ON SPECIFICATION AND STATISTICAL INFERENCE IN SINGLE EQUATION REGRESSION MODELS

NITYANANDA SARKAR

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

1981

Contents

Introduction

This thesis makes an attempt to consider and provide solutions to some problems of what may be broadly described as specification and statistical inference in single equation linear regression models. It is divided into three parts. The first part (i.e., Part I) deals with some problems connected with the Box-Cox transformation of variables. The other two parts (i.e., Parts II and III) are devoted to considerations of autocorrelated linear regression models where the error term consists of two components. While the studies in Part II are concerned with large sample testing and estimation, those in Part III centre around a Monte Carlo investigation of small-sample properties.

In the classical linear regression model one assumes ~~inter~~ alia that (i) the relationship between the regressand (y) and the set of regressors (x's i.e., x_1, x_2, ..., x_k) is linear and (ii) the disturbances are non-autocorrelated. As regards assumption (i), one might argue that there is no reason why a linear relationship rather than some nonlinear relationship should always provide the best fit. In fact, quite often one uses different nonlinear relationships e.g., semi-log, double-log, log-inverse etc., in empirical work. The choice of a specific functional form is generally guided by the researcher's perception about the nature of the relationship derived from preliminary analysis of the sample data, knowledge acquired from previous studies or indications given by the underlying theory. However, such a priori and/or to some extent arbitrary choice about the functional form is not always quite satisfactory since there might exist other forms of relationships which would fit a given

data better. The Box-Cox transformation attempts to provide an unified and satisfactory approach of tackling this problem. It suggests a systematic procedure of choosing the particular power transformation of the original variables which would fit a given set of data best. However, it appears that even this approach of choosing nonlinear relationships between y and the x's is not fully satisfactory since in the Box-Cox transformation it is assumed that there exists a value of the transformation parameter λ for which $y^{(\lambda)}$ (the transformed value of y as per the Box-Cox transformation) and the x's $\{ x^{(\lambda)}$'s (the Box-Cox transformed values of the x's) satisfy the following desirable conditions :

(i) the relationship between $y^{(\lambda)}$ and x's $\{ x^{(\lambda)}$'s is linear,

(ii) the error variance is the same across observations,

and (iii) the distribution of the error is normal.

There have been some attempts to relax these assumptions separately. In our opinion, however, these extensions of the Box-Cox (transformation) model are not quite satisfactory and there is serious need of further work in this area. In Part I of this thesis we have tried to deal with such problems. In Chapter 1, we present a survey of the literature on the Box-Cox transformation and in the next three Chapters we tackle the problems of heteroscedasticity, autocorrelatedness and nonnormality in the context of reasonable extensions of the Box-Cox model. It has also been shown in Chapter 4 that Tobin's limited dependent variable model can be looked upon as a special case of the Box-Cox model under the assumption of a truncated distribution for the errors. Also, our treatment of the Box-Cox model with nonnormality (more specifically, with truncated normality) and

heteroscedasticity provides a generalization of Tobin model to a heteroscedastic situation.

We may now outline the work done in Part II of the thesis. So far as the assumption of uncorrelated disturbances in the standard linear regression model is concerned, it has been recognized for a long time that this is quite a restrictive assumption and that in many situations successive disturbances are likely to be correlated, particularly in time series data. The literature on how to detect the presence of autocorrelation and how to estimate the parameters of the model efficiently in case the disturbances are found to be autocorrelated is voluminous and is still growing. However, in all these studies the disturbance term is considered as having a single component which follows e.g., first order autoregressive process (AR(1)), when there is autocorrelation. It is, however, recognized that the disturbance or the error term in a linear regression model comprises a number of components representing the effects of misspecification, errors-in-observation etc., all or some of which may be autocorrelated. In fact, for a misspecified model, it can be easily shown that there are two components in the disturbance term — one due to misspecification and the other, the disturbance term associated with the true regression equation. Obviously, autocorrelation in such models can be due to autocorrelation in both the components or due to autocorrelation in one of them, while the other component is present but non-autocorrelated. It can be shown that in such situations standard methods of estimation (like Cochrane-Orcutt or Prais-Winsten) no longer remain appropriate as one or more of the properties claimed for these estimators

are lost. It, therefore, seems desirable to have a fresh look at the autocorrelated linear regression model in the light of the above observations. This is what has been attempted in Part II of our thesis. There we consider an autocorrelated linear regression model where the error term has been decomposed into two components and examine the problems of estimation of the parameters and identification of situations in terms of whether an observed autocorrelation in the disturbances results from one or both of the components being autocorrelated. After presenting a survey on the related literature in the first Chapter (i.e., Chapter 5) of this Part, we pass on to consider and provide large sample solutions to be problems of statistical inference in the above-mentioned model for two types of error processes. In Chapter 6, we assume each of the components to follow AR(1), while in Chapter 7, each component is assumed to follow an ARMA(1, 1) process. If one of the two components has its origin in mis-specification of the regression equation, our procedures for identification might enable us to decide whether or not in a given situation mis-specification is present and/or whether or not it is contributing to autocorrelation. The methods of estimation naturally depend on the results of the identification procedures.

Finally, we come to the investigations reported in Part III. When in an actual application the error term of a linear regression model is found to be autocorrelated, one often assumes the errors to follow an AR(1) process and then uses the standard reestimation procedures (Cochrane-Orcutt etc.). Analytical results bearing on the small-sample properties of these estimators are hardly available but there have been

some Monte Carlo studies on the performance of these estimators in small samples. However, no small-sample result seems to be available if one is interested in the performance of the reestimation methods in situations where misspecification is present and may contribute to the autocorrelation of the disturbances. The available large sample results for such situations indicate that these reestimation methods no longer yield even consistent estimates of the regression coefficients excepting for a special case. Also, the error term in such cases usually does not follow AR(1) process even if each of the components is AR(1) or one component is so while the other is random. Hence the use of standard reestimation methods generally give inefficient estimates in the presence of misspecification. We have carried out a Monte Carlo study to investigate how these conventional reestimation methods perform in the presence of misspecification. In Part III of this thesis we have reported these results. Such Monte Carlo studies are important because, as already mentioned, the reestimation methods are used in situations where autocorrelation is observed without paying any attention to the possible sources of autocorrelation which have a bearing on the properties of the reestimation methods. The first chapter of this Part (i.e., Chapter 8) explains the motivation for such a study and summarizes the results of earlier Monte Carlo studies on autocorrelated linear regression models. In the next Chapter (i.e., Chapter 9) we work out some large sample properties of some standard methods of estimation in a misspecified regression model with errors being autocorrelated. In this Chapter we also briefly describe the different reestimation methods like that of Cochrane-Orcutt,

Prais-Winsten, Durbin's two-step, Hildreth-Lu search procedure etc. The last two Chapters (i.e., Chapters 10 and 11) present the results of the Monte Carlo study carried out by the author. Chapter 10 is confined to the description of the plan and scope of the experiment, and Chapter 11 sets out the actual results of the experiment through statistical tables.

Detailed derivations of some of the mathematical results used in some of the Chapters are appended to these Chapters. In the Appendix to Chapter 11 we present a number of statistical tables giving further details about the results of the Monte Carlo study.

I express my sincere gratitude to Dr. Robin Mukherjee for his supervision of this work. The present shape of the thesis owes much to his meticulous editing of the manuscript. I am extremely grateful to Professor Nikhilesh Bhattacharya who, in spite of his very busy time, went through the whole manuscript and gave many useful advice and suggestions. To Dr. Dipankar Coondoo, I owe much because of his constant interest in my work and for his help in resolving many problems. Discussions with him on various problems which frequently came up in course of my work had been very fruitful. I express my sincere thanks to him. Dr. Maitreyi Chaudhuri research inspired me to take up the work presented in Parts II and III of this thesis. She also made some useful comments in the initial stages of my research. Dr. Bikas Kumar Sinha's suggestions were of much help to me in solving some problems in Chapters 4 and 7. Dr. Kesar Singh helped occasionally in the initial stages in clarifying certain statistical concepts. To all of them, I express my gratefulness. I also thank

Professors G. S. Maddala, T. Lancaster and P. K. Trivedi for their comments on some parts of my work. Messors. S. Ramani and Pradip Kumar Maitra helped me in writing the computer programs in the early stages. I thank both of them.

I would also like to thank Dr. Pradip Maiti, Mr. Bireswar Bose, Mr. Rabindranath Mukhopadhya and Miss Amita Mitra for helping me in various ways throughout the period of my work. I express my thanks to the authorities of the Indian Statistical Institute for granting me the facilities for work in the Institute and to the members of the Computer Science Unit, particularly Professor J. Roy, Mr. Ajoy Kumar Adhikari and Mr. Subhas Chandra Kundu for their help and cooperation. Finally, I thank Mr. Khagendra Chandra Neogi for his sincere and excellent typing and Mr. Dilip Chatterjee for rolling the stencils.

Abstract

This thesis attempts to consider and provide solutions to what may be broadly described as some problems of specification and statistical inference in single equation linear regression models. It consists of three parts, each having several chapters. The first part is devoted to some problems connected with the use of the well-known Box-Cox (BC) transformation of variables in single equation regression models. In the other two parts we examine an autocorrelated linear regression model from a rather unconventional angle. Precisely, we consider the problems which arise when the error term in an autocorrelated linear regression model is viewed as decomposed into two additive components, one of which may arise due to misspecification while the other may be the true disturbance. The second part deals with the analysis of such situations in large samples and the third part discusses the results of a Monte Carlo experiment comparing the performances of different estimators in small samples.

There are four chapters in the first part of which the first Chapter is devoted to critically reviewing the literature on the Box-Cox transformation. Box and Cox (1964) proposed a family of power transformations of the dependent variable (y) in a regression model defined as

$$y^{(\lambda)} = \begin{cases} \dfrac{y^{\lambda} - 1}{\lambda} & \text{if } \lambda \neq 0 \\ \ln y & \text{if } \lambda = 0 \, . \end{cases}$$

They assumed that there exists a λ for which the transformed dependent variable i.e., $y^{(\lambda)}$ will be a linear function of the regressors (or

similarly of the transformed regressors) and that the disturbance term in the transformed linear regression will be homoscedastic and normal. They also suggested maximum likelihood (ML) method of estimation for such transformed linear regression models. Subsequently, a number of studies have been made in order to examine the robustness of the ML procedure suggested by Box and Cox and/or generalizing the BC transformation by incorporating the problems of heteroscedasticity, autocorrelation and nonnormality. There are also some studies on the use of the BC transformation in simultaneous equations systems and in applied econometric work. Our review gives an account of these studies and notes their important limitations, particularly those which are relevant to our research. We have also indicated the relationship between the BC transformation with truncated normality of the disturbances and Tobin's limited dependent variable model in this chapter.

In Chapter 2, we consider the problem of heteroscedasticity in the context of the Box-Cox transformation. It has been shown by Zarembka (1974) that the ML method of estimation suggested by Box and Cox is not robust to heteroscedasticity. This led Zarembka (1974), later Egy and Lahiri (1978) and Lahiri and Egy (1981) to try to incorporate heteroscedasticity in the BC model. For this purpose Zarembka assumed

$$V(y_i) = \sigma^2 \left[E(y_i) \right]^h$$

where σ^2 is an unknown parameter and h is a known constant. Lahiri and Egy (1981), on the other hand, considered the structure for variances of $y_i^{(\lambda)}$'s as

$$V(y_i^{(\lambda)}) = \sigma^2 m_i^h$$

where m_i's are exogenously given and h and σ^2 are unknown parameters. Methods of estimation have also been worked out by these authors for the assumptions about variance stated above. Limitations of their assumptions (and hence of the extensions of the Box-Cox model in this direction) have been noted in the first as well as in this Chapter. It may, in particular, be noted that Zarembka assumed the heteroscedasticity parameter to be known and that the estimator suggested by him is only approximately consistent but not efficient. Lahiri and Egy (1981), on the other hand, avoided the complications created by the transformation itself by directly assuming a structure for the variance of $y_i^{(\lambda)}$ rather than for that of y_i.

We first note that the power transformations proposed by Box and Cox would not, in general, lead to homoscedasticity of disturbances. This is obvious from the well-known approximation

$$V(y_i^{(\lambda)}) \simeq V(y_i) \left[E(y_i)\right]^{2\lambda-2}$$

which shows that the heteroscedasticity in $y_i^{(\lambda)}$ depends on the transformation parameter λ and also on the heteroscedasticity in the original values of the dependent variable. Even if $V(y_i)$ is constant for all i (= 1, 2, ..., n), $y_i^{(\lambda)}$ would, in general, be heteroscedastic. We consider the following two structures for variance of y_i :

$$\text{(i) } V(y_i) = \sigma^2 \left[E(y_i)\right]^h$$

$$\text{and (ii) } V(y_i) = \sigma^2 m_i^h .$$

The first assumption is the same as that of Zarembka; however, we assume h to be unknown. The second assumption was made by Lahiri and Egy for $V(y_i^{(\lambda)})$

and not for $V(y_i)$. We suggest ML methods of estimation for such models. For the purpose of obtaining the ML estimates we have proposed two methods. The first method involving a search over λ and h is not new. It has been used, for example, by Lahiri and Egy for their model. The other method developed by us is simpler and can be used to obtain the ML estimates by searching over λ only. We have also shown how likelihood ratio (LR) tests can be used to identify different situations in terms of the transformation parameter and the parameter representing the heteroscedasticity in the transformed linear regression model. Among other things, such tests of hypotheses help us in knowing if for a given data, the generalization of the BC model suggested by us is better than the original BC model. We have also illustrated the application of our methods of estimation and testing of hypotheses to two sets of empirical data. For one of the examples we found that the log-likelihood value for our generalized BC model was much higher than that for the original BC model. For the other example, however, this was not so, and apparently the data represented a situation almost satisfying the assumptions made by Box and Cox.

Chapter 3 is concerned with the generalization of the BC model to take account of autocorrelation in the disturbances. This generalization, however, differs from that of Savin and White (1978) who were the first to incorporate autocorrelation in the BC model. Savin and White did this in a very straightforward manner by assuming an AR(1) process for the disturbances (ε_i's) of the transformed linear regression model without taking into consideration the complications created by the transformation itself. However, it can be seen from the approximate relation

$$\mathrm{Cov}(y_i^{(\lambda)}, y_j^{(\lambda)}) \simeq \mathrm{Cov}(y_i, y_j) \{E(y_i)\ E(y_j)\}^{\lambda-1};$$

$$i, j = 1, 2, \ldots, n$$

that the variance-covariance matrix of ε, apart from being dependent on that of y_i's, depends also on λ. Our approach is based on this relation, and may be considered more reasonable and appropriate since it takes into account, in particular, the transformation parameter in addition to the variances and autocovariances of the original observations on the dependent variable in forming the variance-covariance matrix of the transformed dependent variable. Furthermore, our approach can also accommodate heteroscedasticity. In fact, we have introduced heteroscedasticity not only for the transformed dependent variable but also for the original dependent variable by assuming that

$$V(y_i) = \sigma^2 \left[E(y_i)\right]^h, \quad \text{where } \sigma^2 \text{ and } h \text{ are unknown.}$$

It is shown that the Savin and White formulation comes out as a special case of our model and hence our treatment could be considered more general than that of Savin and White.

For estimation of this model, we have suggested the ML method and have indicated how such estimates could be obtained. We have also indicated how LR tests can be used to identify different situations in terms of the transformation parameter and the parameters of autocorrelation and heteroscedasticity including the situation considered by Savin and White. We have further shown that the ML estimate of the transformation parameter and hence of other parameters in the model will not be consistent if the disturbances in the model for the transformed dependent variable actually

have a structure of variances and autocovariances different from that assumed by Savin and White.

In Chapter 4, we have considered nonnormality along with heteroscedasticity of the transformed dependent variable in the Box-Cox type of set-up. It is obvious from the definition of the transformation that for the transformation to be meaningful, y should be > 0. This in turn imposes obvious restrictions on the range of values which can be assumed by $y^{(\lambda)}$: $y^{(\lambda)} > -\frac{1}{\lambda}$ if $\lambda > 0$, but $y^{(\lambda)} < -\frac{1}{\lambda}$ if $\lambda < 0$. It is only when $\lambda = 0$ that $y^{(\lambda)}$ can have the range of a normal variable. Poirier (1978a) considered this problem and suggested maximum likelihood method of estimation assuming the distribution of $y_i^{(\lambda)}$'s $(i = 1, 2, \ldots, n)$ to be truncated normal, the truncation point depending on the value of λ (i.e., whether it is positive or negative). He, however, retained the homoscedasticity assumption for the variance of the disturbance term. We have already argued that apart from the possibility of heteroscedasticity in y_i's, the BC transformation itself leads, in general, to heteroscedasticity in $y_i^{(\lambda)}$'s. We, therefore, extend the Box-Cox procedure by treating the distribution of $y_i^{(\lambda)}$'s to be truncated normal with unequal variances across observations. Such an extension is all the more important since it has already been shown by Draper and Cox (1969) and by Zarembka (1974) that the Box-Cox method of estimation is not robust to non-symmetric distributions — which a truncated normal distribution obviously is — and to heteroscedasticity. We have made two alternative assumptions about the variance of the original observations (i.e., y_i's) :

$$\text{(i)}\quad V(y_i) = \sigma^2 m_i^h$$

$$\text{and (ii)}\quad V(y_i) = \sigma^2 \left[E(y_i)\right]^h$$

where σ^2 and h are unknown for both the assumptions and m_i's are exogenous and known.

It should be noted that when $\lambda=1$, the BC model with a truncated distribution (say, in particular, normal) is nothing but Tobin's limited dependent variable model excepting for unity being added to the constant term and the (known) truncation point being the same for all the observations. Hence by incorporating heteroscedasticity together with truncated normality in the BC model, we have also extended the Tobin model in the direction of incorporating heteroscedasticity in the disturbances. Since Tobin model has considerable interest and applicability, we have first considered the extention of Tobin model (to be referred to as generalized limited dependent variable model (GLDVM)) and then proceeded to the BC model with our generalizations (to be referred to as Box-Cox nonnormal heteroscedastic model (BCNNHM)).

Our main results in this Chapter can be summarized as follows :

(i) Under the assumption that $V(y_i) = \sigma^2 m_i^h$, the ML equation for GLDVM has a strong consistent root. We have also extended Amemiya's method of obtaining an initial consistent estimator for the model so that it can be used to obtain an estimate asymptotically equivalent to the ML estimate at the second stage of the Newton-Raphson iterative method.

(ii) Under the assumption that $V(y_i) = \sigma^2 \left[E(y_i)\right]^h$, we have shown that the ML equation for GLDVM has a strong consistent root. We have than suggested the use of the standard non-linear methods of estimation to obtain the ML estimate of the parameters for this case.

(iii) Under the assumption that $V(y_i) = \sigma^2 \left[E(y_i)\right]^h$, the ML equation for BCNNHM has been shown to have a strong consistent root. Nonlinear methods of estimation have been suggested to obtain the ML estimate of the parameters for this model.

In Part II of this thesis we deal with some problems of statistical inference in an autocorrelated linear regression model. It is recognized in the econometric literature that autocorrelation in a linear regression model is generally caused by misspecification (by which we mean situations where some of the relevant regressors have been left out) and/or by errors-in-observations. Yet the estimation of such a model is often done by assuming a simple autocorrelation structure for the combined error term (representing the total effect of the different factors) like that implied by an autoregressive (AR), a moving average (MA) or an auto-regressive moving average (ARMA) process. Such treatment tacitly assumes that whatever be the source(s) of autocorrelation (i.e., misspecification, errors-in-observations or both), no differential treatment for the structure of the error term is necessary and that the procedures of statistical inference remain the same. This, however, does not seem to be satisfactory. It is desirable to examine whather an observed autocorrelation in

the disturbances is caused by misspecification or by errors-in-observations or by both because, as we shall demonstrate, the appropriate method of estimation depends on the actual situation. It is desirable to have procedures for identifying the actual situation; and also to investigate how one should go about estimating the model in case both the components are present in the model (and each may contribute to the observed autocorrelation) since in such cases the standard reestimation methods like Cochrane-Orcutt (CO), Prais-Winsten (PW) or Durbin's two-step (D2) may fail to give even consistent estimates. Part II of our thesis is devoted to examining the problems which arise when an autocorrelation in a single equation regression model is viewed in this manner and to propose methods which can be used to tackle these problems.

The first Chapter of Part II i.e., Chapter 5, makes a review of the relevant literature. There is a large volume of literature on the problem of autocorrelation. There are also some studies on the effect of misspecification in the standard linear regression model. While a detailed account of the former is available in recent standard textbooks e.g., Judge et al. (1980), a survey of the latter area is made in Chaudhuri (1979, Chapter 1). Hence, what we have done in this Chapter is to touch upon those studies and results which are not readily available in textbooks and/or would help the reader appreciate the motivation of our work.

So far as misspecification is concerned, we have reviewed the studies on the following aspects of the problem :

(i) The effect of misspecification on the disturbances of the misspecified regression equation.

(ii) The effect of misspecification on the ordinary least squares (OLS) estimator and on the Durbin-Watson (DW) statistic.

(iii) The residual variance criterion and Ramsey's (1969) test for detecting the presence of misspecification.

(iv) The effect of misspecification on the standard reestimation methods (also called two-step procedures) frequently used to efficiently estimate the parameters of an autocorrelated linear regression model where the error term follows an AR(1) process.

As regards the literature on autocorrelation, it is mainly concerned with the problem of testing for the presence of autocorrelation in the disturbances irrespective of whether it is caused by misspecification or errors-in-observations or both, and, in case autocorrelation is detected, developing methods of efficiently estimating the parameters of the model. We have touched upon the following aspects of the literature :

(i) Tests for detecting autocorrelation in the disturbances. These include general tests where the error term follows any process under the alternative hypothesis, tests specifically meant for detecting the alternative given by an AR(1) process and tests against non-AR(1) processes alternatives.

(ii) Estimation of the autocorrelated linear regression models when the error process is assumed to follow different processes viz. AR(1), AR(2), AR(q) with $q \geq 3$, MA(1), MA(p) with $p \geq 2$ and ARMA(q, p) where $q \geq 1$ and $p \geq 1$.

In Chapters 6 and 7, we have presented our work on the problems of statistical inference when the error term is regarded as composed of two additive components — one due to misspecification and the other due to other factors like errors-in-observations — each of which may contribute to the autocorrelation in the disturbances[1]. We consider the linear regression model as

$$y_t = x_t' \beta + \varepsilon_t^+ , \qquad t = 1, 2, \ldots, n$$

and $$\varepsilon_t^+ = \varepsilon_t + Z_t$$

where x_t' is a (1 x k) row vector of the t-th observation on k-regressors, β is a (k x 1) column vector of the associated regression coefficients and y_t is the t-th observation on the dependent variable.

We shall first describe our work in Chapter 6. In order to analyse the problem of autocorrelation in such a model, we assume (apart from making the standard assumptions) that both ε_t and Z_t follow AR(1) processes and these two terms are independent of each other and also of the set of regressors. As a general result, it is known that, in general, the sum of two independent AR(1) processes is an ARMA(2, 1) process and not another AR(1) process.[2] Thus it may seem that there is no need of a fresh

---

1/ In terms of a misspecified regression equation, where some regressors are omitted for one reason or other, it can be shown that the error component of the misspecified equation is the sum of two components, one of which represents the error associated with the true regression equation while the other can be ascribed to the effect of misspecification.

2/ We have in our thesis examined the situations under which the resulting process would, in fact, be some AR(1) process.

investigation of the model presented here since standard methods of estimation and testing of hypotheses for regression models with ARMA error processes can be used [e.g., Pierce (1971a, 1971b)]. We, however, argue that it is still important to analyse this kind of models for a number of reasons. For one thing, if it is possible to ascertain whether an autocorrelation is wholly or partly due to misspecification, then one can try to remove the misspecification to the extent possible by trying to include some relevant regressors in the model. What is more important, if there is misspecification in the model (which might give rise to autocorrelation) then it is likely that some standard assumptions will be violated and this will create some problems for the estimation of the model. In fact, in cases where a component due to misspecification is present in the error term, then, even if this component is non-autocorrelated, the standard reestimation methods (like CO, PW, D2) would, in general, yield inconsistent estimates of the regression coefficients. Thus it is important to know whether the error really consists of two components or not, and whether the observed autocorrelation is due to one or both the components. A straightforward use of methods appropriate for an ARMA(2, 1) error process for the combined error term would fail to provide these insights to the problem of autocorrelation because with such an assumption, the information regarding the behaviour of the individual components is not utilized at all. Also, the estimation for an ARMA(2, 1) error process is quite difficult; our methods of estimation and testing being based on OLS residuals are computionally simple.

To analyse the problems in detail, we have listed seven possible cases where the two components of the error term are present or absent, autocorrelated or non-autocorrelated. We have developed a large sample test that would help us identify cases where the disturbances are non-autocorrelated or autocorrelated due to different reasons. By considering a linear transformation of the OLS residuals i.e.,

$$e_t^+ - \hat{\hat{\rho}}\, e_{t-1}^+ = \theta_t \text{ (say)} \quad \left[\text{where } e_t^+ \text{ is the t-th OLS residual and } \hat{\hat{\rho}} = \frac{\sum_{t=3}^{n} e_t^+ e_{t-2}^+}{\sum_{t=2}^{n} e_t^+ e_{t-1}^+}\right]$$

and defining

$$c_s = \text{Cov}(\theta_t, \theta_{t-s}),$$

we have found that c_s's can be used to discriminate among three broad groups of the seven possible cases. Thus, for example, if the autocorrelation in the disturbances is caused by both the components, then $c_s \neq 0$ for all $s \geq 1$. In situations where the autocorrelation is due to one component only while the other is present but nonautocorrelated, $c_1 \neq 0$ and $c_s = 0$ for all $s \geq 2$. These groups of cases can, therefore, be identified by (say) using Bartlett's (1946) well-known test [vide Box and Jenkins (1976)]. By using DW test we can further divide one of the three broad groups into two sub-groups each consisting of two cases. Thus, the tests enable us to identify the different cases characterized by different combinations of the parameters involved. Although three of these groups/sub-groups contain two cases each, no problem seems to exist for the estimation of the model and for inferences regarding the presence of

misspecification. For those cases where neither the OLS nor the standard reestimation methods can be used, we have suggested the use of the estimated generalized least squares (EGLS) method; the estimated variance-covariance matrix in these cases may be based on OLS residuals only. Under standard assumptions [vide Theil (1971)] these estimators are consistent and asymptotically efficient.

In Chapter 7, we have considered the same problem as in the previous chapter but under a more general assumption regarding the error process. Here we have assumed that both the components ε_t and Z_t follow ARMA(1, 1) process. As the sum of the two independent ARMA(1, 1) processes is an ARMA(p, q) process where $p \leq 2$ and $q \leq 2$, a straightforward use of an ARMA error process with exactly known order cannot be made for the purpose of estimation of this model. This provides an additional argument (apart from the ones given in the previous chapter) for undertaking the present analysis. However, when we assume the ARMA(1, 1) error process for each of the components, the testing and estimation procedures developed in the last Chapter are no longer applicable. For estimation of the model considered in this Chapter, we suggest the use of ML method of estimation which under standard assumptions gives asymptotically efficient estimates. Since the normal equations turn out to be nonlinear, we have advocated the use of Newton-Raphson iterative method for which it is known that the second-stage estimator is asymptotically equivalent to the ML estimator if the initial or first-stage estimator is a consistent one (cf. Rothenberg and Leenders (1964)). We have provided a consistent initial estimator based on OLS residuals. We have also suggested LR tests for identifying

different types of autocorrelated situations characterized by different combinations of values of the parameters.

It may, however, be noted that there is a problem of identification of the parameters involved in the variance-covariance matrix. To overcome this, we have suggested reparametrization of the model. After reparametrization, this identification problem does not pose any difficulty so far as the estimation of the regression coefficients are concerned. But this lack of identifiability poses some problems in the testing of different hypotheses representing different types of autocorrelated situations since all the relevant hypotheses formulated in terms of the original parameters cannot be equivalently stated in terms of the redefined parameters. Thus, we can test hypotheses like ARMA(1,1)– MA(1) (i.e., where one of the components follows an ARMA(1, 1) process and the other an MA(1) process) etc., against the alternative hypothesis ARMA(1, 1) – ARMA(1, 1). However, if the ARMA(1, 1) – MA(1) hypothesis is rejected in favour of ARMA(1,1)– ARMA(1, 1), we are unable to test ARMA(1, 1) – AR(1) against ARMA(1, 1) – ARMA(1, 1) since the additional constraint imposed by the null hypothesis (i.e., the autocorrelation situation represented by ARMA(1,1) – AR(1)) as against the alternative hypothesis (i.e., ARMA(1, 1) – ARMA(1, 1)) cannot be specified in terms of the redefined parameters. To overcome this, we may test hypotheses that can be treated as cases under ARMA(1,1)—AR(1) e.g., AR(1) – AR(1), MA(1) – AR(1) etc., against ARMA(1, 1) – AR(1)(or suitable alternatives) If, however, all these are rejected in favour of ARMA(1,1) – AR(1)), we are unable to conclude if the actual situation is ARMA(1, 1) – ARMA(1, 1) or ARMA(1, 1) – AR(1).

In Part III we continue our investigation of the autocorrelated linear regression model. But here our emphasis is on the small-sample performance of some standard methods of estimation (frequently used under the assumption that the error follows an AR(1) process) in situations where misspecification (which may or may not contribute to autocorrelation in the errors) is explicitly introduced in the model. In the conventional treatment of an autocorrelated linear regression model, the error term is considered as a single component and standard reestimation methods (i.e., CO, PW, D2 and Hildreth-Lu (HL) search procedure) are used if autocorrelation in the error term is found to be significant. No account is taken of the fact that in a misspecified regression model there will, in fact, be two components of the error term. For reasons stated in the previous Part, we consider a model where the error term is decomposed into two independent additive components — one representing the error in the true regression equation and the other arising due to misspecification. Any of the two components, if autocorrelated, is assumed to follow AR(1) process. As already noted, in such a case the composite error would not, in general, follow an AR(1) process and the standard reestimation methods like those of CO, PW and D2 methods would give inconsistent and inefficient estimates. This suggests that one should not blindly apply these methods without trying to ascertain if there has been any misspecification and if so, whether the misspecification has been one of the sources of autocorrelation in the error term. The above-mentioned results, however, refer to the large-sample case. Our interest in this part is in the small sample properties of OLS and the standard reestimation methods viz., CO, PW, D2

and HL. Since these properties cannot be analytically obtained, we carried out a Monte Carlo experiment to examine these properties, confining ourselves to studying the comparative performance of the different methods. It is hoped that the results would throw light on the price one has to pay in terms of efficiency in case these methods are blindly used disregarding possible misspecification of the true model. We have considered two specific situations :

(i) only one of the two components follows AR(1) while the other is present but non-autocorrelated,

and (ii) both the components follow AR(1) error processes.

The four Chapters in Part III are concerned with (i) the motivation of this Monte Carlo experiment, (ii) a review of the connected literature, (iii) an examination of the unbiasedness and consistency properties of some of the standard estimators, (iv) the plan of the experiment and (v) the results of the experiment.

Besides explaining the motivation of the Monte Carlo study, we have also surveyed in Chapter 8 the literature on Monte Carlo studies relating to autocorrelated linear regression models. It may be noted that there has been practically no Monte Carlo study on linear regression models where the error term follows processes other than AR(1). We have also referred to some other important Monte Carlo studies which are not directly related to ours but which helped in formulating and designing our study.

Since the properties (even in large samples) of the estimators given by the standard reestimation methods are not so well-known for

situations marked by the presence of misspecification (which may or may not contribute to the observed autocorrelation), we present in Chapter 9 some results on the unbiasedness and consistency of some of these estimators. In this context we have also briefly described the standard reestimation methods viz., CO, PW, D2 and HL. Thus, this Chapter should help in studying the results of the Monte Carlo study vis-a-vis the known properties of the estimators. The properties brought out in Chapter 9 and those revealed in Chapter 11, taken together, provide a fairly complete picture of the relative performance of the different methods in presence of misspecification.

