ESTIMATION IN ERRORS-IN-VARIABLES MODELS

MANORANJAN PAL

A thesis submitted to the Indian Statistical Institute, Calcutta, in partial fulfilment of the requirements for the award of the degree of Doctor of Philosophy

1981

PREFACE AND ACKNOWLEDGEMENTS

This thesis deals with the problem of estimation of regression parameters in single equation Errors-in-Variables Models (EVM's). Although the literature on EVM's is quite extensive, the available methods of estimation (viz., Maximum Likelihood Method, Instrumental Variable Method etc.) suffer from serious shortcomings.

The structure of the thesis is as follows :

Chapter 1 makes a critical survey of the different assumptions made in the literature on the distribution of errors and reviews the different methods of estimation suggested so far for different types of EVM's.

Chapter 2 proposes some new moment/cumulant-based estimators for the slope parameter and compares their efficiencies *vis-a-vis* OLS estimator assuming lognormality of the true regressor.

In Chapter 3 optimum three-group slope-estimators are found for different combinations of parameter-values in the standard two-variable EVM where the true regressor is assumed to follow lognormal or gamma distribution. Separate examination is made for the case where the disturbances are homoscedastic and the case of heteroseedastic disturbances.

Chapter 4 and 5 are concerned with estimation in more general EVM's, where the classical EVM is extended in the following directions :

(i) The standard deviation of the error term associated with the regressor may vary with the level of the regressor.

(ii) The Errors-in-Variables (EIV's) in the regressor and the regressand may be correlated.

Chapter 4 suggests methods of estimating the parameters of the distribution of error term in the regressor. Chapter 5 makes use of these results and proposes consistent moment estimators for regressions where the algebraic form of the regression is non-linear. Actually, only two such forms are analysed — the semilog and the hyperbola — besides linear regressions, but the method proposed can handle situations where the true values X and Y are related as $Y = \alpha + \beta f(X) + \varepsilon$.

The last chapter (Chapter 6) contains some rather isolated results. It makes some comments on an estimator recently proposed by Kaila, and then discusses how one can tackle the standard two-variable EVM (i) when both the error variances are known and (ii) when there are more than one IV's available.

My greatest debt of gratitude is to my supervisor Professor N. Bhattacharya of the Indian Statistical Institute who guided me throughout the period of my research. But for his help and encouragement from the beginning, it would not ~~be~~ have been possible for me to bring the thesis to its present form. Thanks are also due to Professors J. K. Ghosh, S. K. Chakrabarti, D. Coondoo, R. Mukherjee, M. Chaudhuri (now with the Vishwabharati University) and many other faculty members in the Economic Research Unit of the Indian Statistical Institute who made very useful suggestions and comments at various stages of the work. I specially thank my fellow research workers Mr. M. Bhaumik and Mr. S. Sil who have permitted me to include some of our joint work in this thesis. I have also benefitted from occassional discussions with Messrs. B. Bagchi, S. Chakravarty, N. Sarkar and P. P. Majumder. The work was done with the support of a Research Fellowship of the Indian Statistical Institute. I am also grateful to the staff of the Computer Science Unit for granting me liberal access to the computer and for other assistance. Finally, I thank Mr. K. Neogi of Economic Research Unit for excellent typing and Mr. L. M. Das for rolling the stencils.

CONTENTS

Page

Chapter 0

ABSTRACT

In many econometric investigations, the 'errors-in-variables' (EIV's) are not negligible (Morgenstern, 1963). Examination of 25 series relating to national accounts by Langaskens and Rijckeghem(1974) showed that the standard deviations of the errors ranged from 5 to 77 per cent of the average value of the corresponding variable. Such errors may vitiate 'least-squares' (LS) estimation of regression coefficients (Johnston, 1972). The well-known methods (ML; IV, including grouping method) proposed for handling classical 'EIV model' (EVM) in regression analysis suffer from serious limitations. Some of them make strong distributional assumptions about the errors (and the regressors) and/or assume prior knowledge about the values of the error variances; others need auxiliary variables called 'instrumental-variables' (IV's) which are supposed to be uncorrelated with the error terms, but strongly correlated with the true regressors. The IV's are thus not always handy and, in any case, one can never check the assumptions.

This thesis attempts to find consistent and reasonably efficient estimators of parameters of a variety of EIV models. The classical linear two-variable EVM forms the basis of several investigations and receives considerable attention. This model is specified as

$$Y_i = \alpha + \beta X_i + \varepsilon_i, \quad i = 1, 2, \ldots, n \qquad \ldots (0.1)$$

where α and β are parameters to be estimated; ε_i is the disturbance term distributed normally with mean zero and variance σ_ε^2 for all i, and X_i and Y_i are non-observable true values of the regressor and the

regres and respectively. The ε's are assumed to be independent of X's where X is stochastic. The observed values x_i and y_i are written as

$$\left.\begin{aligned} x_i &= X_i + u_i \\ y_i &= Y_i + v_i \end{aligned}\right\} \qquad \ldots \qquad (0.2)$$

where u_i and v_i are the EIV's which are independent of each other and of the true values X_i and Y_i. u_i's and v_i's have means zero and variances σ_u^2 and σ_v^2 respectively for all i. For $i = 1, 2, \ldots, n$, we assume that $(X_i, Y_i, u_i, v_i, \varepsilon_i)$ are i.i.d. random variables.1/ The main interest centres on estimating β; once β is estimated α can be estimated very easily. Writing

$$\begin{aligned} y_i &= \alpha + \beta X_i + \varepsilon_i + v_i \\ &= \alpha + \beta x_i + \varepsilon_i + v_i - \beta u_i \\ &= \alpha + \beta x_i + w_i \quad \text{(say)} \qquad \ldots \qquad (0.3) \end{aligned}$$

one finds that 'ordinary least squares' (OLS) regression of y on x gives an inconsistent estimator of β essentially because $cov(x_i, w_i) \neq 0$ (Johnston, 1972, p.282). Extension of this model to more than one regressor is obvious. Other important extensions allow u_i and v_i to be correlated or the distribution of u_i to depend on the value of X_i.

Various alternative methods of estimation have been suggested by previous researchers. These are based on different sets of assumptions. Thus, some assume X to be stochastic while others do not. Chapter 1 makes a critical survey of the different assumptions made in the literature on the distribution of errors and of the regressor X

1/ Not all the assumptions are needed for every result.

and reviews the different methods of estimation suggested so far. There are, of course, some models which can not be fully identified at all (vide Section 1.6 of Chapter 1; see also Appendix 4.2 of Chapter 4). It may be mentioned here that some good review articles on EVM's already exist in the literature (Durbin, 1954; Madansky, 1959; Cochran, 1968; Moran, 1971; Pal, 1980a). Among other things, this chapter discusses how one can obtain consistent estimators of θ if (i) one has prior knowledge about the value of the error variances or of their ratio or if (ii) IV's are available. Introduction of lagged values of regressors/regressand may also be helpful in finding consistent estimates of the parameters. Sometimes in the laboratory experiments repeated measurements are available for the same value of the variable. This may help in finding consistent estimates. The problem becomes more difficult if instead of one relation we have many relations in the model[2/], but the variables are affected by EIV's. Apart from economists, sociologists have long been applying such simultaneous equations models in path analysis and multiple indicator analysis. But the progress does not seem to have been satisfactory at all. Chapter 1 also indicates the results obtained by Bayesian econometricians who have tried to find satisfactory answers to this problem. The chapter concludes with brief observations on applications of EVM's to different fields like consumption analysis, geology, management science etc.

Chapter 2 considers the classical two-variable linear EVM specified above, where the true regressor (X) is known to be non-normal. Actually, all that is really needed is that the third order cumulant of

2/ These are known as simultaneous equations models in econometrics.

X is non-zero. We examine the possibilities of consistent moment estimation based on uni- and bi-variate moments or cumulants of third or higher order. If only the sample moments of first and second order are used for estimation, five relations are obtained for seven unknown parameters, viz., α, β, σ_u^2, σ_v^2, σ_ε^2, $\mu_1'(X)$ and $\mu_2'(X)$. However, σ_v^2 and σ_ε^2 always appear in the form of $\sigma_v^2 + \sigma_\varepsilon^2$, so that in effect we have the following five equations for six unknown parameters :

$$m_1'(x) = \mu_1'(X) \qquad \ldots (0.4)$$

$$m_1'(y) = \alpha + \beta\mu_1'(X) \qquad \ldots (0.5)$$

$$m_2'(x) = \mu_2'(X) + \sigma_u^2 \qquad \ldots (0.6)$$

$$m_2'(y) = \alpha + 2\,\alpha\beta\mu_1'(X) + \beta^2\mu_2'(X) + (\sigma_v^2 + \sigma_\varepsilon^2) \qquad \ldots (0.7)$$

$$m_{11}'(x,y) = \alpha\mu_1'(X) + \beta\mu_2'(X) \qquad \ldots (0.8)$$

One may, however, set up similar equations based on third order moments and estimate β from them. Actually one has four such equations :

$$m_3(x) = \mu_3(X) \qquad \ldots (0.9)$$

$$m_{21}(x,y) = \beta\mu_3(X) \qquad \ldots (0.10)$$

$$m_{12}(x,y) = \beta^2\mu_3(X) \qquad \ldots (0.11)$$

$$m_3(y) = \beta^3\mu_3(X) \qquad \ldots (0.12)$$

so that there is a multiplicity of estimators of β. Some of the obvious estimators are given below[3/] :

$$\hat\beta_1 = m_{03}/m_{12}, \quad \hat\beta_2 = m_{12}/m_{21}, \quad \hat\beta_3 = m_{21}/m_{30}, \quad \hat\beta_4 = \sqrt[3]{m_{03}/m_{30}},$$
$$\hat\beta_5 = \pm\sqrt{m_{03}/m_{21}}, \quad \hat\beta_6 = \pm\sqrt{m_{12}/m_{30}},$$

3/ $\hat\beta_3$ was suggested by Durbin (1954), $\hat\beta_4$ was suggested by Drion (1951) and the estimator proposed by Scott (1950) is a function of $\hat\beta_1$, $\hat\beta_2$ and $\hat\beta_3$.

where $m_{ij} = m_{ij}(x,y) = \frac{1}{n} \sum_{k=1}^{n} (x_k - \bar{x})^i \; (y_k - \bar{y})^j$. Signs of $\hat{\beta}_5$ and $\hat{\beta}_6$ may be based on any one of other four estimators. Assuming $\beta \neq 0$, the estimators $\hat{\beta}_1$ to $\hat{\beta}_6$ are consistent if $\mu_3(X) \neq 0$. All the estimators based only on third order moments are functions of $\hat{\beta}_1$, $\hat{\beta}_2$ and $\hat{\beta}_3$, which may be called the 'basic' estimators.

Asymptotic variances of the six estimators of β noted above have been calculated in Chapter 2 and were compared. Efficiencies of these estimators relative to OLS estimator have also been investigated. For this latter examination, X has been assumed to follow a lognormal distribution, which is realistic in some applications, e.g., in engel curve analysis in many countries. The relative efficiencies depend on c_u $(= \sigma_u^2 / \sigma_X^2)$ and $c_{\varepsilon'}$ $(= (\sigma_v^2 + \sigma_\varepsilon^2) / (\beta^2 \sigma_X^2))$ and regions are demarcated in the $(c_u, c_{\varepsilon'})$ - plane where the different estimators happen to be best. The estimators have been found to be fairly efficient even when OLS is valid. It should be mentioned here that the OLS estimator is consistent only when the error in the regressor is absent. The estimators are computationally simple and need milder assumptions than maximum likelihood (ML) or IV estimators.

Consistent estimators of the following type were suggested by Geary (1942) :

$$\hat{\beta} = \hat{K}'(c_1, c_2 + 1) / \hat{K}'(c_1 + 1, c_2) \qquad \ldots \quad (0.13)$$

where c_1 and c_2 are positive integers and $K'(c_1, c_2)$ is the bivariate cumulant of (x, y) of order (c_1, c_2). This estimator is consistent if $K(c_1 + 1, c_2) \neq 0$, where $K(c_1, c_2)$ is the bivariate cumulant of (X,Y) of order (c_1, c_2). Some further estimators of this type are proposed in Chapter 2. These, however, make additional assumptions e.g., u or v is symmetrically distributed. Estimators based on cumulants of fourth or higher order would be useful if the X-distribution is symmetric, for then estimates based on third order moments would fail. The concluding sections of this chapter extend these ideas to the case of $m > 1$ regressors and briefly mention the case where u and v are correlated.

For the standard two-variable EVM, Wald (1940) proposed grouping estimators of the regression coefficients in which the observed values (x_i, y_i), $i = 1, 2, \ldots, n$, are divided into two equal groups according to the rank of the x_i's, and the centres of gravities of the two groups in the scatter diagram are then joined by a straight line to find the slope estimator. Later Bartlett (1949) suggested the use of three groups with equal number of observations in each group according to the order of x_i's. Here, the centres of gravities of the two extreme groups are joined by a straight line to estimate the parameters. Suppose the group means in the two extreme groups are $(\bar{x}_1, \bar{y}_1)$ and $(\bar{x}_3, \bar{y}_3)$ respectively. The grouping estimator of β is then defined by

$$\hat{b} = \frac{\bar{y}_3 - \bar{y}_1}{\bar{x}_3 - \bar{x}_1} \qquad \ldots \quad (0.14)$$

It is not necessary to take equal number of observations in each group. This choice is, however, optimal in the Gauss-Markov set-up if the X's are equispaced. In general, there is a problem of optimal choice of the three groups if one decides to use an estimator of the form of $\hat{b}$. Theil and Van Yzeren (1956) have obtained the optimum group proportions for different distributions of X. The optimum group proportions turned out to be approximately 0.3: 0.4: 0.3 for a variety of distributions considered by them.[4/] The prevailing opinion seems to be that the three-group estimator of Bartlett (with equal groups) is nearly optimum in almost all cases. The conclusion is, however, based on inadequate amount of investigation. The distribution of X examined so far (mainly by Theil and Van Yzeren) are mostly symmetric or negatively skewed. In most empirical applications in economics the X-distribution is highly positively skewed. Hence the rule specified by Bartlett and Theil and Van Yzeren may not be applicable in all cases.

In Chapter 3 we find the optimum group proportions assuming that X follows (i) the lognormal distribution and (ii) the gamma distribution. The optimum group proportions are those for which $V(\hat{b})$ is minimized. The most important finding is that the optimum proportions in the three group (in ascending order of x) are quite stable around the values (0.40, 0.45, 0.15) for the commonly occuring lognormal or gamma type distributions of X. The estimators would be highly efficient (about 80 per cent) relative to OLS if the group proportions are near the optimum values. Further, the gain in efficiency appears to be considerable if one used the optimum group proportions instead of equal groups as in the common Bartlett estimator.

A similar investigation was carried out for the case where the disturbances are heteroscedastic. Actually, we studied the set-up where

$$V(\varepsilon_i \mid X_i) = \lambda X_i^p \qquad \ldots \quad (0.15)$$

4/ For the case where errors in X-values are absent, the Bartlett estimator for three equal groups had been proposed by Nair and Shrivastava (1942), who considered equi-spaced X-values. Nair and Banerjee (1943) later showed that this remains efficient even if errors are present.

as considered by Lancaster (1968), who examined the efficiency of Wald's and Bartlett's (equal-groups) estimators vis-a-vis GLS estimator. In Section 3.3 of Chapter 3, we find the optimum group proportions in this set-up by minimizing the variance of the three-group slope estimator assuming X to follow lognormal and gamma distributions with realistic sets of parameter values. Here also the optimum estimators are found to be highly efficient vis-a-vis 'generalized least squares' (GLS) estimator which is BLUE. Their efficiency is about 80 per cent, and further, the increase in efficiency over Bartlett's estimator is quite considerable. Unlike the homoscedastic case, the optimum group proportions in the heteroscedastic case are highly dependent on p, the degree of heteroscedasticity. As p increases, the first group proportion (lowest X-values) decreases and the third group proportion (highest X-values) increases. (But the sum of the two extreme group proportions is fairly stable.) This means that the choice of optimum group proportions should be made in the light of some approximate idea regarding p. In other respects, the optimum group proportions seem to be nearly stable with respect to the types of distributions or parameters. The efficiencies of the grouping estimators in most of the cases decrease as the coefficient of variation of X increases.

The grouping estimators have some obvious advantages. One little known point is that in the heteroscedastic case mentioned above the variance of the GLS estimator contains terms like $E(X^{-p})$ which may not exist in all cases.

Chapter 4 and 5 are concerned with estimation in more general EVM's, where the classical EVM is extended in the following directions :

(i) The conditional distribution of the error term u give X may vary with the regressor X. To be precise, we stipulate that $V(u|X)$ is proportional to some power of X.

(ii) The errors u_i and v_i may be correlated. Symbolically,

$$\left.\begin{array}{l}\text{(i) } V(u_i|X_i) = a^2 X_i^b \\ \text{(ii) } v_i = \lambda u_i + w_i\end{array}\right\}, \quad i = 1, 2, \ldots, n \qquad \ldots (0.16)$$

where w_i's are i.i.d. random variables independent of u_i.[5/]

In Chapter 4, the main attempt is to estimate parameters of the distribution of X and of u given X, on the basis of observations $x_1, x_2, \ldots, x_n$, making plausible assumptions about the form of these distributions. The techniques used are ML and method of moments. The first three moment equations are

$$m_1' = E(x) = E(X) \qquad \ldots (0.17)$$

$$m_2' = E(x^2) = E(X^2) + a^2 E(X^b) \qquad \ldots (0.18)$$

$$m_3' = E(x^3) = E(X^3) + 3a^2 E(X^{b+1}) \qquad \ldots (0.19)$$

where $m_r' = \frac{1}{n} \Sigma x_i^r$. Cases where b = 0 or 2 are of particular interest. Note that b = 0 implies that u and X are independent while b = 2 implies that the conditional standard deviation of u given X is proportional to X (Cf. Friedman, 1957, p.27)[6/]. For specific values of b, the equations

5/ These problems clearly arise in the analysis of household budget data where both income or total consumer expenditure (x) and item consumption (y) are affected by transitory (seasonal) variation.

6/ Interestingly, if the range of X is from zero to infinity and if the observed variable x is always positive then b can be proved to be equal to 2.

(0.17) to (0.19) have four unknown parameters viz., $E(X)$, $E(X^2)$, $E(X^3)$ and a^2. Hence, if the distribution of X contains only two parameters then the set of equations (0.17) to (0.19) may be solved for three unknowns. As an application of this idea, we assumed X to follow a two-parameter lognormal distribution and found conditions for existence of feasible solution of these equations for specific values of b. Attempts are also made to apply this technique to a situation where X is three-parameter lognormal. In this case, however, one has to use the fourth order moment also.

Assumptions about the forms of the conditional distribution of u given X may also help in the estimation of parameters by method of moments. In the 'methods' suggested here, the range of this conditional distribution is assumed to be finite and so restricted that the observed variable x can not be negative. To achieve this, u give X is assumed to follow a Pearsonian type II distribution which is in many other respects similar to the normal distribution. An important advantage of this set-up is that it allows the use of fractional moments like $E(x^{.5})$, $E(x^{1.5})$ or even moments like $E(\log x)$, so that one need not go to higher order sample moments which are known to have larger sampling errors (Madansky, 1959; Geary, 1942). The distribution of X has been assumed to be two-parameter lognormal. Six sets of moment estimators are proposed for this set-up in Chapter 4. Monte Carlo experiments have been conducted to compare the efficiencies of these estimators. It has been found that estimates based on moments of the lowest order are the best and that the sampling errors of estimates increase with the order of the moments used. Of course, one can use ML method for estimation of parameters in this case. The Monte Carlo experiments, although limited in scope, suggest that one of the moment estimators of (μ, σ^2, m) is nearly as efficient as ML estimator.7/ The method of moments, it should be noted, is obviously far more expeditious than the ML method.

Appendix 4.2 to Chapter 4 cites an EIV model which happens to be non-identifiable.

Chapter 5 tackles the bivariate EVM's with generalizations (0.16) and makes use of the results of Chapter 4. It also considers different algebraic froms of the regression equations, viz.,

(i) Linear : $Y = \alpha + \beta X + \varepsilon$ ... (0.20)

(ii) Hyperbola: $Y = \alpha + \beta / X + \varepsilon$ ... (0.21)

(iii) Semilog : $Y = \alpha + \beta \log X + \varepsilon$ ... (0.22)

In handling these curve-forms, X has been assumed to follow a two-parameter lognormal distribution. If the regression equation is

$$Y = \alpha + \beta f(X) + \varepsilon \qquad \text{... (0.23)}$$

then one may use the following moment equations for estimation of parameters :

$$m_{11} = \beta \operatorname{cov}(X, f(X)) + \lambda a^2 E(X^b) \qquad \text{... (0.24)}$$

$$m_{21} = \beta \operatorname{cov}(X^2, f(X)) - 2\beta E(X) \operatorname{cov}(X, f(X)) + a^2 \beta \operatorname{cov}(X^b, f(X)) + 2\lambda a^2 \operatorname{cov}(X, X^b) \qquad \text{... (0.25)}$$

7/ Here μ and σ^2 are the parameters of lognormal distribution, i.e., $X \sim \Lambda(\mu, \sigma^2)$ and m is the parameter of the conditional density of u given X : $f(u|X) = \text{const.}(1-u^2/X^2)^m$, $-X \leq u \leq X$.

which has the solution

$$\hat{\beta} = \frac{2m_{11}\ \widehat{cov}\ (X,\ X^b) - m_{21}\ \hat{E}\ (X^b)}{2\ \widehat{cov}\ (X,f(X))\ \widehat{cov}\ (X,X^b) - \hat{E}\ (X^b)\ \{\ \widehat{cov}\ (X^b,\ f(X)) - 2\ \hat{E}\ (X)\ \widehat{cov}\ (X,f(X)) + \hat{a}^2\ \widehat{cov}\ (X^b,\ f(X))\ \}} \quad \ldots (0.26)$$

Putting $f(X) = X$, $1/X$ or $\log X$, one gets the estimators for three forms mentioned above. Analogous results should be easy to obtain for other types of distributions of X.

Distributional assumption on X is not always necessary. In the linear set-up (equation (0.20)), with b unknown, the following polynomial

$$\frac{1}{n}\ \Sigma\ \{(y_i - \bar{y}) - d(x_i - \bar{x})\ \}^3 = 0 \quad \ldots (0.27)$$

has a solution $\hat{d}$ which is a consistent estimator of β (Section 5.3, Chapter 5) . This is what Scott (1950) proved, though in a very restrictive set-up.

The problem of identification arises quite often in econometric work. A model can be under-identified, just-identified or over-identified depending on prior information available to the analyst. The classical linear two-variable EVM is known to be under-identified if X is normal. That is why one needs additional information on the error variances σ_u^2 and σ_v^2, or on auxiliary variables called 'instrumental variables' (IV's), to tackle such problems. However, too many IV's again make a model over-identified creating troubles to the researcher.

The last Chapter (Chapter 6) discusses, how one can tackle the standard two-variable EVM in three different situations marked by over- or under-identification. In the first case, no additional information is known, the model being under-identified, since X is normal.

Recently Kaila (1980) proposed an estimator of the slope coefficient for this model. On examination, this estimator turns out to be the geometric mean of the LS regression coefficient of y on x and the reciprocal of the LS regression coefficient of x on y. It is shown that this estimator due to Kaila is optimal for a particular value of $\lambda = \sigma_u^2 / \sigma_v^2$ while for other values of λ, it is less efficient than estimates which could be obtained if λ were known.

The next case considered is where both σ_u^2 and $\sigma_{v'}^2 (= \sigma_v^2 + \sigma_\varepsilon^2)$ are known a priori. Here one has two estimators namely

$$\hat{\beta}_1 = m_{11} / (m_{20} - \sigma_u^2)$$

and $\hat{\beta}_2 = (m_{02} - \sigma_{v'}^2)/m_{11}$

both of which are consistent. One may then define a pooled estimator as

$$\hat{\beta}_a = a\,\hat{\beta}_1 + (1 - a)\,\hat{\beta}_2 , \qquad \ldots (0.28)$$

and minimize the variance of $\hat{\beta}_a$ to find the optimum value of 'a' (vide Section 6.3 of Chapter 6). Since this is an over-identified model, one may as well generalize it by assuming cov.$(u, v') \neq 0$ (here $v' = v + \varepsilon$) and find the ML estimator of β to be the geometric mean of $\hat{\beta}_1$ and $\hat{\beta}_2$ (Madansky, 1959). Comparison between the optimum pooled estimator $\hat{\beta}_0$ (say) and the geometric mean of $\hat{\beta}_1$ and $\hat{\beta}_2$ ($\hat{\beta}_M$, say) does not lead to one clear-cut answer. When the correlation coefficient between u and v' (ρ, say) is zero, $\hat{\beta}_0$ is obviously better. As $|\rho|$ increases from zero to 1, $\hat{\beta}_M$ gradually becomes superior. When $\rho \neq 0$, $\hat{\beta}_0$ is inconsistent, but the asymptotic bias is very small. The difference between the asymptotic variances is also very small.

Finally, we consider a case with too many IV's, each of which is uncorrelated with the error terms, u, v and ε. Then all the estimators

$$\hat{\beta}_j = \frac{\Sigma y z_j}{\Sigma x z_j}, \qquad j = 1, 2, \ldots, K, \qquad \ldots (0.29)$$

where z_j's are IV's, are consistent for β. We consider the pooled estimator

$$\hat{\beta}(a) = \Sigma a_j \hat{\beta}_j \qquad \ldots (0.30)$$

If we minimize the asymptotic variance of $\hat{\beta}(a)$ we arrive at the optimum estimator

$$\hat{\beta}_{01} = \frac{e' \hat{W}^{-1} \hat{\beta}}{e' \hat{W}^{-1} e}, \qquad \ldots (0.31)$$

where $\hat{W} = ((\hat{w}_{ij}))$,

$$\hat{w}_{ij} = \hat{\sigma}_{ij} / (\hat{\sigma}_{x_i} \hat{\sigma}_{x_j}),$$

$$\hat{\sigma}_{ij} = \frac{1}{n} \Sigma z_i z_j,$$

$$\hat{\sigma}_{xj} = \frac{1}{n} \Sigma x z_j,$$

and e is a K x 1 column vector consisting of one's only. All the variables are measured from their respective means.

One may alternatively consider the following class of consistent estimators :

$$\hat{\beta}(c) = \frac{\Sigma y (\Sigma_j c_j z_j')}{\Sigma x (\Sigma_j c_j z_j')}, \qquad \ldots. (0.32)$$

where c_j, j = 1, 2, ..., K, are real constants so that $\hat{\beta}(c)$ is defined, and z'_j's are obtained from z_j's by taking linear transformations such that

(i) $\sum_i z'_{ji} z'_{1i} = 0,$ for $j \neq 1$,

and

(ii) $\sum_i z'^2_{ji} = 1$, for all j.

The optimum estimator in this class is again obtained by minimizing the asymptotic variance. It is found that the optimum estimator thus obtained is the same as the estimator defined in (0.31). That is, the two approaches mentioned here lead to the same consistent estimator. Further investigations show that the common estimator coincides with Theil's (1958) 2SLS estimator.

Chapter 1

INTRODUCTION AND SURVEY OF LITERATURE

1.1 Introduction

"An error is normally viewed to be an expression of imperfection and of incompleteness in description" (Morgenstern, 1963, p. 13). It arises whenever measurable quantities differ from the theoretical counterparts. Hence it may be defined as the difference between an observed (or measured) quantity and the 'true' value of it. But it may not always be easy to define what is a 'true' value of an object. According to Haavelmo (1944, p. 7):

> "The 'true' variables are variables such that, if their behaviour should contradict a theory, the theory could be rejected as false; while 'observational' variables, when contradicting the theory, leave the possibility that we might be trying out the theory on facts for which the theory was not meant to hold, the confusion being caused by the use of same names for quantities that are actually different".

In many investigations we can never get rid of error of measurement. There will always be some imperfections in measuring quantities, however small they may be. In natural sciences or wherever controlled experiements are made, it may sometimes be possible to observe things as accurately as they are needed with the help of modern sophisticated instruments. But even there there is a limit to the accuracy of measurements. Thus, it is not possible to measure both the momentum p of a particle and its position x precisely at the same time, since

$$\Delta x . \Delta p \sim h,$$

where h is Planck's constant. This is known as the " Heisenberg Uncertainty Relation" (see Mittelstaedt, 1976, p. 91). In the social sciences it is even more difficult to have precise measurements. Nowadays it has been widely recognized " ..... that there cannot be absolute accuracy, that there must be error, and that the important thing to do is to try to uncover, remove, or at least limit the error" (Morgenstern, 1963, p.12).

1.2 Classification of Errors and Their Causes

Errors may be divided into two broad classes, viz.,

(i) Errors-in-Variables (EIV) : These occur when the data or observations on variables are not perfect or accurate in the sense that they differ from the 'true' or intended value. They are also sometimes called errors of observations or measurement errors.

(ii) Errors-in-the-Model : These are due to lack of fit of the model even if 'true' measurements are used. Often EIV's are assumed to be small compared to errors in the model, but there are situations where such assumptions may be hazardous.

We will mainly consider errors creeping into statistical data used in economic analysis. We can classify these EIV's into two categories, namely :

(i) Sampling Errors : These errors occur due to the estimation of population parameters based on samples from a population.

(ii) Non-sampling Errors : Non-sampling errors are those which take place while we are collecting and processing the data. These are the errors which are not covered by sampling errors. Complete enumeration may give less accurate results than the sampling approach because non-sampling errors may be much larger in the former than in the latter.

Anyone who is dealing with economic data must be aware of the extent of errors in the observations. Naturally, to have a good idea about the amount of errors that affect the observations it is necessary to know the causes and different sources of errors.

Morgenstern (1963) cited four main sources of errors creeping into different stages :

(i) Any economic model is an attempt towards explaining 'reality' But reality is too complicated to be explained by a mathematically tractable model. This failure of the model forces us to introduce errors in the equations.

(ii) Even if we are sure that error due to (i) is absent and there is no error in the calculations, the parameters derived from observations will be affected with errors due to the errors in the observations or EIV's in the measured or observed values of variables.

(iii) Convergent and limiting processes have to be broken off at some point. Such errors due to approximation when accumulated may cause serious damage to the ultimate results.

(iv) All computing devices and digital machines are either afflicted with 'noise' variables or with round off errors. The elementary operations affected with these errors, when numerous, may pose a threat to the accuracy of the ultimate results.

The above classification made by Morgenstern is based on the presumption of finding sources of errors in the final results. But what we are now interested in is to find the sources or rather causes of EIV's.

Starting from the stage of planning of the data collection project, EIV's may be classified stage by stage as follows :

(i) Errors in the Planning Stage : Setting up vague questions and using incorrect definition of variables leading to faulty classifications, lack of training for investigators etc. are main sources of errors introduced at this stage.

(ii) Errors in the Collecting Stage : These may be divided into two parts :

(a) Errors due to Observer : These may be due to insufficient understanding of the concepts and definitions on the part of the investigator. Bias of interviewer may also affect responses.

(b) Errors due to Observant : Informants often conceal informations due to fear or dislike of such disclosure. Response errors may also be due to recall lapse and lack of understanding of the question etc.

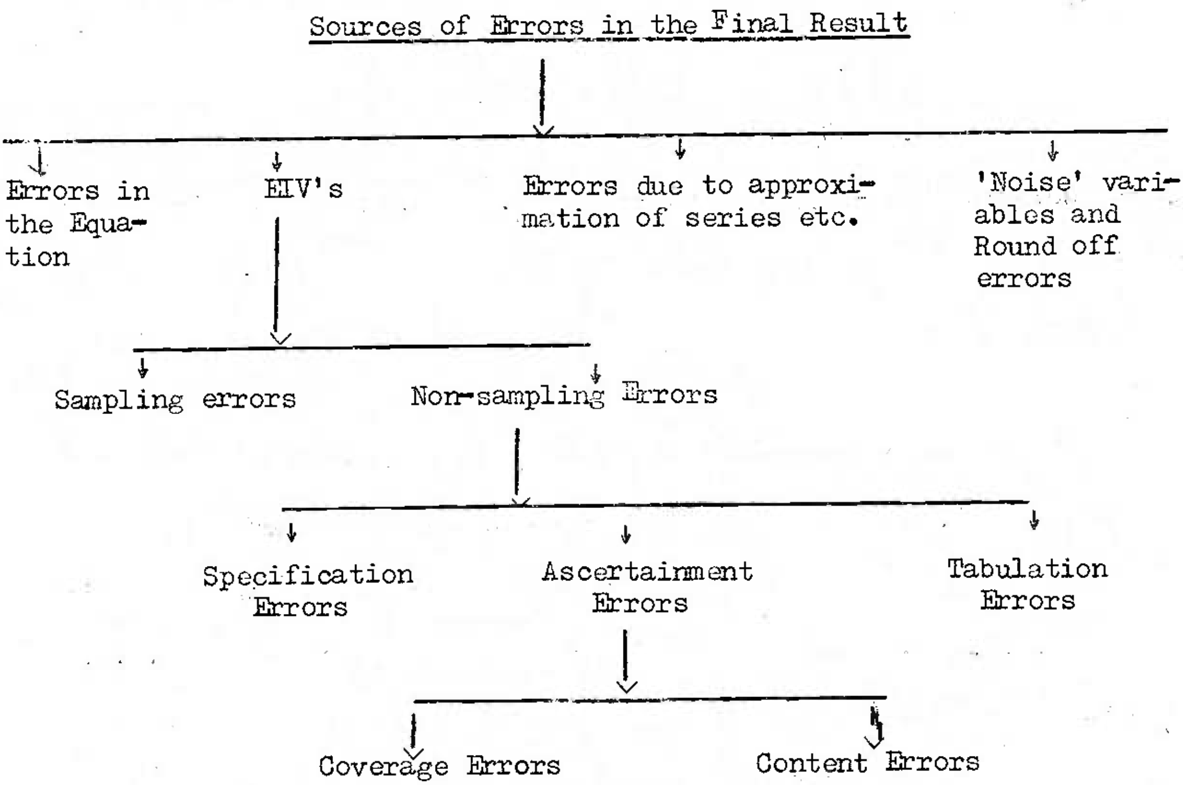
(iii) Errors in the Processing Stage : These comprise errors due to faulty aggregation, omission and duplication, approximation, mistakes in scrutiny of schedules, tabulations etc.

Murthy (1967) named these three categories as (a) specification errors, (b) ascertainment errors, and (c) tabulation errors corresponding to the three stages of census or survey work. He further subdivided ascertainment errors into two parts, namely

(i) Errors due to over-or under-enumeration of the population or sample, duplication and omission of units, and non-response (refusals etc.). These he termed as <u>coverage errors</u>.

(ii) Wrong entries due to errors on the part of investigators and respondents give rise to the errors termed as <u>content errors</u>.

Diagrammatically :



1.3 Empirical Evidence on EIV's

It is essential for any econometrician dealing with data to be aware of the extent and behaviour of possible errors in his analysis of data. This type of study is not only helpful for him to get his results but also necessary in designing programs for the collection of new, improved data.

Without knowledge of possible errors, it may sometimes be impossible to draw any conclusion from the final result. The following striking example is due to Milne (1949, p.30). The system of equations

$$x - y = 1$$

$$x - 1.00001y = 0$$

has the solution x = 100001, y = 100000, while the almost identical system of equations

$$x - y = 1$$

$$x - 0.99999y = 0$$

has the solution x = - 99999, y = - 100000. A slight change in the coefficient of y produces a dramatic change in the solution. Such things may happen to a greater or less extent in all cases of near-exact multicollinearity.

An important study on EIV's was made by Langaskens and Rijckeghem(1974). They had earlier (1967) compared two series of national accounts estimates published by two independent Belgian institutes, D.U.L.B.E.A. and the National Institute of Statistics (N.I.S.). The period of overlap between these competing series was relatively short (1954 to 1960) andit was not possible to have any statistical estimation on the basis of these data. Later, a new opportunity to estimate the variances of the error terms arose when the N.I.S. revised its estimates for the period 1953-1965 entirely, in order to ensure compatibility with the 1965 input-output table. Fortunately enough, the period of overlap in this case was 13 years to enable them to estimate error variances with higher accuracy.

They first assumed that the error terms in the two series were independent of their true components as well as between themselves. They assumed the following model

$$\left.\begin{array}{l} Y_t = \Psi_t + u_t \\ \text{and} \quad Y'_t = \Psi_t + v_t \end{array}\right| \quad t = 1953, \ldots, 1965, \qquad \ldots (1.1)$$

where Y_t and Y'_t are observed valuesof the two series for year t, Ψ_t the true component and u_t and v_t the error components. The variables are assumed to be serially independent. Moreover the expectations of error terms are zero, i.e.,

$$E(u_t) = E(v_t) = 0, \quad \forall t . \qquad \ldots (1.2)$$

Then,

$$\sigma_Y^2 = \sigma_\Psi^2 + \sigma_u^2 \qquad \ldots (1.3)$$

$$\sigma_{Y'}^2 = \sigma_\Psi^2 + \sigma_v^2 \qquad \ldots (1.4)$$

and $$\sigma_{Y-Y'}^2 = \sigma_{u-v}^2 = \sigma_u^2 + \sigma_v^2 . \qquad \ldots (1.5)$$

Hence,

$$\sigma_u^2 = \frac{\sigma_{Y-Y'}^2 + \sigma_Y^2 - \sigma_{Y'}^2}{2} \qquad \ldots (1.6)$$

and

$$\sigma_v^2 = \frac{\sigma_{Y-Y'}^2 + \sigma_{Y'}^2 - \sigma_Y^2}{2} . \qquad \ldots (1.7)$$

A total of 25 different economic series were analysed by them. These series related to variables like Current Expenditures of Enterprises, Current Expenditures of Households, Income from Property and Entrepreneurship, Current Government Expenditure etc. The study provided information regarding the relative importance of measurement errors in the different national account series. They expressed the standard deviations of the error terms as percentage of the average value of the variable to which they correspond. The percentage ranges from 5 to 77. The following table gives a brief summary :

Table 1.1

Error Variance as Percentage of Mean	Number of cases	
	Old series	New series
(1)	(2)	(3)
0 - 20	9	9
20 - 40	9	8
40 - 60	5	6
60 - 80	2	2
Total	25	25

The authors concluded : "Our results indicate that a number of variables in our national accounts are unfit to be used for economic purposes".

It may be remarked here that for a number of years we have two series of National Income estimates for India — the conventional series and the revised series — and a comparison of the two could be extremely revealing. The divergences are quite large and the assumption that $E(u_t - v_t) = 0$ does not seem to be appropriate (Mukherjee, 1969).

Other studies on National Account Statistics also confirmed that estimates of income and product are rarely equal. The discrepancies between them could be studied furtherif sufficient data are available. In

1955 an analysis of this "statistical discrepancy" was published by Gartaganis and Goldberger (1955). The main conclusions about the character of the "statistical discrepancy" were as follows :

(i) The hypothesis of normality was not rejected at the 5 per cent level of significance for any of the series.

(ii) The discrepancy did not indicate (by the Mann-Kendall test) any time trend, though the GNP increased over the sample period.

(iii) First order autocorrelation was not significant for most of the series. The hypothesis of temporal independence was not rejected by Wallis-Moore text.

De Janosi (1961) found no cyclical pattern in the "statistical discrepancy" in a more recent revision of the same data. Adams and De Janosi (1966) later calculated the bias in estimation of regression coefficient in the savings income relationship due to inclusion of this discrepancy in the components of GNP. Murray (1972) also examined the classical assumptions concerning errors in the data.

Numerous studies have been made on "response errors" in survey data collection. In some studies the response errors were seen to be seriously biased while in some studies bias was not found to be appreciable (Keating, Peterson and Stone, 1950; Mosel and Cozan, 1952; NBER, 1958; Lansing, Ginsberg and Braaten, 1961; Ito, 1964; Ferber, 1965; Borus, 1966; Som, 1973; etc.)

1.4 The Classical "Errors-in-Variables Model" (EVM)

In the EIV Models some of the variables must contain some error. These variables have two components, one is the 'true' value and the other, the error. The general two-variable EVM is described by the relation

$$Y = f(X) + \varepsilon \qquad \ldots (1.8)$$

These X and Y are true components of the observed variables x and y respectively, i.e.,

$$x = X + u \qquad \ldots (1.9)$$

$$y = Y + v \qquad \ldots (1.10)$$

where u and v are error components variously called Errors-in-Variables, Errors-in-Observations, Measurement Errors or Observational Errors. X, Y, u and v are non-observable where x and y are observable. ε is the Error-in-the-Equation.

f(X) in (1.8) is a function of X and involves parameters say, α, β etc., which are to be estimated. The distribution of ε may also involve unknown parameters. Statistical estimation of these parameters requires random samples of (x,y) observations from a population where the system of relations is assumed to be valid. Suppose n <u>independent</u> observation pairs $(x_1, y_1), \ldots, (x_n, y_n)$ are drawn from this population. We may then rewrite (1.8) to (1.10) as

$$Y_i = f(X_i) + \varepsilon_i \qquad \ldots (1.11)$$

and $$x_i = X_i + u_i \qquad \ldots (1.12)$$

$$y_i = Y_i + v_i \qquad \ldots (1.13)$$

where i runs from 1 to n. The following assumptions are usually made :-

(i) ε_i's are iid $N(0, \sigma^2_\varepsilon)$ and independent of X_i, u_i and v_i.

(ii) u_i's are iid $N(0, \sigma^2_u)$ independent of v_i, X_i. ...(1.14)

(iii) v_i's are iid $N(0, \sigma^2_v)$ independent of X_i

X_i's are iid.

X may sometimes be assumed to be non-stochastic. The situation is then referred to as classical regression situation with EIV (Lindley, 1947; Kendall, 1951). It is essential that $\varepsilon \neq 0$ for this model. Since there is a great deal of confusion in the literature about the terminology of EIV models (Lindley, 1947; Kendall, 1951, 1952; Moran, 1971), we shall stick to the terminology set out in the following table.

	X stochastic	X non-stochastic
$\varepsilon = 0$	Structural Equation Model	Functional Equation Model
$\varepsilon \neq 0$	Stochastic Regression Equation Model	Non-stochastic Regression Equation Model

Assuming linearity of f(X), the structural or functional relation

$$Y = \alpha + \beta X$$

in the EVM, becomes

$$y = \alpha + \beta x + w$$

where $w = v - \beta u$. An exact relation ($\varepsilon = 0$) between the true variables thus gives rise to a regression relation of the observed variables.

1.5 Assumptions on Errors

The basic duty of an econometrician is to explain the economic reality by means of models involving variables and errors with their behavioural assumptions. But the distribution of errors is different for different situations. Among the three types of errors — Systematic Errors, Extreme or Chaotic Errors and Random Errors — as termed by Velikanov (1965), only Random Errors can be described by stochastic models.

It is often maintained that random errors are symmetrically distributed about zero on both sides. This assumption is made in derivation of noraml distribution in the theory of errors. Thus the assumption of normal distribution of errors is widely used in statistical analysis. In the literature numerous theoretical proofs have been given to explain the wide occurrence of approximately normal distributions of errors. Gnedenko and Kolmogorov (1954); Linnik (1954) etc., however, obtained a form of characteristic functions for errors, which only under restricted conditions gives Gauss' normal distribution.

It is usually assumed in the literature that the size of discrepancy (strictly its variance) is independent of the magnitude of the quantity measured. In economic data it may not always be so.

Friedman (1957) proposed to treat the consumer unit's measured income (y) as the sum of the two components : a permanent component (say y_p), and a transitory component (say y_t); i.e.,

$$y = y_p + y_t.$$

The permanent component, according to Friedman, reflects the effect of those factors determining its capital value or wealth, the nonhuman wealth it owns etc. The transitory component reflects all other factors which may be treated as 'accidental' or 'chance' occurrences. Similarly, he divided the consumer expenditure into two parts :

$$c = c_p + c_t$$

His PIH (Permanent Income Hypothesis) is described by the following three equations :

$$c_p = Ky_p \quad \ldots (1.15)$$

and $$y = y_p + y_t \quad \ldots (1.16)$$

$$c = c_p + c_t \quad \ldots (1.17)$$

where K is a function of the rate of interest, the relative importance of property and non-property income, and the factor determining the consumer unit's tastes and preferences for consumption versus additions to wealth.

He also assumed that

$$\rho_{y_t y_p} = \rho_{c_t c_p} = \rho_{y_t c_t} = 0 \quad \ldots (1.18)$$

where ρ stands for the correlation coefficient. He then argued that the first two assumptions in the equation (1.18) are very mild and highly plausible. To quote from Friedman (op. cit., p. 27) :

> "For a group of individuals, it is plausible to suppose that the absolute size of the transitory component varies with the the size of the permanent component : that a given random event produces the same percentage rather than the same absolute increase or decrease in the income of units with different permanent components. .... it is not, however, inconsistent with zero correlation. ...."

In a different context McIntyre and others (1966) also found the variance in isotope-dilution measurements of Rb^{87}/Sr^{86} in geological samples to be proportional in general to $(Rb^{87}/Sr^{86})^2$. It is important to note that Friedman recognized the possibility that $\rho_{y_t c_t} \angle 0$. This complication is usually assumed away in the literature.

The following model may represent the above phenomenon :

$$x = X + u \qquad \ldots (1.19)$$

with $E(u|X) = 0$ ... (1.20)

and $E(u^2|X) = a^2X^2$. ... (1.21)

We may then write

$$x = X(1 + \frac{u}{X})$$

or $\log x = \log X + \log (1 + \frac{u}{X})$

or $x' = X' + u'$ (say) . ... (1.22)

X' and u' may be taken to be independent for our purpose, because the first two moments are

$$E(\frac{u}{X}) = 0$$

and $V(\frac{u}{X}) = a^2$

which do not depend on X. Hence u' may be taken as independent of the level of the series. Also, log x is homoscedastic though x is not.

It is interesting to compare this with the wellknown Box-Cox transformation suggested in LS regression analysis. Box-Cox (1964) suggested the transformation

$$x^{(\lambda)} = \begin{cases} (x^\lambda - 1)/\lambda, & \text{if } \lambda \neq 0 \\ \log x, & \text{if } \lambda = 0 \end{cases} \qquad \ldots (1.23)$$

in the regressor and/or in the regressand. There is one advantage in taking this type of transformation over simple power transformation x^λ, since it is continuous at $\lambda=0$, because

$$\lim_{\lambda \to 0} (x^\lambda - 1)/\lambda = \log x . \qquad \ldots (1.24)$$

Earlier Tukey (1957) had suggested that transformation of data might make the model more nearly linear, the errors more homoscedastic and more nearly normal. In the bivariate regression analysis if we confine ourselves to the more practically interesting cases of $\lambda=1$ and $\lambda=0$, we end-up with the two functional forms :

$$y = \beta_0 + \beta_1 x + u_1 \qquad \ldots (1.25)$$

$$\log y = \alpha_0 + \alpha_1 \log x + u_2 \qquad \ldots (1.26)$$

One can always choose between these two models by method suggested by Box-Cox.

Katona et al (1954) found "that the standard deviation of savings within each of several income classes is proportional to the average income of each class" (p.203). Lancaster (1968) in his study on company-profit-dividend relationship found the log-variance of dividends and the log-mean profit within each group to be linearly related. An appropriate model for such situations is

$$\left.\begin{aligned} & Y = \alpha + \beta X + \varepsilon \\ \text{where} \quad & E(\varepsilon|X) = 0 \\ \text{and} \quad & V(Y|X) = V(\varepsilon|X) = a^2 b^2 \end{aligned}\right\} \qquad \ldots (1.27)$$

For Katona's example b is obviously 2. Lancaster in his study found b to be approximately equal to 1.5.

These discussions on transformations were really concerned with the errors in the equation, but we can adapt them to include errors in variables. In the errors in variables case one may have

$$\left.\begin{aligned} & E(u|X) = 0 \\ \text{and} \quad & E(u^2|X) = a^2 X^b \end{aligned}\right\} \qquad \ldots (1.28)$$

which is more general assumption than (1.20) and (1.21) stated earlier.

1.6 The Problem of Identification in the Errors-in-Variables Models (EVM's)

It is wellknown that the unobservable Errors-in-Variables (EIV's) are unavoidable factors so far as any statistical data in social, biological or physical sciences are concerned. Some of the statistical problems created by EIV's were recognized during the 19th century. Among the mathematicians who tried their hands on this problem are Adcock (1978); Kummel (1879) and Merriman (1890) (see Roos, 1937). Since then much effort has been made in this particular area, but the progress does not seem to be satisfactory enough. There are mainly three reasons for this slow rate of progress. First, the problem itself is very much complicated. Maravall (1979, p.1) says : "Part of the unpopularity of EVM's was undoubtedly due to the identification problems that unobservable variables could create". Secondly, there are other complications in the regression problem of econometrics which are believed to be more serious, and since very few techniques have been developed for handling more than one complication. So errors-in-variables problems have usually been neglected. The third factor has been the lack of knowledge about the nature and the extent of errors primarily because of the gap among policy-makers, theoreticians and data collecting agencies. To quote Cochran and Rubin (1973) :"This is a difficult field as evidenced by prolonged struggles of the econometricians to find satisfactory methods for coping with such errors in their investigations of relationships between variables, and by the slow rate of progress that has rewarded major efforts to study errors of measurement in sample surveys". Lack of standardization of notations and terminology and hence vague model specification have also created some problems. In this section we shall mainly discuss the problems of identification in the Errors-in-Variables Models (EVM's).

Suppose we wish to estimate a regression equation $Y = \alpha + \beta X + \varepsilon$, from the observed values y and x only, making some reasonable assumptions on true and error components and on their inter-relationship. But, can we always consistently estimate them ? In general, the answer to this question is in the negative, unless the assumptions are chosen carefully. Before going into the details of this problem, let us have a concrete definition of identifiability.

