APPLIED ECONOMETRICS

Potluri Rao   Roger LeRoy Miller

University of Washington

Wadsworth Publishing Company, Inc., Belmont, California

1971

Chapter 3

Bias and Precision of the Regression Estimates

In the previous chapter we presented an intuitive feel for a number of problems of estimation in applied econometrics. We shall now turn to a more rigorous treatment of some of these problems. In many empirical situations the researcher leaves out some independent variables or includes others that do not belong in the true specification of a causal relation between the dependent and the independent variables. In this chapter we shall study the nature of the bias and of the precision of ordinary least squares estimates when the estimated equation is not the truth. Since bias and precision depend crucially on the theoretical distribution of the error terms, we shall analyze the distributional properties of regression estimates under different specifications of the error terms.

To set the stage, let us consider a case in which the truth is given by

$$y_t = \beta_1 x_{1t} + \varepsilon_t, \qquad (3.1)$$

where the lower-case letters refer to variables measured as deviations from their respective means.

The ordinary least squares estimate of β_1 from a set of given values of y and x_1 is obtained as

$$\hat{\beta}_1 = \Sigma x_{1t} y_t / \Sigma x_{1t}^2 \ . \qquad (3.2)$$

The estimate given by (3.2) may not be equal to the parameter value β_1, because

$$\hat{\beta}_1 = \Sigma x_{1t}(\beta_1 x_{1t} + \varepsilon_t)/\Sigma x_{1t}^2 \ , \qquad (3.3)$$

$$\hat{\beta}_1 = \beta_1 + \Sigma x_{1t}\varepsilon_t/\Sigma x_{1t}^2 \ , \qquad (3.4)$$

and the term $\Sigma x_{1t}\varepsilon_t/\Sigma x_{1t}^2$ is not always equal to zero. Different sets of ε's will give different values for the estimate $\hat{\beta}_1$, even though the values of x_1 are the same. Since the values of the ε's are unknown, the researcher has no way of discovering the seriousness of the deviation of the estimates from the parameter, or in which direction this deviation lies. The best he can do is to establish the theoretical distribution of the estimate $\hat{\beta}_1$ when he knows the properties of the error terms. The statistical distribution of $\hat{\beta}_1$ for various specifications of the error term constitutes the problem of precision of the estimates.

There are several ways of approaching this problem. A convenient specification that serves the needs of applied econometricians is to treat the *x*'s as constants (fixed in repeated samples) and to study the statistical properties of $\hat{\beta}_1$ given by equation (3.4). When the error terms are assumed to have been generated by a specified statistical process, the distributional properties of $\hat{\beta}_1$ can be theoretically established.

A simple case is one in which the error terms are assumed to have been generated by a random selection from a statistical distribution with a mean of zero and a constant variance, σ_ε^2. In this specification the error term corresponding to any time period *t*, is generated by the same statistical distribution, and the error term corresponding to one time period does not depend in any systematic way on the error terms of the other time periods. This specification of the error generating process may be stated as

$$E(\varepsilon_t) = 0, \qquad (3.5)$$
$$E(\varepsilon_t^2) = \sigma_\varepsilon^2, \qquad (3.6)$$
$$E(\varepsilon_t \varepsilon_{t'}) = 0 \quad \text{for } t \neq t'. \qquad (3.7)$$

The first two equations, (3.5) and (3.6), specify that the error terms are generated by a statistical distribution with mean zero and variance σ_ε^2 for all time periods, and the last equation, (3.7), specifies that they are generated by a random selection.

Since the error terms have a statistical distribution, the estimate $\hat{\beta}_1$ which depends on these error terms also has a statistical distribution. The mean of the distribution of $\hat{\beta}_1$ may be obtained by taking the expected value of the estimate

$$E(\hat{\beta}_1) = \beta_1 + E(\Sigma x_{1t}\varepsilon_t / \Sigma x_{1t}^2). \qquad (3.8)$$

Since the distribution of $\hat{\beta}_1$ is defined for our purposes under the assumption that the x_1's are fixed in repeated samples, we obtain under assumption (3.5)

$$E(\hat{\beta}_1) = \beta_1 + \Sigma x_{1t} E(\varepsilon_t) / \Sigma x_{1t}^2 = \beta_1, \qquad (3.9)$$

since $E(\varepsilon_t) = 0$ for all *t*.

The ordinary least squares estimate $\hat{\beta}_1$ has a statistical distribution with mean β_1, the parameter value itself. Thus $\hat{\beta}_1$ is an unbiased estimate of β_1 in the true equation.

Let us now turn to the variance of the theoretical distribution of $\hat{\beta}_1$. When $\hat{\beta}_1$, is an estimate with a statistical distribution, its variance is defined as

$$V(\hat{\beta}_1) = E[\hat{\beta}_1 - E(\hat{\beta}_1)]^2 \qquad (3.10)$$
$$= E[\Sigma x_{1t}\varepsilon_t / \Sigma x_{1t}^2]^2 \qquad (3.11)$$
$$= (1/\Sigma x_{1t}^2)^2 . E[\Sigma x_{1t}\varepsilon_t]^2 . \qquad (3.12)$$

Using the relation

$$E[\Sigma x_{1t}\varepsilon_t]^2 = E[x_{11}^2\varepsilon_1^2 + x_{12}^2\varepsilon_2^2 + \ldots + 2x_{11}x_{12}\varepsilon_1\varepsilon_2 + \ldots]$$
$$= x_{11}^2 E(\varepsilon_1^2) + x_{12}^2 E(\varepsilon_2^2) + \ldots + 2x_{11}x_{12}E(\varepsilon_1\varepsilon_2) + \ldots \qquad (3.13)$$

and substituting the specifications (3.6) and (3.7) in equation (3.13) we obtain

$$E[\Sigma x_{1t}\varepsilon_t]^2 = \sigma_\varepsilon^2 . \Sigma x_{1t}^2 . \qquad (3.14)$$

The variance of the estimate $\hat{\beta}_1$, may be obtained by substituting equation (3.14) in equation (3.12) to obtain

$$V(\hat{\beta}_1) = \sigma_\varepsilon^2 / \Sigma x_{1t}^2 . \qquad (3.15)$$

The variance of the estimate $\hat{\beta}_1$ depends on the variance of ε and on the values of the independent variable. The variance of $\hat{\beta}_1$ increases with the variance of the error term and decreases with Σx_{1t}^2 .

The variance of the estimate $\hat{\beta}_1$, provides a measure of the precision of the estimate. The larger the variance of the estimate, the more widespread the distribution and the smaller the precision of the estimate.