Some of the important results presented in Chapter 9 are the following :

(a) OLS, in general, gives biased but consistent estimates of the regression coefficients of the misspecified regression model.

(b) Generalized least squares (GLS) and estimated generalized least squares (EGLS), in general, yield biased and inconsistent estimates of the regression coefficients of the misspecified model.

This result may appear surprising. However, it may be noted that this is a consequence of the definition of β^+ (i.e., the vector of regression coefficients in the misspecified equation) used by Chaudhuri (1977, 1979). Chaudhuri asserted that β^+ cannot, in general, be the sub-vector corresponding to the set of included regressors of β (i.e., the vector of regression coefficients in the true model) and she redefined β^+ in a particular way in order to enable the included regressors to capture as much of the influence of the excluded regressors as possible. However,

one can also redefine β^+ in a way different from that of Chaudhuri and possibly get the opposite result. It seems further research is needed in deciding the proper definition of β^+. In any case, this does not affect our present study since we need only to fix a numerical value of β^+ without bothering about its proper definition.3/

(c) In the special case where the excluded regressors are linearly related to the included ones, the GLS and the EGLS (in the latter case it is also required, apart from the standard conditions, that the variance-covariance matrix is consistently estimated) consistently estimate the regression coefficients of the misspecified regression equation.

(d) Since all the two-step methods mentioned earlier are either GLS or EGLS, the properties of these estimators will be similar to those stated above.

(e) The ML method of estimation can be used only in the special case stated in (c) above and in that case it is as good asymptotically as the reestimation methods under standard conditions (vide Theil (1971)).

In Chapter 10, we describe the scope and design of the Monte Carlo experiment carried out by us. The model used for the purpose of the experiment is

$$y_t = \beta^+ x_{1t} + \varepsilon_t^+ , \qquad t = 1, 2, \ldots\ldots, n$$

$$\varepsilon_t^+ = \varepsilon_t + Z_t .$$

We have considered two different situations each having autocorrelation in the ε_t^+ series :

3/ See, in this connection, the paper by Gupta and Maasoumi (1979).

Situation I : both Z_t and x_{1t} follow AR(1) processes, but ε_t is random,

Situation II : both ε_t and Z_t follow AR(1) processes, but x_{1t} is random.

In the second situation x_1 was taken to be random only to keep the volume of computations manageable and also because the autocorrelation coefficient of the x_1-series was not found to have appreciable effects on the relative performances of the different methods in Situation I.

We examined the relative performances of the OLS, CO, PW, D2 and HL estimators. The set of independent parameters affecting the results, were idenfified by examining the possible interdependence among some of the parameters involved in the model and then a number of values of these parameters were chosen so as to ensure sufficient variation in their values within permissible range. Four such parameters were identified and three values were considered for each of them. There were thus a total of 81 parametric combinations. Random normal deviates were generated by approximating the normal density by a discrete distribution using Hastings approximation. Several criteria were used to assess the relative performances of the different methods. These are : absolute deviation of $\hat{\beta}^+$ ($\hat{\beta}^+$ is the estimate of β^+ by any of the five methods), bias, absolute bias and mean square error of $\hat{\beta}^+$. We also examined the performances of confidence intervals set up by conventional procedures. Since in the case of stochastic regressors, the sampling variances given by conventional formulae may underestimate, correctly estimate or overestimate the actual sampling variance of the estimators, we also examined the usefulness of

these formulae in estimating the true sampling variance.[4/]

The results of our experiment are presented and studied in Chapter 11. Some of the important results may be summarized as follows :

(a) For low values of autocorelation in Z_t, OLS is as good as the others, in general, for Situation I. For Situation II, the results are similar only when both the components ε_t and Z_t have very low values of autocorrelation.

(b) In terms of absolute deviation, mse and ranking by these criteria PW comes out to be the best among the methods examined here for both the situations. For Situation I, CO and D2 are about equally good and they are inferior only to PW. (The CO appears to be slightly better than D2 in come cases.) HL comes next in the ranking, but is clearly inferior to CO and D2. OLS is somewhat erratic, being nearly as good as the others for low values of the autocorrelation in Z_t, but quite bad (and very offen the worst) for other cases. As for Situation II, CO, D2 and HL are very close to one another, and all these come behind PW in the ranking although the superiority of PW vis-a-vis these methods are somewhat reduced in this situation compared to the previous one. In respect of bias also, the picture for both the situations is more or less the same.

(c) Confidence intervals do not throw much light on the relative performances of the different methods for both the situations. However, for high values of the autocorrelation coefficients of Z_t and x_{1t} in

---

4/ It may be mentioned that we used samples of size 15 only; even this meant an enormous volume of computations since there were 81 cases for each of the two situations and 50 such samples were used for computing the estimators for each of the 81 cases i.e., combinations of parameter values.

in Situation I, the confidence intervals for all the methods fail to include the true value of β^+ in a much larger proportion of samples than is specified by the confidence coefficient. For Situation II, this is not so — all the methods seem to perform fairly well for this situation.

(d) For Situation I, the conventional sampling variance formulae associated with the different methods seem to significantly underestimate the true sampling variances of the estimators. The extent of underestimation, on the average, ranges between 17 and 20 per cent for all the methods excepting for HL for which this downward bias seems to be about 25 per cent. For Situation II, the results are quite different. OLS now overestimates by about 8 per cent on the average which, however, may not be statistically significant. For all other methods there seems to be some underestimation but the extent of underestimation is much smaller (between 6 and 9 per cent) than for Situation I.

PART I

ON THE BOX-COX TRANSFORMATION IN SINGLE EQUATION MODELS

Chapter 1

Review of the Literature

1.1 Introduction

In regression analysis one makes the assumption that the expected value of the dependent variable (y) is a linear function of the independent variables. But such an assumption may be untenable in many situations. It could very well be that the underlying relation between the variables is nonlinear but a function of y is linearly related to the independent variables or to some functions of them. One frequently uses different specific nonlinear functional forms like the semilog, doublelog quadratic etc. But any a priori decision about the functional form is not very satisfactory and yet the choice of the form is quite often not based on any objective criterion. Usually a priori choices are dictated by convenience, experience from previous researches and the perception of the researcher regarding the data apart from the suggestions from theory. But in many cases theory does not give any indication regarding the functional form to be chosen. There is, therefore, a general problem of choosing the appropriate functional form for a given set of data. This has been very aptly stated by Kendall and Stuart (1968) in the following lines :

> "Although natural considerations of convenience or technique may dictate that the observations be made on a variable y, it still has to be decided which function of y is to be used for the purpose of the analysis". 1/

1/ Kendall and Stuart (1968), p. 85.

The objective of transforming the variables is precisely to deal with such problems. One seeks to determine the particular transformations of the variables which would ensure linearity of the relationship and also satisfy certain other standard conditions. Use of transformation of variables in econometrics can thus be considered a very powerful and convenient tool for handling the general problem of choosing the functional form of a relationship on the basis of the given data, particularly when no idea about the nature of the relationship can be had from theory or from previous investigations.

The arrangement of the Chapter is as follows. Early work on transformation of variables is briefly described in section 1.2. The Box-Cox (BC) transformation of variables and the problem of heteroscedasticity in it are discussed in sections 1.3 and 1.4 respectively. The problem of nonnormality and the Box-Cox transformation is described in section 1.5. The connection between the BC model and the limited dependent variable model is indicated in section 1.6. Section 1.7 presents a brief summary of the extensions of the BC model. Some concluding observations are given in section 1.8.

1.2 Early Work on Transformation of Variables

Tukey (1957) was perhaps the first person to make a systematic study of the problem of transformation of variables. According to him, the purpose of transformation is to achieve the following three desirable properties :

(i) additivity of the effects,

(ii) constant variability of the error term,

(iii) symmetry and perhaps normality of the distribution of the error term.

He considered a family of transformations $(x+d)^{\lambda}$ defined as $\ln(x+d)$ for $\lambda = 0$. If $d = 0$, the transformation reduces to x^{λ} for $\lambda \neq 0$ and $\ln x$ for $\lambda = 0$. He further showed that if $\lambda = md$ $(m < 0)$ the transformation approaches e^{mx} ad $d \longrightarrow \infty$. Tukey also discussed the properties of the transformation but assumed that the transforming parameter (λ) is either known or can be determined relatively easily.

Subsequently, Turner et al. (1961) and Box and Tidwell (1962) considered the problem of estimation of the transformation parameter but they confined themselves only to transformations of the independent variables in the regression model. They suggested an iterative least squares procedure for estimating λ_i's in the transformation of independent variables to $(x_i + d_i)^{\lambda_i}$ where x_i is the ith $(i = 1, 2, \ldots, k)$ independent variable in the regression equation. For $d_i = 0$ $(i=1,2,\ldots,k)$ their method of estimating the λ_i's may be described in the following steps :

(i) Choose initial values of λ_i's, say, $\lambda_{0i} (i = 1, 2, \ldots, k)$.

(ii) Regress the dependent variable on the regressors $x_i^{\lambda_{0i}}$ and $x_i^{\lambda_{0i}} \ln x_i$ $(i = 1, 2, \ldots, k)$.

(iii) Define the new estimate of each λ_i as $\lambda_{1i} = \lambda_{0i} + b_{0i} / a_{0i}$ where a_{0i} and b_{0i} are the estimates of regression coefficients associated with $x_i^{\lambda_{0i}}$ and $x_i^{\lambda_{0i}} \ln x_i$ respectively.

(iv) Repeat the process with the new estimates of λ_i's till convergence.

Box and Tidwell provided some examples to indicate that the iterative process suggested above converges rapidly. For the case of a single

independent variable the confidence intervals for the parameters of the model were also found out.

1.3 The Box-Cox (BC) Transformation of Variables

All the above mentioned studies focussed attention on the transformation of the independent variables. Transformation of the dependent variable in a regression model is perhaps more interesting, particularly because it gives rise to some special problems. Box and Cox (1964) were the first to systematically study the transformation of the dependent variable. In their well-known paper they proposed a family of power transformations of the dependent variable in a regression model so as to achieve the desirable properties viz., (i) linearity, (ii) homoscedasticity, and (iii) normality of the error distribution, in the transformed regression equation. They defined the transformation as

$$y^{(\lambda)} = \begin{cases} (y^{\lambda} - 1)/\lambda & \text{if } \lambda \neq 0 \\ \ln y & \text{if } \lambda = 0. \end{cases} \quad \ldots (1.3.1)$$

Obviously, the transformation is continuous around $\lambda = 0$. The Box-Cox regression model can then be written in matrix notation as[2/]

$$Y^{(\lambda)} = X\beta + \varepsilon \quad \ldots (1.3.2)$$

where $Y^{(\lambda)} = (y_1^{(\lambda)}, \ldots, y_n^{(\lambda)})'$ is the (n x 1) vector of transformed observations on the dependent variable y defined in (1.3.1), X is a (nxk)

2/ Obviously, independent variables can also be transformed. But since it does not give rise to any additional complication and the maximum likelihood technique (described later in this context) can be used in a way identical to that for the Box-Cox transformation on the dependent variable, we restrict the discussion to the transformation of the dependent variable only.

matrix of observations on the k independent variables including a column of unity for the constant term[3], β is the (k x 1) vector of associated regression coefficients and $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)'$ is a (n x 1) vector of random disturbances which are distributed independently and identically with mean zero and constant variance σ^2.

For the estimation of the parameters of the model viz., λ, β and σ^2, Box and Cox suggested the use of the maximum likelihood (ML) method of estimation. Under the assumption of normality, the log-likelihood function of the original observations $(y_1, \ldots, y_n)$ can be written as

$$L = \text{Const.} - \frac{n}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} (Y^{(\lambda)} - X\beta)' (Y^{(\lambda)} - X\beta) + \ln J \qquad \ldots(1.3.3)$$

where

$$J = \det \left| \frac{\partial y_i^{(\lambda)}}{\partial y_i} \right| = \prod_{i=1}^{n} y_i^{\lambda-1} .$$

The log-likelihood function in (1.3.3) can be maximized directly by using any of the standard nonlinear methods of estimation. However, since such methods are computationally complicated, they suggested maximizing the log-likelihood function separately for each of several values of λ covering a reasonable range and choosing that transformation (value of λ) for which the maximum log-likelihood value is the highest. It can be seen easily that this is equivalent to finding the ordinary least squares (OLS) estimates $\hat{\beta}(\lambda) = (X'X)^{-1} X'Y^{(\lambda)}$ of β for specified values of λ from the transformed model and then choosing

3/ As has been pointed out by Schlesselman (1971), it should be noted that unless a constant term is included in the regression equation the estimation of the transformed model is not invariant to changes in the unit of measurement of y.

that value of λ and the corresponding estimates of β and σ^2 as the required estimates, for which $-\frac{n}{2}\ln\hat{\sigma}^2(\lambda)$,

where $\hat{\sigma}^2(\lambda) = \frac{1}{n}\left[(Y^{(\lambda)} - X\hat{\beta})'(Y^{(\lambda)} - X\hat{\beta})\right]$

is maximum[4/]. Box and Cox also indicated how one can obtain the confidence interval for λ.

Now, so far as the properties of the estimators obtained by the ML method described above are concerned, it is clear that whichever of the two procedures — direct maximization of the log-likelihood function or maximization by searching over λ — is adopted, the estimate of λ and hence those of β and σ^2 will, under standard conditions[5/], be the ML estimates and therefore consistent and efficient. However, the estimates of β's obtained by the search procedure for different λ's are not comparable since these are conditional upon λ's though the estimated elasticities at the sample means are. Furthermore, as Spitzer (1977b) pointed out, one should be careful in using the standard errors of the estimated coefficients which are conditional on λ. These standard errors will underestimate the more relevant unconditional ones. The unconditional standard errors can be obtained by considering the Cramer-Rao lower bound. They can also be obtained by using estimates found by directly maximizing the likelihood function over all the parameters by an appropriately chosen technique of maximizing nonlinear

---

4/ By a suitable choice of the unit of measurement of y_i's defined as $y_i / \left(\prod_{i=1}^{n} y_i\right)^{\frac{1}{n}}$, we can make $\sum_{i=1}^{n} \ln y_i = 0$ and hence $\ln J$ vanishes.

5/ For the standard conditions and other related discussions, see, for example, Wald (1949), Dhrymes (1970, Chapter 3) and Rao (1974, Chapter 5).

functions. The necessary information matrix may be directly evaluated in the latter-mentioned procedure.

The above comments refer to the large sample properties of the estimators. Not much is known about the small sample properties of these estimators. Spitzer (1978) undertook a Monte Carlo study to investigate the small-sample properties of the above-noted estimators of the Box-Cox model. He considered the following model

$$y_i^{(\lambda)} = \alpha + \beta x_{1i}^{(\lambda)} + \gamma x_{2i}^{(\lambda)} + \varepsilon_i, \qquad i = 1, 2, \ldots, n$$

where the dependent and the two independent variables were transformed using the same transformation parameter on all the variables. The values of α, β and γ were fixed at 9.0, -1.5 and 0.5 respectively. Five models corresponding to five different values of λ viz., -1.5, -1.0, -0.15, 1.0 and 1.5 were estimated 50 times for each of two sample sizes viz., $n = 30$ and $n = 60$. ε_i's were assumed to follow independent $N(0, \sigma^2)$ and the value of σ^2 was fixed at 0.426. Their conclusions can be summarized as follows:

(i) In none of the five models bias of the estimators of parameters seemed to be a serious problem. No regularity in over- or under-estimation was observed in any of the models.

(ii) The sampling distributions of the coefficient estimates appeared to be approximately normal in almost all the cases. However, actual sampling distributions appeared to have heavier tail areas than the t-distribution and hence one should be careful in using the t-statistic for hypothesis testing.

(iii) The performance of the models in forecasting was encouraging. The forecasts seemed to be unbiased with small variance and relative errors of forecast.

(iv) The coefficient of variation of y^{λ} seemed to have a relationship with the precision of the estimators of the parameters.

(v) The estimated parameters appeared to be efficient on the basis of the sample asymptotic standard deviations.

1.4 The Problem of Heteroscedasticity in the Box-Cox Transformation

Most of the subsequent studies on the Box-Cox transformation sought to examine more carefully if it is indeed possible to have a transformation which will achieve all the three desirable properties viz., linearity, homoscedasticity and normality. In other words, the question is : Does the λ that linearizes the relationship also lead to normality and homoscedasticity of the errors of the linear regression model with transformed dependent variable ?

Zarembka (1974) considered at length the problem of heteroscedasticity in the context of the Box-Cox transformation. He pointed out that[6] "the transformation on the dependent variable may not lead simultaneously both to additivity of the effect and to homoscedasticity" He also examined to what extent the Box-Cox method of estimation is sensitive to the assumption about variance and proved that the procedure is not robust to (i.e., is vitiated by the presence of) heteroscedasticity.

The bias in estimating λ by the Box-Cox method when errors are, in fact, heteroscedastic has also been obtained by Zarembka in the paper mentioned above. It was shown that there is a tendency of the

---

6/ Zarembka (1974), p. 87.

estimated $\hat{\lambda}$ to approach that transformation of the dependent variable which leads to stabilization of the error variance. In other words, λ and hence other parameters of the transformed model are consistently estimated by the Box-Cox procedure only when the transformation that leads to linearity also leads to homoscedasticity in the error variance.

Zarembka then proposed a method of estimation of the transformed model under the assumption of heteroscedasticity. He assumed the variance of y_i to be given by

$$V(y_i) = \sigma^2 \left[E(y_i)\right]^h , \qquad E(y_i) > 0$$

where σ^2 is an unknown parameter and h is a known constant and developed his method of estimation by using the likelihood equation of the Box-Cox method and the information on the magnitude of bias in estimating λ by the Box-Cox procedure. His method essentially consists in evaluating the expression $E\left(\frac{\partial L}{\partial \lambda}\right)$ for different values of λ (L is as defined in (1.3.3)) when, in fact, the $y_i^{(\lambda)}$'s are heteroscedastic and then taking that value of λ as the estimate of λ where $E\left(\frac{\partial L}{\partial \lambda}\right)$ equals the slope of the maximum likelihood function conditional upon λ.

While Zarembka's paper represents a significant and interesting attempt to tackle the problem of heteroscedasticity in the context of the Box-Cox transformation model, it still has a number of limitations. Firstly, the method of estimation proposed by him presupposes the degree of heteroscedasticity (h) to be known. Secondly, in developing the method he uses a number of assumptions like e.g., $E\left[\ln(y_i)\right] \simeq \ln E(y_i)$, $E\left[\ln(y_i)\right]$ is reasonably symmetric etc., which may not be valid in many situations. Thirdly, consistent estimates of the standard errors of $\hat{\beta}$

can not be obtained from the expression $\hat{\sigma}^2 (X'X)^{-1}$ (where $\hat{\sigma}^2$ is the ML estimator of σ^2 from BC model) as long as ε is heteroscedastic. In fact, there is no convenient formula to calculate the standard errors of $\hat{\beta}$. Finally, and perhaps most importantly, Zarembka's method yields only approximately consistent but not efficient estimates of the parameters. As Zarembka himself has remarked, "an asymptotically efficient procedure for estimating λ and β under heteroscedasticity needs to be developed".7/

Recently, Egy and Lahiri (1978) considered the Box-Cox model under conditions of heteroscedasticity. Assuming that the variance of the disturbances is of the form $\sigma_i^2 \left(= V\left[y_i^{(\lambda)}\right]\right) = \sigma^2 m_i^{\delta}$, where m_i's are exogeneously given and δ, σ^2 are unknown parameters, they proposed ML estimation of the parameters λ, δ, β and σ^2. They suggested a search procedure over (λ, δ) for obtaining the ML estimates of these and hence of the parameters β and σ^2. They also suggested likelihood ratio (LR) tests for testing specific hypotheses regarding λ and δ. The main criticism against such an assumption about $V\left[y_i^{(\lambda)}\right]$ is that apart from assuming m_i's to be exogenously given, it also ignores the complications created by the transformation.

They further claimed to have shown that joint estimation of λ and δ may not be possible when the model is formulated as : $\sigma_i^2 = \sigma^2 \left[E(y_i)\right]^{\delta}$, because in this case a λ-transformation may stabilize the variance completely leading to a near identification problem. We may point out that, even with the questionable approximation used by them, the validity of this conclusion is not beyond doubt (see also, footnote 2 of the next chapter).

7/ Zarembka (1974), p. 102.

1.5 The Problem of Nonnormality and the Box-Cox Transformation

The assumption of normality of the distribution of the error term in the transformed model has also been examined by some researchers. It is obvious that the Box-Cox transformation does not, in general, permit negative values of y_i's and hence in treating the error distribution to be normal one is implicitly assuming that the probability of negative, particularly large negative, values is quite low so that the condition of normality is not seriously affected. In fact, Draper and Cox (1969) examined the robustness of the Box-Cox procedure with respect to nonnormality of the disturbances and showed that so long as the errors (i.e., ε_i's) have reasonably symmetric distributions, the Box-Cox estimation procedure is robust to nonnormality. Thus, the Box-Cox method leads to approximately consistent estimate of λ as long as ε_i is reasonably symmetric.

However, since the condition of normality does not, strictly speaking, hold, it is desirable to examine if things could be improved with more appropriate assumptions about the distribution of the error. It is clear that for the transformation to be well defined, it must be true that $y > 0$ which in turn implies that $y^{(\lambda)} > -\frac{1}{\lambda}$, if $\lambda > 0$, and $y^{(\lambda)} < -\frac{1}{\lambda}$, if $\lambda < 0$. Thus, for non-zero values of λ, the range of $y^{(\lambda)}$ is a proper subset of the real line and hence, strictly speaking, the error term cannot follow a normal distribution. It is only when $\lambda = 0$ that the assumption of normality may be valid. In view of this, recently Poirier (1978a) considered a truncated normal distribution of

the error term in the Box-Cox model, the point of truncation being appropriately defined according as λ is positive or not. However, like Box and Cox, he assumed homoscedasticity of the transformed variable and used standard nonlinear methods to obtain the ML estimates of λ, β and σ^2.

1.6 The Box-Cox Transformation and the Limited Dependent Variable Model

It is interesting to note that for $\lambda = 1$, the Box-Cox model as considered by Poirier (1978a) reduces to the well-known limited dependent variable (LDV) model developed by Tobin (1958) (except for one being added to the constant term)[8]. Although Tobin's model is an important contribution having useful economic applications [see, for instance, the work of Jragg (1971) on demand models

---

[8] This reduced model is slightly different from Tobin's LDV model. Whereas Tobin's model is

$$y_i = \begin{cases} x_i' \beta + \varepsilon_i & \text{if R. H. S.} > \alpha_i \\ 0 & \text{otherwise} \end{cases}$$

(α_i's are known constants and x_i' is the ith row of X)

we have in the reduced model $\alpha_i = 0$ for all i. It may, however, be noted that with a redefinition of y_i, x_i' and β, Tobin's model can easily be seen to be equivalent to our reduced model. The redefined Tobin's model can be written as

$$y_i^* = \begin{cases} x_i^{*\prime} \beta^* + \varepsilon_i & \text{if R. H. S.} > 0 \\ 0 & \text{otherwise} \end{cases}$$

where $y_i^* = y_i - \alpha_i$, $x_i^{*\prime} = (x_i', \alpha_i)$ and $\beta^{*\prime} = (\beta', -1)$.

for durable goods 7, it did not receive sufficient attention from the econometricians for quite some time. A number of years later Amemiya (1973a) considered the model and proved a number of important results for this model. He showed, for example, that for this model there exists a strong consistent root of the normal equations and this root is asymptotically normal. He also proved the inconsistency of the initial estimator proposed by Tobin for use in Newton-Raphson iterative procedure for obtaining the ML estimates, suggested consistent initial estimators of the parameters and proved that the second-stage estimator obtained by Newton-Raphson method will have the same asymptotic distribution as a consistent root of the normal equations if the first-stage estimator is consistent. More recently, Oslen (1978) has shown that the likelihood function for Tobin model has a unique maximum and hence whatever be the initial estimator, not necessarily the one suggested by Amemiya (1973a), if the iterative procedure converges, then the converging values will be the ML estimates corresponding to the global maximum of the likelihood function.

1.7 Other Extensions of the Box-Cox Transformation Model

Attempts have recently been made to extend the Box-Cox model in new directions. Savin and White (1978) extended the model by introducing autocorrelation in the disturbance term of the transformed model. They assumed the disturbances to follow a first-order autoregressive process (AR(1)) and suggested that ML estimators of λ and the autocorrelation coefficient ρ and hence of β and σ^2 may be obtained by a simultaneous

search for the maximum value of the likelihood function over (λ, ρ). They also indicated that LR tests may be used to test various hypotheses regarding λ and ρ. It may be noted, however, that their straightforward extension amounts to ignoring the complications created by the transformation itself. Also, the assumption of AR(1) error process, or for that matter any other standard simple process, implies that the error variance is treated to be homoscedastic which, as will be seen in the next chapter is not always true.

The use of the Box-Cox transformation in a simultaneous equations system has also been made in the recent past. Tintner and Kadekodi (1971) for example, considered a simultaneous equations system where all variables in the system were transformed by the same λ-transformation. More recently, Spitzer (1977a) incorporated the transformation in a simultaneous money demand and supply equations system.

1.8 Conclusions

We may conclude by reiterating that the Box-Cox transformation has been very useful from the point of view of choice of functional form of an economic relationship. A number of applications of this has already been made in empirical studies. As specific examples, we may refer to the studies made by Zarembka (1968) who used the transformation in estimating a relationship among demand for money, current rate of interest and current income, and to those of Mukerji (1963), Dhrymes and Kurz (1964), Ramsey and Zarembka (1971) etc., in estimating production functions. Other notable references are White (1972), Kau and Lee (1976), Spitzer (1976), Chang (1977) etc.

In course of our review we have hinted at some of the inadequacies of the available methodology in this area. We think that there exists considerable scope for further critical examinations and extensions of the model considered by Box and Cox and later researchers. Such improvements are desirable not only from theoretical point of view but also on considerations of applicability in various econometric problems. We attempt to do so in the next three chapters.

Chapter 2

Heteroscedasticity and the Box-Cox Transformation

2.1 Introduction

It has been mentioned in the last Chapter that Box and Cox (1964) defined a family of power transformations of the dependent variable in a a regression model which helps in ensuring a linear relationship between the transformed dependent variable on the one hand, and the set of regressor (fixed) on the other. The approach assumes homoscedasticity and normality of the transformed dependent variable. The transformations considered by them are

$$y^{(\lambda)} = \begin{cases} (y^{\lambda} - 1)/\lambda & \text{if } \lambda \neq 0 \\ \ln y & \text{if } \lambda = 0 \end{cases} \qquad \ldots \quad (2.1.1)$$

where y is the original dependent variable in a regression model[1/]. The central point of the Box-Cox transformation is that there exists a λ for which all the three desirable properties stated above are simultaneously satisfied. The linear regression model with transformed dependent variable (to be mentioned henceforth as the transformed linear regression model) can be described, in matrix notation, as

$$Y^{(\lambda)} = X\beta + \varepsilon \qquad \ldots \quad (2.1.2)$$

where $Y^{(\lambda)} = (y_1^{(\lambda)}, y_2^{(\lambda)}, \ldots, y_n^{(\lambda)})'$ is a $(n \times 1)$ vector of observations on the transformed dependent variable defined according to (2.1.1),

1/ The Box-Cox transformation can also be extended to the independent variables. However, since such transformations do not give rise to any additional problems, we confine ourselves to the case where only the dependent variable is transformed.

X is a (n x k) matrix of observations on k regressors, $\beta = (\beta_1, \ldots, \beta_k)'$ is the (k x 1) vector of associated regression coefficients and $\varepsilon = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)'$ is the (n x 1) vector of disturbances in the transformed linear regression model. The rank of X is assumed to be $k(<n)$. Box and Cox assumed that ε_i's (i = 1, 2, ..., n) have independent normal distributionswith zero mean and same variance σ^2.

The Box-Cox method of estimation of λ and hence of β and σ^2 of the regression model described above is clearly based on the assumption of homoscedasticity of the errors. However, as will be explained in the following section, the transformation would, in general, lead to heteroscedasticity in the dependent variable and hence an estimation procedure depending on the assumption of homoscedasticity is not always tenable. Actually, Zarembka (1974) has shown that the Box-Cox method of estimation is not robust to heteroscedasticity and that estimate of λ is biased towards that power of the dependent variable which leads to a stabilization of the error variance. There is thus a need for developing estimation procedures for a model that would accommodate heteroscedasticity of the dependent variable resulting from the Box-Cox transformation in addition to the heteroscedasticity (if any) in the original dependent variable. Such a model can then enable us to separate out the influence of nonlinearity in an estimate of λ from the influence of stabilizing the error variance. The method of estimation proposed by Zarembka (1974) assumes that the variance of y_i is given by

$$V(y_i) = \sigma^2 \left[E(y_i)\right]^h, \qquad E(y_i) > 0 \qquad \ldots (2.1.3)$$

where σ^2 is an unknown parameter, but h is known. His method is based

on the ML equation of the Box-Cox estimation method. There are several problems with this method : First, the degree of heteroscedasticity (h) must be known; and second, as noted by Zarembka himself, we can get only approximately consistent but not efficient estimators of the parameters. Egy and Lahiri (1978), on the other hand, suggest a model where the pattern of heteroscedasticity is gigen by

$$V\left[y_i^{(\lambda)}\right] \; (= V(\varepsilon_i)) = \sigma^2 \, m_i^{\delta} \qquad \ldots \quad (2.1.4)$$

where m_i's are exogenously given and δ and σ^2 are unknown parameters. They suggest a search procedure over λ and δ for obtaining ML estimates of λ and δ and hence of β and σ^2. They also claim to have shown that Zarembka's assumption mentioned in (2.1.3) above — which leads approximately to the relation

$$V(\varepsilon_i) = \sigma^2 \left[E(y_i)\right]^{\delta}, \quad E(y_i) > 0,$$

δ is, for this case, an unknown parameter involving λ and h — leads to stabilization of error variance and hence to an identification problem[2/]. As will become clear from our analysis, this conclusion does not seem to be valid. The Egy-Lahiri method can be applied only when m_i's are given. Note also that in their formulation $V(y_i^{(\lambda)})$ does not involve λ. But that such straightforward assumption about $V(y_i^{(\lambda)})$ without taking into consideration the complications created by the transformation is not very satisfactory, has also been noted by Zarembka (1974). He points out that "the problem is complicated by

---

2/ It may be noted that in the published version of this paper (1981), they have omitted this particular claim.

the fact that the heteroscedasticity in ε_i depends upon the heteroscedasticity in y_i as well as upon the unknown parameter λ " (page 103).

In this Chapter we reconsider the problem of heteroscedasticity in the context of the Box-Cox transformation and propose a model for tackling heteroscedasticity in the transformed linear regression model. ML method of estimation, which would yield consistent and efficient estimates of the parameters, is suggested. A simpler method of obtaining ML estimates by searching over only λ is also described. Finally, we indicate how likelihood ratio tests may be used for examining various hypotheses regarding the transformation parameter and the parameter indicating the degree of heteroscedasticity.

The Chapter is arranged as follows : While in section 2.2 we discuss the problem of heteroscedasticity in the context of the Box-Cox transformation, the estimation procedure for the model proposed by us is described in section 2.3. Section 2.4 describes the procedure for simultaneous testing of specified values of the transformation and heteroscedasticity parameters characterizing different situations. Section 2.5 presents some numerical results obtained by applying our methods of estimation in our model and a number of other nested models for two different sets of data. Some concluding observations are made in the last section.

2.2 The Problem of Heteroscedasticity in the Box-Cox Model

Let us first discuss the problem of heteroscedasticity in the Box-Cox model more carefully. Given the transformation in (2.1.1), we easily find that[3/]

$$V\left[y_i^{(\lambda)}\right] \simeq V(y_i)\left[E(y_i)\right]^{2\lambda-2} \qquad \ldots \quad (2.2.1)$$

which shows that the transformation affects the heteroscedasticity of the $y_i^{(\lambda)}$'s. Even if $V(y_i)$ is constant for all observations, the transformed observations will, in general, have unequal variances unless, of course, $\lambda=1$ i.e., when there has been no transformation except one being added to the intercept term.