Wald (1950) defined identifiability, but his concept corresponds to what was called multiple identifiability by Koopmans (1950). Though Wald got some conditions of identifiability of parameters, they are stated very generally and cannot be easily varified in any given case. In the present study we adopt the definition stated by Reiersol (1950). In defining identifiability we need the concept of 'structure'. A 'model' becomes a 'structure' if all parameters and all distributions in the model are numerically specified. "A structure is thus a particular realization of the model, and a model is the set of all structures compatible with the given specifications" (Reiersol, 1950). Two structures will be called "observationally equivalent" if both of them lead to the same probability distribution of the observed variables. A parameter is called 'identifiable' if it has the same value in all observationally equivalent structures. In other words : a set of parameters is said to be 'non-identifiable' if there are more than one set of parameters which can give rise to the same probability distribution of the observed random variables. For example, if

$$x = X + u$$

and if we assume that $X \sim N(\mu_1, \sigma_1^2)$ and $u \sim N(0, \sigma_2^2)$ where $\mu_1, \sigma_1^2, \sigma_2^2$ are unknown and X and u are independent of each other then only from the observed values of x we cannot identify σ_1^2 and σ_2^2 separately, i.e., they are non-identifiable. Because $x \sim N(\mu_1, \sigma_1^2 + \sigma_2^2)$ and there are infinitely many sets of parameters $\{(\sigma_1^2, \sigma_2^2) \mid \sigma_1^2 + \sigma_2^2 = \text{constant}\}$ which give rise to the same distribution of x. But μ_1 is identifiable. On the whole, the model is non-identifiable. (A model is said to be identifiable iff all the parameters are identifiable.) In fact, it can be proved that : If X and u are independent and u is normally distributed with mean zero then the model $x = X + u$ is identifiable iff X has no 'normal component'. We say a variable X has a 'normal component' if X can be written as $X = X' + u'$ where X' and u' are independent and u' is normally distributed.

There are wellknown problems of identification in the bivariate EVM's. Madansky (1959) gave an interesting example of such non-identifiability. He considered the usual linear EVM with no disturbance term, i.e.,

$$\left.\begin{aligned} Y &= \alpha + \beta X \\ y &= Y + v \\ x &= X + u \\ E(u) &= E(v) = 0 \end{aligned}\right\} \quad \ldots (1.29)$$

The errors u and v are independent of each other and of true values. Successive observations on X, u and v's are assumed to be independent.

$V(X) = \sigma_X^2$, $V(u) = \sigma_u^2$, $V(v) = \sigma_v^2$, $V(x) = \sigma_x^2$, $V(y) = \sigma_y^2$ and $V(Y) = \sigma_Y^2$. Finally, X, u and v are normally distributed. In the above model the two sets of parameters

σ_X^2	σ_u^2	σ_v^2	β	α
$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	1	$\upsilon-\mu$
$\frac{1}{3}$	$\frac{2}{3}$	$\frac{1}{4}$	$\frac{3}{2}$	$\upsilon-\frac{3}{2}\mu$

lead to the same joint distribution of x and y, namely a bivariate normal distribution with $E(x) = \mu$, $E(y) = \upsilon$, $\sigma_x^2 = \sigma_y^2 = 1$ and $\rho(x, y) = \frac{1}{2}$.

Obviously, a non-identifiable parameter cannot have a consistent estimator and, conversely, if there exists a consistent estimator for a parameter then that parameter is identifiable. In the above model (1.29) it has been proved that α and β are non-identifiable iff X and Y are constants or jointly normally distributed (Reiersol, 1950, see also Koopmans and Reiersol, 1950). When a model is identifiable consistent estimates of the parameters exist. Kiefer and Wolfowitz (1956) proved that in identifiable cases, the ML estimates of the regression parameters are in fact strongly consistent, i.e., with probability one they converge to the true parameters as n, the sample size, approaches infinity. But it is not always easy to solve ML equations.

Other papers written before Reiersol on the identification problem are those by Gini (1921), Koopmans (1937), Geary (1942, 1943, 1949), Tintner (1944, 1945, 1946) etc. Koopmans had actually shown that β is identifiable if the vector X, instead of being normally distributed, can take only two values. But we can easily see that it is a direct corollary to Reiersol's theorem.

One reason why the problem of identification arises is undoubtedly the contemporaneity of the model. Once the model contains dynamic features the results obtained for contemporaneous models do not hold (Geraci, 1977a, p.107). "Indeed, from the results reported by Hurwicz in 1949 until the recent attention, as evidenced in Hsiao (1976b, 1977), Nowak (1977), Maravall (1974) and Maravall and Aigner (1977), little interest was paid to the problem of identification in dynamic EIV models" (Maravall, 1979). Dynamic models with EIV, however, were not unknown to engineers and statisticians (see, for example Whittle, 1963, and Sorenson, 1966; see also Maravall, 1979, for further references).

In the general model discussed by Maravall (1979) some of the exogenous variables may be assumed to be autocorrelated. Assume that the number of exogenous variables that are independently distributed white-noise variables is n_1, and $0 \leq n_1 < n$. The model is described by :

a : $G_p(L)\,\eta_t = \sum_{i=1}^{n} B_{q_i}^{(i)}(L)\ \xi_t^{(i)} + u_t$

b : $R_r(L)u_t = S_s(L)a_t$

c : $M_{m_i}^{(i)}(L)\ \xi_t^{(i)} = H_{h_i}^{(i)}(L)\ b_t^{(i)}, \qquad i = n_1+1, n_1+2, \ldots, n.$

d : $x_t^{(i)} = \xi_t^{(i)} + \delta_t^{(i)}, \qquad i=1, 2, \ldots, n.$

e : $y_t = \eta_t + \varepsilon_t,$

where $R_r(L) = 1 - \rho_1 L - \rho_2 L^2 - \ldots, - \rho_r L^r$

$S_s(L) = 1 + \theta_1 L + \ldots. + \theta_s L^s$

$G_p(L) = 1 - \gamma_1 L - \gamma_2 L^2 - \ldots. - \gamma_p L^p$

$B_{q_i}^{(i)}(L) = \beta_0^{(i)} + \beta_1^{(i)} L + \ldots. + \beta_{q_i}^{(i)} L^{q_i}$

$H_{h_i}^{(i)}(L) = 1 + \alpha_1^{(i)} L + \ldots. + \alpha_{h_i}^{(i)} L^{h_i}$

$M_{m_i}^{(i)}(L) = 1 - \phi_1^{(i)} L - \ldots. - \phi_{m_i}^{(i)} L^{m_i}$

(where L is the lag operator) and the following assumptions (1), (2), (3), (4) and (5) :

Assumption 1 : The charqcteristic roots of the polynomial $G_p(L)$ lie outside the unit circle. (Also, the shock process is assumed to be independent of each one of the exogenous variables.)

Assumption 2: The variable u is independent of the variables $\xi^{(1)}, \ldots, \xi^{(n)}$.

Assumption 3 : The variables $\partial^{(1)}, \ldots, \partial^{(n)}$ and ε are white-noise errors, independent of each other, and independent of $\xi^{(1)}, \ldots, \xi^{(n)}$, η and u.

Assumption 4 : The variable a is white noise. The characteristic roots of the polynomials $R_r(L)$ and $S_s(L)$ lie outside the unit circle.

Assumption 5 : The variables $b^{(1)}, \ldots, b^{(n)}$ are independent white-noise variables. The characteristic roots of the polynomials $H^{(i)}_{h_i}(L)$ and $M^{(i)}_{m_i}(L)$ lie outside the unit circle.

In short, the processes are assumed to be stationary and invertible.

Lemma 1.1 : The coefficients of the polynomials $R_r(L)$, $G_p(L)$, $M^{(i)}_{m_i}(L)$, $H^{(i)}_{h_i}(L)$, $B^{(i)}_{q_i}(L)$ and the variances $\sigma^{(i)}_{\partial\partial}$, $\sigma^{(i)}_{bb}$, for $i = n_1+1, n_1+2, \ldots, n$, are always identified.

Theorem 1.1 : Consider the order $(q_1, q_2, \ldots, q_{n_1})$ of the polynomials of the exogenous variables that are white-noise. Arrange the set of numbers $(p, q_1, q_2, \ldots, q_{n_1})$ in non-decreasing order, and let q^*_j denote the one occupying the jth place in this new sequence (i.e., $q_1^* \leq q_2^* \leq q_3^* \leq \cdots\cdots \leq q_{n_1+1}^*$). The above model is identified iff

(a) when $r > s$, $q_j^* \geq j-1$,

(b) when $r \leq s$, $q_j^* \geq j+s-r$

for $j = 1, 2, \ldots, n_1+1$.

It is clear that the properties of autoregressive models and the moving average models are opposite in nature, but they are not symmetric.

Finding out identifiability conditions in Maravall's models is not very difficult, as it may seem at first sight, because of the complexity of the model; since many of the variables such as u_t, b_t, ξ_t, etc., are assumed to be white noise.

Willasen (1977) compared the identifiability of stochastic difference equations involving EIV's with that of classical EVM. From a Bayesian viewpoint, the problem of identification for parameters of the simultaneous equations model has been treated by Dreze (1974). Kadane (1974) also dealt with this problem from the Bayesian angle. His conclusion is that "identification is a property of the likelihood function and is the same whether considered classically or from the Bayesian approach". Mehra (1974) discussed the relationship between different meanings of identifications.

Further work on identification and estimation of dynamic equations systems with EIV's and related topics can be found in Nerlove (1967), Griliches (1967), Graupe (1972), Hannan (1971, 1976), Grether and Maddala (1973), Pandit (1973), Karni and Weissman (1974), Zellner and Palm (1974), Mehra (1975), Hsiao (1976a), Pierce (1976), Geraci (1977a), Haugh and Box (1977), Newbold (1978), Pesaran (1980), Anderson and Cheng (1980) etc. Other articles on identification and estimation of EVM's are by Geary (1949), Sargan (1958), Halperin (1961), Carlson, Sobel and Watson (1966), Clutton Brock (1967), Solari (1969), Mallios (1969), etc.

1.7 Least Squares Bias and Use of Proxy Variables

The classical linear regression situation (vide section 4 of this chapter) may also be written as

$$y_i = \alpha + \beta x_i + \varepsilon_i + v_i - \beta u_i$$

$$\text{or} \quad y_i = \alpha + \beta x_i + w_i \qquad \ldots (1.30)$$

Since X and Y can not be observed, one usually regresses y on x and 'ordinary least squares' (OLS) yields the slope estimator

$$\hat{b}_1 = \frac{m_{11}}{m_{20}} \qquad \ldots (1.31)$$

where $m_{jk} = \frac{1}{n} \sum_i (x_i - \bar{x})^j (y_i - \bar{y})^k$,

with $\bar{y} = \frac{1}{n} \sum_i y_i$ and $\bar{x} = \frac{1}{n} \sum_i x_i$. The OLS estimate is obviously biased and the bias does not decrease to zero as the sample size increases indefinitely, i.e., the estimator $\hat{b}_1$ is inconsistent. The limiting value of $\hat{b}_1$ is

$$\underset{n \to \infty}{\text{plim}} \; \hat{b}_1 = \frac{\beta}{(1 + \gamma_u)}. \qquad \ldots (1.32)$$

where $\gamma_u = \sigma_u^2 / \sigma_X^2$. Hence β is under-estimated (in the absolute sense) in the limit. The asymptotic bias is

$$B(\hat{b}_1) = -\frac{\gamma_u}{1 + \gamma_u} \beta \qquad \ldots (1.33)$$

The OLS estimate in the classical linear EVM is biased towards the origin.

One can as well regress x on y and get a 'reverse least squares' (RLS) estimator

$$\hat{b}_2 = \frac{m_{02}}{m_{11}}$$

the limiting value of which is

$$\underset{n \to \infty}{\text{plim}} \; \hat{b}_2 = \beta\,(1+ \gamma_{v'})$$

where $\gamma_{v'} = (\sigma_v^2 + \sigma_\varepsilon^2) / (\beta^2 \sigma_X^2)$. The asymptotic bias of $\hat{b}_2$ is

$$B(\hat{b}_2) = \beta\gamma_{v'} \qquad \dots (1.34)$$

Since the asymptotic bias of $\hat{b}_1$ and $\hat{b}_2$ have opposite signs, β lies between these two estimators in the limit. For example, if $\beta > 0$,

$$b_1 < \beta < b_2$$

where b_1 and b_2 are the limiting values of $\hat{b}_1$ and $\hat{b}_2$ respectively. Gini had observed this long ago (Gini, 1921) and thought that an appropriate average of $\hat{b}_1$ and $\hat{b}_2$ would lead to a satisfactory estimate of β. He proposed the arithmetic mean of the two. Recently Kaila (1980) proposed a new estimator of β which on examination turned out to be the geometric mean of the above two estimators. Comparison of biases and MSE's of these two average estimators does not lead to any definite conclusion (vide Chapter 6, Section 6.2, also Pal, 1980a).

The existenceof such bounds seems to be the only comfort in a non-identifiable model. In the multivariate case it is not difficult to get bounds for partial regression coefficients also (Frisch, 1934, Reiersol, 1945). Durbin (1954) in his review article wrote :

"The trouble with this result is that although it is of great interest, it does not help us to improve the accuracy of the

LS estimates of the coefficients. *It is only of value in revealing the limits of variation when we are in a state of complete ignorance about the relative magnitudes of the error variances*" (p.31).

The direction of the bias in the two-variable EVM is always available given the sign of β, the regression coefficient. But the question is : Does this result carry over to the case of more than one explanatory variables ? The answer is unfortunately in the negative. If there is only one regressor with error and all the other regressors are without error, then the results carry through. The direction of the bias especially for the coefficient attached with the regressor containing error can be calculated. All we have to know is the variance-covariance matrix of the observations (Levi, 1973). In the general case, where all the regressors contain errors, no such conclusion is possible. Theil (1961) derived approximate formulae for the LS estimators of the regression coefficients in a linear errors. The formulae

$$\hat{\beta}_1 - \beta_1 \simeq \frac{1}{1-\rho^2} (\theta_1 \beta_1 - \rho\theta_2\beta_2) \quad \ldots (1.35)$$

$$\hat{\beta}_2 - \beta_2 \simeq \frac{1}{1-\rho^2} (\theta_2\beta_2 - \rho\theta_1\beta_1), \quad \ldots (1.36)$$

where $\hat{\beta}_1$ and $\hat{\beta}_2$ are the OLS estimators of β_1 and β_2 from a regression of the measured variable y on the measured variables x_1 and x_2, ρ is the correlation coefficient between the true values X_1 and X_2, and θ_1, θ_2 are the ratios of the error variances to the respective variances of the true values of the regressors.

In a two-regressor EVM where there is only one regressor measured with error (this is called the proxy variable), one may face the problem of choosing between the following two alternatives for estimating the coefficient attached to the variable measured without error (main variable) :

(i) to retain both the variables and to find OLS estimate in the usual manner, and

(ii) to omit the variable measured with error and then find the OLS estimate regressing y on the error-free regressor.

McCallum (1972) studied in detail this question and concluded that the use of proxy variable (i.e., approach (i) above) is likely to be advantageous so far as bias is concerned (see also Wickens, 1972). Aigner (1974) also dealt with this problem and studied MSE's. He found a condition for the superiority of the proxy-variable method (approach (i)). More recently, Lahiri and Chaudhuri investigated such conditions for the case where the regressors are stochastic (vide Chaudhuri, 1979). Giles (1980) also discussed the same problem and found a contour of dominance for the bias squared.

The problem is even more difficult if both the regressors are subject to error. In fact, in this case the direction of bias may be in either way (Welch, 1975). Moreover, no general conclusion is possible as to whether to use the proxy variable or to omit it altogether to estimate the regression coefficient of the main variable. However, depending on conditions on variances and covariances of the variables, there are some cases where definite conclusions are possible (Garber and Klepper, 1980).

1.8 Additional Information and Their Use in Estimation

Since the classical EVM is non-identifiable, it is not possible to get a consistent estimator for the slope parameter unless some additional information is available *a priori* or one is able to assume that either the true variable X or (u, v) is non-normal.

Usually, the following additional informations are used in econometrics :

(i) Either σ_u^2 or σ_v^2 or $\lambda = \sigma_u^2/\sigma_v^2$ /or/ both σ_u^2 and σ_v^2 are known.

(ii) Information on one or more auxiliary variables called instrumental variables (IV's) is available.

(iii) It is given that X is non-normal or at least third-order cumulant is non-zero.[1.1/]

Sometimes a different form of the model is used such as one including a lagged variable. Repeated measurements are also possible at times, especially in laboratory experiments. In laboratory experiments the variable x can be controlled (Berkson, 1950), while X varies at random from one observation to another. Knowledge on higher moments of EIV's may also be utilized (Schneeweiß, 1976).

1.8.1 Knowledge of Error Variances

Cases where one has prior knowledge of σ_u^2 or σ_v^2 or both or of their ratio λ (= σ_u^2/σ_v^2) have been extensively discussed in the literature (Allen, 1939, Seares, 1944, Madansky, 1959, Barnett, 1967, etc.). From the classical EVM one may derive the following equations from exact relation :

1.1/ Sometimes specific distributional assumption is made depending on the situation.

$$E(x) = E(X) = \mu \quad \text{... (1.37)}$$

$$E(y) = E(Y) = \alpha + \beta\mu \quad \text{... (1.38)}$$

$$\sigma_x^2 = \sigma_X^2 + \sigma_u^2 \quad \text{... (1.39)}$$

$$\sigma_y^2 = \beta^2 \sigma_X^2 + \sigma_v^2 \quad \text{... (1.40)}$$

$$\sigma_{xy} = \beta\sigma_X^2 \quad \text{... (1.41)}$$

ML estimation uses sample moments in place of the quantities at the left. We have thus five equations to solve for six unknowns, namely, μ, σ_X^2, σ_u^2, σ_v^2, α and β. So far as variance-covariance parameters are concerned the following three equations are relevant

$$\sigma_x^2 = \sigma_X^2 + \sigma_u^2$$

$$\sigma_y^2 = \beta^2\sigma_X^2 + \sigma_v^2$$

$$\sigma_{xy} = \beta\sigma_X^2.$$

These contain four unknown parameters, σ_X^2, σ_u^2, σ_v^2 and β. Hence knowledge of either σ_u^2, or σ_v^2, or σ_u^2/σ_v^2 would enable us to get the ML estimate of β. In fact,

(i) If σ_u^2 is known, we have the estimate of β as

$$\hat{\beta}_1 = \sigma_{xy} / (\sigma_x^2 - \sigma_u^2), \quad \text{... (1.42)}$$

(ii) If σ_v^2 is known, we have

$$\hat{\beta}_2 = (\sigma_y^2 - \sigma_v^2)/\sigma_{xy}, \quad \text{... (1.43)}$$

(iii) If $\lambda = \sigma_u^2/\sigma_v^2$ is known then the estimate of β becomes (Lindley, 1947)1.2/

1.2/ Lindley's estimate is incorrect. We have written Madansky's (1959) correct version. One may see also Creasy (1956).

$$\hat{\beta}_3 = \frac{(\lambda \hat{b}_1 \hat{b}_2 - 1) + \sqrt{(\lambda \hat{b}_1 \hat{b}_2 - 1)^2 + 4\lambda \hat{b}_1^2}}{2\lambda \hat{b}_1} \qquad \ldots (1.44)$$

where $\hat{b}_1 = m_{11}/m_{20}$ and $\hat{b}_2 = m_{02}/m_{11}$,

where $m_{ij} = \frac{1}{n} \sum_1 (x_1 - \bar{x})^i (y_1 - \bar{y})^j$.

The above three estimates are ML estimates and can also be derived by weighted LS method. Lindley (1947) suggested that one should minimize

$$\sum_{i=1}^{n} w_i(\beta) (y_i - \alpha - \beta x_i)^2 \qquad \ldots (1.45)$$

where the $w_i(\beta)$'s are inversely proportional to the variances of $y_i - \alpha - \beta x_i$ given X_i, i.e., $w_i(\beta) = k/(\sigma_v^2 + \beta^2 \sigma_u^2)$, where k does not depend on i. Now if σ_u^2/σ_v^2 is known to be λ, then $\sigma_v^2 + \beta^2 \sigma_u^2$ becomes $(1 + \beta^2 \lambda) \sigma_v^2$. So now minimizing (1.45) we arrive at equation (1.44) as the estimate of β. This procedure can be followed even when we know only σ_u^2 or σ_v^2 (see also Sprent, 1966).

(iv) If both σ_u^2 and σ_v^2 are known, then we have two estimators $\hat{\beta}_1$ and $\hat{\beta}_2$ defined above. This is like the problem of overidentification, since in our case knowledge of only one of σ_u^2 and σ_v^2 is necessary. We can, of course, define a pooled estimator as

$$\hat{\beta}_a = a \hat{\beta}_1 + (1 - a) \hat{\beta}_2$$

and find out 'a' so as to minimize the variance of $\hat{\beta}_a$ (see Chapter 6, Section 6.3). The ML estimator for this model, where both σ_u^2 and σ_v^2 are known, exist (Barnett, 1967). The ML estimator of β here coincides with $\hat{\beta}_3$. But the other estimators (σ_X^2 etc.) differ from those for the model where $\lambda = \sigma_u^2/\sigma_v^2$ is known.

1.8.2 IV Estimation

Information on one or more auxiliary variables, called instrumental variables (IV's), which are independent of (or at least uncorrelated in the limit with) the error terms and highly correlated with the regressor(s) are sometimes useful for consistent estimation for β (Reiersol, 1945; Geary, 1949; Sargan, 1958; Liviatan, 1961; Mallios, 1969; Carter and Fuller, 1980).

If z is such an IV then

$$\hat{b} = \{ \frac{1}{n} \Sigma z_i(y_i - \bar{y})\}/\{ \frac{1}{n} \Sigma z_i(x_i - \bar{x})\} \qquad \ldots \quad (1.46)$$

is consistent for β provided

$$\underset{n \to \infty}{\text{plim}} \; m_{11} (v + \varepsilon, z) = 0,$$

$$\underset{n \to \infty}{\text{plim}} \; m_{11} (u, z) = 0,$$

and $\underset{n \to \infty}{\text{plim}} \; m_{11} (x, z) \neq 0$ [1.3]

If z is uncorrelated with x the sampling variance of the IV estimator is infinitely large. With only a small correlation between z and x the sampling variance is large and we may be paying a very high price for consistency if we use IV estimator instead of other available estimators. In general, one should look for a variable z which is fairly strongly correlated with X but uncorrelated with u, ε and v.

1.3/ In the case of multiple regressors where some of the variables are free from errors while the others are subject to errors, the variables which are free from errors may be used as IV's, provided they are uncorrelated in the limit with the errors in the regressors and in the regressand and also with the disturbance ε.

An IV, if based on x's affected by errors, is likely to be correlated with u's also, since x is correlated with u. So, use of such IV's may not be safe even for consistent estimation of parameters. In practice, IV's are extremely difficult to obtain and there is no means for checking whether they are really uncorrelated in the limit with each of the error terms and with the disturbance ε.

The difficulty of finding IV's may be illustrated with the situation in engel curve analysis (estimation of demand-expenditure relationship). Data on both regressor (total household consumer expenditure) and regressand (item consumption) are affected by seasonal and other transitory elements apart from data collection errors. Liviatan (1961) suggested the use of recorded income as IV. But recorded income is hardly available in survey data collected in developing countries. Actually, recorded income even if available cannot be trusted, because informants most of the time are reluctant to disclose their true incomes. Again, the errors in the variables may have a definite relation with the recorded income. Liviatan himself admitted that if recorded income includes current borrowing then it is likely to be correlated with the error part of the total expenditure. Again, if expenditure influences income then also we face the same difficulty.

We may now consider multiple regression situations and suppose that several IV's are available. Given a set of IV's, the model can be under-identified, just-identified or over-identified, depending on the number of instruments available for the purpose. In an under-identified model we can do nothing but explore other types of information. An over-identified model can be tackled efficiently either by generalizing the model making it just-identified or by minimizing the variance of the pooled estimator (Blalock, 1969, 1970; Joreskog, 1970, 1973; Werts, Joreskog and Linn, 1973; Joreskog and Goldberger, 1975). Chapter 6, Section 6.4 discusses this case in detail.

The well-known grouping methods due to Wald (1940), and Bartlett (1949) or the method based on ranks due to Durbin (1954) are special cases of IV method (vide Johnston, 1972). The use of equifrequency groups in the three-group method due to Bartlett is widely believed to be optimal or near optimal in a large variety of situations. A careful scrutiny of the supporting evidence, however, shows that most of the X-distributions so far examined are symmetrical or negatively skewed. Pal and Bhaumik (1979) showed that for commonly occuring lognormal and gamma distributions of X the optimal proportions in the three groups should be 0.40, 0.45, and 0.15, approximately, for a wide range of parameter values (vide Chapter 3, Section 3.2 for details). The above group proportions, however, does not hold good in the case of heteroscedastic disturbances. The optimum proportions in different groups vary with the degree of heteroscedasticity (vide Chapter 3, Section 3.3).

It is also not recognized by many users that the estimators proposed by Wald, Bartlett and Durbin are, in general inconsistent, unless the errors in the X-variable are too small to affect the grouping or ranking of the regressor values (Neyman, J., and Scott, E. L., 1951). Durbin actually proposed that the x-values be arranged into broad groups after ranking, and observations in the ith group be given rank i, to reduce the bias of the Durbin's estimator. Clearly, there is scope for further investigations into the bias of all these estimators and for modifications for reduction of bias.

It has been found that the OLS estimator of β in the EVM has a smaller variance compared to the IV estimator. But OLSE is biased. To compensate for this Feldstein (1974) considered the following WAIVE (Weighted Average IVE).

$$\text{WAIVE} = \lambda(\text{OLSE}) + (1-\lambda)\text{IVE},$$

and minimized the MSE of WAIVE to get the optimum λ. Sargan and Mikhail (1971) tried to arrive at some approximate formula for the cumulative distribution function of the IV estimators.

1.8.3 The Case of Non-normal Regressor : Estimation Via Moments and Cumulants

In view of the problem of identification mentioned in section (1.6) of this chapter some work has been done for the case of non-normal regressors (Neyman, 1951; Wolfowitz, 1952, 1953a, 1953b; Spiegelman, 1979). These deep investigations have not, however, led to any useful procedures. There are some simpler estimators based on moments and cumulants which are not optimal, but which may have moderately high efficiency.

Geary (1942, 1943) showed that

$$\hat{\beta} = \frac{K(c_1, c_2 + 1)}{K(c_1 + 1, c_2)}, \quad c_1, c_2 > 0 \qquad \ldots (1.47)$$

where $K(c_1, c_2)$ is the sample cumulant of order (c_1, c_2) of (x, y), is a consistant estimator for β if $\operatorname*{plim}_{n \to \infty} K(c_1 + 1, c_2) \neq 0$. Thus, one has an infinite class of consistent estimators. The sampling errors of sample cumulants (or moments) generally increase rapidly with their order (Geary, 1942; Madansky, 1959). Hence, one should confine oneself to estimators based on moments/cumulants of the lowest feasible order. If X is asymmetric one should not go beyond third order cumulants.

Scott (1950) and Drion (1951) pursued this idea and presented consistent moment estimators for β. Durbin (1954) also suggested IV estimator which reduces to a moment estimator, because the IV suggested by him happens to be some power of the regressor x. The estimators based on third order moments proposed by Geary, Scott, Drion and Durbin are members of a more general class considered by the author (Pal, 1980b). The proposed general class was defined as

$$\hat{\beta}_s = f(\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3) \quad \ldots (1.48)$$

such that

(i) $f(c\hat{\beta}_1, c\hat{\beta}_2, c\hat{\beta}_3) = c.f(\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3)$ for all $c \neq 0$,

and (ii) $f(1, 1, 1) = 1$,

where $\hat{\beta}_1$, $\hat{\beta}_2$ and $\hat{\beta}_3$ are three 'basic' moment estimators defined by

$$\hat{\beta}_1 = m_{03}/m_{12}, \quad \hat{\beta}_2 = m_{12}/m_{21} \quad \text{and} \quad \hat{\beta}_3 = m_{21}/m_{30}.$$

Every consistent estimator based only on third order moments must be a member of this class. The asymptotic efficiencies of six estimators (taking three more members from this class) relative to OLS estimator were studied by Pal (1980) assuming lognormality of the regressor. The estimators were found to be fairly efficient even when OLS is fully valid. The best of them may be chosen in a given situation and this would have moderate or high efficiency (vide Chapter 2).

Scott proved that the equation

$$m_{03} - 3bm_{12} + 3b^2m_{21} - b^3m_{30} = 0 \qquad \ldots (1.49)$$

has a root which is a consistent estimator for β. This result, however, can be arrived at even if one takes a more general model where

(i) u given X has a symmetric distribution with mean 'zero' and variance $a^2 X^b$ where 'a' and 'b' are constants; and

(ii) v is assumed to be correlated with u by the following relation

$$v = \theta u + w \qquad \ldots (1.50)$$

where θ is constant and w is independent of u and is distributed 'normally' with mean zero and variance σ_w^2 (vide Chapter 5, Section 5.3). The second generalization is quite understandable and its need has been stressed by many writers in the EIV literature (vide Rao and Miller, 1972). The first generalization has already been discussed in section 1.5 of this chapter and can be traced to Friedman's book (1957, p.27) where he says : "For a group of individuals, it is plausible to suppose that the absolute size of the transitory component (error component) varies with the size of the permanent component (true component)". McIntyre and others (1966) in geological samples also found the same for isotope-dilution measurements.

In some applications one may exploit prior knowledge of the shape of X-distribution based on analysis of x-values. Thus in engel curve analysis one can assume that true income or expenditure (per capita) is lognormally distributed (Aitchison and Brown, 1957; Bhattacharya and Iyengar, 1961; Iyengar, 1967; Iyengar and Jain, 1974; Bhattacharya, 1978) and have moment estimators of specific nature. At first the parameters of the distribution of income may be estimated by method of moments or by ML method since the specific nature of the model is known. Then the fitted parameter may be used to get consistent estimators for β. The specific nature of the problem also allows one to incorporate generalizations (i) and (ii) above (Pal, 1977). In fact, one can take three-parameter lognormal or even take different types of engel curves. The ML estimation requires assumptions on specific distribution of u given X. This may be taken as normally distributed with mean zero and variance a^2X^2 or — to make its range finite — following (say) a Pearsonian type II distribution in the range $-cX$ to cX where c is a positive constant. It may be mentioned here that the Pearsonian type II distribution is a closed approximation to the normal distribution. Also, the first three moment estimators give the same type of equations. If we want to emphasize the fact that the observed variable is always non-negative, c can be chosen to be less than or equal to one. Taking c=1 does of course simplify the task.

1.9 The Berkson's Model

The model due to Berkson (1950) discusses a situation in which the observed values (x's) are controlled (see also Kendall, 1952; Lindley, 1953), i.e., instead of trying to observe a given X_i one fixes x_i $(= X_i + u_i)$ and tries to observe y_i. This model can particularly be applied in laboratory experiments where a researcher can fix the x-value to a certain level and observe the results. Here x is fixed and X is a random variable giving rise to x. The model is

$$\left.\begin{aligned} Y_i &= \alpha + \beta X_i \\ y_i &= Y_i + v_i \\ x_i &= X_i + u_i \end{aligned}\right\} , \quad i = 1, 2, \ldots, n, \ldots \qquad (1.51)$$

or

$$\left.\begin{aligned} y_i &= \alpha + \beta x_i + v_i - \beta u_i \\ y_i &= Y_i + v_i \\ x_i &= X_i + u_i \end{aligned}\right\} , \quad i = 1, 2, \ldots, n \qquad \ldots (1.52)$$

Since both u and v are independent of x in this case (since x is non-stochastic); OLS procedure gives a consistent estimate of β.

Federov (1974) later generalized this model to the case of multiple regressors. He obtained moment estimators which are consistent, provided they have a limit. Analysis of controlled variables specially to the non-linear case have also been discussed by Geary (1953) and Scheffe (1958).

1.10 Effects of Errors

(a) On Regression Lines : Conditions have been established under which the regression relationship will continue to be linear even when the variables are affected with errors (Lindley, 1947; see also, Allen, 1938; Fix, 1949; Laha, 1956). Cochran (1971) examined the effect of errors on the linearity of relationship between two variables where only the regressor contains errors. The conclusion, for the situation considered by him, is that the relation can be approximated by (i) a quadratic equation if X is skewed or by (ii) a cubic equation if X is symmetric. The linear component obviously dominates the relation.

(b) On the Estimates : Besides Berkson's M del there are some models where the introduction of error terms in the regressors do not have direct influence so far as consistency of LS regression coefficient in concerned (Durbin, 1954). Under the classical functional relationship assumption Richardson and Wu (1970) obtained the exact bias and MSE of the OLS estimate of β. Assuming that the EIV and the errors in the equation are correlated Halperin and Gurian (1971) found the same for the two variable model. They showed that the downward bias does not necessarily hold if the measurement errors u and v are correlated (see also Rao and Miller, 1972; Levi, 1977; Reed and Wu, 1977). Robertson (1974) also worked along this line and found asymptotic variances of ML estimators when (i) σ_u^2, (ii) σ_v^2, (iii) σ_u^2/σ_v^2 or (iv) both σ_u^2 and σ_v^2 are known. He also found asymptotic bias of these estimates up to order n^{-1}. Assuming that an estimator of the covariance matrix of the measurement errors is available Fuller (1980) investigated the limiting behaviour of estimators for several EIV models.

(c) On the Correlation Coefficients and Others : The presence of EIV's causes the variance of the variable to increase its value. Hence any statistic using variance estimates will show some effect. Cochran (1968, 1970) calculated the effect of EIV on multiple correlation coefficient. He assumed a model where EIV's are uncorrelated with the true values and with each other. Introducing the following symbols :

ρ_i = correlation coefficient between Y and X_i

ρ_i' = correlation coefficient between y and x_i

$= \rho \sqrt{R_y R_i}$

where R_y and R_i are reliability coefficients, i.e., $\sigma_i^2 = R_i(\sigma_i^2 + \sigma_{\varepsilon_i}^2)$. (y and x_i's are measured variables and Y and X_i's are true variables.)

Γ = squared multiple correlation coefficient between Y and X's

Γ' = squared multiple correlation coefficient between y and x's.

Assume the simplest case where x_i and x_j are uncorrelated for all i and j such that $i \neq j$, then

$$\Gamma' = \Gamma R_y \left(\Sigma \rho_i^2 R_i / \Sigma \rho_i^2 \right).$$

EIV's induce a lower limiting value of the value of "t" statistic (Bloch, 1978). Thus it has effects on inferences also. The author (Pal and Chakravarty, 1978) introduced a model where income is subject to error. In that it was proved that the presence of error induces the measures of inequality to increase their values, which is, however, not surprising. Lankipalle (1973) examined the effect of error on partial correlation coefficients.

1.11 Repeated Measurements with Error

The problems with repeated observations containing errors have been extensively discussed in Madansky (1959) and Cochran (1968). If we have N_i observations on each (X_i, Y_i) with

$$\left. \begin{aligned} y_{ij} &= Y_i + v_{ij} \\ x_{ij} &= X_i + u_{ij} \end{aligned} \right\} \quad \begin{aligned} i &= 1, 2, \ldots, n \\ j &= 1, 2, \ldots, N_i \end{aligned} \qquad \ldots (1.53)$$

and if the usual assumptions on independence are made, then one can perform an analysis of variance on the x's and the y's and obtain estimates of β (Madansky, 1959). We can explain it by the following table :

Source		Mean Squares	Expected Mean Squares
Between Sets	I	$\dfrac{\sum_1^n N_i(\bar{x}_{io} - \bar{x}_{oo})^2}{n-1}$	$\sigma_u^2 + \left(\dfrac{N^2 - \sum_1^n N_i^2}{nN - N}\right) \sigma_X^2$
	II	$\dfrac{\sum_1^n N_i(\bar{x}_{io} - \bar{x}_{oo})(\bar{y}_{io} - \bar{y}_{oo})}{n-1}$	$Cov(u,v) + \left(\dfrac{N^2 - \sum_1^n N_i^2}{nN-N}\right) \beta\sigma_X^2$
	III	$\dfrac{\sum_1^n N_i(\bar{y}_{io} - \bar{y}_{oo})^2}{n-1}$	$\sigma_v^2 + \left(\dfrac{N^2 - \sum_1^n N_i^2}{nN-N}\right)\beta^2 \sigma_X^2$
Within Sets	IV	$\dfrac{\sum_{i=1}^n \sum_{j=1}^{N_i} (x_{ij} - \bar{x}_{io})^2}{N - n}$	σ_u^2
	V	$\dfrac{\sum_{i=1}^n \sum_{j=1}^{N_i} (x_{ij} - \bar{x}_{io})(y_{ij} - \bar{y}_{io})}{N-n}$	Cov (u, v)
	VI	$\dfrac{\sum_{i=1}^n \sum_{j=1}^{N_i} (y_{ij} - \bar{y}_{io})^2}{N-n}$	σ_v^2

where $\bar{x}_{io} = \sum_{j=1}^{N_i} x_{ij}/N_i$, $\bar{y}_{io} = \sum_{j=1}^{N_i} y_{ij}/N_i$, $\bar{x}_{oo} = \sum_{i=1}^{n} \sum_{j=1}^{N_i} x_{ij}/N$,

$\bar{y}_{oo} = \sum_{i=1}^{n} \sum_{j=1}^{N_i} y_{ij}/N$, and $N = \sum_{i=1}^{n} N_i$

The estimates of β are (II-V)/(I-IV), (III-VI)/(II-V), and $\sqrt{(III-VI)/(I-IV)}$, all of which are consistent. Housner and Brennan (1948) gave another estimate of β in the same model.

The above model can also be looked at from the point of view of grouping, i.e., we have n groups, each groups having N_i observations. In this case also we have the same three estimates. These three estimates are due to Tukey (1951). IV's can also be used in this situation (Madansky, 1959).

Following Lord (1960), Degracie and Fuller (1972) proposed an estimate of the following functionally related covariance model :

$$y_{ij} = \alpha + \tau_i + \beta x_{ij}, \quad i = 1, 2, \ldots, n \text{ and } j = 1, 2, \ldots, r \qquad \ldots (1.54)$$

where n is the number of treatments and r is the number of replications, and τ_i is the ith treatment effect with $\Sigma\tau_i = 0$. x_{ij} and y_{ij} are the observed values with errors following a bivariate normal distribution with zeromeans. Assuming that the estimates of the variances of the observational errors are available, they developed the estimator of β that are unbiased to $O(r^{-1})$ where r is the number of observations on each treatment. But these types of models are not useful in econometrics. They are usually used in experimental design.

Ord (1969) assumed a model where replicated observations (only two) are possible for fixed true values of the variables, and obtained the ML estimators for the functional relationship

$$\left.\begin{aligned} Y_i &= \alpha + \beta X_i \\ \text{where } \xi_{ij} &= X + \delta_{ij} \\ \eta_i &= Y_i + \varepsilon_i \end{aligned}\right\} \quad \begin{aligned} i &= 1, 2, \ldots, n \\ j &= 1, 2 \end{aligned} \qquad \ldots (1.56)$$

with usual assumptions. This may be relevant when observations are based on two independent situations.

Dolby and Freeman (1975) dealt with ML estimation of linear and non-linear functional relationships with replicated observations. The analysis for bivariate case was extended to multivariate situations and the error variance was allowed to be known. Previous articles dealing with replicated observations are Villegas (1961), Dolby (1972), Dolby and Lipton (1972) etc. Dolby (1976) later worked on structural relations of this type. Very recently, Chan and Mak (1979) assumed a linear structural relation of the type:

$$\left.\begin{aligned} Y_i &= \alpha + \beta X_i \\ y_{ij} &= Y_i + \varepsilon_{ij} \\ x_{ij} &= X_i + \partial_{ij} \end{aligned}\right\} \quad \begin{aligned} i &= 1, 2, \ldots, n \\ j &= 1, 2, \ldots, r \end{aligned} \qquad \ldots \quad (1.57)$$

with usual assumptions. He found the ML solution to be a root of a fourth-degree polonomial. However, it is consistent as the number of replications increases.

1.12 Simultaneous Equations Models

Some of the characteristics of simultaneous equations models are shared by EVM's. As for example, in the simplest case of EVM, we have

$$\begin{aligned} y &= \alpha + \beta x + v - \beta u + \varepsilon \\ &= \alpha + \beta x + \varepsilon' \end{aligned} \qquad \ldots \quad (1.58)$$

In the simultaneous equations models the regressor(s) become(s) correlated with the disturbance term. Here also x is correlated with ε', because both variables involve u. In fact, the IV approach in EVM can be treated as one method of estimation for simultaneous equations model. Thus, when the regression of X on z (z is the instrument) is linear (i.e., $X = \gamma z + \eta$) then we have

$$\left.\begin{array}{l} y = \alpha + \beta x + \varepsilon' \\ \text{and} \quad x = \gamma z + \eta' \\ \text{with } \varepsilon' = v - \beta u + \varepsilon, \quad \eta' = \eta + u \end{array}\right\} \quad \ldots (1.59)$$

The IV approach to simultaneous equations models has been discussed in detail by Madansky (1976). Avery (1977) showed how a two- or three-component error structure can be used with seemingly unrelated regression which has an application to large panel data sets. Garaci (1977b) developed a theory for the asymptotically efficient LS estimation procedure of simultaneous equations models with measurement error. Hausman (1977) proposed FIML and IV approaches to simultaneous equations models with EIV. Mariano (1977) found the finite sample properties of IV estimators where the instruments are assumed to be non-stochastic. Lahiri and Schmidt (1978) discussed estimation of Triangular Structural Systems. Among other work on simultaneous equations models Robinson (1974), Geraci (1980), Hausman and Taylor (1980), Phillips (1980) etc., may be mentioned.

Sociologists have long been applying simultaneous equations models in path analysis and multiple indicators analysis (Boudon, 1965, 1967, 1968; Blalock, 1970; Wiley and Wiley, 1970, Werts, Joreskog and Linn, 1973; Duncan, 1975). Joreskog and Goldberger (1975) developed ML and other procedures to an widely approached MIMIC model (Multiple Indicators and Multiple Causes) of a single latent variables. The model is as follows :

$$y^* = \alpha_1 x_1 + \ldots + \alpha_K x_K + \varepsilon \quad \ldots (1.60)$$

The latent variable y^* determines m endogenous indicators as

$$y_1 = \beta_1 y^* + u_1, \ldots, y_m = \beta_m y^* + u_m \qquad \ldots (1.61)$$

The disturbances are all assumed to be mutually independent. Work along this line of causal analysis may be found in Blalock,(1969), Costner,(1969), Zellner, (1970), Hauser and Goldberger, (1971), Hauser, (1972), Duncan and Featherman, (1972), Griliches and Mason, (1972), Goldberger, (1972), Joreskog,(1970, 1973).

1.13 Bayesian Methods of EVM's

A great impetus to the analysis of EVM was given by the introduction of Bayesian methods. Zellner's book (1971, Chapter 5) contains an excellent review on this approach. The problems of identification from Bayesian angles have been studied by Dreze,(1974) and Kadane, (1974). Following Zellner, (1971), Lindley and El-Sayyad, (1968), and Lindley and Smith, (1972); Florens and others, (1974) have developed a theory for EVM estimations and inferences in a Bayesian framework.

To provide the Bayesian analogue of the ML results of the functional form of EVM, Zellner employed the following prior pdf :

$$p(X, \alpha, \beta, \sigma_u, \sigma_v) \propto \frac{1}{\sigma_u \sigma_v} \qquad \ldots (1.62)$$

with $-\infty < \alpha, \beta, X < \infty$, $0 < \sigma_u, \sigma_v < \infty$.

In practice prior pdf's for β and λ $(= \sigma_v^2 / \sigma_u^2)$ must be assigned. Since β lies between $\hat{b}_1$ and $\hat{b}_2$ in the limit one may a assign a beta pdf of the following form

$$p(z|a,b) = \frac{1}{B(a,b)} z^{a-1} (1-z)^{b-1}, \quad a, b > 0, \quad 0 < z < 1, \qquad \ldots (1.63)$$

where $z = (\beta - \hat{b}_1)/(\hat{b}_2 - \hat{b}_1)$, a and b are prior parameters to be assigned by the investigator, and B(a, b) denotes the beta function with arguments a and b. λ may be taken to follow an inverted gamma distribution. In his illustration Zellner found this procedure to give a very good posterior distribution which is more peaked, unimodel and the modal values are very close to the actual values. He also studied structural form of the EVM from the Bayesian point of view.

The study of Florens and others (1974) emphasised the use of "uninformative" prior distribution. They also obtained posterior distributions (i) through an informative prior distribution on the incidental parameters, and (ii) through an informative prior distribution on the covariance matrix.

1.14 Miscellaneous

Testing and Inferences : One may test whether two regression lines have equal slopes where the variables under study contain measurement errors, if error-variances are known. Lord (1960) did the same using duplicate measurements of the predictor variable to estimate the measurement error variance. Stroud (1972) compared conditional means and variances of the two sets of observations following EVM where the variances of the measurement errors are known. Rogot's work was on morbidity rates of two groups. He compared the two groups in terms of morbidity rates in the presence of errors of misclassification.

In a functional relationship Halperin (1964) showed how one gets confidence interval for the slope parameter which is different from that obtained by t-statistic. Before that he, in 1961, dealt with confidence interval around estimates of the form $\Sigma l_i y_i / \Sigma l_i x_i$ (where $\Sigma l_i = 0$ and $\Sigma l_i^2 = 1$) in a structural equation set up.

Applications : An obvious application of EVM's is to the theory of consumption function introduced by Friedman (1957). Attfield (1977) tried this model where only grouped means are available to analyse the impact of consumption. A related work is by Ware (1972) where he assumed that only ranks of the means are known. Applications can also be found in Sibling Models (Griliches, 1979), in Geology (McIntyre and others, 1966), in Management Sciences (Warren et al, 1974). There are cases (e.g., dummy variables in case of detecting a desease correctly or incorrectly) where regression equation contains binary independent variable measured with error. Aigner (1973) examined the problem of getting a consistent estimator in this situation.

Others : In a structural EVM (with one IV available) the ML estimate is the median of the LS, Reverse LS and the IV estimate if all these have same sign (Leamer, 1978). Zellner tried the MELO (Minimum Expected Loss) approach to structural equations models with errors. The article by Zellner and Park (1979) contains a review on developments along this line. Kunitomo (1980) derived asymptotic expansions of the distributions of the ML estimator and the OLS estimator in a linear functional relationship model where the ratic of the error variances are assumed to be known. ML Estimations in linear functional and structural relationships were also discussed in Villegas (1961), Gleser and Watson (1973), Bhargava (1977), Healy (1980).

It may be mentioned here that some good review articles on the topic of EVM's already exist in the literature. Particular mention may be made of papers by Durbin (1954), Madansky (1959), Cochran (1968), Moran (1971). Griliches' (1974) review was mainly concerned with simultaneous equations Model. Goldberger (1972b, 1974) stressed recent trends in research in the area. Lankipalle (1975) and Chaudhuri (1979) also did some review work on this line. The review work of the author (Pal, 1980a) mainly explores the possibility of finding moment estimators in the EVM's where the regressor is non-normal; through critical inspections of the past work along this line.

Chapter 2

CONSISTENT MOMENT ESTIMATORS OF REGRESSION COEFFICIENTS IN THE STANDARD EVM

2.1 Introduction

In many econometric investigations, the errors in variables (EIV) are not negligible (Morgenstern, 1963) and vitiate LS estimation of regression coefficients (Johnston, 1972). Thus, examination of 25 series relating to national accounts by Langaskens and Rijckeghem (1974) showed that the standard deviations of the errors ranged from 5 to 77 per cent of the average value of the corresponding variable.

The well-known methods proposed for handling the classical EIV model (EVM) in regression analysis suffer from serious limitations :

(a) ML estimation requires strong assumptions about the distribution of the errors and also some knowledge of the covariance matrix of the error terms.

(b) The technique of IV estimation is not always handy because suitable instruments may not be available, and in any case, one can never check the assumptions that the instrument is uncorrelated in the limit with each of the error terms. The Wald Bartlett grouping methods as well as the method due to Durbin (1954) tacitly assume that the errors affecting the regressor values are too small to alter their grouping or ranking.

This chapter examines the possibilities of an approach made by Geary (1942) and others apparently neglected by later researchers. The approach yields consistent and reasonably efficient estimators of regression coefficients based on uni- and bi-variate moments of third or higher order which are computationally simple and need milder assumptions than those mentioned in (a) and (b) above. It is assumed that the errors in the variables are independent of their true values (see, however, section 2.7). The case where they are dependent will be dealt with in latter chapters which consider the situation where errors in the variables are possibly correlated.

Section (2.2) specifies the two-variable regression model under investigation and reviews the work done by Geary and others. Section (2.3) compares the asymptotic variances of six moment-based estimators mentioned in section (2.2). Section (2.4) compares the asymptotic variance of the moment-based estimator with the least asymptotic variance with that of OLS assuming that the regressor is error free, under specific distributional assumptions. Section (2.5) compares the asymptotic efficiencies of the six estimators relative to OLS estimator assuming that the regressor is lognormally distributed. Section (2.6) deals with estimators based on higher moments which would be useful if the estimators based on third-order moments fail, because the distribution of the 'true regressor' is symmetric. Section (2.7) extends these ideas to the case of $m > 1$ regressors and briefly mentions the case where the error terms are correlated. Section (2.8) makes some concluding observations on the limitations of the results reached.