When the error terms are generated randomly by a statistical distribution with mean zero and variance σ_ε^2, the ordinary least squares estimate has a theoretical distribution with mean β_1 and variance $\sigma_\varepsilon^2/\Sigma x_{1t}^2$. The variance of the estimate involves the variance of the error term from the regression equation, which is generally unknown. We can still use these results for comparison of alternative estimation procedures if all procedures involve the same unknowns.

3.1 Irrelevant Variables

Consider a situation in which the researcher could have estimated the parameter β_1, by estimating the regression equation

$$y_t = \tilde{\beta}_1 x_{1t} + \tilde{\beta}_2 x_{2t} + e_t \ . \qquad (3.16)$$

When equation (3.1) is the truth, estimated equation (3.16) is a misspecification of the model, because it includes an irrelevant variable, (x_2).

In the ordinary least squares estimation procedure the estimate $\tilde{\beta}_1$ is obtained as

$$\tilde{\beta}_1 = \frac{\Sigma x_2^2 . \Sigma x_1 y - \Sigma x_1 x_2 . \Sigma x_2 y}{\Sigma x_1^2 . \Sigma x_2^2 - \Sigma x_1 x_2 . \Sigma x_1 x_2} \ . \qquad (3.17)$$

By substituting the true relation (3.1) for *y* in equation (3.17) we obtain

$$\tilde{\beta}_1 = \beta_1 + \frac{\Sigma x_2^2 . \Sigma x_1 \varepsilon - \Sigma x_1 x_2 . \Sigma x_2 \varepsilon}{\Sigma x_1^2 . \Sigma x_2^2 - \Sigma x_1 x_2 . \Sigma x_1 x_2} \ . \qquad (3.18)$$

Using the specification of the error terms (3.5) and assuming that the *x*'s are fixed in repeated samples (nonstochastic), we can show that

$$E(\tilde{\beta}_1) = \beta_1 \ . \qquad (3.19)$$

The distribution of $\tilde{\beta}_1$, the estimate of β_1, obtained from the misspecified model (3.16), has a mean of β_1. Even though x_2 is an irrelevant variable (does not appear in the true equation) the estimate $\tilde{\beta}_1$ is an unbiased estimate.

Now the researcher has two different ways of estimating β_1, both of which yield unbiased estimates, and he can base his choice on the precision of each.

The variance of estimate $\hat{\beta}_1$ is given by (3.15), so we turn to the variance of estimate $\tilde{\beta}_1$:

$$V(\tilde{\beta}_1) = E[\tilde{\beta}_1 - E(\tilde{\beta}_1)]^2 \qquad (3.20)$$

$$= E\left[\frac{\Sigma x_2^2 . \Sigma x_1 \varepsilon - \Sigma x_1 x_2 . \Sigma x_2 \varepsilon}{\Sigma x_1^2 . \Sigma x_2^2 - \Sigma x_1 x_2 . \Sigma x_1 x_2}\right]^2 . \qquad (3.21)$$

Using the specification of errors (3.6) and (3.7), and evaluating the terms as in (3.10) through (3.15), equation (3.21) can be simplified to

$$V(\tilde{\beta}_1) = \sigma_\varepsilon^2 / [\Sigma x_1^2 (1 - r_{x1x2}^2)] . \qquad (3.25)$$

The variance of the estimate $\hat{\beta}_1$ does not depend on the correlation between the two variables. It depends on Σx_1^2 only.

The variance of the estimate $\tilde{\beta}_1$ depends on Σx_1^2 as well as on the correlation between variables x_1 and x_2. Since the square of the correlation coefficient is always a positive fraction (non-zero), the variance of $\tilde{\beta}_1$ is usually larger than the variance of $\hat{\beta}_1$.

However, when the correlation between the two independent variables (r_{x1x2}) is zero, both $\hat{\beta}_1$ and $\tilde{\beta}_1$ have the same variance. In this case both the estimates are identical, because zero correlation implies ($\Sigma x_1 x_2 = 0$). When the term ($\Sigma x_1 x_2$) is zero, equation (3.17) reduces to

$$\tilde{\beta}_1 = \Sigma x_1 y / \Sigma x_1^2 , \qquad (3.26)$$

which is the same as the estimate $\hat{\beta}_1$, given by equation (3.2).

Now let us turn to the distributional properties of $\tilde{\beta}_2$, the regression coefficient of the irrelevant variable. The ordinary least squares estimate $\tilde{\beta}_2$ is obtained as

$$\tilde{\beta}_2 = \frac{\Sigma x_1^2 . \Sigma x_2 y - \Sigma x_1 x_2 . \Sigma x_1 y}{\Sigma x_1^2 . \Sigma x_2^2 - \Sigma x_1 x_2 . \Sigma x_1 x_2} . \qquad (3.27)$$

By substitution of the true relation (3.1) for *y*,

$$\tilde{\beta}_2 = \frac{\Sigma x_1^2 . \Sigma x_2 \varepsilon - \Sigma x_1 x_2 . \Sigma x_1 \varepsilon}{\Sigma x_1^2 . \Sigma x_2^2 - \Sigma x_1 x_2 . \Sigma x_1 x_2} . \qquad (3.28)$$

The value of $\tilde{\beta}_2$ also depends on the error terms (ε 's). Since the error terms follow a statistical distribution, so does $\tilde{\beta}_2$. The mean value of the theoretical distribution of $\tilde{\beta}_2$ is

$$E(\tilde{\beta}_2) = 0 . \qquad (3.29)$$

The true value of the parameter β_2 is, however, zero, because for analytical purposes the true relation (3.1) may also be written as

$$y_t = \beta_1 x_{1t} + 0 . x_{2t} + \varepsilon_t . \qquad (3.30)$$

The regression coefficient of the irrelevant variable has a statistical distribution with mean zero which is its true parameter value; $\tilde{\beta}_2$ is, therefore, an unbiased estimate.

This result may be generalized to a case of several irrelevant variables by rewriting the truth as

$$y_t = \beta_1 x_{1t} + 0.x_{2t} + \ldots + 0.x_{kt} + \varepsilon_t , \tag{3.31}$$

where the independent variables x_2 through x_k are irrelevant variables. The estimates of regression coefficients of all the irrelevant variables (x_2 through x_k) have theoretical distributions with zero mean, which is their true value. The estimate of the regression coefficient of x_1 is still unbiased.

The variance of the distribution of $\tilde{\beta}_2$ is

$$V(\tilde{\beta}_2) = E[\tilde{\beta}_2 - E(\tilde{\beta}_2)]^2 . \tag{3.32}$$

Following the derivations in equations (3.21) through (3.25), we may derive the variance of $\tilde{\beta}_2$ in equation (3.16) as

$$V(\tilde{\beta}_2) = \sigma_\varepsilon^2 / [\Sigma x_2^2 (1 - r_{x1\,x2}^2)] . \tag{3.33}$$

Even though x_2 is an irrelevant variable according to the specification of the truth, its regression coefficient has nonzero variance. That is, when the regression is estimated the researcher may observe a nonzero value for $\tilde{\beta}_2$.

