7648</td>
<td>0.8116</td>
<td>0.8624</td>
<td>0.9430</td>
<td>0.9500</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\xi$</td>
<td>2.8600</td>
<td>2.4170</td>
<td>2.1000</td>
<td>1.7410</td>
<td>1.0720</td>
<td>1.0000</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

0.0625  | 3.3818 | 2.8952 | 2.5227 | 2.1432 | 1.7001 | 1.6753 | 1.7018 |     |
0.2500  | 2.1133 | 1.8758 | 1.7027 | 1.5375 | 1.3924 | 1.3912 | 1.5536 |     |
0.5625  | 1.3296 | 1.2309 | 1.1678 | 1.1217 | 1.1609 | 1.1662 | 1.3751 |     |
1.0000  | 0.8918 | 0.8601 | 0.8504 | 0.8643 | 0.9914 | 1.0128 | 1.2108 |     |
1.5625  | 0.6406 | 0.6436 | 0.6620 | 0.7087 | 0.8917 | 0.9176 | 1.0791 |     |
2.2500  | 0.4897 | 0.5135 | 0.5491 | 0.6171 | 0.8374 | 0.8659 | 0.9820 |     |
3.0625  | 0.3960 | 0.4344 | 0.4829 | 0.5677 | 0.8161 | 0.8457 | 0.9146 |     |
4.0000  | 0.3371 | 0.3879 | 0.4480 | 0.5488 | 0.8188 | 0.8481 | 0.8709 |     |
5.0625  | 0.3011 | 0.3641 | 0.4363 | 0.5536 | 0.8379 | 0.8655 | 0.8453 |     |
6.2500  | 0.2814 | 0.3579 | 0.4438 | 0.5782 | 0.8667 | 0.8910 | 0.8330 |     |
7.5625  | 0.2747 | 0.3671 | 0.4685 | 0.6198 | 0.8988 | 0.9188 | 0.8303 |     |
9.0000  | 0.2793 | 0.3909 | 0.5094 | 0.6747 | 0.9290 | 0.9442 | 0.8340 |     |

( Note: $\xi^* = (\xi^2 + n)^{-1}$. n and $\theta = t_{x}^2$ )

While for $\theta = 2.25$ the two estimators are approximately equivalent. Further, with increase in the degree of collinearity, the efficiency of $\hat{\beta}$ increases for $\theta \leq 1.5625$ while it decreases, though less rapidly as for $\beta$(COV $\xi$), for $\theta \geq 2.25$. It is also interesting to note that $\hat{\beta}$ dominates $\beta$(COV $\xi$) for $\theta \geq 1$, the dominance being more and more prominent as $\theta$ becomes larger or $\xi^*$ becomes smaller. For $\theta = 0.5625$, the two estimators are almost equally disposed and for $\theta < 0.5625$, the $\beta$(COV $\xi$) usually dominates $\hat{\beta}$, of course more prominently if the degree of collinearity is higher and $\xi^*$ is small.
Thus, on the whole, we observe that a WTD estimator compares more favourably with the OLSE though it is not always better. If a researcher can confidently bound the true value $t_\gamma$ to be below 3.00 or so, but has no more information, he may find WTD estimator more suitable than the OLSE, some modifications in the ranges wherein either estimator (the $\hat{\beta}$ or $\beta(\text{COV} \ \hat{x})$) is claimed better notwithstanding. These findings corroborate Toro-Vizcarrondo and Wallace (1968, 1969) for the fact that MSE of an estimator of $\beta$ can be reduced by omitting $z$ if and only if the 'true $t_\gamma$' is less than one.
5. PROOF OF THEOREMS

Before proving the results in Theorem I, we will prove the following Lemma.