To examine the situation where the original observations are heteroscedastic, we consider, for example, a specific form of heteroscedasticity of y_i's, viz.,

$$V(y_i) = \sigma^2\left[E(y_i)\right]^h, \qquad E(y_i) > 0, \quad i=1, 2, \ldots, n \qquad \ldots(2.2.2)$$

where h is unknown. Then from (2.2.1), we have, under the assumption stated in (2.2.2),

$$V\left[y_i^{(\lambda)}\right] \simeq \sigma^2\left[E(y_i)\right]^{2\lambda-2+h}. \qquad \ldots (2.2.3)$$

By expressing $E(y_i)$ in terms of $E(y_i^{(\lambda)})$, (2.2.3) can be reduced to

---

3/ We make use of the well-known approximation

$$V\left[f(y)\right] \simeq V(y)\left[\frac{df(y)}{dy}\right]^2_{y = E(y)}$$

where $f(y)$ is a function of a random variable y.

$$V(y_i^{(\lambda)}) \simeq \begin{cases} \sigma^2 \left[1 + \lambda E(Z_i)\right]^{\delta} & \text{for } \lambda \neq 0 \\ \sigma^2 \left[e^{E(Z_i)}\right]^{\delta} & \text{for } \lambda = 0 \end{cases} \quad \ldots (2.2.4)$$

where $\delta = \left\{\frac{2\lambda-2+h}{\lambda}\right\}$ for $\lambda \neq 0$ and $(h-2)$ for $\lambda = 0$ and $Z_i = y_i^{(\lambda)}$. 4/

Thus, given that y_i's are heteroscedastic of the type considered, $y_i^{(\lambda)}$'s will have approximately constant variance across observations only when $h = (2-2\lambda)$ which, of course, is a very special case. For other types of heteroscedasticity of y_i's, $V(y_i^{(\lambda)})$'s will obviously be different. We, therefore, conclude that, in general, the transformed dependent variable will have different variances across observations irrespective of whether the original observations i.e., y_i's, were

---

4/ $\underline{\lambda \neq 0}$: Here $Z_i = \frac{y_i^{\lambda} - 1}{\lambda}$ i.e. $y_i = (1 + \lambda Z_i)^{\frac{1}{\lambda}}$.

Using Taylor expansion around $E(Z_i)$ upto the second term, we have

$$y_i \simeq \left[1 + \lambda E(Z_i)\right]^{\frac{1}{\lambda}} + \left[1 + \lambda E(Z_i)\right]^{\left(\frac{1}{\lambda} - 1\right)} (Z_i - E(Z_i)) .$$

Taking expectations on both sides, we have

$$E(y_i) \simeq \left[1 + \lambda E(Z_i)\right]^{\frac{1}{\lambda}} .$$

Equation (2.2.3), therefore, becomes

$$V\left[y_i^{(\lambda)}\right] \simeq \sigma^2 \left[1 + \lambda E(Z_i)\right]^{(2\lambda-2+h)/\lambda} .$$

$\underline{\lambda = 0}$: Here $Z_i = y_i^{(\lambda)} = \ln y_i$, then $y_i = e^{Z_i}$.

Taylor expansion around $E(Z_i)$ upto the second term yields

$$y_i \simeq e^{E(Z_i)} + (Z_i - E(Z_i)) e^{E(Z_i)} .$$

Taking expectations we, therefore, have

$$E(y_i) \simeq e^{E(Z_i)}$$

and hence from (2.2.3), we find

$$V\left[y_i^{(\lambda)}\right] \simeq \sigma^2 \left[e^{E(Z_i)}\right]^{h-2} .$$

homoscedastic or not and that the pattern of heteroscedasticity will, involve the transformation parameter λ besides the parameter h indicating the degree of heteroscedasticity, if any, in the original dependent variable.

2.3 Estimation of the Box-Cox Model Under Heteroscedasticity

From the discussion of the previous section, it is now clear that one should consider, for the purpose of estimation, the Box-Cox model with heteroscedasticity of the transformed dependent variable as given in (2.2.1). Since the methods available for estimating the Box-Cox heteroscedastic (BCH) model (viz., Zarembka's method and Egy-Lahiri method) are limited in applicability for reasons stated in the previous section, we propose here ML method of estimation of the BCH model given by (2.1.2) and (2.2.1) and suggest two different ways of obtaining the ML estimates of the relevant parameters. In our study we make two different assumptions about $V(y_i)$:

$$\text{(i)}\quad V(y_i) = \sigma^2 \left[E(y_i)\right]^h$$

$$\text{and (ii)}\quad V(y_i) = \sigma^2 \, m_i^h$$

where σ^2 and h are unknown parameters and m_i's are exogenously given. Assumption (i) is the same as that made by Zarembka (1974) but he treated h to be known for his estimation method. As can be seen from (2.1.4), Egy and Lahiri assumed the form (ii) for $V(y_i^{(\lambda)})$. It may be pointed out, however, that by considering a model given by (2.1.2) and

(2.2.1), we are implicitly assuming that there exists a value of λ i.e., a power transformation, which linearizes the relationship and ensures normality of the disturbances in the transformed model but in general, leads to heteroscedasticity, the exact nature of which will depend upon the nature of heteroscedasticity in the original y_i's and on the transformation parameter λ.

Let us first consider assumption (i). Then from section 2.2 we have from (2.2.4),

$$V\left[y_i^{(\lambda)}\right] \ (= \sigma_i^2, \text{ say}) \simeq \begin{cases} \sigma^2 \left[1 + \lambda E(Z_i)\right]^{\delta}, & \lambda \neq 0 \\ \sigma^2 \left[e^{E(Z_i)}\right]^{\delta}, & \lambda = 0. \end{cases}$$

Under the assumption of normality, the likelihood function of $(y_1, \ldots, y_n)$ can be written for $\lambda \neq 0$ as 5/

$$L = \text{Const.} - \frac{n}{2} \ln \sigma^2 - \frac{\delta}{2} \sum_{i=1}^{n} \ln (1 + \lambda \mu_i) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} \left\{ \frac{(y_i^{(\lambda)} - x_i' \beta)^2}{(1 + \lambda \mu_i)^{\delta}} \right\} + \ln J \quad \ldots (2.3.1)$$

where

$$J = \det \left| \frac{\partial y_i^{(\lambda)}}{\partial y_i} \right| = \prod_{i=1}^{n} y_i^{\lambda - 1} \ 6/,$$

$\mu_i = E(Z_i)$ and x_i' is the (1 x k) vector of ith observation on k regressors i.e., x_i' is the ith row of X.

5/ For the purpose of the present analysis we consider the likelihood equation only for $\lambda \neq 0$; for $\lambda = 0$, the likelihood equation can similarly be written.

6/ By choosing a suitable unit of measurement for y_i's, $\sum_{i=1}^{n} \ln y_i$ can be made to be zero in which case ln J vanishes in (2.3.1). However, as Schlesselman (1971) has pointed out, all the elements of the first column of X must then be assumed to be unity.

Simultaneous estimation of all the parameters of the model viz., λ, β, δ and σ^2 by ML procedure is evidently very complicated because of the nonlinearities involved. One may, however, find ML estimates of the parameters by using a search procedure over λ and δ.

This procedure can be described in the following steps :

(a) Fix a value for each of λ and δ.

(b) Use ordinary least squares (which obviously means that δ is taken to be zero) for the fixed λ and estimate β as

$$\hat{\beta}(\lambda, 0) = (X'X)^{-1} X'Y(\lambda) \qquad \ldots (2.3.2)$$

and hence $\hat{\mu}_i(\lambda, 0) = x_i' \hat{\beta}(\lambda, 0)$, $i = 1, 2, \ldots, n$.

(c) Find estimates of β and σ^2 in the second stage by

$$\hat{\beta}(\lambda, \delta) = (X' \hat{V}^{-1}(\lambda, \delta) X)^{-1} X' \hat{V}^{-1}(\lambda, \delta) Y(\lambda) \qquad \ldots (2.3.3)$$

where $\hat{V}^{-1}(\lambda, \delta)$ is the inverse of the matrix

$$\hat{V}(\lambda, \delta) = \text{diag}\left(\hat{v}_1(\lambda, \delta), \hat{v}_2(\lambda, \delta), \ldots, \hat{v}_n(\lambda, \delta) \right),$$

$$\hat{v}_i(\lambda, \delta) = \left[1 + \lambda \hat{\mu}_i(\lambda, 0)\right]^{\delta}, \quad i = 1, 2, \ldots, n \qquad \ldots (2.3.4)$$

and

$$\hat{\sigma}^2(\lambda, \delta) = \frac{1}{n} \sum_{i=1}^{n} \frac{(y_i(\lambda) - x_i' \hat{\beta}(\lambda, \delta))^2}{\left[1 + \lambda \hat{\mu}_i(\lambda, 0)\right]^{\delta}} \qquad \ldots (2.3.5)$$

(d) Log-likelihood value at (λ, δ) [1] is then computed as

$$L(\lambda, \delta) = \text{Const.} - \frac{\delta}{2} \sum_{i=1}^{n} \ln \left[1 + \lambda \hat{\mu}_i(\lambda, 0)\right] - \frac{n}{2} \ln \hat{\sigma}^2(\lambda, \delta) + (\lambda - 1) \sum_{i=1}^{n} \ln y_i \qquad \ldots (2.3.6)$$

---

[1] It is clear from (2.3.6) that there is no a priori reason to conclude that there may be more than one (λ, δ) for which $L(\lambda, \delta)$ will have approximately the same value and hence an identification problem as claimed by Egy and Lahiri (1978). See also footnote 2.

(e) Calculate now μ_i's as

$$\hat{\mu}_i(\lambda, \delta) = x_i' \hat{\beta}\ (\lambda, \delta), \qquad i = 1, 2, \ldots, n.$$

(f) Define new $\hat{v}_i\ (\lambda, \delta)$ $(i = 1, 2, \ldots, n)$ with $\hat{\mu}_i\ (\lambda, 0)$ in (2.3.4) being replaced by $\hat{\mu}_i(\lambda, \delta)$ and repeat steps (a) to (e) till the maximum of the log-likelihood function for the given (λ, δ) is attained (in computing new $\hat{\sigma}^2\ (\lambda, \delta)$ and $L(\lambda, \delta)$, $\hat{\mu}_i(\lambda, 0)$'s are now replaced by $\hat{\mu}_i(\lambda, \delta)$'s in (2.3.5) and (2.3.6)).

Steps (a) to (f) are repeated for different combinations of plausible values of λ and δ. For each pair of values (λ, δ), we thus obtain a maximum log-likelihood value. The ML estimates $\hat{\lambda}$ and $\hat{\delta}$ are those values of λ and δ for which the log-likelihood value is maximum among the maximized log-likelihood values over the different (λ, δ)'s and the corresponding estimates of β and σ^2, say $\hat{\beta}$ and $\hat{\sigma}^2$, are the ML estimates of β and σ^2 under standard conditions (see, Dhrymes (1970, Ch. 3) and Oberhofer and Kmenta (1974) for relevant conditions and results).

This search procedure is simple but obviously expensive and time consuming. We, therefore, suggest another method of obtaining ML estimates which is based upon searching over λ only. This method is an extension of that developed by Coondoo and Sarkar (1979) in the context of the standard (i.e., non-transformed) linear regression model having error variance structure of the type $V(y_i) = \left[E(y_i) \right]^{\delta}$ where δ is an unknown parameter and $E(y_i) > 0$. What we essentially do is to fix a value of λ and then estimate the parameter δ which maximizes the log-likelihood function conditional on the value of λ.

Suppose for a given λ, $\hat{\beta}(\lambda)$ is some consistent estimate of β for the given λ [8]. Then one can derive estimates of σ^2 and δ conditional on λ and $\hat{\beta}(\lambda)$ as follows :

$$\frac{\partial L}{\partial \sigma^2}\Big|_{\beta=\hat{\beta}(\lambda)} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n} e_i^2(\lambda)\,\tilde{\mu}_i^{-\delta}(\lambda) \qquad \ldots (2.3.7)$$

and

$$\frac{\partial L}{\partial \delta}\Big|_{\beta=\hat{\beta}(\lambda)} = -\frac{1}{2}\sum_{i=1}^{n} \ln \tilde{\mu}_i(\lambda) + \frac{1}{2\sigma^2}\sum_{i=1}^{n} e_i^2(\lambda)\,\tilde{\mu}_i^{-\delta}(\lambda)\ln \tilde{\mu}_i(\lambda) \qquad \ldots (2.3.8)$$

where $e_i(\lambda) = y_i^{(\lambda)} - x_i'\hat{\beta}(\lambda)$

and $\tilde{\mu}_i(\lambda) = \left[1 + \lambda \hat{\mu}_i(\lambda)\right]$, $\hat{\mu}_i(\lambda) = x_i'\hat{\beta}(\lambda)$.

Setting (2.3.7) and (2.3.8) equal to zero,

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n} e_i^2(\lambda)\,\tilde{\mu}_i^{-\delta}(\lambda) \qquad \ldots (2.3.9)$$

and $$\sum_{i=1}^{n} \ln \tilde{\mu}_i(\lambda) = \frac{n\sum_{i=1}^{n} e_i^2(\lambda)\,\tilde{\mu}_i^{-\delta}(\lambda)\ln \tilde{\mu}_i(\lambda)}{\sum_{i=1}^{n} e_i^2(\lambda)\,\tilde{\mu}_i^{-\delta}(\lambda)} . \qquad \ldots (2.3.10)$$

Let us now normalize $\tilde{\mu}_i(\lambda)$ by defining

$$\tilde{\tilde{\mu}}_i(\lambda) = \tilde{\mu}_i(\lambda) \Big/ \prod_{i=1}^{n}\left\{\tilde{\mu}_i(\lambda)\right\}^{\frac{1}{n}}$$

so that

$$\sum_{i=1}^{n} \ln \tilde{\tilde{\mu}}_i(\lambda) = 0.$$

---

[8] For the purpose of computation,

$\hat{\beta}(\lambda) = (X'X)^{-1}X'Y^{(\lambda)}$ is the obvious choice.

It is easy to see that (2.3.10) is satisfied even after a change in the units of the $\tilde{\mu}_i(\lambda)$'s. Hence we have for the $\tilde{\tilde{\mu}}_i(\lambda)$'s

$$\sum_{i=1}^{n} \ln \tilde{\tilde{\mu}}_i(\lambda) = \frac{n \sum_{i=1}^{n} e_i^2(\lambda) \tilde{\tilde{\mu}}_i^{-\delta}(\lambda) \ln \tilde{\tilde{\mu}}_i(\lambda)}{\sum_{i=1}^{n} e_i^2(\lambda) \tilde{\tilde{\mu}}_i^{-\delta}(\lambda)} \cdot \quad \ldots (2.3.11)$$

But since $\sum_{i=1}^{n} \ln \tilde{\tilde{\mu}}_i(\lambda) = 0$, (2.3.11) reduces to

$$\sum_{i=1}^{n} e_i^2(\lambda) \tilde{\tilde{\mu}}_i^{-\delta}(\lambda) \ln \tilde{\tilde{\mu}}_i(\lambda) \quad (= S(\delta(\lambda)), \text{ say}) \qquad \ldots (2.3.12)$$

Now, one can see that $S(\delta(\lambda))$ has a unique root for δ for a given λ because

$$\lim_{\delta \to \infty} S(\delta(\lambda)) = -\infty, \qquad \lim_{\delta \to -\infty} S(\delta(\lambda)) = \infty \quad 2/$$

and

$$\frac{\partial S(\delta(\lambda))}{\partial \delta} = -\sum_{i=1}^{n} e_i^2(\lambda) \tilde{\tilde{\mu}}_i^{-\delta}(\lambda) (\ln \tilde{\tilde{\mu}}_i(\lambda))^2 < 0$$

for $-\infty < \delta < \infty$.

This root can, therefore, be obtained by using standard techniques used for solving nonlinear equations, say, Newton-Raphson iterative procedure, as follows :

2/ By our change of units of measurement, some of the $\tilde{\tilde{\mu}}_i$'s are greater than 1 while others are less than 1. Thus

$$\lim_{\delta \to \infty} S(\delta(\lambda)) = \lim_{\delta \to \infty} \sum_{\tilde{\mu}_i(\lambda) < 1} e_i^2(\lambda) \tilde{\tilde{\mu}}_i^{-\delta}(\lambda) \ln \tilde{\tilde{\mu}}_i(\lambda)$$

$$+ \lim_{\delta \to \infty} \sum_{\tilde{\mu}_i(\lambda) > 1} e_i^2(\lambda) \tilde{\tilde{\mu}}_i^{-\delta}(\lambda) \ln \tilde{\tilde{\mu}}_i(\lambda) .$$

Obviously, the 1st part goes to $-\infty$ and the 2nd part to zero as $\delta \to \infty$. Hence $\lim_{\delta \to \infty} S(\delta(\lambda)) = -\infty$. Similarly $\lim_{\delta \to -\infty} S(\delta(\lambda)) = \infty$.

$$\delta_r = \delta_{r-1} + \frac{\sum_{i=1}^{n} e_i^2(\lambda)\, \tilde{\tilde{\mu}}_i^{-\delta_{r-1}}(\lambda)\, \ln \tilde{\tilde{\mu}}_i(\lambda)}{\sum_{i=1}^{n} e_i^2(\lambda)\, \tilde{\tilde{\mu}}_i^{-\delta_{r-1}}(\lambda)\, (\ln \tilde{\tilde{\mu}}_i(\lambda))^2} \qquad \ldots (2.3.13)$$

where δ_r and δ_{r-1} are solutions of δ at the rth and $(r-1)$th stages respectively. Once the converging δ, for the given λ, say $\hat{\delta}(\lambda)$ is available, estimate of σ^2, say, $\hat{\sigma}^2(\lambda, \hat{\delta}(\lambda))$ may be obtained from (2.3.9) by replacing δ by $\hat{\delta}(\lambda)$.

Now, the value of the log-likelihood function is computed as

$$L(\lambda, \hat{\delta}(\lambda)) = \text{Const.} - \frac{n}{2} \ln \hat{\sigma}^2(\lambda, \hat{\delta}(\lambda)) - \frac{\hat{\delta}(\lambda)}{2} \sum_{i=1}^{n} \ln \tilde{\mu}_i(\lambda) + (\lambda - 1) \sum_{i=1}^{n} \ln y_i \qquad \ldots (2.3.14)$$

Then, one gets improved estimates of β, say, $\hat{\beta}(\lambda, \hat{\delta}(\lambda))$, defined as

$$\hat{\beta}(\lambda, \hat{\delta}(\lambda)) = (X' \hat{V}^{-1}(\lambda, \hat{\delta}(\lambda)) X)^{-1} X' \hat{V}^{-1}(\lambda, \hat{\delta}(\lambda)) Y^{(\lambda)} \qquad \ldots (2.3.15)$$

where $\hat{V}^{-1}(\lambda, \hat{\delta}(\lambda))$ is the inverse of the matrix

$$\hat{V}(\lambda, \hat{\delta}(\lambda)) = \text{diag}\left(\hat{v}_1(\lambda, \hat{\delta}(\lambda)), \hat{v}_2(\lambda, \hat{\delta}(\lambda)), \ldots, \hat{v}_n(\lambda, \hat{\delta}(\lambda))\right),$$

where $\hat{v}_i(\lambda, \hat{\delta}(\lambda)) = \left[\tilde{\mu}_i(\lambda)\right]^{\hat{\delta}(\lambda)}$, $i = 1, 2, \ldots, n$.

Defining new $\tilde{\mu}_i$'s and e_i's as

$$\tilde{\mu}_i(\lambda, \hat{\delta}(\lambda)) = \left[1 + \lambda x_i' \hat{\beta}(\lambda, \hat{\delta}(\lambda))\right] \qquad \ldots (2.3.16)$$

and $$e_i(\lambda, \hat{\delta}(\lambda)) = \left[y_i^{(\lambda)} - x_i' \hat{\beta}(\lambda, \hat{\delta}(\lambda))\right]$$

another unique estimate of δ is obtained from (2.3.13) and new estimate of σ^2 i.e., $\hat{\sigma}^2(\lambda, \hat{\delta}(\lambda))$ is obtained from (2.3.9) with $\tilde{\mu}_i(\lambda)$, δ and $e_i(\lambda)$ replaced by $\tilde{\mu}_i(\lambda, \hat{\delta}(\lambda))$, $\hat{\delta}(\lambda)$ and $e_i(\lambda, \hat{\delta}(\lambda))$ respectively

from (2.3.16). The new value of $L(\lambda, \hat{\delta}(\lambda))$ is accordingly calculated from (2.3.14) with $\tilde{\mu}_i(\lambda)$ being replaced by $\tilde{\mu}_i(\lambda, \hat{\delta}(\lambda))$ and σ^2 and δ by their new estimates.

With these new $\tilde{\mu}_i(\lambda, \hat{\delta}(\lambda))$'s, new $\hat{V}(\lambda, \hat{\delta}(\lambda))$ is obtained and hence new $\hat{\beta}(\lambda, \hat{\delta}(\lambda))$ from (2.3.15) and then new estimate of δ and so on. The process is repeated till the maximum value of the log-likelihood function for the given λ is obtained, and the corresponding solution of δ obtained from (2.3.13) is taken as the estimate of δ for the given λ.

In order to maximize over the entire parameter space, we only need to choose alternative values of λ over a reasonable range and obtain the unique $\hat{\delta}(\lambda)$'s and the corresponding maximum log-likelihood values for the given λ values. The value of λ, say $\hat{\lambda}$, for which the log-likelihood function is maximum among all the maximum log-likelihood values for the given λ's, is, in fact, the ML estimate of λ. At $\lambda=\hat{\lambda}$, the corresponding estimates of δ, β and σ^2 denoted by $\hat{\delta}$, $\hat{\beta}$ and $\hat{\sigma}^2$ accordingly obtained from (2.3.13), (2.3.15) and (2.3.9) respectively[10], are, in fact, the ML estimates. As has been pointed out by Savin and White (1978) in the context of Box-Cox procedure with autocorrelation that it does not appear to be feasible to analytically derive the

---

10/ This will be so because of $\lambda=\hat{\lambda}$, the iterative process with new $\tilde{\mu}_i(\lambda)$'s will converge at the maximum log-likelihood value and the converging estimates of β, δ and σ^2 i.e., $\hat{\beta}$, $\hat{\delta}$ and $\hat{\sigma}^2$ will satisfy the normal equations under standard conditions. See Dhrymes (1970. Ch. 3) and Oberhofer and Kmenta (1974) for conditions and other details.

variance-covariance matrix of the limiting distribution of the ML estimates of all the parameters. However, an approximation to the estimated information matrix can be obtained numerically from the second derivatives.

Now, if we consider the second assumption about the variance of y_i i.e., $V(y_i) = \sigma^2 m_i^h$, we get

$$V\left[y_i^{(\lambda)}\right] = \begin{cases} \sigma^2 \left[E(y_i)\right]^{2\lambda-2} m_i^h & \text{if } \lambda \neq 0 \\ \sigma^2 \left[E(y_i)\right]^{-2} m_i^h & \text{if } \lambda = 0 \end{cases}$$

which after approximations similar to the one done in case of first assumption about $V(y_i)$, reduces to

$$V\left[y_i^{(\lambda)}\right] \simeq \begin{cases} \sigma^2 \left[1 + \lambda E(Z_i)\right]^{(2\lambda-2)/\lambda} m_i^h & \text{if } \lambda \neq 0 \\ \sigma^2 \left[e^{E(Z_i)}\right]^{-2} m_i^h & \text{if } \lambda = 0. \end{cases} \quad \ldots (2.3.17)$$

The estimation methods presented above can be restated by changing the relevant equations in accordance with this new expression given for $V\left[y_i^{(\lambda)}\right]$ in (2.3.17). Even without going into details, one can conclude that for this case also the ML estimates of λ, h, β and σ^2 can be obtained by either searching over λ and h or by searching over λ only (see Appendix 2.1 for necessary steps).

2.4 Likelihood Ratio (LR) Tests

We now consider LR test of hypotheses concerning different values of λ and δ [11/] in either of the models. Such hypotheses could provide information regarding the appropriate functional form of the model and the degree of heteroscedasticity embodied in it. If $L(H_o)$ and $L(H_1)$ are the maximum values of the log-likelihood function under the null hypothesis (H_o) and the alternative hypothesis (H_1), then for the large-sample LR test of H_o against H_1, the test statistic defined as

$$l(.) = -2\left[L(H_o) - L(H_1)\right] \qquad \ldots \qquad (2.4.1)$$

has a limiting chi-square distribution with degrees of freedom p, the number of additional restrictions imposed by H_o.

Hypotheses of different types may be tested by this method. Firstly, we may be interested in testing the hypothesis that the functional form is linear under the condition that the disturbances are homoscedastic. The conditional null hypothesis is H_o : $\lambda=1$ and $\delta = 0$ and the alternative is H_1 : $\lambda \neq 1$ and $\delta = 0$. The appropriate test statistic is

$$l(1) = -2\left[L(\lambda=1,\ \delta=0) - L(\hat{\lambda}(0),\ \delta = 0)\right] \sim \chi^2_1 \text{ under } H_o$$

where $\hat{\lambda}(0)$ denotes the ML estimate of λ given $\delta = 0$ i.e., estimate of λ in the Box-Cox model. The limitation of this test, however, is that it assumes homoscedastic disturbances which is not usually true

---

11/ We write down the different likelihood ratio tests for assumption (i) about variance of y_i i.e., in terms of λ and δ. In case of assumption (ii) about the variance of y_i, δ is to be replaced by h.

for the transformed model. Hence, we may be interested in testing a more general hypothesis H_o : $\lambda=1$ against H_1 : $\lambda \neq 1$. The relevant test statistic then is

$$l(2) = -2\left[L(\lambda=1, \hat{\delta}(1)) - L(\hat{\lambda}, \hat{\delta})\right] \sim \chi^2_1 \text{ under } H_o.$$

where $\hat{\delta}(1)$ denotes ML estimate of δ given $\lambda=1$.

In some cases we may like to test for a specific value of λ different from unity, say, $\lambda=0$. Obviously, depending upon whether $\delta = 0$ or not, the test statistis would be similar to the above.

We may as well be interested to test the hypothesis H_o : $\delta = 0$ given $\lambda=1$ against H_1 : $\delta \neq 0$ given $\lambda = 1$ i.e., whether the heteroscedasticity is due to heteroscedasticity in y_i's only. The test statistic in this case becomes

$$l(3) = -2\left[L(\lambda=1, \delta=0) - L(\lambda=1, \hat{\delta}(1))\right] \sim \chi^2_1 \text{ under } H_o.$$

As before a more general hypothesis will be H_o : $\delta = 0$ against H_1 : $\delta \neq 0$; for this the test statistic is

$$l(4) = -2\left[L(\hat{\lambda}(0), \delta = 0) - L(\hat{\lambda}, \hat{\delta})\right] \sim \chi^2_1 \text{ under } H_o.$$

Finally, we may wish to test the joint hypothesis H_o : $\lambda=1$, $\delta = 0$ against H_1 : $\lambda \neq 1$ and/or $\delta \neq 0$. The test statistic is obviously given by

$$l(5) = -2\left[L(\lambda=1, \delta = 0) - L(\hat{\lambda}, \hat{\delta})\right] \sim \chi^2_2 \text{ under } H_o.$$

2.5 Numerical Illustrations

In this section we report the results obtained by using maximum likelihood (ML) method of estimation suggested here for the BCH model for two different sets of data. The first one is taken from O'Hara and McClelland (1964) and relates to radio-sales (y) and income (x) for

49 states in the United States for the year 1954. This was first analysed by Rutemiller and Bowers (1968) in terms of a linear regression model with heteroscedastic errors. The second example is based on the data available from Feigl and Zelen (1965) on survival time (y) of 17 leukemia patients and their while blood cell-counts (x). They used this data to estimate the linear regression of y on ln x with exponential errors. This set of data was also used by Cox and Snell (1968) and Amemiya (1973b) for their studies.

We have used these two sets of data to estimate the following regression equations

$$y_i^{(\lambda)} = \beta_1 + \beta_2 x_i + \varepsilon_i \quad (i = 1,2,.., 49), \text{ for the first example} \quad \ldots (2.5.1)$$

and

$$y_i^{(\lambda)} = \beta_1 + \beta_2 x_i^{(\lambda)} + \varepsilon_i \quad (i=1,2,.., 17), \text{ for the second example} \quad \ldots (2.5.2)$$

under two alternative assumptions about the variances of the y_i's viz., (i) $V(y_i) = \sigma^2 \left[E(y_i)\right]^h$ and (ii) $V(y_i) = \sigma^2 x_i^h$, where in each case h and σ^2 are unknown. It then follows from (2.2.1) that the variance for $y_i^{(\lambda)}$ corresponding to these cases are (a) $\sigma_i^2 = V(y_i^{(\lambda)}) \simeq \sigma^2 \left[E(y_i)\right]^{\delta_1}$ where $\delta_1 = \{2\lambda - 2 + h\}$ and (b) $\sigma_i^2 \simeq \sigma^2 \left[E(y_i)\right]^{2\lambda-2} x_i^h$. In addition to these two, we have made use of a third assumption which is the one used by Egy and Lahiri (1978) viz., $\sigma_i^2 = \sigma^2 x_i^\delta$. It should, however, be noted that these assumptions do not make much of a difference in the ultimate conclusions for single-regressor models like ours since we can easily see that $\sigma^2 \left[E(y_i)\right]^{\delta_1} \simeq \sigma^{*2} x_i^\delta$ (where σ^{*2} is suitably defined) under the condition that the value of $(1 + \lambda\beta_1)$ is much smaller

compared to the value of $\lambda\beta_2 x_i$[12/] which we have, in fact, found to hold in our examples. In such cases, the difference in assumption about the variances will, however, be reflected in the estimate of σ^2. This is obvious from footnote 12. In our numerical examples also, we have found the estimate of σ^2 to be different for the different assumptions about the variance.

In Table 2.1, we present the maximum values of the log-likelihood function for each of the different assumed values of one or both of λ and δ/h. More specifically, in addition to the maximum log-likelihood values for BCH and the Box-Cox (BC) models, we also report the maximum log-likelihood values for $\lambda=1$ and 0 and/or $\delta/h = 0$. For each of the two examples we have used both the methods described in section 2.3 for obtaining ML estimates i.e., (i) search over λ and δ/h only and (ii) search over λ only. These two methods produced almost identical results (though the second one took much less computation time) and hence, for convenience, we have reported the results for the first method only. Results of the likelihood ratio (LR) tests for different hypotheses concerning λ and δ are given in Table 2.2. We note from Table 2.1 that for both the examples the maximum values of the log-likelihood function for the three different assumptions about the variance are quite close to one another for each of the six models (values of λ and δ/h) considered. The LR test statistic values in Table 2.2 are, therefore, reported only for case (a) i.e., for the assumption $V\left[y_i^{(\lambda)}\right] = \sigma^2\left[E(y_i)\right]^{\delta_1}$

12/ For (2.5.1), we have from section 2.2

$$\sigma^2\left[E(y_i)\right]^{\delta_1} \simeq \sigma^2\left[1 + \lambda E(y_i^{(\lambda)})\right]^{\delta}$$

$$= \sigma^2\left[1 + \lambda\beta_1 + \lambda\beta_2 x_i\right]^{\delta}$$

$$\simeq \sigma^{*2} x_i^{\delta} \quad \text{(where } \sigma^{*2} = \sigma^2(\lambda\beta_2)^{\delta}\text{) under the stated condition.}$$

The same relation obviously holds for (2.5.2) also.