In the general case of a regression equation with *k* independent variables, whether they are relevant or not, the variance of $\tilde{\beta}_i$, is given by the *i*th diagonal element, and the covariance between $\tilde{\beta}_i$ and $\tilde{\beta}_j$; by the element in the *i*th column and *j*th row, in the following matrix:

$$\sigma_\varepsilon^2 \begin{bmatrix} \Sigma x_1^2 & \Sigma x_1 x_2 & \ldots & \Sigma x_1 x_k \\ \Sigma x_1 x_2 & \Sigma x_2^2 & \ldots & \Sigma x_2 x_k \\ . & . & \ldots & . \\ . & . & \ldots & . \\ . & . & \ldots & . \\ \Sigma x_1 x_k & \Sigma x_2 x_k & \ldots & \Sigma x_k^2 \end{bmatrix}^{-1}$$

3.2 Left-Out Variables

A similar situation frequently found in empirical research is one in which a parameter can be estimated from two different regression equations, one of which has a left-out variable. To analyze this case let us consider that the true relation is

$$y_t = \beta_1 x_{1t} + \beta_2 x_{2t} + \varepsilon_t \,, \tag{3.34}$$

where the error term (ε) follows the specification given by equations (3.5), (3.6), and (3.7).

When the researcher estimates the true relation (3.34), the estimate of β_1 by ordinary least squares is obtained as

$$\hat{\beta}_1 = \frac{\Sigma x_2^2 . \Sigma x_1 y - \Sigma x_1 x_2 . \Sigma x_2 y}{\Sigma x_1^2 . \Sigma x_2^2 - \Sigma x_1 x_2 . \Sigma x_1 x_2} \,. \tag{3.35}$$

By substituting the true relation (3.34) for y in equation (3.35) we obtain

$$\hat{\beta}_1 = \beta_1 + \frac{\Sigma x_2^2 . \Sigma x_1 \varepsilon - \Sigma x_1 x_2 . \Sigma x_2 \varepsilon}{\Sigma x_1^2 . \Sigma x_2^2 - \Sigma x_1 x_2 . \Sigma x_1 x_2} \,. \tag{3.36}$$

Since the error terms follow specification (3.5), it is easily seen that

$$E(\hat{\beta}_1) = \beta_1 \,. \tag{3.37}$$

The theoretical distribution of $\hat{\beta}_1$, has the mean β_1, the true parameter value.

The variance of the estimate $\hat{\beta}_1$ is

$$V(\hat{\beta}_1) = E[\hat{\beta}_1 - E(\hat{\beta}_1)]^2 \,. \tag{3.38}$$

The algebra involved in evaluating expression (3.38) is similar to that used in evaluating equation (3.20), hence

$$V(\hat{\beta}_1) = \sigma_\varepsilon^2 / [\Sigma x_1^2 (1 - r_{x1x2}^2)] \,. \tag{3.39}$$

The estimate $\hat{\beta}_1$ obtained by estimating the true relation (3.34) has a statistical distribution with mean β_1 and variance (3.39) when the error terms follow specifications (3.5), (3.6), and (3.7).

Suppose the researcher, by misspecification, estimates the following regression equation:

$$y_t \;=\; \tilde{\beta}_1 x_{1t} \;+\; e_t \,. \qquad (3.40)$$

The ordinary least squares estimate of β_1 in the misspecified model is obtained as

$$\tilde{\beta}_1 \;=\; \Sigma x_1 y / \Sigma x_1^2 \,. \qquad (3.41)$$

By substituting the true relation (3.34) for y in equation (3.41) we obtain

$$\tilde{\beta}_1 \;=\; \beta_1 + \beta_2 \Sigma x_1 x_2 / \Sigma x_1^2 + \Sigma x_1 \varepsilon / \Sigma x_1^2 \;\;. \qquad (3.42)$$

The mean value of the statistical distribution of $\tilde{\beta}_1$ when the errors follow specification (3.5) is

$$E(\tilde{\beta}_1) \;=\; \beta_1 + \beta_2 (\Sigma x_1 x_2 / \Sigma x_1^2) \;\;. \qquad (3.43)$$

Using the Yule notation, we may rewrite (3.43) as

$$E(\tilde{\beta}_1) \;=\; \beta_1 + \beta_2 b_{21} \;\;. \qquad (3.43a)$$

When the true relation is (3.34) and the errors are generated by specification (3.5), the researcher has two alternative estimators, namely $\hat{\beta}_1$, and $\tilde{\beta}_1$. The first is unbiased, whereas $\tilde{\beta}_1$ is biased as a result of the misspecification of the model. If the researcher is interested only in unbiased estimators, his choice is obvious. But if he wants to consider the precision of the estimates as well, then he needs the variance of the distribution of the estimate.

The variance of $\tilde{\beta}_1$ is

$$\begin{aligned} V(\tilde{\beta}_1) \;&=\; E[\tilde{\beta}_1 - E(\tilde{\beta}_1)]^2 \\ &=\; E\left[\beta_1 + \beta_2 b_{21} + \frac{\Sigma x_1 \varepsilon}{\Sigma x_1^2} - \beta_1 - \beta_2 b_{21}\right]^2 \\ &=\; E\left[\frac{\Sigma x_1 \varepsilon}{\Sigma x_1^2}\right]^2 \end{aligned}$$

The algebra involved in evaluating the expression $E[\Sigma x_1 \varepsilon / \Sigma x_1^2]^2$ is the same as in equations (3.10) through (3.15); hence

$$V(\tilde{\beta}_1) \;=\; \sigma_\varepsilon^2 / \Sigma x_1^2 \;\;. \qquad (3.45)$$

The variance of $\tilde{\beta}_1$ in equation (3.45) is therefore smaller than the variance of $\hat{\beta}_1$ in equation (3.39), even though the estimate $\tilde{\beta}_1$ corresponds to a misspecified model, (3.40). If the researcher can live with bias in the estimates and wants only estimates with smaller variance, then he will chose $\tilde{\beta}_1$ instead of $\hat{\beta}_1$ as an estimate of β_1.

In this case he has the option of choosing between a less efficient estimate of the true parameter β_1 and a more efficient estimate of a wrong (biased) value ($\beta_1+\beta_2.b_{21}$). If lack of bias is the prime criterion, he will choose one estimate; he will choose the other if precision takes precedence. The better choice is not obvious unless he has some additional criterion of selection.