Lemma 1

If \( \hat{\gamma} \sim N(\gamma, \sigma_\gamma) \) and \( \hat{\xi}_\gamma = (\hat{\gamma} / \sigma_\gamma) \) is distributed as non-central Student's t-distribution with \( n \) degrees of freedom and the non-centrality parameter \( (\gamma / \sigma_\gamma) \). Then

\[
dF(\hat{\gamma}, \hat{\xi}_\gamma) = \frac{(n/\sigma_\gamma)^{n/2} e^{-a/2}}{2^{n/2} \Gamma(n/2) \sqrt{\sigma_\gamma} 2^n} \cdot \frac{1}{(\hat{\xi}_\gamma)^{n+1}} \left( \frac{\gamma}{\sigma_\gamma} \right)^{j+n} \left( \sigma_\gamma \right)^{j} \exp \left\{ -\frac{1}{2} \left( \frac{\hat{\gamma} - \gamma}{\sigma_\gamma} \right)^2 \right\} \frac{\partial^j}{\partial \hat{\xi}_\gamma^j} \left[ A(\hat{\xi}_\gamma) \right] d\gamma d\xi_\gamma
\]

where

\[
A(\hat{\xi}_\gamma) = \frac{1}{2 \sigma_\gamma^2} \left[ 1 + \frac{n}{\hat{\xi}_\gamma^2} \right]
\]

Proof:

Since \( \sigma_\gamma^2 \sim \chi^2(n) \sigma_\gamma^2 \), we have

\[
dF(\hat{\gamma}, \hat{\xi}_\gamma^2) = \frac{1}{\sigma_\gamma \sqrt{2\pi}} \exp \left\{ -\frac{1}{2} \left( \frac{\hat{\gamma} - \gamma}{\sigma_\gamma} \right)^2 \right\} \cdot \frac{(n/\sigma_\gamma^2)}{2^{n/2} \Gamma(n/2) \sqrt{(n/2)}} \left( \sigma_\gamma^2 \right)^{-1} \exp \left\{ -\frac{1}{2} \left( \sigma_\gamma^2 / \sigma_\gamma^2 \right) \right\} d\sigma_\gamma^2
\]
employing the following transformation in (34)

\[
\hat{\tau}^2 = \left( \hat{\tau} \hat{\xi} \right) \quad -\infty \leq \hat{\xi} \leq \infty
\]

so that the jacobian of the transformation is given by

\[
\left( 2\hat{\tau}^2 / \hat{\xi}^3 \right)
\]

we obtain

\[
dF(\hat{\tau}, \hat{\xi}) = \frac{1}{\sigma_{\gamma} / 2} \exp \left\{ -\frac{1}{2} \left( \frac{\hat{\tau} - \gamma}{\sigma_{\gamma}} \right)^2 \frac{(n/\sigma_{\gamma}^2)^{n/2}}{(n/2)} \right\} \left( \frac{\hat{\tau}}{\hat{\xi}} \right)^{n-2} \exp \left\{ -\frac{1}{2} \frac{\hat{\tau}^2}{(\sigma_{\gamma} \hat{\xi})^2} \right\}
\]

\[
\times \frac{2\hat{\tau}^2}{\hat{\xi}^3} \, d\hat{\tau} \, d\hat{\xi}
\]

employing (17) and (20) leads to, after little algebraic manipulations,

\[
dF(\hat{\tau}, \hat{\xi}) = \frac{(n/\sigma_{\gamma}^2)^{n/2} \exp \left\{ -\frac{1}{2} \left( \frac{\hat{\tau} - \gamma}{\sigma_{\gamma}} \right)^2 \frac{(n/\sigma_{\gamma}^2)^{n/2}}{(n/2)} \right\}}{\frac{2^n}{2^2 - 1} \left[ (n/2) \sigma_{\gamma} \sqrt{2\pi} \hat{\xi} \right]}
\]

\[
\times \exp \left\{ -\frac{1}{2} \frac{\hat{\tau}^2}{(\sigma_{\gamma} \hat{\xi})^2} \right\} \left( \frac{\hat{\tau}}{\hat{\xi}} \right)^{n} \exp \left\{ \frac{\hat{\tau}^2}{2\sigma_{\gamma}} \right\} \exp \left\{ -\frac{n}{2\sigma_{\gamma}^2} \left( 1 + \frac{n}{\hat{\xi}^2} \right) \right\} \exp \left\{ \hat{\tau}/\sigma_{\gamma} \right\}
\]

expanding the second exponential term on the right hand side of the above leads, after some further rearrangement of terms, and use of notation (33), to the result (32).