Table 2.1 : Maximum log-likelihood values (L) for different values of the transformation parameter (λ) and degree of heteroscedasticity (δ or h).[a]

Model*	Case (a): $V\left[y_i^{(\lambda)}\right] = \sigma^2\left[E(y_i)\right]^{\delta}1 \simeq \sigma^2\left[1+\lambda E(y_i^{(\lambda)})\right]^{\delta}$			Case (b) : $V\left[y_i^{(\lambda)}\right] = \sigma^2 x_i^{\delta}$			Case (c) : $V\left[y_i^{(\lambda)}\right] = \sigma^2\left[E(y_i)\right]^{2\lambda-2} x_i^{h}$		
	λ	δ	L	λ	δ	L	λ	h	L
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Radio-Sales Data A	1**	0**	-159.524	1**	0.**	-159.524	1**	0**	-159.524
B	1**	1.501	-131.989	1**	1.501	-132.660	1**	1.501	-132.060
C	0.923	1.326	-129.686	0.923	1.273	-129.588	0.923	1.439	-129.596
	(0.923)[a]	(1.325)	(-129.732)	(0.923)	(1.277)	(-129.609)	(0.923)	(1.438)	(-129.595)
D	0.786	0**	-146.375	0.786	0**	-146.375	[b]	[b]	[b]
E	0**	0**	-187.682	0**	0**	-187.682	0**	0**	-192.277
F	0**	0.613	-184.316	0**	-0.923	-183.067	0**	-0.428	-191.205
Leukemia Data A	1**	0**	- 64.526	1**	0**	- 64.526	1**	0**	- 64.526
B	1**	0.582	- 64.188	1**	-0.281	- 64.054	1**	-0.281	- 64.054
C	0.124	[c]	- 56.210	0.163	0.587	- 55.842	0.00	-0.299	- 56.363
	(0.124)[a]	[c]	(- 56.209)	(0.163)	(0.587)	(- 55.841)	(0.01)	(-0.332)	(- 56.366)
D	0.282	0**	- 57.340	0.299	0**	- 57.347	[b]	[b]	[b]
G	0**	0**	- 59.642	0**	0**	- 59.642	0**	0**	- 56.408
H	0**	-1.500	- 56.367	0**	0.851	- 56.368	0**	-0.282	- 56.367

Notes : * A : Linear Homoscedastic, B : Linear Heterescedastic, C : Box-Cox Heteroscedastic (BCH), D : Box-Cox (BC), E : Semi-log Homoscedastic, F : Semi-log Heteroscedastic, G : Double-log Homoscedastic and H : Double-log Heteroscedastic.

** indicates that the value of the parameter in question is given a priori from the model assumed.

[a] As both the methods for obtaining ML estimates give identical results, we have reported the results of the first method only. For the sake of illustration, however, the results obtained by the second method are given in brackets for the BCH model only.

[b] Entires corresponding to BC model for case (c) are omitted since for this case even for h=0, the model does not reduce to that of BC.

[c] Since the estimate of δ in BCH model for the Leukemia data in case (a) lies in a range where L is insensitive to changes in δ, we have kept the corresponding entry in the table blank.

Table 2.2 : Results of likelihood ratio tests for different null hypotheses for case (a).

Statistic	H_o	H_1	Radio-Sales data		Leukemia data	
			value of LR-test statistic	conclusions*	value of LR-test statistic	conclusions*
(1)	(2)	(3)	(4)	(5)	(6)	(7)
1(1)	$\lambda=1, \delta=0$	$\hat{\lambda}(0), \delta=0$	26.298	reject H_o	14.372	reject H_o
1(2)	$\lambda=1, \hat{\delta}(1)$	$\hat{\lambda}, \hat{\delta}$	4.606	accept H_o at 1 % but reject at 5 %	15.956	reject H_o
1(3)	$\lambda=1, \delta=0$	$\lambda=1, \hat{\delta}(1)$	55.070	reject H_o	0.676	accept H_o
1(4)	$\hat{\lambda}(0), \delta=0$	$\hat{\lambda}, \hat{\delta}$	33.378	reject H_o	2.260	accept H_o
1(5)	$\lambda=1, \delta=0$	$\hat{\lambda}, \hat{\delta}$	59.676	reject H_o	16.632	reject H_o
1(6)	$\lambda=0, \delta=0$	$\lambda=0, \hat{\delta}(0)$	6.732	reject H_o	6.550	accept H_o at 1 % but reject at 5 %
1(7)	$\lambda=0, \delta=0$	$\hat{\lambda}, \hat{\delta}$	116.244	reject H_o	6.864	reject H_o
1(8)	$\lambda=0, \hat{\delta}(0)$	$\hat{\lambda}, \hat{\delta}$	109.260	reject H_o	0.314	accept H_o
1(9)	$\lambda=0, \delta=0$	$\hat{\lambda}(0), \delta=0$	82.866	reject H_o	4.604	accept H_o at 1 % but reject at 5 %

At 5 % level of significance $\chi^2_1 = 3.84$ and $\chi^2_2 = 5.99$.

At 1 % level of significance $\chi^2_1 = 6.64$ and $\chi^2_2 = 9.21$.

* Unless otherwise mentioned, 'reject/accept H_o' means that the null hypothesis H_o is rejected/accepted in favour of/against the alternative hypothesis H_1 at both 5 % and 1 % levels of significance.

For the first example, the likelihood method of estimation of BCH model has yielded estimates of λ and δ as $\hat{\lambda} = 0.923$ and $\hat{\delta} = 1.326$ and the maximum log-likelihood value is -129.686. The corresponding value for BC model is only -146.375. We find from 1(4) that BC model is clearly rejected in favour of BCH model. This, therefore, indicates that one would have chosen a wrong model by straightforwardly using Box-Cox procedure. As the model actually turns out to be approximately linear, 1(2) suggests that we may accept $H_o : \lambda=1, \hat{\delta}(1)$ for the sake of simplicity against $H_1 : \hat{\lambda}, \hat{\delta}$ because when tested H_o is rejected at 5 per cent but not at 1 per cent level of significance. It is important to note that the advantage of using BCH model is that by rejecting the inappropriate null hypotheses against the unrestricted hypothesis, it helps us in choosing the proper model. This is shown by 1(5), 1(7) and 1(8). It may also be seen from 1(6), for example, how conditional hypotheses (where values of one parameter are a priori assumed to be known) may lead to wrong conclusion about the proper model. In this case λ is a priori fixed at $\lambda=0$ and we find that the null hypothesis ($H_o : \lambda=0, \delta=0$) is rejected in favour of the alternative ($H_1 : \lambda=0, \hat{\delta}(0)$) at 5 per cent level of significance and H_o is almost accepted against H_1 at 1 per cent level of significance though the maximum values of the log-likelihood function are much less for both the hypotheses as compared to the maximum value for the BCH model. Thus we find that choice of functional form appears to be crucial in discriminating among different models and also, as in this examples, that estimation of λ seems to be influenced by heteroscedasticity.

In the second example of survival time of leukemia patients the ML method of estimation gives $\hat{\lambda} = 0.124$ for the BCH model. However, as the value of λ is very small and the values of $\left[E(y_i^{(\lambda)})\right]$'s are not very large which is clear from the data, it may be seen from our approximation for $E(y_i)$ i.e., $E(y_i) \simeq \left[1 + \lambda E(y_i^{(\lambda)})\right]^{1/\lambda}$, that the variances σ_i^2's become insensitive to $\delta = \{\frac{2\lambda-2+h}{\lambda}\}$ because $\left[1 + \lambda E(y_i^{(\lambda)})\right] \simeq 1$ for all i. Hence we find that for the estimate of δ in the range -1 to -9 the maximum value of the log-likelihood function is almost the same being around -56.[13/] In such cases of small values of λ, we suggest that one should test for $H_o : \lambda = 0, \hat{\delta}(0)$ against $H_1 : \hat{\lambda}, \hat{\delta}$ and if the null hypothesis is accepted, one should proceed (for further studies) with the model where $\lambda = 0$, and for this case there is a separate expression for approximating $E(y_i)$ given in footnote 4. We find from Table 2.1 that maximum log-likelihood values for the case $(\lambda=0, \hat{\delta}(0))$ and for the BCH model (i.e., $\hat{\lambda}, \hat{\delta}$) are quite close, being -56.367 and -56.210 respectively, and δ has a unique estimate at $\hat{\delta} = -1.500$ in the former case. The fact that $\hat{\delta}$ and hence $\hat{\delta}_1$ has a negative value indicates that the variance of $y_i^{(\lambda)}$'s decrease with increase in $E(y_i)$. We, however, note that for $\lambda = 0$ and $\hat{\delta} = -1.5$, the estimate of h comes out to be 0.5 which means that the original observations of the dependent variable i.e., y_i's, have variance increasing with increases in $E(y_i)$. This clearly shows how transformation may affect the variance of the

---

13/ This is obviously not going to happen for the other two assumptions about the variance i.e., for cases (b) and (c) and this is clear from Table 2.1. However, as it should be, the maximum log-likelihood values for the three cases are almost same.

transformed observations compared to that of the original observations. This negative value of $\hat{\delta}$ also implies that heteroscedasticity is relatively less important. This is also corroborated by 1(4). How conditional hypotheses may lead to acceptance of wrong models is revealed here also by 1(3), for example, which indicates that linear homoscedastic model is accepted against linear heteroscedastic model though both have much less maximum log-likelihood values as compared to the BCH model and both are, as indicated by 1(5) and 1(2), rejected in favour of the BCH model.

Thus we find that in a practical situation the doparture of the estimate of λ obtained from a model with heteroscedasticity from that of a model assuming homoscedasticity and hence the consequences of choosing an inappropriate model may depend upon the data. But since the actual situation cannot be known a priori, it is, in general, advisable to estimate λ within the framework of the BCH model.

2.6 Conclusions

Box and Cox (1964) suggested a transformation of the dependent variable in a regression model in order to achieve linearity, homoscedasticity and normality of the transformed dependent variable and proposed a maximum likelihood method of estimation of the parameters of such a model. Zarembka (1974) however showed that the ML method of estimation suggested by Box and Cox is not robust to heteroscedasticity and that the estimate of λ will be biased towards the direction of stabilizing the error variance.

In this Chapter we have asserted that the transformation that leads to linearity does not necessarily lead to homoscedasticity also. This is evident from the transformation itself. We have argued that since heteroscedasticity in the transformed dependent variable is due to both the transformation and the heteroscedasticity in the original values of the dependent variable, one should estimate the parameters in the framework of heteroscedasticity as given by the transformation and the heteroscedasticity in the original dependent variable. We have advocated the ML method of estimation of such a model and have also suggested an easier and computationally convenient method of obtaining ML estimates by searching over only the transformation parameter. We have also indicated how likelihood ratio tests can be used to compare the fits provided by different models corresponding to different combinations of the transformation parameter and the parameter indicating the degree of heteroscedasticity. Our analysis also shows incidentally that the assumption of variance of the type $\sigma^2 \left[E(y_i)\right]^{\delta}$ for the transformed dependent variable does not lead to any identification problem as claimed by Egy and Lahiri (1978) (see footnote 2).

Appendix 2.1

Methods of Estimation under the Alternative Assumption : $V(y_i) = \sigma^2 m_i^h$

First method (i.e., searching over λ and h)

Steps (a) and (b) remain unchanged with δ being replaced by h.

In step (c),

$\hat{v}_i(\lambda, \delta)$ changes to

$$\hat{v}_i(\lambda, h) = \left[1 + \lambda \hat{\mu}_i (\lambda, 0)\right]^{(2\lambda-2)/\lambda} m_i^h ;$$

$$i = 1, 2, \ldots, n$$

and $\hat{\sigma}^2(\lambda, \delta)$ to

$$\hat{\sigma}^2(\lambda, h) = \frac{1}{n} \sum_{i=1}^{n} \frac{(y_i^{(\lambda)} - x_i' \hat{\beta}(\lambda, h))^2}{\left[1 + \lambda \hat{\mu}_i(\lambda, 0)\right]^{(2\lambda-2)/\lambda} m_i^h} .$$

In step (d), $L(\lambda, \delta)$ changes to

$$L(\lambda, h) = \text{Const.} - \frac{h}{2} \sum_{i=1}^{n} \ln m_i - \frac{\lambda-1}{\lambda} \sum_{i=1}^{n} \ln (1+\lambda \hat{\mu}_i(\lambda, 0))$$

$$- \frac{n}{2} \ln \hat{\sigma}^2 (\lambda, h) + (\lambda-1) \sum_{i=1}^{n} \ln y_i .$$

Other steps remain unchanged excepting that δ is now to be replaced by h.

Second method (i.e., searching over λ only)

In expression (2.3.7) to (2.3.13), the following changes are to be made :

δ, $\tilde{\mu}_i^{-\delta}(\lambda)$ and $\ln \tilde{\mu}_i(\lambda)$ will now have to be replaced by h, $\tilde{\mu}_i^{-(2\lambda-2)/\lambda}\, m_i^{-h}$ and $\ln m_i$ respectively.

Normalization will now have to be done with respect to m_i's. $L(\lambda, \hat{\delta}(\lambda))$ in (2.3.14) will change to

$$L(\lambda, \hat{h}(\lambda)) = \text{Const.} - \frac{\hat{h}(\lambda)}{2} \sum_{i=1}^{n} \ln m_i - \frac{\lambda-1}{\lambda} \sum_{i=1}^{n} \ln \tilde{\mu}_i(\lambda)$$

$$- \frac{n}{2} \ln \hat{\sigma}^2(\lambda, \hat{h}(\lambda)) + (\lambda-1) \sum_{i=1}^{n} \ln y_i \,.$$

Chapter 3

The Box-Cox Transformation and Non-spherical Disturbances

3.1 Introduction

In the last Chapter we discussed the problem of heteroscedasticity in the context of the Box-Cox transformation and proposed a model incorporating both the transformation and heteroscedasticity due to it and also possible heteroscedasticity in the original values of the dependent variable. We also suggested estimation of the parameters of such models. The present Chapter is devoted to a similar analysis for non-spherical disturbances, in particular, that for autocorrelation among disturbances in the context of the Box-Cox transformation.

Generalization of the Box-Cox model to take into account autocorrelation of the disturbances was first attempted by Savin and White (1978). They adopted a straightforward and somewhat convenient approach of assuming the disturbances of the transformed model to follow some standard error process. More specifically, they assumed a first-order autoregressive (AR(1)) process[1/] for the disturbances of the transformed linear regression model and used this generalized model to

1/ As mentioned by Savin and White, one can also assume higher order AR or moving average processes for the disturbances. Furthermore, as in Savin and White (1978), it is possible to consider transformation of independent variables as well. We, however, confine ourselves to the transformation of the dependent variable only since transformations of independent variables do not give rise to any real problem and can be easily tackled in a similar fashion.

test for autocorrelation and functional form simultaneously.

In this Chapter we propose an alternative and more general approach for incorporating autocorrelation of disturbances in the Box-Cox model. It will be seen that our approach can, in fact, tackle the problems of both heteroscedasticity and autocorrelation (i.e., non-sphericalness, in general) in the Box-Cox model. Instead of assuming some standard error process for the disturbances in the transformed linear regression equation, we suggest starting with an assumed autocorrelation function for the original observations on the dependent variable. In contrast to the simple generalization by Savin and White, our approach gives due consideration to the complications created by the fact that the variances and autocovariances of the transformed dependent variable depend on those of the original variable and also on the unknown transformation parameter. Savin and White's generalization comes out to be a special case of our model.

In section 3.2, we describe the problem and our generalization. A method of estimation of our generalized model is suggested in section 3.3. We also examine the consistency of the estimate of the transformation parameter obtained through the Savin and White procedure when actually the transformed dependent variable has alternative variance-covariance structures. This is presented in section 3.4. Like Savin and White, we also indicate in section 3.5 how likelihood ratio tests can be used to identify situations characterized by different combination of values of the transformation parameter and parameters relating to heteroscedasticity and autocorrelation. The last section gives some concluding observations.

3.2 The Box-Cox Transformation and the Problem of Autocorrelation

Let us recall the Box-Cox model described as

$$y_i^{(\lambda)} = x_i'\beta + \varepsilon_i , \qquad i = 1, 2, \ldots, n \qquad \ldots (3.2.1)$$

where $y_i^{(\lambda)}$ is the ith observation on the transformed dependent variabl and is defined as

$$y_i^{(\lambda)} = \begin{cases} (y_i^{\lambda} - 1)/\lambda , & \text{if } \lambda \neq 0 \\ \ln y_i , & \text{if } \lambda = 0 \end{cases} \qquad \ldots (3.2.2)$$

and $x_i' = (x_{1i}, \ldots, x_{ki})$ is the ith row vector of observations on k fixed regressors, $\beta = (\beta_1, \ldots, \beta_k)'$ is the vector of associated regression coefficients and ε_i is the disturbance in the ith transformed observation on the dependent variable in the regression model.

Given the above definition in (3.2.2), we can now expand $y_i^{(\lambda)}$ around the expected value of y_i using Taylor expansion and approximate it as follows :

$$y_i^{(\lambda)} \simeq \left[E(y_i)\right]^{(\lambda)} + (y_i - E(y_i)) \left(\frac{dy_i^{(\lambda)}}{dy_i} \right)_{y_i = E(y_i)}$$

$$= \left[E(y_i)\right]^{(\lambda)} + (y_i - E(y_i)) \left[E(y_i)\right]^{\lambda-1}$$

whence it follows that

$$\mathrm{Cov}(y_i^{(\lambda)}, y_j^{(\lambda)}) \simeq \mathrm{Cov}(y_i, y_j) \left[E(y_i)\, E(y_j)\right]^{\lambda-1}, \qquad \ldots (3.2.3)$$

$$i, j = 1, \ldots, n.$$

The above relation shows how the autocovariances of the transformed observations i.e., $y_i^{(\lambda)}$'s, are related to those of the y_i's and to the transformation parameter λ. It is also clear that the autocovariance structures of the transformed and the original observations are not identical excepting for the case $\lambda = 1$ i.e., when there has been no transformation except one being added to the intercept term.

We may get further insight into the nature of autocovariances of the transformed variable by examining the relationship in (3.2.3) more closely. When either the values of $\left[E(y_i)\ E(y_j) \right] = d_{ij}$, say, are very close to unity for all i and j, or when λ is very close to unity or both, the variances and autocovariances of $y_i^{(\lambda)}$'s will be almost the same as those of the y_i's. Obviously, the first case is very unrealistic and the second case (i.e., $\lambda = 1$) represents the standard (i.e., non-transformed) autocorrelated linear regression model. If, on the other hand, the underlying relation between the original dependent variable and the independent variables is highly nonlinear and/or d_{ij}'s are not very close to unity, the autocovariance structuresfor $y_i^{(\lambda)}$'s will be different from that of y_i's. Finally, if $\lambda < 1$, say equal to -1, in particular, and d_{ij}'s are greater than unity, then $d_{ij}^{\lambda-1}$'s may, in fact, affect the autocovariances of $y_i^{(\lambda)}$'s quite significantly compared to those of y_i's. Thus, we find that the autocovariances of $y_i^{(\lambda)}$'s may be the same, differ only slightly or differ appreciably from those of the y_i's depending upon the values of λ and the d_{ij}'s.[2/]

2/ It may, however, be noted that the correlation coefficients between $y_i^{(\lambda)}$ and $y_j^{(\lambda)}$ are approximately the same as those between y_i and y_j for all i, j = 1, ..., n (i ≠ j).

The difference between our approach and that of Savin and White may now be stated with reference to the above relationship. Whereas Savin and White generalized the Box-Cox formulation by directly assuming a simple error process for the disturbances of the transformed linear regression model, our proposal is to consider instead a structure for $V(y_i)$ and an autocorrelation function for y_i's and then use the relationship in (3.2.3) to derive the variance-covariance matrix for $y_i^{(\lambda)}$'s. Assuming that $V(y_i) = \sigma_y^2 \left[E(y_i)\right]^{2h}$, where $E(y_i) > 0$ and σ_y^2 and h are unknown parameters, we find from (3.2.3), that

$$\text{Cov}(y_i^{(\lambda)}, y_j^{(\lambda)}) \simeq \sigma_y^2 \text{ Corr } (y_i, y_j) \left[E(y_i)\ E(y_j)\right]^{\lambda-1+h}.$$

Let us further assume that

$$\text{Corr } (y_i, y_j) = \rho^{*\,|i-j|}, \quad i, j = 1, 2, \ldots, n\ (i \neq j)$$

where $|\rho^*| < 1$ is unknown.

Then we have,

$$\text{Cov}(y_i^{(\lambda)}, y_j^{(\lambda)}) \simeq \sigma_y^2\, \rho^{*\,|i-j|} \left[E(y_i)\ E(y_j)\right]^{\lambda-1+h}; \quad \ldots(3.2.4)$$

$$i, j = 1, 2, \ldots, n.$$

This alternative approach is more general and appropriate as compared to that of Savin and White for a number of reasons. First of all, it may be noted that an implication of the Savin and White's assumption of AR(1) process, or for that matter any other standard error process like the higher order AR processes or moving average processes, for the

transformed linear regression model is that the disturbances become homoscedastic after the transformation. But, as we have shown in Chapter 2, it is not, in general, possible to have a transformation which would ensure both linearity of the relation and homoscedasticity of the disturbances. In other words, apart from possible heteroscedasticity in y_i's, the transformation itself introduces, in general, heteroscedasticity in the transformed model. This is obvious from (3.2.3) when i=j. Clearly, Savin and White's generalization cannot tackle such situations and hence should be, from this point of view, considered as somewhat restrictive. Our generalization, on the other hand, can easily accommodate such cases. To be more precise, our generalization allows for heteroscedasticity in the transformed linear regression model due not only to the transformation but also to the heteroscedasticity in the original observations i.e., y_i's. This is obvious from our assumption about $V(y_i)$ and (3.2.4). Hence our generalization can be claimed to be more reasonable and appropriate.

Secondly, our formulation (approximately) reduces to that of Savin and White when $(\lambda-1+h) = 0$. Thus, the case when the disturbances in the transformed linear regression actually follow AR(1) is nothing but a special case of our generalization.

Thirdly, in any practical situation the observations relate to y_i's and x_i's and not to the transformed values of the variables. Hence any guess about the nature of the variances and autocovariances in the data must necessarily refer to that based on the original variables and no <u>a priori</u> idea can be made about the variances and

autocovariances of the errors in the transformed model on the basis of the given data.

Fourthly, it is clear from (3.2.3) that there is not much plausibility of assuming a simple structure for Cov $(y_i^{(\lambda)}, y_j^{(\lambda)})$ independent of λ even when Cov (y_i, y_j) has a simple structure.

Finally, our approach brings into focus the complications created by the transformation and its role in modifying the variances and autocovariances of the original dependent variable.

3.3 A Suggested Method of ML Estimation

In this section we describe a method of estimation of the Box-Cox model as generalized above and given by (3.2.1) and (3.2.4). It is clear that since, apart from nonlinearities, the variance-covariance matrix of the disturbances involves $E(y_i)$ etc., estimation of this model is quite difficult. We nevertheless propose a ML method of estimation and indicate how much estimates can be obtained.

Let $\sigma_y^2 V$ denote the variance-covariance matrix of ε obtained from (3.2.4).[3/] Then

$$V = ((v_{ij}))$$

where

$$v_{ij} = \rho^{*|i-j|}\left[E(y_i)\,E(y_j)\right]^{\lambda-1+h}; \qquad \ldots (3.3.1)$$

$$i, j = 1, \ldots, n.$$

3/ Other standard assumptions may be made about Corr (y_i, y_j) and also for $V(y_i)$. The method of estimation, however, will remain, in principle, the same.

The log-likelihood function of $y_1, y_2, \ldots, y_n$ can now be written as

$$L = \text{Const.} - \frac{n}{2} \ln \sigma_y^2 - \frac{1}{2} \ln |V| + \frac{1}{\sigma_y^2}(Y^{(\lambda)} - X\beta)' V^{-1}(Y^{(\lambda)} - X\beta) + (\lambda-1) \sum_{i=1}^{n} \ln y_i \quad \ldots (3.3.2)$$

where

$$Y^{(\lambda)} = (y_1^{(\lambda)}, \ldots, y_n^{(\lambda)})', \quad X = (x_1, x_2, \ldots, x_n)'$$

and V is assumed to be non-singular.

Considering the complications created by the nonlinearity, we suggest a search procedure over λ, ρ^* and δ $(=\lambda-1+h)$ for obtaining the ML estimates of the parameters involved. The procedure may be described in the following steps.

Step 1 : Fix a value for each of λ, ρ^* and δ.

Step 2 : Apply ordinary least squares (which obviously means V=I) and obtain estimates of $E\left[y_i^{(\lambda)}\right]$'s from (3.2.1).

Step 3 : Obtain estimates of $E(y_i)$'s from the following relationship[4]

$$E(y_i) \simeq \begin{cases} \left[1 + \lambda E(y_i^{(\lambda)})\right]^{1/\lambda}, & \text{if } \lambda \neq 0 \\ e^{E(y_i^{(\lambda)})}, & \text{if } \lambda = 0 \end{cases}$$

4/ See footnote 4 of Chapter 2 for derivation of the relationship.

Step 4 : Using the estimates of $E(y_i)$'s in Step 3, form the estimated variance-covariance matrix $\hat{V}(\lambda, \rho^*, \delta)$ for the fixed $(\lambda, \rho^*, \delta)$.

Step 5 : Find estimates of β and σ_y^2 as

$$\hat{\beta}(\lambda,\rho^*, \delta) = (X'\hat{V}^{-1}(\lambda, \rho^*, \delta)X)^{-1} X'\hat{V}^{-1}(\lambda, \rho^*, \delta)Y^{(\lambda)} \quad \ldots (3.3.3)$$

and

$$\hat{\sigma}_y^2 (\lambda, \rho^*, \delta) = \frac{1}{n} (Y - X\hat{\beta}(\lambda, \rho^*, \delta))'\hat{V}^{-1}(\lambda,\rho^*, \delta) (Y - X\hat{\beta}(\lambda, \rho^*, \delta)).$$

Step 6 : Calculate the value of the log-likelihood function at $\hat{\beta}(\lambda, \rho^*, \delta)$ and $\hat{\sigma}_y^2 (\lambda, \rho^*, \delta)$ from (3.3.2) as

$$L(\lambda,\rho^*, \delta) = \text{Const.} - \frac{n}{2} \ln \hat{\sigma}_y^2 (\lambda, \rho^*, \delta) - \frac{1}{2} \ln \left|\hat{V}(\lambda, \rho^*, \delta)\right| + (\lambda-1) \sum_{i=1}^{n} \ln y_i . \quad \ldots (3.3.4)$$

Step 7 : Obtain new estimates of $E(y_i^{(\lambda)})$'s using $\hat{\beta}(\lambda, \rho^*, \delta)$ in (3.3.3) and repeat Steps (3) to (6) till the maximum value of the log-likelihood function for the given $(\lambda, \rho^*, \delta)$ is obtained.

Now, in order to locate the ML estimates of λ, ρ^* and δ one must carry out the above steps for all plausible combinations of values of $(\lambda, \rho^*, \delta)$. For each combination of values one gets the maximum value of the log-likelihood function and the maximum over all these

maxima is indeed the maximum value of the log-likelihood function of the model and the corresponding values of λ, ρ^* and δ, say $\hat{\lambda}$, $\hat{\rho}^*$ and $\hat{\delta}$ and hence of $\hat{\beta}(\hat{\lambda}, \hat{\rho}^*, \hat{\delta})$ and $\hat{\sigma}^2_y(\hat{\lambda}, \hat{\rho}^*, \hat{\delta})$ are under standard conditions[5] the ML estimates of the parameters of the model.

3.4 Consistency of Estimates by Savin-White Procedure under Alternative Formulations

It would be interesting to examine if the ML estimates of the parameters obtained by Savin and White's method for their generalization of the Box-Cox model remain consistent if, in fact, the situation is as described by our formulation. If it happens that the ML estimates in the Savin and White generalization are consistent even under our assumptions for the autocovariance structures, then it does not really matter, from the point of view of consistency, in what way the Box-Cox model is generalized to take account of autocorrelation. In what follows, we shall show that this does not, in general, happen.

To prove this, let us assume, as has been done by Savin and White (1978), that ε_i's follow an AR(1) process given by

$$\varepsilon_i = \rho\varepsilon_{i-1} + u_i, \qquad i = 1, 2, \ldots, n \qquad \ldots (3.4.1)$$

where $|\rho| < 1$ and u_i's are independent with zero mean and same variance σ^2 across observations. We can write the log-likelihood equation of

---

[5] For details, see Dhrymes (1970, Ch. 3), Rao (1973, Ch. 5) etc.

$y_1, y_2, \ldots, y_n$ under the model as given by (3.2.1) and (3.4.1) as[6]

$$L = \text{Const.} - \frac{n}{2} \ln \sigma^2 + \frac{1}{2} \ln (1-\rho^2) - \frac{Q}{2\sigma^2} + (\lambda-1) \sum_{i=1}^{n} \ln y_i \quad \ldots \quad (3.4.2)$$

where

$$Q = \sum_{i=1}^{n} (Z_i^* - \mu_i^*)^2$$

$$Z_1^* = \sqrt{1-\rho^2}\, Z_1, \qquad Z_1 = y_1^{(\lambda)}$$

$$\mu_1^* = \sqrt{1-\rho^2}\, \mu_1, \qquad \mu_1 = x_1'\beta$$

$$Z_i^* = Z_i - \rho Z_{i-1}, \qquad Z_i = y_i^{(\lambda)} \quad \text{for } i = 2, 3, \ldots, n$$

and

$$\mu_i^* = \mu_i - \rho \mu_{i-1}, \qquad \mu_i = x_i'\beta \quad \text{for } i = 2, 3, \ldots, n.$$

When ε_i's have variances and autocovariances different from those given by AR(1), the condition for consistent estimation of λ by ML method of estimation[7] is

$$E\left(\frac{\partial L}{\partial \lambda_o}\right)_{\lambda_o = \lambda} = 0, \text{ at least asymptotically,}$$

where λ_o refers to some arbitrary choice of the transformation patameter.

6/ Without any loss of generality, we carry out the analysis for $\lambda \neq 0$ only because for $\lambda=0$ similar derivation can easily be done.

7/ If λ is consistently estimated, then other parameters are obviously consistently estimated under standard conditions and therefore we check for the consistent estimation of λ only.

Now from (3.4.2),

$$\frac{\partial L}{\partial \lambda_o} = -\frac{1}{\sigma^2} \sum_{i=1}^{n} (Z_i^* - \mu_i^*) \frac{\partial Z_i^*}{\partial \lambda_o} + \sum_{i=1}^{n} \ln y_i$$

where

$$\frac{\partial Z_i^*}{\partial \lambda_o} = \begin{cases} \sqrt{1-\rho^2} \dfrac{\partial Z_1}{\partial \lambda_o} & \text{for } i = 1. \\ \dfrac{\partial Z_i}{\partial \lambda_o} - \rho \dfrac{\partial Z_{i-1}}{\partial \lambda_o} & \text{for } i = 2,3,\ldots, n. \end{cases}$$

Thus,

$$E\left(\frac{\partial L}{\partial \lambda_o}\right)_{\lambda_o=\lambda} = E\left[-\frac{(1-\rho^2)}{\sigma^2} (Z_1 - \mu_1) \frac{\partial Z_1}{\partial \lambda} - \frac{1}{\sigma^2} \sum_{i=2}^{n} \{(Z_i - \mu_i) - \rho(Z_{i-1} - \mu_{i-1})\} \left\{ \frac{\partial Z_i}{\partial \lambda} - \rho \frac{\partial Z_{i-1}}{\partial \lambda} \right\} \right] + E\left(\sum_{i=1}^{n} \ln y_i \right)$$

$$= E\left[-\frac{1-\rho^2}{\sigma^2} (Z_1 - \mu_1) \frac{\partial Z_1}{\partial \lambda} - \frac{1}{\sigma^2} \sum_{i=2}^{n} (Z_i - \mu_i) \frac{\partial Z_i}{\partial \lambda} - \frac{\rho^2}{\sigma^2} \sum_{i=2}^{n} (Z_{i-1} - \mu_{i-1}) \frac{\partial Z_{i-1}}{\partial \lambda} + \frac{\rho}{\sigma^2} \sum_{i=2}^{n} (Z_i - \mu_i) \frac{\partial Z_{i-1}}{\partial \lambda} + \frac{\rho}{\sigma^2} \sum_{i=2}^{n} (Z_{i-1} - \mu_{i-1}) \frac{\partial Z_i}{\partial \lambda} \right]$$

$$+ \frac{1}{\lambda} E\left[\sum_{i=1}^{n} \ln (1 + \lambda Z_i) \right], \qquad \ldots (3.4.3)$$

Now, without any loss of generality, the units of measurement for y_i can be supposed to be such that $\sum_{i=1}^{n} \ln y_i = 0$ and hence,

$$E\left[\sum_{i=1}^{n} \ln (1 + \lambda Z_i)\right] = 0.$$

However, as noted by Schlesselman (1971), in such a case we have to assume that $x_{1i} = 1$ for all $i = 1, 2, \ldots, n$.