Since there is a trade-off between bias and precision of the estimates, separate consideration of either may not be desirable. One rule that weighs both these aspects is the concept of "quadratic loss." Whenever an estimate of a parameter differs from its true value, a loss proportional to the square of the difference between the estimate value and the parameter value is associated with that value of the estimate. Since the loss is assessed proportional to the square of the difference, it is called the "quadratic loss." Thus defined, the quadratic loss has a statistical distribution because its value depends on the value of the estimate, which has a statistical distribution. The expected value of the distribution of quadratic loss is called the *mean quadratic loss*, or the *mean square error*. The estimator having the least mean quadratic loss tends to minimize the quadratic loss to a decision maker if he uses the same estimator in repeated trials.

The mean square error, or the mean quadratic loss, may be formally defined as

$$MSE(\hat{\theta}) \;=\; E(\hat{\theta}-\theta)^2\,, \qquad (3.46)$$

where θ is the parameter and $\hat{\theta}$ is an estimate of that parameter.

Mean square error can be expressed in terms of the variance and the bias of the estimate by first adding and then subtracting $E(\hat{\theta})$ in (3.46),

$$\begin{aligned} MSE(\hat{\theta}) \;&=\; E(\hat{\theta}-E(\hat{\theta})+E(\hat{\theta})-\theta)^2 && (3.47)\\ &=\; E[\hat{\theta}-E(\hat{\theta})]^2+[E(\hat{\theta})-\theta]^2 && (3.48)\end{aligned}$$

because the cross-product term has a zero expected value. Hence, the mean square error may be written as

$$MSE(\hat{\theta}) \;=\; Variance(\hat{\theta})+[Bias(\hat{\theta})]^2\,. \qquad (3.49)$$

Using the concept of mean square error, we can compare the two previous estimates of β_1, namely $\hat{\beta}_1$ and $\tilde{\beta}_1$. Since $\hat{\beta}_1$ is an unbiased estimate of β_1, the mean square error of $\hat{\beta}_1$ is merely the variance of the estimate itself:

$$MSE(\hat{\beta}_1) = \sigma_\varepsilon^2 / [\Sigma x_1^2 (1 - r_{x1x2}^2)] . \tag{3.50}$$

However, $\tilde{\beta}_1$ is biased, hence the mean square error is

$$MSE(\tilde{\beta}_1) = \sigma_\varepsilon^2 / \Sigma x_1^2 + \beta_2^2 . b_{21}^2 . \tag{3.51}$$

Although $\tilde{\beta}_1$ is obtained from a misspecified model, its mean square error can be smaller than that of $\hat{\beta}_1$. The condition under which this may occur may be obtained as

$$MSE(\hat{\beta}_1) > MSE(\tilde{\beta}_1) . \tag{3.52}$$

By using the corresponding expressions for the mean square errors,

$$\sigma_\varepsilon^2 / [\Sigma x_1^2 (1 - r_{x1x2}^2)] > \sigma_\varepsilon^2 / \Sigma x_1^2 + \beta_2^2 (\Sigma x_1 x_2 / \Sigma x_1^2)^2 , \tag{3.53}$$

$$\sigma_\varepsilon^2 (1 - r_{x1x2}^2) > \sigma_\varepsilon^2 + \beta_2^2 \Sigma x_2^2 . r_{x1x2}^2 , \tag{3.54}$$

$$\sigma_\varepsilon^2 \left(\frac{1}{1 - r_{x1x2}^2} - 1 \right) > \beta_2^2 \Sigma x_2^2 . r_{x1x2}^2 , \tag{3.55}$$

$$1 > \beta_2^2 / \left[\frac{\sigma_\varepsilon^2}{\Sigma x_2^2 (1 - r_{x1x2}^2)} \right] . \tag{3.56}$$

The expression

$$\frac{\sigma_\varepsilon^2}{\Sigma x_2^2 (1 - r_{x1x2}^2)}$$

is the theoretical variance of the estimate $\hat{\beta}_2$ in the true equation (3.34). By using this relationship we can rewrite the condition (3.56) for smaller MSE($\tilde{\beta}_1$) compared to MSE($\hat{\beta}_1$) as

$$1 > \beta_2^2 / V(\hat{\beta}_2) \tag{3.57}$$

or

$$1 > |\tau| , \tag{3.58}$$

where

$$\tau = \beta_2 / \sqrt{V(\hat{\beta}_2)} = \beta_2 \cdot \sqrt{\frac{\Sigma x_2^2 (1 - r_{x1x2}^2)}{\sigma_\varepsilon^2}} \ . \tag{3.59}$$

The researcher should note that τ is not the same as the *t*-ratio corresponding to $\hat{\beta}_2$. The *t*-ratio has estimates of the parameter and variance in the numerator and denominator, whereas τ has the theoretical values.

When τ is less than unity in magnitude the researcher can obtain a smaller mean square error for the estimate of β_1 by misspecifying the model rather than estimating the true regression equation. The value of τ is, however, based on the true values of the parameter and of the variance of the estimate, both of which are generally unknown. This result can be successfully employed only when the researcher has reason to believe that the parameter value cannot take on certain prespecified values. For example, it is usually assumed that the marginal propensity to consume is always a positive fraction, and so is elasticity of output with respect to labor in a production function. If the researcher believes that even the extreme value of this parameter cannot make τ exceed unity, then he can discard the variable if the only objective is to minimize the mean square error of the estimate of β_1.

Note that as the sample size increases the quantity Σx_2^2 also increases. Since this quantity appears in the numerator of equation (3.59), τ is an increasing function of sample size. With a large enough sample size the value of τ will exceed unity, therefore rule (3.58) will dictate that no variable should be deleted from the true equation. Thus we may state that the gain in the mean square error of an estimate is essentially a small sample property of the ordinary least squares estimates due to deletion of a variable for which $|\tau| < 1$.

The choice of an estimator in applied econometrics is necessarily dictated by many considerations beyond the textbook properties of best linear unbiased estimates. The researcher seldom, if ever, knows the true specification and frequently cannot afford to include all variables that might seem relevant in explaining the movements of a dependent variable. When he includes a variable which does not belong in the true specification, he is increasing the variance of the estimates without biasing them. When he discards a variable he may be biasing the estimates but gaining in precision. The net effect of precision and bias can be more, or less, than with estimating the true relation. How much bias a researcher is willing to accept or how much precision he can forego depends on the specific situation and on the seriousness of the consequences of the results; no general guidelines can be set up.