Now, we note from the definition of \( \beta(\text{COV } \xi) \) given by (7) that

\[
E \{ \beta(\text{COV } \xi) \} = E \{ \hat{\tau} \mid \hat{\xi} \geq \xi \} P(\hat{\xi} \geq \xi) + E \{ \hat{\tau} \mid \hat{\xi} < \xi \} P(\hat{\xi} < \xi)
\]

\[
+ E \{ \hat{\tau} \mid \hat{\xi} \leq \xi \} P(\hat{\xi} \leq \xi)
\]
Employing (15) in above, we get

\[ 1 \beta (\text{COV} \xi) \leq 1 \beta (\xi) + (x^2 / x^2) \beta (\xi) \frac{1}{x^2} (\xi, \xi \gamma) \]

we consider the conditional expectation term of the right hand side of (38) i.e.

\[ 1 \beta (\xi, \xi \gamma) \]

\[ = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{(n/\gamma^2) \gamma / 2}{\epsilon / 2} e^{-\epsilon / 2} \left( \frac{1}{2 \gamma} \right)^{n+1} \]

\[ = \sum_{j=0}^{\infty} \left( \frac{\gamma j}{\sigma^2} \right) j \left( \frac{1}{\gamma} \right)^{j+1} \exp \left( - \frac{\gamma^2 \Lambda (\xi, \xi \gamma) d_{\gamma} \epsilon \right) \int \int_{-\infty}^{\infty} \left( \frac{1}{2 \gamma} \right)^{n+1} \]

\[ = \frac{(n/\gamma^2) \gamma / 2}{\epsilon / 2} \left( \frac{1}{\gamma} \right)^{n+1} \sum_{j=0}^{\infty} \left( \frac{\gamma j}{\sigma^2} \right) j \left( \frac{1}{\gamma} \right)^{j+1} \exp \left( - \frac{\gamma^2 \Lambda (\xi, \xi \gamma) d_{\gamma} \right) \int \int_{-\infty}^{\infty} \left( \frac{1}{2 \gamma} \right)^{n+1} \]

\[ = \frac{(n/\gamma^2) \gamma / 2}{\epsilon / 2} \left( \frac{1}{\gamma} \right)^{n+1} \sum_{j=0}^{\infty} \left( \frac{\gamma j}{\sigma^2} \right) j \left( \frac{1}{\gamma} \right)^{j+1} \exp \left( - \frac{\gamma^2 \Lambda (\xi, \xi \gamma) d_{\gamma} \right) \int \int_{-\infty}^{\infty} \left( \frac{1}{2 \gamma} \right)^{n+1} \]

\[ = \frac{1}{2 \gamma} \left( \frac{1+n+2}{2} \right) \left( \frac{1}{\gamma} \right)^{n+1} \int \int_{-\infty}^{\infty} \left( \frac{1}{2 \gamma} \right)^{n+1} \left( 1 + \frac{1}{n+2} \right) \frac{1}{2 \gamma} \left( \frac{1}{\gamma} \right)^{n+1} \exp \left( - \frac{\gamma^2 \Lambda (\xi, \xi \gamma) d_{\gamma} \right) \int \int_{-\infty}^{\infty} \left( \frac{1}{2 \gamma} \right)^{n+1} \]
A little manipulation reduces the above to

\[ P(\xi \mid \xi_{\gamma} < \xi) = \frac{\delta_{\gamma} e^{-\theta/2}}{2} \sum_{j=0}^{\infty} \frac{j/2}{2^j} \frac{(n/2)\Gamma(j+1/2)}{(n)(j+1/2)^{n+1}} \]

Applying the following transformation

\[ 1 + \frac{\xi_{\gamma}^2}{n} = \frac{1}{h} \Rightarrow \xi_{\gamma}^2 = \frac{n(1-h)}{h} ; 0 \leq h \leq 1^* \]

where \( 1^* = \{(1^2+n)^{-1} \cdot n\} \) and the jacobian of transformation is given by

\[ -2\xi_{\gamma} d\xi_{\gamma} = \frac{n}{h^2} \quad \text{dh} \]

the integral in (40) converts to the well-known incomplete beta function defined by (29). Employing it, the equation (40) thus reduces to

\[ P(\xi \mid \xi_{\gamma} < \xi) = \frac{\delta_{\gamma} e^{-\theta/2}}{2} \sum_{j=0}^{\infty} \frac{(\theta/2)^{j/2}}{j!} \frac{(n/2)\Gamma(j+1/2)}{(n)(j+1/2)^{n+1}} \]