Following Zarembka (1974), we also use the approximation[8/]

$$E(Z_i - \mu_i) \frac{\partial Z_i}{\partial \lambda} \simeq \sigma_i^2 E(\ln y_i), \quad \sigma_i^2 = V(\varepsilon_i) \qquad \ldots (3.4.4)$$

Also using derivations shown in Appendix 3.1 we have,

$$\sum_{i=2}^{n} (Z_{i-1} - \mu_{i-1}) \frac{\partial Z_i}{\partial \lambda} \simeq \sum_{i=2}^{n} \sigma_{(i-1)i} E(\ln y_i) + \sum_{i=2}^{n} t_i (\sigma_{(i-1)i} - \rho_i \sigma_i^2) \qquad \ldots (3.4.5)$$

and similarly

$$\sum_{i=2}^{n} (Z_i - \mu_i) \frac{\partial Z_{i-1}}{\partial \lambda} \simeq \sum_{i=2}^{n} \sigma_{i(i-1)} E(\ln y_{i-1}) + \sum_{i=2}^{n} t_{i-1} (\sigma_{i(i-1)} - \bar{\rho}_i \sigma_{i-1}^2)$$

where

$$\sigma_{(i-1)i} = E(\varepsilon_{i-1} \varepsilon_i) = \mathrm{Cov}(\varepsilon_{i-1}, \varepsilon_i) \quad (\text{since } E(\varepsilon_i) = 0 \text{ for all } i)$$

$$t_i = \frac{E(p_i^2)}{2\lambda(1+\lambda \mu_i)^2}$$

$$p_i = \lambda \varepsilon_i$$

$$\rho_i = E(\varepsilon_{i-1} \varepsilon_i^3) / E(\varepsilon_i^4)$$

and

$$\bar{\rho}_i = E(\varepsilon_i \varepsilon_{i-1}^3) / E(\varepsilon_{i-1}^4).$$

8/ The algebra is similar to the one shown in Appendix 3.1 and hence is omitted.

Then substituting (3.4.4) and (3.4.5) in (3.4.3), we have,

$$E\left(\frac{\partial L}{\partial \lambda_o}\right)_{\lambda_o=\lambda} \simeq -\frac{1-\rho^2}{\sigma^2}\,\sigma_1^2\,E(\ln y_1) - \frac{1}{\sigma^2}\sum_{i=2}^{n}\sigma_i^2\,E(\ln y_i)$$

$$-\frac{\rho^2}{\sigma^2}\sum_{i=2}^{n}\sigma_{i-1}^2\,E(\ln y_{i-1})$$

$$+\frac{\rho}{\sigma^2}\sum_{i=2}^{n}\left[\sigma_{(i-1)i}\,E(\ln y_i)+\sigma_{i(i-1)}\,E(\ln y_{i-1})\right]$$

$$+\frac{\rho}{\sigma^2}\sum_{i=2}^{n}\Big[t_i\,(\sigma_{(i-1)i}-\rho_i\,\sigma_i^2)$$

$$+\,t_{i-1}(\sigma_{i(i-1)}-\bar{\rho}_i\,\sigma_{i-1}^2)\Big] \quad ..(3.4.6)$$

$\neq 0$, in general.

Thus we find that the ML estimators of the parameters of the model are, in general, not consistent if the variance-covariance matrix is different from that generated by AR(1). Obviously, for our kind of assumption of autocovariance structures for $y_i^{(\lambda)}$'s, this conclusion remains valid.2/

3.5 Likelihood Ratio Tests

To test different hypotheses concerning the parameters, we suggest the use of likelihood ratio (LR) test. As noted in Chapter 2 the test statistic is defined as

$$l(.) = -2\left[L(H_o) - L(H_1)\right]$$

2/ Similar algebra can be carried out for other types of standard error processes for the disturbance of the transformed linear regression model. However, the conclusions will obviously be the same.

where $L(H_o)$ and $L(H_1)$ are maximum log-likelihood values under the null (H_o) and alternative (H_1) hypotheses. $l(.)$ follows a χ^2 distribution asymptotically with degree of freedom r, the number of additional restrictions imposed by H_o. From the point of view of the present analysis, one would be interested mainly in examining the effect of transformation on autocorrelation and heteroscedasticity and hence one would like to test hypotheses involving λ, ρ^* and δ. Like Savin and White (1978), one can formulate different conditional and unconditional hypotheses. For unconditional null hypotheses, the alternative is always $H_1 : \lambda = \hat{\lambda}, \rho^* = \hat{\rho}^*, \delta = \hat{\delta}$. For conditional hypotheses, the alternatives are suitably defined. Without going into obvious details, we list down the kind of unconditional hypotheses that might be of interest.

Tests	H_o	H_1	χ^2 with d.f.
l(1)	$\lambda = 1$ (i.e., linear)	$\lambda = \hat{\lambda}$	1
l(2)	$\rho^* = 0$ (i.e., no autocorrelation)	$\rho^* = \hat{\rho}^*$	1
l(3)	$\delta = 0$ (i.e., homoscedastic and $\varepsilon_i \sim$ AR(1)).	$\delta = \hat{\delta}$	1
l(4)	$\lambda = 1, \rho^* = 0$ (i.e., linear and no autocorrelation)	$\lambda = \hat{\lambda}, \rho^* = \hat{\rho}^*$	2
l(5)	$\lambda = 1, \delta = 0$ (i.e., linear, homoscedastic and $\varepsilon_i \sim$ AR(1))	$\lambda = \hat{\lambda}, \delta = \hat{\delta}$	2
l(6)	$\rho^* = 0, \delta = 0$ (homoscedastic and no autocorrelation i.e., original Box-Cox case)	$\rho^* = \hat{\rho}^*, \delta = \hat{\delta}$	2
l(7)	$\lambda = 1, \rho^* = 0, \delta = 0$ (i.e., linear, homoscedastic and no autocorrelation)	$\lambda = \hat{\lambda}, \rho^* = \hat{\rho}^*, \delta = \hat{\delta}$	3

In a similar manner, we can set up different conditional hypotheses representing alternative situations.

It is thus clear that one can test for the functional form, and for different patterns of autocorrelation and heteroscedasticity to draw inferences about the true situation for a given set of data on the basis of the generalized model suggested here. In the present context, mention may, in particular, be made about 1(3) which represents the Savin and White formulation i.e., the errors in the transformed linear regression model follows an AR(1) process under the null hypothesis. This test will help one decide if the Savin and White formulation is, in fact, true in a given situation.

3.6 Conclusions

In this Chapter we have argued that the generalization of the Box-Cox model suggested by Savin and White by incorporating a standard error process for the disturbances in the transformed linear regression model is not always proper and appropriate. Like the problem of heteroscedasticity in the Box-Cox model, the problem of autocorrelation also cannot be straightforwardly tackled since it is affected by the transformation parameter as well. We have proposed an alternative generalization which considers the structure of the variance-covariance matrix of the disturbances in the transformed linear regression model as derived from the assumed autocorrelation and variance structures of the original observations and the transformation parameter. Savin and White generalization turns out to be a special case of our model. We have also indicated [illegible] method of estimation of the parameters of such a model and have

shown that the estimate of λ and hence of other parameters obtained through the Savin and White procedure are not, in general, consistent when the disturbances in the transformed dependent variable have a structure of variances and autocovariances different from that in AR(1). We have also pointed out that standard likelihood ratio tests can be used in our model for testing hypotheses concerning λ, ρ^* and δ i.e., for identifying situations having different functional forms, and patterns of autocorrelation and heteroscedasticity. As a specific case, one may test if the Savin and White formulation is realistic in a given situation.

We may repeat that exactly similar conclusions will be obtained if independent variables are also assumed to be transformed. However, if the transformation parameter(s) for independent variable(s) is (are) different from that of the dependent variable, the algebra is section 3.4 becomes more cumbersome though the conclusion remains unchanged. Also the estimation procedure becomes more lengthy and time-consuming.

Appendix 3.1

Derivation of the Relationships in (3.4.5)

By definition

$$Z_i = \frac{y_i^{\lambda} - 1}{\lambda} \qquad \ldots \quad (A\ 3.1.1)$$

i.e., $y_i^{\lambda} = 1 + \lambda Z_i$.

Thus,

$$y_i^{\lambda} \ln y_i = Z_i + \lambda \frac{\partial Z_i}{\partial \lambda} .$$

Hence

$$\frac{\partial Z_i}{\partial \lambda} = \frac{1}{\lambda^2} \left[(1+\lambda Z_i) \ln(1+\lambda Z_i) - \lambda Z_i \right] \qquad \ldots \quad (A\ 3.1.2)$$

Let $p_i = \lambda \varepsilon_i$,

then

$$E\left[\sum_{i=2}^{n} (Z_{i-1} - \mu_{i-1}) \frac{\partial Z_i}{\partial \lambda} \right]$$

$$= \frac{1}{\lambda} E\left[\sum_{i=2}^{n} p_{i-1} \frac{\partial Z_i}{\partial \lambda} \right]$$

$$= \frac{1}{\lambda^3} E\left[\sum_{i=2}^{n} p_{i-1} \{(1+\lambda\mu_i+p_i)\ln(1+\lambda\mu_i+p_i) - (\lambda\mu_i + p_i)\} \right]$$

since from (3.2.1), $\lambda Z_i = \lambda\mu_i + \lambda\varepsilon_i = \lambda\mu_i + p_i$.

Letting

$$q_j = \frac{p_i}{(1 + \lambda\mu_i)}$$

we have from above,

$$E\left[\sum_{i=2}^{n}(Z_{i-1}-\mu_{i-1})\frac{\partial Z_i}{\partial\lambda}\right]$$

$$=\frac{1}{\lambda^3}E\left[\sum_{i=2}^{n}p_{i-1}\,p_i\,\frac{(1+q_i)}{q_i}\{\ln(1+q_i)+\ln(1+\lambda\mu_i)\}\right]$$

$$-\frac{1}{\lambda}\sum_{i=2}^{n}\sigma_{(i-1)i} \qquad \ldots(A\ 3.1.3)$$

where $\sigma_{(i-1)i}=\mathrm{Cov}(\varepsilon_{i-1},\varepsilon_i)=E(\varepsilon_{i-1}\,\varepsilon_i)$.

Since for the transformation to be sensible, $y_i>0$ that is, $Z_i+\frac{1}{\lambda}>0$ so that $(\mu_i+\frac{1}{\lambda})/\sigma_i=(1+\lambda\mu_i)/\lambda\sigma_i\gg 1$, [Zarembka (1974), pp. 89-90] one can treat q_i's to be small and hence we can expand $\ln(1+q_i)$ in power series. Keeping terms upto q_i^2, we have from (A 3.1.3)

$$E\left[\sum_{i=2}^{n}(Z_{i-1}-\mu_{i-1})\frac{\partial Z_i}{\partial\lambda}\right]$$

$$\simeq\frac{1}{\lambda^3}E\left[\sum_{i=2}^{n}p_{i-1}\,p_i\left(1+\frac{q_i}{2}-\frac{q_i^2}{6}\right)+\sum_{i=2}^{n}p_{i-1}\,p_i\ln(1+\lambda\mu_i)\right]$$

$$-\frac{1}{\lambda}\sum_{i=2}^{n}\sigma_{(i-1)i}$$

$$=\frac{1}{\lambda^3}E\left[\sum_{i=2}^{n}\frac{p_{i-1}\,p_i^2}{2(1+\lambda\mu_i)}-\sum_{i=2}^{n}\frac{p_{i-1}p_i^3}{6(1+\lambda\mu_i)^2}\right]+\frac{1}{\lambda}\sum_{i=2}^{n}\sigma_{(i-1)i}\ln(1+\lambda\mu_i)$$

$$=\sum_{i=2}^{n}\frac{E(\varepsilon_{i-1}\,\varepsilon_i^2)}{2(1+\lambda\mu_i)}-\lambda\sum_{i=2}^{n}\frac{E(\varepsilon_{i-1}\,\varepsilon_i^3)}{6(1+\lambda\mu_i)^2}$$

$$+\frac{1}{\lambda}\sum_{i=2}^{n}\sigma_{(i-1)i}\ln(1+\lambda\mu_i).$$

Since ε_i's are assumed to be autocorrelated, we use the approximation

$$E(\varepsilon_{i-1}\ \varepsilon_i^2) \simeq \delta_i\ E(\varepsilon_i^3)$$

and $$E(\varepsilon_{i-1}\ \varepsilon_i^3) \simeq \rho_i\ E(\varepsilon_i^4)$$

where ρ_i and δ_i are accordingly defined.

Then the above expression becomes

$$E\left[\sum_{i=2}^{n}(Z_{i-1}-\mu_{i-1})\frac{\partial Z_i}{\partial\lambda}\right]$$

$$=\sum_{i=2}^{n}\frac{\delta_i\ E(\varepsilon_i^3)}{2(1+\lambda\mu_i)}-\lambda\sum_{i=2}^{n}\frac{\rho_i\ E(\varepsilon_i^4)}{6(1+\lambda\mu_i)^2}+\frac{1}{\lambda}\sum_{i=2}^{n}\sigma_{(i-1)i}\ln(1+\lambda\mu_i)$$

$$=\sum_{i=2}^{n}\frac{\delta_i\ E(\varepsilon_i^3)}{2(1+\lambda\mu_i)}-\lambda\sum_{i=2}^{n}\left[\frac{E(\varepsilon_i^4)}{\{E(\varepsilon_i^2)\}^2}-3\right]\frac{\rho_i\{E(\varepsilon_i^2)\}^2}{6(1+\lambda\mu_i)^2}$$

$$-\frac{3\lambda}{6}\sum_{i=2}^{n}\frac{\rho_i\{E(\varepsilon_i^2)\}^2}{(1+\lambda\mu_i)^2}+\frac{1}{\lambda}\sum_{i=2}^{n}\sigma_{(i-1)i}\ln(1+\lambda\mu_i).$$

... (A 3.1.4)

But since for normal distribution, third central moment and kurtosis are zero, we have from (A 3.1.4)

$$E\left[\sum_{i=2}^{n}(Z_{i-1}-\mu_{i-1})\frac{\partial Z_i}{\partial\lambda}\right]=-\frac{\lambda}{2}\sum_{i=2}^{n}\frac{\rho_i\{E(\varepsilon_i^2)\}^2}{(1+\lambda\mu_i)^2}+\frac{1}{\lambda}\sum_{i=2}^{n}\sigma_{(i-1)i}\ln(1+\lambda\mu_i)$$

$$=-\frac{1}{2\lambda}\sum_{i=2}^{n}\frac{\sigma_i^2\rho_i\{E(p_i^2)\}}{(1+\lambda\mu_i)^2}$$

$$+\frac{1}{\lambda}\sum_{i=2}^{n}\sigma_{(i-1)i}\ln(1+\lambda\mu_i)\ldots.\ (A\ 3.1.5)$$

Now, since

$$\lambda E(\ln y_i) = E\left[\ln (1 + \lambda Z_i)\right]$$

$$= E\left[\ln \left(1 + \frac{p_i}{1 + \lambda\mu_i}\right)\right] + \ln(1 + \lambda\mu_i)$$

$$\simeq E\left[\frac{p_i}{(1 + \lambda\mu_i)}\right] - E\left[\frac{p_i^2}{2(1 + \lambda\mu_i)^2}\right] + \ln (1 + \lambda\mu_i),$$

we have,

$$\frac{1}{\lambda} \ln(1 + \lambda\mu_i) = \frac{1}{\lambda} \frac{E(p_i^2)}{2(1 + \lambda\mu_i)^2} + E(\ln y_i).$$

Therefore, we have from (A 3.1.5),

$$E\left[\sum_{i=2}^{n} (Z_{i-1} - \mu_{i-1}) \frac{\partial Z_i}{\partial \lambda}\right] = -\frac{1}{2\lambda} \sum_{i=2}^{n} \frac{p_i^2 \rho_i \{E(p_i^2)\}}{(1 + \lambda\mu_i)^2}$$

$$+ \frac{1}{2\lambda} \sum_{i=2}^{n} \sigma_{(i-1)i} \frac{\{E(p_i^2)\}}{(1 + \lambda\mu_i)^2}$$

$$+ \sum_{i=2}^{n} \sigma_{(i-1)i} E(\ln y_i)$$

$$= \frac{1}{2\lambda} \sum_{i=2}^{n} (\sigma_{(i-1)i} - \sigma_i^2 \rho_i) \frac{E(p_i^2)}{(1+\lambda\mu_i)^2}$$

$$+ \sum_{i=2}^{n} \sigma_{(i-1)i} E(\ln y_i).$$

The other relation in (3.4.5) can be derived in an exactly similar fashion.

Chapter 4

Nonnormality, Heteroscedasticity and the Box-Cox Transformation

4.1 Introduction

We discussed at length the problems of heteroscedasticity and autocorrelation in the context of the Box-Cox transformation in the last two Chapters. The present Chapter is devoted to studying the Box-Cox model with particular reference to the assumption of normality of the distribution of the transformed dependent variable. Although most of the studies on the Box-Cox transformation assumes normality of this distribution, it is easy to see that, strictly speaking, $y^{(\lambda)}$ as defined by the transformation

$$y^{(\lambda)} = \begin{cases} (y^{\lambda}-1)/\lambda, & \text{if } \lambda \neq 0 \\ \ln y, & \text{if } \lambda = 0 \end{cases}$$

cannot follow a normal distribution for $\lambda \neq 0$. It has been rightly pointed out by Poirier (1978a) that for the transformation to be well-defined, y must be greater than zero and hence

$$y^{(\lambda)} > -\frac{1}{\lambda}, \quad \text{if } \lambda > 0$$

$$\text{and } y^{(\lambda)} < -\frac{1}{\lambda}, \quad \text{if } \lambda < 0.$$

Clearly then $y^{(\lambda)}$ cannot, in general, have values in the range $-\infty$ to ∞ and therefore it is not justified to assume $y^{(\lambda)}$ to follow a normal distribution when $\lambda \neq 0$. For $\lambda = 0$, however, ln y can take values in

the range $-\infty$ to ∞ and hence the assumption of normality is justified in this case. In view of this, Poirier (1978a) used ML method of estimation in the Box-Cox model by assuming truncated normality of the distribution of $y_i^{(\lambda)}$'s.

However, Poirier, like Box and Cox, assumed homoscedasticity of the $y_i^{(\lambda)}$'s across observations in estimating the model. But, as we have already stressed, this assumption also is not, in general, valid. We want to emphasize, therefore, that one ought to consider the transformed dependent variable in the Box-Cox transformation model to follow a truncated normal distribution with heteroscedasticity. The main purpose of this Chapter is to consider and study such a model, henceforth referred to as the Box-Cox nonnormal heteroscedastic model (BCNNHM).

We may point out in this connection that when $\lambda=1$, the Box-Cox model as generalized by Poirier reduces to the well-known limited dependent variable model developed by Tobin (1958) (excepting for the fact that one gets added to the constent term). Some years after the publication of Tobin's paper, Amemiya (1973a) reconsidered the model and proved a number of results for this model. He proved that there exists a strong consistent root[1/] of the normal equations of Tobin's model which is also asymptotically normal. He also showed that the initial estimators proposed by Tobin for the iterative procedure for finding the ML estimator are inconsistent and suggested a consistent initial

1/ Throughout this Chapter, by a 'root of the normal equations' we shall mean that it is a solution of the normal equations corresponding to a local maximum of the log-likelihood function.

estimator of the parameters. It is further shown that the second-round estimator obtained from the Newton-Raphson method would, in fact, be asymptotically equivalent to the ML-estimator if the first round estimator is consistent. More recently, Oslen (1978) has shown that there is a unique single maximum of the ML function of Tobin's model and hence, whatever be the initial estimator, if the iterative process converges, then the converging value will correspond to the global maximum of the ML function and hence will be the ML estimator.

In generalizing the Box-Cox model by introducing heteroscedasticity of the transformed dependent variable whose distribution is truncated normal, we shall also be generalizing Tobin model. However, since Tobin model can be used without any direct reference to the question of transformation and since a relatively simple method of obtaining consistent initial estimates (under a specific assumption about heteroscedasticity) for use in Newton-Raphson method can be developed even when heteroscedasticity is intorduced in this model, we propose to consider Tobin's limited dependent variable model as generalized by incorporating heteroscedasticity separately in view of its distinct interest. This model will be referred to as the GLDVM (i.e., generalized limited dependent variable model).

The order of presentation in this Chapter is as follows : In section 4.2 we describe the generalization of the Box-Cox model in the direction stated above and also state the assumptions and earlier results to be used in our analysis. In section 4.3, we consider the GLDVM and prove, following Amemiya's approach, that the normal equations

of the GLDVM have a stong consistent root which follows asymptotically normal distribution. We prove this result for two different assumptions regarding the heteroscedasticity of the original observations. We then extent Amemiya's method of obtaining a consistent initial estimator of the parameters under one of the two specific assumptions about the heteroscedasticity of y_i's. Section 4.4 gives a proof of the existence of a strong consistent root of the normal equations for the BCNNHM. This root also has an asymptotically normal distribution. We also suggest the use of standard techniques of nonlinear estimation for obtaining the ML estimates. The Chapter ends with some concluding observations in section 4.5.

4.2 The Box-Cox Nonnormal Heteroscedastic Model (BCNNHM)

4.2.1 The Model

We define the Box-Cox generalized model[2/, 3/] as

$$y_i^{(\lambda_o)} = \begin{cases} \tilde{\beta}_o' x_i + \varepsilon_i & \text{if R.H.S.} > -\frac{1}{\lambda_o} \text{ when } \lambda_o > 0 \\ & \text{or } < -\frac{1}{\lambda_o} \text{ when } \lambda_o < 0 \\ 0 & \text{otherwise} \end{cases} \quad \ldots(4.2.1)$$

$$(i = 1, 2, \ldots, n)$$

2/ Here λ_o is the value of λ for the correct transformation. The subscript 'o' is added to other parameters for the same consideration.

3/ Alternatively, one can use the same model but analyse only those observations for which

$$y_i^{(\lambda_o)} = \tilde{\beta}_o' x_i + \varepsilon_i \quad \text{if R.H.S.} > -\frac{1}{\lambda_o} \text{ when } \lambda_o > 0 \text{ or } < -\frac{1}{\lambda_o} \text{ when } \lambda_o < 0.$$

However, it can easily be seen that such a treatment would not change any of the principal results presented here.

Since $y_i^{(\lambda_o)}$ at $\lambda_o = 0$ can reasonably be assumed to follow normal distribution, this case does not come under the perview of our study.

where $y_i^{(\lambda_o)} = (y_i^{\lambda_o} - 1)/\lambda_o$ is the transformed dependent variable on the ith observation,

$x_i = (x_{1i}, x_{2i}, \ldots, x_{ki})'$ is a (k x 1) column vector of the ith observation on k (fixed) regressors,

$\tilde{\beta}_o' = (\beta_{1o}, \beta_{2o}, \ldots, \beta_{ko})$ is the (1 x k) row vector of associated regression coefficients,

and ε_i's are independent disturbances following normal distributions with zero mean and variance σ_{oi}^2.

We assume $x_{1i} = 1$ for all $i = 1, 2, \ldots, n$ [cf. Schlesselman (1971)]

For convenience, we rewrite (4.2.1) as

$$\tilde{y}_i^{(\lambda_o)} = \begin{cases} \beta_o' x_i + \varepsilon_i & \text{if R. H. S. } > 0 \text{ when } \lambda_o > 0 \text{ or} \\ & \quad < 0 \text{ when } \lambda_o < 0 \\ 0 & \text{otherwise} \end{cases} \quad \ldots (4.2.2)$$

$(i = 1, 2, \ldots, n)$

where $\tilde{y}_i^{(\lambda_o)} = y_i^{\lambda_o}/\lambda_o$

and $\beta_o' = (\beta_{1o} + \frac{1}{\lambda_o}, \beta_{2o}, \ldots, \beta_{ko})$.

As noted in Chapter 2,

$$\sigma_{oi}^2 \left[= V(y_i^{(\lambda_o)}) \right]$$

will be of the form

$$\sigma_{oi}^2 \simeq V(y_i) \left[E(y_i)\right]^{2\lambda_o - 2} \quad \ldots (4.2.3)$$

Let us assume that

$$V(y_i) = \sigma_o^2 \left[E(y_i)\right]^{h_o}, \quad E(y_i) > 0 \text{ [4]}$$

where σ_o^2 and h_o are unknown constants. Then

---

[4] Other standard assumptions about $V(y_i)$ can also be made.

$$\sigma_{oi}^2 \simeq \sigma_o^2 \left[E(y_i)\right]^{2\lambda_o - 2 + h_o} . \qquad \ldots (4.2.4)$$

(4.2.4) can be reduced further to[5]

$$\sigma_{oi}^2 \simeq \sigma_o^2 Z_{oi}^{\delta_o} \qquad \ldots (4.2.5)$$

where $Z_{oi} = \lambda_o \mu_{oi}$,

$$\mu_{oi} = E\left[\tilde{y}_i^{(\lambda_o)}\right],$$

and $\delta_o = \dfrac{2\lambda_o - 2 + h_o}{\lambda_o}$.

Our problem then is to estimate $\theta_o' = (\beta_o', \sigma_o^2, \delta_o, \lambda_o)$ on the basis of the observations $(y_1, y_2, \ldots, y_n)$.

5/ $\tilde{y}_i^{(\lambda)} = y_i^{\lambda} / \lambda$

i.e. $y_i = \left[\lambda \tilde{y}_i^{(\lambda)}\right]^{\frac{1}{\lambda}}$.

Using Taylor expansion around $E(\tilde{y}_i^{(\lambda)})$, we have

$$y_i \simeq \left[\lambda E(\tilde{y}_i^{(\lambda)})\right]^{\frac{1}{\lambda}} + \left[\tilde{y}_i^{(\lambda)} - E(\tilde{y}_i^{(\lambda)})\right] \frac{\partial \left[\lambda \tilde{y}_i^{(\lambda)}\right]^{\frac{1}{\lambda}}}{\partial \tilde{y}_1^{(\lambda)}} \Bigg|_{\tilde{y}_i^{(\lambda)} = E(\tilde{y}_i^{(\lambda)})}$$

and hence

$$E(y_i) \simeq \left[\lambda E(\tilde{y}_i^{(\lambda)})\right]^{\frac{1}{\lambda}}.$$

4.2.2 Assumptions

(i) The parameter space Ψ of $\theta = (\beta', \sigma^2, \delta, \lambda)'$ is compact. It does not contain the region $\sigma^2 \leq 0$ but contains an open neighbourhood of θ_0.

(ii) x_i is bounded and the empirical distribution function, say H_n, defined as

$$H_n(x) = j/n$$

where j is the number of points $x_1, x_2, \ldots, x_n$ less than or equal to x, converges to a distribution function, say H.

(iii) $\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} x_i x_i'$ is positive definite.

(iv) $\lambda \mu_i > 0$ for all $i = 1, 2, \ldots, n$.

4.2.3 Previous Results

We now set down some results and lemmas which will be used in our study. Two of these lemmas are originally due to Jennrich (1969) and have been used by Amemiya (1973a). The proofs can be found in Jennrich (1969) and Amemiya (1973a).

Lemma 4.1 : Let X be a Euclidean space and Ψ a compact subset of a Euclidean space. If g is a bounded and continuous function on X x Ψ and if $\{H_n\}$ is a sequence of distribution functions on X which converge to a distribution function H, then

$$\int g(x, \theta)\, dH_n(x) \longrightarrow \int g(x, \theta)\, dH(x)$$

uniformly for all θ in Ψ.

Lemma 4.2 : Let $\{\eta_i\}$, i = 1, 2, ..., n, be independent random variables with mean $\bar{\eta}_i$ and first four moments uniformly bounded in i. Let $\{g_i\}$ be a sequence of continuous functions on a compact set Ψ such that

$$\frac{1}{n} \sum_{i=1}^{n} g_i(\theta_1)\, g_i(\theta_2)$$

converges uniformly for θ_1 and θ_2 in Ψ. Then

$$\frac{1}{n} \sum_{i=1}^{n} g_i(\theta)\, (\eta_i - \bar{\eta}_i)$$

converges to 0 a. e. uniformly for all θ in Ψ.

Lemma 4.3 : Let $Q_n(\omega, \theta)$ be a measurable function on a measurable space Ω and for each ω in Ω a continuous function for θ in a compact set Ψ. Then there exists a measurable function $\hat{\theta}_n(\omega)$ such that

$$Q_n\left[\omega, \hat{\theta}_n(\omega)\right] = \sup_{\theta \in \Psi} Q_n(\omega, \theta) \text{ for all } \omega \text{ in } \Omega .$$

If $Q_n(\omega, \theta)$ converges to $Q(\theta)$ a.e. uniformly for all θ in Ψ, and if $Q(\theta)$ has a unique maximum at $\theta_o \in \Psi$, then $\hat{\theta}_n$ converges to θ_o a.e.

Lemma 4.4 : Let $f_n(\omega, \theta)$ be a measurable function on a measurable space Ω and for each ω in Ω a continuous function for θ in a compact set Ψ. If $f_n(\omega, \theta)$ converges to $f(\theta)$ a.e. uniformly for all θ in Ψ, and if $\hat{\theta}_n(\omega)$ converges to θ_o a.e., then $f_n\left[\omega, \hat{\theta}_n(\omega)\right]$ converges to $f(\theta_o)$ a.e.

Both Lemmas 4.3 and 4.4 hold if we change the expression "a. e." to "with probability approaching 1".