3.3 Serial Correlation in the Errors

In some econometric studies the specification that the error terms in each observation are drawn independently of other error terms may be inappropriate. For example, in cases of expectation models and distributed lag specifications, even though the errors in the theoretical model may be independent of each other, those in the estimated equation may not be. A frequent case of dependence in error terms occurs in the context of time series studies where the errors in one time period are dependent on errors in previous periods.

When error terms are serially dependent, ordinary least squares estimation does not yield the best linear unbiased estimates, even if the estimated equation is the truth. In order to study the distributional properties of ordinary least squares estimates in such a case, let us consider a simple specification in which the error terms are generated by a first order Markov scheme:

$$\varepsilon_t \ = \ \rho \, \varepsilon_{t-1} \ + \ \upsilon_t \, , \qquad (3.60)$$

where υ is assumed to be drawn at random from a distribution with mean zero and variance σ_υ^2 and is distributed independently of past values of the ε's.

By squaring equation (3.60) and taking the expected value we see that

$$V(\varepsilon) \ = \ \rho^2 V(\varepsilon) + V(\upsilon) \ , \qquad (3.61)$$

or

$$\sigma_\varepsilon^2 \ = \ \rho^2 \sigma_\varepsilon^2 + \sigma_\upsilon^2 \qquad (3.62)$$

$$\sigma_\varepsilon^2 \ = \ \sigma_\upsilon^2 / (1-\rho^2) \, . \qquad (3.63)$$

Given the value of the parameter ρ, the variance of ε is related to the variance of υ.

The specification of error terms in the case of a first order Markov scheme may be written as

$$E(\varepsilon_t) \ = \ 0 \, , \qquad (3.64)$$

$$E(\varepsilon_t \varepsilon_{t-s}) \ = \ \rho^s \sigma_\varepsilon^2 \ . \qquad (3.65)$$

Let us consider the simple case in which the truth is

$$y_t \ = \ \beta \, x_t \ + \ \varepsilon_t \ . \qquad (3.66)$$

The ordinary least squares estimation yields

$$E(\hat{\beta}) = \beta + E(\Sigma x_t \varepsilon_t / \Sigma x_t^2) . \qquad (3.67)$$

Since $E(\varepsilon_t) = 0$ and the x's are assumed to be fixed in repeated samples, the estimate $\hat{\beta}$ is still unbiased. The serial correlation in the error terms does not introduce any bias in the regression estimates as long as we retain specification (3.64).

However, when the error terms are serially correlated the precision of the estimate $\hat{\beta}$ depends on the serial correlation parameter as well as on the process generating the independent variable. Consider the variance of $\hat{\beta}$:

$$V(\hat{\beta}) = E[\hat{\beta} - E(\hat{\beta})]^2 \qquad (3.68)$$
$$= E[\Sigma x_t \varepsilon_t / \Sigma x_t^2]^2 \qquad (3.69)$$
$$= (1/\Sigma x_t^2)^2 . E[x_1^2 \varepsilon_1^2 + x_2^2 \varepsilon_2^2 + ... + 2. x_1 x_2 \varepsilon_1 \varepsilon_2 + ...] \qquad (3.70)$$

Since the x's are assumed to be fixed, using the specifications of the error terms (3.64) and (3.65) we obtain

$$V(\hat{\beta}) = (1/\Sigma x_t^2)^2 . \sigma_\varepsilon^2 . [\Sigma x_t^2 + 2 \Sigma \rho^s \Sigma x_t . x_{t+s}] . \qquad (3.71)$$

When the error terms are serially independent (ρ = 0), then the variance of the estimate depends only on Σx^2 and σ_ε^2; but when the errors are serially correlated the variance of $\hat{\beta}$ depends also on the terms $\Sigma x_t . x_{t+s}$.

To incorporate the information on the independent variable as well as the error term, let us consider a simple case in which the independent variable is also generated by a first order Markov scheme:

$$x_t = \lambda x_{t-1} + w_t , \qquad (3.72)$$

where *w*, is serially independent and has a statistical distribution with mean zero and variance σ_w^2. By assuming that the *w*'s are independent of past values of the *x*'s, we can use the following approximation to simplify the algebra:

$$\Sigma x_t x_{t+s} \simeq \lambda^s \Sigma x_t^2 . \qquad (3.73)$$

Since λ is a fraction λ^s converges to zero as *s* increases.

When the *x*'s are generated by a first order Markov scheme with parameter λ, expression

(3.71) may be written as

$$
\begin{aligned}
V(\hat{\beta}) &= (1/\Sigma x_t^2)^2 . \sigma_\varepsilon^2 . [\Sigma x_t^2 (1+2\Sigma \rho^s \lambda^s)] \\
&= (\sigma_\varepsilon^2 / \Sigma x_t^2)[1+2\rho\lambda(1+\rho\lambda+\rho^2\lambda^2+...)] \\
&= (\sigma_\varepsilon^2 / \Sigma x_t^2)[1+2\rho\lambda/(1-\rho\lambda)] \\
&= (\sigma_\varepsilon^2 / \Sigma x_t^2)(1+\rho\lambda)/(1-\rho\lambda) . \qquad (3.74)
\end{aligned}
$$

In this example the precision of the least squares estimate depends also on how the independent variable and the error terms are generated. The precision increases with the magnitude of ρ and λ when they are of the opposite sign and decreases when they are of the same sign.

When the error terms are serially correlated, the ordinary least squares estimate, though unbiased, does not have minimum variance. According to the well known Gauss-Markov theorem, the ordinary least squares estimate is the minimum-variance unbiased estimate only when the error terms are serially independent and have the same variance for all observations. If the researcher knows the value of the parameter, ρ, then, by a suitable linear transformation of the variables, he can reduce equation (3.66) to a form in which the ordinary least squares estimation provides such minimum-variance estimates.

To explain this procedure: when the parameter ρ is known, the following transformation on the error terms generates a new error term which is serially independent:

$$\varepsilon^* = \varepsilon_t - \rho\,\varepsilon_{t-1} \ . \qquad (3.75)$$

Since ε^* is nothing but υ in equation (3.60), which is by specification serially independent and has the same variance for all observations, it satisfies all requirements for ordinary least squares estimation to yield the minimum variance estimate. Consider equation (3.66) corresponding to time periods *t* and *t*-1.

$$y_t = \beta x_t + \varepsilon_t \ , \qquad (3.76)$$
$$y_{t-1} = \beta x_{t-1} + \varepsilon_{t-1} \ . \qquad (3.77)$$

By multiplying equation (3.77) by ρ and then subtracting it from equation (3.76) we obtain

$$(y_t - \rho\, y_{t-1}) = \beta(x_t - \rho\, x_{t-1}) + (\varepsilon_t - \rho\,\varepsilon_{t-1}) \ . \qquad (3.78)$$