2.2 The Model and Available Moment Estimators

Consider the following set of relations :

$$Y_i = \alpha + \beta X_i + \varepsilon_i, \qquad i = 1, 2, \ldots, n \qquad \ldots (2.1)$$

where X and Y are true but non-observable magnitudes of the regressor and the regressand respectively; α and β are unknown parameters; and ε is the disturbance which is normally distributed. The assumptions of the Classical Normal Linear Regression Model (Goldberger, 1964) hold excepting that X is stochastic and fully independent of ε. The observed values of regressor and regressand are

$$x_i = X_i + u_i \quad \text{and} \quad y_i = Y_i + v_i \qquad \ldots, \quad (2.2)$$

where u_i and v_i are EIV's assumed to be independent of true values and between themselves. ε, u, v, are assumed to be serially i.i.d. with

$$E(u_i) = E(v_i) = 0, \quad V(u_i) = \sigma_u^2, \quad V(v_i) = \sigma_v^2, \quad \forall i.$$

and

$$E(\varepsilon_i) = 0, \quad V(\varepsilon_i) = \sigma_\varepsilon^2, \quad \forall i.$$

Let us write for sample moments

$$m_{rs}(x,y) = \frac{1}{n} \sum_i (x_i - \bar{x})^r (y_i - \bar{y})^s ,$$

and

$$m'_{rs}(x,y) = \frac{1}{n} \sum_i x_i^r y_i^s ,$$

where

$$\bar{x} = \frac{1}{n} \Sigma x_i \quad \text{and} \quad \bar{y} = \frac{1}{n} \Sigma y_i .$$

We may also write for simplicity

$$m_{io}(x,y) = m_i(x), \qquad m_{oj}(x,y) = m_j(y).$$

Correspondingly true moments will be denoted μ_{rs}, μ'_{rs}, $\mu_i(x)$ or $\mu_j(y)$, as the case may be.

It is well-known that under certain conditions the sample moments are consistent estimators of corresponding true moments which are functions of α, β, the variances σ_u^2, σ_v^2, σ_ε^2 and the true moments of X. If only the moments of the first and the second order are considered, five relations are obtained for seven unknown parameters, viz., α, β, σ_u^2, σ_v^2, σ_ε^2, $\mu_1'(X)$ and $\mu_2'(X)$ [or $\mu_2(X)$]. In fact, σ_v^2 and σ_ε^2 always appear in the form of $\sigma_v^2 + \sigma_\varepsilon^2$, so that in effect we have five relations for six unknown parameters. The first five equations considered by Drion (1951) are 2.1/

(i) $m_1'(x) = \mu_1'(X)$,

(ii) $m_1'(y) = \alpha + \beta\mu_1'(X)$,

(iii) $m_2'(x) = \mu_2'(X) + \sigma_u^2$,

(iv) $m_2'(y) = \alpha^2 + 2\alpha\beta\,\mu_1'(X) + \beta^2\,\mu_2'(X) + (\sigma_\varepsilon^2 + \sigma_v^2)$,

(v) $m_{11}'(x,y) = \alpha\,\mu_1'(X) + \beta\mu_2'(X)$.

One may, then, include similar equations based on third-order moments, if u and v are further assumed to be symmetrically distributed or rather having zero thrid-order moment. One can choose from the four equations given below :

(vi) $m_3(x) = \mu_3(X)$,

(vii) $m_3(y) = \beta^3\,\mu_3(X)$,

(viii) $m_{21}(x,y) = \beta\mu_3(X)$,

(ix) $m_{12}(x,y) = \beta^2\,\mu_3(X)$

2.1/ Drion assumed a functional relationship between X and Y so that X is non-stochastic and $\sigma_\varepsilon^2 = 0$. However, the introduction of σ_ε^2 does not alter the picture.

Inclusion of any two of these equations introduces only one new unknown parameter, namely $\mu_3(X)$. Drion used the equation for $m_3(x)$ and $m_3(y)$ and solved the system of seven equations for the seven unknown parameters to get $\hat{\beta}_4 = \sqrt[3]{m_3(y)/m_3(x)}$ as an estimator of β which is consistent under the mild condition [that $\underset{n \to \infty}{\text{plim}}\ m_3(x) \neq 0$ [or simply $\mu_3(X) \neq 0$].

For each pair of equations from the set (vi) to (ix) we get a separate set of estimators. Thus, for estimation of β we have six choices :

$$\hat{\beta}_1 = m_{03}/m_{12}, \quad \hat{\beta}_2 = m_{12}/m_{21}, \quad \hat{\beta}_3 = m_{21}/m_{30}$$ 2.2/

$$\hat{\beta}_4 = \sqrt[3]{m_{03}/m_{30}}, \quad \hat{\beta}_5 = \pm\sqrt{m_{03}/m_{21}}, \quad \hat{\beta}_6 = \pm\sqrt{m_{12}/m_{30}}$$

The choice of signs for $\hat{\beta}_4$ and $\hat{\beta}_5$ can be based on the sign of any one of the other four estimates. Obviously from eqs. (vi) to (ix) it follows that each of the estimators is consistent if $\mu_3(X) \neq 0$. However, for the estimators $\hat{\beta}_1$, $\hat{\beta}_2$ and $\hat{\beta}_5$ we have to assume, in addition, that $\beta \neq 0$.2.3/

The first three estimators may be regarded as basic estimators. All other estimators, based only on moments up to third order, must be functions of these three estimators. Thus,

$$\hat{\beta}_4 = \sqrt[3]{\hat{\beta}_1 \cdot \hat{\beta}_2 \cdot \hat{\beta}_3}, \quad \hat{\beta}_5 = \pm\sqrt{\hat{\beta}_1 \cdot \hat{\beta}_2}, \quad \hat{\beta}_6 = \pm\sqrt{\hat{\beta}_2 \cdot \hat{\beta}_3}.$$

2.2/ Durbin (1954).

2.3/ If $\beta = 0$, OLS estimate $\hat{\beta}_0$ is a consistent estimate of β and the asymtotic variance is $O(1/n)$ — like that of $\hat{\beta}_3$ — and this can be used to test $H_0 : \beta = 0$ in large samples. However, $\beta \neq 0 \Leftrightarrow x$ and y are fully independent (in the present model) and one can test $H_0 : \beta = 0$ in finite samples using rank correlation methods or through the t-test for sample correlation coefficient assuming conditional distributions of y given x to be normal.

In fact we can find infinitely many consistent estimates forming weighted arithmetic or geometric means of these estimators. More generally, suppose

$$\hat{\beta}_s = f(\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3),$$

so that

$$f(c\hat{\beta}_1, c\hat{\beta}_2, c\hat{\beta}_3) = cf(\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3) \text{ for all } c \neq 0$$

and

$$f(1, 1, 1) = 1$$

then $\hat{\beta}_s$ is consistent since $\hat{\beta}_1$, $\hat{\beta}_2$ and $\hat{\beta}_3$ are consistent. Note that Scott's (1950) estimate 2.4/ can be shown to belong to this class.

It may be mentioned that only $\hat{\beta}_2$ out of the estimators mentioned above is a member of the class of estimators proposed by Geary (1942).

2.3 Comparative Asymptotic Variances in the General Case

Each of the above six estimators under mild conditions have asymptotic normal distribution. Asymptotic variances of the six estimators can easily be obtained. If in addition u and v are normally distributed the expressions for the asymptotic variances reduce to the following :

2.4/ Scott proved that $m_{03} - 3bm_{12} + 3b^2m_{21} - b^3m_{30} = 0$ has a root which will be a consistent estimator of β. But she does not give any method to find out the particular root which will be consistent. Both Scott and Drion assumed that $\sigma_\varepsilon^2 = 0$. But obviously if $\sigma_\varepsilon^2 > 0$ the approaches remain valid; only the estimate of σ_v^2 now estimates $\sigma_v^2 + \sigma_\varepsilon^2$. However, Scott also assumed that u and v are normally distributed which is not necessary for the consistency property. Symmetry of u and v serves the purpose.

$$V_1 = R\{(C_u + C_{\varepsilon'})(b_2 - 1) + 4C_uC_{\varepsilon'} + 8C^2_{\varepsilon'} + 2C_uC^2_{\varepsilon'} + 6C^3_{\varepsilon'}\}$$

$$V_2 = R\{(C_u + C_{\varepsilon'})(b_2 - 1) + 2C^2_u + 2C_{\varepsilon'}C^2_u + 2C^2_{\varepsilon'} + 2C_uC^2_{\varepsilon'}\}$$

$$V_3 = R\{(C_u + C_{\varepsilon'})(b_2 - 1) + 8C^2_u + 6C^3_u + 4C_uC_{\varepsilon'} + 2C_{\varepsilon'}C^2_u\}$$

$$V_4 = R\{(C_u + C_{\varepsilon'})(b_2 - 1) + 2C^2_u + \tfrac{2}{3}C^3_u + 2C^2_{\varepsilon'} + \tfrac{2}{3}C^3_{\varepsilon'}\}$$

$$V_5 = R\{(C_u + C_{\varepsilon'})(b_2 - 1) + 0.5C^2_u + C_uC_{\varepsilon'} + 0.5C_{\varepsilon'}C^2_u + 4.5C^2_{\varepsilon'} + 1.5C^3_{\varepsilon'}\}$$

$$V_6 = R\{(C_u + C_{\varepsilon'})(b_2 - 1) + 4.5C^2_u + 1.5C^3_u + C_uC_{\varepsilon'} + 0.5C^2_{\varepsilon'}C_u + 0.5C^2_{\varepsilon'}\}$$

and for OLS estimator we have

$$V_o = R'\,[\,C_u\{b_2C_u + 1 - C_u + C^2_u\} + C_{\varepsilon'}(1 + C_u)^3\,]$$

where

$$R = \beta^2/(nb_1),\quad C_u = \sigma^2_u/\sigma^2_X,\quad C_{\varepsilon'} = \sigma^2_{\varepsilon'}/(\beta^2\sigma^2_X) = (\sigma^2_\varepsilon + \sigma^2_v)/(\beta^2\sigma^2_X)$$

and

$$b_1 = \mu^2_3(X)/\sigma^6_X,\quad b_2 = \mu_4(X)/\sigma^4_X \text{ and } R' = \beta^2/\{n(1+C_u)^4\}$$

Three interesting specific cases may be investigated here.

Case 1 : $C_u = 0$. Here OLS estimation is optimal. In this case

$$V_o \leqq V_3 \leqq V_6 \leqq V_2 \leqq V_4 \leqq V_5 \leqq V_1$$

The equality between V_o and V_3 holds iff $b_2 - b_1 - 1 = 0$, i.e., if the variable X takes only two distinct values. V_1 to V_6 are all equal iff $C_{\varepsilon'} = 0$.

Case 2 : $C_{\varepsilon'} = 0$. Here also we get straightforward inequalities between V_1 to V_6 as follows :

$$V_1 \leqq V_5 \leqq V_2 \leqq V_4 \leqq V_6 \leqq V_3.$$

In general we cannot say which one of V_1 and V_0 is larger. But there are cases for which we can say something such as the following theorem :

Theorem 2.1 : $V_0 < V_1$ if $b_1 < 4 + 5C_u + 4C_u^2 + C_u^3$

Proof. Writing $g(b_1, b_2) = V_0/V_1$ we can easily see that

$$\max_{b_2} g(b_1, b_2) < 1 \Longleftrightarrow b_1 < 4 + 5C_u + 4C_u^2 + C_u^3 \qquad \text{Q.E.D.}$$

$C_{\varepsilon'} = 0$ is the structural case where the regressand is free from error. Hence reverse least squares yields MLE

$$\hat{\beta}_7 = m_{02}/m_{11}$$

asymptotic variance of which is, in general

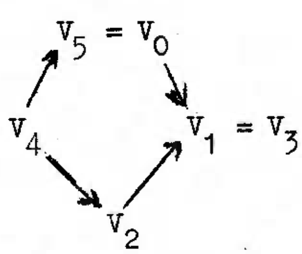
$$V_7 = (\beta^2/n)\left[C_{\varepsilon'}\{C_{\varepsilon'}/b_2 + 1 - C_{\varepsilon'} + C_{\varepsilon'}^2\} + C_u(1 + C_{\varepsilon'})^3 \right]$$

The efficiency of $\hat{\beta}_1$ relative to $\hat{\beta}_7$, in the present case, is

$$E(\hat{\beta}_1 | C_{\varepsilon'} = 0) = b_1/(b_2 - 1).$$

Observe that this is the efficiency of $\hat{\beta}_3$ relative to $\hat{\beta}_0$ where $C_u = 0$ (section 4).

Case 3 : $C_u = C_{\varepsilon'} = C$ (say). If $C_u = C_{\varepsilon'}$ the relative magnitudes of the variances can be shown diagrammatically as under



where '$\longrightarrow$' means '$\leq$' .

Here the MLE is

$$\hat{\beta}_8 = \pm\sqrt{m_{02}/m_{20}}\ ,$$

where the sign depends on the sign of m_{11}. The asymptotic variance of $\hat{\beta}_8$ can be shown to be, in general

$$V_8 = \frac{\beta^2}{4n(1 + C_{\varepsilon'})(1 + C_u)^3} \left[(b_2 - 1)\ (C_u - C_{\varepsilon'})^2 + 2C_{\varepsilon'}(2 + C_u)(1 + C_{\varepsilon'})^2 + 2C_u(2 + C_{\varepsilon'})(1 + C_u)^2 \right]$$

which reduces, in the present case, to

$$V_8 = \beta^2 C(2 + C)/(1 + C)^2 n$$

So, efficiency of $\hat{\beta}_4$ relative to $\hat{\beta}_8$ is

$$E(\hat{\beta}_4 | C_u = C_{\varepsilon'} = C) = \frac{(2+C)b_1}{2(1+C)^2\ \{(b_2-1) + 2C + \frac{2}{3}C_2\}}$$

$$\simeq b_1/(b_2 - 1),\ \text{ for small } C.$$

General Conclusion : If error in X is zero then use the estimator which involves the highest powers of x, i.e., $\hat{\beta}_3$ ($= m_{21}/m_{30}$). If error in Y is zero then it is best to use $\hat{\beta}_1$ ($= m_{03}/m_{12}$) where the y values have the highest power. If the two relative errors are equal then use $\hat{\beta}_4$ ($= \sqrt[3]{m_{03}/m_{30}}$) which have equal influences of x and y.

<u>General Case</u> : $C_u \geq 0$, $C_{\varepsilon'} \geq 0$. It is clear that V_1 and V_3 are symmetric in the sense that

$$V_1(C_u, C_{\varepsilon'}) = V_3(C_{\varepsilon'}, C_u),$$

i.e., from V_1 we get V_3 simply by interchanging roles of C_u and $C_{\varepsilon'}$. This is also true for V_5 and V_6. Moreover

(i) $V_1 < V_3$ if $C_{\varepsilon'} < C_u$ and (ii) $V_5 < V_6$ if $C_{\varepsilon'} < C_u$
We als- note that if $C_u > C_{\varepsilon'}$, then $V_3 = \max(V_1, V_2, \ldots, V_6)$; and if $C_u < C_{\varepsilon'}$, then $V_1 = \max(V_1, V_2, \ldots, V_6)$

2.4 Efficiency of $\hat{\beta}_3$ Where OLS is Valid :

We have seen that if $C_u = 0$, $\hat{\beta}_3$ is the best among the six moment estimators. In order to study the efficiency of $\hat{\beta}_3$ relative to the OLS estimator, we assume plausible forms for the distribution of X.

Asymptotic efficiency of $\hat{\beta}_3$ relative to $\hat{\beta}_0$, if $\sigma_u^2 = 0$, is

$$E(\hat{\beta}_3 \mid \sigma_u^2 = 0) = b_1/(b_2 - 1).$$

Observe that the efficiency does not depend on $C_{\varepsilon'}$, and is a function of b_1 and b_2 only of the distribution of the true regressor.

Lognormal Distribution : Lex $X \sim \Lambda(\mu, \sigma^2)$. For simplicity take $\mu = 1$, since efficiency remains unaffected by change of μ.

$$\therefore E(\hat{\beta}_3 \mid \sigma_u^2 = 0) = \frac{(w^2 - 1)(w^2 + 2)^2}{w^8 + 2w^6 + 3w^4 - 4}$$

where $w = \exp(\sigma^2/2)$. Denoting $E(\hat{\beta}_3 \mid \sigma_u^2 = 0)$ by $E(\sigma^2)$ we may present few of the values computed :

$\lim_{\sigma^2 \to 0} E(\sigma^2) = 0$, $E(0.01) = 0.042$, $E(0.1) = 0.263$

$E(0.5) = 0.421$, $E(1) = 0.339$, $E(2) = 0.143$

Efficiency of $\hat{\beta}_3$ increases from zero to slightly over 0.42 reaching a peak between $\sigma^2 = 0.5$ and 0.6 and then slowly decreases to zero.

It may be mentioned here that for empirical size distributions of population by per capita household consumption expenditure estimated for rural and urban India from different roundsof NSS the fitted LN distributions have σ^2 in the region of (0.25, 0.5) corresponding to Lorenz ratio in the range of (0.28, 0.38) (Roy and Dhar, 1961).

Gamma Distribution : Let $X \sim G(\alpha, p)$. Putting $\alpha = 1$ to simplify calculations, we find the efficiency of $\hat{\beta}_3$ as

$$E(p) = 2/(p + 3).$$

Efficiency is maximum, 2/3, when p= 0 and decreases to zero as $p \longrightarrow \infty$.

Salem and Mount (1974) fitted Gamma distribution to personal income data for the United States for the years 1960-69, and found that the parameter p lay in the interval (1.94, 2.51) so that the asymptotic efficiency of $\hat{\beta}_3$ falls in the interval (0.365, 0.405).

2.5 Comparative Asymptotic Efficiencies Where $X \sim \Lambda(\mu, \sigma^2)$:

We now examine the comparative asymptotic efficiencies of $\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_6$ in the general case where neither C_u nor $C_{\varepsilon'}$ is necessarily zero. We assume that X is lognormally distributed, which is realistic in engel curve analysis in many countries (vide Aitchison and Brown, 1957, Bhattacharya and Iyengar, 1961, Roy andDhar, 1960, Iyengar, 1967). It should be noted that if $C_u \neq 0$, the OLS estimator is inconsistent, while $\hat{\beta}_1$ to $\hat{\beta}_6$ are consistent.

Symbolically, let $X \sim \Lambda(\mu, \sigma^2)$, and

$$E_i = V_o/V_i = \text{the efficiency of } \hat{\beta}_i \text{ relative to } \hat{\beta}_o \quad \ldots(2.3)$$

Obviously E_i is a function of σ^2, C_u and $C_{\varepsilon'}$. We calculated the asymptotic efficiency of each estimator as defined in (2.3) for each combination of values of σ^2, C_u and $C_{\varepsilon'}$,

C_u : 0, 0.01, 0.02, 0.05, 0.07, 0.1, 0.15, 0.2, 0.5, 1.0

$C_{\varepsilon'}$: 0, 0.01, 0.05, 0.1, 0.2, 0.5, 1.0, 2.0, 5.0, 10.0

σ^2 : 0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0, 1.5, 2.0, 5.0[2.5/].

The results are presented briefly through Table (2.1) and Fig. (2.1).

(a) As could be expected, the ranking of estimators is independent of σ^2, though the actual asymptotic efficiency is influenced by σ^2.

(b) When $C_{\varepsilon'}$ is zero, $\hat{\beta}_1$ ranks first. As $C_{\varepsilon'}$ increases (C_u remaining constant) $\hat{\beta}_5$, $\hat{\beta}_4$, $\hat{\beta}_6$ and then $\hat{\beta}_3$ take the first rank sequentially. Fig.2.1 gives a broad picture of their relative positions for different combinations of C_u and $C_{\varepsilon'}$.

(c) We must note that as σ^2 increases asymptotic efficiencies of the six estimators approach equality and for $\sigma^2 > 0.7$ the estimators are practically equally efficient.

(d) It is seen from the table that when $C_{\varepsilon'}$ and σ^2 are held constant, the efficiency of the various $\hat{\beta}_i$ relative to OLS does not always improve as C_u increases. The picture would of course be different if MSE were considered.

2.5/ These values of σ^2 correspond to LR's ranging from 0.06 to 0.89.

Table 2.1 : Asymptotic efficiencies of $\hat{\beta}_i$ $(i = 1, 2, \ldots, 6)$ with respect to $\hat{\beta}_0$ for different values of σ^2, C_u and C_ε

		$\hat{\beta}_1$			$\hat{\beta}_2$			$\hat{\beta}_3$		
C_u	C_ε / σ^2	0.1	0.5	1.0	0.1	0.5	1.0	0.1	0.5	1.0
0.0	0.0	-	-	-	-	-	-	-	-	-
	0.1	0.22	0.40	0.34	0.25	0.42	0.34	0.26	0.42	0.34
	0.5	0.11	0.33	0.32	0.21	0.40	0.34	0.26	0.42	0.34
	1.0	0.06	0.25	0.30	0.17	0.38	0.33	0.26	0.42	0.34
0.1	0.0	0.25	0.88	2.85	0.24	0.87	2.84	0.21	0.85	2.83
	0.1	0.21	0.61	2.65	0.23	0.63	1.57	0.21	0.61	1.57
	0.5	0.11	0.37	1.25	0.19	0.45	0.73	0.21	0.46	0.73
	1.0	0.05	0.26	0.79	0.16	0.29	0.53	0.22	0.42	0.54
0.5	0.0	0.17	0.96	3.86	0.13	0.91	3.83	0.07	0.76	3.68
	0.1	0.15	0.82	3.24	0.13	0.81	3.23	0.07	0.68	3.12
	0.5	0.08	0.52	1.97	0.12	0.58	2.02	0.08	0.52	1.97
	1.0	0.04	0.33	1.31	0.10	0.45	1.41	0.09	0.43	1.40
1.0	0.0	0.10	0.59	2.43	0.06	0.54	2.39	0.02	0.35	2.17
	0.1	0.09	0.55	2.22	0.07	0.51	2.19	0.02	0.34	1.90
	0.5	0.06	0.40	1.63	0.06	0.41	1.64	0.03	0.30	1.53
	1.0	0.03	0.27	1.20	0.06	0.34	1.26	0.03	0.27	1.30

contd...../-

Table 2.1 (contd.)

C_u	C_ε/σ^2	$\hat{\beta}_4$ 0.1	$\hat{\beta}_4$ 0.5	$\hat{\beta}_4$ 1.0	$\hat{\beta}_5$ 0.1	$\hat{\beta}_5$ 0.5	$\hat{\beta}_5$ 1.0	$\hat{\beta}_6$ 0.1	$\hat{\beta}_6$ 0.5	$\hat{\beta}_6$ 1.0
0.0	0.0	-	-	-	-	-	-	-	-	-
	0.1	0.25	0.42	0.34	0.24	0.41	0.34	0.26	0.42	0.34
	0.5	0.20	0.40	0.34	0.16	0.37	0.33	0.25	0.42	0.34
	1.0	0.16	0.37	0.33	0.10	0.33	0.32	0.23	0.41	0.34
0.1	0.0	0.24	0.87	2.84	0.25	0.88	2.85	0.22	0.86	2.84
	0.1	0.23	0.63	1.57	0.23	0.62	1.57	0.23	0.62	1.57
	0.5	0.19	0.44	0.73	0.15	0.42	0.72	0.22	0.46	0.73
	1.0	0.15	0.38	0.53	0.10	0.34	0.51	0.21	0.42	0.54
0.5	0.0	0.13	0.91	3.82	0.16	0.95	3.85	0.10	0.85	3.77
	0.1	0.13	0.81	3.23	0.15	0.83	3.24	0.11	0.76	3.19
	0.5	0.13	0.59	2.02	0.12	0.57	2.01	0.12	0.57	2.01
	1.0	0.11	0.46	1.41	0.08	0.42	1.38	0.12	0.47	1.42
1.0	0.0	0.06	0.52	2.38	0.09	0.58	2.42	0.04	0.46	2.31
	0.1	0.06	0.50	2.18	0.09	0.54	2.22	0.04	0.44	2.12
	0.5	0.07	0.42	1.65	0.08	0.43	1.65	0.05	0.38	1.61
	1.0	0.07	0.36	1.27	0.06	0.34	1.26	0.06	0.34	1.26

In table 2.2 we present some typical situations in Engel curve analysis.

Table 2.2 : Efficiencies of $\hat{\beta}_i$ (i = 1, 2, ..., 6) with respect to $\hat{\beta}_o$: some typical situations

C_u	σ^2	$C_{\varepsilon'}$	Efficiency of $\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$	$\hat{\beta}_6$
		0.5	0.306	0.391	0.411	0.389	0.357	0.409
	0.4	1	0.209	0.359	0.403	0.347	0.292	0.394
		2	0.108	0.314	0.399	0.278	0.200	0.376
0.05								
		0.5	0.341	0.409	0.424	0.407	0.383	0.422
	0.5	1	0.250	0.378	0.411	0.370	0.324	0.404
		2	0.140	0.340	0.404	0.310	0.238	0.388

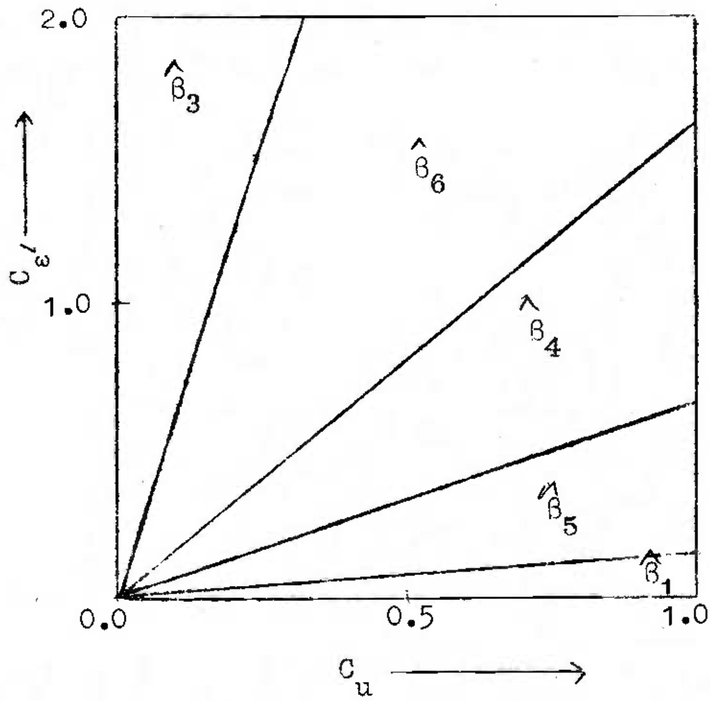


Fig.2.1. Showing best estimator for different combinations of C_u and $C_{\varepsilon'}$.

2.6 Estimation Via Cumulants :

We first give some results on bivariate cumulants which throw up a series of estimators of β including some of those proposed earlier.

Let $K(m,n)$ or $K_{m,n}(X,Y)$ denote bivariate cumulant of order (m,n) of the joint distribution of (X,Y) : $K'(m,n)$ denotes the same quantity for (x,y) 2.6/. The following theorem was proved by Geary (1942) :

Theorem 2.2. Suppose (u, v) is jointly independent of (X, Y). Also suppose that u is independent of v. Then

$$\hat{\beta} = \hat{K}'(c_1, c_2 + 1)/ \hat{K}'(c_1 + 1, c_2), \quad c_1 \text{ and } c_2 > 0,$$

is a consistent estimator of β if $K(c_1 + 1, c_2) \neq 0$.

The only estimator of this type based on cumulants of order three is $\hat{K}'_{12}/ \hat{K}'_{21}$. If in addition to above, u and v are symmetrically distributed then there exist two more estimators via cumulants of order three, namely,

$$\hat{K}'_{03} / \hat{K}'_{12} \quad \text{and} \quad \hat{K}'_{21}/ \hat{K}'_{30}$$

which are consistent under the same condition. Observe that $K'_{ij} = \mu_{ij}$ for $i + j = 3$. Hence these estimators are nothing but our $\hat{\beta}_1$, $\hat{\beta}_2$ and $\hat{\beta}_3$ considered in section 2.2.

Evidently, our method of estimation via third-order cumulants fails when X is symmetrically distributed. For a symmetric distribution of X, provided $K_4(X) \neq 0$, estimation via fourth-order cumulants is possible. In general, the following results follow from properties of cumulants listed in appendix:

2.6/ Properties of cumulants are stated in the appendix (2.1).

(1) For $r > 2$

$$K'(0,r) = \beta K'(1, r-1) + K_r(v).$$

Now, if v is symmetric and r is odd (≥ 3) then $K_r(v) = 0$. Hence $\hat{\beta} = \hat{K}'(0, r)/\hat{K}'(1, r-1)$ is consistent if $K(1, r-1) \neq 0$.

(2) Similarly, if u is symmetric and r is odd (≥ 3) then $\hat{\beta} = \hat{K}'(r-1, 1)/\hat{K}'(r, 0)$ is consistent if $K(r, 0) \neq 0$.

(3) If u is normally distributed then $K_r(u) = 0$ for $r \geq 3$. Hence $\hat{\beta} = \hat{K}'(r-1, 1)/\hat{K}'(r, 0)$ is consistent for any $r \geq 3$ if $K'(r, 0) \neq 0$.

(4) If v is normally distributed then $K_r(v) = 0$, for $r \geq 3$. Hence $\hat{\beta} = \hat{K}'(0, r)/\hat{K}'(1, r-1)$ is consistent for any $r \geq 3$ if $K(1, r-1) \neq 0$.

In particular, in addition to Geary's estimators <u>via</u> fourth-order cumulants, we may have two more consistent estimators if we assume that u and v are normally distributed.

$$\hat{K}'(3, 1)/\hat{K}'(4, 0) \quad \text{and} \quad \hat{K}'(0, 4)/\hat{K}'(1, 3),$$

both estimators being consistent if $K(4, 0) \neq 0$. For the second estimator we must have in addition $\beta \neq 0$.

2.7 <u>Some Extensions and Comments</u> :

We may consider the case of m regressors each subject to error besides the regressand, making the same assumptions as in the previous sections. Our model is

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_m X_m + \varepsilon \qquad \dots \quad (2.4)$$

$$x_i = X_i + u_i \qquad i = 1, 2, \dots, m \qquad \dots \quad (2.5)$$

$$y = Y + v \qquad \dots \quad (2.6)$$

Suppose

$$Y' = Y - \bar{Y} \quad \text{and} \quad X_i' = X_i - \bar{X}_i, \qquad i = 1, 2, \ldots, m,$$

and similarly

$$y' = y - \bar{y} \quad \text{and} \quad x_i' = x_i - \bar{x}_i, \qquad i = 1, 2, \ldots, m;$$

v and u_i's are assumed to be symmetrically distributed.

Then

$$E(y' x_i'^2) = \sum_{j=1}^{m} \beta_j \, E(x_j' \, x_i'^2), \quad i = 1, 2, \ldots, m,$$

Hence

$$A\beta = B$$

where

$$A = ((a_{ij})) \quad \text{with } a_{ij} = E(x_i'^2 \, x_j'),$$

$$\beta' = (\beta_1, \beta_2, \ldots, \beta_m),$$

and

$$B' = (E(y' x_1'^2), \ldots, E(y' x_m'^2)).$$

$\therefore$ $\hat{\beta} = \hat{A}^{-1} \hat{B}$ is consistent provided that $|A| \neq 0$, where $\hat{A}$ and $\hat{B}$ are consistent estimates of A and B.

In many empirical studies we must relax the assumption that u and v are independent. We may take (u, v) to be bivariate normally distributed with unknown correlation coefficient ρ. Interestingly enough, our estimation procedure does not differ at all in either way. Both sets of assumptions give us same estimate of β. This is due to the peculiarity of bivariate normal distribution which says that $K_{ij}(u, v)$, for $i + j > 2$, is zero. The multivariate extension of the problem of correlated errors is also similar. This is due to the fact that marginally each u_i is normally distributed and each pair (u_i, u_j) is bivariate normal.

The method of estimation <u>via</u> cumulants originally introduced by Geary may be looked upon as application of the IV method; instruments being taken from x and y itself. As for example, the three third-order cumulant estimators may be viewed to have instruments $z_1 (= y'^2)$, $z_2 (= x'y')$ and $z_3 (= x'^2)$ respectively.

2.8 Conclusion :

This chapter examines the possibilities of moment/cumulant based estimators of the kind first proposed by Geary. It proposes some new estimators of that class which have smaller asymptotic variances in some specific situations. It also compares the asymptotic efficiencies of various estimators based on third-order cumulants and finds the best estimator in that class in different regions of the parametric space assuming lognormality of the regressor which is realistic for some economic data, e.g., in the Engel curve analysis. Efficiencies of these estimators relative to OLS have been investigated.

As soon as errors affect observations on the regressor, comparison of the variances of these six estimators ($\hat{\beta}_1, \ldots, \hat{\beta}_6$) with the variance of the OLS estimator ($\hat{\beta}_0$) does not seem to be justifiable on the ground that the OLS estimator is biased and inconsistent. So the two MSE's should be compared and since the $\hat{\beta}_i$'s $(i = 1, 2, \ldots, 6)$ are consistent, the efficiency of the $\hat{\beta}_i$'s relative of $\hat{\beta}_0$ based on the MSE criterion goes to infinity as $n \to \infty$.

The choice of one out of the class of estimators mentioned in this chapter may be difficult in many situations, because this requires estimation of the variance of every estimator. It is well-known that standard errors of estimators of cumulants generally increase with their order (Madansky, 1959, Geary, 1942). Hence one should not take higher-order cumulants when estimation is possible by taking lower-order cumulants. So if X is asymmetric ($\mu_3(X) \neq 0$) one should base the estimate on cumulants of order three. The assumption that $\mu_3(X) \neq 0$ is very important in this case. If X is symmetric non-normal (so that $K_4(X) \neq 0$) then one should use fourth-order cumulants.

Geary admitted that this method is inapplicable if X is normally distributed. In fact, if all the cumulants of order three or more vanish, then one can conclude that either (i) the variates X and Y are independent or (ii) they are normally distributed. It may be recalled that if (X, Y, u, v, ε) are normally distributed, then the parameter, β, of this model is not identifiable (Reiersol, 1950).

The assumption of independence of true and error components may sometimes be inappropriate. Many economic variables like income or consumer expenditure are seasonally affected and if the reference period of the enquiry is short (say, a month preceding date of interview) then the error components are likely to be dependent on the true components. This problem is considered in Chapters 4 and 5.

In general, methods of tackling econometric problems with errors in variables should depend heavily on what is known about these errors. To reiterate Morgenstern (1963) : 'As long as theory has not been sufficiently developed to cover the complicated cases of many simultaneous sources of error and their shifting nature of interdependence, one must proceed on an heuristic and common sense basis'.

Appendix 2.1 : Properties of Cumulants

Property 1 : Cumulants are invariant under changes of origin, except the first ($K_1 = \mu_1'$).

Property 2 : If the variate values are multiplied by a constant C, K_r is multiplied by C^r.

Property 3 : The cumulant of a sum of independent variables is the sum of the cumulants of the variables.

Property 4 : K_r ($r \geqq 3$) = 0 for normal distribution.

Property 5 : The bivariate cumulant K_{c_1, c_2} ($c_1 > 0$, $c_2 > 0$) of independent random variables is zero.

Property 6 : If (u, v) is independent of (X, Y) then

$$K_{c_1,c_2}(X + u, Y + v) = K_{c_1,c_2}(X,Y) + K_{c_1,c_2}(u,v).$$

Property 7 : $K_{c_1,c_2}(X_1, a+bX_2)$ equals $a + bK_{c_1,c_2}(X_1, X_2)$ if $c_1 = 0$ and $c_2 = 1$, and $b^{c_2} K_{c_1,c_2}(X_1, X_2)$ otherwise.

Property 8 : Any odd ($\geqq 3$) order cumulant of a symmetric distribution is zero.

Property 9 : For a bivariate normal distribution, the bivariate cumulant $K_{c_1,c_2} = 0$ if $c_1 + c_2 > 2$.

Appendix 2.2 : Derivation of Asymptotic Variances

$$\mu_{20} = \mu_2(X) + \mu_2(u) = \mu_2(x)$$

$$\mu_{11} = \beta\mu_2(X) = \beta\mu_2(x) - \beta\mu_2(u)$$

$$\mu_{02} = \beta^2\mu_2(X) + \mu_2(\varepsilon) + \mu_2(v) = \beta^2\mu_2(X) + \mu_2(\varepsilon')$$

$$\mu_{40} = K_4(x) + 3K_2^2(x) = K_4(X) + K_4(u) + 3\left\{K_2(X) + K_2(u)\right\}^2$$

$$= \mu_4(X) + \mu_4(u) + 6\mu_2(X)\,\mu_2(u)$$

$$\mu_{31} = K_{31} + 3K_{20}\,K_{11} = K_{31}(X,Y) + K_{31}(u,\varepsilon') + 3\left\{\mu_2(X) + \mu_2(u)\right\}\beta\mu_2(X)$$

$$= \beta\mu_4(X) + 3\beta\mu_2(X)\,\mu_2(u)$$

$$\mu_{22} = K_{22} + K_{20}\,K_{02} + 2K_{11}^2$$

$$= K_{22}(X,Y) + K_{20}\,K_{02} + 2K_{11}^2 + \left\{K_{22}(u,\varepsilon') + K_{22}(X,\varepsilon') + K_{22}(u,Y)\right\}$$

$$= \beta^2K_4(X) + \left\{\mu_2(X) + \mu_2(u)\right\}\left\{\beta^2\mu_2(X) + \mu_2(\varepsilon')\right\} + 2\beta^2\mu_2^2(X)$$

$$= \beta^2\mu_4(X) + \beta^2\mu_2(X)\,\mu_2(u) + \mu_2(\varepsilon')\left\{\mu_2(X) + \mu_2(u)\right\}$$

$$\mu_{13} = K_{13} + 3K_{02}\,K_{11}$$

$$= \beta^3K_4(X) + 3\beta\mu_2(X)\left\{\beta^2\mu_2(X) + \mu_2(\varepsilon')\right\}$$

$$= \beta^3\mu_4(X) + 3\beta\mu_2(X)\,\mu_2(\varepsilon')$$

$$\mu_{04} = K_{04} + 3K_{02}^2$$

$$= K_{04}(X,Y) + 3\left\{\beta^4\mu_2^2(X) + \mu_2^2(\varepsilon') + 2\beta^2\mu_2(X)\mu_2(\varepsilon')\right\} + K_4(\varepsilon')$$

$$= \beta^4\mu_4(X) + \mu_4(\varepsilon') + 6\beta^2\mu_2(X)\mu_2(\varepsilon')$$

$$\mu_{50} = K_{50} + 10\ K_{30}\ K_{20}$$

$$= K_5(X) + 10\ K_3(X) \{ K_2\ (X) + K_2\ (u) \}$$

$$= \mu_5(X) + 10\mu_3\ (X)\ \mu_2(u)$$

$$\mu_{41} = K_{41} + 4K_{30}K_{11} + 6K_{21}\ K_{20}$$

$$= K_{41}(X,Y) + 4K_3(X)\ \beta\mu_2(X) + 6\beta K_3(X) \{ \mu_2(X) + \mu_2(u) \}$$

$$= \beta\mu_5(X) + 6\beta\mu_3(X)\ \mu_2(u).$$

$$\mu_{32} = K_{32} + K_{30}K_{02} + 6K_{21}K_{11} + 3K_{20}\ K_{12}$$

$$= \beta^2 K_5(X) + K_3(X) \{ \beta^2\mu_2(X) + \mu_2(\varepsilon') \} + 6\beta K_3(X)\beta\mu_2(X)$$

$$+ 3 \{ \mu_2(X) + \mu_2(u) \}\ \beta^2 K_3(X)$$

$$= \beta^2\mu_5(X) + 3\beta^2\mu_3(X)\ \mu_2(u) + \mu_3(X)\ \mu_2(\varepsilon')$$

$$\mu_{23} = K_{23} + K_{03}K_{20} + 6K_{12}K_{11} + 3K_{02}\ K_{21}$$

$$= \beta^3 K_5(X) + \beta^3 K_3(X) \{ \mu_2(X) + \mu_2(u) \} + 6\beta^3 K_3(X)\mu_2(X)$$

$$+ 3\beta K_3(X) \{ \beta^2\mu_2(X) + \mu_2(\varepsilon') \}$$

$$= \beta^3\mu_5(X) + \beta^3\mu_3(X)\mu_2(u) + 3\beta\mu_3(X)\mu_2(\varepsilon')$$

$$\mu_{14} = K_{14} + 4K_{03}K_{11} + 6K_{12}K_{02}$$

$$= \beta^4 K_5(X) + 4\beta^3 K_3(X)\ \beta\mu_2(X) + 6\beta^2 K_3(X) \{ \beta^2\mu_2(X) + \mu_2(\varepsilon') \}$$

$$= \beta^4\mu_5(X) + 6\beta^2\mu_3(X)\ \mu_2(\varepsilon')$$

$$\mu_{05} = K_{05} + 10K_{03}K_{02}$$

$$= \beta^5 K_5(X) + 10\beta^3 K_3(X) \{ \beta^2\mu_2(X) + \mu_2(\varepsilon') \}$$

$$= \beta^5\mu_5(X) + 10\beta^3\mu_3(X)\ \mu_2(\varepsilon')$$

$$\mu_{60} = K_{60} + 15K_{40}K_{20} + 10K_{30}^2 + 15K_{20}^3$$

$$= K_6(X) + K_6(u) + 15\{K_4(X) + K_4(u)\}\{K_2(X) + K_2(u)\}$$

$$+ 10K_3^2(X) + 15\{K_2(X) + K_2(u)\}^3$$

$$= \mu_6(X) + K_6(u) + 15K_4(u)K_2(u) + 15K_2^3(u) + 15K_4(X)K_2(u)$$

$$+ 15K_2(X)K_4(u) + 45K_2^2(X)K_2(u) + 45K_2(X)K_2^2(X).$$

$$= \mu_6(X) + \mu_6(u) + 15\mu_4(X)\mu_2(u) + 15\mu_2(X)\mu_4(u).$$

$$\mu_{51} = K_{51} + 5K_{40}K_{11} + 10K_{31}K_{20} + 10K_{30}K_{21} + 15K_{20}^2K_{11}$$

$$= \beta K_6(X) + 5\{K_4(X) + K_4(u)\}\beta\mu_2(x) + 10\beta K_4(X)$$

$$\{\mu_2(X) + \mu_2(u)\} + 15\beta\mu_2(X)\{\mu_2(X) + \mu_2(u)\}^2$$

$$= \beta\mu_6(X) + 5\beta\mu_2(X)K_4(u) + 10\beta K_4(X)\mu_2(u) + 15\beta\mu_2(X)\mu_2^2(u)$$

$$+ 30\beta\mu_2^2(X)\mu_2(u)$$

$$= \beta\mu_6(X) + 5\beta\mu_2(X)\mu_4(u) + 10\beta\mu_2(u)\mu_4(X).$$

$$\mu_{42} = K_{42} + K_{40}K_{02} + 8K_{31}K_{11} + 4K_{30}K_{12} + 6K_{22}K_{20} + 6K_{21}^2$$

$$+ 3K_{20}^2K_{02} + 12K_{20}K_{11}^2.$$

$$= \beta^2K_6(X) + \{K_4(X) + K_4(u)\}\{\beta^2\mu_2(X) + \mu_2(\varepsilon')\} + 8\beta K_4(X)\beta\mu_2(X)$$

$$+ 4K_3(X)\beta^2K_3(X) + 6\beta^2K_4(X)\{\mu_2(X) + \mu_2(u)\} + 6\beta^2K_3^2(X)$$

$$+ 3\{K_2(X) + \mu_2(u)\}^2\{\beta^2\mu_2(X) + \mu_2(\varepsilon')\} + 12\{\mu_2(X) + \mu_2(u)\}\beta^2\mu_2^2(X)$$

$$= \beta^2\mu_6(X) + \beta^2K_4(u)\mu_2(X) + K_4(u)\mu_2(\varepsilon') + K_4(X)\mu_2(\varepsilon') + 6\beta^2K_4(X)\mu_2(u)$$

$$+ 3\mu_2(\varepsilon')\{\mu_2^2(X) + 2\mu_2(X)\mu_2(u) + \mu_2^2(u)\} + \beta^2\, 6\mu_2^2(X)\mu_2(u)$$

$$+ 3\beta^2\mu_2(X)\mu_2^2(u) + 12\beta^2\mu_2^2(X)\mu_2(u)$$

$$= \beta^2\mu_6(X) + \beta^2\mu_2(X)\mu_4(u) + \mu_2(\varepsilon')\mu_4(u) + \mu_2(\varepsilon')\mu_4(X)$$

$$+ 6\beta^2\mu_2(u)\mu_4(X) + 6\beta^2\mu_2^2(X)\mu_2(u) + 6\mu_2(X)\mu_2(u)\mu_2(\varepsilon').$$

$$\mu_{33} = K_{33}+3K_{31}K_{02}+K_{30}K_{03}+\bullet K_{22}K_{11} + 9K_{21}K_{12} + 3K_{20}K_{13} + 9K_{20}K_{11}K_{02} + 6K_{11}^{3}$$

$$= \beta^{3}K_{6}(X) + 3\beta K_{4}(X)\left\{\beta^{2}\mu_{2}(X) + \mu_{2}(\varepsilon')\right\} + 9\beta^{2}K_{4}(X)\ \ \beta\mu_{2}(X) + 9\beta^{3}K_{3}^{2}(X)$$

$$+ \beta^{3}K_{3}^{2}(X) + 3\left\{\mu_{2}(X) + \mu_{2}(u)\right\}\ \beta^{3}K_{4}(X) + 9\left\{\mu_{2}(X) + \mu_{2}(u)\right\}\beta\mu_{2}(X)$$

$$\left\{\beta^{2}\mu_{2}(X) + \mu_{2}(\varepsilon')\right\} + 6\beta^{3}\mu_{2}^{3}(X).$$

$$= \beta^{3}\mu_{6}(X) + 3\beta K_{4}(X)\mu_{2}(\varepsilon') + 3\beta^{3}\ K_{4}(X)\mu_{2}(u) + 9\beta\mu_{2}(X)\left\{\mu_{2}(X)\mu_{2}(\varepsilon')\right.$$

$$\left. + \beta^{2}\mu_{2}(u)\mu_{2}(X) + \mu_{2}(u)\mu_{2}(\varepsilon')\right\}$$

$$= \beta^{3}\mu_{6}(X) + 3\beta\mu_{4}(X)\mu_{2}(\varepsilon') + 3\beta^{3}\mu_{4}(X)\mu_{2}(u) + 9\beta\mu_{2}(X)\mu_{2}(u)\mu_{2}(\varepsilon').$$

$$\mu_{24} = K_{24} + K_{04}\ K_{20}+8K_{13}K_{11} + 4K_{03}K_{21} + 6K_{22}K_{02} + 6K_{12}^{2}+ 3K_{02}^{2}K_{20} + 12K_{02}K_{11}^{2}$$

$$= \beta^{4}K_{6}(X) + \beta^{4}K_{4}(X)\left\{\mu_{2}(X) + \mu_{2}(u)\right\} + K_{4}(\varepsilon')\left\{\mu_{2}(X) + \mu_{2}(u)\right\}$$

$$+ 8\beta^{4}K_{4}(X)K_{2}(X) + 4\beta^{4}K_{3}^{2}(X) + 6\beta^{2}K_{4}(X)\left\{\beta^{2}\mu_{2}(X) + \mu_{2}(\varepsilon')\right\}$$

$$+ \quad 6\beta^{4}K_{3}^{2}(X) + 3\left\{\beta^{2}K_{2}(X) + \mu_{2}(\varepsilon')\right\}^{2}\quad \left\{\mu_{2}(X) + \mu_{2}(u)\right\}$$

$$+ 12\left\{\beta^{2}\mu_{2}(X) + \mu_{2}(\varepsilon')\right\}\beta^{2}\mu_{2}^{2}(X)$$

$$= \beta^{4}\mu_{6}(X) + \beta^{4}K_{4}(X)\mu_{2}(u) + K_{4}(\varepsilon')\mu_{2}(u) + 6\beta^{2}K_{4}(X)\mu_{2}(\varepsilon')$$

$$+ 3\mu_{2}(u)\left\{\beta^{4}\mu_{2}^{2}(X) + \mu_{2}^{2}(\varepsilon') + 2\beta^{2}\mu_{2}(X)\mu_{2}(\varepsilon')\right\}$$

$$+ 3\mu_{2}(X)\left\{2\beta^{2}\mu_{2}(X)\mu_{2}(\varepsilon') + \mu_{2}^{2}(\varepsilon')\right\} + 12\beta^{2}\mu_{2}^{2}(X)\ \ \mu_{2}(\varepsilon')$$

$$= \beta^{4}\mu_{6}(X) + \beta^{4}\mu_{2}(u)\ \ \mu_{4}(X) + \mu_{2}(u)\ \ \mu_{4}(\varepsilon') + 6\beta^{2}\mu_{2}(\varepsilon')\mu_{4}(X)$$

$$+ 12\beta^{2}\mu_{2}(X)\ \ \mu_{2}(u)\ \ \mu_{2}(\varepsilon') + 3\mu_{2}(X)\ \ \mu_{2}^{2}(\varepsilon')$$

$$\mu_{15} = K_{15} + 5K_{04}K_{11} + 10K_{13}K_{02} + 10K_{03}K_{12} + 15K_{02}^2K_{11}$$

$$= \beta^5 K_6(X) + 5\{\beta^4 K_4(X) + K_4(\varepsilon')\}\ \beta\mu_2(X) + 10\beta^3 K_4(X)\{\beta^2\mu_2(X) + \mu_2(\varepsilon')\}$$
$$+ 10\beta^3 K_3(X)\ \beta^2 K_3(X) + 15\beta K_2(X)\{\beta^2\mu_2(X) + \mu_2(\varepsilon')\}^2$$