We shall also need the following expressions for differential coefficients to differentiate the log-likelihood function with respect to the parameters.[6/]

$$\frac{\partial F_i}{\partial \beta} = f_i x_i \left(1 - \frac{\delta}{2}\right)$$

$$\frac{\partial f_i}{\partial \beta} = \left[- \frac{\mu_i f_i}{\sigma_i^2} + \frac{\delta \mu_i f_i}{2\sigma_i^2} - \frac{\delta f_i}{2\mu_i} \right] x_i$$

$$\frac{\partial F_i}{\partial \sigma^2} = - \frac{\mu_i f_i}{2\sigma^2}$$

$$\frac{\partial f_i}{\partial \sigma^2} = \frac{\mu_i^2 - \sigma_i^2}{2\sigma_i^2 \sigma^2} f_i$$

$$\frac{\partial F_i}{\partial \delta} = - \frac{1}{2} \mu_i f_i \ln Z_i$$

$$\frac{\partial f_i}{\partial \delta} = \frac{\mu_i^2 - \sigma_i^2}{2\sigma_i^2} f_i \ln Z_i$$

$$\frac{\partial F_i}{\partial \lambda} = \frac{-\delta \mu_i f_i}{2\lambda} = \frac{-\delta \mu_i^2 f_i}{2Z_i}$$

$$\frac{\partial f_i}{\partial \lambda} = \frac{\delta f_i}{2\lambda}\left(\frac{\mu_i^2 - \sigma_i^2}{\sigma_i^2}\right) = \frac{\delta f_i \mu_i}{2Z_i}\left(\frac{\mu_i^2 - \sigma_i^2}{\sigma_i^2}\right)$$

... (4.2.6)

6/ It should be noted that σ_i^2 here is that defined in (4.2.5) and therefore depends on β. More generally, all the symbols appearing in F_i and f_i have the same meaning as in the model formulated here.

where

$$F_i = F_i(\beta' x_i, \sigma_i^2) = \int_{-\infty}^{\beta' x_i} \frac{1}{\sqrt{2\pi\sigma_i^2}} e^{-\frac{1}{2}(\alpha/\sigma_i)^2} d\alpha$$

and $$f_i = f_i(\beta' x_i, \sigma_i^2) = \frac{1}{\sqrt{2\pi\sigma_i^2}} e^{-\frac{1}{2}(\beta' x_i/\sigma_i)^2}$$

4.3 The Generalized Limited Dependent Variable Model (GLDVM)

We shall study in this section the limited dependent variable model of Tobin as generalized here by incorporating heteroscedasticity in the disturbances. It is evident from (4.2.2) that this GLDVM is a special case of BCNHM with $\lambda_o = 1$.[7] It is also clear from (4.2.3) that for this case $\sigma_{oi}^2 = V(y_i)$. Following Kmenta (1971) we first assume that $\sigma_{oi}^2 = \sigma_o^2 m_i^{\delta_o}$ where m_i's are exogenously given and σ_o^2 and δ_o are unknown parameters. Our problem is to estimate $\theta_o' = (\beta_o', \sigma_o^2, \delta_o)$ from $(y_1, y_2, \ldots, y_n)$. The strong consistency of the ML estimator is proved in subsection 4.3.1 while the estimation

---

7/ Tobin (1958), in fact, considered in his model

$$y_i = \begin{cases} \beta' x_i + \varepsilon_i & \text{if R. H. S.} > \alpha_i \\ 0 & \text{otherwise} \end{cases}$$

where α_i's are known constants. But as Amemiya (1973a) has noted such a model can easily be analysed with slight modification of our results where instead of y_i we now have $y_i^* = y_i - \alpha_i$, $x_i^{*\prime} = (x_i', \alpha_i)$ and $\beta^{*\prime} = (\beta', 1)$.

procedure is outlined in subsection 4.3.2. In subsection 4.3.3 we demonstrate the strong consistency of the ML estimator for a different structure of the variance viz., that given by (4.2.5).[8/]

4.3.1 Strong Consistency of ML Estimator When $V(y_i) = \sigma_o^2 m_i^{\delta_o}$

Let s denote the set of observations for which $y_i = 0$ i.e., $s = \{y_i : y_i = 0\}$. Suppose further that s has n_1 elements. Then the complementary set of s, say $\bar{s}$, where $\bar{s} = \{y_i : y_i > 0\}$ has $n - n_1$ elements. We shall, like Tobin (1958) write the log-likelihood function as[9/]

$$L = \text{Const.} + \sum_{s} \ln (1 - F_i) - \frac{1}{2} \sum_{\bar{s}} \ln \sigma_i^2 - \frac{1}{2} \sum_{\bar{s}} \left(\frac{y_i - \beta' x_i}{\sigma_i} \right)^2 . \qquad \ldots (4.3.1)$$

With our assumption about σ_i^2, (4.3.1) can be written as

$$L = \text{Const.} + \sum_{s} \ln (1 - F_i) - \frac{1}{2} \sum_{\bar{s}} \ln \sigma^2 - \frac{\delta}{2} \sum_{\bar{s}} \ln m_i - \frac{1}{2\sigma^2} \sum_{\bar{s}} \left(\frac{y_i - \beta' x_i}{m_i^{\delta/2}} \right)^2 . \qquad \ldots (4.3.2)$$

8/ In this Chapter the symbols δ_o and θ_o (as also δ and θ) are being used without any distinction for both the GLDVM and BCNNHM according to definitions in sections 4.3 and 4.2 respectively.

9/ Obviously, for GLDVM λ is positive and hence the log-likelihood function has been written for $(y_1, \ldots, y_n)$ where

$$y_i = \begin{cases} \beta' x_i + \varepsilon_i & \text{if R. H. S.} > 0 \\ 0 & \text{otherwise.} \end{cases}$$

Following Amemiya (1973a) and using the Assumptions and Lemmas stated in the previous section and expressions for the first two raw moments of the truncated normal distribution $\frac{1}{n}L$ can be shown from (4.3.2) to converge a.e. uniformly for all θ in Ψ to Q defined as

$$Q = \lim_{n\to\infty} \frac{1}{n} \sum_{1}^{n} \Big[\{\ln (1 - F_i)\} (1 - F_{oi}) - \frac{1}{2} (\ln \sigma^2) F_{oi} - \frac{\delta}{2} (\ln m_i) F_{oi} - \frac{m_i^{-\delta}}{2\sigma^2} \{((\beta_o - \beta)' x_i)^2 F_{oi} + 2\sigma_{oi}^2 (\beta_o - \beta)' x_i f_{oi} + \sigma_{oi}^2 (F_{oi} - \beta_o' x_i f_{oi})\} \Big] \quad \ldots(4.3.3)$$

Because of Assumptions (i) and (ii) and Lemma 4.1, the series of both the first and second derivatives of all the terms of Q converge uniformly and we can, therefore, interchange the summation and derivative operation. We thus have the following first-order differential coefficients of Q with respect to the parameters. To evaluate these derivatives we use (4.2.6) with Z_i's being replaced by m_i's and with $\frac{\partial F_i}{\partial \beta} = f_i x_i$ and $\frac{\partial f_i}{\partial \beta} = \frac{-\mu_i f_i x_i}{\sigma_i^2}$.

$$\frac{\partial Q}{\partial \beta} = \lim_{n\to\infty} \frac{1}{n} \sum_{1}^{n} \Big[- \frac{1 - F_{oi}}{1 - F_i} f_i x_i + \frac{m_i^{-\delta}}{\sigma^2} \{ (\beta_o - \beta)' x_i F_{oi} x_i + \sigma_{oi}^2 f_{oi} x_i \} \Big] \quad \ldots(4.3.4)$$

$$\frac{\partial Q}{\partial \sigma^2} = \frac{1}{2\sigma^2} \lim_{n\to\infty} \frac{1}{n} \sum_{1}^{n} \Big[\frac{1-F_{oi}}{1-F_i} \beta' x_i f_i - F_{oi} + \frac{m_i^{-\delta}}{\sigma^2} \{ ((\beta_o - \beta)' x_i)^2 F_{oi} + 2\sigma_{oi}^2 (\beta_o - \beta)' x_i f_{oi} + \sigma_{oi}^2 (F_{oi} - \beta_o' x_i f_{oi}) \} \Big]. \quad \ldots \quad (4.3.5)$$

and $$\frac{\partial Q}{\partial \delta} = \lim_{n\to\infty} \frac{1}{n} \sum_{1}^{n} \left[\frac{1 - F_{oi}}{1 - F_i} \frac{\ln m_i}{2} \beta' x_i f_i - \frac{1}{2} (\ln m_i) F_{oi} \right.$$

$$+ \frac{m_i^{-\delta}}{2\sigma^2} \ln m_i \{ ((\beta_o - \beta)' x_i)^2 F_{oi} + 2\sigma_{oi}^2 (\beta_o - \beta)' x_i f_{oi}$$

$$\left. + \sigma_{oi}^2 (F_{oi} - \beta_o' x_i f_{oi}) \} \right] \quad \ldots (4.3.6)$$

From (4.3.4), (4.3.5) and (4.3.6) it easily follows that

$$\frac{\partial Q(\theta_o)'}{\partial \theta} = 0. \quad \ldots (4.3.7).$$

We now obtain the second-order derivatives[10/] and evaluate them at θ_o to obtain the following relations :

$$\frac{\partial^2 Q(\theta_o)}{\partial \beta \partial \beta'} = \lim_{n\to\infty} \frac{1}{n} \sum_{1}^{n} \left[\frac{\beta_o' x_i}{\sigma_{oi}^2} f_{oi} - \frac{1}{1-F_{oi}} f_{oi}^2 \right.$$

$$\left. - \frac{1}{\sigma_o^2} m_i^{-\delta_o} F_{oi} \right] x_i x_i' \quad \ldots (4.3.8)$$

$$\frac{\partial^2 Q(\theta_o)}{\partial (\sigma^2)^2} = \frac{1}{2\sigma_o^2} \lim_{n\to\infty} \frac{1}{n} \sum_{1}^{n} \left[-\frac{1}{\sigma_o^2} \beta_o' x_i f_{oi} - \frac{1}{1-F_{oi}} \frac{(\beta_o' x_i f_{oi})^2}{2\sigma_o^2} \right.$$

$$+ \beta_o' x_i f_{oi} \frac{(\beta_o' x_i)^2 - \sigma_{oi}^2}{2\sigma_{oi}^2 \sigma_o^2} + \frac{F_{oi}}{\sigma_o^2}$$

$$\left. - \frac{2}{\sigma_o^4} m_i^{-\delta_o} \sigma_{oi}^2 (F_{oi} - \beta_o' x_o f_{oi}) \right]. \quad \ldots (4.3.9)$$

10/ For the exact form of the derivatives, see Appendix 4.1.

$$\frac{\partial^2 Q(\theta_o)}{\partial \delta^2} = \lim_{n\to\infty} \frac{1}{n} \sum_{1}^{n} \left[- \frac{1}{1-F_{oi}} \left(\frac{\ln m_i}{2} \right)^2 (\beta_o' x_i f_{oi})^2 \right.$$

$$+ \left(\frac{\ln m_i}{2} \right) \beta_o' x_i + \frac{(\beta_o' x_i)^2 - \sigma_{oi}^2}{2\sigma_{oi}^2} f_{oi} \ln m_i$$

$$\left. - \frac{1}{2\sigma_o^2} m_i^{-\delta_o} (\ln m_i)^2 \sigma_{oi}^2 (F_{oi} - \beta_o' x_i f_{oi}) \right] \quad \ldots(4.3.10)$$

$$\frac{\partial^2 Q(\theta_o)}{\partial \sigma^2 \partial \delta} = \frac{1}{2\sigma_o^2} \lim_{n\to\infty} \frac{1}{n} \sum_{1}^{n} \left[- \frac{1}{1-F_{oi}} \frac{\ln m_i}{2} (\beta_o' x_i f_{oi})^2 \right.$$

$$+ \beta_o' x_i \frac{(\beta_o' x_i)^2 - \sigma_{oi}^2}{2\sigma_{oi}^2} f_{oi} (\ln m_i)$$

$$\left. - \frac{1}{\sigma_o^2} \{ m_i^{-\delta_o} (\ln m_i) \sigma_{oi}^2 (F_{oi} - \beta_o' x_i f_{oi}) \} \right] \quad \ldots(4.3.11)$$

$$\frac{\partial^2 Q(\theta_o)}{\partial \beta \partial \sigma^2} = \frac{1}{2\sigma_o^2} \lim_{n\to\infty} \frac{1}{n} \sum_{1}^{n} \left[\frac{\beta_o' x_i f_{oi}^2 x_i}{1 - F_{oi}} - \frac{(\beta_o' x_i)^2 - \sigma_{oi}^2}{\sigma_{oi}^2} f_{oi} x_i \right.$$

$$\left. - \frac{2}{\sigma_o^2} m_i^{-\delta_o} \sigma_{oi}^2 f_{oi} x_i \right] \quad \ldots(4.3.12)$$

and

$$\frac{\partial^2 Q(\theta_o)}{\partial \beta \partial \delta} = \lim_{n\to\infty} \frac{1}{n} \sum_{1}^{n} \left[\frac{1}{1-F_{oi}} \frac{\ln m_i}{2} \beta_o' x_i f_{oi}^2 x_i \right.$$

$$- \frac{(\beta_o' x_i)^2 - \sigma_{oi}^2}{2\sigma_{oi}^2} f_{oi} (\ln m_i) x_i$$

$$\left. - \frac{1}{\sigma_o^2} m_i^{-\delta_o} (\ln m_i) \sigma_{oi}^2 f_{oi} x_i \right]. \quad \ldots (4.3.13)$$

Defining

$$w_i = \frac{\beta_o' x_i}{\sigma_{oi}} = \frac{\beta_o' x_i}{\sigma_o m_i^{\delta_o/2}}$$

and writing the standard normal density and distribution functions evaluated at w_i as $\Phi_i (\equiv \sigma_{oi} f_{oi})$ and $\phi_i (\equiv F_{oi})$ respectively, we have, after some algebraic manipulation, from (4.3.8) to (4.3.13)

$$\frac{\partial^2 Q(\theta_o)}{\partial\beta\partial\beta'} = -\lim_{n\to\infty} \frac{1}{n} \sum_1^n a_i x_i x_i' \qquad \ldots (4.3.14)$$

$$\frac{\partial^2 Q(\theta_o)}{\partial(\sigma^2)^2} = -\lim_{n\to\infty} \frac{1}{n} \sum_1^n b_i \qquad \ldots (4.3.15)$$

$$\frac{\partial^2 Q(\theta_o)}{\partial\delta^2} = -\lim_{n\to\infty} \frac{1}{n} \sum_1^n c_i \qquad \ldots (4.3.16)$$

$$\frac{\partial^2 Q(\theta_o)}{\partial\sigma^2\partial\delta} = -\lim_{n\to\infty} \frac{1}{n} \sum_1^n k_i \qquad \ldots (4.3.17)$$

$$\frac{\partial^2 Q(\theta_o)}{\partial\beta\,\partial\sigma^2} = -\lim_{n\to\infty} \frac{1}{n} \sum_1^n d_i x_i \qquad \ldots (4.3.18)$$

and $$\frac{\partial^2 Q(\theta_o)}{\partial\beta\partial\delta} = -\lim_{n\to\infty} \frac{1}{n} \Sigma e_i x_i \qquad \ldots (4.3.19)$$

where

$$a_i = -\frac{1}{\sigma_{oi}^2}\left(w_i\Phi_i - \frac{\Phi_i^2}{1-\phi_i} - \phi_i\right)$$

$$b_i = -\frac{1}{4\sigma_o^4}\left(w_i^3\Phi_i + w_i\Phi_i - \frac{w_i^2\Phi_i^2}{1-\phi_i} - 2\phi_i\right)$$

$$c_i = -\left(\frac{\ln m_i}{2}\right)^2\left(w_i^3\Phi_i + w_i\Phi_i - \frac{w_i^2\Phi_i^2}{1-\phi_i} - 2\phi_i\right)$$

$$d_i = \frac{1}{2\sigma_{oi}\sigma_o^2}\left(w_i^2\Phi_i + \Phi_i - \frac{w_i\Phi_i^2}{1-\phi_i}\right)$$

$$e_i = \frac{\ln m_i}{2\sigma_{oi}}\left(w_i^2\Phi_i + \Phi_i - \frac{w_i\Phi_i^2}{1-\phi_i}\right)$$

$$k_i = \frac{\ln m_i}{4\sigma_o^2}\left(\frac{w_i^2\Phi_i^2}{1-\phi_i} + 2\phi_i - w_i\Phi_i - w_i^3\Phi_i\right). \qquad \ldots(4.3.20)$$

Therefore,

$$\frac{\partial^2 Q(\theta_o)}{\partial\theta\partial\theta'} = -\lim_{n\to\infty}\frac{1}{n}\begin{pmatrix} \sum_1^n a_i x_i x_i' & \sum_1^n d_i x_i & \sum_1^n e_i x_i \\ \sum_1^n d_i x_i' & \sum_1^n b_i & \sum_1^n k_i \\ \sum_1^n e_i x_i' & \sum_1^n k_i & \sum_1^n c_i \end{pmatrix} \qquad \ldots(4.3.21)$$

$$\underline{\text{def}} \quad -A, \text{ say.}$$

In order to show that $\frac{\partial^2 Q(\theta_o)}{\partial\theta\partial\theta'}$ is negative definite, we show that A is positive definite (p. d.). Now, for any k-component vector p and scalars q and r,

$P'AP$ [where $P' = (p', q, r)$] can be reduced to

$$P'AP = \lim_{n\to\infty} \frac{1}{n} \sum_{1}^{n} S_i' B_i S_i \qquad \ldots (4.3.22)$$

where $S_i' = (p'x_i, q, r)$

and $B_i = \begin{pmatrix} a_i & d_i & e_i \\ d_i & b_i & k_i \\ e_i & k_i & c_i \end{pmatrix}$.

Now, in order to show A to be p. d. let us first study B_i.

Lemma 4.5: B_i is positive semidefinite (Assumption (ii) is sufficient to ensure this).

Proof : Let

$$\left.\begin{aligned} \overline{a_i} &= w_i\Phi_i - \frac{\Phi_i^2}{1-\phi_i} - \phi_i \\ \overline{b_i} &= w_i^3\Phi_i + w_i\Phi_i - \frac{w_i^2\Phi_i^2}{1-\phi_i} - 2\phi_i \\ \overline{d_i} &= w_i^2\Phi_i + \Phi_i - \frac{w_i\Phi_i^2}{1-\phi_i} \end{aligned}\right| \qquad \ldots (4.3.23)$$

Then from (4.3.20)

$$a_i = -\frac{\bar{a}_i}{\sigma_{oi}^2}, \qquad b_i = -\frac{\bar{b}_i}{4\sigma_o^4}$$

$$c_i = -\frac{(\ln m_i)^2}{4}\bar{b}_i, \qquad d_i = \frac{\bar{d}_i}{2\sigma_{oi}\sigma_o^2} \qquad \ldots (4.3.24)$$

$$e_i = \frac{(\ln m_i)\,\bar{d}_i}{2\sigma_{oi}} \quad \text{and} \quad k_i = -\frac{\ln m_i}{4\sigma_o^2}\bar{b}_i .$$

Now, from (4.3.23)

$$\bar{a}_i = w_i \Phi_i - \Phi_i^2 / (1 - \phi_i) - \phi_i$$

$$= \Phi_i \left[w_i - \Phi_i / (1 - \phi_i) \right] - \phi_i .$$

Since $\frac{\Phi_i}{1-\phi_i} \to w_i$ [11/] (if w_i is bounded which is ensured by Assumption (ii)), we obviously have $\bar{a}_i < 0$ and hence $a_i > 0$.

Again, since $w_i \Phi_i - 2\Phi_i < 0$ [12/], we have

$$\bar{b}_i = w_i^2 \Phi_i \left(w_i - \frac{\Phi_i}{1-\phi_i} \right) + \left(w_i \Phi_i - 2\phi_i \right) < 0,$$

and therefore from (4.3.24), both b_i and c_i are positive.

11/ See Feller (1972), p. 175 or Amemiya (1973a), p. 1007.

12/ $2\phi_i - w_i\Phi_i \to 0$ as $w_i \to -\infty$ and its derivative with respect to w_i i.e., $2\Phi_i \cdot \Phi_i + w_i^2 \Phi_i \cdot \Phi_i = \Phi_i + w_i^2 \Phi_i$ is always positive. Hence $2\phi_i - w_i\Phi_i$ is positive and, therefore, $w_i\Phi_i - 2\phi_i$ is negative.

Let us now look at the second-order principal minors :

$$\begin{vmatrix} b_i & k_i \\ k_i & c_i \end{vmatrix} = b_i c_i - k_i^2 = 0 \text{ (from 4.3.24)} .$$

$$\begin{vmatrix} a_i & d_i \\ d_i & b_i \end{vmatrix} = a_i b_i - d_i^2 = \frac{1}{4\sigma_{oi}^2 \sigma_o^4} (\bar{a}_i \bar{b}_i - \bar{d}_i^2) > 0.$$ [13]

Similarly,

$$\begin{vmatrix} a_i & e_i \\ e_i & c_i \end{vmatrix} = a_i c_i - e_i^2 = \frac{(\ln m_i)^2}{4\sigma_{oi}^2} (\bar{a}_i \bar{b}_i - \bar{d}_i^2) > 0 .$$

It can also be seen using relations in (4.3.24), that

$$\begin{vmatrix} a_i & d_i & e_i \\ d_i & b_i & k_i \\ e_i & k_i & c_i \end{vmatrix} = 0$$

Thus we have proved that B_i is positive semidefinite.

Q. E. D.

Lemma 4.6 : Under Assumptions(ii) and (iii) $\frac{\partial^2 Q(\theta_o)}{\partial \theta \, \partial \theta'}$ is negative definite.

Proof : We can write from (4.3.22)

$$P' A P \geq \inf_i (\xi_i) \lim_{n \to \infty} \frac{1}{n} \sum_1^n S_i' S_i$$

where ξ_i is the smallest characteristic root of B_i. [14]

13/ See Amemiya (1973a), pp.1007-1008.

14/ Vide Rao (1974), p. 62.

Now, by Assumption (iii), $\lim_{n\to\infty} \frac{1}{n} \sum_{1}^{n} S_i' S_i$ is positive unless all the elements of p, q and r are zero. But since B_i is positive semidefinite for all i = 1, 2, ..., n, we have,

$$\inf_i (\xi_i) = 0.$$

Thus, $P'A\,P \geq 0$.

Now, if we look at $P'AP$ from (4.3.22)

$$P'AP = \lim_{n\to\infty} \frac{1}{n} \sum_{1}^{n} S_i' B_i S_i.$$

Obviously $P'AP$ is zero when $S_i' B_i S_i = 0$ for all i = 1, 2, ..., n. If at least one of the quadratic forms under the summation is positive, then $P'AP$ is also positive. Now, excepting for the obvious case when $S_i = 0$ (which can not be true for all i = 1, 2, ..., n), $S_i' B_i S_i = 0$ if and only if $B_i S_i = 0$. Let us then examine the solution vector $S_i' = (S_{1i}, S_{2i}, S_{3i})$ such that $B_i S_i = 0$.

It is clear from Appendix 4.2 that whatever be the second and third elements of the solution vector i.e., S_{2i} and S_{3i}, the first element of the solution vector i.e., S_{1i} is always zero. This cannot be so for all i = 1, 2, ..., n since S_1 has the form $S_i' = (p\,x_i, q, r)$. Thus we find $P'AP > 0$ for $P \neq 0$ and hence the result.

Since $\frac{\partial^2 Q}{\partial\theta\partial\theta'}$ is continuous, Lemma 4.6 implies there is a closed set

$$B(\theta_o) = \{\, \theta : \left| \theta - \theta_o \right| \leq \eta \,\}, \quad B \leftarrow \Psi$$

such that $\frac{\partial^2 Q}{\partial\theta\partial\theta'}$ is negative definite for all θ in $B(\theta_o)$.

Now, following Amemiya (1973a, pp.1008-1010), we can state the following theorem:

Theorem 4.1 : Under Assumptions (i) to (iii), the normal equations have a strongly consistent root, say $\hat{\theta}_n$, and

$$\sqrt{n}\,(\hat{\theta}_n - \theta_o) \longrightarrow N\left[0, \left(-\frac{\partial^2 Q(\theta_o)}{\partial\theta\,\partial\theta'}\right)^{-1}\right]. \quad 15/$$

4.3.2 ML Estimation When $V(y_i) = \sigma_o^2 m_i^{\delta_o}$

It is obvious that the ML estimator of θ (i.e., a root of the normal equations) is difficult to obtain because of the nonlinearities in the normal equations. There are a number of standard methods to obtain the solution of such nonlinear equations. Hausman and Wise (1977), for example, suggest the use of 'gradient method' which involves calculation of only the first derivatives of the ML function. Poirier (1978a), on the other hand, recommends the use of 'modified quadratic hill climbing method' originally developed by Goldfeld and Quandt (1972). These methods guarantee convergence to a local maximum only and hence one has to try different initial values of the parameters to obtain the global maximum. One can use any of these methods. We, however, suggest the use of Newton-Raphson method which, although suffers from the same shortcoming viz., that there is no guarantee of the convergence of the iterative procedure to the root that corresponds to the global maximum of the likelihood function, has the advantage that the second-round

15/ It seems that Oslen's (1978) result that the ML function of Tobin's limited dependent variable model has a unique maximum, does not hold for GLDVM.

estimator will have the same asymptotic distribution as a consistent root of the normal equations under general conditions provided the initial estimator of θ is consistent (vide Amemiya (1973a)).

In what follows we extend Amemiya's method of obtaining the consistent initial estimator for θ to our case. Noting that

$$E(y_i) = \beta_o' x_i + \sigma_{oi}^2 \frac{f_{oi}}{F_{oi}}, \quad i \in \bar{s} \qquad \ldots (4.3.25)$$

and $$E(y_i^2) = (\beta_o' x_i)^2 + \sigma_{oi}^2 \beta_o' x_i \frac{f_{oi}}{F_{oi}} + \sigma_{oi}^2, \quad i \in \bar{s}. \qquad \ldots (4.3.26).$$

We can write from (4.3.25) and (4.3.26)

$$E(y_i^2) = \beta_o' x_i \; E(y_i) + \sigma_{oi}^2, \quad i \in \bar{s}.$$

Hence

$$y_i^2 = \beta_o' x_i y_i + \sigma_{oi}^2 + v_i, \quad i \in \bar{s} \qquad \ldots (4.3.27)$$

where $$v_i = \beta_o' x_i \left[E(y_i) - y_i \right] + \left[y_i^2 - E(y_i^2) \right].$$

Unlike Amemiya, we find (4.3.27) is not linear in all the parameters because by assumption $\sigma_{oi}^2 = \sigma_o^2 m_i^{\delta_o}$. By a series expansion of $m_i^{\delta_o} = e^{\delta_o \ln m_i}$, we can, however, approximate $m_i^{\delta_o}$ by

$$\{ 1 + \delta_o \ln m_i + \frac{\delta_o^2 (\ln m_i)^2}{2} + \ldots \}.$$

Then we have from[16] (4.3.27),

16/ One can retain further terms in the series expansion but there is then a problem of choosing proper estimates for δ_o and σ_o^2. Since we are interested only in consistent estimates of δ_o and σ_o^2, we can choose them from the estimates of σ_o^2, $\sigma_o^2 \delta_o$, $\sigma_o^2 \delta_o^2$, ... obtained from (4.3.29). Even if we stop at the term in $\sigma_o^2 \delta_o^2$, we have a multiplicity of consistent estimates. Nevertheless one may not stop at the term in $\sigma_o^2 \delta_o^2$.

$$y_i^2 = \beta_o' x_i y_i + \sigma_o^2 + \sigma_o^2 \delta_o \ln m_i + \sigma_o^2 \delta_o^2 \frac{(\ln m_i)^2}{2}$$

$$+ \ldots\ldots + v_i , \quad i \varepsilon \bar{s} \cdot \qquad \ldots \quad (4.3.28)$$

From (4.3.28) and the definition of v_i, it is clear that we have a errors-in-variables problem in estimating the parameters in (4.3.28) by regression method. Assuming that $\left(\sum_{\bar{s}} x_i x_i' \right)^{-1}$ exists, let us define

$$\hat{y}_i = \left[x_i' \left(\sum_{\bar{s}} x_i x_i' \right)^{-1} \sum_{\bar{s}} x_i y_i \right] .$$

We then suggest $\left(x_i' \hat{y}_i, 1, \ln m_i, \frac{(\ln m_i)^2}{2}, \ldots\ldots \right)$ as instrumental variables for estimating the regression equations (4.3.28) and define $\hat{\theta}_1$, the first stage estimator of

$$\theta = (\beta_o', \sigma_o^2, \sigma_o^2 \delta_o, \sigma_o^2 \delta_o^2, \ldots\ldots)'$$

as

$$\hat{\theta}_1 = \begin{bmatrix} \hat{\beta}_o \\ \hat{\sigma}_o^2 \\ \hat{\sigma}_o^2 \hat{\delta}_o \\ \hat{\sigma}_o^2 \hat{\delta}_o^2 \\ \vdots \end{bmatrix} = \left[\sum_{\bar{s}} \begin{bmatrix} x_i \hat{y}_i \\ 1 \\ \ln m_i \\ \frac{(\ln m_i)^2}{2} \\ \vdots \end{bmatrix} (x_i' y_i, 1, \ln m_i, \frac{(\ln m_i)^2}{2}, \ldots) \right]^{-1}$$

$$\sum_{\bar{s}} \begin{bmatrix} x_i \hat{y}_i \\ 1 \\ \ln m_i \\ \frac{(\ln m_i)^2}{2} \\ \vdots \end{bmatrix} y_i^2 \qquad \ldots \quad (4.3.29)$$

assuming that the matrix to be inverted is non-singular.

As this estimator is similar to that of Amemiya suggested in the context of LDV models, we can easily use Amemiya's Theorem 3 (1973a, p.1013) to conclude that under assumptions (i) to (iii), $\hat{\theta}_1$ defined above is weakly consistent provided the matrix whose inverse is required in (4.3.29) is non-singular. It can easily be argued on the same lines as Amemiya's that excepting for the case where there is no regressor that varies with i, the required inverse will, in general exist. Once a consistent estimate of θ i.e., $\hat{\theta}_1$ is obtained, we can easily obtain a consistent initial estimate of θ from $\hat{\theta}_1$.

4.3.3 Strong Consistency of ML Estimator When $V(y_i) = \sigma_o^2 \left[E(y_i)\right]^{\delta_o}$

Let us assume an alternative structure for σ_{oi}^2 defined in (4.2.5) i.e.,

$$\sigma_{oi}^2 = \sigma_o^2 \left[E(y_i)\right]^{\delta_o} = \sigma_o^2 \mu_{oi}^{\delta_o}$$

where $\mu_{oi} = E(y_i) > 0$ for all $i = 1, 2, \ldots, n$.

For calculating first-order and second-order derivatives of the likelihood function with respect to the parameters we use derivatives in (4.2.6) with Z_i's replaced by μ_i's. In order to show that the normal equations under the present assumption for variance possess a strong consistent root, we write the appropriate log-likelihood function which is similar to (4.3.2) excepting that m_i is now replaced by μ_i. Obviously $\frac{\partial Q}{\partial \sigma^2}$ and $\frac{\partial Q}{\partial \delta}$ will be the same as in (4.3.5) and (4.3.6) respectively with m_i being replaced by μ_i. $\frac{\partial Q}{\partial \beta}$ will now become

$$\frac{\partial Q}{\partial \beta} = \lim_{n\to\infty} \frac{1}{n} \sum_{1}^{n} \left[-\frac{1-F_{oi}}{1-F_i} f_i x_i \left(1 - \frac{\delta}{2}\right) - \frac{\delta x_i F_{oi}}{2\mu_i} \right.$$

$$+ \frac{1}{2\sigma^2} \mu_i^{-\delta} \{2(\beta_o - \beta)' x_i F_{oi} x_i + 2\sigma_{oi}^2 x_i f_{oi}\}$$

$$+ \frac{\delta}{2\sigma^2} \mu_i^{-(\delta+1)} x_i \{((\beta_o - \beta)' x_i)^2 F_{oi}$$

$$\left. + 2\sigma_{oi}^2 (\beta_o - \beta)' x_i f_{oi} + \sigma_{oi}^2 (F_{oi} - \beta_o' x_i f_{oi})\} \right], \quad \ldots(4.3.30)$$

Clearly $\frac{\partial Q(\theta_o)}{\partial \beta} = 0$ and hence $\frac{\partial Q(\theta_o)}{\partial \theta} = 0$ in this case also.