By defining variables y^* and x^* as ($y_t - \rho\, y_{t-1}$) and ($x_t - \rho\, x_{t-1}$) respectively, we may rewrite equation (3.78) as

$$y_t^* = \beta x_t^* + \varepsilon_t^* . \qquad (3.79)$$

In this equation the error terms are serially independent and are drawn from statistical distributions with the same variance. However, the equation is based on only T - 1 observations, whereas there are T observations in all. The observation corresponding to y_1^* and x_1^*, is not defined because it involves observation corresponding to time period 0. This problem can be overcome by noting that the expected value of ε_1 is zero with variance $\sigma_v^2/(1 - \rho^2)$. By a suitable transformation on ε_1 we can obtain ε_1^*, which will have the same statistical properties as that of the other ε^* 's. Since (1- ρ^2) is a constant, the following transformation on ε_1 will provide the required ε_1^* :

$$\varepsilon_1^* = \sqrt{(1-\rho^2)}\varepsilon_1 . \qquad (3.80)$$

To obtain ε_1^* in the regression equation we have to transform the dependent and the independent variables as well. The y_1^* and x_1^* are therefore $\sqrt{1-\rho^2}\, y_1$ and $\sqrt{1-\rho^2}x_1$ respectively.

By including the first observation also in equation (3.79) we have *T* observations. The ordinary least squares estimation based on the *T* transformed variables yields the unbiased minimum variance estimate of β.

The estimation of parameter β from these transformed variables is called *generalized least squares*, to distinguish it from the estimation using untransformed variables in equation (3.66). This estimation procedure is very general and can be applied to a regression with several independent variables.

Consider the general case in which

$$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \ldots + \beta_k X_{kt} + \varepsilon_t , \qquad (3.81)$$

where the error term ε is generated by a first order Markov scheme with parameter ρ. The generalized least squares estimates of the β's, which are the minimum-variance unbiased estimates, are obtained by applying ordinary least squares estimation to the following regression equation in which all the variables, dependent and independent, are transformed:

$$Y_t^* = \beta_0^* + \beta_1 X_{1t}^* + \beta_2 X_{2t}^* + \ldots + \beta_k X_{kt}^* + \varepsilon_t^*. \qquad (3.82)$$

In (3.82) the transformed variables are obtained as

$$Y_t^* = Y_t - \rho Y_{t-1}, \qquad (3.83)$$
$$X_{it}^* = X_{it} - \rho X_{it-1}, \qquad (3.84)$$
$$Y_1^* = \sqrt{(1-\rho^2)} Y_1, \qquad (3.85)$$
$$X_{i1}^* = \sqrt{(1-\rho^2)} X_{i1} \qquad (3.86)$$

Although the generalized least squares estimates are minimum-variance unbiased estimates, they cannot be attained, because they involve the parameter value ρ which is rarely known to the researcher.

When the researcher suspects that the error terms are serially correlated and believes that the value is some specified value, say ρ^*, he can obtain the estimates from the transformed variables by treating the specified ρ^* as though it were the true parameter. When the true parameter value ρ is different from ρ^*, the estimates of the β's from the transformed variables are no longer the minimum-variance unbiased estimates.

Sometimes, even when the error terms were generated by a first order Markov scheme, the use of a value ρ^* different from ρ can yield estimates that are even less efficient than those yielded by ordinary least squares estimation from the untransformed variables.

To study the consequences of using the wrong value ρ^* for ρ in such a case, let us consider a situation having one independent variable (3.66), in which the errors were generated by the first order Markov scheme with parameter ρ, and the independent variable is also generated by a first order Markov scheme with parameter λ as in (3.72). To simplify the algebra, let us consider the case with only T-1 observations, in which the transformations on dependent and independent variables using a value ρ^* for ρ are

$$y_t^* = y_t - \rho^* y_{t-1} \qquad (3.87)$$
$$x_t^* = x_t - \rho^* x_{t-1} \qquad (3.88)$$

After transforming the variables as in (3.87) and (3.88), the researcher estimates the regression equation:

$$y_t^* = \hat{\beta}^* x_t^* + e_t, \qquad (3.89)$$

where

$$\hat{\beta}^* = \Sigma y_t^* x_t^* / \Sigma (x_t^*)^2. \qquad (3.90)$$

By expressing the transformed variables in the form of the original variables the researcher obtains

$$\hat{\beta}^* = \beta + \frac{\Sigma(x_t - \rho^* x_{t-1})(\varepsilon_t - \rho^* \varepsilon_{t-1})}{\Sigma(x_t - \rho^* x_{t-1})^2} . \tag{3.91}$$

Since the expected value of ε is zero, $\hat{\beta}^*$ is an unbiased estimate of β irrespective of the value of ρ^*. The estimate $\hat{\beta}^*$ has a theoretical distribution with mean β and variance given by

$$V(\hat{\beta}^*) = E\left[\frac{\Sigma(x_t - \rho^* x_{t-1})(\varepsilon_t - \rho^* \varepsilon_{t-1})}{\Sigma(x_t - \rho^* x_{t-1})^2}\right]^2 \tag{3.92}$$

$$= \frac{1}{[\Sigma(x_t - \rho^* x_{t-1})^2]^2} . \{\Sigma(x_t - \rho^* x_{t-1})^2 \; E(\varepsilon_t - \rho^* \varepsilon_{t-1})^2$$

$$+ 2\Sigma\Sigma(x_t - \rho^* x_{t-1})(x_{t+s} - \rho^* x_{t+s-1}) . E[(\varepsilon_t - \rho^* \varepsilon_{t-1})(\varepsilon_{t+s} - \rho^* \varepsilon_{t+s-1})]\} . \tag{3.93}$$

To simplify the algebra involved, we may use the following approximations:

$$\Sigma(x_t - \rho^* x_{t-1})^2 \simeq \Sigma x_t^2 (1 + \rho^{*2} - 2\rho^* \lambda) \tag{3.94}$$

$$\Sigma(x_t - \rho^* x_{t-1})(x_{t+s} - \rho^* x_{t+s-1}) \simeq \lambda^{s-1}(\lambda - \rho^*)(1 - \lambda\rho^*)\Sigma x_t^2 \tag{3.95}$$

$$E(\varepsilon_t - \rho^* \varepsilon_{t-1})^2 = \sigma_\varepsilon^2 (1 + \rho^{*2} - 2\rho^* \rho) \tag{3.96}$$

$$E[(\varepsilon_t - \rho^* \varepsilon_{t-1})(\varepsilon_{t+s} - \rho^* \varepsilon_{t+s-1})] = \rho^{s-1}(\rho - \rho^*)(1 - \rho\rho^*)\sigma_\varepsilon^2 . \tag{3.97}$$