$$= \beta^5\mu_6(X) + 5\beta K_4(\varepsilon')\mu_2(X) + 10\beta^3 K_4(X)\mu_2(\varepsilon') + 30\beta^3\mu_2^2(X)\mu_2(\varepsilon')$$
$$+ 15\mu_2(X)\ \mu_2^2(\varepsilon')\beta$$

$$= \beta^5\mu_6(X) + 5\beta\mu_4(\varepsilon')\ \mu_2(X) + 10\beta^3\mu_2(\varepsilon')\ \mu_4(X)$$

$$\mu_{06} = K_{06} + 16K_{04}K_{02} + 10K_{03}^2 + 15K_{02}^3$$

$$= \beta^6 K_6(X) + K_6(\varepsilon') + 15\{\beta^4 K_4(X) + K_4(\varepsilon')\}\{\beta^2\mu_2(X) + \mu_2(\varepsilon')\}$$
$$+ 10\beta^6 K_3^2(X) + 15\{\beta^6\mu_2^3(X) + 3\beta^4\mu_2^2(X)\mu_2(\varepsilon') + 3\beta^2\mu_2(X)\mu_2^2(\varepsilon') + \mu_2^3(\varepsilon')\}$$

$$= \beta^6\mu_6(X) + K_6(\varepsilon') + 15K_4(\varepsilon')\ K_2(\varepsilon') + 15K_2^3(\varepsilon') + 15\beta^4 K_4(X)\mu_2(\varepsilon')$$
$$+ 15\beta^2\ K_4(\varepsilon')\ \mu_2(X) + 45\beta^4\mu_2^2(X)\mu_2(\varepsilon') + 45\beta^2\mu_2(X)\mu_2^2(\varepsilon')$$

$$= \beta^6\mu_6(X) + \mu_6(\varepsilon') + 15\beta^4\mu_2(\varepsilon')\ \mu_4(X) + 15\beta^2\mu_4(\varepsilon')\ \mu_2(X)$$

$$V(m_{12}) = \frac{1}{n}(\mu_{24} - \mu_{12}^2 + 4\mu_{02}\mu_{11}^2 + \mu_{20}\mu_{02}^2 + 4\mu_{11}^2\mu_{02} - 4\mu_{11}\mu_{13} - 2\mu_{02}\mu_{22})$$

$$= \frac{1}{n}\Big[\{K_{24} + K_{04}K_{20} + 8K_{13}K_{11} + 4K_{03}K_{21} + 6K_{22}K_{02} + 6K_{12}^2 + 3K_{02}^2K_{20}$$
$$+ 12K_{02}K_{11}^2\} - K_{12}^2 + 8K_{11}^2K_{02} + K_{20}K_{02}^2 - 4K_{11}(K_{13} + 3K_{02}K_{11})$$
$$- 2K_{02}(K_{22} + K_{20}K_{02} + 2K_{11}^2\Big]$$

$$= \frac{1}{n}(K_{24} + K_{04}K_{20} + 4K_{13}K_{11} + 4K_{03}K_{21} + 4K_{22}K_{02} + 5K_{12}^2$$
$$+ 2K_{02}^2K_{20} + 4K_{02}K_{11}^2)$$

$$= \Big[\beta^4 K_{60}(X) + \beta^4 K_{40}(X) \{ \mu_2(X) + \mu_2(u) \} + K_{04}(\varepsilon') \{ \mu_2(X) + \mu_2(u) \}$$

$$+ 4\beta^3 K_{40}\ \beta\mu_2(X) + 4\beta^3\ \beta K_{30}^2 + 4\beta^2 K_{40} \{ \beta^2\mu_2(X) + \mu_2(\varepsilon') \}$$

$$+ 5\beta^4 K_{30}^2 + 2\{ \mu_2(X) + \mu_2(u) \} \beta^4 \mu_2^2(X)$$

$$+ 4\{ \mu_2(X) + \mu_2(u) \} \beta^2 \mu_2(X)\ \mu_2(\varepsilon')$$

$$+ 2\{ \mu_2(X) + \mu_2(u) \} \mu_2^2(\varepsilon') + 4\beta^2 \mu_2^2(X) \{ \beta^2 \mu_2(X) + \mu_2(\varepsilon') \} \Big]$$

$$= \frac{1}{n} \Big[\beta^4 \{ K_{60}(X) + 9K_{40}(X)\ K_{20}(X) + 9K_{30}^2(X) + 6K_2^3(X) \}$$

$$+ \beta^4 \mu_2(u) \{ K_{40}(X) + 2\mu_2^2(X) \} + 4\beta^2 \mu_2(\varepsilon') \{ K_{40}(X) + 2\mu_2^2(X) \}$$

$$+ 4\beta^2 \mu_2(X) \mu_2(u) \mu_2(\varepsilon') + \{ \mu_2(X) + \mu_2(u) \} \{ K_4(\varepsilon') + 2\mu^2(\varepsilon') \} \Big]$$

$$V(m_{03}) = \frac{1}{n} (\mu_{06} - \mu_{03}^2 + 9\mu_{02}^3 - 6\mu_{02}\mu_{04})$$

$$= \frac{1}{n} \Big[K_{06} + 15K_{04}K_{02} + 10K_{03}^2 + 15K_{02}^3 - K_{03}^2 + 9K_{02}^3$$

$$- 6K_{02} \{ K_{04} + 3K_{02}^2 \} \Big]$$

$$= \frac{1}{n} (K_{06} + 9K_{04}K_{02} + 9K_{03}^2 + 6K_{02}^3)$$

$$= \frac{1}{n} \Big[\beta^6 K_6(X) + K_6(\varepsilon') + 9\beta^4 K_4(X) \{ \beta^2 \mu_2(X) + \mu_2(\varepsilon') \}$$

$$+ 9K_4(\varepsilon') \{ \beta^2 \mu_2(X) + \mu_2(\varepsilon') \} + 9\beta^6 K_3^2(X)$$

$$+ 6\{ \beta^6 \mu_2^3(X) + 3\beta^4 \mu_2^2(X) \mu_2(\varepsilon') + 3\beta^2 \mu_2(X) \mu_2^2(\varepsilon') + \mu_2^3(\varepsilon') \} \Big]$$

$$= \frac{1}{n} \Big[\beta^6 \{ K_6(X) + 9K_4(X)\ K_2(X) + 9K_3^2(X) + 6K_2^3(X) \}$$

$$+ \{ K_6(\varepsilon') + 9K_4(\varepsilon')\ K_2(\varepsilon') + 6K_2^3(\varepsilon') \}$$

$$+ 9\beta^4\ \mu_2(\varepsilon') \{ K_4(X) + 2\mu_2^2(X) \}$$

$$+ 9\beta^2\ \mu_2(X) \{ K_4(\varepsilon') + 2\mu_2^2(\varepsilon') \} \Big]$$

$$V(m_{21}) = \frac{1}{n}\,(\mu_{42} - \mu_{21}^2 + 8\mu_{20}\mu_{11}^2 + \mu_{02}\mu_{20}^2 - 4\mu_{31}\mu_{11} - 2\mu_{22}\mu_{20})$$

$$= \frac{1}{n}\Big[(K_{42} + K_{40}K_{02} + 8K_{31}K_{11} + 4K_{30}K_{12} + 6K_{22}K_{20} + 6K_{21}^2$$
$$+ 3K_{20}^2\,K_{02} + 12K_{20}\,K_{11}^2) - K_{21}^2 + 8K_{20}\,K_{11}^2 + K_{02}\,K_{20}^2$$
$$- 4K_{11}\,(K_{31} + 3K_{20}\,K_{11}) - 2K_{20}(K_{22} + K_{20}\,K_{02} + 2K_{11}^2)\Big]$$

$$= \frac{1}{n}\,(K_{42} + K_{40}\,K_{02} + 4K_{31}\,K_{11} + 4K_{30}\,K_{12} + 4K_{22}\,K_{20}$$
$$+ 5K_{21}^2 + 2K_{20}^2\,K_{02} + 4K_{20}\,K_{11}^2)$$

$$= \frac{1}{n}\Big[\beta^2\,\{\mu_6(X) - 6\mu_4(X)\,\mu_2(X) - \mu_3^2(X) + 9\mu_2^3(X)\}$$
$$+ \beta^2\mu_2(X)\,\{K_4(u) + 2\mu_2^2(u)\} + 4\beta^2\mu_2(u)\{K_4(X) + 2\mu_2^2(X)\}$$
$$+ \mu_2(\varepsilon')\,\{K_4(X) + K_4(u) + 2\mu_2^2(X) + 2\mu_2^2(u) + 4\mu_2(X)\,\mu_2(u)\}\Big]$$

$$= \frac{1}{n}\Big[\beta^2\,\{\mu_6(X) - 6\mu_4(X)\,\mu_2(X) - \mu_3^2(X) + 9\mu_2^3(X)\}$$
$$+ \beta^2\mu_2(X)\{\mu_4(u) - \mu_2^2(u)\} + 4\beta^2\mu_2(u)\,\{\mu_4(X) - \mu_2^2(X)\}$$
$$+ \mu_2(\varepsilon')\,\{\mu_4(X) - \mu_2^2(X) + \mu_4(u) - \mu_2^2(u) + 4\mu_2(X)\mu_2(u)\}\Big]$$

$$V(m_{30}) = \frac{1}{n}\,(\mu_6 - \mu_3^2 + 9\mu_2^3 - 6\mu_2\mu_4)$$

$$= \frac{1}{n}\Big[\{\mu_6(X) + \mu_6(u) + 15\mu_4(X)\,\mu_2(u) + 15\mu_2(X)\,\mu_4(u)\}$$
$$- \mu_3^2(X) + 9\{\mu_2^3(X) + \mu_2^3(u) + 3\mu_2^2(X)\,\mu_2(u) + 3\mu_2(X)\,\mu_2^2(u)\}$$
$$- 6\,\{\mu_2(X) + \mu_2(u)\}\,\{\mu_4(X) + \mu_4(u) + 6\mu_2(X)\,\mu_2(u)\}\Big]$$

$$= \frac{1}{n}\Big[\ \mu_6(X) - \mu_3^2(X) + 9\mu_2^3(X) - 6\mu_4(X)\ \mu_2(X)$$

$$+ \{\mu_6(u) + 9\mu_2^3(u) - 6\mu_4(u)\ \mu_2(u)\}$$

$$+ 9\mu_2(u)\{\mu_4(X) - \mu_2^2(X)\} + 9\mu_2(X)\{\mu_4(u) - \mu_2^2(u)\}\Big]$$

$$\mathrm{Cov}(m_{03}, m_{12}) = \frac{1}{n}(\mu_{15} - \mu_{03}\mu_{12} + 6\mu_{02}^2\ \mu_{11} + 3\mu_{11}\ \mu_{02}^2 - 2\mu_{04}K_{11} - 4\mu_{13}\mu_{02})$$

$$= \frac{1}{n}\Big[\ K_{15} + 5K_{04}K_{11} + 10K_{13}K_{02} + 10K_{03}K_{12} + 15K_{02}^2K_{11}$$

$$- 4K_{02}\{K_{13} + 3K_{02}K_{11}\} - 2K_{11}\{K_{04} + 3K_{02}^2\}$$

$$- K_{03}K_{12} + 9K_{02}^2\ K_{11}\Big]$$

$$= \frac{1}{n}(K_{15} + 3K_{04}K_{11} + 6K_{13}K_{02} + 9K_{03}K_{12} + 6K_{02}^2\ K_{11})$$

$$= \frac{1}{n}\Big[\beta^5\ K_6(X) + 3\beta\mu_2(X)\ \{\beta^4K_4(X) + K_4(\varepsilon')\}$$

$$+ 6\beta^3\ K_4(X)\ \{\beta^2\mu_2(X) + \mu_2(\varepsilon')\} + 9\beta^3K_3(X)\ \beta^2K_3(X)$$

$$+ 6\beta\mu_2(X)\{\beta^4\mu_2^2(X) + 2\beta^2\mu_2(X)\mu_2(\varepsilon') + \mu_2^2(\varepsilon')\}\Big]$$

$$= \frac{1}{n}\Big[\beta^5\{K_6(X) + 9K_4(X)\ K_2(X) + 9K_3^2(X) + 6K_2^3(X)\}$$

$$+ 3\beta\ \mu_2(X)\ \{K_4(\varepsilon') + 2K_2^2(\varepsilon')\}$$

$$+ 6\beta^3\ \mu_2(\varepsilon')\ \{K_4(X) + 2\mu_2^2(X)\}\Big]$$

$$\mathrm{Cov}(m_{12}, m_{21}) = \frac{1}{n}(\mu_{33} - \mu_{12}\mu_{21} + 5\mu_{20}\ \mu_{02}\ \mu_{11} + 4\mu_{11}^3 - 4\mu_{22}\mu_{11} - \mu_{13}\mu_{20} - \mu_{02}\mu_{31})$$

$$= \frac{1}{n}\Big[\{K_{33} + 3K_{31}K_{02} + K_{30}K_{03} + 9K_{22}K_{11} + 9K_{21}K_{12} + 3K_{20}K_{13}$$

$$+ 9K_{20}K_{02}K_{11} + 6K_{11}^3\} - K_{12}K_{21} + 5K_{20}K_{02}K_{11} + 4K_{11}^3$$

$$- 4K_{11}\{K_{22} + K_{20}K_{02} + 2K_{11}^2\} - K_{20}\{K_{13} + 3K_{02}\ K_{11}\}$$

$$- K_{02}\{K_{31} + 3K_{20}\ K_{11}\}\Big]$$

$$= \frac{1}{n} \left(K_{33} + 2K_{31}K_{02} + K_{30}K_{03} + 5K_{22}K_{11} + 8K_{12}K_{21} + 2K_{20}K_{13} + 4K_{20}K_{02}K_{11} + 2K_{11}^{3} \right)$$

$$= \frac{1}{n} \Big[\beta^3 K_6(X) + 2\beta K_4(X) \left\{ \beta^2 \mu_2(X) + \mu_2(\varepsilon') \right\} + \beta^3 K_3^2(X)$$

$$+ 5\beta^3 K_4(X) K_2(X) + 8\beta^3 K_3^2(X) + 2\beta^3 K_4(X) \left\{ \mu_2(X) + \mu_2(u) \right\}$$

$$+ 4\beta\mu_2(X) \left\{ \mu_2(X) + \mu_2(u) \right\} \beta^2 \mu_2(X)$$

$$+ 4\beta\mu_2(X)\, \mu_2(\varepsilon') \left\{ \mu_2(X) + \mu_2(u) \right\} + 2\beta^3 K_2^3(X) \Big]$$

$$= \frac{1}{n} \Big[\beta^3 \left\{ K_6(X) + 9K_4(X)\, K_2(X) + 9K_3^2(X) + 6K_2^3(X) \right\}$$

$$+ 2\beta\mu_2(\varepsilon') \left\{ K_4(X) + 2K_2^2(X) \right\} + 2\beta^3 \mu_2(u) \left\{ K_4(X) + 2K_2^2(X) \right\}$$

$$+ 4\beta\mu_2(X)\, \mu_2(u)\, \mu_2(\varepsilon') \Big]$$

$$\text{Cov}\,(m_{03}, m_{21}) = \frac{1}{n} \left(\mu_{24} - \mu_{03}\mu_{21} + 3\mu_{02}^2\mu_{20} + 6\mu_{11}^2\mu_{02} - 2\mu_{13}\mu_{11} - \mu_{04}\mu_{20} - 3\mu_{02}\mu_{22} \right)$$

$$= \frac{1}{n} \Big[\left(K_{24} + K_{04}K_{20} + 8K_{13}K_{11} + 4K_{03}K_{21} + 6K_{22}K_{02} + 6K_{12}^2 + 3K_{20}K_{02}^2 + 12K_{02}K_{11}^2 \right) - K_{03}K_{21}$$

$$+ 3K_{02}^2 K_{20} + 6K_{11}^2 K_{02} - 2K_{11}\left(K_{13} + 3K_{02}K_{11} \right)$$

$$- K_{20}\left(K_{04} - 3K_{02}^2 \right) - 3K_{02}\left(K_{22} + K_{20}K_{02} + 2K_{11}^2 \right) \Big]$$

$$= \frac{1}{n} \left(K_{24} + 6K_{13}K_{11} + 3K_{03}K_{21} + 3K_{22}K_{02} + 6K_{12}^2 + 6K_{11}^2 K_{02} \right)$$

$$= \frac{1}{n} \Big[\beta^4 K_6(X) + 6\beta^4 K_4(X) K_2(X) + 3\beta^4 K_3^2(X)$$

$$+ 3\left\{ \beta^2 \mu_2(X) + \mu_3(\varepsilon') \right\} \beta^2 K_4(X) + 6K_3^2(X)$$

$$+ 6\beta^2 \mu_2^2(X) \left\{ \beta^2 \mu_2(X) + \mu_2(\varepsilon') \right\} \Big]$$

$$= \frac{1}{n}\Big[\ \beta^4\{K_4(X) + 9K_4(X)\ K_2(X) + 9K_3^2(X) + 6K_2^3(X)\}$$

$$+ 3\mu_2(\varepsilon')\ \beta^2\{K_4(X) + 2K_2^2(X)\}\Big]$$

$$Cov(m_{12}, m_{30}) = \frac{1}{n}(K_{42} + 6K_{31}K_{11} + 3K_{30}\ K_{12} + 3K_{22}\ K_{20} + 6K_{21}^2 + 6K_{11}^2K_{20})$$

$$= \frac{1}{n}\Big[\beta^2K_6(X) + 6\beta^2K_4\ (X)\ K_2(X) + 3\beta^2K_3^2(X)$$

$$+ 3\beta^2K_4(X)\{K_2(X) + K_2(u)\} + 6\beta^2\ K_3^2(X)$$

$$+ 6\beta^2K_2^2(X)\ \{K_2(X) + K_2(u)\}\ \Big]$$

$$= \frac{1}{n}\Big[\ \beta^2\{K_6(X) + 9K_4(X)\ K_2(X) + 9K_3^2(X) + 6K_2^3(X)\}$$

$$+ 3\beta^2\mu_2(u)\ \{K_4(X) + 2K_2^2(X)\}\Big]$$

$$Cov\ (m_{03}, m_{30}) = \frac{1}{n}(\ \mu_{33} - \mu_{30}\mu_{03} + 9\mu_{11}\mu_{20}\mu_{02} - 3\mu_{31}\mu_{02} - 3\mu_{20}\mu_{13})$$

$$= \frac{1}{n}\Big[K_{33} + 3K_{31}K_{02} + K_{30}K_{03} + 9K_{22}K_{11} + 9K_{21}K_{12} + 3K_{20}K_{13}$$

$$+ 9K_{20}K_{11}K_{02} + 6K_{11}^3 - K_{30}K_{03} + 9K_{11}K_{02}K_{20}$$

$$- 3K_{02}\{K_{31} + 3K_{20}K_{11}\} - 3K_{20}\{K_{13} + 3K_{02}K_{11}\}\Big]$$

$$= \frac{1}{n}(K_{33} + 9K_{22}\ K_{11} + 9K_{21}K_{12} + 6K_{11}^3\ )$$

$$= \frac{1}{n}\ \Big[\beta^3K_6(X) + 9\beta^3K_4(X)K_2(X) + 9\beta^3K_3^2(X) + 6\beta^3K_2^3(X)\ \Big]$$

$$= \frac{1}{n}\Big[\beta^3\{\ K_6(X) + 9K_4(X)K_2(X) + 9K_3^2(X) + 6K_2^3(X)\}\Big]$$

$$C_{ov.}(m_{30}, m_{21}) = \frac{1}{n}\left(\mu_{51} - \mu_{30}\mu_{21} + 9\mu_{20}^2\ \mu_{11} - 2\mu_{40}\mu_{11} - 4\mu_{31}\mu_{20}\right)$$

$$= \frac{1}{n}\Big[\beta\left\{\mu_6(X) + 6\mu_2(X)\mu_4(u) + 10\mu_2(u)\ \mu_4(X)\right\}$$

$$- \beta\mu_3^2(X) + 9\left\{\mu_2^2(X) + \mu_2^2(u) + 2\mu_2(X)\mu_2(u)\right\}\ \beta\mu_2(X)$$

$$- 2\beta\mu_2(X)\ \left\{\mu_4(X) + \mu_4(u) + 6\mu_2(X)\mu_2(u)\right\}$$

$$- 4\beta\left\{\mu_2(X) + \mu_2(u)\right\}\ \left\{\mu_4(X) + 3\mu_2(X)\ \mu_2(u)\right\}\Big]$$

$$= \frac{1}{n}\Big[\beta\ \left\{\mu_6(X) - \mu_3^2\ (X) + 9\mu_2^3(X) - 6\mu_4(X)\ \mu_2(X)\right\}$$

$$+ \beta\mu_4(X)\ \left\{10\mu_2(u) - 4\mu_2(u)\right\} + \beta\mu_2(X)\ \left\{5\mu_4(u) + 9\mu_2^2(u)\right.$$

$$\left.- 2\mu_4(u) - 12\mu_2^2(u)\right\} + \beta\mu_2^2(X)\ \left\{18\mu_2(u) - 12\mu_2(u) - 12\mu_2(u)\right\}$$

$$= \frac{\beta}{n}\Big[\left\{\mu_6(X) - \mu_3^2(X) + 9\mu_2^3(X) - 6\mu_4(X)\mu_2(X)\right\} + 6\mu_4(X)\mu_2(u)$$

$$+ 3\mu_2(X)\ \left\{\mu_4(u) - \mu_2^2(u)\right\} - 6\mu_2^2(X)\ \mu_2(u)\Big]$$

$$V\left(\frac{m_{21}}{m_{30}}\right) = \frac{\beta^2}{n\mu_3^2(X)}\Big[4\mu_4(X)\mu_2(u) + 2\mu_2^2(X)\mu_2(u) + 4\mu_2(X)\ \left\{\mu_4(u) - \mu_2^2(u)\right\}$$

$$- 3\mu_2(u)\left\{\mu_4(X) - \mu_2^2(X)\right\} + \left\{\mu_6(u) + 9\mu_2^3(u) - 6\mu_2(u)\mu_4(u)\right.$$

$$+ \frac{1}{\beta^2}\ \mu_2(\varepsilon')\left\{\mu_4(u) - \mu_2^2(u) + \mu_4(X) - \mu_2^2(X) + 4\mu_2(X)\mu_2(u)\right\}\Big]$$

$$= \frac{\beta^2}{n\mu_3^2(X)}\Big[4\mu_4(X)\mu_2(u) + 2\mu_2^2(X)\mu_2(u) + 8\mu_2(X)\ \mu_2^2(u)$$

$$- 3\mu_2(u)\left\{\mu_4(X) - \mu_2^2(X)\right\} + \left\{25\mu_2^3(u) + 9\mu_2^3(u) - 18\mu_2^3(u)\right\}$$

$$+ \frac{1}{\beta^2}\ \mu_2(\varepsilon')\left\{2\mu_2^2(u) + \mu_4(X) - \mu_2^2(X) + 4\mu_2(X)\mu_2(u)\right\}\Big]$$

$$= \frac{\beta^2}{n\mu_3^2(X)} \Big[\mu_2(u) \{ \mu_4(X) + 2\mu_2^2(X) + 3\mu_2^2(X) \}$$

$$+ 8\mu_2^2(u)\, \mu_2(X) + 6\mu_2^3(u)$$

$$+ \frac{1}{\beta^2}\mu_2(\varepsilon') \{ 2\mu_2^2(u) + 4\mu_2(X)\mu_2(u) + \mu_4(X) - \mu_2^2(X) \} \Big]$$

$$V\left(\frac{m_{12}}{m_{21}}\right) = \frac{\beta^2}{n\mu_3^2} \Big[\mu_2(u)\{K_4(X) + 2\mu_2^2(X)\} + \frac{4\mu_2(\varepsilon')\{K_4(X) + 2K_2^2(X)\}}{\beta^2}$$

$$+ \frac{4\mu_2(X)\,\mu_2(u)\,\mu_2(\varepsilon')}{\beta^2} + \frac{\mu_2(X)\{K_4(\varepsilon') + 2\mu_2^2(\varepsilon')\}}{\beta^4}$$

$$+ \frac{\mu_2(u)\{K_4(\varepsilon') + 2\mu_2^2(\varepsilon')\}}{\beta^4} + \mu_2(X)\{\mu_4(u) - \mu_2^2(u)\}$$

$$+ 4\mu_2(u)\{\mu_4(X) - \mu_2^2(X)\} + \frac{\mu_4(\varepsilon')\{\mu_4(X) - \mu_2^2(X)\}}{\beta^2}$$

$$+ \frac{\mu_2(\varepsilon')\{\mu_4(u) - \mu_2^2(u)\}}{\beta^2} + \frac{4\mu_2(X)\mu_2(u)\mu_2(\varepsilon')}{\beta^2}$$

$$- 4\mu_2(u)\{K_4(X) + 2K_2^2(X)\} - \frac{4\mu_2(\varepsilon')\{K_4(X) + 2K_2^2(X)\}}{\beta^2}$$

$$- \frac{8\mu_2(X)\,\mu_2(u)\,\mu_2(\varepsilon')}{\beta^2} \Big]$$

$$= \frac{\beta^2}{n\mu_3^2} \Big[\mu_2(u)\{\mu_4(X) - \mu_2^2(X)\} + \mu_2(X)\{\mu_4(u) - \mu_2^2(u)\}$$

$$+ \frac{\mu_2(\varepsilon')\{\mu_4(X) - \mu_2^2(X)\}}{\beta^2} + \frac{\mu_2(X)\{\mu_4(\varepsilon') - \mu_2^2(\varepsilon')\}}{\beta^4}$$

$$+ \frac{\mu_2(u)\{\mu_4(\varepsilon') - \mu_2^2(\varepsilon')\}}{\beta^4} + \frac{\mu_2(\varepsilon')\{\mu_4(u) - \mu_2^2(u)\}}{\beta^4} \Big]$$

$$= \frac{\beta^2}{n\mu_3^2}\Bigg[\mu_2(u)\left\{\mu_4(X) - \mu_2^2(X)\right\} + \mu_2(X)\cdot 2\mu_2^2(u) + \frac{\mu_2(\varepsilon')\left\{\mu_4(X)-\mu_2^2(X)\right\}}{\beta^2} + \frac{\mu_2(\varepsilon')\,2\mu_2^2(u)}{\beta^2} + \frac{\mu_2(X)\,2\mu_2^2(\varepsilon')}{\beta^4} + \frac{\mu_2(u)\,2\mu_2^2(\varepsilon')}{\beta^4}\Bigg]$$

$$V\left(\frac{m_{03}}{m_{12}}\right) = \frac{\beta^2}{n\mu_3^2}\Bigg[\mu_2(u)\left\{\mu_4(X) - \mu_2^2(X)\right\} + \frac{4\mu_2(\varepsilon')\left\{\mu_4(X)-\mu_2^2(X)\right\}}{\beta^2} + \frac{4\mu_2(X)\,\mu_2(u)\,\mu_2(\varepsilon')}{\beta^2} + \frac{\left\{\mu_2(X)+\mu_2(u)\right\}\left\{\mu_4(\varepsilon')-\mu_2^2(\varepsilon')\right\}}{\beta^4} + \frac{K_6(\varepsilon')+9K_4(\varepsilon')K_2(\varepsilon')+6K_2^3(\varepsilon')}{\beta^6} + \frac{9\mu_2(\varepsilon')\left\{\mu_4(X)-\mu_2^2(X)\right\}}{\beta^2} + \frac{9\mu_2(X)\left\{\mu_4(\varepsilon')-\mu_2^2(\varepsilon')\right\}}{\beta^4} - \frac{6\mu_2(X)\left\{\mu_4(\varepsilon')-\mu_2^2(\varepsilon')\right\}}{\beta^4} - \frac{12\mu_2(\varepsilon')\left\{\mu_4(X)-\mu_2^2(X)\right\}}{\beta^2}\Bigg]$$

$$= \frac{\beta^2}{n\mu_3^2}\Bigg[\mu_2(u)\left\{\mu_4(X) - \mu_2^2(X)\right\} + \frac{\mu_2(\varepsilon')\left\{\mu_4(X)-\mu_2^2(X)\right\}}{\beta^2} + \frac{4\mu_2(X)\left\{\mu_4(\varepsilon')-\mu_2^2(\varepsilon')\right\}}{\beta^4} + \frac{\mu_2(u)\left\{\mu_4(\varepsilon')-\mu_2^2(\varepsilon')\right\}}{\beta^4} + \frac{4\mu_2(X)\,\mu_2(u)\,\mu_2(\varepsilon')}{\beta^2} + \frac{K_6(\varepsilon')+9K_4(\varepsilon')K_2(\varepsilon')+6K_2^3(\varepsilon')}{\beta^6}$$

$$= \frac{\beta^2}{n\mu_3^2}\Bigg[\mu_2(u)\left\{\mu_4(X) - \mu_2^2(X)\right\} + \frac{\mu_2(\varepsilon') \left\{\mu_4(X) - \mu_2^2(X)\right\}}{\beta^2}$$

$$+ \frac{4\mu_2(X)\ \mu_2(u)\ \mu_2(\varepsilon')}{\beta^2} + \frac{4\mu_2(X)\ 2\mu_2^2(\varepsilon')}{\beta^4}$$

$$+ \frac{\mu_2(u)\ 2\mu_2^2(\varepsilon')}{\beta^4} + \frac{6\mu_2^3\ (\varepsilon')}{\beta^6}\Bigg]$$

$$V\left(\frac{m_{12}}{m_{30}}\right)^{\frac{1}{2}} = \frac{\beta^2}{4n\mu_3^2}\Bigg[-6\mu_2(u)\left\{K_4(X) + 2K_2^2(X)\right\}$$

$$+ \mu_2(u)\left\{K_4(X) + 2K_2^2\ (X)\right\}$$

$$+ \frac{4\mu_2(\varepsilon')\ \left\{K_4(X) + 2K_2^2\ (X)\right\}}{\beta^2}.$$

$$+ \frac{4\mu_2(X)\ \mu_2(u)\ \mu_2(\varepsilon')}{\beta^2}$$

$$+ \frac{\left\{\mu_2(X) + \mu_2(u)\right\}\left\{K_4(\varepsilon') + 2K_2^2(\varepsilon')\right\}}{\beta^4}$$

$$+ \mu_6(u) - 9\mu_2^3(u) - 6\mu_4(u)\ \mu_2(u)$$

$$+ 9\mu_2(u)\left\{\mu_4(X) - \mu_2^2(X)\right\} + 9\mu_2(X)\ \left\{\mu_4(u)-\mu_2^2(u)\right\}\Bigg]$$

$$= \frac{\beta^2}{4n\mu_3^2}\Bigg[4\mu_2(u)\left\{\mu_4(X) - \mu_2^2(X)\right\} + 9\mu_2(X)\left\{\mu_4(u)- \mu_2^2(u)\right\}$$

$$+ \mu_6(u) - 9\mu_2^3(u) - 6\mu_4(u)\ \mu_2(u)$$

$$+ \frac{4\mu_2(\varepsilon')\ \left\{K_4(X) + 2K_2^2(X)\right\}}{\beta^4}$$

$$+ \frac{4\mu_2(X)\ \mu_2(u)\ \mu_2(\varepsilon')}{\beta^2}$$

$$+ \frac{\left\{\mu_2(X) + \mu_2(u)\right\}\left\{K_4(\varepsilon') + 2K_2^2(\varepsilon')\right\}}{\beta^4}$$

$$= \frac{\beta^2}{4n\mu_3^2} \left[4\mu_2(u)\left\{ \mu_4(X) - \mu_2^2(X) \right\} + 9\mu_2(X)2\mu_2^2(u) + 6\mu_2^3(u) \right.$$

$$+ \frac{4\mu_2(\varepsilon')\left\{ \mu_4(X) - \mu_2^2(X)\right\}}{\beta^2} + \frac{4\mu_2(X)\mu_2(u)\mu_2(\varepsilon')}{\beta^2}$$

$$\left. + \frac{\left\{\mu_2(X) + \mu_2(u)\right\} \; 2\mu_2^2(\varepsilon')}{\beta^4} \right]$$

$$V\left(\frac{m_{03}}{m_{21}}\right)^{\frac{1}{2}} = \left\{ \frac{1}{2}\left(\frac{\mu_{03}}{\mu_{21}}\right)^{\frac{1}{2}-1} \right\}^2 V\left(\frac{m_{03}}{m_{21}}\right)$$

$$= \frac{\beta^2}{4n\mu_3^2} \left[\mu_2(X)\left\{K_4(u) + 2K_2^2(u)\right\} + 4\mu_2(u)\left\{K_4(X)+2K_2^2(X)\right\} \right.$$

$$+ \frac{4\mu_2(\varepsilon')\left\{K_4(X)+2K_2^2(X)\right\}}{\beta^2} + \frac{\mu_2(\varepsilon')\left\{K_4(u)+2K_2^2(u)\right\}}{\beta^2}$$

$$+ \frac{4\mu_2(X)\mu_2(u)\mu_2(\varepsilon')}{\beta^2} + \frac{9\mu_2(X)\left\{K_4(\varepsilon') + 2K_2^2(\varepsilon')\right\}}{\beta^4}$$

$$\left. + \frac{K_6(\varepsilon') + 9K_4(\varepsilon')K_2(\varepsilon') + 6K_2^3(\varepsilon')}{\beta^6} \right]$$

$$= \frac{\beta^2}{4n\mu_3^2} \left[\mu_2(X)\; 2\mu_2^2(u) + 4\mu_2(u)\left\{ \mu_4(X) - \mu_2^2(X)\right\} \right.$$

$$+ \frac{4\mu_2(\varepsilon')\left\{ \mu_4(X) - \mu_2^2(X)\right\}}{\beta^2} + \frac{\mu_2(\varepsilon')\; 2\mu_2^2(u)}{\beta^2}$$

$$+ \frac{4\mu_2(X)\;\mu_2(u)\;\mu_2(\varepsilon')}{\beta^2} + \frac{9\mu_2(X)\; 2\mu_2^2(\varepsilon')}{\beta^4}$$

$$\left. + \frac{6\mu_2^3(\varepsilon')}{\beta^6} \right]$$

$$V\left\{\left(\frac{m_{03}}{m_{30}}\right)^{\frac{1}{3}}\right\} = \frac{1}{9}\left\{(\beta^3)^{\frac{1}{3}} - 1\right\}^2 \frac{\beta^6}{n\mu_3}\left[\frac{K_6(\varepsilon')+9K_4(\varepsilon)K_2(\varepsilon')+6K_2^3(\varepsilon')}{\beta^6}\right.$$

$$+\frac{9\mu_2(\varepsilon')\left\{K_4(X)+2K_2^2(X)\right\}}{\beta^2} + \frac{9\mu_2(X)\left\{K_4(\varepsilon')+2K_2^2(\varepsilon')\right\}}{\beta^4}$$

$$+ K_6(u) + 9K_4(u)K_2(u) + 6K_2^3(u) + 9\mu_2(u)\left\{K_4(X)+2K_2^2(X)\right\}$$

$$\left. + 9\mu_2(X)\left\{K_4(u) + 2K_2^2(u)\right\}\right]$$

$$= \frac{\beta^2}{9n\mu_3}\left[9\mu_2(X)\, 2\mu_2^2(u) + 9\mu_2(u)\left\{K_4(X) + 2K_2^2(X)\right\} + 6\mu_2^3(u)\right.$$

$$\left. + \frac{9\mu_2(\varepsilon')\left\{K_4(X) + 2K_2^2(X)\right\}}{\beta^2} + \frac{9\mu_2(X)2\mu_2^2(\varepsilon')}{\beta^4} + \frac{6\mu_2^3(\varepsilon')}{\beta^6}\right]$$

$$= \frac{\beta^2}{n\mu_3}\left[2\mu_2^2(u)\mu_2(X) + \mu_2(u)\left\{\mu_4(X) - \mu_2^2(X)\right\} + \frac{2}{3}\mu_2^3(u)\right.$$

$$\left. + \frac{\mu_2(\varepsilon')\left\{\mu_4(X) - \mu_2^2(X)\right\}}{\beta^2} + \frac{2\mu_2(X)\mu_2^2(\varepsilon')}{\beta^4} + \frac{2}{3}\frac{\mu_2^3(\varepsilon')}{\beta^6}\right]$$

$$V(\hat{\beta}_7) = V\left(\frac{m_{02}}{m_{11}}\right)$$

$$= \left(\frac{\mu_{02}}{\mu_{11}}\right)^2\left[\frac{V(m_{02})}{E^2(m_{02})} + \frac{V(m_{11})}{E^2(m_{11})} - 2\frac{Cov(m_{02}, m_{11})}{E(m_{02}).\,E(m_{11})}\right]$$

$$= \frac{(\beta^2\mu_2(X) + \sigma_v^2 + \sigma_{\varepsilon'}^2)^2}{\beta^2\,\mu_2^2(X)}\,\frac{1}{n}\left[\frac{\mu_{04} - \mu_{02}^2}{(\beta^2\mu_2(X) + \sigma_{\varepsilon'}^2)} + \frac{\mu_{22} - \mu_{11}^2}{\beta^2\,\mu_2^2(X)}\right.$$

$$\left. - 2\,\frac{\mu_{13} - \mu_{02}\,\mu_{11}}{\mu_{02}\,\mu_{11}}\right]$$

$$= \frac{\beta^2}{n}(1+c_{\varepsilon'})^2 \left[\frac{K_{04} + 2K_{02}^2}{(\beta^2\mu_2(X) + \sigma^2_{\varepsilon'})^2} + \frac{K_{22} + K_{20}K_{02} + K_{11}^2}{\beta^2 \mu_2^2(X)} - 2\frac{K_{13} + 2K_{02} K_{11}}{\mu_{02}\mu_{11}} \right]$$

$$= \frac{\beta^2}{n}(1 + c_{\varepsilon'})^2 \left[\frac{\beta^4 K_4(X) + 2K_{02}^2}{(\beta^2\mu_2(X) + \sigma^2_{\varepsilon'})^2} + \frac{\beta^2 K_4(X) + \{K_2(X) + K_2(u)\}\{\beta^2 K_2(X) + \sigma^2_{\varepsilon'}\} + \beta^2 K_2^2(X)}{\beta^2 \mu_2^2(X)} - 2\frac{\beta^3 K_4(X) + 2\beta K_2(X)\{\beta^2 K_2(X) + \sigma^2_{\varepsilon'}\}}{\{\beta^2\mu_2(X) + \sigma^2_{\varepsilon'}\}\beta\mu_2(X)} \right]$$

$$= \frac{\beta^2(1+ c^2_{\varepsilon'})}{n} \left[\frac{\beta^4(\mu_4 - \mu_2^2) + 4\beta^2\mu_2\sigma^2_{\varepsilon'} + 2\sigma^4_{\varepsilon'}}{(\beta^2\mu_2 + \sigma^2_{\varepsilon'})^2} + \frac{\beta^2(\mu_4 - \mu_2^2) + \beta^2\mu_2\sigma^2_u + (\mu_2 + \sigma^2_u)\sigma^2_{\varepsilon'}}{\beta^2\mu_2^2} - 2\frac{\beta^3(\mu_4 - \mu_2^2) + 2\beta\mu_2\sigma^2_{\varepsilon'}}{(\beta^2\mu_2 + \sigma^2_{\varepsilon'})\beta\mu_2} \right]$$

$$= \frac{\beta^2}{n}(1+ c^2_{\varepsilon'}) \left[\frac{b_2 - 1 + 4c_{\varepsilon'} + 2c^2_{\varepsilon'}}{(1 + c_{\varepsilon'})^2} + \frac{b_2 - 1 + c_u + (1+c_u)c_{\varepsilon'}}{1} - 2\frac{b_2 - 1 + 2c_{\varepsilon'}}{(1 + c_{\varepsilon'})} \right]$$

$$= \frac{\beta^2}{n}\left[c_{\varepsilon'}\{c_{\varepsilon'} b_2 + 1 - c_{\varepsilon'} + c^2_{\varepsilon'}\} + c_u(1 + c_{\varepsilon'})^3 \right]$$

Similarly,

$$V\left(\frac{m_{11}}{m_{20}}\right) = \frac{\beta^2}{n}\left[c_u\left\{c_u b_2 + 1 - c_u + c_u^2\right\} + c_{\varepsilon'}(1 + c_u)^3\right]$$

$$V(\hat{\beta}_8) = \left\{\frac{1}{2}\left(\frac{\mu_{02}}{\mu_{20}}\right)^{\frac{1}{2}-1}\right\}^2 \left(\frac{\mu_{02}}{\mu_{20}}\right)^2 \left[\frac{V(m_{02})}{E^2(m_{02})} + \frac{V(m_{20})}{E^2(m_{20})} - 2\,\frac{\mathrm{cov}(m_{02}, m_{20})}{E(m_{02})\,E(m_{20})}\right]$$

$$= \frac{1}{4n}\left(\frac{\mu_{02}}{\mu_{20}}\right)\left[\frac{\beta^4(\mu_4 - \mu_2^2) + 4\beta^2\mu_2\sigma_{\varepsilon'}^2 + 2\sigma_{\varepsilon'}^4}{(\beta^2\mu_2 + \sigma_{\varepsilon'}^2)^2} + \frac{K_4(X) + 2(\mu_2^2 + 2\mu_2\sigma_u^2 + \sigma_u^4)}{(\mu_2 + \sigma_u^2)^2} - 2\,\frac{K_{22} + K_{20}K_{02} + 2K_{11}^2 - K_{02}K_{20}}{(\beta^2\mu_2 + \sigma_{\varepsilon'}^2)(\mu_2 + \sigma_u^2)}\right].$$

$$= \frac{\beta^2(1 + c_{\varepsilon'})}{4n\,(1 + c_u)}\left[\frac{b_2 - 1 + 4c_{\varepsilon'} + 2c_{\varepsilon'}^2}{(1 + c_{\varepsilon'})^2} + \frac{b_2 - 1 + 4c_u + 2c_u^2}{(1 + c_u)^2} - 2\,\frac{b_2 - 1}{(1+c_u)(1+c_{\varepsilon'})}\right]$$

$$= \frac{\beta^2}{4n(1+c_{\varepsilon'})(1+c_u)^3}\left[(b_2-1)(c_u-c_{\varepsilon'})^2 + 2c_{\varepsilon'}(1+c_{\varepsilon'})(1+c_u)^2 + 2c_u(1+c_u)(1+c_{\varepsilon'})^2 + 2c_{\varepsilon'}(1 + c_u)^2 + 2c_u(1+c_{\varepsilon'})^2\right]$$

If $c_u = c_{\varepsilon}' = c$ (say), then

$$V(\hat{\beta}_8) = \frac{\beta^2}{4n(1+c)^3}\left[4c(1+c)^3 + 4c(1+c)^2\right]$$

$$= \frac{\beta^2 c \quad (2+c)}{(1+c)n}$$

$$\frac{V_8}{V_4} = \frac{c(2+c)}{1+c} \cdot \frac{b_1}{2c\{(b_2-1) + 2c + \frac{2}{3}c^2\}}$$

$$\simeq \frac{b_1}{b_2 - 1} \text{ (for small c).}$$

CHAPTER 3

OPTIMUM GROUPING ESTIMATORS IN THE STANDARD TWO-VARIABLE EVM

3.1 Introduction

A. Wald (1940) suggested a grouping method for estimating the slope of a straight line regression where the regressor X is subject to errors of observation. His method consists in dividing the observations (x_i, y_i), i = 1, 2, ..., n, into two equal groups according to the order of x_i's. The centre of gravities of the two groups in the scatter diagram are then joined by a straight lineto get the desired slope estimator.

Later Bartlett (1949) proposed as an improvement upon this method a three-group estimator. Here the scatter diagram is divided into three equal groups according to the order of x_i's; and the middle group is neglected for estimation of slope (β), i.e., the centre of gravities of the two extreme groups are joined by a straight line to get the estimator of β (See also Nair and Shrivastava, 1942 and Nair and Banerjee, 1943). Symbolically,

$$b_B = \frac{\bar{y}_1 - \bar{y}_2}{\bar{x}_1 - \bar{x}_2} \qquad \text{.... (3.1)}$$

where $(\bar{y}_1, \bar{x}_1)$ and $(\bar{y}_2, \bar{x}_2)$ are means of the first and last $33\frac{1}{3}$ per cent observations. In the case of Wald's estimator (b_W) they consist of first and last 50 per cent observations, i.e., the middle group consists of no observation.

These grouping estimators[3.1] have two distinct advantages over the OLS estimator. First, they are very simple and hence convenient from the view point of application. Second, they can allow for errors in the regressor to a certain extent : If the ranking of x_i's is the same as that of the underlying values (X_i's) then these estimators are consistent. It is true, however, that there is some loss of efficiency if LS method is valid (i.e., the regressor is free from errors). For the case of equispaced X-values, Barlett showed that his method based on three equal groups is optimal within the class of three-group estimators and hence more efficient than the Wald estimator.

Later, Theil and Van Yzeren (1956) suggested that the extreme groups should each contain 30 per cent of the observations (i.e., $p_1 = 0.3$, $p_2 = 0.4$ and $p_3 = 0.3$, where p_i = proportion of observations in the ith ordinal group). They got this result by assuming X, the regressor variable, to follow a Beta-distribution. They found that these are approximately the optimum proportions of observations in the three groups. The optimum proportions were determined by minimizing the sampling variance of the slope estimator which is proportional to

$$V = \frac{1/F(X_1) + 1/F(X_2)}{(\bar{X}_1 - \bar{X}_2)^2}, \qquad \ldots \quad (3.2)$$

where
$$\bar{X}_1 = \frac{1}{F(X_1)} \int_0^{X_1} tf(t)\,dt,$$
$$\bar{X}_2 = \frac{1}{1-F(X_2)} \int_{X_2}^{1} tf(t)\,dt,$$

and $f(t) = \frac{1}{B(p,\, q)}\, t^{p-1} (1-t)^{q-1}$, $0 \leq t \leq 1$ represents density

3.1/ These are in fact Instrumental Variables estimators (See Johnston, 1972).

function of X and F(.) the corresponding distribution function, and X_1 and X_2 the upper and lower limits, respectively, of the first and the third groups. However, the only pairs of parameter values (p, q) considered by them were (1.5, 1.5), (2, 1.5), (8, 8), (10, 4); the corresponding skewness coefficient ($\gamma_1 = \mu_3 / \mu_2^{3/2}$ where μ_r = rth central moment) being 0, - 0.22, 0, - 0.46.[3.2] The optimum efficiencies _vis-a-vis_ OLS ranged from 0.826 to 0.870 and the efficiencies for the suggested grouping were almost equal ranging from 0.809 to 0.870.

The motivation for the present study is the need for allowing for transitory (seasonal) elements in the regressor and regressand in Engel curve analysis. In Engel curve analysis the explanatory variable, viz., per capita income and total expenditure are clearly positively skewed and they are often found to be lognormally distributed (Aitchison and Brown, 1957, Roy and Dhar, 1960, Bhattacharya and Iyengar, 1967, Bhattacharya and Chatterjee, 1971, Iyengar and Jain, 1974, Jain, 1977) or sometimes gamma (Salem and Mount, 1974). Since, in these situations the regressor (X) is quite different from those considered by Bartlett (1949) and Theil and Van Yzeren (1956), the grouping recommended by Theil and Van Yzeren is not likely to be efficient in such cases.

Researches by Theil and Van Yzeren (1956), Nair and Shrivastava (1942), Nair and Banerjee (1943) have created the impression that in most cases the 'optimal' grouping is obtained if the three groups contain equal number of x-values. A careful scrutiny, however, shows that

3.2/ Both Bartlett (1949) and Theil and Van Yzeren (1956) considered the classical set-up for two-variable linear regression where x is free from errors. The same approach is followed in this investigation also.

these investigations hardly considered any skewed distribution of X. Most of the distributions considered by them, e.g., Normal, Uniform, Traingular etc., are not realistic for many situations. Only Gibson and Jowett (1957) studied, among other types, two particular forms of gamma distribution 3.3/.

The aim of this investigation is to find optimal groups for log-normal and gamma type distributions of X for a wide range of parameters, covering situations likely to be encountered in Engel curve analysis. Here it may be mentioned that Salem and Mount (1974) found the skewness coefficients and the coefficients of Kurtosis ($\gamma_2 = \mu_4/\mu_2^2 - 3$) to lie in the ranges (1.3, 1.4) and (2.4, 3.0) corresponding to the empirically found values of the parameter r in the range (2.0, 2.5). For the empirical size distributions of population by per-capita household consumption expenditure estimated for rural and urban India from successive rounds of the National Sample Survey (NSS) the fitted lognormal distributions seem to have σ^2 (the variance of the logarithms) in the range (0.25, 0.50) corresponding to Lorenz ratio (LR) between 0.28 and 0.38 (Roy and Dhar, 1960). Corresponding ranges of skewness and Kurtosis coefficients are (1.75, 2.94) and (5.90, 18.51) respectively — still higher than those found by Salem and Mount. More recently Ahmed and Bhattacharya (1972) found it to be three-parameter lognormal, the third parameter being the threshold parameter.

3.3/ The probability density function of gamma distribution is

$$f(X) = \frac{1}{\Gamma(r)c^r} X^{r-1} e^{-X/c}, \quad 0 < X < \infty, \quad c > 0, r > 0$$

The parameter c is a scale-parameter and does not matter in the choice of optimal groups. The other parameter r is important. Gibson and Jowett (1957) chose the values 1 and 4 for this parameter.

In section 3.2 we find he optimum proportions of the three-group procedure when the regressor follows a lognormal or gamma type distribution. It is found that for commonly occuring distributions of regressor values the optimum proportions differ considerably from those recommended by Bartlett, Theil and Van Yzeren. Roughly speaking, the optimal allocations are about $p_1 = 0.4$, $p_2 = 0.45$ and $p_3 = 0.15$. Grouping estimators based on such groups may be appreciably more efficient than the Bartlett estimator or the Theil and Van Yzeren estimator (Pal and Bhaumik, 1979).