The expressions for $\frac{\partial^2 Q}{\partial(\sigma^2)^2}$, $\frac{\partial^2 Q}{\partial \delta^2}$ and $\frac{\partial^2 Q}{\partial \sigma^2 \partial \delta}$ will be the same as in (4.3.9), (4.3.10) and (4.3.11) respectively with m_i being replaced by μ_i. The expressions for $\frac{\partial^2 Q}{\partial \beta \partial \beta'}$, $\frac{\partial^2 Q}{\partial \beta \partial \sigma^2}$ and $\frac{\partial^2 Q}{\partial \beta \partial \delta}$ for this case are given in Appendix 4.3. Evaluating these derivatives at θ_o and, as before, defining w_i as

$$w_i = \frac{\beta_o' x_i}{\sigma_{oi}} = \frac{\mu_{oi}}{\sigma_{oi}}$$

and writing the standard normal density and distribution functions evaluated at w_i by Φ_i and ϕ_i respectively, we have the following :

$$\frac{\partial^2 Q(\theta_o)}{\partial \beta \partial \beta'} = \lim_{n\to\infty} \frac{1}{n} \sum_{1}^{n} \frac{1}{\sigma_{oi}^2} \left[\left(1 - \frac{\delta_o}{2}\right)^2 \left(w_i \Phi_i - \frac{\Phi_i^2}{1-\phi_i}\right) \right.$$

$$\left. - \left(\frac{\delta_o \Phi_i}{w_i} + \phi_i + \frac{\delta_o^2 \phi_i}{2w_i^2} - \frac{\delta_o^2 \Phi_i}{4w_i}\right) \right] x_i x_i' \ldots \quad (4.3.31)$$

$$\frac{\partial^2 Q(\theta_o)}{\partial\beta\partial\sigma^2} = \lim_{n\to\infty} \frac{1}{n}\sum_{1}^{n} \frac{1}{2\sigma_o^2\,\sigma_{oi}} \left[(1-\frac{\delta_o}{2}) \left(\frac{\Phi_i^2 w_i}{1-\phi_i} - w_i^2 \Phi_i \right) - \Phi_i(1-\frac{\delta_o}{2}) - \frac{\delta_o \phi_i}{w_i} \right] x_i \qquad \ldots (4.3.32)$$

and

$$\frac{\partial^2 Q(\theta_o)}{\partial\beta\partial\delta} = \lim_{n\to\infty} \frac{1}{n}\sum_{1}^{n} \frac{\ln \mu_{oi}}{2\sigma_{oi}} \left[(1-\frac{\delta_o}{2}) \left(\frac{\Phi_i^2 w_i}{1-\phi_i} - w_i^2 \Phi_i \right) - \Phi_i(1-\frac{\delta_o}{2}) - \frac{\delta_o \phi_i}{w_i} \right] x_i . \qquad \ldots (4.3.33)$$

Proceeding as before we can show that

$$P'\bar{A}P = \lim_{n\to\infty} \frac{1}{n} \sum_{1}^{n} S_i' \bar{B}_i S_i$$

where $\bar{A} \overset{\text{def}}{=\!=} - \dfrac{\partial^2 Q(\theta_o)}{\partial\theta\partial\theta'}$, $P' = (p', q, r)$, $S_i' = (p' x_i, q, r)$.

$$\bar{B}_i = \begin{pmatrix} a_i' & d_i' & e_i' \\ d_i' & b_i' & k_i' \\ e_i' & k_i' & c_i' \end{pmatrix}$$

$$a_i' = \frac{\bar{a}_i'}{\sigma_{oi}^2}, \quad \bar{a}_i' = (1-\frac{\delta_o}{2})^2 \left(\frac{\Phi_i^2}{1-\phi_i} - w_i\Phi_i \right) + \left(\frac{\delta_o \Phi_i}{w_i} + \phi_i + \frac{\delta_o^2 \phi_i}{2w_i^2} - \frac{\delta_o^2 \Phi_i}{4w_i} \right)$$

$$d_i' = \frac{\bar{d}_i'}{2\sigma_o^2\sigma_{oi}}, \quad \bar{d}_i' = (1-\frac{\delta_o}{2}) \left(w_i^2\Phi_i - \frac{\Phi_i^2 w_i}{1-\phi_i} \right) + \Phi_i(1-\frac{\delta_o}{2}) + \frac{\delta_o\phi_i}{w_i} \qquad (4.3.34)$$

$$e_i' = \frac{\ln \mu_{oi}}{2\sigma_{oi}}\, \bar{d}_i'$$

and b_i', c_i' and k_i' are the same as b_i, c_i and k_i in (4. 3. 20) respectively with m_i being replaced by μ_{oi}. As in the previous case, we have the following lemma:

Lemma 4.7 : Under Assumptions (ii), (iii) and (iv) $\bar{A}$ is positive definite.

Proof : It is quite clear that to prove the lemma it is enough to show that

$$a_i' > 0 \quad \text{and} \quad \begin{vmatrix} a_i' & d_i' \\ d_i' & b_i' \end{vmatrix} > 0 .$$

From (4.3.34), we can write

$$\bar{a}_i' = (1 - \frac{\delta_o}{2})^2 \{ \Phi_i (\frac{\Phi_i}{1-\phi_i} - w_i) \} + \frac{\delta_o^2}{4w_i^2} (2\phi_i - w_i\Phi_i)$$

$$+ \frac{1}{w_i} (\delta_o\Phi_i + \phi_i w_i) .$$

Since we have already shown in the previous case that

$$\frac{\Phi_i}{1-\phi_i} > w_i \quad \text{and} \quad (2\phi_i - w_i\Phi_i) > 0,$$

it is enough to show that $(\delta_o \Phi_i + \phi_i w_i)$ is positive if $\bar{a}_i'$ is to be positive. If $\delta_o \geq 0$, $(\delta_o \Phi_i + \phi_i w_i)$ is obviously positive since by assumption (iv) $w_i > 0$. For $\delta_o < 0$, it can be seen[17] that

$$(\delta_o \Phi_i + \phi_i w_i) > 0.$$

Hence it is proved that $\bar{a}_i' > 0$, so that $a_i' > 0$.

17/ $\delta_o\Phi_i + w_i \phi_i \longrightarrow 0$ as $w_i \to -\infty$, for all finite values of δ_o and the derivative of $(\delta_o\Phi_i + w_i\phi_i)$ with respect to w_i is $(\phi_i + \Phi_i w_i(1 - \delta_o))$. Thus for $\delta_o < 0$, the derivative is positive. Hence $\delta_o\Phi_i + w_i\phi_i > 0$ for all finite $\delta_o < 0$.

Now $\begin{vmatrix} a_i' & d_i' \\ d_i' & b_i' \end{vmatrix}$ can be reduced after substituting the values of a_i', b_i' and d_i' from (4.3.34) to

$$\frac{D_i}{4\sigma_o^4\,\sigma_{oi}^2}\left(\phi_i w_i + \frac{2\phi_i}{w_i} + \Phi_i\right) + \frac{1}{4\sigma_o^4\,\sigma_{oi}^2}\left(2\phi_i^2 - \Phi_i^2 - w_i\Phi_i\phi_i\right) \quad \ldots(4.3.35)$$

where $D_i = \Phi_i w_i \left(\frac{\phi_i}{1-\phi_i} - w_i\right) > 0.$

Thus for (4.3.35) to be positive it is sufficient to show that

$$(2\phi_i^2 - \Phi_i^2 - w_i\Phi_i\phi_i) > 0$$

which, in fact, is true, because

$$\{2\phi_i^2 - \Phi_i^2 - w_i\Phi_i\phi_i\} \to 0 \text{ as } w_i \to -\infty$$

and its derivative with respect to w_i i.e.,

$$\{3\phi_i\Phi_i + w_i\Phi_i^2 + w_i^2\Phi_i\phi_i\}$$

is obviously positive. Thus, for this assumption regarding σ_{oi}^2, we have proved that $\bar{A}$ is positive definite. Hence $\frac{\partial^2 Q(\theta_o)}{\partial\theta\,\partial\theta'}$ is negative definite. We can, therefore, conclude that under Assumptions (i) to (iv), Theorem 4.1 holds for this assumption regarding σ_{oi}^2 also.

4.3.4 ML Estimation When $V(y_i) = \sigma_o^2\left[E(y_i)\right]^{\delta_o}$

Estimation of the model under this assumption regarding the variance structure however appears to be difficult. It does not seem possible to extend Amemiya's method for obtaining the consistent initial estimates of the parameters. We, therefore, cannot use Newton-Raphson method to take advantage of the result that the second stage estimator

is asymptotically equivalent to the ML estimator in the sense of having the same asymptotic distribution as the ML estimator. For this case we suggest the use of any standard method of nonlinear estimation like Hausman and Wise's 'gradient method' (1977) and Goldfeld and Quandt's (1972) 'modified quadratic hill-climbing method' (or Newton-Raphson method) to obtain the ML estimates. Since only convergence to a local maximum can be guaranteed, repeated applications starting from different initial values is necessary to locate the global maximum and hence the ML estimator of θ.

4.4 ML Estimation of the Box-Cox Nonnormal Heteroscedastic Model (BCNNHM)

We now come back to the BCNNHM defined in (4.2.2) and (4.2.5). We first show that a root of the normal equations for this model is strongly consistent.

As for the GLDVM, we write the log-likelihood function of $(y_1, \ldots, y_n)$ as (assuming $\lambda > 0$)[18/]

$$L = \text{Const.} + \sum_s \ln (1 - F_i) - \frac{1}{2} \sum_s \ln \sigma^2 - \frac{\delta}{2} \sum_s \ln Z_i - \frac{1}{2\sigma^2} \sum_s \left(\frac{\tilde{y}_i^{(\lambda)} - \beta' x_i}{Z_i^{\delta/2}} \right)^2 + (\lambda - 1) \sum_{i=1}^{n} \ln y_i \quad \ldots (4.4.1)$$

18/ Without any loss of generality, we write down the likelihood function and carry out further analysis for $\lambda > 0$ only since for $\lambda < 0$, the likelihood function will have only one change viz., instead of $\sum_s (1-F_i)$, it will now have $\sum_s F_i$. Obviously, the derivatives etc. will accordingly change and similar analysis can be done with this change in the likelihood function.

As in the previous case, we can show that $\frac{1}{n} L$ converges uniformly for all θ in Ψ to Q defined by

$$Q = \lim_{n \to \infty} \frac{1}{n} \sum_{1}^{n} \Big[\{\ln(1 - F_i)\} (1 - F_{oi}) - \frac{1}{2} (\ln\sigma^2) F_{oi} - \frac{\delta}{2} (\ln Z_i) F_{oi} - \frac{1}{2\sigma^2} Z_i^{-\delta} \{ ((\beta_o - \beta)' x_i)^2 F_{oi} + 2\sigma_{oi}^2 (\beta_o - \beta)' x_i f_{oi} + \sigma_{oi}^2 (F_{oi} - \beta_o' x_i f_{oi}) \} \Big] + (\lambda - 1) \sum_{i=1}^{n} \ln y_i \,. \qquad \ldots (4.4.2)$$

Note that since we have assumed $x_{1i} = 1$ for all $i = 1, 2, \ldots, n$, we can make $\sum_{i=1}^{n} \ln y_i = 0$ by a change in the unit of measurement of y_i, and drop the last term of (4.4.2). In order to check whether $\frac{\partial Q(\theta_o)}{\partial \theta}$ is zero, it is obvious that we have only to check additionally if $\frac{\partial Q(\theta_o)}{\partial \lambda} = 0$. Recalling (4.2.6), we find

$$\frac{\partial Q}{\partial \lambda} = \lim_{n \to \infty} \frac{1}{n} \sum_{1}^{n} \Big[\frac{1-F_{oi}}{1-F_i} \frac{\delta \mu_i^2 f_i}{2Z_i} - \frac{\delta F_{oi} \mu_i}{2Z_i} + \frac{\delta Z_i^{-(\delta+1)} \mu_i}{2\sigma^2} \{ ((\beta_o - \beta)' x_i)^2 F_{oi} + 2\sigma_{oi}^2 (\beta_o - \beta)' x_i f_{oi} + \sigma_{oi}^2 (F_{oi} - \mu_{oi} f_{oi}) \} \Big] . \qquad \ldots (4.4.3)$$

It is easy to check that $\frac{\partial Q(\theta_o)}{\partial \lambda} = 0$ and hence $\frac{\partial Q(\theta_o)}{\partial \theta} = 0$.

As for the second derivatives, expressions for

$$\frac{\partial^2 Q(\theta_o)}{\partial\beta\,\partial\beta'}, \quad \frac{\partial^2 Q(\theta_o)}{\partial(\sigma^2)^2}, \quad \frac{\partial^2 Q(\theta_o)}{\partial\delta^2},$$

$$\frac{\partial^2 Q(\theta_o)}{\partial\beta\partial\sigma^2}, \quad \frac{\partial^2 Q(\theta_o)}{\partial\beta\,\partial\delta} \quad \text{and} \quad \frac{\partial^2 Q(\theta_o)}{\partial\sigma^2\partial\delta}$$

are as given in (4.3.31), (4.3.9), (4.3.10), (4.3.32), (4.3.33) and (4.3.11) respectively with both $\ln m_i$ and $\ln \mu_{oi}$ (whichever appears) replaced by $\ln Z_{oi}$. We have only to derive the expressions for

$$\frac{\partial^2 Q}{\partial\lambda^2}, \quad \frac{\partial^2 Q}{\partial\lambda\partial\beta}, \quad \frac{\partial^2 Q}{\partial\lambda\partial\delta} \quad \text{and} \quad \frac{\partial^2 Q}{\partial\lambda\partial\sigma^2}.$$

These derivatives are given in Appendix 4.4. Denoting $w_i = \frac{\mu_{oi}}{\sigma_{oi}}$, we evaluate these at θ_o and simplify them to obtain

$$\frac{\partial^2 Q(\theta_o)}{\partial\lambda^2} = -\lim_{n\to\infty} \frac{1}{n} \sum_1^n l_i \qquad \ldots (4.4.4)$$

$$\frac{\partial^2 Q(\theta_o)}{\partial\lambda\partial\beta} = -\lim_{n\to\infty} \frac{1}{n} \sum_1^n m_i x_i' \qquad \ldots (4.4.5)$$

$$\frac{\partial^2 Q(\theta_o)}{\partial\lambda\partial\sigma^2} = -\lim_{n\to\infty} \frac{1}{n} \sum_1^n r_i \qquad \ldots (4.4.6)$$

and

$$\frac{\partial^2 Q(\theta_o)}{\partial\lambda\partial\delta} = -\lim_{n\to\infty} \frac{1}{n} \sum_1^n n_i \qquad \ldots (4.4.7)$$

where

$$l_i = -\frac{\sigma_{oi}^2 \delta_o^2 w_i^2}{4Z_{oi}^2}\bar{b}_i, \quad \bar{b}_i \text{ is as defined in (4.3.23)}$$

$$m_i = \frac{\delta_o w_i}{2Z_{oi}}\bar{d}_i, \quad \bar{d}_i = \bar{d}_i' \text{ in (4.3.34)}$$

$$n_i = -\frac{\delta_o \sigma_{oi} w_i \ln Z_{oi}}{4Z_{oi}}\bar{b}_i$$

$$r_i = -\frac{\delta_o \sigma_{oi} w_i}{4Z_{oi}\sigma_o^2}\bar{b}_i \qquad \ldots (4.4.8)$$

and Φ_i and ϕ_i are normal density and distribution functions evaluated at w_i. Now denoting a_i, b_i, c_i, k_i, d_i and e_i as given below, we have from (4.3.15), (4.3.16), (4.3.17), (4.3.31), (4.3.32), (4.3.33) and (4.4.4) to (4.4.7), as before,

$$\mathfrak{P}'\mathfrak{A}\mathfrak{P} = \lim_{n\to\infty} \frac{1}{n} \sum_{1}^{n} \mathfrak{S}_i' \mathfrak{B}_i \mathfrak{S}_i$$

where $\mathfrak{P}' = (p, q, r, u)$, $\mathfrak{A} \underline{\underline{\text{def}}} -\frac{\partial^2 Q(\theta_o)}{\partial\theta\partial\theta'}$ ... (4.4.9)

$\mathfrak{S}_i' = (p' x_i, q, r, u)$

$$\mathfrak{B}_i = \begin{pmatrix} a_i & d_i & e_i & m_i \\ d_i & b_i & k_i & r_i \\ e_i & k_i & c_i & n_i \\ m_i & r_i & n_i & l_i \end{pmatrix} \qquad \ldots (4.4.10)$$

$$a_i = \frac{1}{\sigma_{oi}^2}\,\bar{a}_i\,, \quad \bar{a}_i = \bar{a}_i' \quad \text{in (4.3.34)}$$

$$b_i = -\frac{1}{4\,\sigma_o^4}\,\bar{b}_i\,, \quad \bar{b}_i \text{ is as defined in (4.3.23)}$$

$$c_i = -\left(\frac{\ln Z_{oi}}{2}\right)^2 \bar{b}_i$$

$$k_i = \frac{\ln Z_{oi}}{4\sigma_o^2}\left(-\bar{b}_i\right)$$

$$d_i = \frac{1}{2\sigma_o^2\,\sigma_{oi}}\,\bar{d}_i\,, \quad \bar{d}_i = \bar{d}_i' \quad \text{in (4.3.34)}$$

and $$e_i = \frac{\ln Z_{oi}}{2\sigma_{oi}}\,\bar{d}_i\,.$$

Since $\bar{b}_i$ has been shown to be < 0, we conclude from (4.4.8) the $l_i > 0$. Also, all the elements in $\tilde{B}_i$ being in terms of $\bar{a}_i$, $\bar{b}_i$, $\bar{d}_i$ and other constants, one can easily check, as before, that all principal minors of $\tilde{B}_i$ are greater than or equal to zero. Then arguing in the same manner as for GLDVM, it is easily established that $\tilde{A}$ is positive definite and hence $\dfrac{\partial^2 Q(\theta_o)}{\partial\theta\partial\theta'}$ is negative definite and therefore a theorem similar to Theorem 4.1 holds under Assumptions (i) to (iv).

As for the estimation of the parameters, one has to use standard nonlinear techniques of estimation like Hausman and Wise's (1977) gradient method etc., to obtain the estimates. It appears that under standard conditions [vide Cox and Hinkley (1974)] the converging values would, in fact, be a root of the normal equations. In order

to obtain the global maximum, one has to repeat this with different initial values of the parameters.

4.5 Conclusions

In this Chapter we have extended the Box-Cox model by simultaneously treating the distribution of the transformed dependent variable as truncated normal and variances of the original dependent variable varying across observations. This is an improvement over Poirier's (1978a) work in the sense that we have considered heteroscedasticity as well. Although, Tobin's limited dependent variable model comes out as a special case of such a generalized model, we have considered here the generalization of Tobin model separately because of its distinct interest.

Appendix 4.1

Second-Order Derivatives of Q in (4.3.3)

We evaluate the second-order derivatives of Q with respect to the parameters. From expressions (4.3.4) to (4.3.6), we have,

$$\frac{\partial^2 Q}{\partial\beta\partial\beta'} = \lim_{n\to\infty} \frac{1}{n} \sum_{1}^{n} \left[\frac{1-F_{oi}}{1-F_i} \frac{\beta' x_i}{\sigma_i^2} f_i - \frac{1-F_{oi}}{(1-F_i)^2} f_i^2 \right.$$

$$\left. - \frac{1}{\sigma^2} m_i^{-\delta} F_{oi} \right] x_i x_i' \qquad \ldots(A\ 4.1.1)$$

$$\frac{\partial^2 Q}{\partial(\sigma^2)^2} = \lim_{n\to\infty} \frac{1}{n} \sum_{1}^{n} \left[-\frac{1}{2(\sigma^2)^2} \frac{1-F_{oi}}{1-F_i} \beta' x_i f_i \right.$$

$$- \left(\frac{1}{2\sigma^2}\right)^2 \left(\frac{1}{1-F_i}\right)^2 (\beta' x_i f_i)^2 (1-F_{oi})$$

$$+ \frac{1}{2\sigma^2} \frac{1-F_{oi}}{1-F_i} \beta' x_i \frac{(\beta' x_i)^2 - \sigma_i^2}{2\sigma_i^2 \sigma^2} f_i + \frac{F_{oi}}{2(\sigma^2)^2}$$

$$- \frac{m_i^{-\delta}}{(\sigma^2)^3} \{ ((\beta_o - \beta)' x_i)^2 F_{oi} + 2\sigma_{oi}^2 (\beta_o - \beta)' x_i f_{oi}$$

$$\left. + \sigma_{oi}^2 (F_{oi} - \beta_o' x_i f_{oi}) \} \right] \qquad \ldots (A\ 4.1.2)$$

$$\frac{\partial^2 Q}{\partial \delta^2} = \lim_{n\to\infty} \frac{1}{n} \sum_{1}^{n} \left[\frac{1-F_{oi}}{(1-F_i)^2} \left(-\frac{\ln m_i}{2} \right) \beta' x_i f_i \frac{\ln m_i}{2} \beta' x_i f_i \right.$$

$$+ \frac{1-F_{oi}}{1-F_i} \frac{\ln m_i}{2} \beta' x_i \frac{(\beta' x_i)^2 - \sigma_i^2}{2\sigma_i^2} f_i \ln m_i$$

$$- \frac{m_i^{-\delta} (\ln m_i)^2}{2\sigma^2} \{ ((\beta_o - \beta)' x_i)^2 F_{oi} + 2\sigma_{oi}^2 (\beta_o - \beta)' x_i f_{oi}$$

$$\left. + \sigma_{oi}^2 (F_{oi} - \beta_o' x_i f_{oi}) \} \right] \qquad \ldots \text{(A 4.1.3)}$$

$$\frac{\partial^2 Q}{\partial \sigma^2 \partial \delta} = \lim_{n\to\infty} \sum_{1}^{n} \left[-\frac{1}{2\sigma^2} \frac{1-F_{oi}}{(1-F_i)^2} \frac{\ln m_i}{2} (\beta' x_i f_{oi})^2 \right.$$

$$+ \frac{1}{2\sigma^2} \frac{1-F_{oi}}{1-F_i} \beta' x_i \{ \frac{(\beta' x_i)^2 - \sigma_i^2}{2\sigma_i^2} \} f_i \ln m_i$$

$$- \frac{m_i^{-\delta} \ln m_i}{2\sigma^4} \{ ((\beta_o - \beta)' x_i)^2 F_{oi} + 2\sigma_{oi}^2 (\beta_o - \beta)' x_i f_{oi}$$

$$\left. + \sigma_{oi}^2 (F_{oi} - \beta_o' x_i f_{oi}) \} \right] \qquad \ldots \text{(A 4.1.4)}$$

$$\frac{\partial^2 Q}{\partial \beta \partial \sigma^2} = \lim_{n\to\infty} \frac{1}{n} \sum_{1}^{n} \left[\frac{1-F_{oi}}{(1-F_i)^2} \frac{1}{2\sigma^2} \beta' x_i f_i^2 x_i \right.$$

$$- \frac{1-F_{oi}}{1-F_i} \{ \frac{(\beta' x_i)^2 - \sigma_i^2}{2\sigma^2 \sigma_i^2} \} f_i x_i$$

$$\left. - \frac{m_i^{-\delta}}{\sigma^4} \{ (\beta_o - \beta)' x_i F_{oi} x_i + \sigma_{oi}^2 f_{oi} x_i \} \right] \qquad \ldots \text{(A 4.1.5)}$$

and $$\frac{\partial^2 Q}{\partial\beta\partial\delta} = \lim_{n\to\infty} \frac{1}{n} \sum_{1}^{n} \left[\frac{1-F_{oi}}{(1-F_i)^2} \frac{\ln m_i}{2} \beta' x_i f_i^2 x_i \right.$$

$$- \frac{1-F_{oi}}{1-F_i} \left\{ \frac{(\beta' x_i)^2 - \sigma_i^2}{2\sigma_i^2} \right\} f_i (\ln m_i) x_i$$

$$\left. - \frac{m_i^{-\delta} \ln m_i}{\sigma^2} \left\{ (\beta_o - \beta)' x_i F_{oi} x_i + \sigma_{oi}^2 f_{oi} x_i \right\} \right] \quad \ldots (A\ 4.1.6)$$

Evaluating the second-order derivatives at θ_o, we find from (A 4.1.1) to (A 4.1.6),

$$\frac{\partial^2 Q(\theta_o)}{\partial\beta\,\partial\beta'} = \lim_{n\to\infty} \frac{1}{n} \sum_{1}^{n} \left[\frac{\beta_o' x_i}{\sigma_{oi}^2} f_{oi} - \frac{f_{oi}^2}{1-F_{oi}} - \frac{1}{\sigma_o^2} m_i^{-\delta_o} F_{oi} \right] x_i x_i' \quad \ldots (A\ 4.1.7)$$

$$\frac{\partial^2 Q(\theta_o)}{\partial(\sigma^2)^2} = \frac{1}{2\sigma_o^2} \lim_{n\to\infty} \frac{1}{n} \sum_{1}^{n} \left[-\frac{1}{\sigma_o^2} \beta_o' x_i f_{oi} - \frac{1}{1-F_{oi}} \frac{(\beta_o' x_i f_{oi})^2}{2\sigma_o^2} \right.$$

$$+ \beta_o' x_i \left\{ \frac{(\beta_o' x_i)^2 - \sigma_{oi}^2}{2\sigma_{oi}^2 \sigma_o^2} \right\} f_{oi} + \frac{F_{oi}}{\sigma_o^2}$$

$$\left. - \frac{2m_i^{-\delta_o} \sigma_{oi}^2}{\sigma_o^4} (F_{oi} - \beta_o' x_i f_{oi}) \right] \quad \ldots (A\ 4.1.8)$$

$$\frac{\partial^2 Q(\theta_o)}{\partial \delta^2} = \lim_{n\to\infty} \frac{1}{n} \sum_{1}^{n} \Bigg[- \frac{1}{1 - F_{oi}} \left(\frac{\ln m_i}{2} \right)^2 \left(\beta_o' x_i f_{oi} \right)^2 + \frac{(\ln m_i)}{2} \beta_o' x_i \left\{ \frac{(\beta_o' x_i)^2 - \sigma_{oi}^2}{2\sigma_{oi}^2} \right\} f_{oi} \ln m_i - \frac{1}{2\sigma_o^2} m_i^{-\delta_o} (\ln m_i)^2 \sigma_{oi}^2 (F_{oi} - \beta_o' x_i f_{oi}) \Bigg] \quad \ldots (A\ 4.1.9)$$

$$\frac{\partial^2 Q(\theta_o)}{\partial \sigma^2 \partial \delta} = \frac{1}{2\sigma_o^2} \lim_{n\to\infty} \frac{1}{n} \sum_{1}^{n} \Bigg[- \frac{1}{1 - F_{oi}} \frac{\ln m_i}{2} (\beta_o' x_i f_{oi})^2 + \beta_o' x_i \left\{ \frac{(\beta_o' x_i)^2 - \sigma_{oi}^2}{2\sigma_{oi}^2} \right\} f_{oi} \ln m_i - \frac{1}{\sigma_o^2} m_i^{-\delta_o} (\ln m_i) \sigma_{oi}^2 (F_{oi} - \beta_o' x_i f_{oi}) \Bigg] \quad \ldots (A\ 4.1.10)$$

$$\frac{\partial^2 Q(\theta_o)}{\partial \beta \partial \sigma^2} = \frac{1}{2\sigma_o^2} \lim_{n\to\infty} \frac{1}{n} \sum_{1}^{n} \Bigg[\frac{\beta_o' x_i f_{oi}^2 x_i}{(1 - F_{oi})} - \left\{ \frac{(\beta_o' x_i)^2 - \sigma_{oi}^2}{\sigma_{oi}^2} \right\} f_{oi} x_i - \frac{2}{\sigma_o^2} m_i^{-\delta_o} \sigma_{oi}^2 f_{oi} x_i \Bigg] \quad \ldots (A\ 4.1.11)$$

and $$\frac{\partial^2 Q(\theta_o)}{\partial \beta \partial \delta} = \lim_{n\to\infty} \frac{1}{n} \sum_{1}^{n} \Bigg[\frac{1}{1-F_{oi}} \frac{\ln m_i}{2} \beta_o' x_i f_{oi}^2 x_i - \left\{ \frac{(\beta_o' x_i)^2 - \sigma_{oi}^2}{2\sigma_{oi}^2} \right\} f_{oi} (\ln m_i) x_i - \frac{m_i^{-\delta_o}}{\sigma_o^2} (\ln m_i) \sigma_{oi}^2 f_{oi} x_i \Bigg] \quad \ldots (A4.1.12)$$

Appendix 4.2

Solution Vector of the System of Equations $B_i S_i = 0$

$$B_i S_i = 0$$

i.e.

$$\begin{pmatrix} a_i & d_i & e_i \\ d_i & b_i & k_i \\ e_i & k_i & c_i \end{pmatrix} \begin{pmatrix} S_{1i} \\ S_{2i} \\ S_{3i} \end{pmatrix} = 0 \quad .$$

Thus we have the following three equations :

$$a_i S_{1i} + d_i S_{2i} + e_i S_{3i} = 0 \qquad \ldots (A\ 4.2.1)$$

$$d_i S_{1i} + b_i S_{2i} + k_i S_{3i} = 0 \qquad \ldots (A\ 4.2.2)$$

and $$e_i S_{1i} + k_i S_{2i} + c_i S_{3i} = 0 \ . \qquad \ldots (A\ 4.2.3)$$

Now, from (A 4.2.1), we have

$$S_{1i} = \frac{-(d_i S_{2i} + e_i S_{3i})}{a_i} \ . \qquad \ldots (A\ 4.2.4)$$

Substituting S_{1i} from (A 4.2.4) into (A 4.2.2) and (A 4.2.3) we find,

$$S_{2i} = \frac{-S_{3i}(a_i k_i - d_i e_i)}{(a_i b_i - d_i^2)} \qquad \ldots (A\ 4.2.5)$$

and $$S_{2i} = \frac{-S_{3i}(a_i c_i - e_i^2)}{(a_i k_i - e_i d_i)} \ . \qquad \ldots (A\ 4.2.6)$$

From (A 4.2.5) and (A 4.2.6), we have

$$S_{3i}\left[\frac{a_i c_i - e_i^2}{a_i k_i - d_i e_i} - \frac{a_i k_i - d_i e_i}{a_i b_i - d_i^2}\right] = 0 \cdot \qquad \ldots (A\ 4.2.7)$$

The expression in the bracket is nothing but $a_i |B_i|$ where $|B_i|$ denotes the determinent of B_i.

Thus (A 4.2.7) becomes

$$S_{3i}\, a_i |B_i| = 0 \cdot \qquad \ldots (A\ 4.2.8)$$

Excluding the trivial solution, we find from (A 4.2.8), $S_{3i} \neq 0$ since it has already been shown that $a_i > 0$ and $|B_i| = 0$.

Now, from (A 4.2.4) and (A 4.2.5), we have

$$S_{1i} = \frac{-S_{3i}}{a_i}\left\{\frac{-d_i(a_i k_i - d_i e_i)}{(a_i b_i - d_i^2)} + e_i\right\} \cdot \qquad \ldots (A.1.2.9)$$

But the expression in the curly bracket in (A 4.2.9) can easily be shown to be zero; thus even when $S_{3i} \neq 0$, $S_{1i} = 0$. Since this solution holds for all $i = 1, 2, \ldots, n$, we have for the system of equations $\{B_i S_i = 0,\ i = 1, 2, \ldots, n\}$ the nontrivial solution vector as

$$S_i' = (0, S_{2i}, S_{3i})$$

where $S_{2i} \neq 0$, $S_{3i} \neq 0$ for all $i = 1, 2, \ldots, n$.