By use of the expressions in (3.94) through (3.97), the variance of $\hat{\beta}^*$ may be simplified to

$$V(\hat{\beta}^*) \simeq \frac{\sigma_\varepsilon^2}{\Sigma x_t^2} . \frac{(1 + \rho^{*2} - 2\rho^* \lambda)(1 + \rho^{*2} - 2\rho^* \rho) + 2(\lambda - \rho^*)(\rho - \rho^*)(1 - \lambda\rho^*)...}{(1 + \rho^{*2} - 2\rho^* \lambda)^2} \tag{3.98}$$

Since $\Sigma\rho^{s-1}\lambda^{s-1} = (1 - \rho\lambda)^{-1}$ we can write (3.98) as

$$V(\hat{\beta}^*) \simeq \frac{\sigma_\varepsilon^2}{\Sigma x_t^2} . \left[\frac{1 + \rho^{*2} - 2\rho^* \rho}{1 + \rho^{*2} - 2\rho^* \lambda} + \frac{2(\lambda - \rho^*)(\rho - \rho^*)(1 - \lambda\rho^*)(1 - \rho\rho^*)}{(1 - \rho\lambda)(1 + \rho^{*2} - 2\rho^* \lambda)^2}\right] \tag{3.99}$$

Since ρ and λ are fractions, their higher powers converge to zero.

Though this expression for the variance of $\hat{\beta}^*$ is messy and probably incomprehensible to most readers, it can answer a few specific questions.

When ρ^* = 0 the estimate $\hat{\beta}^*$ is nothing but the ordinary least squares estimate based on T-1 observations. The variance (3.99) for ρ^* = 0 is the same as the expression (3.74) obtained previously for ordinary least squares.

The least minimum variance obtainable corresponds to the generalized least squares estimate when $\rho^* = \rho$. The expression for minimum variance is obtained by replacing ρ^* by ρ in (3.99) to obtain

$$V(\hat{\beta}^*)_\rho = \frac{\sigma_\varepsilon^2}{\Sigma x_t^2} \left[\frac{1-\rho^2}{1+\rho^2-2\rho\lambda}\right]. \qquad (3.100)$$

Similarly, the variance corresponding to first-difference estimates can be obtained by replacing ρ^* with 1.

The information contained in the expression for variance (3.99) can also be summarized by defining the relative efficiency of an estimate obtained through use of a value of ρ^* different from ρ as

$$EFF = V(\hat{\beta}^*)_\rho \;/\; V(\hat{\beta}^*)_{\rho^*} . \qquad (3.101)$$

Since the relative efficiency of the estimate depends on ρ^*, ρ, and λ, let us plot the relative efficiency with respect to ρ^* and λ for two arbitrarily selected values of ρ, namely ρ = 0.2 and 0.6. These plots are given in Figures 3.1 and 3.2.

When the ρ^* is different from the true value of the parameter ρ, the loss in precision of the estimation from the transformed variables can be extremely large. To take a few examples, let us consider the situation in which the true ρ = 0.2. The researcher suspects serial correlation in the errors but does not know the true value of ρ. Suppose he suspects the parameter to be “high” and selects ρ^* = 0.8. In this situation he has the following alternatives: (1) to stop with the ordinary least squares, (2) to estimate the parameters from the transformed variables using ρ^* = 0.8, or (3) to use first-difference estimates.

As can be seen from Figure 3.1, the relative efficiencies of these alternatives are different for different values of λ, the serial correlation in the independent variable. When the independent variable is “trending,” the parameter λ is large. Consider an economic series with λ = 0.8 as an independent variable. Of all the three estimates, ordinary least squares has maximum precision.

When researchers suspect high serial correlation in the errors they usually go ahead and estimate the parameters from the first-difference estimates.

Figures 3.1 and 3.2 show that a considerable amount of precision can be gained by resisting this temptation. The gain is substantial when the independent variable has "high" serial correlation, which is generally the case with economic time series data. Suppose the researcher suspects high serial correlation in the errors in the neighborhood of, say, 0.6 and instead of going to the first-difference estimates he uses an arbitrary value of ρ^* from the interval (0.4 - 0.99); he will then obtain estimates with higher precision than the first-difference estimates.

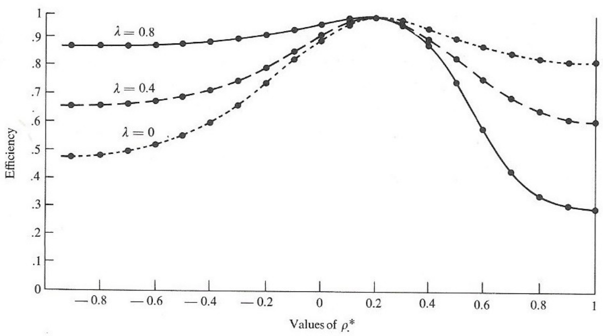

Figure 3.1. Efficiency of GLS for Various Values of ρ^* When True $\rho = 0.2$

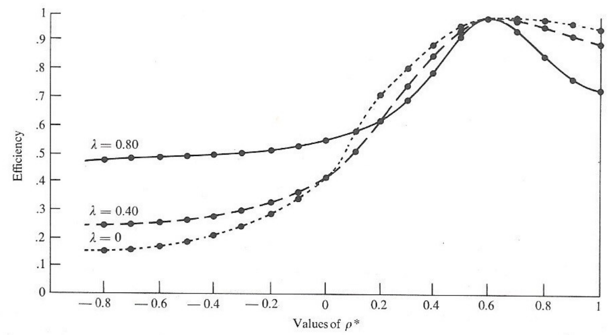

Figure 3.2. Efficiency of GLS for Various Values of ρ^* When True $\rho = 0.6$

In a practical situation, however, he rarely knows the parameter values. Unless the theory explicitly states it, there may be no strong reason to suspect any serial correlation in the error terms. Some researchers tend to blame serial correlation whenever their results are difficult to interpret, but such an attitude sometimes leads to serious consequences.

To illustrate the point, let us consider the case in which there is no serial correlation in the errors and the independent variables; that is, $\rho = \lambda = 0$.

In this case $\hat{\beta}^*$ is an unbiased estimate of β and the variance obtained from (3.99) is

$$V(\hat{\beta}^*) = \frac{\sigma_\varepsilon^2}{\Sigma x_t^2}\left[1 + \frac{\rho^{*2}}{(1+\rho^{*2})^2}\right]. \qquad (3.102)$$

The least variance is obtained when the arbitrary value ρ^* is equal to the true value of the parameter 0. (Remember, the true model has no serial correlation.) Whenever the value of ρ^* differs from zero, the variance is larger than the estimate using the true value of zero. When the true errors are, in fact, serially independent and the researcher suspecting serial correlation employs an arbitrary value to gain precision, he will actually be losing precision by so doing. Conversely, when the errors are serially correlated he may be able to increase the precision of the estimate (compared to the ordinary least squares approach) provided that he can obtain a “good” value of ρ^*. If the ρ^* value increases the variance of the estimate relative to ordinary least squares, then he may prefer to settle for the latter estimation procedure even though it does not provide the “best” estimate, for the alternative is even worse.