In the section 3.3 of this Chapter we extend the same idea to the case considered by Lancaster (1968) where the disturbance term is heteroscedastic. Lancaster considered the following set-up with heteroscedastic disturbances :

$$Y_i = \alpha + \beta X_i + \varepsilon_i, \quad i = 1, 2, \ldots, n \qquad \ldots (3.3)$$

where α and β are parameters to be estimated. The standard Gauss-Markov model is assumed **except** that

$$V(\varepsilon_i | X_i) = \lambda X_i^p, \quad \lambda > 0 \qquad \ldots (3.4)$$

where λ and p are constants. The case p = 0 reduces the situation to homoscedasticity and in this case the OLS estimators are BLUE. Though p may take negative values, it is plausible to assume p to be greater than zero. Katona and others (1954) found the standard deviation of savings within each of several income classes to be proportional to the average income of each class. Goldberger (1964) also found p, which may be called the degree of heteroscedasticity, to be about 2 in the savings-income relationship. Lancaster (1968), however, found p to be approximately equal to 1.5 in the company divident payments and company profit relationship.

The BLUE's of the coefficients in this model are those given by Generalized Least Squares (GLS); these are the same as OLS estimators obtained from (3.3) after dividing through by the appropriate power of X. The interest mainly centres on estimating β, since the parameter α is easily estimated once we estimate β.

Lancaster examined the efficiencies of the three estimators of β viz., the LS estimator, b_L, Wald's estimator b_W and Bartlett's three-group estimator with equal number of observations in each group b_B, relative to GLS estimator b_G. Lancaster assumed the regressor X to follow a two-parameter lognormal distribution $\Lambda(\mu, \sigma^2)$ where $E(\log X) = \mu$ and $V(\log X) = \sigma^2$. He found some interesting results from his study:

(i) As expected, the efficiency of OLS estimator relative to GLS estimator monotonically decreases as p increases from zero. As against this, Wald's and Bartlett's estimator attain maximum efficiency when p is near 1.

(ii) For moderately high coefficient of variation of X and with high value of p grouping estimators become superior to OLS estimator. Lancaster concludes : "Our results suggest that in additon to its well-known computional simplicity, the grouping estimator performs better than OLS in the presence of heteroscedasticity of a type which the scanty evidence suggests is common with economic data" (Lancaster, 1968, p.188).

In section (3.3) we find the optimum allocations for the three-group estimator in Lancaster's set-up assuming X to follow lognormal or gamma distribution. The efficiency of the optimal three-group estimator vis-a-vis GLS estimator is compared with those of OLS, Wald's and Bartlett's estimators. Here again, the optimal allocations are far from those suggested by Bartlett and Theil and Van Yzeren. As expected, the allocations depend heavily on p. For high values of p, the optimum values of p_1 and p_2 are nearly the reverse of the optimum values for the homoscedastic case. In any case, the efficiency can be considered enhanced by use of optimal groups instead of equal or nearly equal groups (Sil et al, 1981). It also discussed the results and shows their relevance to problem with errors-in-variables.

3.2 The Homoscedastic Case

Given a distribution of X, one can compare different allocations of the observations into three groups using the expressions for V given in equation (3.2). It is easy to see that the ratio of V-values for two such allocations is invariant under a linear transformation of X. The comparisons for the gamma distribution can be done for a fixed value of the scale parameter c; and those for the three-parameter lognormal for fixed values of τ (the threshold parameter) and μ.

Lognormal : $X \sim \Lambda(\tau, \mu, \sigma^2)$

We take $\tau = 0$ and $\mu = 1$. For different values of σ^2 in the range (0.20 - 0.65) we found the optimal proportions in the three groups and the corresponding efficiencies of Bartlett type estimator relative to OLS estimator. The results are summarized in Table 3.1.

Table 3.1 : Efficiencies of different estimates where regressor $X \sim \Lambda(., \sigma^2)$

σ^2	Lorenz ratio (LR)	γ_1^*	γ_2^{**}	p_1	p_3	Efficiency vis-a-vis OLS: Wald	Bartlett	Optimum	Approximations to optimum: p_1=0.4, p_3=0.1	p_1=0.4, p_3=0.15
0.20	0.25	1.52	4.35	0.35	0.17	0.54	0.68	0.77	0.72	0.72
0.25	0.28	1.75	5.90	0.37	0.16	0.51	0.66	0.76	0.73	0.73
0.30	0.30	1.98	7.71	0.38	0.15	0.50	0.63	0.76	0.74	0.74
0.35	0.32	2.21	9.81	0.39	0.14	0.48	0.60	0.75	0.73	0.73
0.40	0.34	2.45	12.27	0.42	0.13	0.46	0.58	0.73	0.72	0.72
0.45	0.36	2.69	15.14	0.43	0.12	0.43	0.56	0.73	0.72	0.72
0.50	0.38	2.94	18.51	0.43	0.12	0.42	0.54	0.72	0.72	0.72
0.55	0.40	3.20	22.45	0.43	0.11	0.40	0.51	0.71	0.70	0.70
0.60	0.42	3.47	27.08	0.44	0.10	0.38	0.50	0.70	0.70	0.70
0.65	0.43	3.75	32.53	0.44	0.10	0.37	0.48	0.69	0.69	0.69

* $\gamma_1 = \mu_3 / \mu_2^{3/2}$,

** $\gamma_2 = (\mu_4 / \mu_2^2) - 3$

Efficiencies of the optimal three-group method is found to be in the range (0.69 - 0.77), whereas the efficiencies of Wald's and Bartlett's equal-groups estimators are in the ranges (0.37 - 0.54) and (0.48 - 0.68) respectively. Clearly the optimum allocation leads to a substantial gain in efficiency over Bartlett's procedure.

A close look at Table (3.1) suggests that the following procedure is nearly optimal in most cases : Take the lowest 40 per cent observations in the first group and the highest 10 (or 15) per cent in the last group and then join the two centres of gravities in the scatter diagram by a straight line to get the slope estimator. Compared to the strictly optimal allocation this procedure entails a loss of efficiency not exceeding 5 per cent. The percentage allocation 40:45:15 seems to be the most advantageous in the light of the results reported below for the gamma distribution.

Gamma : $X \sim G(c, r)$

The parameter c is taken to be 1. We calculated optimum group proportions and associated efficiencies for the following values of p (Table 3.2) :

$$p = 1.00, 1.25, 1.50, 1.75, 2.00, 2.25, 2.50, 2.75, 3.00, 3.25, 3.50, 3.75, 4.00.$$

Table (3.2) also suggests a practical procedure which is nearly optimal in most cases : Here the percentage of observations in the three groups may be taken as 40:40:20 or 40:45:15. The loss of efficiency with either of the allocations is less than 3 per cent in all the cases considered in Table (3.2).

Table 3.2 : Efficiencies of different estimates where regressor $X \sim G(., r)$

r	Lorenz ratio (LR)	γ_1 *	γ_2 **	p_1	p_3	Efficiency vis-a-vis OLS: Wald	Bartlett	Optimum	Approximations to optimum: $p_1=0.4$, $p_3=0.2$	$p_1=0.4$, $p_3=0.15$
1.00	0.50	2.00	6.00	0.45	0.22	0.43	0.53	0.61	0.58	0.58
1.25	0.46	1.79	4.80	0.43	0.19	0.49	0.61	0.72	0.71	0.71
1.50	0.42	1.63	4.00	0.40	0.17	0.52	0.66	0.76	0.75	0.75
1.75	0.40	1.51	3.43	0.40	0.17	0.54	0.68	0.78	0.78	0.77
2.00	0.37	1.41	3.00	0.39	0.17	0.55	0.69	0.79	0.78	0.78
2.25	0.35	1.33	2.67	0.38	0.18	0.56	0.70	0.79	0.79	0.78
2.50	0.34	1.26	2.40	0.38	0.18	0.57	0.71	0.80	0.80	0.78
2.75	0.32	1.20	2.18	0.37	0.18	0.58	0.72	0.80	0.80	0.78
3.00	0.31	1.15	2.00	0.36	0.20	0.58	0.72	0.80	0.80	0.78
3.25	0.30	1.12	1.85	0.36	0.20	0.58	0.73	0.80	0.80	0.79
3.50	0.29	1.07	1.72	0.36	0.20	0.59	0.74	0.80	0.80	0.80
3.75	0.28	1.03	1.60	0.36	0.20	0.59	0.74	0.80	0.80	0.80
4.00	0.27	1.00	1.50	0.35	0.20	0.59	0.74	0.80	0.79	0.78

* $\gamma_1 = \mu_3/\mu_2^{3/2}$

** $\gamma_2 = (\mu_4/\mu_2^2) - 3$

We may draw some more conclusions regarding Tables (3.1) and (3.2). Efficiencies of Wald, Bartlett and Optimum estimators decrease as σ^2 for lognormal distribution increases and r for gamma distribution decreases. It should be noted that as σ^2 increases the coefficient of variation (CV) increases, but as r increases the coefficient of variation decreases. Hence CV and efficiency have trends in opposite directions to each other in both the cases. Efficiencies based on approximate proportional allocations to optimum, however, first increase and then decrease for lognormal. But for gamma distribution these efficiencies more or less increase as r increases. Again, as CV increases p_1 increases and p_3 more or less decreases in both the cases.

3.3 The Heteroscedastic Case

3.3.1 The Estimators and Their Variances

We consider the heteroscedastic set-up defined by (3.3) and (3.4) as done by Lancaster. Bartlett's estimator with equal groups can also be improved here by taking optimum proportions of observations in the three groups. The optimum proportions are obtained through minimizing the variance of the estimator

$$b = (\bar{Y}_1 - \bar{Y}_2)/(\bar{X}_1 - \bar{X}_2) \qquad \ldots \qquad (3.5)$$

$$= \beta + (\bar{\varepsilon}_1 - \bar{\varepsilon}_2)/(\bar{X}_1 - \bar{X}_2),$$

where

$$\bar{X}_1 = \frac{1}{n_1} \sum_{i \varepsilon I_1} X_i \qquad \bar{X}_2 = \frac{1}{n_2} \sum_{i \varepsilon I_2} X_i$$

$$\bar{Y}_1 = \frac{1}{n_1} \sum_{i \varepsilon I_1} Y_i, \qquad \bar{Y}_2 = \frac{1}{n_2} \sum_{i \varepsilon I_2} Y_i$$

$$I_1 = \{1, 2, \ldots, n_1\}, \qquad I_2 = \{n-n_2+1, n-n_2+2, \ldots, n\},$$

$X_1, X_2, \ldots, X_n$ being in incre sing order and n_1 and n_2 the number of observations in the first and in the third groups. The sampling variance of b is obviously

$$V(b) = \frac{\lambda}{(\bar{X}_1-\bar{X}_2)^2}\{(\sum_{i\varepsilon I_1} X_i^p)/n_1^2 + \sum_{i\varepsilon I_2} X_i^p/n_2^2 \} \qquad \ldots. (3.6)$$

Now, following Lancaster, we assume that $X \sim \Lambda(\mu, \sigma^2)$ with distribution function $\Lambda(\mu, \sigma^2)$. Let $\Lambda(S|\mu, \sigma^2) = p_1$ and $\Lambda(T|\mu, \sigma^2) = 1 - p_3$. Then, for large samples

$$V(b) = \frac{\lambda}{n} \frac{\int_0^S X^p d\Lambda(X)/\Lambda^2(S) + \int_T^\infty X^p d\Lambda(X)/\{1-\Lambda(T)\}^2}{\left[\int_0^S Xd\Lambda(X)/\Lambda(S) - \int_T^\infty Xd\Lambda(X)/\{1-\Lambda(T)\}\right]^2} \qquad \ldots (3.7)$$

To obtain optimum values of S and T or rather the optimum proportions $p_1 (= \Lambda(S))$ and p_3 $(= 1 - \Lambda(T))$ we varied S and T and found the values that minimize V(b). Comparing V(b) in equation (3.7) above with $V(b_G)$ in (3.11), $V(b_L)$ in (3.14) we find that μ does not influence the relative efficiencies of different estimators and hence its value was fixed at 1 for this search for optimal proportions.

A similar procedure was followed to obtain proportions in the three groups and corresponding sampling variances assuming $X \sim G(c, r)$ with the following density

$$g(X) = c^r X^{r-1} e^{-cX}/\Gamma(r), \qquad X \geq 0,\ c,\ r > 0 \qquad \ldots (3.8)$$

Here also the scale parameter c does not affect the relative efficiencies of the different estimators, and so its value was fixed to be 1.

The BLUE of β is

$$b_G = \frac{m'_{11}(Y, X^{1-p})\, m'_1(X^{-p}) - m'_{11}(Y, X^{-p})\, m'_1(X^{1-p})}{m'_1(X^{2-p})\, m'_1(X^{-p}) - m'^{\,2}_1(X^{1-p})} \qquad \ldots (3.9)$$

where $m'_{11}(X_1, X_2) = \frac{1}{n} \sum_{i=1}^{n} X_{1i}\, X_{2i}$,

$$m'_1(X_1) = \frac{1}{n} \sum_{i=1}^{n} X_{1i} \quad \text{etc.}$$

The variance of this estimator is for large samples

$$V(b_G) = \frac{\lambda}{n} \frac{E(X^{-p})}{E(X^{-p})\, E(X^{2-p}) - E^2(X^{1-p})} \qquad \ldots (3.10)$$

Assuming lognormality, the expression for $V(b_G)$ reduces to

$$V(b_G) = \frac{\lambda}{n} \frac{e^{\,p + \frac{1}{2} p^2 \sigma^2}}{e^{\,\mu + \sigma^2 (p-1)^2}\,(e^{\sigma^2} - 1)} \qquad \ldots (3.11)$$

The OLS estimator is

$$b_L = \frac{n^{-1} \sum_i (X_i - \bar{X})(Y_i - \bar{Y})}{n^{-1} \sum_i (X_i - \bar{X})^2} \qquad \ldots\ldots (3.12)$$

$$= \beta + \frac{n^{-1} \sum_i (X_i - \bar{X})\, \varepsilon_i}{n^{-1} \sum_i (X_i - \bar{X})^2},$$

which has the variance (for large samples)

$$V(b_L) = \frac{\lambda}{n} \frac{E(X^{2+p}) - 2E(X)E(X^{1+p}) + E^2(X)E(X^p)}{\mu_2^2(X)} \qquad \ldots (3.13)$$

where $\mu_2(X) = E(X^2) - E^2(X)$. Assuming X to follow a lognormal distribution, expression (3.13) becomes

$$V(b_L) = \frac{\lambda}{n} \frac{e^{p + \frac{1}{2}p^2\sigma^2}\left(e^{\sigma^2(1+2p)} - 2e^{p\sigma^2} + 1\right)}{(e^{\sigma^2} - 1)^2 \, e^{2\mu + \sigma^2}} \qquad \ldots (3.14)$$

The variance of Wald's estimator can be found simply by putting $\Lambda(S) = 1 - \Lambda(T) = \frac{1}{2}$ in equation (3.7) i.e.,

$$V(b_W) = \frac{\lambda}{n} \frac{E(X^p)}{\left\{ \int_0^S X d\Lambda(X) - \int_S^\infty X d\Lambda(X) \right\}^2} \qquad \ldots (3.15)$$

The expressions for sampling variances of b_G and b_L where $X \sim G(1, r)$ are

$$V(b_G) = \frac{\lambda}{n} \frac{\Gamma(r-p)\,\Gamma(r)}{\Gamma(r-p)\,\Gamma(r-p+2) - \Gamma^2(r+1-p)} \qquad \ldots (3.16)$$

and

$$V(b_L) = \frac{\lambda}{n} \frac{\Gamma(p+r)}{\Gamma(r) \cdot r^2} (p^2 + p + r) \qquad \ldots (3.17)$$

We calculated the efficiencies of different estimators — b_L, b_W, b_B (the equal three-group Bartlett estimator) and b_o (the optimum three-group estimator) — vis-a-vis the GLS estimator b_G for different parametric values of p and σ^2 when $X \sim \Lambda(\mu, \sigma^2)$ or of r when $X \sim G(c, r)$. The parametric values chosen are

$p = 0.0, 0.5, 1.0, 1.5, 2.0$

σ^2 (when $X \sim \Lambda(1, \sigma^2)$) : 0.2, 0.3, 0.4, 0.5, 0.6, 0.7

r (when $X \sim G(1, r)$) : 1.00, 1.50, 2.00, 2.50, 3.00, 3.50, 4.00.

The ranges of values of coefficient of variation (CV), Lorenz ratio (LR) and skewness coefficient $\gamma_1 = \mu_3 / \mu_2^{3/2}$ are (0.47, 1.01), (0.25, 0.45) and (1.52, 4.04) respectively when σ^2 of lognormal distribution varies from 0.2 to 0.7. The corresponding ranges are (0.25, 1.00), (0.27, 0.50) and (1.00, 2.00) respectively when r of gamma distribution ranges from 1 to 4.

The results are summarized in the Tables (3.3) and (3.4). Some of the interesting conclusions emerging from the study are stated in the next sub-section.

Table 3.3 : Efficiencies of different estimators vis-a-vis GLS estimator where $X \sim \Lambda(., \sigma^2)$.

σ^2	p	LR	CV*	γ_1^*	γ_2^*	Optimum proportions		Efficiency vis-a-vis GLS			
						First group	Third group	OLS	Wald	Bartlett	Optimum
	0.0					0.35	0.17	1.00	0.54	0.68	0.77
	0.5					0.28	0.25	0.91	0.62	0.78	0.80
0.2	1.0	0.25	0.47	1.52	4.35	0.21	0.33	0.71	0.66	0.79	0.83
	1.5					0.15	0.44	0.49	0.62	0.72	0.84
	2.0					0.09	0.57	0.30	0.54	0.59	0.82
	0.0					0.38	0.15	1.00	0.50	0.63	0.76
	0.5					6.29	0.24	0.88	0.62	0.77	0.80
0.3	1.0	0.30	0.59	1.98	7.71	0.20	0.36	0.62	0.67	0.80	0.83
	1.5					0.11	0.48	0.37	0.62	0.69	0.84
	2.0					0.07	0.65	0.19	0.49	0.51	0.81
	0.0					0.42	0.13	1.00	0.46	0.58	0.73
	0.5					0.29	0.23	0.85	0.61	0.77	0.80
0.4	1.0	0.34	0.70	2.45	12.27	0.18	0.37	0.55	0.68	0.80	0.83
	1.5					0.10	0.53	0.29	0.61	0.67	0.83
	2.0					0.06	0.71	0.12	0.46	0.46	0.79
	0.0					0.43	0.12	1.00	0.42	0.54	0.72
	0.5					0.29	0.22	0.82	0.61	0.76	0.79
0.5	1.0	0.38	0.81	2.94	18.51	0.17	0.38	0.49	0.69	0.80	0.85
	1.5					0.10	0.57	0.23	0.61	0.65	0.86
	2.0					0.04	0.77	0.08	0.42	0.41	0.71
	0.0					0.44	0.10	1.00	0.38	0.50	0.70
	0.5					0.29	0.22	0.80	0.60	0.75	0.79
0.6	1.0	0.42	0.91	3.47	27.08	0.17	0.39	0.44	0.70	0.80	0.86
	1.5					0.08	0.61	0.18	0.60	0.63	0.86
	2.0					0.03	0.83	0.06	0.38	0.37	0.71
	0.0					0.44	0.09	1.00	0.35	0.46	0.6[illegible]
	0.5					0.29	0.21	0.77	0.59	0.74	0.78
0.7	1.0	0.45	1.01	4.04	38.94	0.16	0.41	0.40	0.71	0.81	0.86
	1.5					0.07	0.64	0.15	0.59	0.61	0.86
	2.0					0.02	0.86	0.04	0.35	0.33	0.66

* $\gamma_1 = \sqrt{\mu_3^2/\mu_2^3}$ and $\gamma_2 = (\mu_4/\mu_2^2) - 3$

Table 3.4 : Efficiencies of different estimators vis-a-vis GLS estimator where $X \sim G(., r)^{**}$

r	p	LR	CV	γ_1^*	γ_2^*	Optimum proportions: First group	Optimum proportions: Third group	Efficiency vis-a-vis GLS: OLS	Wald	Bartlett	Optimum
	0.0					0.45	0.22	1.00	0.43	0.53	0.61
	0.5					0.23	0.31	0.73	0.55	0.68	0.69
1.0	1.0	0.50	1.0	2.00	6.00	-	-	-	-	-	-
	1.5					-	-	-	-	-	-
	2.0					-	-	-	-	-	-
	0.0					0.40	0.17	1.00	0.52	0.66	0.76
	0.5					0.23	0.30	0.79	0.61	0.76	0.78
1.5	1.0	0.42	0.82	1.63	4.00	0.11	0.50	0.43	0.52	0.59	0.74
	1.5					-	-	-	-	-	-
	2.0					-	-	-	-	-	-
	0.0					0.39	0.17	1.00	0.55	0.69	0.79
	0.5					0.23	0.29	0.82	0.62	0.78	0.79
2.0	1.0	0.37	0.71	1.41	3.00	0.11	0.46	0.50	0.55	0.64	0.78
	1.5					0.03	0.70	0.24	0.37	0.40	0.72
	2.0					-	-	-	-	-	-
	0.0					0.38	0.18	1.00	0.57	0.71	0.80
	0.5					0.24	0.29	0.85	0.63	0.78	0.80
2.5	1.0	0.34	0.63	1.26	2.40	0.12	0.43	0.56	0.57	0.66	0.79
	1.5					0.05	0.62	0.29	0.42	0.45	0.74
	2.0					0.01	0.86	0.13	0.24	0.22	0.63
	0.0					0.36	0.20	1.00	0.58	0.72	0.80
	0.5					0.24	0.29	0.87	0.63	0.78	0.80
3.0	1.0	0.31	0.58	1.15	2.00	0.14	0.42	0.60	0.58	0.69	0.79
	1.5					0.06	0.58	0.34	0.45	0.50	0.76
	2.0					0.01	0.77	0.17	0.29	0.30	0.69
	0.0					0.36	0.20	1.00	0.59	0.74	0.80
	0.5					0.24	0.28	0.88	0.63	0.78	0.80
3.5	1.0	0.29	0.53	1.07	1.70	0.15	0.41	0.64	0.59	0.70	0.79
	1.5					0.07	0.55	0.39	0.47	0.53	0.77
	2.0					0.03	0.72	0.20	0.33	0.35	0.71
	0.0					0.35	0.20	1.00	0.59	0.74	0.80
	0.5					0.24	0.28	0.90	0.63	0.79	0.80
4.0	1.0	0.27	0.50	1.00	1.50	0.16	0.40	0.67	0.59	0.71	0.80
	1.5					0.09	0.52	0.43	0.49	0.56	0.77
	2.0					0.04	0.68	0.24	0.36	0.39	0.73

* $\gamma_1 = \sqrt{\mu_3^2/\mu_2^3}$ and $\gamma_2 = (\mu_4/\mu_2^2) - 3$

** The cases where efficiencies do not exist are kept blank.

3.3.2 Discussion of Results

(i) The efficiency of the optimum three-group estimator is quite high, about 80 per cent, for most of the situation covered in the study.

(ii) As in the homoscedastic case, the efficiency of the optimum three-group estimator is often much higher than that of the equal-group Bartlett estimator. This is found for both lognormal and gamma type distributions of X. The difference is in some cases between 30 and 40 per cent. The difference increases as p moves away from 1 (when X is lognormal) and from 0.5 (when X is gamma). However, Bartlett's estimator is generally more efficient than Wald's estimator in both the cases.3.4/

(iii) On the whole, the efficiencies of the grouping estimators decrease as the coefficient of variation of X increases. Lancaster's statement that "the efficiency of the estimators diminishes as the variance of the distribution of X-observations increases" is generally confirmed here for lognormal distribution, though strictly speaking it is the coefficient of variation which determines efficiency. Since when $X \sim \Lambda(\mu, \sigma^2)$ efficiency does not depend on μ, it depends on CV and it is wrong to relate it to variance which involves μ. In the case of gamma distribution our results show that the efficiencies increase as the variance of X distribution increases. But here unlike lognormal the CV decreases as the variance increases.

3.4/ Lancaster observed this when X is lognormal.

(iv) The efficiencies of Wald's, Bartlett's and of the optimum three-group estimator are the highest near $p = 1$ when X is lognormal and near $p = 0.5$ when X is gamma.3.5/

(v) As found by Lancaster, the efficiency of OLS estimator is of course the highest (=1) when p=0, i.e., when the data are homoscedastic and as p increases the efficiency rapidly decreases for both the cases. At some point ($p \simeq 1.5$) it becomes less than that of other grouping estimators.

(vi) The behaviour of the optimum group-proportions is quite remarkable. The optimum proportions steadily change as p changes. As p increases, the first group proportion decreases and the third group proportion increases very rapidly. For $p \geq 2$, at some point the first group proportion becomes less than 10 per cent when X is lognormal and less than 5 per cent when X is gamma. In those cases the third group proportion becomes very high (about 70 per cent or more in almost all the cases). The sum of the two proportions is, however, more or less constant. Moreover, the proportions do not seem to vary much with σ^2 for lognormal or with r for gamma.

To conclude, given the value of p, the optimum proportions are fairly robust with respect to changes in r or σ^2. Hence if p is approximately known, optimum three-group procedure can be applied which is possibly much more efficient than Bartlett's equal three-group

---

3.5/ The efficiency of Wald's estimator is symmetric around p=1 and 0.5 respectively. This can be proved theoretically, since the efficiencies are proportional to $e^{-\sigma^2(p-1)^2}$ and $\dfrac{1}{\Gamma(r+p)\,\Gamma(r-p+1)}$ respectively.

estimator and simpler than OLS and GLS estimator. The efficiency of this optimum estimator is fairly high throughout the ranges of parameters considered in the study. Unlike other grouping estimators, the efficiency of the optimum estimator is about 70 per cent or more in almost all the cases and in most of the cases it is about 80 per cent.

Lancaster showed the superiority of grouping estimators over OLS estimator in the heteroscedastic set-up for some values of p. There is one more reason why one should sometimes prefer grouping estimators. The distribution of X may be such that $E(X^{1+p})$ or $E(X^{2+p})$ does not exist whereas $E(X^{p})$ exists. There may be trouble even with the BLUE for $V(b_G)$ contains $E(X^{-p})$ which also may not exist. An example of the latter type occurs if X has a gamma distribution with crucial parameter $r < p$. The gamma distribution is sometimes taken as a good fit for income distribution (Salem and Mount, 1974).

In the errors-in-variables models with homoscedastic disturbances OLS estimator has been proved to be inconsistent. The grouping estimators are consistent if the groups are unaltered by the introduction of errors. In the section 3.2 of this chapter the problem of getting optimum grouping estimators in the case of homoscedastic disturbances has been discussed. The aim of this study is to show how one can extend their procedure to the present set-up to improve the Bartlett-tupe estimators and obtain optimum group proportions with considerably higher efficiency.

If many types of distributions arising in practice are analysed in this way, it may be possible to find a relation between the optimum proportions and the measures of skewness and kurtosis of the distributions.

Chapter 4

ESTIMATION PROBLEMS IN MORE GENERAL EIV MODELS

PART I

4.1 Introduction

In this and the following chapter we try to get some consistent estimators of the parameter β in two variable EVM when the assumptions are relaxed in several directions :

(i) The conditional distribution of the error term u associated with regressor X given X is allowed to have variance proportional to a power of X which may be unknown;4.1/

(ii) The errors u and v (the error associated with the regressand variable) may be correlated (*vide* Pal, 1977).

These relaxations are suggested by empirical applications, e.g., those in engel curve analysis based on cross-section data. It may be mentioned here that such generalizations are often ignored in econometric literature. In practice they are very important and appear to be realistic for many applications. The object of the study is to tackle problem of estimating the regression coefficient in EIV models more general than those usually considered. While the solutions have obvious applications to engel curve analysis, the problem tackled is one of the classical problems of econometric literature.

Liviatan (1961) considered the problem of EIV in the context of engel curve analysis. The problem is specially important in India where a great deal of work on engel curve estimation is based on National Sample Survey (NSS) data on household budgets collected through the

---

4.1/ The homoscedastic situation is also considered as a special case of this general model.

interview method. The interviews are uniformly spread over the survey period, usually one year, but any particular sample household furnishes data on consumer expenditure for the reference period of last month, that is, the last thirty days preceding the date of interview. As a result, seasonal variation is superimposed on between households variation of total expenditure (x) and item consumption (y). Conceptually, in engel curve analysis one is interested in the relationship between the stable or permanent component (X and Y) of these variables, whereas in common practice one regresses the observed value y on the observed value x. One thus faces a typical EIV problem where the errors in the two variables are likely to be correlated. Bhattacharya (1967) cites two instances suggesting that the usual LS regression approach may be giving seriously biased estimates of engel elasticities in India.

In solving the estimational problems arising in these more general set ups, we have frequently assumed that X is lognormally distributed. This is very realistic in engel curve analysis (vide, Aitchison and Brown, 1957; Bhattacharya and Iyengar, 1961; Ahmed and Bhattacharya, 1972; Iyengar and Jain, 1974; Bhattacharya, 1978). The distribution of u given X has sometimes been assumed to be normal and sometimes to follow Pearsonian type II distribution; the latter may be more logical since $x = X+u$ is necessarily ≥ 0. Particular attention has been given to moment estimators but ML estimation has also been attempted.

The OLS estimator of the parameter β in the standard two-variable EVM has the limiting value

$$\underset{n \to \infty}{\text{plim}} \; \hat{\beta}_L = \beta \frac{\sigma_X^2}{\sigma_X^2 + \sigma_u^2} , \qquad \ldots \quad (4.1)$$

which is not equal to β. Clearly, consistent estimation of β would be possible if one had a consistent estimate of σ_X^2 or σ_u^2 or σ_u^2/σ_X^2. Once we find such a consistent estimate, we get the consistent estimate of β as

$$\hat{\beta} = \hat{\beta}_L \frac{\sigma_x^2}{\hat{\sigma}_X^2} \quad \text{if } \sigma_X^2 \text{ is estimated consistently}$$

$$= \hat{\beta}_L \frac{\sigma_x^2}{\sigma_x^2 - \hat{\sigma}_u^2} \quad \text{if } \sigma_u^2 \text{ is estimated consistently}$$

$$= \hat{\beta}_L \left(1 + \left(\frac{\widehat{\sigma_u^2}}{\sigma_X^2}\right)\right) \quad \text{if } \sigma_u^2/\sigma_X^2 \text{ is estimated consistently}$$

In the present chapter we consider only the observations on x viz., $x_1, x_2, \ldots, x_n$, generated as

$$x_i = X_i + u_i, \quad i = 1, 2, \ldots, n \qquad \ldots (4.2)$$

where X_i is the true value and u_i the error. We examine the possibility of estimating separately the parameters of the distributions of X and u. Clearly, the separation is impossible unless some assumptions are made. Our attempt is to get some useful results making relistic assumptions. These results are utilized in the following chapter where we consider bivariate data and tackle the problems of two variable EVM.

In equation (4.2) we assume that X_i's are i.i.d. and u_i given X_i is distributed such that

(i) $E(u_i|X_i) = 0$ for all i = 1, 2, ..., n ... (4.3)

(ii) $V(u_i|X_i) = a^2 X_i^b$ for all i = 1, 2, ..., n ... (4.4)

where a and b are positive constants.

and (iii) $E(u_i^{2r+1}|X_i) = 0$ for all i = 1, 2, ..., n ... (4.5)

and r = 1, 2, 3, ... etc.

The special cases b=0 and b=2 are of particular interest. Observe that in the case b=0, X and u may be assumed to be independent of each other. The case b=2, i.e., where the standard deviation of the error term is proportional to the magnitude of X has already been referred to by Friedman (vide section 1.5, Chapter 1). It can also be shown that if X is stretched from zero to infinity as in lognormal or gamma distributions and if x, the observed variate, is always positive then b must be equal to 2.

Theorem 4.1. In the model defined by equations (4.2), (4.3), (4.4) and (4.5). if the absolute value of u given X does not exceed X (so that no x is negative) and if X is distributed from 0 to infinity, then b must be equal to 2.

Proof. Since the absolute value of u given X is less than or equal to X,

$$E(u_i^2 | X) \leq X^2 \quad \text{or} \quad a^2 X^b \leq X^2 \quad \text{or} \quad a^2 \leq X^{2-b}$$

If b >2, by taking X sufficiently large and if b <2, by taking X sufficiently small, we get $a^2=0$, which is unacceptable. Hence b=2. Q.E.D.

In fact, there is another justification for taking b=2 which can be seen from the following theorem.

Theorem 4.2. Suppose that in the assumptions of theorem (4.1), the assumption on V(u|X) is replaced by

$$V(u|X) = a_0 + a_1 X + \cdots\cdots + a_k X^k \qquad \ldots \qquad (4.6)$$

where $a_0, a_1, \ldots, a_k$ are constants. Then

$$a_0 = a_1 = a_3 = \cdots\cdots = a_k = 0 \text{ and } a_2 > 0.$$

Proof. By the same argument as stated in the proof of theorem (4.1), we have

$$a_0 + a_1 X + a_2 X^2 + \cdots\cdot + a_k X^k \leq X^2$$

Taking X sufficiently small we get $a_0 = 0$, so,

$$a_1 X + a_2 X^2 + \cdots\cdot + a_k X^k \leq X^2$$

$$\text{or} \quad a_1 + a_2 X + \cdots\cdots + a_k X^{k-1} \leq X$$

Applying the same argument one gets $a_1 = 0$. Hence

$$a_2 X + \cdots\cdots + a_k X^{k-1} \leq X$$

$$\text{or} \quad a_2 + a_3 X + \cdots\cdot + a_k X^{k-2} \leq 1,$$

$$\text{or} \quad \frac{a_2}{X^{k-2}} + \frac{a_3}{X^{k-3}} + \cdots\cdots + a_k \leq \frac{1}{X^{k-2}} \qquad (\text{for } X > 0)$$

Now take X sufficiently large and get $a_k = 0$. Continuing this way we get $a_3 = \dots\dots = a_{k-1} = a_k = 0$. $a_2 > 0$ follow since variance is always positive if X is not degenerate. Q. E. D.

Although $b \neq 2$ is theoretically impossible in the circumstances considered above, in practice we may not find X to be very small or very large, as assumed in the proof, in the sample data. So, we may not always confine ourselves to b=2 only. We shall in fact see what happens if we take b to be some other specific value.

In the next section the method of moment estimation of different parameters of the univariate model has been discussed. In the section (4.3) we derive the p.d.f. of the observed variable x for models with different distributional assumptions and obtain the ML equation for one particular model of interest. Section (4.4) discusses moment estimators for the particular model. In section (4.5) the results of some Monte-Carlo experiments for one model are presented. One set of moment estimators is found to be highly efficient in this case. The concluding section (section 4.7) discusses the possibilities of other moment estimators which are computationally simple. Appendix (4.2) examines one interesting model which is proved to be non-identifiable.

4.2 The Moment Estimators of Univariate Models

4.2.1 The General Set-up

We have the following moment equations from the model (4.2) to (4.5) :

$$E(x) = E(X) = \mu_1' \text{ (say)} \qquad \dots (4.6)$$

$$E(x^2) = E(X^2) + a^2E(X^b) = \mu_2' \text{ (say)} \qquad \dots (4.7$$

$$E(x^3) = E(X^3) + 3a^2E(X^{b+1}) = \mu_3' \text{ (say)} \qquad \dots (4.8)$$

and so on.

To estimate parameters of the model we replace μ'_r, $(r = 1, 2, \ldots)$ by the corresponding sample moments m'_r, $(r = 1, 2, \ldots\ldots)$, and set

$$m'_1 = E(X) \qquad \ldots \quad (4.9)$$

$$m'_2 = E(X^2) + a^2 E(X^b) \qquad \ldots \quad (4.10)$$

$$m'_3 = E(X^3) + 3a^2 E(X^{b+1}) \qquad \ldots \quad (4.11)$$

If we assume X to follow a two-parameter lognormal distribution $\Lambda(\mu, \sigma^2)$ and b to be known, then the three parameters μ, σ^2 and a^2 may be estimated by method of moments using equations (4.9), (4.10), and (4.11).[4.2] If we want to tackle the more general problem where b is unknown, we may assume u given X to follow some standard distribution, say normal, and add another equation setting m'_4 equal to $E(x^4)$ to the equations (4.9) to (4.11). If $u|X$ is normal, it has an infinite range and, in principle, $x = X+u$ can be negative. That is why we avoid this assumption in the greater part of this chapter. Though we shall discuss here the problem of finding solutions for the case b=0 and b=2 only, the case where b is known but different from 0 or 2 may be tackled in a similar manner (vide subsection (4.2.5)).

4.2.2 Case 1 : b = 0

In this case, assuming $X \sim \Lambda(\mu, \sigma^2)$, the equations (4.9) to (4.11) reduce to

$$m'_1 = \exp(\mu + \sigma^2/2) \qquad \ldots \quad (4.12)$$

$$m'_2 = m_1'^2 \exp(\sigma^2) + a^2 \qquad \ldots \quad (4.13)$$

$$m'_3 = m_1'^3 \exp(3\sigma^2) + 3a^2 m'_1 \qquad \ldots \quad (4.14)$$

Substituting a^2 of (4.13) in (4.14), we get,

$$m'_3 = m_1'^3 e^{3\sigma^2} - 3m_1'^3 e^{\sigma^2} + 3m'_1 m'_2 = f(\sigma^2), \quad \text{(say)} \ \ldots . \ (4.15)$$

[4.2] Except in section (5.3), X has been assumed to follow a lognormal distribution throughout these two chapters. Subsection (4.2.4) discusses estimation problems where X is three-parameter lognormal.

<u>Theorem 4.3</u>. It is possible to get positive solutions for σ^2 and a^2 from equations (4.12) to (4.14) if $m_1' > 0$, the third central moment $m_3 > 0$ and $m_3' < (m_2'/m_1')^3$.

<u>Proof</u>. $f(\sigma^2)$ is an increasing function of σ^2 for $\sigma^2 > 0$, since

$$f'(\sigma^2) = 3m_1'^3 e^{\sigma^2} (e^{2\sigma^2} - 1) > 0$$

Now,

$$f(0) = -2m_1'^3 + 3m_1' m_2' = 2m_1' \{ (m_2' - m_1'^2) + m_2'/2 \} > 0,$$

and

$$m_3' - f(0) = m_3' - 3m_1'm_2' + 2m_1'^3 = m_3 > 0 \quad \text{(by assumption)}.$$

Hence the possibility of getting a solution $\sigma^2 > 0$ follows from the figure (4.1) given below.

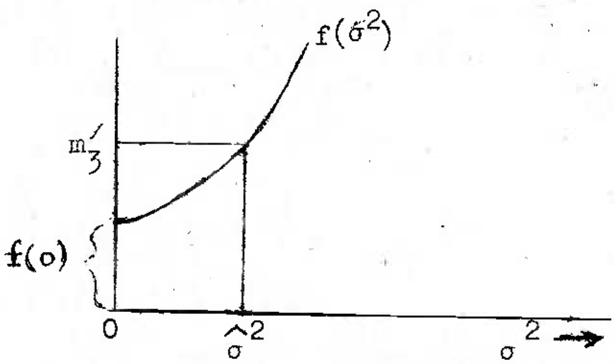


Fig. 4.1

Again rearranging (4.15), we get

$$m_3' - m_1'^3 e^{3\sigma^2} = 3m_1' (m_2' - m_1'^2 e^{\sigma^2})$$

or

$$g_1(\sigma^2) = g_2(\sigma^2) \quad \text{(say)} \quad \text{for } \sigma^2 = \hat{\sigma}^2$$

We know that $g_1(\sigma^2)$ and $g_2(\sigma^2)$ can intersect at one point only (figure (4.2)). Both $g_1(\sigma^2)$ and $g_2(\sigma^2)$ decrease as σ^2 increases. To ensure $a^2 > 0$, we must have $g_1(\hat{\sigma}^2) = g_2(\hat{\sigma}^2) > 0$, i.e., $g_1(\sigma^2)$ and $g_2(\sigma^2)$ must intersect in the first quadrant only. Now since $g_1(0) > g_2(0)$ (because $m_3 > 0$), the condition reduces to

$$g_2^{-1}(0) > g_1^{-1}(0)$$

$$\text{or} \quad \sigma_2^2 > \sigma_1^2 \quad \text{(say).}$$

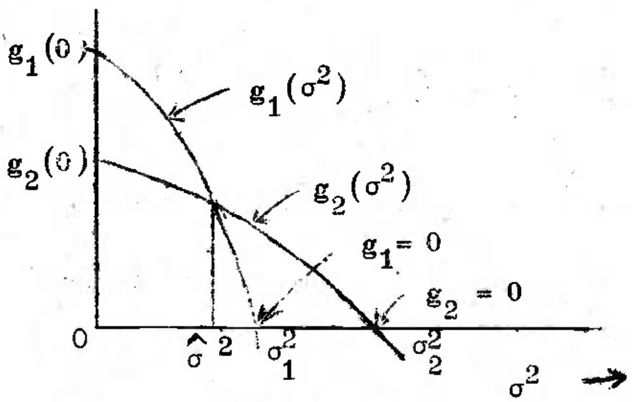


Fig. 4.2

Now,

$$g_1(\sigma^2) = 0 \Rightarrow \frac{m_3'}{m_1'^3} = e^{3\sigma^2},$$

$$\text{or} \quad \sigma_1^2 = \frac{1}{3} \log \left(\frac{m_3'}{m_1'^3} \right).$$

Similarly,

$$g_2(\sigma^2) = 0 \Rightarrow \sigma_2^2 = \log \left(\frac{m_2'}{m_1'^3} \right).$$

Condition for $a^2 > 0$ in therefore

$$\log \left(\frac{m_2'}{m_1'^2} \right) > \frac{1}{3} \log \left(\frac{m_3'}{m_1'^3} \right)$$

$$\text{or} \quad m_3' < \left(\frac{m_2'}{m_1'} \right)^3.$$

Combining the two conditions for $a^2 > 0$ and $\sigma^2 > 0$, we get

$$3m_1' m_2' - 2m_1'^3 < m_3' < \left(\frac{m_2'}{m_1'}\right)^3 \qquad \ldots \quad (4.16)$$

These conditions are not inconsistent, since

$$3m_1' m_2' - 2m_1'^3 < \left(\frac{m_2'}{m_1'}\right)^3 ,$$

because

$$m_2'^3 - 3m_1'^4 m_2' + 2m_1'^6 = (m_2' - m_1'^2)^2 \; (m_2' + 2m_1'^2) > 0.$$

It also follows from the proof that the feasible solution is unique. Q.E.D.

4.2.3 : Case 2 : b = 2

In this case assuming $X \sim \Lambda(\mu, \sigma^2)$, the equations are

$$m_1' = \exp(\mu + \tfrac{1}{2}\sigma^2) \qquad \ldots \quad (4.17)$$

$$m_2' = m_1'^2 \, (1 + a^2) \, \exp(\sigma^2) \qquad \ldots \quad (4.18)$$

$$m_3' = m_1'^3 \, (1 + 3a^2) \, \exp(3\sigma^2). \qquad \ldots \quad (4.19)$$

From (4.18) and (4.19), we get,

$$a^2 = \frac{m_2'}{m_1' \exp(\sigma^2)} = \frac{1}{3}\left(\frac{m_3'}{m_1'^3 \exp(3\sigma^2)} - 1\right)$$

$$\text{or} \quad m_3' = 3m_1' m_2' e^{2\sigma^2} - 2m_1'^3 e^{3\sigma^2} = g(\sigma^2), \text{ say.} \qquad \ldots. \quad (4.20)$$

Theorem 4.4 : We have unique feasible solutions for σ^2 and a^2 from equations (4.17) to (4.19) under the same conditions as stated in the theorem (4.3).

<u>Proof</u> : $g'(\sigma^2) = 6m'_1 m'_2 \exp(2\sigma^2) - 6m'^3_1 \exp(3\sigma^2)$

$$\therefore \quad g'(\sigma^2) = 0 \iff e^{\sigma^2} = m'_2/m'^2_2 .$$

Let, $\qquad e^{\sigma^{*2}} = m'_2 \,/\, m'^2_1$

$$g''(\sigma^2)\Big|_{\sigma^2 = \sigma^{*2}} = - 6m'^3_1 \, e^{3\sigma^2} < 0 .$$

So, $\qquad \underset{\sigma^2 > 0}{\text{Max}} \; g(\sigma^2) = g(\sigma^{*2}) = (m'_2 \,/\, m'_1)^3 .$

If $\sigma^2 < \sigma^{*2}$, then $\quad g'(\sigma^2) > 0$, $\qquad$ ... (4.21)

since $\quad g'(\sigma^2) = 6m'^3_1 \, e^{2\sigma^2} (e^{\sigma^{*2}} - e^{\sigma^2})$.

Again, since $a^2 = m'_2 \,/\, (m'^2_1 \, e^{\sigma^2}) - 1$,

$$a^2 > 0 \iff \sigma^2 < \sigma^{*2} \qquad \text{... (4.22)}$$

From (4.21) and (4.22) it follows that for the solution σ^2 to be positive m'_3 must lie between $g(0)$ and $g(\sigma^{*2})$ [see figure 4.3],

or $\quad 3m'_1 m'_2 - 2m'^3_1 < m'_3 < (m'_2 \,/\, m'_1)^3$

or $\quad m'_3 > 0$ and $m'_3 < (m'_2 \,/\, m'_1)^3$.

The conditions are the same as in the case where b is taken to be zero. $\qquad$ Q.E.D.

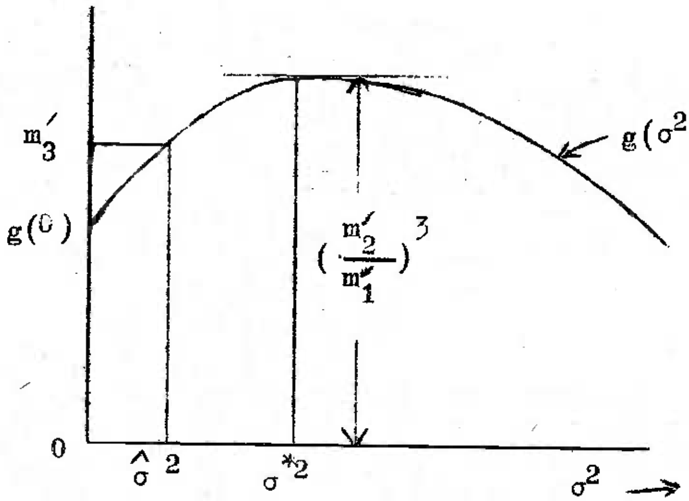


Fig. 4.3

We may now make some observations on the conditions for feasibility of the solutions :

(i) $m_3 > 0$ is to be expected, since X is lognormal which is a positively skewed distribution and all odd order raw moments of u vanish.

(ii) $m_3' < (m_2' / m_1')^3$: After some manipulations the condition becomes

$$\frac{m_3}{m_1'^3} < \left(\frac{m_2}{m_1'^2}\right)^3 + 3\left(\frac{m_2}{m_1'^2}\right)^2 .$$

In case of two parameter lognormal distribution we have the equality :

$$\frac{m_3}{m_1'^3} = \left(\frac{m_2}{m_1'^2}\right)^3 + 3\left(\frac{m_2}{m_1'^2}\right)^2 .$$

Since we are adding a variate, odd order moments of which vanish, to a lognormal variate both m_1' and m_3 remain unchanged while the variance m_2 is increased. So the above inequality is expected to hold in practice.

4.2.4: Case 3 : $b = 0$ and $X \sim \Lambda(\theta, \mu, \sigma^2)$

In some situations a more appropriate assumption on X may be that it follows the three-parameter lognormal distribution $\Lambda(\theta, \mu, \sigma^2)$, where θ signifies the threshold value (Ahmed and Bhattacharya, 1972, Jain, 1977). But the model, then, becomes very difficult to tackle. So for mathematical simplicity we shall assume $b=0$.

In this case we estimate four parameters, namely, θ, μ, σ^2, and a^2. Hence we need four moment equations. But 4th order moment of x involves 4th order moment of u, so that a distributional assumption of u becomes necessary. We assume it to be distributed normal with mean zero and variance a^2. So our model becomes :

$$\left.\begin{aligned} x &= X + u \\ X &\sim \Lambda(\theta, \mu, \sigma^2) \\ u &\sim N(0, a^2) \end{aligned}\right\} \qquad \ldots (4.23)$$

To eliminate θ we take central moments. Now

$$x - E(x) = \{ X - E(X) \} + u,$$

where the distribution of $X - E(X)$ is same as that of $X' - E(X')$, X' being distributed $\Lambda(\mu, \sigma^2)$. Denoting central moments of the population by μ_r and of the sample by m_r, $r = 2, 3, \ldots,$ etc., we arrive at the following equations :

$$m_2(x) = c_1^2 \, (e^{\sigma^2} - 1) + a^2 \qquad \ldots (4.24)$$

$$m_3(x) = c_1^3 \, (e^{\sigma^2} - 1)^2 \, (e^{\sigma^2} + 2) \qquad \ldots (4.25)$$

$$m_4(x) = c_1^4 (e^{\sigma^2} - 1)^3 (e^{3\sigma^2} + 3e^{2\sigma^2} + 6e^{\sigma^2} + 6)$$
$$+ 3\{c_1^2(e^{\sigma^2} - 1) + a^2\}^2 \qquad \ldots (4.26)$$

where $c_1 = e^{\mu + \sigma^2/2}$

The equation (4.26) may be written as

$$m_4(x) - 3m_2^2(x) = c_1^4(e^{\sigma^2} - 1)^3 \, (e^{3\sigma^2} + 3e^{2\sigma^2} + 6e^{\sigma^2} + 6) \ldots (4.27)$$

Eliminating c_1 from (4.25) and (4.27), one gets,

$$\frac{m_3^4(x)}{\{m_4(x) - 3m_2^2(x)\}^3} = \frac{(e^{\sigma^2} + 2)^4}{(e^{\sigma^2}-1)(e^{3\sigma^2}+3e^{2\sigma^2}+6e^{\sigma^2}+6)^3} = f(\sigma^2), \text{say} \ldots (4.28)$$

From (4.25) and (4.27) it is evident that both $m_3(x)$ and $m_4(x) - 3m_2^2(x)$ are positive.