Employing (43) in the equation (38) leads to, after some adjustment of terms, the result (25) of Theorem I.
To evaluate the mean-squared error of $\beta(\text{COV} \xi)$, we again note from the definition given by (7) that

$$E[\beta(\text{COV} \xi) - \beta]^2 = E[(\hat{\beta} - \beta)^2 | \xi > t] P(\xi > t)$$

$$+ E[(\hat{\beta} - \beta)^2 | \xi \leq t] P(\xi \leq t)$$

Utilizing the relation (15) in (44), we get

$$E[\beta(\text{COV} \xi) - \beta]^2 = E(\hat{\beta} - \beta)^2 +$$

$$(x'z/x'x) E[\hat{\gamma}^2 | \xi > t] P(\xi > t)$$

$$+ 2(x'z/x'x) E[\hat{\gamma} (\hat{\beta} - \beta) | \xi \leq t] P(\xi \leq t)$$

Considering the second conditional expectation term of (45), it is noted that

$$E[\hat{\gamma} (\hat{\beta} - \beta) | \xi > t] = E[\hat{\gamma} E(\hat{\beta} - \beta | \xi) | \xi > t]$$

$$= E[(-x'z/x'x) \hat{\gamma} (\hat{\beta} - \gamma) | \xi > t]$$

$$= -(x'z/x'x) E[\hat{\gamma}^2 | \xi > t] + \gamma (x'z/x'x).$$

Utilizing (46) in (45), in turn, provides after some adjustment of terms, the following

$$E[\beta(\text{COV} \xi) - \beta]^2 = E(\hat{\beta} - \beta)^2$$

$$- (x'z/x'x)^2 E[\hat{\gamma}^2 | \xi > t] P(\xi > t)$$

$$+ 2 \gamma (x'z/x'x)^2 E[\hat{\gamma} | \xi \leq t] P(\xi \leq t)$$
On the right hand side of (47), the conditional expectation of the middle term is taken up as under

\[ E(\hat{\gamma}^2 | \hat{\gamma} < \xi)P(\hat{\gamma} < \xi) = \int_{-\infty}^{\xi} \int_{-\infty}^{\infty} \hat{\gamma}^2 dF(\hat{\gamma}, \hat{\gamma}) \]

\[ = \frac{(n/\sigma_{\gamma}^2/n)^{n/2} e^{-\bar{\vartheta}/2}}{2^{n/2} \Gamma(n/2) \sigma_{\gamma} \sqrt{2\pi}} \sum_{j=0}^{\infty} \left( \frac{\gamma}{\sigma_{\gamma}^2} \right)^j \frac{1}{j!} \]

\[ \times \int_{-\infty}^{\xi} \left[ \int_{0}^{\infty} (\hat{\gamma}^2)^{j+1} \exp(-\hat{\gamma}^2) A(\hat{\gamma}) d\hat{\gamma}^2 \right] d\hat{\gamma} \]

\[ = \frac{(n/\sigma_{\gamma}^2/n)^{n/2} e^{-\bar{\vartheta}/2}}{2^{n/2} \Gamma(n/2) \sigma_{\gamma} \sqrt{2\pi}} \sum_{j=0}^{\infty} \left( \frac{\gamma}{\sigma_{\gamma}^2} \right)^j \frac{1}{j!} \]

\[ \times \int_{-\infty}^{\xi} \left[ \Gamma\left(\frac{j+3}{2}\right) \left\{ \frac{1}{2\sigma_{\gamma}^2} \left( 1 + \frac{n}{\hat{\gamma}^2} \right) \right\}^{j/2} \right] d\hat{\gamma} \]