Typically the researcher obtains a ρ^* from the residuals of ordinary least squares estimation of (3.81) as

$$\rho^* = \Sigma e_t e_{t-1} / \Sigma e_{t-1}^2 \ . \qquad (3.103)$$

Estimate ρ^* is not the true value ρ and has a statistical distribution. It is consistent but generally biased in small samples, the bias being negative and of the order of (ρ/T) in magnitude. When the independent variables in the estimated regression are also serially correlated, then the bias depends also on the parameters that generated their serial correlation. In the present case, in which ρ and λ are the parameters of serial correlation in the errors and in the independent variable respectively, the bias in ρ^* is of the order of magnitude of [(ρ +λ))/T].

Estimate ρ^* as obtained in (3.103) has a variance. The variance of ρ^* is of the order of (1/7). For example, with a sample size of 49, the variance of ρ is approximately (1/49) and the standard deviation is approximately (1/7).

In the case of serially independent errors the two-sigma limits for ρ^* cover the range (-0.3 to +0.3). When the sample size is small, even though the true errors are serially independent, the chance is very high that ρ^* (obtained from the residuals in (3.103)) will show a sizable value. The researcher trying to improve the precision of his estimates should therefore be judicious in his selection of the arbitrary value for ρ^*.

3.4 Heteroscedasticity in Errors

Another situation in which the ordinary least squares method does not produce the best linear unbiased estimates occurs when the variance of the error term differs among various observations—that is, is non-constant. In such cases even though the errors in different observations are drawn at random, they are drawn from different distributions with zero means but different variances.

Consider the specification

$$y_t = \beta_1 x_{1t} + \beta_2 x_{2t} + \ldots + \beta_k x_{kt} + \varepsilon_t \ , \qquad (3.104)$$

$$E(\varepsilon_t) = 0 , \tag{3.105}$$
$$E(\varepsilon_t^2) = \sigma_{\varepsilon_t}^2 . \tag{3.106}$$

In this specification all the errors are drawn from statistical distributions with zero mean but having different variances as indicated by the subscript of σ_ε^2 in (3.106).

Estimation of equation (3.104) by ordinary least squares, therefore, does not yield the best linear unbiased estimates because of the lack of constant variance in the error term. However, if the researcher knows the individual variances, $\sigma_{\varepsilon t}^2$'s, then he can use the following transformation on the variables:

$$y_t^/ = y_t / \sigma_{\varepsilon_t} , \tag{3.107}$$
$$x_t^/ = x_t / \sigma_{\varepsilon_t} . \tag{3.108}$$

Instead of estimating equation (3.104), he may then estimate the following:

$$y_t^/ = \beta_1 x_{1t}^/ + \beta_2 x_{2t}^/ + \ldots + \beta_k x_{kt}^/ + \varepsilon_t^/ . \tag{3.109}$$

This is a reduced form of equation (3.104), obtained by dividing each of the observations by the corresponding standard deviation of its error term.

Hence, the error term in (3.109) is

$$\varepsilon_t^/ = \varepsilon_t / \sigma_{\varepsilon_t} \tag{3.110}$$

By assumption (3.106), the variance of the error term ε_t is $\sigma_{\varepsilon t}^2$; hence,

$$V(\varepsilon_t^/) = \sigma_{\varepsilon_t}^2 / \sigma_{\varepsilon_t}^2 = 1 . \tag{3.111}$$

That is, the transforming of (3.104) as indicated yields a regression equation with constant variance.

In the transformation version the error terms are serially independent and have the same variance, satisfying the assumptions under which the ordinary least squares estimation is best. Estimation of equation (3.109) yields the best linear unbiased estimates for the β's.

This technique has little relevance to empirical work because the researcher rarely knows the variances of the error terms, except that in some situations he may know the variance up to a constant of proportionality. That is, he may believe that the error term has a variance proportional to a quantifiable variable.

Consider, for example, the following linear curve of crude oil requirements for refineries:

$$Y_t \ = \ \beta_0 \ + \ \beta_1 X_{1t} \ + \ \beta_2 X_{2t} \ + \ \beta_3 X_{3t} \ + \ \varepsilon_t \ , \qquad (3.112)$$

where *Y* is the crude oil and *X1*, *X2* , and *X3* are gasoline, kerosene, and fuel oil respectively.

The refineries are of different sizes, and small refineries may perhaps be expected to exhibit small variance in the error term and large refineries to exhibit larger variance, even though the requirement function is assumed to be the same for all refineries. The nature of the variance of the error term may be specified as

$$V(\varepsilon_t) \ = \ X_{4t}^2 . k \ , \qquad (3.113)$$

where *k* is a constant of proportionality and X_{4t}, is the capacity of the *t*th refinery.

In this case the transformation of the variables is of the form

$$Y_t^{/} \ = \ Y_t / X_{4t} \ , \qquad (3.114)$$

$$X_{\mathrm{it}}^{/} \ = \ X_{\mathrm{it}} / X_{4t} \qquad i = 1,2,3 \qquad (3.115)$$

This transformation adjusts the equation in such a way that its reduced form satisfies all conditions for the least squares estimation to yield best linear unbiased estimates. The data on the capacity of the refineries are available, hence equation (3.109) can be estimated.

The estimated equation is

$$(Y_t / X_{4t}) \ = \ \hat{\beta} + \hat{\beta}_0 (1/X_{4t}) + \hat{\beta}_1 (X_{1t}/X_{4t}) + \hat{\beta}_2 (X_{2t}/X_{4t}) + \hat{\beta}_3 (X_{3t}/X_{4t}) + e_t \ \ , \qquad (3.116)$$

where the constant term (β) is introduced even though it is not present in equation (3.112) to make the summary statistics meaningful.

Another frequent example occurs when the data come from a published source in which the agency compiling information reports aggregate results rather than individual observations. For example, the Bureau of the Census reports only the aggregate income of all families in given geographic localities.

If the researcher believes that the variance of the error term is the same to each individual, then the variance of error in the aggregate corresponding to each locality cannot be the same unless the number of individuals in each is the same. When data on the number of individuals in each aggregate are also given, which is generally true in such reports, the researcher can use that information in transforming the variables to improve the precision of his estimates.