Now, $f(\sigma^2)$ strictly decreases from ∞ to 0 as σ^2 moves from 0 to ∞. (This can be seen by taking derivatives and limits)4.3/. So for a given sample we may set $m_3^4(x)/\{m_4(x) - 3m_2^2(x)\}^3$ equal to $f(\sigma^2)$ and get a unique estimate $\hat{\sigma}^2$ of σ^2 and hence unique estimates for μ and a^2 from (4.24) and (4.25) subject to the following conditions which are likely to be fulfilled if the model is correct and if n is sufficiently large :

(i) $m_3(x) > 0$

(ii) $m_4(x) - 3m_2^2(x) > 0$

(iii) $m_2(x) - \hat{c}_1 \, (e^{\hat{\sigma}^2} - 1) > 0$.

θ can now be estimated using $m_1'(x) = \theta + \hat{c}_1$

4.3/ $\dfrac{df(\sigma^2)}{d\sigma^2} = - \dfrac{6e^{2\sigma^2} \, (e^{\sigma^2}+2)^3 \, (e^{2\sigma^2} + 4e^{\sigma^2} + 3)}{(e^{\sigma^2}-1)^2 \, (e^{3\sigma^2} + 3e^{2\sigma^2} + 6e^{\sigma^2} + 6)^4} < 0$

4.2.5 $u|X \sim P_{II}(-X, X, m)$

For later use in connection with bivariate problems, we present here some results for the case where $u|X$ follows the perfectly symmetrical Pearsonian type II distribution in the range $-X$ to X with p.d.f. (*vide* Elderton, 1953, p.86-89).

$$p(u|X)\,du = X^{-1}K_m\,(1-u^2/X^2)^m\,du; \quad -X \leq u \leq X, \qquad \ldots. (5.29)$$

where $K_m = B^{-1}(m + \frac{1}{2}, \frac{1}{2}) = B^{-1}(m+1, m+1).\ 2^{-2m-1}$

We may safely assume $m > 0$ since errors near zero are likely to be more probable than errors near $\pm X$. Clearly, the mean is zero and all odd order moments about zero vanish. The 2 *K.th* order moment is

$$\mu_{2K}(u|X) = \frac{1.3.5 \ldots. (2K-1)\ X^{2K}}{(2m+3)\ (2m+5) \ldots. (2m+2K+1)} \qquad \ldots. (4.30)$$

Specifically,

$$\mu_2(u|X) = X^2/(2m+3)$$

and

$$\mu_4(u|X) = 3X^4/\{(2m+3)(2m+5)\}.$$

Clearly, $m \geq 0 \iff a^2 \leq 1/3$, since here $a^2 = 1/(2m+3)$ in the expression $E(u^2) = a^2E(X^2)$.

In fact, the above result that $m \geq 0 \iff a^2 \leq 1/3$ may be generalized. That is, for any unimodal symmetric continuous distribution of $u|X$ from $-X$ to X, we have $a^2 \leq 1/3$. It is assumed that the p.d.f. increases monotonically from 0 at $-X$ to a maximum value at 0 and again decreases monotonically to 0 at X (*vide* Appendix 4.1).

The model is easily tackled as before once we recognize it as a special case of the general model described by (4.2) to (4.5) with

$b=2$ and $a^2 = 1/(2m+3)$, so that one can use the equations (4.17) to (4.19). The assumption on specific distribution of $u|X$ made above enables us to utilize higher order moments, and hence fit more elaborate models. For example, take $u|X$ to follow a Pearsonian type II distribution in the range ($-cX$, cX), where c is a positive constant. We may take the first four moment equations to estimate the parameters. Another possibility is to assume that X follows a three parameter lognormal distribution. The assumption on $u|X$ to follow this particular distribution that it is Pearsonian type II has one more advantage. We not only can go to higher order moments, but also take fractional moments (i.e., moments of fractional power). In section (4.4) we discuss how the fractional moments can be utilized and later discuss the Monte-Carlo results carried out for these moment estimators along with ML estimators.

4.2.6 The General Case : $b = b_o$ and $0 \leq b_o \leq 2$

In this case

$$m_1' = e^{\mu + \frac{1}{2}\sigma^2} = E(X) \qquad \ldots (4.31)$$

$$m_2' = m_1'^2 \, e^{\sigma^2} + a^2 e^{b\mu + \frac{1}{2} b^2 \sigma^2} \qquad \ldots (4.32)$$

$$m_3' = m_1'^3 \, e^{3\sigma^2} + 3m_1'(m_2' - m_1'^2 \, e^{\sigma^2}) \, e^{b\sigma^2} = f_b(\sigma^2), \text{ say} \ldots (4.33)$$

We assume that the moments m_1', m_2', m_3' and m_3 of x are all positive.

We then prove some results about the behaviour of the function $f_b(\sigma^2)$.

Result 1 : $a^2 > 0$ iff $m_2' - m_1'^2 e^{\sigma^2} > 0$ or $\sigma^2 < \sigma^{*2} = \log\left(\frac{m_2'}{m_1'^2}\right)$.

Proof: Use equation (4.32).

Since $a^2 > 0$ is a necessary condition for our purpose, we shall henceforth investigate the properties of $f_b(\sigma^2)$ for $\sigma^2 < \sigma^*$ only.

Result 2 : $\left.\frac{\partial f_b(\sigma^2)}{\partial \sigma^2}\right|_{\sigma^2 = \sigma^{*2}} \begin{cases} > 0 & \text{if } b < 2 \\ = 0 & \text{if } b = 2 \\ < 0 & \text{if } b > 2 \end{cases}$

Result 3 : $\left.\frac{\partial f_b(\sigma^2)}{\partial \sigma^2}\right|_{\sigma^2=0} \geq 0$ if $b \geq 0$ which is already assumed.

Result 4 : $f_{b_1}(\sigma^2) > f_{b_2}(\sigma^2)$ iff $b_1 > b_2$ for $\sigma^2 < \sigma^{*2}$.

Result 5 : $\frac{\partial f_b(\sigma^2)}{\partial \sigma^2} > 0$ for $\sigma^2 < \sigma^{*2}$ and $b \leq 2$.

The behaviour of the function $f_b(\sigma^2)$ for different b, as evident from the above results, is drawn in the diagram (4.4).

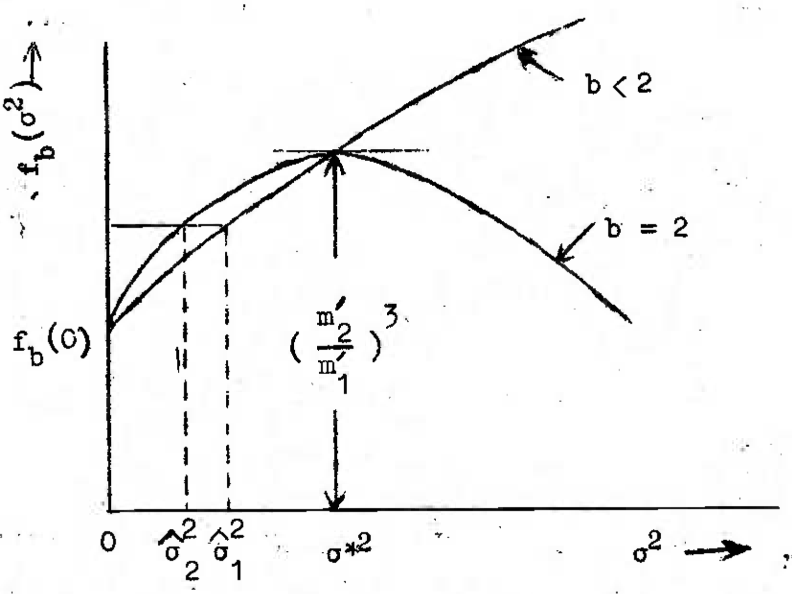


Fig. 4.4

Solutions for σ^2 must be sought in the interval $(0, \log \frac{m_2'}{m_1'^2})$ to ensure $a^2 > 0$. Hence, if $b \leq 2$ we have unique solution for σ^2 if

$f_b(0) < m_3' < (\frac{m_2'}{m_1'})^3$

or (i) $m_3 > 0$

and (ii) $m_3' < (\frac{m_2'}{m_1'})^3$.

Observe that the conditions thus arrived at are same as that of the special case where b = 0 or b = 2.

If $b > 2$ then there may not be any unique solution. However, $b > 2$ is quite unlikely.

4.3 The Derivation of the p.d.f. for x and the Maximum Likelihood Equations

For the sake of convenience, we reformulate the model defined by equations (4.2), (4.3) and (4.4) as

$$x_i = X_i + u_i = X_i(1 + \frac{u_i}{X_i}) = X_i(1 + u_i') \quad \ldots \quad (4.34)$$

Also assume that b=2. Note that $E(u_i'|X_i)$ and $V(u_i'|X_i)(=a^2)$ do not depend on X_i. We shall assume that X_i and u_i' are independent. Notationally, if we assume that X has a distribution function F and density function f and u' has a symmetric and unimodal density function g, then

$$\begin{aligned} P(c) &= P(X \leq c) \\ &= P\{X(1 + u') \leq c\} \\ &= \int_0^\infty P\{X(1 + u') \leq c | X\} f(X)dX \\ &= \int_0^\infty P\{u' < \frac{c}{X} - 1\} f(X)dX \\ &= \int_0^\infty G(\frac{c}{X} - 1) f(X)dX \quad \ldots \quad (4.35) \end{aligned}$$

Specific distributional assumptions for u and X may now be made to get the distribution function of x.

Model 1 : $u \sim P_{II}(-X, X, m)$ i.e., $u' \sim P_{II}(-1, 1, m)$

$$g(u') = K_m(1 - u'^2)^m \quad \text{where } K_m = B^{-1}(m + 1, \tfrac{1}{2}) .$$

$$\therefore G(\tfrac{c}{X} - 1) = \begin{cases} 1 & \text{if } \frac{c}{2} \geq X \\ \int_{-1}^{\frac{c}{X}-1} K_m(1-v^2)^m \, dv & \text{if } \frac{c}{2} < X \end{cases}$$

From (4.35), we have,

$$P(c) = P(x \leq c)$$

$$= \int_0^{c/2} f(X)\,dX + \int_{c/2}^{\infty} \left(\int_{-1}^{\frac{c}{X}-1} K_m(1-v^2)^m \, dv \right) f(X)\,dX$$

$$= F(X/2) + \int_{-1}^{1} K_m(1-v^2)^m \left(\int_{c/2}^{\frac{c}{v+1}} f(X)\,dX \right) dv$$

$$= \int_{-1}^{1} K_m(1-v^2)^m \, F\left(\frac{c}{v+1}\right) dv$$

Writing x for c

$$P(x) = \int_{-1}^{1} K_m(1-v^2)^m \, F\left(\frac{x}{v+1}\right) dv \qquad \ldots (4.36)$$

$$\therefore \quad p(x) = \int_{-1}^{1} K_m(1-v^2)^m \, f\left(\frac{x}{v+1}\right) \frac{1}{v+1} \, dv$$

$$= \int_{-1}^{1} K_m(1-v)^m (1+v)^{m+1} \, f\left(\frac{x}{v+1}\right) dv \qquad \ldots (4.37)$$

Putting $f = \frac{1+v}{2}$,

$$p(x) = K_m \int_0^1 2^m (1-t)^m \, 2^{m-1} \, t^{m-1} \, f\left(\frac{x}{2t}\right) 2\,dt$$

$$= K_m \, 2^{2m} \int_0^1 t^{m-1} (1-t)^m \, f\left(\frac{x}{2t}\right) dt \qquad \ldots (4.38)$$

Taking a different transformation, one gets,

$$p(x) = K_m (2x)^m \int_{x/2}^{\infty} \frac{f(t)}{t^{m+1}} \left(1 - \frac{x}{2t}\right)^m dt \qquad \ldots \quad (4.39)$$

The rth order moment about zero is

$$\mu_r' = E(x^r) = E\{X^r (1 + u')^r\}$$

$$= E(X^r)\ E(1+u')^r \quad \text{since X and } u' \text{ are independent}$$

$$= E(X^r) \int_{-1}^{1} K_m (1-u'^2)^m (1+u')^r du'$$

$$= E(X^r) \int_{-1}^{1} K_m (1-u')^m (1+u')^{r+m} du'$$

$$= E(X^r) \int_0^1 K_m\, 2^{2m+r+1}\, t^{m+r} (1-t)^m dt$$

$$= E(X^r)\ K_m\ 2^{2m+r+1}\ B(m+r+1,\ m+1)^{4.4/} \qquad \ldots \quad (4.40)$$

To find the likelihood function we assume X to follow a lognormal distribution with parameters μ and σ^2. Symbolically, let $X \sim \Lambda(\mu, \sigma^2)$ with distribution function $\Lambda(X \mid \mu, \sigma^2)$ and density function $\lambda(X \mid \mu, \sigma^2)$. the likelihood function is

$$L = \prod_{i=1}^{n} p(x_i) \qquad \ldots \quad (4.41)$$

or

$$\log L = \sum_{i=1}^{n} \log p(x_i) \qquad \ldots \quad (4.42)$$

4.4/ Alternatively also from the density function, one gets,

$$E(x^r) = 2^{2m} Km \int_0^1 w^{m-1} (1-w)^m \{\int_0^{\infty} x^r f(x/2w) dx\} dw$$

$$= 2^{2m} K_m \int_0^1 w^{m-1}(1-w)^m \{\int_0^{\infty} \left(\frac{x}{2w}\right)^r (2w)^r f\left(\frac{x}{2w}\right) d\left(\frac{x}{2w}\right) 2w\} dw$$

$$= E(X^r)\ 2^{2m} K_m \int_0^1 2^{r+1}\ w^{m+r} (1-w)^m dw$$

$$= E(X^r)\ K_m\ 2^{2m+r+1}\ B(m+r+1,\ m+1).$$

We maximize log L with respect to μ, σ^2 and m. Taking partial derivatives and equating them to zero, we get,

$$\frac{\partial \log L}{\partial \mu} = 0$$

$$\text{or} \quad \mu = \frac{1}{n}\sum_{i=1}^{n} \frac{\int_{x_i/2}^{\infty} \frac{\lambda(X)}{X^{m+1}}\left(1 - \frac{x_i}{2X}\right)^m \log X \, dX}{\int_{x_i/2}^{\infty} \frac{\lambda(X)}{X^{m+1}}\left(1 - \frac{x_i}{2X}\right)^m dX} \quad \ldots (4.43)$$

$$\frac{\partial \log L}{\partial \sigma} = 0$$

$$\text{or} \quad \sigma^2 = \frac{1}{n}\sum_{i=1}^{n} \frac{\int_{x_i/2}^{\infty} \frac{\lambda(X)}{X^{m+1}}\left(1 - \frac{x_i}{2X}\right)^m (\log X - \mu)^2 \, dX}{\int_{x_i/2}^{\infty} \frac{\lambda(X)}{X^{m+1}}\left(1 - \frac{x_i}{2X}\right)^m dX} \quad \ldots. (4.44)$$

and $\frac{\partial \log L}{\partial m} = 0$

$$\text{or} \quad \frac{1}{2}\sum_{k=0}^{\infty} \frac{1}{(K+m+1)\,(K+m+\frac{3}{2})} \ \underline{4.5/}$$

$$= -\frac{1}{n}\sum_{i=1}^{n} \frac{\int_{x_i/2}^{\infty} \frac{\lambda(X)}{X^{m+1}}\left(1 - \frac{x_i}{2X}\right)^m \log\left(1 - \frac{x_i}{2X}\right) dX}{\int_{x_i/2}^{\infty} \frac{\lambda(X)}{X^{m+1}}\left(1 - \frac{x_i}{2X}\right)^m dX} \quad \ldots. (4.45)$$

where $\lambda(X) = \frac{1}{\sqrt{2\pi}\,\sigma X} \, e^{-\frac{1}{2\sigma^2}(\log X - \mu)^2}$

$\underline{4.5/}$ $-\log K_m = \log\sqrt{(m+1)} - \log\sqrt{(m+3/2)} + \log\sqrt{(1/2)}$ since

$$K_m = \sqrt{(m+1)}\ \sqrt{(1/2)} \,/\, \Gamma(m+3/2)$$

$$\therefore \frac{\partial \log K_m}{\partial m} = \left(\gamma + \frac{1}{m+1} + \sum_{k=1}^{\infty}\frac{1}{K+m+1} - \sum_{k=1}^{\infty}\frac{1}{K}\right)$$

$$- \left(\gamma + \frac{1}{m+3/2} + \sum_{k=1}^{\infty}\frac{1}{K+m+3/2} - \sum_{k=1}^{\infty}\frac{1}{K}\right)$$

$$= \frac{1}{2}\sum_{k=0}^{\infty}\frac{1}{(K+m+1)(K+m+3/2)}$$

Model 2: $u' \sim N(0, a^2)$:

$$P(x) = \int_0^\infty \left(\int_{-\infty}^{\frac{x}{X}-1} \frac{1}{\sqrt{2\pi}\, a} e^{-t^2/2a^2} dt \right) f(X)\, dX$$

and

$$p(x) = \frac{dP(x)}{dx}$$

$$= \int_0^\infty \frac{1}{\sqrt{2\pi}\, aX} e^{-\frac{1}{2a^2 X^2}(x - X)^2} f(X)\, dX \quad .... (4.46)$$

From this or otherwise we have

$$E(x) = E(X)$$

$$E(x^2) = (1 + a^2)\, E(X^2)$$

andso on.

4.4 Other Moment Estimators of Model 1 :

Since the sampling variance increases very rapidly as the order of the moments increases, it may be advisable not to utilize moments of order more than 2 if we can avoid them. From this point of view, the moment estimators suggested above may not be the most advantageous ones. Since, however, the number of parameters to be estimated is 3 or 4, we need some sample moments other than m'_1 and m'_2 for estimation by method of moments. If we make distributional assumptions on $u|X$, such as $u|X \sim P_{II}(-X, X, m)$, we can use m'_r where r is fractional like 0.5 or 1.5 (say). In the set up given by Model 1, one can take any three of such moments with order r not exceeding 2 and solve the three equations for three unknowns. One can even allow r to tend to zero and take

$$\frac{1}{n} \Sigma \lim_{r \to 0} \frac{x^r - 1}{r} = \frac{1}{n} \Sigma \log x = \log G_x$$

where G_x is the geometric mean of x. We will call this the moment of order zero although $m_0' = 1$. Let us write m_r' for the equation based on the sample moment of order r in the estimation by method of moments. Obviously for r = 0, we take

$$m_1'(\log x) = E(\log x) = E(\log X) + E\{\log (1+u')\}$$

$$= \mu + \int_{-1}^{1} K_m(1-u'^2)^m \log(1+u')du'$$

$$= \mu + \int_{-1}^{1} K_m(1-u'^2)^m \left[u' - \frac{u'^2}{2} - \frac{u'^3}{3} \ldots \right] du'$$

$$= \mu - \int_{-1}^{1} K_m(1-u'^2)^m \left\{ \frac{u'^2}{2} + \frac{u'^4}{4} + \ldots \right\} du'$$

$$= \mu - \sum_{r=1}^{\infty} \mu_{2r}'(u')/(2r) \qquad \ldots \quad (4.47)$$

where $\mu_{2r}'(u') = K_m \; 2^{2m+2r+1} \; B(m+2r+1, m+1)$. Below are reported the results of Monte-Carlo study where ML estimation were compared with various types of method of moments estimators, including some where the order of moments used was fractional or zero.

4.5 The Monte-Carlo Results

The Monte-Carlo experiment was conducted to compare ML estimators and various moment estimators for model (1) of section (4.3) of this chapter. For this purpose, 50 samples of hundred observations each were generated for x values assuming the parameters (μ, σ^2, m) to be (1, 0.5, 2). Since

$$x = X(1 + u')$$

and X and u' are independent; the samples were generated for X and u' separately. The distribution of u' is

$$P(u' < u^*) = \int_{-1}^{u^*} K_2 (1 - u'^2)^2 \, du'$$

$$= K_2 \int_{-1}^{u^*} (1 + u'^4 - 2u'^2) \, du'$$

$$= K_2 \left[u^* + 1 + \frac{u^{*5}}{5} + \frac{1}{5} - 2\frac{u^{*3}}{3} - \frac{2}{3} \right]$$

$$= \frac{15}{16} \left[\frac{8}{15} + u^* - \frac{2u^{*3}}{3} + \frac{u^{*5}}{5} \right]$$

$$= \frac{1}{2} + \frac{u^*}{16} (15 - 10u^{*3} + 3u^{*5}) .$$

Thus for $P_{II}(-1, 1, 2)$, random numbers were converted into u' values using explicit expression for distribution function of u'. Generation of X values was not very difficult since $(\log X-1)/\sqrt{0.5}$ is standard normal and tables of standard normal deviates can be found in many books (The Rand Corporation, 1955, Wold, 1955).

For each sample we computed the following estimators of (μ, σ^2, m):

(a) ML estimators

(b) Moment estimators using equations :

(i) m_1' (log x), m_1' and m_2'

(ii) $m'_{.25}$, m_1' and m_2'

(iii) $m'_{.5}$, m_1' and m_2'

(iv) $m'_{.75}$, m_1' and m_2'

(v) $m'_{1.5}$, m_1' and m_2'.

The likelihood equations (4.43), (4.44) and (4.45) were solved through iterative process. Given the initial values μ_o, σ_o^2, and m_o, which can be estimated from moment equations m_1', m_2' and m_3', one can straight find μ_1 and σ_1^2. To find m_1 we substitute μ_o, σ_o^2 and m_o in the right hand side which equates them. This again needs another iteration. Since left hand side of (4.45) is a monotone function of m, there exist a unique m which equates them. The proof for the convergence of the main iterative process, could not be found out. However, in the Monte-Carlo study the process converged for all the 20 samples. The initial values for (μ, σ^2, m) were taken to be (1, 0.5, 2), the true values of the population parameters. For some of the samples the initial values were taken to be different to see whether they converged to the same solution. The process was stopped as soon as it was found that the absolute value of the successive solutions for μ and σ^2 were less than .001, which was taken to be the criterion of convergence.

For solving the likelihood equations the integrals were computed by trapezoidal rule, after computing ordinates at intervals varying from .01 to .10 depending on the second order differences of the densities. The method of moments is far more expeditious. To solve the likelihood equations one must make use of electronic computer whereas the moment equations can be solved much more easily, though here also we need to use iterative process. Considering the amount of time it takes, we covered only 20 samplesfor ML estimation. The results are summarized in Tables (4.1) and (4.2).

Table 4.1 : Comparison of Different Estimation Procedures : A Monte-Carlo Study

Procedures	No. of Samples	Bias		Standard Error		Mean Squared Error		Bias/S.E.	
		$\bar{\mu}-\mu$	$\bar{\sigma^2}-\sigma^2$	S.E. (μ)	S.E (σ^2)	MSE (μ)	MSE (σ^2)	μ	σ^2
ML Method	20	-.0001	.0050	.0831	.1049	.0069	.0110	-.0012	.0477
Method of Moments Using									
(i) $m'_1(\log x)$, m'_1 and m'_2	20	.0045	-.0015	.0878	.1296	.0077	.0168	.0513	-.0116
(ii) $m'_{.25}$, m'_1 and m'_2	20	.0205	-.0350	.0985	.1371	.0101	.0200	.2081	-.2553
(iii) $m'_{.5}$, m'_1 and m'_2	20	.0235	-.0385	.0990	.1421	.0104	.0217	.2374	-.2709
(iv) $m'_{.75}$, m'_1 and m'_2	20	.0255	-.0440	.1010	.1460	.0109	.0232	.2525	-.3015
(v) $m'_{1.5}$, m'_1 and m'_2	20	.0415	-.0615	.1058	.1631	.0129	.0304	.3921	-.3771

Table 4.2 : Comparison of Different Moment Estimators : A Monte-Carlo Study

Procedures	No. of Samples	Bias		Standard Error		Mean Squared Error		Bias/S.E.	
		$\bar{\mu}-\mu$	$\bar{\sigma}^2-\sigma^2$	S.E. (μ)	S.E. (σ^2)	MSE (μ)	MSE (σ^2)	μ	σ^2
Method of Moments Using									
(i) m_1' (log x), m_1' and m_2'	50	.0168	−.0172	.0933	.1261	.0090	.0162	.1801	−.1364
(ii) $m_{.25}'$, m_1' and m_2'	50	.0386	−.0516	.0995	.1261	.0114	.0197	.3879	−.4885
(iii) $m_{.5}'$, m_1' and m_2'	50	.0412	−.0664	.1010	.1296	.0119	.0218	.4079	−.6123
(iv) $m_{.75}'$, m_1' and m_2'	50	.0434	−.0714	.1025	.1364	.0124	.0237	.4234	−.5235
(v) $m_{1.5}'$, m_1' and m_2'	50	.0562	−.0906	.1058	.1483	.0144	.0302	.5312	−.6109

It is merely a coincidence that the bias of ML estimate of μ has been so small. Otherwise, if we compare the MSE's they don't seem to differ very much, at least between a and b(i). Biases as well as MSE's go on increasing as we go down in the table. Observe that in the moment equations m_1' and m_2' have been taken common for all the procedures. The third equation with moments of fractional order (or zero order, i.e., the mean of logarithms) have been taken differently for different sets. As one goes down in the table this power increases. The efficiencies of the estimators by b(i) method is not much less than those of ML estimators. Considering the difficulties one faces in solving likelihood equations it is advisable to adopt the procedure b(i).

4.6 Concluding Observations

The efficiencies of different set of moment estimators may be compared with that of ML estimators for other models also. Simplicity of moment estimators should be stressed while comparing the efficiencies. Time and cost considerations may prove to be prohibitive for ML estimation.

In the above Monte-Carlo experiment, we could as well take the following two sets of moment estimators :

(i) m_1', m_2' and m_3'

(ii) m_1' (log x), m_2'(log x), m_3' (log x)

The set (m_1', m_2', m_3') would obviously give very low efficiency as can be predicted from the above table. But, procedure based on (m_1', m_2', m_3')

does not assume any distribution for $u|X$. Regarding the second procedure one should note that

$$\mu_3 (\log x) = K_3 (\log x),$$

where K_3 denotes the third order cumulant. If $X \sim \Lambda(\mu, \sigma^2)$ and $u|X \sim P_{II} (-X, X, m)$, then $\log X$ and $\log (1 + \frac{u}{X})$ are independent. Thus one gets

$$\begin{aligned} K_3(\log x) &= K_3 (\log X) + K_3 (\log (1+ \tfrac{u}{X})) \\ &= \mu_3 (\log X) + \mu_3 (\log (1 + \tfrac{u}{X})) \\ &= \mu_3 (\log (1 + \tfrac{u}{X})). \end{aligned}$$

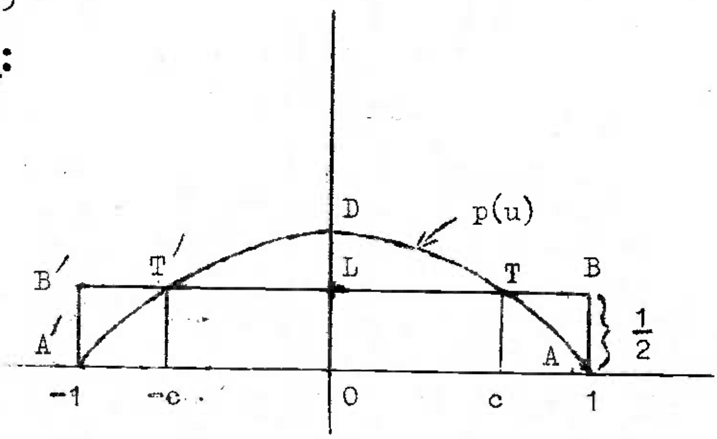
$\mu_3 (\log (1 + \frac{u}{X}))$ is a function of m only (say f(m)). So, given the sample estimate of $\mu_3 (\log x)$ one can easily estimate m. If the table of values f(m) corresponding to m is at hand, this procedure gives a simple way of estimating m. The other two parameters can easily be estimated by taking m_1' (log x) and m_2'(log x). The use of m_1' (log x) and m_2' (log x) is very efficient especially for small values of a^2. This idea, though very promising, could not be pursued further, since this came into view after analyzing Monte-Carlo results.

Appendix 4.1

Let u be a random variable distributed symmetrically about zero.

__Theorem A4.1__ : If p.d.f. of u increases monotonically from zero at -1 to a maximum value at 0 and again decreases monotonically to 0 at 1 then $V(u) \leq \frac{1}{3}$.

__Proof:__



As in the above diagram, consider another random variable v which follows a uniform distribution from -1 to 1. Let us suppose that the two density functions intersect at points $-c$ and c $(c > 0)$. From symmetry and the property of p.d.f. it follows that

The area of the segment $A'B'T'$

= The area of the segment $T'LD$

= The area of the segment DLT

= The area of the segment TAB

= Δ, say.

$$\int_{-1}^{1} u^2\, p(u)\, du - \frac{1}{3}$$

$$= \int_{-1}^{1} u^2 \{ p(u) - \tfrac{1}{2} \}\, du$$

$$= 2 \int_{0}^{1} u^2 \{ p(u) - \tfrac{1}{2} \}\, du$$

$$= 2 \int_{0}^{c} u^2 \{ p(u) - \tfrac{1}{2} \}\, du + 2 \int_{c}^{1} u^2 \{ p(u) - \tfrac{1}{2} \}\, du$$

$$\leq 2c^2\Delta - 2c^2\Delta = 0 \qquad \text{Q.E.D.}$$

Appendix 4.2

The Problem of Identification in an Univariate Model

It has already been mentioned that if $u|X \sim N(0, a^2)$ (so that if u and X are independent) and if $X \sim N(\mu, \sigma^2)$ then the model $x = X + u$ is not identifiable if only x's are observed. We shall now show that if in the same model u and X are not independent; to be more precise, if $u|X \sim N(0, a^2 X^2)$ then also the model is not identifiable. In particular, the parameters a^2 and σ^2 are not identifiable. The moment equations in this case are

$$m_1' = \mu \quad \ldots \quad (A4.1)$$

$$m_2' = (\mu^2 + \sigma^2)(1 + a^2) \quad \ldots \quad (A4.2)$$

$$m_3' = (\mu^3 + 3\mu\sigma^2)(1 + 3a^2) \quad \ldots \quad (A4.3)$$

From the symmetry in the right hand side of (A4.2) and (A4.3) one can easily verify that if (a_1^2, σ_1^2) is a solution of (a^2, σ^2) then $\left(\frac{\sigma_1^2}{m_1'^2}, a_1^2 m_1'^2\right)$ is a solution of (a^2, σ^2). But this is not a proof of identifiability. To prove that the model is not identifiable, we derive the p.d.f. of x to be

$$p(x|a^2, \sigma^2) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\, aX} e^{-\frac{1}{2a^2}\left(\frac{x}{X}-1\right)^2} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2\sigma^2}(X-\mu)^2} dX \quad \ldots \quad (A4.5)$$

The Likelihood function $L(a^2, \sigma^2)$ is the product of n such integrals with $x = x_1, x_2, \ldots, x_n$. Let (a_1^2, σ_1^2) be the true value of the parameters, then

Theorem A4.2 : The likelihood function $L(a_1^2, \sigma_1^2) = L(\frac{\sigma_1^2}{\mu^2}, a_1^2 \mu^2)$.

Proof: Write $(\frac{\sigma_1^2}{\mu^2}, a_1^2 \mu^2)$ instead of (a^2, σ^2) in (A4.4) and get

$$p(x \mid \frac{\sigma_1^2}{\mu^2}, a_1^2 \mu^2) = \int_{-\infty}^{\infty} \frac{\mu}{\sqrt{2\pi}\sigma_1 X} e^{-\frac{\mu^2}{2\sigma_1^2}(\frac{x}{X}-1)^2} \frac{1}{\sqrt{2\pi}a_1\mu} e^{-\frac{(X-\mu)^2}{2a_1^2\mu^2}} dX$$

Put $t = X/\mu$, $dt = dX/\mu$

$$\therefore p(x \mid \frac{\sigma_1^2}{\mu^2}, a_1^2\mu^2) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}a_1 t} e^{-\frac{(t-1)^2}{2a_1^2}} \frac{1}{\sqrt{2\pi}\sigma_1} e^{(\frac{x}{t}-\mu)^2/(2\sigma_1^2)} dt$$

Put $w = x/t$, $dw = -(x/t^2)$, $dt = -\frac{dt}{x} w^2$

$$\therefore p(x \mid \frac{\sigma_1^2}{\mu^2}, a_1^2\mu^2) = \int_{-\infty}^{\infty} \frac{w}{x} \frac{1}{\sqrt{2\pi}a_1} e^{-\frac{1}{2a_1^2}(\frac{x}{w}-1)^2} \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{(w-\mu)^2}{2\sigma_1^2}} \frac{x}{w^2} dw$$

$$= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}a_1 X} e^{-\frac{1}{2a_1^2}(\frac{x}{X}-1)^2} \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{1}{2\sigma_1^2}(X-\mu)^2} dX$$

$$= p(x/a_1^2, \sigma_1^2). \qquad \text{Q.E.D.}$$

We are getting the same density of observed x's for two different set of parameters. Hence the model is not identifiable.

Chapter 5

ESTIMATION PROBLEMS IN MORE GENERAL EIV MODELS

PART II

5.1 The Bivariate EIV Model

As a continuation from the previous chapter here we discuss the estimation problem of some bivariate models with EIV's through moments only. Except in section (5.3), the true regressor variable X has been assumed to follow a two parameter lognormal distribution throughout this chapter.

To see how the estimation problem arises in the bivariate set-up, let us consider the linear model

$$Y_i = \alpha + \beta X_i + \varepsilon_i, \qquad i = 1, 2, \ldots, n, \qquad \ldots. (5.1)$$

where X_i and Y_i are true but non-observable magnitudes of regressor and regressand respectively; α and β are unknown parameters to be estimated from data; ε_i is the disturbance assumed to be independent of other variables not involving ε_i and

$$E(\varepsilon_i) = 0, \quad V(\varepsilon_i) = \sigma_\varepsilon^2; \qquad i = 1, 2, \ldots, n. \qquad \ldots. (5.2)$$

The regressor X_i is stochastic. The observed values of the regressor and the regressand are respectively, x_i and y_i with

$$\left.\begin{aligned} x_i &= X_i + u_i \\ y_i &= Y_i + v_i \end{aligned}\right\} \qquad \ldots. (5.3)$$

where u_i and v_i are EIV's with zero means and variances σ_u^2 and σ_v^2 respectively. They are also uncorrelated with the true components X_i and Y_i. All variables are assumed to be i.i.d. for $i = 1, 2, \ldots, n$.

The above model, assuming that u's and v's are not necessarily independent of each other, leads to the following expression for the limiting value of the OLS estimate of β :

$$\beta_L = \underset{n \to \infty}{\text{plim}} \; \hat{\beta}_L = \frac{\beta\sigma_X^2 + \text{Cov}\,(u, v)}{\sigma_X^2 + \sigma_u^2}, \qquad \ldots (5.4)$$

where Cov (u,v) is not necessarily zero. If one assumes a linear dependence between the two errors, i.e.,

$$v_i = \lambda u_i + w_i, \qquad \ldots (5.5)$$

and if one makes the usual assumptions of classical linear regression model for (5.5) (Goldberger, 1964, p.162) excepting that u_i is stochastic, then one gets

$$\beta_L = \frac{\sigma_X^2\,\beta + \sigma_u^2\,\lambda}{\sigma_X^2 + \sigma_u^2}.$$

$$= \frac{\sigma_X^2}{\sigma_X^2 + \sigma_u^2}\,\beta + \frac{\sigma_u^2}{\sigma_X^2 + \sigma_u^2}\,\lambda. \qquad \ldots (5.6)$$

So, β_L is the weighted average of β and λ, so that the OLS estimate will not be consistent unless either $\sigma_u^2 = 0$, which is contrary to our assumptions, or $\beta = \lambda$, which is unlikely (Rao and Miller, 1972, p.181). Similar problems arise if nonlinear equations, such as semilog, hyperbola etc., connect X and Y in place of (5.1).

From (5.6) it is clear that consistent estimation of β is possible if we have consistent estimates of σ_X^2, σ_u^2 and λ. In the previous chapter we have tackled the problem of finding consistent estimates of σ_X^2 and σ_u^2 from the assumptions on x, viz., $x_1, x_2, \ldots, x_n$, by the method of moments and also by the method of ML under the following assumptions :

(i) $E(u|X) = 0$

(ii) $E(u^2|X) = a^2X^b$

(iii) $E(u^3|X) = 0$

and (iv) $X \sim \Lambda(\mu, \sigma^2)$

for usual moment estimation, and

(i) $u|X \sim P_{II}(-X, X, m)$

(ii) $X \sim \Lambda(\mu, \sigma^2)$

for fractional moment and ML estimation. Under further assumptions consistent estimation of b and a^2 is possible in the same situation. We shall now see how we can make use of these results in bivariate problems (such as those arising in engel curve analysis) for estimating β (and λ) from observations (x_i, y_i), $i = 1, 2, \ldots, n$.

In the next section of this chapter we discuss how one may get, by method of moments, consistent estimates of parameters for different algebraic forms of the relationship between the true values. The results are based on the assumption that X is lognormally distributed. Other distributional forms can also be handled in the same manner. Some results further assume that the errors in the regressor follows a Pearsonian type II distribution. Section (5.3) attempts to get consistent estimates of parameters without making any distributional assumptions about the true regressor. Here however, the relationship between the true values is assumed to be linear. The concluding section points out some limitations of the results obtained and by making some suggestions for further work along this line (vide Pal, 1977).

5.2 Estimation Assuming Specific Distributions for X and u

As done in the previous chapter we first assume some specific values for b (= 0 and 2). Later on, we tackle the problem for the more general case, where b is specified but has some other values. It should be recalled that b can be estimated from the observations $x_1, x_2, \ldots, x_n$. While the approach is quite general, the concrete results are generally based on the assumption of lognormality of X-values.

5.2.1 The Case Where b = 0

We shall consider three algebraic forms of the regression of Y on X.

Linear : Here equation (5.1) is appropriate. The moment equations are

$$m_{11} = \beta\, \mu_2(X) + \lambda a^2 \quad \ldots (5.7)$$

$$m_{21} = \beta\, \mu_3(X) = \beta\mu_3(x) \quad \ldots (5.8)$$

$$m_{12} = \beta^2\mu_3(X) = \beta^2\mu_3(x) \quad \ldots (5.9)$$

$$m_{03} = \beta^3\mu_3(X) = \beta^3\mu_3(x) \quad \ldots (5.10)$$

$$m_{30} = \mu_3(X) = \mu_3(x) \quad \ldots (5.11)$$

where

$$m_{ij} = \frac{1}{n}\sum_{K=1}^{n} (x_K - \bar{x})^i\, (y_K - \bar{y})^j$$

with

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

Hence, we get the same set of estimators of β as those in the case where $\lambda = 0$ (vide Chapter 3), since equations (5.8) to (5.10) do not contain λ. The three "basic" estimators are

$$\hat{\beta}_1 = \frac{m_{03}}{m_{12}}, \quad \hat{\beta}_2 = \frac{m_{12}}{m_{21}}, \quad \text{and} \quad \hat{\beta}_3 = \frac{m_{21}}{m_{30}}.$$

Observe that contrary to what is implied in the section heading no distributional assumptions about X is necessary to estimate β and other parameters like $\mu_3(X)$. But for estimation of λ from (5.7) distributional assumption would be needed. λ being different from zero the comparative efficiency of the estimators will of course be different from those where $\lambda=0$.

Hyperbola : Here the relationship concerning the true components is

$$Y = \alpha + \beta / X + \varepsilon . \qquad \dots (5.12)$$

We also make the assumption that $X \sim \Lambda(\mu, \sigma^2)$. The following moment equations may be used in this case :

$$m_{11} = \beta \operatorname{Cov}(X, \tfrac{1}{X}) + \lambda a^2 \qquad \dots (5.13)$$

$$m_{21} = \beta \left[\operatorname{Cov}(X^2, \tfrac{1}{X}) - 2E(x) \operatorname{Cov}(X, \tfrac{1}{X}) \right]. \qquad \dots (5.14)$$

So,

$$\hat{\beta} = m_{21} \Big/ \left[\widehat{\operatorname{Cov}}(X^2, \tfrac{1}{X}) - 2 \bar{x} \widehat{\operatorname{Cov}}(X, \tfrac{1}{X}) \right] \dots (5.15)$$

and,

$$\hat{\lambda} = \{ m_{11} - \hat{\beta} \widehat{\operatorname{Cov}}(X, \tfrac{1}{X}) \} / \hat{a}^2 . \qquad \dots (6.16)$$

It is not difficult to estimate $\operatorname{Cov}(X, \frac{1}{X})$ and $\operatorname{Cov}(X^2, \frac{1}{X})$ as we already know the estimates of parameters μ and σ^2 from the corresponding univariate model.

Semilog : The semilog equation is

$$Y = \alpha + \beta \log X + \varepsilon \qquad \ldots (5.17)$$

The assumptions involved are the same as those in previous case. The moment equations are

$$m_{11} = \beta \, Cov\,(X, \log X) + \lambda a^2 \qquad \ldots (5.18)$$

and $$m_{21} = \beta \, \{Cov\,(X^2, \log X) - 2E(X)\, Cov.(X, \log X)\}. \quad \ldots (5.19)$$

β can be estimated from (5.19) exploiting the lognormality of X and the estimates of μ and σ^2. Then λ is obtained from (5.18) using the estimates $\hat{\beta}$, $\hat{a}^2$ and $\widehat{Cov}.\,(X, \log X)$.

5.2.2 The Case Where b=2

Again we consider three algebraic forms.

Linear : Assuming b=2 the following moment equations may be used to estimate β :

$$m_{11} = \beta\mu_2(X) + \lambda a^2\, E(X^2) \qquad \ldots (5.20)$$

$$m_{21} = \beta\,\{\mu_3(X) + a^2\, Cov\,(X^2, X)\} + \lambda.\, 2a^2\, Cov\,(X^2, X). \quad \ldots (5.21)$$

One can solve (5.20) and (5.21) to get β as

$$\beta = \frac{2m_{11}\, Cov\,(X^2, X) - m_{21}\, E(X^2)}{2\mu_2(X)\, Cov\,(X^2, X) - E(X^2)\,\{\mu_3(X) + a^2\, Cov\,(X^2, X)\}} \qquad \ldots (5.22)$$

To estimate β, we substitute estimates of $Cov\,(X^2, X)$, $\mu_2(X)$, $E(X^2)$ and $\mu_3(X)$.

Here it may be noted that if we estimate β by OLS method, then

$$\beta_L = \underset{n \to \infty}{\text{plim}}\ \hat{\beta}_L = \frac{\beta\mu_2(X) + \lambda\sigma_u^2}{\mu_2(X) + \sigma_u^2}. \qquad \ldots (5.23)$$

or $$\beta = \frac{\beta_L\, \mu_2(x) - \lambda\, \sigma_u^2}{\mu_2(X)}. \qquad \ldots (5.24)$$

If $\lambda \to 0$ then $\beta = \frac{\beta_L \cdot \mu_2(X)}{\mu_2(X)}$ which is nothing but the case where Cov (u, v) $\neq 0$; and in that case we estimate β from (5.20) only. Of course, we take the estimate of $\mu_2(X)$ from the corresponding univariate model.

<u>Hyperbola</u> : The moment equations to be used are

$$m_{11} = \beta \operatorname{Cov}(X, \tfrac{1}{X}) + \lambda a^2 E(X^2) \qquad \ldots (5.25)$$

$$m_{21} = \beta \{ \operatorname{Cov}(X^2, \tfrac{1}{X}) - 2E(X) \operatorname{cov}(X, \tfrac{1}{X}) + a^2 \operatorname{Cov}(X^2, \tfrac{1}{X}) \} + \lambda . 2a^2 \operatorname{Cov}(X, X^2) \qquad \ldots (5.26)$$

Solving (5.25) and (5.26), we get,

$$\beta = \frac{2m_{11} \operatorname{Cov}(X, X^2) - m_{21} E(X^2)}{2 \operatorname{Cov}(X, \frac{1}{X}) \operatorname{Cov}(X, X^2) - E(X^2) \{ \operatorname{Cov}(X^2, \frac{1}{X}) - 2E(X) \operatorname{Cov}(X, \frac{1}{X}) + a^2 \operatorname{Cov}(X^2, \frac{1}{X}) \}} \qquad \ldots (5.27)$$

<u>Semilog</u> : Here also we take the same moment equations and get,

$$m_{11} = \beta \operatorname{Cov}(X, \log X) + \lambda a^2 E(X^2) \qquad \ldots (5.28)$$

and

$$m_{21} = \beta \{ \operatorname{Cov}(X^2, \log X) - 2E(X) \operatorname{Cov}(X, \log X) + a^2 \operatorname{Cov}(X^2, \log X) \} + 2\lambda a^2 \operatorname{Cov}(X, X^2) \qquad \ldots (5.29)$$

The solution for β is

$$\beta = \frac{2m_{11} \operatorname{Cov}(X, X^2) - m_{21} E(X^2)}{2 \operatorname{Cov}(X, \log X) \operatorname{Cov}(X, X^2) - E(X^2) \{ \operatorname{Cov}(X^2, \log X) - 2E(X) \operatorname{Cov}(X, \log X) + a^2 \operatorname{Cov}(X^2, \log X) \}} \qquad \ldots (5.30)$$

5.2.3 The Case Where $b \neq 0$ or 2 :

In this section we assume that b is known apriori or is estimated from the univariate model. We may take a more general algebraic form

$$Y = \alpha + \beta f(X) + \varepsilon , \qquad \dots (5.31)$$

which becomes straight line, hyperbola and semilog as special cases. The same moment equations give

$$m_{11} = \beta \, Cov\,(X, f(X)) + \lambda a^2 E(X^b), \qquad \dots (5.32)$$

$$m_{21} = \beta \, Cov\,(X^2, f(X)) - 2\beta E(X)\, Cov\,(X, f(X)) + 2\lambda a^2 Cov\,(X, X^b) + a^2 \beta \, Cov\,(X^b, f(X)) \qquad \dots (5.33)$$

which has the solution

$$\beta = \frac{m_{11} \cdot 2\,Cov\,(X, X^b) - m_{21}\, E(X^b)}{2\,Cov\,(X, f(X))\, Cov\,(X, X^b) - E(X^b)\, \{ Cov\,(X^b, f(X)) - 2E(X)\, Cov(X, f(X)) + a^2\, Cov\,(X^b, f(X)) \}} \qquad \dots (5.34)$$

Observe that putting b=0 or 2, we get the expressions for β given earlier.

The other three third order moment equations give

$$m_{30} = \mu_3(X) + 3a^2 Cov\,(X, X^b) \qquad \dots (5.35)$$

$$m_{12} = \beta^2 \{ Cov\,(X, f^2(X)) - 2E\,(f(X))\, Cov\,(X, f(X)) \} + 2\beta\lambda a^2 Cov\,(X^b, f(X)) + \lambda^2 a^2 Cov\,(X^b, X) \qquad \dots (5.36)$$

and $$m_{03} = \beta^2 \mu_3\,(f(X)) + 3\beta\lambda^2 a^2 Cov\,(f(X), X^b). \qquad \dots (5.37)$$

Special cases where b=0 or b=2 or where f(X) is X, log X, or $\frac{1}{X}$ are obtained in a straightforward manner from these equations. One may take the help of these equations also in estimating β for special cases, but the solution may be more difficult; and further, the efficiencies of the resultant estimators may be lower.

5.2.4 Other Moment Equations Assuming $u|X \sim P_{II}$ (-X, X, m)

So far we have not assumed anything specifically about the distribution of u given X except for its symmetry around zero (or rather that odd order moments vanish). Below we illustrate how the assumption of a specific distribution of $u|X$ may be helpful in some cases.

Let us take the case where b=2 and suppose that we are to estimate a semilog equation. We assume that

$$u|X \sim P_{II}\ (-X,\ X,\ m).$$

The p.d.f. of this distribution is

$$p(u|X)\,du = X^{-1}\ K_m (1-u^2/X^2)^m\ du,\quad -X \leq u \leq X, \qquad \ldots\ (5.38)$$

where $K_m = B^{-1}\ (m + \frac{1}{2}, \frac{1}{2}) = B^{-1}\ (m + 1, m + 1) 2^{-2m-1}$. Now,

$$\begin{aligned} x &= X + u \\ &= X\ (1 + u/X) \\ &= Xu'\ \text{(say)}. \qquad \ldots\ (5.39) \end{aligned}$$

X and u′ are independent of each other if u|X follows the Pearsonian type II distribution specified above. Again,

$$\begin{aligned} E(u' \log u') &= E\{u'(-(1-u') - \frac{(1-u')^2}{2} - \frac{(1-u')^3}{3} - \ldots\ldots)\} \\ &= E\,[-u'\{(1-u') + \frac{(1-u')^2}{2} + \frac{(1-u')^3}{3} + \ldots\ldots\}] \\ &= -E\{(1-u') + \frac{(1-u')^2}{2} + \frac{(1-u')^3}{3} \ldots\}+ E\{(1-u')^2 + \frac{(1-u')^3}{2} + \ldots\} \\ &= -E\{\frac{(1-u')^2}{2} + \frac{(1-u')^4}{4} + \ldots\} + E\{(1-u')^2 + \frac{(1-u')^4}{3} + \ldots\} \\ &= \mu_2(u')\ (1 - \frac{1}{2}) + \mu_4(u')\ (\frac{1}{3} - \frac{1}{4}) + \ldots\ldots \end{aligned}$$

$$= \frac{\mu_2(u')}{1.2} + \frac{\mu_4(u')}{3.4} + \frac{\mu_6(u')}{5.6} + \ldots \qquad \ldots (5.40)$$

and $E(u \log u' | X) = XE(u' \log u')$.