\[ = \frac{\sigma_{\gamma}^2 e^{-\bar{\vartheta}/2}}{\Gamma(n/2) \sqrt{2\pi}} \sum_{j=0}^{\infty} \left( \frac{\vartheta/2}{\sigma_{\gamma}^2} \right)^j/2 \frac{j+3/2}{n/2} \Gamma\left(\frac{j+3}{2}\right) \]

\[ \times \int_{-\infty}^{\xi} (\hat{\gamma})^{j+2} \left( 1 + \frac{\hat{\gamma}^2}{n} \right)^{-j+3/2} d\hat{\gamma} \]

\[ = \frac{\sigma_{\gamma}^2 e^{-\bar{\vartheta}/2}}{\Gamma(n/2) \sqrt{2\pi}} \sum_{j=0}^{\infty} \left( \frac{\vartheta/2}{\sigma_{\gamma}^2} \right)^j/2 \frac{j+3/2}{n/2} \Gamma\left(\frac{j+3}{2}\right) B_{\frac{n}{2}} \left( \frac{j+3}{2} \right) \]

where \( B_{\frac{n}{2}}(\cdot, \cdot) \) is the incomplete beta function defined at (29).

Employing (43) and (48) in (47) easily leads to the result (26) of the Theorem I.
For proving the results stated in Theorem 11 about the WTD estimator, we first state the following lemma.

**Lemma 2**

Let \( w \sim N(e^2, 1) \) and \( v \sim \chi^2(n) \) and they be independent. Then, for integer values of \( p \) and \( q \), we have

\[
E \left\{ \frac{w^p v^q}{(w^2 + v)^q} \right\} = \frac{2^{p/2} \Gamma(q + \frac{n}{2}) e^{-\phi/2}}{\Gamma(n/2)} \cdot \frac{\sum_{j=0}^{\infty} \frac{\Gamma(j + \frac{p+q-1}{2}) \Gamma(j + \frac{p+1}{2})}{\Gamma(j + q + \frac{p+1}{2}) \Gamma(j + \frac{1}{2})} (\phi/2)^j}{j!}
\]

if \( p \) is an even integer; and

\[
\frac{2^{p-1} \Gamma(q + \frac{n}{2}) \sum_{j=0}^{\infty} \frac{\Gamma(j + \frac{p+q+2}{2}) \Gamma(j + \frac{p+2}{2})}{\Gamma(j + q + \frac{p+2}{2}) \Gamma(j + \frac{3}{2})} (\phi/2)^j}{j!}
\]

if \( p \) is an odd integer.

For proof see Dwivedi, Srivastava and Hall (1980), Appendix.

Employing (15) in (11), the WTD estimator \( \hat{\beta} \) reduces to the following form:

\[
\hat{\beta} = \hat{\beta} + (1 - \hat{\lambda}^*) (x'z/x'x) \hat{y}
\]

where \( \hat{\lambda}^* \) is defined by (22). By virtue of (19) and (22), we get

\[
\tilde{\beta} = \tilde{\beta} + (x'z/x'x) \frac{\sigma_{x'z}}{n} \left( \frac{w^2}{w^2 + \frac{v}{n}} \right)^{-1} \frac{w}{w + v}
\]
Employing the expansion for the inverse of the term in (52), leads to after little algebraic manipulations,

\[
\hat{\beta} = \hat{\beta} + \left(x^t z / x^t x\right) \frac{\sigma_y}{n} wv \sum_{i=0}^{\infty} \left(\frac{n-1}{n}\right)^{\alpha} \frac{v^\alpha}{(w^2 + v)^{\alpha+1}}
\]

Thus, the bias of \( \beta \) is given by

\[
E(\hat{\beta} - \beta) = \left(x^t z / x^t x\right) \frac{\sigma_y}{n} \frac{\sum_{i=0}^{\infty} \left(\frac{n-1}{n}\right)^{\alpha} E\left(\frac{w^2 v^2}{(w^2 + v)^{\alpha+1}}\right)}
\]