We know that

$$\mu_{2K}(u/X) = \frac{1.3.5.\ \ldots\ldots\ (2K-1)}{(2m+3)(2m+5)\ \ldots\ (2m+2K+1)}. \qquad \ldots (5.41)$$

Since m can be estimated from the univariate model, μ_{2K} can be estimated consistently. Instead of the following equation

$$m_{11} = \beta \text{ Cov } (X, \log X) + \lambda.\ a^2\ E(X^2), \qquad \ldots (5.42)$$

we may use the equation

$$m_{11}\ (\log x, y) = \beta\ V(\log X) + \lambda\ E(\ u \log u'). \qquad \ldots (5.43)$$

Cov (X, log X) is simply $\sigma^2\ e^{\mu + \sigma^2/2}$ and $V(\log X) = \sigma^2$ assuming $X \sim \Lambda(\mu, \sigma^2)$. After substituting estimates of $E(u \log u')$, a^2, Cov (X, log X), V (log X) etc., in (5.42) and (5.43) we can easily solve them to get consistent estimates of β and λ.

5.3 Estimation Without Distributional Assumption on X

In this section we arrive at Scott's (1950) estimate under a more general set-up than that assumed by her. Like Scott, we need not make any distributional assumption about the true regressor. Moreover b may take any value and this value may be unknown. We write down all the moment equations of the linear model up to order three for the sake of convenience.

$$m'_{10} = E(X) \qquad \ldots (5.44)$$

$$m'_{01} = \alpha + \beta\ E(X) \qquad \ldots (5.45)$$

$$m_{20} = \mu_2(X) + a^2 E(X^b) \qquad \ldots (5.46)$$

$$m_{11} = \beta\mu_2(X) + \lambda a^2\, E(X^b) \qquad \dots (5.47)$$

$$m_{02} = \beta^2\mu_2(X) + \lambda a^2\, E(X^b) + \mu_2(w) + \mu_2(\varepsilon) \qquad \dots (5.48)$$

$$m_{30} = \mu_3(X) + 3a^2\, B \qquad \dots (5.49)$$

$$m_{21} = \beta\mu_3(X) + a^2\,(2\lambda + \beta)B \qquad \dots (5.50)$$

$$m_{12} = \beta^2\mu_3(X) + a^2\lambda\,(\lambda + 2\beta)B \qquad \dots (5.51)$$

$$m_{03} = \beta^3\mu_3(X) + 3a^2\lambda^2\,\beta B \qquad \dots (5.52)$$

where $B = \mathrm{Cov}\,(X^b, X) = E(X^{b+1}) - E(X^b)\,E(X)$. We have nine equations to solve for ten unknowns, namely, $E(X)$, α, β, $\mu_2(X)$, λ, b, a^2, $\mu_2(w)$, $\mu_2(\varepsilon)$, and $\mu_3(X)$. However, $\mu_2(w)$ and $\mu_2(\varepsilon)$ appears only once and in the form of $\mu_2(w) + \mu_2(\varepsilon)$. So in effect, we have nine equations to solve for nine unknowns. The following theorem reaches the same result as Scott's, but, does so under a general set-up.

<u>Theorem 5.1</u> : Under the set-up defined by (4.2), (4.3), (4.4), (4.5), and (5.1), (5.2), (5.3) and (5.5)

$$\frac{1}{n}\sum_{i=1}^{n}\{(y_i - \bar{y}) - d(x_i - \bar{x})\}^3 = 0\,,$$

where $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$ and $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$, has a solution $\hat{d}$ which is a consistent estimator of β.

<u>Proof</u> : The proof is obvious since

$$\operatorname*{plim}_{n\to\infty}\{m_{03} - 3\beta m_{12} + 3\beta^2 m_{21} - \beta^3 m_{30}\} = 0 \text{ [5.1/]} \qquad \text{Q.E.D.}$$

5.1/ For the general form $Y = \alpha + \beta f(X) + \varepsilon$, we have

$$m_{03} - 3\beta m_{12} + 3\beta^2 m_{21} - \beta^3 m_{30} = \beta^3[\mu_3\{f(X)-X\} + 3a^2\,\mathrm{Cov}\,\{X^b, f(X)-X\}] + 2\lambda a^2\beta^2\,\mathrm{Cov}\{X^b, f(X)-X\} + 3\beta\lambda^2 a^2\,\mathrm{Cov}\,\{f(X)-X, X^b\},$$

and this vanishes if $f(X) = X$.

Scott proved the same thing under the following assumptions, which are unduly restrictive :

(i) $\mu_2(\varepsilon) = 0$;

(ii) u and v are normally distributed and independent of each other;

(iii) u and v are independent of their corresponding true components (X and Y); and

(iv) X is non-normal.

The major limitation of this estimation procedure is that it does not specify which solution is to be taken in case there are more than one real solutions. However, one can always check the feasibility conditions of it using equations (5.42), (5.43) and (5.44) by seeing for which solution the estimates of $\mu_2(\varepsilon) + \mu_2(w)$, a^2, $\mu_2(X)$ etc., are positive. But more than one solution may satisfy these checks. Existence of only one feasible solution would of course make the situation very smooth.

5.4 Concluding Remarks

One major limitation of the results in this chapter is that the estimators are obtained by method of moments and nothing is known about their efficiency vis-a-vis ML and other optimal estimators. It has not been possible to apply ML method for solving the problems.

In some applications, e.g., in engel curve analysis, the algebraic form of the regression equation may be different from the three covered in this chapter. While the algebraic form is not altered by the presence of EIV's some work may be necessary to evolve procedures for choice of algebraic form in the presence of EIV's. Also, work like that reported in this chapter must be done for other important algebraic forms (e.g., double-log relation). When estimators have been found for all the major engel curve forms, the situation will be ripe for an application to empirical budget data. Moreover, for a given algebraic form one can take different types of moments to get more than one consistent estimates of β and thus again it creates the problem of choice among the different estimates.

Household budget data are often available in a grouped form and further work would be necessary to adapt the moment estimators reported in this chapter for purposes of application to such grouped data.

Throughout the chapters 4 and 5 except in the last section of chapter 5, the distribution of X has been assumed to be lognormal. Some other distributions, e.g., Pareto, Gamma, Log-logistic etc., may be appropriate depending on the specific situations. The moment estimators will then have to be changed accordingly and similar approach can be taken in those situations. The assumption on the distribution of u given X, that it follows a Pearsonian type II distribution, is also subject to criticism. It is necessary to see how robust the moment estimates are due to changes in the distributional assumptions on X, u given X or the distrubance term ε.

Chapter 6

SOME FURTHER RESULTS FOR THE STANDARD TWO-VARIABLE EVM

6.1 Introduction

In this concluding chapter we present some results of a miscellaneous nature on statistical inference in the standard two-variable EVM (*vide* Pal, 1980a). We first give some comments on the estimator proposed by Kaila (1980). Next, the standard EVM where both the error variances are known *a priori* has been treated with a view to getting an improved estimator by pooling the two earlier known estimators both of which are consistent. Lastly, we consider the estimator due to Boudon (1965, 1967, 1968) in the context of EVM and discuss how Boudon's estimator can be improved further. The three problems discussed in this chapter may seem to be disconnected. However, in all the three cases we are trying to estimate the same coefficient of the EVM. In the first case, that is for Kaila's estimator, no other additional information is available. In the second case, two items of additional information are available, viz., the values of σ_u^2, the variance of the error associated with the regressor and σ_v^2, the variance of the error associated with the regressand. In the third case, more than one additional information in the form of IV's are available (*vide* Pal, 1980a).

6.2 Remarks on Kaila's Estimator

In the standard EVM (*vide* Chapter 1, Section 1.4) the limiting values of the OLS slope estimator ($\hat{b}_1 = m_{11}/m_{20}$) and of the reverse LS slope estimator ($\hat{b}_2 = m_{02}/m_{11}$) are related as

$$b_1 < \beta < b_2 \qquad \text{(assuming } \beta > 0 \text{)},$$

where b_1 and b_2 are the limiting values of $\hat{b}_1$ and $\hat{b}_2$ respectively (see Chapter 1, Section 1.7). Hence an appropriate averaging of $\hat{b}_1$ and $\hat{b}_2$ might lead to a satisfactory estimate of β. Gini proposed the arithmetic mean in this context.

Recently Kaila (1980) proposed a new method of estimating β where "..... the best slope is defined as that which lies between two extreme slopes $\hat{b}_1$ and $\hat{b}_2$ and for which the percentage deviations from their respective minima of the functions L(y) and L(x) are equal, where $L(y) = \sum_i (\alpha + \beta x_i - y_i)^2$, $L(x) = \frac{1}{\beta^2} \sum_i (\alpha + \beta x_i - y_i)^2$".

Kaila showed that the "best" slope is thus given by

$$\hat{b}_k = \sqrt{L^*(y)/L^*(x)}, \qquad \ldots\ (6.1)$$

where $L^*(y)$ and $L^*(x)$ are the minimum values of the functions L(y) and L(x) corresponding to the slopes $\hat{b}_1$ and $\hat{b}_2$ respectively.

It is interesting to see that $\hat{b}_K$ is nothing but the geometric mean of the two extreme slope estimators.

$$\hat{b}_K = \sqrt{L^*(y)/L^*(x)}$$

$$= \left[\frac{\hat{b}_2^{\,2} \sum_i \{ \hat{b}_1 (x_i - \bar{x}) - (y_i - \bar{y}) \}^2}{\sum_i \{ \hat{b}_2 (x_i - \bar{x}) - (y_i - \bar{y}) \}^2} \right]^{\frac{1}{2}}$$

$$= \left[\frac{\hat{b}_2 (\hat{b}_1^2 \, m_{20} - 2\hat{b}_1 \, m_{11} + m_{02}}{\hat{b}_2^{\,2} m_{20} - 2 \, \hat{b}_2 m_{11} + m_{02}} \right]^{\frac{1}{2}}.$$

Now, substituting $\frac{m_{11}}{m_{20}}$ for $\hat{b}_1$ and $\frac{m_{02}}{m_{11}}$ for $\hat{b}_2$, we get,

$$\hat{b}_K = \sqrt{\frac{m_{02}}{m_{20}}} = \sqrt{\frac{m_{11}}{m_{20}} \; \frac{m_{02}}{m_{11}}} \; .$$

Hence, $$\hat{b}_K = \sqrt{\hat{b}_1 \; \hat{b}_2} \; . \qquad \ldots (6.2)$$

Now we may compare the new estimator with the conventional estimator $\hat{b}_A$, the arithmetic mean of $\hat{b}_1$ and $\hat{b}_2$. The bias of $\hat{b}_K$ is asymptotically

$$B(\hat{b}_K) \simeq \frac{\beta}{2} \left[(r_{v'} - r_u) + r_u^2 - r_u r_{v'} - (r_{v'} - r_u)^2 / 4 \right] \quad \ldots (6.3)$$

retaining terms up to second degree in r_u and $r_{v'}$. The asymptotic bias of $\hat{b}_A$ is, up to the same order,

$$B(\hat{b}_A) \simeq \frac{\beta}{2} \left[(r_{v'} - r_u) + r_u^2 \right] \qquad \ldots (6.4)$$

Hence, $$B^2(\hat{b}_A) - B^2(\hat{b}_K) \simeq \frac{\beta^2}{8} (r_{v'} - r_u) \left[(r_{v'} - r_u)^2 + 4 r_{v'} r_u \right] \quad \ldots (6.5)$$

retaining terms up to order three. The above difference is positive, zero or negative according as $r_{v'}$ is greater than, equal to or less than r_u. Similarly, comparing the two asymptotic variances we get (vide Appendix 6.1)

$$V(\hat{b}_A) - V(\hat{b}_K) \simeq \frac{\beta^2 (r_u + r_{v'})^2}{4n} \left[(B_2 - 1)(r_{v'} - r_u) + 3r_u - r_{v'} \right] \quad \ldots (6.6)$$

where B_2 $(= \mu_4(X) / \mu_2^2(X))$ is the coefficient of Kurtosis. Thus, $V(\hat{b}_A) - V(\hat{b}_K) \gtrless 0$ (in large samples) according as

$$r_{v'} / r_u \gtrless 1 - \frac{2}{B_2 - 2} \; . \qquad \ldots (6.7)$$

It may be mentioned here that the X-distribution in Engel curve analysis (e.g. Household income or total expenditure) is highly positively skewed as well as very peaked so that the Kurtosis coefficient is more than 10 in most of the cases.

Comparison of "mean square errors" (MSE's) reduces to the comparison of biases in large samples. It may be noted that, in general, both $\hat{b}_K$ and $\hat{b}_A$ are inconsistent.

If an a priori estimate for $\sigma_u^2 / \sigma_v^2 (= \lambda$, say) is given and if u, v and ε are normally distributed, then the MLE for β is known to be (Lindley, 1947; Madansky, 1959)

$$\hat{\beta}_\lambda = \frac{(\lambda \hat{b}_1 \hat{b}_2 - 1) + \sqrt{(\lambda \hat{b}_1 \hat{b}_2 - 1)^2 + 4\lambda \hat{b}_1^2}}{2\lambda \hat{b}_1} . \quad .... (6.8)$$

Obviously $\hat{b}_A$ corresponds to

$$\lambda_A = \frac{2}{\hat{b}_1 (\hat{b}_1 + \hat{b}_2)} \quad .... (6.9)$$

whereas $\hat{b}_K$ corresponds to

$$\lambda_K = \frac{1}{\hat{b}_1 \hat{b}_2} \quad .... (6.10)$$

so that

$$\lambda_K - \lambda_A = \frac{\hat{b}_1 - \hat{b}_2}{\hat{b}_1 \hat{b}_2 (\hat{b}_1 + \hat{b}_2)} . \quad .... (6.11)$$

Hence $\lambda_K - \lambda_A < 0$ in thelimit. Otherwise also it is obvious that Kaila's estimator corresponds to a lower λ than the conventional estimator $\hat{b}_A$. On the whole, any slope estimator intermediate between $\hat{b}_1$ and $\hat{b}_2$ is optimal for some λ and not optimal for others. Discussion of themerits of any such estimator (like Kaila's) cannot be made in the absence of any a priori knowledge of the magnitude of λ.

6.3 Estimation of β Where Both Error Variances are Known

In the last section we discussed some problems of estimating β where neither σ_u^2 nor $\sigma_{v'}^2$ $(= \sigma_v^2 + \sigma_\varepsilon^2)$ is known *a priori*. If, on the other hand, both σ_u^2 and $\sigma_{v'}^2$ are known *a priori*, then we have two estimators $\hat{\beta}_1$ and $\hat{\beta}_2$ defined in chapter 1 section 1.8. The problem can be looked upon as a problem of overidentification, since the knowledge of only one of σ_u^2 and $\sigma_{v'}^2$ is necessary in this case. If we define a pooled estimator as

$$\hat{\beta}_a = a\,\hat{\beta}_1 + (1 - a)\,\hat{\beta}_2$$

and find out 'a' so as to minimize the variance of $\hat{\beta}_a$, then the optimum estimator thus obtained is obviously better than $\hat{\beta}_1$ and $\hat{\beta}_2$, since they are members of the above considered class.

Now the asymptotic dispersion matrix of $(\hat{\beta}_1, \hat{\beta}_2)$ is

$$V(\hat{\beta}_1) \simeq \frac{\beta^2}{n}(r_u + r_{v'} + r_u r_{v'} + 2r_u^2) \qquad \ldots (6.12)$$

$$V(\hat{\beta}_2) \simeq \frac{\beta^2}{n}(r_u + r_{v'} + r_u r_{v'} + 2r_{v'}^2) \qquad \ldots (6.13)$$

$$\text{Cov}(\hat{\beta}_1, \hat{\beta}_2) \simeq \frac{\beta^2}{n}(r_u + r_{v'} - r_u r_{v'}) \qquad \ldots (6.14)$$

where $r_u = \sigma_u^2 / \sigma_X^2$ and $r_{v'} = (\sigma_v^2 + \sigma_\varepsilon^2)/(\beta^2 \sigma_X^2)$. Hence the optimum 'a' is

$$a_0 = \hat{r}_{v'} / (\hat{r}_u + \hat{r}_{v'}), \qquad \ldots (6.15)$$

where $\hat{r}_u = \sigma_u^2 / (m_{20} - \sigma_u^2)$ and $\hat{r}_{v'} = (\sigma_v^2 + \sigma_\varepsilon^2)/(m_{02} - \sigma_v^2 - \sigma_\varepsilon^2)$.

The optimum $\hat{\beta}_a$ is then

$$\hat{\beta}_0 = \frac{\hat{r}_{v'}\,\hat{\beta}_1 + \hat{r}_u\,\hat{\beta}_2}{\hat{r}_{v'} + \hat{r}_u}. \qquad \ldots (6.16)$$

The asymptotic variance of $\hat{\beta}_0$ is

$$V(\hat{\beta}_0) \simeq \frac{\beta^2}{n}(r_u + r_{v'} + r_u r_{v'}) . \qquad \ldots (6.17)$$

Madansky (1959) suggests the use of all available information for reducing the sampling variance, but did not actually propose any pooled estimator for this problem. He writes (p. 179)

> "The case in which both σ_u^2 and σ_v^2 are known is an over-identified situation (since only knowledge of their ratio is necessary for identifiability). In this case, it would seem reasonable to use all the available information in the hope of achieving a small variance of the estimate of β ".

He then generalized the model assuming Cov $(u, v') \neq 0$ $(v' = v + \varepsilon)$ and found the MLE of β to be

$$\hat{\beta}_M = \sqrt{\hat{\beta}_1 \hat{\beta}_2} . \qquad \ldots (6.18)$$

If Cov $(u, v') = 0$, we get

$$V(\hat{\beta}_M) \simeq \frac{\beta^2}{n}\left(r_u + r_{v'} + \frac{r_u^2 + r_{v'}^2}{2}\right) . \qquad \ldots (6.19)$$

$\hat{\beta}_0$ is the optimum among the class of linear combinations of $\hat{\beta}_1$ and $\hat{\beta}_2$, while $\hat{\beta}_M$ does not fall in this class. So it is necessary to compare the variances and MSE's of $\hat{\beta}_0$ and $\hat{\beta}_M$ to judge which estimator is superior. Obviously, when Cov $(u, v') = 0$, asymptotic variances satisfy the following inequality :

$$V(\hat{\beta}_0) \leq V(\hat{\beta}_M),$$

since $(r_u^2 + r_{v'}^2)/2 \geq r_u r_{v'}$. When Cov $(u, v') \neq 0$ we have,

$$V(\hat{\beta}_M) \simeq \frac{\beta^2}{n}\left[r_u + r_{v'} + \frac{r_u^2 + r_{v'}^2}{2} + \rho\sqrt{r_u r_{v'}}\,(r_u + r_{v'} - 2) - \rho^2 r_u r_{v'} \right], \quad \ldots (6.20)$$

where ρ is the correlation coefficient between u and v'. $V(\hat{\beta}_M)$ is the smallest when ρ is + 1 as it is expected.

For $\rho \neq 0$, we get asymptotically

$$B(\hat{\beta}_0) \simeq \beta\rho^2 r_u r_{v'} \Big/ (1 + \rho\sqrt{r_u r_{v'}}) \quad \ldots (6.21)$$

$$V(\hat{\beta}_0) \simeq \frac{\beta^2}{n(r_u + r_{v'})^2}\Big[(B_2 - 1)\,\rho^2 r_u r_{v'}(r_u - r_{v'})^2 + (r_u + r_{v'})^3 + r_u r_{v'}(r_u + r_{v'})^2 - 2\rho\sqrt{r_u r_{v'}}(r_u + r_{v'})^2 - 4 r_u\,\rho\sqrt{r_u r_{v'}}(r_u - r_{v'})(r_u + r_{v'}) + \rho^2 r_u r_{v'}\,(5r_u^2 - 3r_{v'}^2 - 6r_u r_{v'})\Big]. \quad \ldots (6.22)$$

Terms of first degree in r_u and $r_{v'}$ disappear in the expression $V(\hat{\beta}_0) - V(\hat{\beta}_M)$ i.e., terms of order two or above in r_u and $r_{v'}$ remain. But the bias is of the order two, bias squared contain terms of degree four or more in r_u and $r_{v'}$. Hence the effect of bias of $\hat{\beta}_0$ is negligible so far as comparison between MSE's of $\hat{\beta}_0$ and $\hat{\beta}_M$ are concerned for moderate values of n. Comparison between $\hat{\beta}_0$ and $\hat{\beta}_M$ does not lead to one clear-cut answer. When $\rho = 0$, $\hat{\beta}_0$ is obviously better. When $\rho \neq 0$, $\hat{\beta}_0$ is inconsistent, but asymptotic bias is very small. Here the difference between the asymptotic variance of $\hat{\beta}_0$ and $\hat{\beta}_M$ is also small. In short, for moderate values of n, $\hat{\beta}_0$ may be preferable to $\hat{\beta}_M$. For large values of n, $\hat{\beta}_M$ is recommended, though the actual difference between the MSE's may not be large.

The motivation behind introducing this example is to demonstrate that there may be a way out of an overidentification problem viz., the introduction of a more general model. However, much depends upon how usefully the model is generalized.

6.4 Remarks on Boudon's Estimator and Further Improvements

In this section two IV methods of consistent estimation of the slope parameter in a two variable EVM are discussed. One of them is based on ideas from Boudon (1965, 1967, 1968) and Goldberger (1970) and is a modification of Boudon's method which was primarily proposed in a different context. Both the method leads to optimum estimators in the sense that the estimators are derived by minimizing asymptotic variances within some classes of consistent estimators. It is also shown that the two approaches lead to the same consistent estimator and the common estimator coincides with Theil's (1958, pp.347-351) two-stage-least-squares (2SLS) estimator.

6.4.1 Boudon's Set-up

Boudon (1968, pp.209-211) considered the following relation

$$Y_i = \beta X_i + \varepsilon_i , \qquad i = 1, 2, \dots, n \qquad \dots (6.23)$$

where X_i's, $i = 1, 2, \dots, n$ are explanatory variables; β is the parameter to be estimated; and Y_i and X_i are measured from their respective means. The standard assumptions of CLS regression are made throughout, excepting that the X_i's may be stochastic. There are, in addition, observations on some auxiliary stochastic variables, called Instrumental Variables (IV's) denoted $z_1, z_2, \dots, z_K$, also measured from their respective means and assumed to be independent of the error terms. Throughout this paper observations on X_i's and z_i's are assumed to be independently drawn from the same population.

For every IV z_j we have the relation

$$\Sigma Yz_j = \beta\Sigma Xz_j + \Sigma\varepsilon z_j \text{ [6.1]}, \qquad j = 1, 2, \ldots, K \qquad \ldots (6.24)$$

from which β can be estimated consistently (since $\underset{n\to\infty}{\text{plim}} \frac{1}{n}\Sigma\varepsilon z_j = 0$) using

$$\hat{\beta}_j = \frac{\Sigma Yz_j}{\Sigma Xz_j}, \qquad j = 1, 2, \ldots, K. \qquad \ldots. (6.25)$$

We will assume that the IV's are stochastic.

In general, the basic estimators $\hat{\beta}_j$ for different j will not coincide. Boudon proposed a pooled estimator which can be derived by minimizing the expression

$$Q = \sum_j (\Sigma Yz_j - \beta\Sigma Xz_j)^2 \qquad \ldots (6.26)$$

with respect to β. The solution is obviously

$$\hat{\beta}_B = \frac{\sum_j (\Sigma X z_j)(\Sigma Yz_j)}{\sum_j (\Sigma Xz_j)^2}. \qquad \ldots (6.27)$$

This is a weighted averageof the $\hat{\beta}_j$'s with weights proportional to $(\Sigma Xz_j)^2$. The estimate $\hat{\beta}_B$ is not invariant under unequal scale changes of IV's. Thus, if we multiply only one of the IV's by a constant which is different from 'one' keeping the others unchanged we get a different estimate $\hat{\beta}_B$. To remove this difficulty, let us minimize

---

6.1/ Unless specially required, we shall omit the subscript which ranges from 1 to n.

a different quantity Q^* in place of Q :

$$Q^* = \sum_j (\hat{\beta}_j - \beta)^2 d_j \qquad \ldots (6.28)$$

where d_j = weight of $\hat{\beta}_j$

$$= \frac{(\Sigma X z_j)^2}{\Sigma z_j^2} .$$

Observe that the weights are now unitfree so far as the IV's are concerned. The revised estimator turns out to be

$$\hat{\beta}_R = \frac{\sum_j d_j \hat{\beta}_j}{\sum_j d_j}$$

$$= \frac{\sum_j \{ (\Sigma X z_j)(\Sigma Y z_j)/ (\Sigma z_j^2) \}}{\sum_j \{ (\Sigma X z_j)^2 / (\Sigma z_j^2) \}} . \qquad \ldots (6.29)$$

Goldberger (1970) examines Boudon's estimator and points out that the conventional LS estimator (LSE) of β is BLUE and hence preferable to Boudon's estimator. The revised estimator also is inferior to LSE for the same reason. If, however, we introduce errors in the regressor-observations, the situation becomes completely different. Goldberger's claim, that the LSE is better than Boudon's estimator, ceases to be valid. The LSE is, in general, inconsistent while Boudon's estimator is consistent.

In the next subsection we reformulate the whole model bringing in EIV's and propose an alternative approach for estimating β. The subsection (6.4.3) deals with yet another method of estimating β consistently and shows that the two approaches lead to the same estimator. The subsection (6.4.4) shows that the common estimator can be arrived at if one follows Theil's suggestion of applying 2SLS method (Theil, 1958, pp.347-51) for tackling the EVM in a situation where more than one IV are available. The last subsection concludes the chapter by pointing out how one of the approaches may be generalized and how one can utilize the overidentified restrictions by generalizing the model.

6.4.2 The First Approach to Estimation of β

Let us assume that we observe x_i and y_i in place of X_i and Y_i, respectively, where

$$x_i = X_i + u_i, \qquad i = 1, 2, \ldots, n \qquad \ldots (6.30)$$

$$y_i = Y_i + v_i, \qquad i = 1, 2, \ldots, n . \qquad \ldots (6.31)$$

The EIV's u and v are i.i.d. random variables with mean zero and variances σ_u^2 and σ_v^2 respectively. It will be assumed that the IV's are independent of the u's and the v's. Our regression equation is

$$Y_i = \beta X_i + \varepsilon_i, \qquad i = 1, 2, \ldots, n . \qquad \ldots (6.32)$$

The EIV's are also independent of the errors-in-equation ε. A consistent estimator of β is 6.2/

$$\hat{\beta}_R = \frac{\sum_j \left(\left(\Sigma xz_j \right) \left(\Sigma yz_j \right) / \Sigma z_j^2 \right)}{\sum_j \left(\left(\Sigma \; xz_j \right)^2 / \Sigma z_j^2 \right)} . \qquad \ldots (6.33)$$

This is a weighted average of the basic estimators

$$\hat{\beta}_j = \frac{\Sigma yz_j}{\Sigma xz_j}, \qquad j = 1, 2, \ldots, K, \qquad \ldots (6.34)$$

6.2/ In Boudon's approach the regressor itself was included in the set of IV's. In the EVM this would destroy the consistency property of $\hat{\beta}_R$.

In his discussion on Boudon's estimator Goldberger mentioned that the asymptotic variance of $\hat{\beta}_B$ may not be smaller than that of every basic estimator $\hat{\beta}_j$. This is because the $\hat{\beta}_j$'s may be positively correlated among themselves and hence minimization of Q loses its significance. The same remark applies to $\hat{\beta}_R$ in our case. In these circumstances, we may seek that linear combination of $\hat{\beta}_j$'s for which the asymptotic variance is minimum. We therefore consider the class of linear combinations

$$\hat{\beta}(a) = \Sigma a_j \hat{\beta}_j = a' \hat{\beta} \quad \text{(say)} \qquad \ldots (6.35)$$

such that

$$\Sigma a_j = 1 . \qquad \ldots (6.36)$$

Every member in the class of estimators defined by (6.35) and (6.36) is consistent and so is the optimum estimator derived by minimizing the asymptotic variance of $\hat{\beta}(a)$ subject to $\Sigma a_j = 1$. The optimum a is (see Appendix 6.2)

$$\hat{a}_o = \frac{\hat{W}^{-1} e}{e' \hat{W}^{-1} e} \qquad \ldots (6.37)$$

where $\hat{a}_o$ is a column vector of estimated coefficients of $a_1, a_2, \ldots, a_K$ in (6.35); W is the KxK estimated asymptotic covariance matrix of the basic estimators, i.e.,

$$\hat{W} = ((\hat{w}_{ij})), \qquad \ldots (6.38)$$

where $\hat{w}_{ij} = \hat{\sigma}_{ij} / (\hat{\sigma}_{xi} \hat{\sigma}_{xj})$, ... (6.39)

where $\hat{\sigma}_{ij} = \frac{1}{n} \Sigma z_i z_j$, ... (6.40)

and $\hat{\sigma}_{xj} = \frac{1}{n} \Sigma x z_j$, ... (6.41) and e is a $K \times 1$ column vector consisting of one's only. This leads to the optimum estimator for β as

$$\hat{\beta}_{01} = \frac{e'\hat{W}^{-1}\hat{\beta}}{e'\hat{W}^{-1}e} . \qquad \ldots (6.42)$$

The estimate of the asymptotic variance of the above estimator is

$$V_1 = \frac{1}{(e'\hat{W}^{-1}e)} . \qquad \ldots (6.43)$$

The expression (6.42) can be simplified to some extent. Define a diagonal matrix $\Delta_{K \times K}$, the jth diagonal element of which is

$$\delta_j = \frac{1}{\sigma_{x_j}} . \qquad \ldots (6.44)$$

Let $\hat{\Sigma}$ be the dispersion matrix of $z_1, \ldots, z_K$.

Hence $$\hat{W} = \Delta \hat{\Sigma} \Delta \qquad \ldots (6.45)$$

or $$\hat{W}^{-1} = \Delta^{-1} \hat{\Sigma}^{-1} \Delta^{-1} . \qquad \ldots (6.46)$$

The estimator $\hat{\beta}_{01}$ may now be simplified as

$$\begin{aligned} \hat{\beta}_{01} &= \frac{e'\hat{W}^{-1}\hat{\beta}}{e'\hat{W}^{-1}e} \\ &= \frac{e'\Delta^{-1}\hat{\Sigma}^{-1}\Delta^{-1}\hat{\beta}}{e'\Delta^{-1}\hat{\Sigma}^{-1}\Delta^{-1}e} \\ &= \frac{\hat{\sigma}_x'\hat{\Sigma}^{-1}\hat{\sigma}_y}{\hat{\sigma}_x'\hat{\Sigma}^{-1}\hat{\sigma}_x} , \end{aligned} \qquad \ldots (6.47)$$

where $\hat{\sigma}_x$ is the column vector of estimated covariances of $z_1, z_2, \ldots, z_K$, each with x and $\hat{\sigma}_y$ is the same for y.

If $\hat{\beta}_j$'s are uncorrelated among themselves, right hand side (RHS) of (6.47) reduces to the RHS of (6.29), i.e., $\hat{\beta}_{01}$ becomes $\hat{\beta}_R$ ensuing validity of Goldberger's comment and the approach.

6.4.3 The Second Approach

There is yet another class of estimators which will be discussed in this subsection. Let us first subject the IV's $z_1, z_2, \ldots, z_K$ to a linear transformation and obtain a new set of IV's $z_1', z_2', \ldots, z_K'$ as

$$z_j' = \sum_{l=1}^{K} q_{lj} z_l , \qquad j = 1, 2, \ldots, K \qquad \ldots (6.48)$$

such that

$$\text{(i)} \quad \sum_i z_{ji}' z_{li}' = 0 \qquad \text{for } j \neq l \qquad \ldots (6.49)$$

and

$$\text{(ii)} \quad \sum_i z_{ji}' = 1 \quad \text{for all } j . \qquad \ldots (6.50)$$

Our new class of consistent estimators can now be defined as

$$\hat{\beta}(c) = \frac{\Sigma y(\sum_j c_j z_j')}{\Sigma x(\sum_j c_j z_j')} , \qquad \ldots (6.51)$$

where c_j, $j = 1, 2, \ldots, K$, are any real constants such that $\hat{\beta}(c)$ is defined. We may minimize the asymptotic variance of $\hat{\beta}(c)$ with respect to c_j's (see Appendix 6.3) and get the optimum c_j to be, apart from a constant of proportionality, which does not influence $\hat{\beta}(c)$:

$$c_j^* = \frac{1}{n} \Sigma X z_j' \qquad \ldots (6.52)$$

$$= \frac{1}{n} \Sigma x z_j' - \frac{1}{n} \Sigma u z_j' . \qquad \ldots (6.53)$$

Since $\underset{n \to \infty}{\text{plim}} \frac{1}{n} \Sigma u z_j' = 0$, $\underset{n \to \infty}{\text{plim}} c_j^* = \underset{n \to \infty}{\text{plim}} \frac{1}{n} \Sigma x z_j'$ so that c_j^* can be replaced by $\frac{1}{n}\Sigma x z_j'$ without affecting asymptotic properties. Thus, the optimum estimator in the class $\hat{\beta}$ (c) is

$$\hat{\beta}_{02} = \frac{\Sigma(\Sigma x z_j')(\Sigma y z_j')}{\Sigma(\Sigma x z_j')^2} \qquad \ldots (6.54)$$

This expression may be compared with (6.27) and (6.29), noting that

$$\Sigma z_j'^2 = 1 \quad \forall j.$$

To simplify (6.54), let Z, Z^* denote two matrices with row vectors of observations on z_j's and transformed z_j''s for j = 1, 2, ..., K.

Write

$$Z^* = Z.\ Q, \qquad Q \text{ nonsingular} \qquad \ldots (6.55)$$

Since $Z^{*\prime} Z^* = I_{KxK}$, $Z'Z$ is proportional to $\hat{\Sigma}$, and since $Z'Z = (QQ')^{-1}$, we have $\hat{\Sigma} \propto (QQ')^{-1}$. Hence,

$$\hat{\beta}_{02} = \frac{(x'Z^*)\ (Z^{*\prime}y)}{(x'Z^*)\ (Z^{*\prime}x)}$$

$$= \frac{x'Z\ QQ'\ Z'y}{x'Z\ QQ'\ Z'x}$$

$$= \frac{\hat{\sigma}_x'\ \hat{\Sigma}^{-1}\ \hat{\sigma}_y}{\hat{\sigma}_x'\ \hat{\Sigma}^{-1}\ \hat{\sigma}_x} \qquad \ldots (6.56)$$

The expression (6.56) is same as the expression (6.47).

$$\therefore \quad \hat{\beta}_{01} = \hat{\beta}_{02}$$

The two approaches discussed above thus lead to the same consistent estimator which has optimal properties demonstrated above and which is clearly asymptotically more efficient than any of $\hat{\beta}_1, \ldots, \hat{\beta}_K$.

6.4.4 The 2SLS Approach

Theil (1958, pp. 347-351) suggested the use of 2SLS method of estimation for tackling the EIV model. The method may be explained briefly in terms of the following model where the symbols have their usual meaning :

$$y = \beta x + \varepsilon', \quad \ldots (6.57)$$

$$x = X + u, \quad \ldots (6.58)$$

and $$\varepsilon' = -\beta u + v + \varepsilon \quad \ldots (6.59)$$

x and ε' are correlated with each other. Thatis why the OLS estimator of β is not consistent. Theil suggested that we estimate $\hat{x}$ from a multiple regression on $z_1, z_2, \ldots, z_K$ and then regress y on $\hat{x}$ to get a LS estimate for β. Such 2SLS method of estimation is usually applied to simultaneous equation systems where some of the regressors are correlated with the disturbance term in the same equation. Even though our model is not atypical simultaneous equation model, Theil's idea of 2SLS estimation seems to be applicable and possesses many desirable properties.

To estimate x from $z_1, z_2, \ldots, z_K$, consider the linear equation

$$x = Zp + u' \quad \ldots (6.60)$$

with usual assumptions and where Z consists of column vectors of observations for each z_j. The estimated $\hat{p}$ is

$$\hat{p} = (Z'Z)^{-1} Z'x \quad \ldots (6.61)$$

and hence the estimated $\hat{x}$ is

$$\hat{x} = Z(Z'Z)^{-1} Z'x \quad \ldots (6.62)$$

The 2SLS slope estimator is then

$$\hat{\beta}_{2SLS} = (\hat{x}' \hat{x})^{-1} \hat{x}' y$$

$$= \frac{x'Z(Z'Z)^{-1} Z'y}{x'Z(Z'Z)^{-1} Z'x}$$

$$= \frac{\hat{\sigma}_x' \hat{\Sigma}^{-1} \hat{\sigma}_y}{\hat{\sigma}_x' \hat{\Sigma}^{-1} \hat{\sigma}_x} \quad \ldots (6.63)$$

And thus we see that the optimum estimator obtained by two approaches in subsections (6.4.2)and(6.4.3) coincides with Theil's 2SLS estimator.

6.4.5 Discussions

The first approach may be generalized to the case where number of regressers is more than one (say m). Each basic estimator is then a vector obtained by regressing yz_j on $x_1z_j, \ldots, x_mz_j$. It is not, however, clear how one can generalize the second approach in that direction.

In social science literature the problem discussed in this chapter is named as 'multiple indicator" problem. This arises only in overidentified models. Though many attempts have been made (vide for example Joreskog, 1970; Wiley and Wiley, 1970; Fisher, 1966; Wert, Joreskog and Linn, 1973; Blalock, 1970 etc.) to find a satisfactory solution to overidentification problem, this subject is yet to reach a satisfactory level. For our particular example 2SLS seems to be the only answer justified from different angles two of which have been cited in this chapter.

We can of course utilize the overidentified restrictions in generalizing the model. At some stage in the process of generalization we shall reach a just-identified model which allows a ML solution. This, however, may lead one to a difficult situation. The success in this approach depends on how efficiently the generalization is made.

Appendix 6.1

$$\mu_2(x) = \mu_2(X) + \sigma_u^2 = \mu_2(X)\ (1 + r_u)$$

$$\mu_2(y) = \beta^2 \mu_2(X)\ (1 + r_v)$$

$$\mu_{11}(x,y) = \beta\mu_2(X)(1 + \rho\ \sqrt{r_u\ r_{v'}})$$

$$\mu_{22}(x,y) = K_{22}(x,y) + K_{20}(x,y)\ K_{02}(x,y) + 2K_{11}^2(x,y)$$

$$= K_{22}(x,y) + \mu_2(X)(1+r_u)\ \beta^2\mu_2(X)(1+r_{v'}) + 2\beta^2\mu_2^2(X)(1+\rho\sqrt{r_u r_{v'}})^2$$

$$= \beta^2\mu_2^2(X)(B_2 + r_u) + \beta^2\mu_2^2(X) r_u r_{v'} + \beta^2\mu_2^2(X) r_{v'} + 2\rho^2\ \beta^2\mu_2^2(X)\ r_u r_{v'}$$

$$+ 4\beta^2\rho\mu_2^2(X)\ \sqrt{r_u r_{v'}}$$

$$= \beta^2\mu_2^2(X)\ \{ B_2 + r_u + r_u r_{v'} + r_{v'} + 2\rho^2\ r_u r_{v'} + 4\rho\ \sqrt{r_u r_{v'}} \}.$$

$$\mu_{31}(x,y) = K_{31}(x,y) + 3K_{20}(x,y)\ K_{11}(x,y)$$

$$= \beta K_4(X) + 3\mu_2^2(X)\ \beta(1 + r_u)\ (1 + \rho\ \sqrt{r_u r_{v'}})$$

$$= \beta\ \mu_2^2(X)\ \{ B_2 + 3r_u + 3\ \rho\sqrt{r_u r_{v'}} + 3r_u\ \rho\sqrt{r_u r_{v'}} \}$$

$$\mu_{13}(x,y) = \beta^3\mu_2^2(X)\ \{ B_2 + 3r_{v'} + 3\rho\ \sqrt{r_u r_{v'}} + 3r_{v'}\rho\sqrt{r_u r_{v'}} \}$$

$$\mu_4(x) = \mu_4(X) + 6\mu_2(X)\ \sigma_u^2 + \mu_4(u)$$

$$= \mu_2^2(X)\ \{ B_2 + 6r_u + 3r_u^2 \}$$

$$\mu_4(y) = \beta^4\ \mu_2^2(X)\ \{ B_2 + 6r_{v'} + 3r_{v'}^2 \}$$

$$V(m_{20}) = \frac{1}{n}\{ \mu_{40} - \mu_{20}^2 \}$$

$$= \frac{\mu_2^2(X)}{n}\left[(B_2 - 1) + 4r_u + 2r_u^2\right]$$

$$V(m_{02}) = \frac{\beta^4 \mu_2^2(X)}{n}\left[(B_2 - 1) + 4r_{v'} + 2r_{v'}^2\right]$$

$$V(m_{11}) = \frac{1}{n}(\mu_{22} - \mu_{11}^2)$$

$$= \frac{\beta^2 \mu_2^2(X)}{n}\left[(B_2 - 1) + r_u + r_u r_{v'} + r_{v'} + \rho^2 r_u r_{v'} + 2\rho\sqrt{r_u r_{v'}}\right]$$

$$Cov(m_{20}, m_{11}) = \frac{1}{n}(\mu_{31} - \mu_{20}\mu_{11})$$

$$= \frac{\beta\mu_2^2(X)}{n}\left[(B_2 - 1) + 2r_u + 2\rho\sqrt{r_u r_{v'}}(1 + r_u)\right]$$

$$Cov(m_{02}, m_{11}) = \frac{\beta^3 \mu_2^2(X)}{n}\left[(B_2 - 1) + 2r_{v'} + 2\rho\sqrt{r_u r_{v'}}(1 + r_{v'})\right]$$

$$Cov(m_{20}, m_{02}) = \frac{1}{n}(\mu_{22} - \mu_{02}\mu_{20})$$

$$= \frac{\beta^2 \mu_2^2}{n}\left[(B_2 - 1) + 2\rho^2 r_u r_{v'} + 4\rho\sqrt{r_u r_{v'}}\right]$$

$$V\left(\frac{m_{11}}{m_{20}}\right) = \frac{\beta^2}{n(1+r_u)^4}\left[\frac{V(m_{11})}{E^2(m_{11})} + \frac{V(m_{20})}{E^2(m_{20})} - \frac{2\,Cov(m_{11}, m_{20})}{E(m_{11})\,E(m_{20})}\right]$$

$$= \frac{\beta^2}{n(1+r_u)^4}\left[(B_2 - 1)r_u^2 + r_u + r_{v'} + 3r_u r_{v'} + r_u^3 + 3r_u^2 r_{v'} + r_u^3 r_{v'}\right]$$

(if $\rho = 0$)

$$V\left(\frac{m_{02}}{m_{11}}\right) = \frac{\beta^2}{n}\left[(B_2 - 1)r_{v'}^2 + r_u + r_{v'} + 3r_u r_{v'} + r_{v'}^3 + 3r_u r_{v'}^2 + r_u r_{v'}^3\right]$$

(from symmetry if $\rho = 0$)

$$\mathrm{Cov}\left(\frac{m_{02}}{m_{11}}, \frac{m_{11}}{m_{20}}\right) = \frac{\mathrm{Cov}(m_{02}, m_{11})}{\mu_{20}\,\mu_{11}} - \frac{\mathrm{Cov}(m_{02}, m_{20})}{\mu_{20}^2}$$

$$- \frac{\mu_{02}}{\mu_{20}\,\mu_{11}^2} V(m_{11}) + \frac{\mu_{02}}{\mu_{11}\,\mu_{20}^2} \mathrm{Cov}(m_{11}, m_{20})$$

$$= \frac{\beta^2}{n(1+r_u)^2} \Big[-(B_2 - 1) r_u r_{v'} + r_u + r_{v'} + r_u r_{v'} - r_u^2 - r_{v'}^2$$

$$-2r_u^2\, r_{v'} - 2r_u r_{v'}^2 - r_u^2\, r_{v'}^2 \Big]$$

$$= \frac{\beta^2}{n(1+r_u)^4} \Big[-(B_2 - 1)\ (r_u r_{v'} + 2r_u^2\, r_{v'}) + r_u + r_{v'}$$

$$+ r_u^2 + 3r_u r_{v'} - r_{v'}^2 - r_u^3 + r_u^2 r_{v'} - 4r_u r_{v'}^2 \Big]$$

(if $\rho = 0$)

$$V\left(\frac{m_{02}}{m_{11}}\right) = \frac{\beta^2}{n(1+ r_u)^4} \Big[(B_2 - 1)\ (r_{v'}^2 + 4r_u r_{v'}) + r_u + r_{v'} + 4r_u^2 + 7r_u r_{v'}$$

$$+ 18r_u^2 r_{v'} + 6r_u^3 + 3r_u\, r_{v'}^2 + r_{v'}^3 \Big]$$

(if $\rho = 0$)

$$V\left(\frac{\hat{b}_1 + \hat{b}_2}{2}\right) = \frac{1}{4} \{ V(\hat{b}_1) + V(\hat{b}_2) + 2\,\mathrm{Cov}\,(\hat{b}_1, \hat{b}_2) \}$$

$$= \frac{1}{4} \{ V\left(\frac{m_{11}}{m_{20}}\right) + V\left(\frac{m_{02}}{m_{11}}\right)\ 2\,\mathrm{Cov}\left(\frac{m_{11}}{m_{20}}, \frac{m_{02}}{m_{11}}\right) \}$$

$$\simeq \frac{\beta^2}{4n(1 + r_u)^4} \Big[4(r_u + r_{v'}) + (B_2 - 1)\ (r_u^2 + r_{v'}^2 - 2r_u r_{v'}$$

$$- 4r_u^2\, r_{v'} + 4r_u r_{v'}^2) + 6r_u^2 + 16r_u r_{v'} - 2r_{v'}^2$$

$$+ 5r_u^3 + 23r_u^2\, r_{v'} - 5r_u r_{v'}^2 + r_{v'}^3 \Big]$$

$$V(\sqrt{\hat{b}_1 \hat{b}_2}) = \frac{E(\hat{b}_2)}{4E(\hat{b}_1)} V(\hat{b}_1) + \frac{E(\hat{b}_1)}{4E(\hat{b}_2)} V(\hat{b}_2) + \frac{2\,Cov(\hat{b}_1, \hat{b}_2)}{4}$$

$$= \frac{1}{4}\left[(1+r_u)(1+r_{v'})\, V(\hat{b}_1) + \frac{V(\hat{b}_2)}{(1+r_u)(1+r_{v'})} + 2\,Cov(\hat{b}_1, \hat{b}_2)\right]$$

$$\simeq \frac{\beta^2}{4n(1+r_u)^4}\Big[4(r_u+r_{v'}) + (B_2-1)(r_u^2 + r_{v'}^2 - 3r_u^2 r_{v'} + 3r_u r_{v'}^2 + r_u^3 - r_{v'}^3 - 2r_u r_{v'}) + 6r_u^2 + 16r_u r_{v'} - 2r_{v'}^2 + 2r_u^3 + 18r_u^2 r_{v'} - 6r_u r_{v'}^2 + 2r_{v'}^3\Big]$$

$$V(\hat{b}_A) - V(\hat{b}_K) \simeq \frac{\beta^2 (r_u+r_{v'})^2}{4n}\left[(B_2-1)(r_{v'} - r_u) + 3r_u - r_{v'}\right]$$