Utilizing (50) in (54) easily provides the result (30) of the Theorem II. To evaluate the mean-squared error of \( \hat{\beta} \), it is observed from (52) that

\[
E(\hat{\beta} - \beta)^2 = E(\hat{\beta} - \beta)^2 + \left(x^t z / x^t x\right) \frac{\sigma_y}{n} \frac{\sum_{i=0}^{\infty} \left(\frac{n-1}{n}\right)^{\alpha} E\left(\frac{w^2 v^2}{(w^2 + v)^{\alpha+1}}\right)}
\]

\[
= \left(1 - \frac{n-1}{n} \frac{v}{w^2 + v}\right)^{-2} + \frac{2 \sigma_y}{n} \frac{x^t z}{x^t x}
\]

\[
E\left\{\frac{w^2 v^{2+2}}{w^2 + v} \left(1 - \frac{n-1}{n} \frac{v}{w^2 + v}\right)^{-1} (\hat{\beta} - \beta)^2\right\}
\]

\[
= E(\hat{\beta} - \beta)^2 + \left(x^t z / x^t x\right) \frac{\sigma_y}{n} \frac{\sum_{i=0}^{\infty} \left(\frac{n-1}{n}\right)^{\alpha} E\left(\frac{w^2 v^{2+2}}{w^2 + v} (\hat{\beta} - \beta)^2\right)}
\]

\[
= \sum_{\alpha=0}^{\infty} \left(\frac{n-1}{n}\right)^{\alpha} \frac{w^2 v^{\alpha+1}}{(w^2 + v)^{\alpha+1}} E\left(\frac{w^2 v^{\alpha+1}}{(w^2 + v)^{\alpha+1}} (\hat{\beta} - \beta)^2\right)
\]

In order to evaluate the expectation of the last
term on the right hand side of (55), it is split into conditional expectation and by virtue of (13) and (14), it is observed that

\[(56) \quad \mathbb{E}(\hat{\beta} - \beta \mid \hat{\gamma}) = -\left(\frac{x^t z}{x^t x}\right) (\hat{\gamma} - \gamma)\]

Employing (56) we get

\[(57) \quad \mathbb{E}(\hat{\beta} - \beta)^2 = \mathbb{E}(\hat{\beta} - \beta)^2 + \left(\frac{x^t z}{x^t x} \frac{\sigma_\gamma}{n}\right)^2 \cdot \sum_{\alpha=0}^{\infty} (\alpha+1) \left(\frac{n-1}{n}\right)^\alpha \left(\frac{\sum w^2 \cdot v^{\alpha+2}}{(w^2 + v)^{\alpha+2}}\right) \cdot \left\{\frac{\sigma_\gamma}{n} \left(\frac{2 \cdot v^{\alpha+1}}{(w^2 + v)^{\alpha+1}}\right) - \gamma \mathbb{E}\left(\frac{w \cdot v^{\alpha+1}}{(w^2 + v)^{\alpha+1}}\right)\right\}
\]

Use of (49) and (50) in (57) leads, after some adjustment of terms, to the result (31) of the Theorem II.

6. SUMMARY

Multicollinearity introduces a problem in estimation of regression models because the moment matrix, say \(X^t X\) (\(X\) being the matrix of explanatory variables) is no longer non-singular. A variety of solutions leading to alternative estimators have been proposed in literature and most of them are biased. Some of them use additional information. Feldstein (1973) analysed properties of two such biased estimators for a regression model with only two parameters, but his conclusions
were based on simulation experiments which are manifested with well-known limitations. Explaining these two estimators including the operational ones in each case, finite sample properties are discussed in this chapter providing the biases and mean-squared errors for both the operational estimators. In order to compare our findings with those of Feldstein (1973), we assumed a similar format of sampling experiments with same values considered for parameters. By far, we concur with Feldstein (1973) regarding appropriateness of these estimators - the COV and the WTD, but it is indicated that the WTD estimator in fact dominates the OLSE for $\theta \leq 1.5625$, the dominance becoming stronger with the increase in the degree of collinearity. Similarly, it is also indicated that the COV estimator dominates OLSE only for $\theta \leq 0.5625$. As a concluding remark, it is stated that, on the whole, neither of these estimators have proved to be superior to the OLSE and a word of caution is advised for workers. However, since they behave nicely vis-à-vis in certain ranges, these estimators should not be ignored altogether.