Now,

$$\hat{\beta}_1 = \frac{m_{11}}{m_{20} - \sigma_u^2}, \quad \hat{\beta}_2 = \frac{m_{02} - \sigma_{v'}^2}{m_{11}}, \quad \hat{\beta}_M = \pm\sqrt{\hat{\beta}_1 \hat{\beta}_2} = \pm\sqrt{\frac{m_{02} - \sigma_{v'}^2}{m_{20} - \sigma_u^2}}$$

$$V(\hat{\beta}_1) = \frac{\beta^2 (1+\rho\sqrt{r_u r_{v'}})^2}{n}\Big[\{(B_2-1) + r_u + r_u r_{v'} + r_{v'} + \rho^2 r_u r_{v'} + 2\rho\sqrt{r_u r_{v'}}\} / (1+\rho\sqrt{r_u r_{v'}})^2 + (B_2 - 1 + 4r_u + 2r_u^2) - 2\frac{(B_2-1) + 2r_u + 2\rho\sqrt{r_u r_{v'}}(1+r_u)}{1+\rho\sqrt{r_u r_{v'}}}\Big]$$

$$= \frac{\beta^2}{n}\Big[\rho^2 r_u r_{v'}(B_2-1) + r_u + r_{v'} + 2r_u^2 + r_u r_{v'} + \rho^2 r_u r_{v'}(2r_u^2 - 3) + \rho\sqrt{r_u r_{v'}}(4r_u^2 - 2)\Big]$$

$$V(\hat{\beta}_2) = \frac{\beta^2}{n(1+\rho\sqrt{r_u r_{v'}})} \Big[\frac{B_2 - 1 + 4r_{v'} + 2r_{v'}^2}{1}$$

$$+ \frac{B_2 - 1 + r_u + r_{v'} + r_u r_{v'} + \rho^2 r_u r_{v'} + 2\rho\sqrt{r_u r_{v'}}}{(1+\rho\sqrt{r_u r_{v'}})^2}$$

$$- 2\,\frac{B_2 - 1 + 2r_{v'} + 2\rho\sqrt{r_u r_{v'}}\,(1 + r_{v'})}{1+\rho\sqrt{r_u r_{v'}}} \Big]$$

$$= \frac{\beta^2}{n(1+\rho\sqrt{r_u r_{v'}})^4} \Big[(B_2 - 1)\,\rho^2 r_u r_{v'} + r_u + r_{v'} + r_u r_{v'} + 2r_{v'}^2$$

$$+ \rho^2 r_u r_{v'}(2r_{v'}^2 - 3) + \rho\sqrt{r_u r_{v'}}\,(4r_{v'}^2 - 2) \Big]$$

$$\simeq \frac{\beta^2}{n} \Big[(B_2 - 1)\rho^2 r_u r_{v'} + r_u + r_{v'} + r_u r_{v'} + 2r_{v'}^2$$

$$+ \rho\sqrt{r_u r_{v'}}\,(-4r_u - 4r_{v'} - 2) + 5\rho^2 r_u r_{v'} \Big]$$

$$Cov(\hat{\beta}_1, \hat{\beta}_2) = \frac{Cov(m_{02} - \sigma_v^2, m_{11})}{E(m_{20} - \sigma_u^2)\, E(r_1)} - \frac{Cov\,(m_{02} - \sigma_v^2, m_{20} - \sigma_u^2)}{E^2(m_{20} - \sigma_u^2)}$$

$$- \frac{V(m_{11})E(m_{02} - \sigma_v^2)}{E(m_{20} - \sigma_u^2)E^2(m_{11})} + \frac{E(m_{02} - \sigma_v^2)\, Cov(m_{11}, m_{20} - \sigma_u^2)}{E(m_{11})\, E^2(m_{20} - \sigma_u^2)}$$

$$\simeq \frac{\beta^2}{n(1+\rho\sqrt{r_u r_{v'}})^2} \Big[-\rho r_u r_{v'}(B_2 - 1) + r_u + r_{v'} - r_u r_{v'}$$

$$+ \rho\sqrt{r_u r_{v'}}(4r_u + 4r_{v'} - 2) - 7\rho^2 r_u r_{v'} \Big]$$

$$\simeq \frac{\beta^2}{n} \Big[-(B_2 - 1)\,\rho^2 r_u r_{v'} + r_u + r_{v'} - r_u r_{v'}$$

$$+ \rho\sqrt{r_u r_{v'}}\,(2r_u + 2r_{v'} - 2) + \rho^2 r_u r_{v'}(2r_u + 2r_{v'} - 3) \Big]$$

When $\rho = 0$, $\dfrac{V(\hat{\beta}_1) - Cov(\hat{\beta}_1, \hat{\beta}_2)}{V(\hat{\beta}_1) + V(\hat{\beta}_2) - 2\,Cov(\hat{\beta}_1, \hat{\beta}_2)} = \dfrac{r_u}{r_u + r_{v'}}$

$$\therefore \quad \hat{\hat{\beta}}_0 = \frac{r_{v'}\,\hat{\beta}_1}{r_u + r_{v'}} + \frac{r_u\,\hat{\beta}_2}{r_u + r_{v'}}$$

$$V(\hat{\beta}_0) = \left(\frac{r_{v'}}{r_u + r_{v'}}\right)^2 V(\hat{\beta}_1) + \left(\frac{r_u}{r_u + r_{v'}}\right)^2 V(\hat{\beta}_2) + 2\frac{r_u\,r_{v'}}{(r_u + r_{v'})^2}\,Cov(\hat{\beta}_1, \hat{\beta}_2)$$

$$\simeq \frac{\beta^2}{n(r_u + r_{v'})^2}\Big[(B_2 - 1)\,\rho^2 r_u r_{v'}\,(r_u + r_{v'})^2 + (r_u + r_{v'})^3$$

$$+ r_u r_{v'}(r_u + r_{v'})^2 - 2\rho\sqrt{r_u r_{v'}}\,(r_u + r_{v'})^2$$

$$+ \rho\sqrt{r_u r_{v'}}\,(-4r_u)\,(r_u^2 - r_{v'}^2)$$

$$+ \rho^2 r_u r_{v'}\,(5r_u - 3r_{v'}^2 - 6r_u r_{v'})$$

$$V\left(\pm\sqrt{\hat{\beta}_1\,\hat{\beta}_2}\right) = \frac{1}{4}\Big[\frac{V(\hat{\beta}_1)}{(1 + \rho\sqrt{r_u r_{v'}})^2} + V(\hat{\beta}_2)\,(1 + \rho\sqrt{r_u r_{v'}})^2$$

$$+ 2\,Cov(\hat{\beta}_1, \hat{\beta}_2)\Big]$$

$$\simeq \frac{\beta^2}{n}\Big(r_u + r_{v'} + \frac{r_u^2 + r_{v'}^2}{2} + \rho\sqrt{r_u\,r_{v'}}\,(r_u + r_{v'} - 2)$$

$$- \rho^2\,r_u\,r_{v'}\Big)$$

Appendix 6.2

Our problem is to minimize the asymtotic variance of $\hat{\beta}(a)$ subject to $\Sigma a_i = 1$ where $\hat{\beta}(a) = \Sigma a_i \hat{\beta}_i$. Consider

$$V = \Sigma a^2 \hat{w}_{ii} + \sum_{i \neq j} a_i a_j \hat{w}_{ij} - \lambda(\Sigma a_i - 1),$$

where λ is the Lagrangean multiplier and $\hat{w}_{ij}$ is the asymptotic covariance between $\hat{\beta}_i$ and $\hat{\beta}_j$. Taking partial derivative of V with respect to a_i and equating it to zero, we get

$$2a_i \hat{w}_{ii} + 2 \sum_{j \neq i=1}^{K} a_j \hat{w}_{ij} - \lambda = 0 \quad \forall i$$

Adding over all i,

$$2 \hat{W} \hat{a} = \lambda e$$

or $$\hat{a} = \frac{\hat{\lambda}}{2} \hat{W}^{-1} e .$$

Substituting the above expression in $\Sigma a_i = 1$ we get

$$\hat{a}' e = \frac{\hat{\lambda}}{2} e' \hat{W}^{-1} e = 1$$

or $$\frac{\hat{\lambda}}{2} = (e' \hat{W}^{-1} e)^{-1}$$

Hence, $$\hat{a} = \frac{\hat{W}^{-1} e}{e' \hat{W}^{-1} e} .$$

Now (i,j)th element of W is

$$w_{ij} = \text{Asym. Cov}\left(\frac{\Sigma yz_i}{\Sigma xz_i}, \frac{\Sigma yz_j}{\Sigma xz_j} \right)$$

$$= \frac{1}{n} \text{plim} \frac{n(\Sigma z_i \varepsilon')(\Sigma z_j \varepsilon')}{(\Sigma xz_i)(\Sigma xz_j)} \quad (\text{where } \varepsilon' = \varepsilon + v - \beta u)$$

$$= \frac{1}{n} \text{plim} \frac{\frac{1}{n}(\Sigma z_i \varepsilon')(\Sigma z_j \varepsilon')}{\frac{1}{n^2}(\Sigma xz_i)(\Sigma xz_j)}$$

$$= \frac{1}{n} \frac{\text{plim} \frac{1}{n}(\Sigma z_i \varepsilon')(\Sigma z_j \varepsilon')}{\text{plim} \frac{1}{n^2}(\Sigma xz_i)(\Sigma xz_j)} = \left(\frac{1}{n} \sigma^2_{\varepsilon'}\right) \frac{\sigma_{ij}}{\sigma_{xi}\, \sigma_{xj}}$$

$$\therefore \hat{w}_{ij} \alpha \frac{\hat{\sigma}_{ij}}{\hat{\sigma}_{xi}\, \hat{\sigma}_{xj}}$$

Appendix 6.3

$$\hat{\beta}_{02} = \frac{\Sigma c_j\,(\Sigma y z_j')}{\Sigma c_j(\Sigma x z_j')}$$

$$= \frac{\beta \sum_j c_j(\Sigma X z_j') + \sum_j c_j(\Sigma \varepsilon^* z_j')}{\sum_j c_j(\Sigma X z_j') + \sum_j c_j(\Sigma u z_j')} \quad ; \quad (\varepsilon^* = \varepsilon + v)$$

$$= \beta\left\{1 + \frac{\Sigma c_j(\Sigma u z_j')}{\Sigma c_j(\Sigma X z_j')}\right\}^{-1} + \frac{\Sigma c_j(\Sigma \varepsilon^* z_j')}{\Sigma c_j(\Sigma X z_j')}\left\{1 + \frac{\Sigma c_j(\Sigma u z_j')}{\Sigma c_j(\Sigma X z_j')}\right\}^{-1}$$

$$= \beta\left\{1 - \frac{\Sigma c_j(\Sigma u z_j')}{\Sigma c_j(\Sigma X z_j')} + \dots\right\} + \frac{\Sigma c_j(\Sigma \varepsilon^* z_j')}{\Sigma c_j(\Sigma X z_j')}\left\{1 - \frac{\Sigma c_j(\Sigma u z_j')}{\Sigma c_j(\Sigma X z_j')} + \dots\right\}$$

Hence, asymptotically

$$V(\hat{\beta}_{02}) \simeq \frac{1}{n} \operatorname{plim} n \Big[-\beta\left\{\frac{\Sigma c_j(\Sigma u z_j')}{\Sigma c_j(\Sigma X z_j')} - \dots\right\} + \frac{\Sigma c_j(\Sigma \varepsilon^* z_j')}{\Sigma c_j(\Sigma X z_j')}\left\{1 - \frac{\Sigma c_j(\Sigma u z_j')}{\Sigma c_j(\Sigma X z_j')} + \dots\right\} \Big]^2$$

$$\simeq \frac{1}{n}\beta^2 \operatorname{plim} n\left\{\frac{\Sigma c_j(\Sigma u z_j')}{\Sigma c_j(\Sigma X z_j')}\right\}^2 + \frac{1}{n}\operatorname{plim} n\left\{\frac{\Sigma c_j(\Sigma \varepsilon^* z_j')}{\Sigma c_j(\Sigma X z_j')}\right\}^2$$

(neglecting terms of higher order)

$$= \frac{1}{n}\beta^2 \frac{\operatorname{plim} \frac{1}{n}\Sigma c_k c_l(\Sigma u z_k')(\Sigma u z_l')}{\operatorname{plim}\{\Sigma c_j(\frac{1}{n}\Sigma X z_j')\}^2}$$

$$+ \frac{1}{n} \frac{\operatorname{plim} \frac{1}{n}\Sigma c_k c_l(\Sigma \varepsilon^* z_k')(\Sigma \varepsilon^* z_l')}{\operatorname{plim}\{\Sigma c_j(\frac{1}{n}\Sigma X z_j')\}^2}$$

$$= \frac{1}{n} \left[\frac{\beta^2 \sigma_u^2 \Sigma c_k c_l \sigma'_{kl} + \sigma_\varepsilon^{*2} \Sigma c_k c_l \sigma'_{kl}}{(\Sigma c_j \sigma'_{Xj})^2} \right]$$

$$= \frac{1}{n} (\beta^2 \sigma_u^2 + \sigma_\varepsilon^{*2}) \frac{\Sigma c_k c_l \sigma'_{kl}}{(\Sigma c_j \sigma'_{Xj})^2} .$$

Taking partial derivatives and equating it to zero, assuming that the variance-covariance matrix of z' is I, one gets

$$c_i \alpha \sigma'_{Xi}$$

Hence the estimate of c_i is proportional to $\frac{1}{n} \Sigma Xz_i$, or asymptotically to $\frac{1}{n} \Sigma \bar{x} z_i$, since

$$\underset{n \to \infty}{\text{plim}} \frac{1}{n} \Sigma u z_i = 0$$

REFERENCES

1. Adams, G. F., and De Janosi, P. E. (1966) : On the Statistical Discrepancy in the Revised U. S. National Accounts, Journal of American Statistical Association, Vol. 61, pp. 1219-1229.

2. Adcock, R. J. (1878) : A Problem in Least Squares, Analyst, Vol. 5, pp. 53-54.

3. Ahmed, M., and Bhattacharya, N. (1972) : Size Distribution of Per Capita Personal Income in India, Economic and Political Weekly, Special Number, Vol. 7, pp. 1581-1588.

4. Aigner, D. J. (1973) : Regression With A Binary Independent Variable Subject to Errors of Observations, Journal of Econometrics, Vol. 1, pp. 49-60.

5. ____________ (1974) : MSE Dominance of L. S. with Errors of Observation, Journal of Econometrics, Vol. 2, pp.365-372.

6. Aitchison, J., and Brown, J. A. C. (1957) : The Lognormal Distribution, Cambridge University Press, Cambridge.

7. Allen, H. V. (1938) : A Theorem Concerning the Linearity of Regression, Statistical Research Memoire, Vol. 2, pp. 60-68.

8. Allen, R. G. D. (1939) : Assumptions on Linear Regression, Econometrica, Vol. 6, pp. 199-204.

9. Anderson, T. W., and Cheng, H. (1980) : Estimation of Dynamic Models With Error Components, Presented to the 4th World Congress, Econometric Society, Aix- En Province, France.

10. Attfield, C. L. F. (1977) : Estimation of a Model Containing Unobservable Variables Using Grouped Observations, Journal of Econometrics, Vol. 6, pp. 51-63.

11. Avery, R. B. (1977) : Error Components and Seemingly Unrelated Regressions, Econometrica, Vol. 45, pp. 199-209.

12. Barnett, V. D. (1967) : A Note on Linear Structural Relationships When Both Residual Variances are Known, Biometrika, Vol. 54, pp. 670-672.

13. Bartlett, M. S. (1949) : Fitting of Straight Lines When Both Variables are Subject to Error, Biometrics, Vol. 5, pp. 207-212.

14. Berkson, J. (1950) : Are There Two Regressors ? Journal of American Statistical Association, Vol. 45, pp.164-180.

15. Bhargava, A. K. (1977) : ML Estimation in a Multivariate EIV Regression Model With Unknown Covariance Matrix, Communications in Statistics, Vol. A.6, pp.587-601.

16. Bhattacharya, N. (1967) : Consumer Behaviour in India : An Application of the Linear Expenditure System, Economic and Political Weekly, Vol. 2, pp.2093-2098.

17. ______________ (1978) : A Study on Consumer Behaviour in India, in the Series, A Survey of Research in Economics, Vol.7, Econometrics, Ed. of the Series: D. T. Lakdawala, Eds. of the Vol. : N. S. Iyengar and N. Bhattacharya, Allied Publishers, Bombay.

18. ______________ and Chatterjee, G. S. (1971) : On Rural-Urban Differential in Consumer Prices and Per Capita Household Expenditure in India by Levels of Living, Sankhyā, Series B, Vol. 33, pp. 355-370.

19. ______________ and Iyengar, N. S. (1961) : Changes in the Pattern of Distribution of Persons by Monthly Per-Capita Total Consumer Expenditure Over NSS Rounds, Working Paper No.218, Planning Division, Indian Statistical Institute, Calcutta.

20. Blalock, H. M. (Jr.) (1969) : Multiple Indicators and the Causal Approach to Measurement Error, American Journal of Sociology, Vol. 75, pp. 264-272.

. ______________ (1970) : Estimating Measurement Error Using Multiple Indicator and Several Po..nts in Time, American Sociological Review, Vol. 35, pp.101-111.

2. Bloch, F. E. (1978) : Measurement Error and Statistical Significance of an Independent Variable, The American Statistician, Vol. 32, pp. 26-27.

3. Borus, M. E. (1966) : Response Error in Survey Reports of Earnings Information, Journal of American Statistical Association, Vol. 61, pp. 729-738.

24. Boudon, R. (1965) : A Method of Linear Causal Analysis : Dependence Analysis, American Sociological Review, Vol. 30, pp. 365-373.

25. __________ (1967) : L'analyse Mathematique Des Faits Sociaux, Pair, Plom.

26. __________ (1968) : A New Law at Correlation Analysis, Chapter 6, pp. 199-235, in Methodology in Social Reserch; (eds.) H. B. Blalock and A. B. Blalock, New York.

27. Box, G. E. P., and Cox, D. R. (1964) : An Analysis of Transformations, Journal of Royal Statistical Society, Series B, Vol. 26, pp. 211-243.

28. Carlson, F. D., Sobel, E., and Watson, G. S. (1966) : Linear Functional Relationships Between Variables Affected by Errors, Biometrics, Vol. 22, No. 252-267.

29. Carter, R. L., and Fuller, W. A. (1980) : Instrumental Variable Estimation of the Simple EVM, Journal of American Statistical Association, Vol. 75, pp. 687-692.

30. Chan, L. K., and Mak, T. K. (1979) : ML Estimation of Linear Structural Relationship With Replication, Journal of Royal Statistical Society, Series B, Vol. 41, pp.263-268.

31. Chaudhuri, M. (1979) : Some Problems on Econometric Regression Analysis, Ph. D. Thesis, Indian Statistical Institute, Calcutta.

32. Clutton-Brock, M. (1967) : ML Distribution for Estimation of Functions When Both Variables are Subject to Error, Technometrics, Vol. 9, pp. 261-269.

33. Cochran, W. G. (1968) : Errors of Measurement in Statistics, Technometrics, Vol. 10, pp.637-666.

34. ______________ (1970) : Some Effects of Errors of Measurement on Multiple Correlation, Journal of American Statistical Association, Vol. 65, pp. 22-34.

35. ______________ (1971) : Some Effects of Errors of Measurement on Linear Regression, Proceedings of the Sixth Berkely Symposium on Mathematical Statistics and Probability, pp. 527-539.

36. ______________ and Rubin, D. B. (1973) : Controlling Bias in Observational Studies, Sankhyā, Series A, Vol. 35, pp. 417-446.

37. Costner, H. L. (1969) : Theory, Deduction and Rules of Correspondence, American Journal of Sociology, Vol. 75, pp.245-263.

38. Creasy, M. A. (1956) : Confidence Limits for the Gradient in the Linear Functional Relationship, Journal of Royal Statistical Society, Series B, Vol. 18, pp.65-69.

39. Degracie, J. S. and Fuller, W. A. (1972) : Estimation of the Slope and Analysis of Covariance When the Concomitant Variable is Measured With Error, Journal of American Statistical Association, Vol. 67, pp.930-937.

40. De Janosi, P. E. (1961) : The Statistical Discrepancy in the National Accounts Revisited, Econometrica, Vol. 29, pp. 427-429.

41. Dolby, G. R. (1972) : Generalized Least Squares and MLE of Non-linear Functional Relationship With Replicated Observations and Correlated Errors, Journal of Royal Statistical Society, Series B, Vol. 34, pp. 393-400.

42. ____________ (1976) : The Ultra-Structural Relation : A Synthesis of the Functional and Structural Relations, Biometrika, Vol. 63, pp. 39-50.

43. ____________ and Freeman, T. G. (1975) : Functional Relationships Having Many Independent Variables and Errors With Multivariate Normal Distribution, Journal of Multivariate Analysis, Vol. 5, pp. 466-479.

44. ____________ and Lipton, S. (1972) : MLE of the General Non-linear Functional Relationship With Replicated Observations and Correlated Errors, Biometrika, Vol. 59, pp.121-129.

45. Drèze, J. H. (1974) : Bayesian Theory of Identification in Simultaneous Equations Models, in studies in Bayesian Econometrics and Statistics, (eds.) S. E. Fienberg and A. Zellner, North Holland, pp. 159-174.

46. Drion, E. F. (1951) : Estimation of the Parameters of a Straight Line and of the Variances of the Variables, if they are Both Subject to Error, Indagationes Mathematicae, Vol. 13, pp. 256-260.

47. Duncan, O. D. (1975) : Introduction to Structural Equation Models, Academic Press, New York.

48. ____________ and Featherman, D. L. (1972) : Psychological and Cultural Factors in the Process of Educational Achievement, Social Science Research, Vol. 1, pp.121-145.

49. Durbin, J. (1954) : Errors in Variables, Revue de L'Institute International de Statistique, Vol. 22, pp.23-32.

50. Elderton, W. P. (1953) : Frequency Curves and Correlation, Cambridge University Press, Cambridge.

51. Federov, V. V. (1974) : Regression Problems With Controllable Variables Subject to Error, Biometrica, Vol. 61, pp.49-56.

52. Feldstein, M. S. (1974) : Errors in Variables : A Consistent Estimator With Small MSE in Finite Samples, Journal of American Statistical Association, Vol. 69, pp.990-996.

53. Ferber, R. (1965) : The Reliability of Consumer Surveys of Financial Holdings : Time Deposits, Journal of American Statistical Association, Vol.60, pp.148-163.

54. Fisher, F. M. (1966) : The Identification Problem in Econometrics, McGraw Hill, New York.

55. Fix, E. (1949) : Distributions Which Lead to Linear Regressions, Proceedings of First Berkley Symposium on Mathematical Statistics and Probability, University of California Press.

56. Florens, J. P., Mouchart, M., and Richard, J. F. (1974) : Bayesian Inference in EVM, Journal of Multivariate Analysis, Vol. 4, pp.419-452.

57. Friedman, M. (1957) : A Theory of the Consumption Function, Princeton University Press, Princeton.

58. Frisch, R. (1934) : Statistical Confluence Analysis by Means of Complete Regression Systems, Oslo.

59. Fuller, W. A. (1980) : Properties of Some Estimators For the Errors-in-Variables Model, The Annals of Statistics, Vol. 8, pp. 407-422.

60. Garber, S., and Klepper, S. (1980) : Extending the Classical Normal EVM, Econometrica, Vol. 48, pp.1541-1546.

61. Geary, R. C. (1942) : Inherent Relations Between Random Variables, Proceedings of Royal Irish Academy, Vol. 47, pp.63-76.

62. ________ (1943) : Relations Between Statistics : The General and the Sampling Problem When the Samples are Large, Proceedings of Royal Irish Academy, Vol. 49, pp.177-196.

63. ______________ (1949) : Determination of Linear Relations Between Systematic Parts of Variables With Errors of Observations the Variances of Which are Unknown, Econometrica, Vol. 17, pp. 30-59.

64. ______________ (1953) : Non-linear Functional Relationship Between Two Variables When One Variable is Controlled, Journal of American Statistical Association, Vol. 48, pp. 94-103.

65. Geraci, V. J. (1977a) : Identification of Simultaneous Equations Models with Measurement Errors, in Latent Variables in Socio-Economic Models, (eds.) D. J. Aigner and A. S. Goldberger, North Holland Publishing Co., Also in Journal of Econometrics, 1976, Vol. 4, pp.262-283.

66. ____________ (1977b) : Estimation of Simultaneous Equation Models With Measurement Error, Econometrica, Vol. 45, pp.1243-1255.

67. ____________ (1980) : EIV and the Individual Structural Equation, Presented to the 4th World Congress, Econometric Society, Aix-En-Province, France.

68. Gibson, W. M., and Jowett, G. H. (1957) : Three-Group Regression Analysis : Part 1 Applied Statistics, Vol. 6, pp.114-122.

69. Giles, R. L. (1980) : Errors in Omission and Measurement : Estimating the Parameters of the Variable Subject to Error, Presented to the 4th World Congress, Econometric Society, Aix-En-Province, France.

70. Gini, C. (1921) : Sull'interpolazione di Una Retta Quando I Valori Della Variabile Independente Sono Affetti da Errori Accidentali, Metron, Vol. 1, pp. 63-82.

71. Gleser, L. J., and Watson, G. S. (1973) : Estimation of a Linear Transformation, Biometrika, Vol. 60, pp.525-534.

72. Gnedenko, B. V., and Kolmogorov, A. N. (1954) : Limit Distributions of Sums of Independent Random Variables, Addison Wesley, Cambridge.

73. Goldberger, A. S. (1964) : Econometric Theory, Wiley, New York.

74. ______________ (1970) : On Boudon's Method of Linear Causal Analysis, American Sociological Review, Vol.35, pp.97-101.

75. ______________ (1972a) : Structural Equation Methods in the Social Sciences, Econometrica, Vol. 40, pp.979-1001.

76. ______________ (1972b) : ML Estimation of Regressions Containing Unobservable Independent Variables, International Economic Review, Vol. 13, pp.1-15.

77. ______________ (1974) : Unobservable Variables in Econometrics, In Frontiers of Econometrics, (ed.) P. Zarembka, Academic Press, New York.

78. Graupe, D. (1972) : Identification of Systems, Van Nostrand Reinhold Co., New York.

79. Grether, D. M., and Maddala, G. S. (1973) : Errors in Variables and Serially Correlated Distributed Lag Models, Econometrica, Vol. 41, pp. 255-262.

80. Griliches, Z. (1967) : Distributed Lags : A Survey, Econometrica, Vol. 35, pp. 16-4?.

81. ______________ (1974) : Errors in Variables and Other Unobservables, Econometrica, Vol. 42, pp. 971-998.

82. ______________ (1979) : Sibling Models and Data in Economics: Beginnings of a Survey, Journal of Political Economy, Vol. 87, pp. S37-S64.

83. ______________ and Mason, W. (1972) : Income and Ability, Journal of Political Economy, Vol. 80, S74-S103.

84. Haavelmo, T. (1944) : The Probability Approach to Econometrics, Econometrica, Vol. 12, Supplement.

85. Halperin, M. (1961) : Fitting of Straight Lines and Prediction When Both Variables are Subject to Error, Journal of American Statistical Association, Vol. 56, pp. 657-669.

86. ______________ (1964) : Interval Estimation in Linear Regression When Both Variables are Subject to Error, Journal of American Statistical Association, Vol. 59, pp.1112-1120.

87. ______________ and Gurian, J. (1971) . A Note on Estimation in Straight Line When Both Variables are Subject to Error, Journal of American Statistical Association, Vol. 66, pp. 587-589.

88. Hannan, E. J. (1971) : The Identification Problem for Multiple Equations Systems With Moving Average Errors, Econometrica, Vol. 39, pp. 751-766.

89. ____________ (1976) : The Identification and Parametrization of Armax and State Space Forms, Econometrica, Vol. 44, pp. 713-723.

90. Haugh, L. D. and Box, G. E. P. (1977) : Identification in Dynamic Regression (Distributed Lag) Models Connecting Two Time Series, Journal of American Statistical Association, Vol. 72, pp. 121-130.

91. Hauser, R. M. (1972) : Disaggregating a Social-Psychological Model of Educational Attainment, Social Science Research, Vol. 1, pp. 159-218.

92. __________ and Goldberger, A. S. (1971) : The Treatment of Unobservable Variables in Path Analysis, in Sociological Methodology 1971, (ed.) H. L. Costner, San Francisco, Jossey-Bass, Chapter 4, pp. 81-177.

93. Hausman, J. A. (1977) : Errors in Variables in Simultaneous Equations Models, Journal of Econometrics, Vol. 5, pp.389-401.

94. ____________ and Taylor, W. E. (1980) : Panel Data and Unobservable Individual Effects : An Efficient IV's Approach, Presented to the 4th World Congress, Econometric Society, Aix-En-Province, France.

95. Healy, J. D. (1980) : ML Estimation of a Multivariate Linear Functional Relationship, Journal of Multivariate Analysis, Vol. 10, pp. 243-251.

96. Housner, G. W., and Brennan, J. F. (1948) : Estimation of Linear Trends, Annals of Mathematical Statistics, Vol. 19, pp. 380-388.

97. Hsiao, C. (1976a) : Identification and Estimation of Simultaneous Equation Models With Measurement Error, International Economic Review, Vol. 17, pp. 319-339.

98. ________ (1976b) : Measurement Error in a Dynamic Simultaneous Equations Models With Stationary Disturbances, Institute of Business and Economic Research, University of California, Berkeley.

99. ________ (1977) : Identification for a Linear Dynamic Simultaneous Error-Shock Model, International Economic Review, Vol. 18, pp. 181-194.

100. Iyengar, N. S. (1967) : Some Estimates of Engel Elasticities Based on NSS data, Journal of Royal Statistical Society, Series A, Vol. 130, pp. 84-101.

101. ________ and Jain, L. R. (1974) : On a Method of Estimating Income Distribution, Economic and Political Weekly, Vol.9, pp. 2103-2109.

102. Jain, L. R. (1977) : On Fitting a Three-Parameter Lognormal Distribution to Consumer Expenditure Data, Sankhyā, Series C, Vol. 39, pp. 61-73.

103. Johnston, J. (1960, 1972) : Econometric Methods, McGraw Hill Book Co., New York.

104. Joreskog, K. G. (1970) : A General Method for the Analysis of Covariance Structure, Biometrika, Vol. 57, pp. 239-251.

105. ____________ (1973) : Analysis of Covariance Structure, in Multivariate Analysis - III, (ed.) P. R. Krishnaiah, Academic Press, New York.

106. ____________ and Goldberger, A. S. (1975) : Estimation of A Model With Multiple Indicators and Multiple Causes of a Single Latent Variable, Journal ofAmerican Statistical Association, Vol. 70, pp. 631-639.

107. Kadane, J. B. (1974) : The Role of Identification in Bayesian Theory, in Studies in Bayesian Econometrics and Statistics, (eds.) S. E. Fienberg and A. Zellner, North Halland, 1974, pp. 175-191, Also Presented to the Winter Meeting of the Econometric Society, Toronto, December, 1972.

108. Kaila, K. L. (1980) : A New Method for Linear Curve Fitting With Both Variables in Error, Sankhyā, Series B.

109. Karni, E., and Weissman, I. (1974) : A Consistent Estimator of the Slope in a Regression in EIV, Journal of American Statistical Association, Vol. 69, pp.211-213.

110. Katona, G., Klein, L. R., Lansing, J. B., and Morgan, J. N. (1954) : Contributions of Survey Methods to Econometrics, Columbia, Referred to in Lancaster (1968).

111. Keating, E., Peterson, D. G., and Stone, C. H. (1950) : Validity of Work Histories Obtained by Interview, Journal of Applied Psychology, Vol. 34, pp.6-11.

112. Kendall, M. G. (1951) : Regression Structure and Functional Relationship. Part I, Biometrika, Vol. 38, pp. 11.25.

113. ____________ (1952) : Regression, Structure, and Functional Relationship, Part II, Biometrika, Vol. 39, pp.96-108.

114. Keifer, J., and Wolfowitz, J. (1956) : Consistency of the Maximum Likelihood Estimator in the Presence of Infinitely Many Incidental Parameters, Annals of Mathematical Statistics, Vol. 27, pp. 887-906.

115. Koopmans, T. C. (1937) : Linear Regression Analysis of Econometric Time Series, Haarlem, F. Bohn.

116. ______________ (1950) : Statistical Inference in Dynamic Economic Models, Cowles Commission Monograph, 10, John Wiley and Sons.

117. ______________ and Reiersol, O. (1950) : The Identification of Structural Characteristics, Annals of Mathematical Statistics, Vol. 21, pp. 165-181.

118. Kummel, C. H. (1979) : Reduction of Observation Equations Which Contain More Than One Observed Quantity, Analyst, Vol. 6, pp. 97-105.

119. Kunitomo, N. (1980) . Asymptotic Expansions of the Distributions of Estimators in a Linear Functional Relationship and Simultaneous Equations, Journal of American Statistical Association, Vol. 75, pp.693-700.

120. Laha, R. G. (1956) : On a Characterization of the Stable Law With Finite Expectation, Annals of Mathematical Statistics, Vol. 27, pp.187-195.

121. Lahiri, K., and Schmidt, P. (1978): On the Estimation of Triangular Structural Systems, Econometrica, Vol. 46, pp.1217-1221.

122. Lancaster, T. (1966) : A Note on an EIV Model, Journal of American Statistical Association, Vol. 61, pp.128-135.

123. ______________ (1968) : Grouping Estimators on Heteroscedastic Data, Journal of American Statistical Association, Vol.63, pp. 182-191.

124. Langaskens, Y., and Van Rijckeghem, W. (1967) : A Statistical and Econometric Comparison of Two Systems of National Accounts; DULBEA and NIS, (In Dutch) Cahiers Economiques de Bruxelles, 4th Quarter, pp. 551-587, Referred to in Langasken and Van Rijckeghem (1974).

125. ______________ and ______________ (1974) : A New Method to Estimate Measurement Errors in National Income Account Statistics : The Belgian Case, International Statistical Review, Vol. 42, pp. 283-290.

126. Lankipalle, K. N. (1973) : Some Effects of Errors in Variables on Correlations, Presented to the 13th Conference of the Indian Econometric Society, Referred to in Lankipalle (1975).

127. ______________ (1975) : Errors in Variables Linear Models : A Survey, Presented to the Fifteenth Annual Econometric Conference, Indian Statistical Institute, Calcutta.

128. Lansing, J. B., Ginsburg, G. P., and Braaten, K. (1961) : An Investigation of Response Error : Studies in Consumer Savings No. 2, Bureau of Economic and Business Research, University of Illinois, Illinois.

129. Leamer, E. E. (1978) . Least Squares Versus IV's Estimation in a Simple EVM, Econometrica, Vol. 46, pp. 961-968.

130. Levi, M. D. (1973) : Errors in Variables Bias in the Presence of Correctly Measured Variables, Econometrica, Vol. 41, pp. 985-986.

131. ______________ (1977) : Measurement Errors and Bounded OLS Estimates, Journal of Econometrics, Vol. 6, pp. 165-171.

132. Lindley, D. V. (1947) : Regression Lines and the Linear Functional Relationship, Journal of Royal Statistical Society, Supplementary, Vol. 9, pp. 219-244.

133. ______________ (1953) : Estimation of a Functional Relationship, Biometrika, Vol. 40, pp. 47-49.

134. ______________ and El-Sayyad, G. M. (1968) : The Bayesian Estimation of a Linear Functional Relationship, Journal of Royal Statistical Society, Series B, Vol. 30, pp. 190-202.

135. _____________ and Smith, A. F. M. (1972) : Bayes Estimates for the Linear Model, Journal of Royal Statistical Association, Series B, Vol. 34, pp. 1-41.

136. Linnik, Y. V. (1954) : On the Stability of Probability Laws With An Exponent Below Unit, DAN SSSR (Reports of the Academy of Sciences of the USSR), Vol. 24.

137. Liviatan, N. (1961) : Errors in Variables and Engel Curve Analysis, Econometrica, Vol. 29, pp. 339-362.

138. Lord, F. M. (1960) : Large Sample Covariance Analysis, When the Control Variable is Fallible, Journal of American Statistical Association, Vol. 55, pp. 307-321.

139. Madansky, A. (1959) : The Fitting of Straight Lines When Both Variables are Subject to Error, Journal of American Statistical Association, Vol. 54, pp. 173-205.

140. ___________ (1976) : Foundations of Econometrics, in the Series Advanced Textbooks in Economics, (eds.) C. J. Bliss and M. D. Intriligator, Vol. 7, North-Holland.

141. Mallios, U. S. (1969) : A General Application of IV's Estimation of Straight Line Relations When Both Variables are Subject to Error, Technometrics, Vol. 11, pp. 255-263.

142. Maravall, A. (1974) : Identification in Shock-Error Univariate Autoregressive Distributed Lag Models, Social Systems Research Institute. The University of Wisconsin, Madison.

143. ___________ (1979) . Identification in Dynamic Shock-Error Models, Lecture Notes in Economics and Mathematical Systems, (Managing Editors) M. Beckmann and H. P. Kunzi, Vol. 165.

144. ___________ and Aigner, D. J. (1977) : Identification of the Dynamic Shock-Error Model : The Case of Dynamic Regression, In Latent Variables in Socio-Economic Models, (Eds.) D. J. Aigner and A. S. Goldberger, North Holland Publishing Co.

145. Mariano, R. S. (1977) : Finite Sample Properties of IV Estimators of Structural Coefficients, Econometrica, Vol. 45, pp. 487-496.

146. McCallum, B. T. (1972) : Relative Asymptotic Bias From Errors of Ommission and Measurement, Econometrica, Vol. 40, pp. 757-758.

147. McIntyre, G. A., Brooks, C., Comston, W., and Turek, A. (1966) : The Statistical Assessment of Rb-Sr Isochrons, Journal of Geophysical Research, Vol.71, pp. 5459-5468.

148. Mehra, R. K. (1974) : Identification in Control and Econometrics : Similarities and Differences, Annals of Economic and Social Measurement, Vol.3, Referred to in Mehra (1975).

149. __________ (1975) : Identification and Estimation of the EVM in Structural Form, Research Memorandum, RM-75-38, International Institute for Applied Systems Analysis, Austria.

150. Merriman, M. (1890) : The Determination by the Method of Least Squares of the Relation Between Two Variables Connected by the Equation Y = AX+B, Both Variables Being Liable to Errors of Observation, Report of U. S. Coast and Geodetic Survey, p.687.

151. Milne, W. E. (1949) : Numerical Calculus, Princeton.

152. Mittelstaedt, P. (1976) : Philosophical Problems of Modern Physics, Vol. XVIII of Boston Studies in the Philosophy of Science, (ed.) R. S. Cohen and M. W. Wartofsky.

153. Moran, P. A. P. (1971) : Estimating Structural and Functional Relationships, Journal of Multivariate Analysis, Vol. 1, pp. 232-255.

154. Morgenstern, O. (1963) : On the Accuracy of Economic Observations, 2nd edition, Princeton University Press, Princeton.

155. Mosel, J. N., Cozan, L. W. (1952) : The Accuracy of Application Blank Work Histories, Journal of Applied Psychology, Vol. 36, pp.365-369.

156. Mukherjee, M. (1969) : National Income of India : Trends and Structure, Statistical Publishing Society, Calcutta.

157. Murray, T. W. (1972) : An Empirical Examination of the Classical Assumptions Concerning Errors in Data, Journal of American Statistical Association, Vol. 67, pp. 530-537.

158. Murthy, M. N. (1967) : Sampling Theory and Methods, Statistical Publishing Society, Calcutta.

159. Nair, K. R., and Banerjee, K. S. (1943) : A Note on Fitting of Straight Lines if Both Variables are Subject to Error, Sankhyā, Vol. 6, p. 331.

160. __________ and Shrivastava, M. P. (1942) : On a Simple Method of Curve Fitting, Sankhyā, Vol.6, pp. 121-132.

161. N. B. E. R. (1958) : An Appraisal of the 1950 Census Data, Princeton University Press, Princeton.

162. Nerlove, M. (1967) : Distributed Lags and Unobserved Components in Economic Time Series, Cowles Foundation Discussion Paper No. 221, Yale University.

163. Newbold, P. (1978) : Feedback Induced by Measurement Errors, International Economic Review, Vol. 19, pp. 787-791.

164. Neyman, J. (1951) : Existence of Consistence Estimate of the Directional Parameter in a Linear Structural Relation Between Two Variables, Annals of Mathematical Statistics, Vol. 22, pp. 496-512.

165. __________ and Scott, E. L. (1951) : On Certain Methodsof Estimating Linear Structural Relation, Annals of Mathematical Statistics, Vol. 22, pp. 352-361.

166. Nowak, E. (1977) : Identification of the Dynamic Shock-Error Model with Auto-Correlated Errors, Unpublished Manuscript (Referred to in Maravall (1979)).

67. Ord, J. K. (1969) : A New Approach to the Estimation of Parameters in Linear Functional Relationships, Proceedings of the 37th Session, Bulletin of the International Statistical Institute, pp. 169-171.

68. Pal, M. (1977) : Estimation of Parameters of Engel Curves in the Presence of Errors-in-Variables, Presented to the 17th Indian Econometric Conference, Trivandrum.

69. ______ (1980a) : Some Observations on Estimation of Bivariate Relationships in the Presence of Errors-in-Variables, Forthcoming, Sankhyā, Series B.

70. ______ (1980b) : Consistent Moment Estimators of Regression Coefficients in the Presence of Errors-in-Variables, Journal of Econometrics, Vol. 14, pp. 349-364.

171. ______ and Bhaumik, M. (1979) : A Note on Bartlett's Method of Grouping in Regression Analysis, Tech. Report No.ERU/5/79, Indian Statistical Institute, Calcutta, Forthcoming in Sankhyā D.

172. ______ and Chakravarty, S. (1978) : Some Results on Measurement of Inequality, Tech. Report No. ERU/3/78, Indian Statistical Institute, Calcutta.

173. Pandit, S. M. (1973) : Data Dependent Systems : Modeling Analysis and Optimal Control Via Time Series, Doctoral Dissertation, The University of Wisconsin, Madison.

174. Pesaran, M. H. (1980) : Identification of Rational Expectation Models, Presented to the 4th World Congress, Econometric Society, Aix-En-Province, France.

175. Phillips, G. (1980) : The Bias and Moment Matrix of the Two-Stage-Least-Squares (2SLS) Estimator in Simultaneous Equation Models With Dependent Disturbances, Presented to the 4th World Congress, Econometric Society, Aix-En-Province, France.

176. Pierce, D. A. (1976) : Seasonal Adjustment When Both Deterministic and Stochastic Seasonality are Present, Presented to the NBER, Bureau of the Census Conference on Seasonal Analysis, Washington, D. C.

177. (The) Rand Corporation (1955) : A Million Random Digits With 100,000 Normal Deviates, The Free Press Publishers, Glencoe, Illinois.

178. Rao, P., and Miller, R. L. (1972) : Applied Econometrics, Prentice Hall of India Private Limited, New Delhi.

179. Reed, A. H., and Wu, G. T. (1977) : Estimation of Bias in CLR Slope When the Proper Model is Functional Linear Regression, Communicationsin Statistics, Vol. A6(5), pp.405-416.

180. Reiersol, O. (1945) : Confluence Analysis by Means of Instrumental Sets of Variables, Ph. D. Thesis, Stockholm, Arkiv for Matematik, Astronomi Och Fysik, Vol. 32, pp. 1-119.

181. ___________ (1950) : Identifiability of a Linear Relation Between Variables Which are Subject to Error, Econometrica, Vol. 18, pp. 375-389.

182. Richardson, D. H., and Wu, De-Min. (1970) : Least Squares and Grouping Method Estimators in the EVM, Journal of American Statistical Association, Vol.65, pp.724-748.

183. Robertson, C. A. (1974) : Large Sample Theory for the Linear Structural Relation, Biometrika, Vol. 61, pp. 353-359.

184. Robinson, P. M. (1974) : Identification, Estimation and Large Sample Theory for Regressions Containing Unobservable Variables, International Economic Review, Vol.15, pp. 680-692.

185. Rogot, E. (1961) : A Note on Measurement Errors and Detecting Real Differences, Journal of American Statistical Association, Vol. 56, pp. 314-319.

186. Roos, C. F. (1937) : A General Invariant Criterion of Fit for Lines and Planes Where all Variables are Subject to Error, Metron, Vol. 12, pp.1-20.

187. Roy, J., and Dhar, S. K. (1960) : A Study on the Pattern of Consumer Expenditure in Rural and Urban India, in Studies on Consumer Behaviour, Asia Publishing House, Bombay.

188. ________ and __________ (1961) : A Study on the Consumer Expenditure in Rural and Urban India, Part II : Changes During the Period July 1954 to June 1959, Presented to the Second Indian Econometric Conference, Waltair.

189. Salem, A. B. Z., and Mount, T. D. (1974) : A Convenient Descriptive Model of Income Distribution : The Gamma Density, Econometrica, Vol. 42, pp. 1115-1127.

190. Sargan, J. D. (1958) : The Estimation of Economic Relationships Using Instrumental Variables, Econometrica, Vol.26, pp.393-415.

191. ____________ and Mikhail, W. M. (1971) : A General Approximation to the Distribution of IV's Estimator, Econometrica, Vol.39, pp. 131-169.

192. Scheffe, H. (1958) : Fitting Straight Lines When One Variable is Controlled, Journal of American Statistical Association, Vol. 53, pp.106-117.

193. Schneeweiß, H. (1976) : Consistent Estimation of a Regression With Errors-in-Variables, Metrika, Vol. 23, pp.101-115.

194. Scott, E. L. (1950) : Note on Consistent Estimates of the Linear Structural Relation Between Two Variables, Annals of Mathematical Statistics, Vol. 21, pp. 284-288.

195. Seares, F. H. (1944) : Regression Lines and the Functional Relation, Astrophysical Journal, Vol. 100, 255-263.

196. Sil, S., Bhaumik, M., and Pal, M. (1981) : A Note on Bartlett's Grouping Estimators in an Heteroscedastic Model, Tech. Report No. ERU/2/81, Indian Statistical Institute, Calcutta.

197. Solari, M. E. (1969) : The ML Solution of the Problem of Estimating a Linear Functional Relationship, Journal of Royal Statistical Society, Series B, Vol. 31, pp. 372-375.

198. Som, R. K. (1973) : Recall Lapse in Demographic Enquiries, Asia Publishing House, Bombay,

199. Sorenson, H. (1966) : Kalman Filtering Techniques, In Advances in Control Systems, (Ed.) C. Leondes, Academic Press, New York.

200. Spiegelman, C. (1979) : On Estimating the Slope of a Straight Line When Both Variables are Subject to Error, The Annals of Statistics, Vol. 7, pp. 201-206.

201. Sprent, P. (1966) : A Generalized L. S. Approach to Linear Functional Relationships, Journal of Royal Statistical Society, Series B, Vol. 28, pp. 278-297.

202. Stroud, T. W. F. (1972) : Comparing Conditional Means and Variances in a Regression Model With Measurement Errors of Known Variances, Journal of American Statistical Association Vol. 67, pp. 407-412.

203. Theil, H. (1958, 1961) : Economic Forecasts and Policy, North-Holland Publishing Co., Amsterdam.

204. ________ and Van Yzeren, J. (1956) : On the Efficiency of Wald's Method of Fitting Straight Lines, Review of International Statistical Institute, Vol. 24, pp. 17-26.

205. Tintner, G. (1944) : An Application of the Variate Difference Method to Multiple Regression, Econometrica, Vol. 12, pp. 97-113.

)6. ______________ (1945) : A Note Concerning Hotelling's Method of Inverting a Partitioned Matrix, Annals of Mathematical Statistics, Vol. 16, pp. 304-308.

)7. ______________ (1946) : Multiple Regression For Systems of Equations, Econometrica, Vol. 14, pp.5-36.

)8. Tukey, J. W. (1951) : Components of Regression, Biometrics, Vol.7, pp. 33-70.

)9. ____________ (1957) : On the Comparative Anatomy of Transformation, Annals of Mathematical Statistics, Vol.28, pp. 602-632.

10. Velikanov, M. A. (1965) : Measurement Errors and Empirical Relations, Israel Program for Scientific Translations, Jerusalem.

11. Villegas, C. (1961) : MLE of a Linear Functional Relationship, Annals of Mathematical Statistics, Vol. 32, pp.1048-1062.

12. Wald, A. (1940) : Fitting of Straight Lines if Both Variables are Subject to Error, Annals of Mathematical Statistics, Vol. 11, pp.284-300.

13. ________ (1950) : Note on the Identification of Economic Relations, in Statistical Inference in Dynamic Economic Models, Cowles Commission Monograph 10, John Wiley and Sons, New York.

14. Ware, J. H. (1972) : The Fitting of Straight Lines When Both Variables are Subject to Error and the Ranks of the Means are Known, Journal of American Statistical Association, Vol. 67, pp. 891-897.

15. Warren, R. D., Whittle, J. K., and Fuller, W. A. (1974) : An EIV's Analysis of Managerial Role Performance, Journal of American Statistical Association, Vol.64, pp.886-893.

216. Welch, F. (1975) : Human Capital Theory : Education, Discrimination and Life Cycles, American Economic Review, Vol. 65, pp. 63-73.

217. Werts, C. E., Joreskog, K. G., and Linn, R. L. (1973) : Identification and Estimation in Path Analysis With Unmeasured Variables, American Journal of Sociology, Vol. 78, pp. 1469-1484.

218. Whittle, P. (1963) : Prediction and Regulation by Linear Least Squares Methods, The English University Press Ltd., London.

219. Wickens, M. R. (1972) : A Note on the Use of Proxy Variables, Econometrica, Vol. 40, pp.759-760.

220. Wiley, D. E., and Wiley, J. A. (1970) : The Estimation of Measurement Error in Panel Data, American Sociological Review, Vol. 35, pp.112-117.

221. Willasen, Y. (1977) : On Identifiability of Stochastic Difference Equations With EIV in Relation to Identifiability of the Classical EIV Models, Scandinavian Journal of Statistics, Vol. 4, pp. 119-124.

222. Wold, H. (1955) : Random Normal Deviates, Cambridge University Press.

223. Wolfowitz, J. (1952) : Consistent Estimators of the Parameters of a Linear Structural Relation, Skandinavisk Acturietidaskrift, Vol. 35, pp. 132-151.

224. ____________ (1953a) : Estimation by the Minimum Distance Method, Annals of the Institute of Statistical Mathematics, Vol. 5, pp. 9-23.

225. ____________ (1953b) : The Minimum Distance Method, Annals of Mathematical Statistics, Vol. 28, pp. 89-110.

226. Zellner, A. (1970) : Estimation of Regression Relationships Containing Unobservable Variables, International Economic Review, Vol.11, pp. 441-451.

227. __________ (1971) : An Introduction to Bayesian Inference in Econometrics, John Wiley and Sons, New York.

228. __________ and Palm, F. (1974) : Time Series Analysis and Simultaneous Equation Econometric Models, Journal of Econometrics, Vol. 2, pp. 17-54.

229. __________ and Park, S. B. (1979) : Minimum Expected Loss (MELO) Estimators for Functions of Parameters and Structural Coefficients of Econometric Models, Journal of American Statistical Association, Vol. 74, pp. 185-193.