Appendix 4.3

Second-Order Derivatives of Q When $V(y_i) = \sigma_o^2 \left[E(y_i)\right]^{\delta_o}$

$$\frac{\partial^2 Q}{\partial\beta\partial\beta'} = \lim_{n\to\infty} \frac{1}{n} \sum_{1}^{n} \Bigg[- \frac{1-F_{oi}}{(1-F_i)^2} f_i^2 \left(1 - \frac{\delta}{2}\right)^2$$

$$- \frac{1-F_{oi}}{1-F_i}\left(1 - \frac{\delta}{2}\right)\left(- \frac{\beta' x_i f_i}{\sigma_i^2} + \frac{\delta\beta' x_i f_i}{2\sigma_i^2} - \frac{\delta f_i}{2\mu_i}\right)$$

$$+ \frac{\delta F_{oi}}{2\mu_i^2} - \frac{\delta\mu_i^{-\delta-1}}{2\sigma^2}\{\quad 2(\beta_o - \beta)' x_i F_{oi} + 2\sigma_{oi}^2 f_{oi}\}$$

$$- \frac{1}{2\sigma^2}\mu_i^{-\delta}(2F_{oi}) - \frac{\delta(\delta+1)}{2\sigma^2}\mu_i^{-\delta-2}\{((\beta_o - \beta)' x_i)^2 F_{oi}$$

$$+ 2\sigma_{oi}^2(\beta_o - \beta)' x_i f_{oi} + \sigma_{oi}^2 (F_{oi} - \beta_o' x_i f_{oi})\}$$

$$+ \frac{\delta}{2\sigma^2}\mu_i^{-\delta-1}\{-2(\beta_o - \beta)' x_i F_{oi} - 2\sigma_{oi}^2 f_{oi})\}\Bigg] x_i x_i'$$

.... (A 4.3.1)

$$\frac{\partial^2 Q}{\partial \beta \partial \sigma^2} = \lim_{n\to\infty} \frac{1}{n} \sum_{1}^{n} \Bigg[\frac{1 - F_{oi}}{(1-F_i)^2} \frac{f_i \beta' x_i f_i}{2\sigma^2} (1 - \frac{\delta}{2})$$

$$- \frac{1 - F_{oi}}{1 - F_i} (1 - \frac{\delta}{2}) f_i \mu_i^{\delta} \{ \frac{(\beta' x_i)^2 - \sigma_i^2}{2\sigma_i^4} \}$$

$$- \frac{\mu_i^{-\delta}}{\sigma^4} \{ (\beta_o - \beta)' x_i F_{oi} + \sigma_{oi}^2 f_{oi} \} - \frac{\delta \mu_i^{-\delta-1}}{2\sigma^4} \{ ((\beta_o - \beta)' x_i)^2 F_{oi}$$

$$+ 2\sigma_{oi}^2 (\beta_o - \beta)' x_i f_{oi} + \sigma_{oi}^2 (F_{oi} - \beta_o' x_i f_{oi}) \} \Bigg] x_i \quad \ldots(A.4.3.2)$$

and $$\frac{\partial^2 Q}{\partial \beta \partial \delta} = \lim_{n\to\infty} \frac{1}{n} \sum_{1}^{n} \Bigg[\frac{1-F_{oi}}{(1-F_i)^2} f_i (1 - \frac{\delta}{2}) \frac{\ln \mu_i}{2} \beta' x_i f_i + \frac{1 - F_{oi}}{1 - F_i} \frac{f_i}{2}$$

$$- \frac{1-F_{oi}}{1-F_i} (1 - \frac{\delta}{2}) \{ \frac{(\beta' x_i)^2 - \sigma_i^2}{2\sigma_i^2} \} (\ln \mu_i) f_i - \frac{F_{oi}}{2\mu_i}$$

$$+ \frac{\mu_i^{-\delta}}{\sigma^2} (-\ln \mu_i) \{ (\beta_o - \beta)' x_i F_{oi} + \sigma_{oi}^2 f_{oi} \}$$

$$+ \{ \frac{1}{2\sigma^2} \mu_i^{-\delta-1} + \frac{\delta}{2\sigma^2} \mu_i^{-\delta-1} (-\ln \mu_i) \} \{ ((\beta_o - \beta)' x_i)^2 F_{oi}$$

$$+ 2\sigma_{oi}^2 (\beta_o - \beta)' x_i f_{oi} + \sigma_{oi}^2 (F_{oi} - \beta_o' x_i f_{oi}) \} \Bigg] x_i \quad \ldots(A\ 4.3.3)$$

Thus, from (A 4.3.1) to (A 4.3.3), the second-order derivatives evaluated at θ_o will be the following:

$$\frac{\partial^2 Q(\theta_o)}{\partial\beta\,\partial\,\beta'} = \lim_{n\to\infty} \frac{1}{n} \sum_{1}^{n} \Bigg[- \frac{f_{oi}^2}{1-F_{oi}} \left(1 - \frac{\delta_o}{2}\right)^2$$

$$- \left(1 - \frac{\delta_o}{2}\right) \left\{ - \frac{\mu_{oi} f_{oi}}{\sigma_{oi}^2} + \frac{\delta_o \mu_{oi} f_{oi}}{2\sigma_{oi}^2} - \frac{\delta_o f_{oi}}{2\mu_{oi}} \right\}$$

$$+ \frac{\delta_o F_{oi}}{2\mu_{oi}^2} - \frac{\delta_o f_{oi}}{\mu_{oi}} - \frac{F_{oi}}{\sigma_{oi}^2} - \frac{\delta_o(\delta_o + 1)}{2\mu_{oi}^2}(F_{oi} - \mu_{oi} f_{oi})$$

$$- \frac{\delta_o f_{oi}}{\mu_{oi}} \Bigg] x_i x_i'$$

$$= \lim_{n\to\infty} \frac{1}{n} \sum_{1}^{n} \Bigg[\left(1 - \frac{\delta_o}{2}\right)^2 \left(\frac{\mu_{oi} f_{oi}}{\sigma_{oi}^2} - \frac{f_{oi}^2}{1 - F_{oi}} \right) - \frac{\delta_o f_{oi}}{\mu_{oi}}$$

$$- \frac{F_{oi}}{\sigma_{oi}^2} - \frac{\delta_o^2 F_{oi}}{2\mu_{oi}^2} + \frac{\delta_o^2 f_{oi}}{4\mu_{oi}} \Bigg] x_i x_i' \qquad \ldots (A\ 4.3.4)$$

$$\frac{\partial^2 Q(\theta_o)}{\partial\beta\partial\sigma^2} = \lim_{n\to\infty} \frac{1}{n} \sum_{1}^{n} \Bigg[\frac{f_{oi}^2}{1-F_{oi}} \frac{\mu_{oi}}{2\sigma_o^2} \left(1 - \frac{\delta_o}{2}\right)$$

$$- f_{oi} \mu_{oi}^{\delta_o} \left(1 - \frac{\delta_o}{2}\right) \left\{ \frac{\mu_{oi}^2 - \sigma_{oi}^2}{2\sigma_{oi}^4} \right\} - \frac{\mu_{oi}^{-\delta_o}}{\sigma_o^4} \delta_{oi}^2 f_{oi}$$

$$- \frac{\delta_o}{2\mu_{oi}\sigma_o^2} \left(F_{oi} - \mu_{oi} f_{oi} \right) \Bigg] x_i$$

$$= \frac{1}{2\sigma_o^2} \lim_{n\to\infty} \frac{1}{n} \sum_{1}^{n} \Bigg[\left(1 - \frac{\delta_o}{2}\right) \left\{ \frac{f_{oi}^2 \mu_{oi}}{1 - F_{oi}} - \frac{f_{oi} \mu_{oi}^2}{\sigma_{oi}^2} \right\}$$

$$- f_{oi}\left(1 - \frac{\delta_o}{2}\right) - \frac{\delta_o F_{oi}}{\mu_{oi}} \Bigg] x_i \qquad \ldots (A\ 4.3.5)$$

and $$\frac{\partial^2 Q(\theta_o)}{\partial\beta\partial\delta} = \lim_{n\to\infty} \frac{1}{n} \sum_{1}^{n} \left[\frac{f_{oi}^2\ \mu_{oi}}{1 - F_{oi}} \left(1 - \frac{\delta_o}{2}\right) \left(\frac{\ln \mu_{oi}}{2}\right) + \frac{f_{oi}}{2} \right.$$

$$- f_{oi}\left(1 - \frac{\delta_o}{2}\right) \ln \mu_{oi} \left\{ \frac{\mu_{oi}^2 - \sigma_{oi}^2}{2\sigma_{oi}^2} \right\} - \frac{F_{oi}}{2\mu_{oi}}$$

$$\left. - (\ln \mu_{oi})\ f_{oi} + \left\{ \frac{1}{2\mu_{oi}} - \frac{\delta_o}{2\mu_{oi}} (\ln \mu_{oi}) \right\} (F_{oi} - \mu_{oi}\ f_{oi}) \right] x_i$$

$$= \lim_{n\to\infty} \frac{1}{n} \sum_{1}^{n} \left[\frac{\ln \mu_{oi}}{2} \left[\left(1 - \frac{\delta_o}{2}\right) \left\{ \frac{f_{oi}^2\ \mu_{oi}}{1 - F_{oi}} - \frac{f_{oi}\ \mu_{oi}^2}{\sigma_{oi}^2} \right\} \right. \right.$$

$$\left. \left. - f_{oi}\left(1 - \frac{\delta_o}{2}\right) - \frac{\delta_o F_{oi}}{\mu_{oi}} \right] \right] x_i \qquad \cdots (A\ 4.3.6)$$

Appendix 4.4

Second-Order Derivatives of Q in (4.4.2)

$$\frac{\partial^2 Q}{\partial \lambda^2} = \lim_{n\to\infty} \frac{1}{n} \sum_{1}^{n} \Bigg[- \frac{1 - F_{oi}}{(1- F_i)^2} \left(\frac{\delta \mu_i^2 f_i}{2 Z_i} \right)^2$$

$$+ \frac{1-F_{oi}}{1-F_i} \frac{\delta \mu_i^2}{2Z_i} \left\{ \frac{\mu_i^2 - \sigma_i^2}{2Z_i \sigma_i^2} \right\} \delta \mu_i f_i - \frac{1- F_{oi}}{1- F_i} \frac{\delta \mu_i^2 f_i}{2Z_i^2} \mu_i$$

$$+ \frac{\delta F_{oi} \mu_i^2}{2Z_i^2} - \frac{\delta(\delta+ 1)Z_i^{-\delta - 2}}{2\sigma^2} \mu_i^2 \{ ((\beta_o - \beta)' x_i)^2 F_{oi}$$

$$+ 2\sigma_{oi}^2 (\beta_o - \beta)' x_i f_{oi} + \sigma_{oi}^2 (F_{oi} - \mu_{oi} f_{oi}) \} \Bigg] \quad \ldots (A\ 4.4.1)$$

$$\frac{\partial^2 Q}{\partial \lambda \partial \beta} = \lim_{n\to\infty} \frac{1}{n} \sum_{1}^{n} \Bigg[\frac{1 - F_{oi}}{(1- F_i)^2} \left(1 - \frac{\delta}{2}\right) \frac{f_i^2 \delta \mu_i^2}{2Z_i} + \frac{1-F_{oi}}{1-F_i} \frac{\delta f_i \mu_i}{Z_i}$$

$$- \frac{1-F_{oi}}{1-F_i} \frac{\delta \mu_i f_i}{2Z_i} + \frac{1-F_{oi}}{1-F_i} \frac{\delta \mu_i^2}{2Z_i} \left(- \frac{\mu_i f_i}{\sigma_i^2} + \frac{\delta \mu_i f_i}{2\sigma_i^2} - \frac{\delta f_i}{2\mu_i} \right)$$

$$+ \left\{ \frac{(-\delta)(\delta+ 1)Z_i^{-(\delta+2)} \lambda \mu_i}{2\sigma^2} + \frac{\delta Z_i^{-\delta-1}}{2\sigma^2} \right\} \{ ((\beta_o - \beta)' x_i)^2 F_{oi}$$

$$+ 2\sigma_{oi}^2 (\beta_o - \beta)' x_i f_{oi} + \sigma_{oi}^2 (F_{oi} - \mu_{oi} f_{oi}) \}$$

$$+ \frac{\delta Z_i^{-(\delta+1)} \mu_i}{2\sigma^2} \{ -2(\beta_o - \beta)' x_i F_{oi} - 2\sigma_{oi}^2 f_{oi} \} \Bigg] x_i'$$

$$\ldots (A\ 4.4.2)$$

$$\frac{\partial^2 Q}{\partial\lambda\partial\sigma^2} = \lim_{n\to\infty} \frac{1}{n} \sum_{1}^{n} \left[\frac{1-F_{oi}}{(1-F_i)^2} \frac{\delta\mu_i^2 f_i}{2Z_i} \left(-\frac{\mu_i f_i}{2\sigma^2}\right) \right.$$

$$+ \frac{1-F_{oi}}{1-F_i} \frac{\delta\mu_i^2}{2Z_i} \left(\frac{\mu_i^2}{\sigma_i^2} - 1\right) \frac{f_i}{2\sigma^2}$$

$$- \frac{\delta Z_i^{-(\delta+1)} \mu_i}{2(\sigma^2)^2} \{((\beta_o-\beta)'x_i)^2 F_{oi} + 2\sigma_{oi}^2(\beta_o-\beta)'x_i f_{oi}$$

$$\left. + \sigma_{oi}^2 (F_{oi} - \mu_{oi} f_{oi})\} \right] \qquad \ldots (A\ 4.4.3)$$

and $$\frac{\partial^2 Q}{\partial\lambda\partial\delta} = \lim_{n\to\infty} \frac{1}{n} \sum_{1}^{n} \left[\frac{1-F_{oi}}{(1-F_i)^2} \frac{\delta\mu_i^2 f_i}{2Z_i} \left(-\frac{1}{2}\mu_i f_i \ln Z_i\right) \right.$$

$$+ \frac{1-F_{oi}}{1-F_i} \frac{\delta\mu_i^2}{2Z_i} \left(\frac{\mu_i^2 - \sigma_i^2}{2\sigma_i^2}\right) f_i \ln Z_i + \frac{1-F_{oi}}{1-F_i} \frac{\mu_i^2 f_i}{2Z_i}$$

$$- \frac{F_{oi}\mu_i}{2Z_i} + \{ \frac{Z_i^{-(\delta+1)} \mu_i}{2\sigma^2} - \frac{\delta Z_i^{-(\delta+1)} \mu_i \ln Z_i}{2\sigma^2} \}$$

$$\{((\beta_o-\beta)'x_i)^2 F_{oi} + 2\sigma_{oi}^2 (\beta_o-\beta)'x_i f_{oi}$$

$$\left. + \sigma_{oi}^2(F_{oi} - \mu_{oi} f_{oi})\} \right] \qquad \ldots (A\ 4.4.4)$$

Evaluating them at θ_o we have, from (A 4.4.1) to (A 4.4.4),

$$\frac{\partial^2 Q(\theta_o)}{\partial \lambda^2} = \lim_{n\to\infty} \frac{1}{n} \sum_{1}^{n} \left[-\frac{1}{1-F_{oi}} \frac{\delta_o^2 \mu_{oi}^4 f_{oi}^2}{4Z_{oi}^2} + \frac{\delta_o^2 \mu_{oi}^3}{4Z_{oi}^2} \left(\frac{\mu_{oi}^2}{\sigma_{oi}^2} - 1 \right) f_{oi} \right.$$

$$\left. - \frac{\delta_o \mu_{oi}^3 f_{oi}}{2Z_{oi}^2} + \frac{\delta_o F_{oi} \mu_{oi}^2}{2Z_{oi}^2} - \frac{\delta_o(\delta_o+1)\mu_{oi}^2}{2Z_{oi}^2} (F_{oi} - \mu_{oi} f_{oi}) \right]$$

$$= \lim_{n\to\infty} \frac{1}{n} \sum_{1}^{n} \left[\frac{\delta_o^2 \mu_{oi}^2}{4Z_{oi}^2} \left\{ - \frac{\mu_{oi}^2 f_{oi}^2}{1-F_{oi}} + \frac{\mu_{oi}^3 f_{oi}}{\sigma_{oi}^2} \right.\right.$$

$$\left.\left. + \mu_{oi} f_{oi} - 2F_{oi} \right\} \right] \quad \ldots \text{(A 4.4.5)}$$

$$\frac{\partial^2 Q(\theta_o)}{\partial \lambda \partial \beta} = \frac{\delta_o}{2\lambda_o} \lim_{n\to\infty} \frac{1}{n} \sum_{1}^{n} \left[\left(1 - \frac{\delta_o}{2}\right) \frac{f_{oi}^2 \mu_{oi}}{1-F_{oi}} + f_{oi} \right.$$

$$- \frac{\mu_{oi}^2 f_{oi}}{\sigma_{oi}^2} \left(1 - \frac{\delta_o}{2}\right) - \frac{\delta_o f_{oi}}{2} - \frac{\delta_o}{\mu_{oi}} (F_{oi} - \mu_{oi} f_{oi})$$

$$\left. - 2f_{oi} \right] x_i' \quad \ldots \text{(A4.4.6)}$$

$$= \frac{\delta_o}{2\lambda_o} \lim_{n\to\infty} \frac{1}{n} \sum_{1}^{n} \left[\left(1 - \frac{\delta_o}{2}\right) \frac{f_{oi}^2 \mu_{oi}}{1-F_{oi}} - f_{oi} \right.$$

$$\left. - \frac{\mu_{oi}^2 f_{oi}}{\sigma_{oi}^2} \left(1 - \frac{\delta_o}{2}\right) + \frac{\delta_o f_{oi}}{2} - \frac{\delta_o F_{oi}}{\mu_{oi}} \right] x_i'$$

$$\frac{\partial^2 Q(\theta_o)}{\partial\lambda\partial\sigma^2} = \lim_{n\to\infty} \frac{1}{n} \sum_{1}^{n} \left[- \frac{\delta_o \mu_{oi}^3 f_{oi}^2}{(1-F_{oi}).\, 4Z_{oi}\, \sigma_o^2} + \frac{\delta_o \mu_{oi}^4 f_{oi}}{4Z_{oi}\, \sigma_o^2\, \sigma_{oi}^2} \right.$$

$$\left. - \frac{\delta_o \mu_{oi}^2 f_{oi}}{4Z_{oi}\, \sigma_o^2} - \frac{\delta_o F_{oi}\, \mu_{oi}}{2Z_{oi}\, \sigma_o^2} + \frac{\delta_o \mu_{oi}^2 f_{oi}}{2Z_{oi}\, \sigma_o^2} \right] \quad \ldots (A\ 4.4.7)$$

$$= \lim_{n\to\infty} \frac{1}{n} \sum_{1}^{n} \left[\frac{\delta\mu_{oi}}{4Z_{oi}\, \sigma_o^2} \left\{ - \frac{\mu_{oi}^2 f_{oi}^2}{1 - F_{oi}} + \frac{\mu_{oi}^3 f_{oi}}{\sigma_{oi}^2} \right.\right.$$

$$\left.\left. + \mu_{oi} f_{oi} - 2F_{oi} \right\} \right]$$

and $$\frac{\partial^2 Q(\theta_o)}{\partial\lambda\partial\delta} = \lim_{n\to\infty} \frac{1}{n} \sum_{1}^{n} \left[- \frac{\delta_o}{1-F_{oi}} \frac{\mu_{oi}^3 f_{oi}^2}{4Z_{oi}} \ln Z_{oi} + \frac{\mu_{oi}^2 f_{oi}}{2Z_{oi}} \right.$$

$$+ \frac{\delta_o \mu_{oi}^2}{2Z_{oi}} \left(\frac{\mu_{oi}^2}{2\sigma_{oi}^2} - \frac{1}{2} \right) (\ln Z_{oi}) f_{oi} - \frac{\mu_{oi} F_{oi}}{2Z_{oi}}$$

$$\left. + \frac{\mu_{oi}}{2Z_{oi}} (F_{oi} - \mu_{oi} f_{oi}) - \frac{\delta_o \mu_{oi}}{2Z_{oi}} (F_{oi} - \mu_{oi} f_{oi}) \ln Z_{oi} \right]$$

$$= \lim_{n\to\infty} \frac{1}{n} \sum_{1}^{n} \left[\frac{\delta_o \mu_{oi} \ln Z_{oi}}{4Z_{oi}} \left\{ - \frac{\mu_{oi}^2 f_{oi}^2}{1 - F_{oi}} \right.\right.$$

$$\left.\left. + \frac{\mu_{oi}^3 f_{oi}}{\sigma_{oi}^2} - 2F_{oi} + \mu_{oi} f_{oi} \right\} \right] \quad \ldots (A\ 4.4.8)$$

PART II

SOME LARGE SAMPLE RESULTS IN AUTOCORRELATED MODELS WITH DECOMPOSED ERROR TERM

Chapter 5

Review of the Literature

5.1 Introduction

One of the important assumptions of the classical linear regression model is that the disturbances in the regression equation are uncorrelated. But quite often this assumption is unrealistic, i.e., the disturbances happen to be correlated among themselves. In the econometric literature this is referred to as the problem of autocorrelation. It is generally considered to be a problem for time-series-based investigations but it can also occur in cross-section data which are, for instance, based on some kind of natural ordering.

When the problem of autocorrelation arises, the classical model becomes inapplicable and so one has to use some alternative framework to tackle the problem. Extensive studies have been made in this area; in fact, the literature on autocorrelation has by now grown very large, a detailed account of which can be found in any standard textbook. What is basically done in these studies is to propose tests for detecting the presence of autocorrelation and to develop methods for consistent and efficient estimation of the parameters of the model in case autocorrelation is found to be present.

A review of the literature on autocorrelation is presented in this Chapter in order to make the dissertation self-contained. In doing so our objective is not to make a comprehensive survey of the existing literature in this area. A more detailed account of it may be found

in Judge et al. (1980). We shall here try to touch upon only those studies and results that are not readily available in textbooks and that would also help in appreciating the motivation of our work presented in the next two Chapters. This Chapter is arranged as follows. In section 5.2, the problem of misspecification as a source of autocorrelation is discussed. Studies regarding testing the presence of autocorrelation and the methods of estimation under different assumptions about the error process is described in section 5.3. Some concluding observations are given in section 5.4.

5.2 Autocorrelation and Misspecification

We start with the explanation often given for the existence of autocorrelation in the disturbances of a linear regression equation. It is stated in the literature that the sources of autocorrelation are omission of variables and/or errors-in-observations. Let us first discuss the consequences of the former being a source of autocorrelation.

Theil (1957) was the first to consider the problem of misspecification in a linear regression model. He primarily examined the consequences of misspecification for the estimates of the parameters. Although the term 'misspecification' refers to any kind of incorrect specification of the model, Theil used it (in fact, he used the term 'specification error') to mean situations where (i) some relevant explanatory variables have been excluded, (ii) some irrelevant variables have been included, and (iii) the form of the relationship has been wrongly specified. The reasons usually given for misspecification of a model are :

(i) ignorance of the analyst about the role of the particular regressors (or some functions of the regressors) which influence the regressand;

(ii) non-availability of reliable data on the regressors; and

(iii) to avoid the problem of multicollinearity.

Now, if misspecification is to give rise to autocorrelation in the disturbances, it should be true that some explanatory variables which should have been included in the regression have been omitted and that these variables are themselves autocorrelated. There is, however, one problem with this interpretation. As noted by Maddala (1977, p.274) : "It is that we should also be saying that these autocorrelated omitted variables, whatever they are, are independent of the explanatory variables included in the equation or else we run into the problem that the residuals are correlated with the explanatory variables."

5.2.1 Misspecification and OLS Estimation

Let us examine very briefly the consequences of misspecification (by which we shall henceforth mean situations where either some relevant regressors or some functions of some of the included regressors have been wrongly excluded) on the ordinary least squares(OLS) method of estimation of the parameters of the model. For this purpose, we consider the following model

$$Y = X\beta + \varepsilon \qquad \ldots (5.2.1)$$

where Y is a $(n \times 1)$ vector of observations on the dependent variable, X is a $(n \times k)$ matrix of fixed observations on k regressors, β is a

$(k \times 1)$ vector of associated regression coefficients and ε is the $(n \times 1)$ vector of disturbances with $E(\varepsilon) = 0$ and $V(\varepsilon) = \sigma_\varepsilon^2 I_n$. We further assume that rank of $X = k (< n)$. For such a model, the least squares estimator of β defined as $\hat{\beta} = (X'X)^{-1} X'Y$ is the best linear unbiased estimator of β.

Following Theil (1957), we assume that instead of using X, one takes some other nonstochastic matrix $\mathfrak{X}$ [1/] which is of order $(n \times k^+)$ with rank k^+ $(\leq k)$. The misspecified model can then be written as

$$Y = \mathfrak{X} \tilde{\beta} + \tilde{\varepsilon} \qquad \ldots \quad (5.2.2)$$

where $$\tilde{\varepsilon} = X\beta - \mathfrak{X} \tilde{\beta} + \varepsilon \qquad \ldots \quad (5.2.3)$$

and $\tilde{\beta}$ is the $(k^+ \times 1)$ vector of regression coefficients associated with the k^+ regressors in the misspecified equation. Clearly $E(\tilde{\varepsilon})$ is not, in general, equal to zero. If, however, the columns of X and $\mathfrak{X}$ are linearly related, then one may have $E(\tilde{\varepsilon}) = 0$ for some choice of $\tilde{\beta}$.

Now, the OLS estimator of $\tilde{\beta}$ for the model in (5.2.2) is given as

$$\hat{\tilde{\beta}} = (\mathfrak{X}'\mathfrak{X})^{-1} \mathfrak{X}'Y \qquad \ldots \quad (5.2.4)$$

$$= (\mathfrak{X}'\mathfrak{X})^{-1} \mathfrak{X}'X\beta + (\mathfrak{X}'\mathfrak{X})^{-1} \mathfrak{X}'\varepsilon \quad (\text{from}(5.2.1))$$

and hence

$$E(\hat{\tilde{\beta}}) = \mathfrak{P}\beta$$

where $$\mathfrak{P} = \{(\mathfrak{X}'\mathfrak{X})^{-1} \mathfrak{X}'X\}$$

is the $(k^+ \times k)$ matrix of (estimated) regression coefficients in the

---

1/ $\mathfrak{X}$ may be identical with X except that one or more columns of X have been deleted in $\mathfrak{X}$. It is also possible that X and $\mathfrak{X}$ are of the same order and are identical except for one or two columns.

auxillary regression equations of the true regressors (i.e., columns of X) on the included ones (i.e., columns of $\tilde{X}$). There is thus a simple linear relationship between the expectation of the estimated coefficients in the misspecified regression equation used by the analyst and the unknown vector β. In general, $\hat{\tilde{\beta}}$ is not an unbiased estimator of β if some relevant regressors have been omitted. In the same manner one can explicitly work out the bias when there is a misspecification of the form.

Theil also showed that the usual residual variance estimator from the misspecified model in (5.2.2) i.e.,

$$\hat{\sigma}^2_{\varepsilon} = \tilde{e}'\tilde{e}/(n - k^+)$$

where $\tilde{e} = Y - X\hat{\tilde{\beta}}$

and $\hat{\tilde{\beta}}$ is as defined in (5.2.4), overestimates the variance of the disturbance term i.e., σ^2_{ε} which is equal in both the true and misspecified models. This result is considered to provide a basis for selection of a model i.e., a specification with a smaller residual variance should be preferred to some other specification. Clearly, however, this criterion is not suitable in situations where the specifications being compared are both incorrect. It has also been pointed out by Koerts and Abrahamse (1970) that an analyst may make a wrong decision with a large probability by choosing a model with smaller residual variance if the sample size is small. Kloek (1975), on the other hand, showed that the probability of adopting a wrong model on the basis of residual variance criterion converges to zero as the

sample size increases. This criterion has also been justified by Schmidt (1974) who proved that even when the disturbances are generated by a first-order autoregressive process the residual variance criterion is valid asymptotically.

Recently Chaudhuri (1977, 1979) made further studies on the problem of misspecification. She was concerned with misspecification being a source of autocorrelation in the disturbances. She considered a k-variable linear regression model as described in (5.2.1) as the true model and defined the misspecified model as the one where m regressors (without any loss of generality, these may be treated as the last m regressors) have been omitted. Thus, the mispecified model considered by her is given by (5.2.2) with $k^+ = k-m$. She argued that β^+ i.e., the $(k^+ \times 1)$ vector of regression coefficients associated with the k^+ regressors included in the misspecified model, should be redefined in order that the included regressors may explain as much of the variation in y as possible.[2/]

She first considers the case when X is nonstochastic. It is easy to see then that $E\{(Y - X^+\beta^+)'(Y - X^+\beta^+)\}$ is minimized when $\beta^+ = P\beta$, $P = (X^{+\prime}X^+)^{-1} X^{+\prime}X$ i.e., we get, on the average, the best fit when β^+ is so chosen. We can then rewrite the misspecified model as

$$Y = X^+P\beta + \varepsilon^+ \qquad \ldots (5.2.5)$$

where $\varepsilon^+ = Y - X^+P\beta$

$$= Z + \varepsilon, \qquad Z=(X - X^+ P)\beta \qquad \ldots (5.2.6).$$

2/ For a critique of Chaudhuri's approach, see Gupta and Maasoumi (1979).

It is obvious that $E(\varepsilon^+) \neq 0$, in general, and $V(\varepsilon^+) = \sigma_\varepsilon^2 I_n$ though $E(\varepsilon^+ \varepsilon^{+\prime}) \neq \sigma_\varepsilon^2 I_n$. Also note that ε^+ cannot, in general, follow a AR(1) process even if ε_i's are so.

Now the OLS estimator in (5.2.2) is given as

$$\hat{\beta}^+ = (X^{+\prime} X^+)^{-1} X^{+\prime} Y$$

$$= \beta^+ + (X^{+\prime} X^+)^{-1} X^{+\prime} \varepsilon + (X^{+\prime} X^+)^{-1} X^{+\prime} Z$$

and therefore $E(\hat{\beta}^+) = \beta^+$ since $E(\varepsilon) = 0$ and $X^{+\prime} Z = 0$. The latter is obvious from the definition of Z in (5.2.6). Thus β^+ is unbiasedly estimated. But, as before, it can be shown that

$$E\left(\frac{e^{+\prime} e^+}{n-k^+}\right) > \sigma_\varepsilon^2$$

where $e^+ = Y - X^+ \hat{\beta}^+$

and hence the sampling variance of estimated coefficients will in such cases be overestimated.

Chaudhuri (1977, 1979) also examined the performance of the Durbin-Watson (DW) test in this situation allowing ε_t's to follow AR(1) process. She showed in particular that

$$\operatorname{plim}_{n \to \infty} d \simeq 2(1 - \rho_o) \qquad \ldots (5.2.7)$$

where

d = value of the DW statistic

$$\rho_o = \frac{\rho_\varepsilon \sigma_\varepsilon^2 + \rho_Z \sigma_o^2}{\sigma_\varepsilon^2 + \sigma_o^2}$$

$$\sigma_o^2 = \lim_{n \to \infty} \frac{1}{n} \sum_{t=1}^{n} Z_{\infty, t}^2$$

$$\rho_Z = \frac{\lim_{n\to\infty} \frac{1}{n} \sum_{t=2}^{n} Z_{\infty,\,t}\, Z_{\infty,\,(t-1)}}{\lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} Z^2_{\infty,\,t}}$$

$Z_{\infty,\,t}$ is the t-th element of the vector $(X - X^+ P_\infty)\beta$

$$P_\infty = \lim_{n\to\infty} P$$

and

ρ_ε is the first-order autocorrelation coefficient of ε_t's.

It is clear from (5.2.7) that $\operatorname{plim}_{n\to\infty} d$ is likely to be less than 2 for $\rho_\varepsilon > 0$ unless ρ_Z is negative and sufficiently large numerically. But even if $\rho_\varepsilon = 0$, the value of $\operatorname{plim}_{n\to\infty} d$ will come out to be significantly less than 2 if ρ_Z is positive and $\sigma_\varepsilon^2/\sigma_o^2$ is such that ρ_o is appreciably greater than zero. Thus DW test may indicate the presence of autocorrelation in the disturbances of the misspecified equation though these disturbances are in fact mutually uncorrelated.

When X is stochastic but independent of ε her observations with regard to the performance of OLS estimator and of the DW statistic can be summarised as follows :

(a) OLS would yield biased but consistent estimate of β^+.

(b) The usual formulae for estimating the sampling variances of these OLS estimates of β^+ may give underestimates, overestimates or unbiased estimates depending upon the relative effects of (i) omission of regressors, (ii) the autocorrelation among the disturbances in the true model (if any), and (iii) the autocorrelations of the included regressors.

(c) The disturbances in the misspecified model will, in general, have non-zero means and will not, in general, follow an AR(1) process. But these disturbances will be autocorrelated and if the first-order autocorrelation coefficient of the disturbances of the true model (i.e., ε_t's) and/or the autocorrelation coefficient of the vector Z in the misspecified model are positive, then the usual DW statistic would, in large samples, come out to be significantly less than 2 more often than the chosen level of significance.

5.2.2 Tests for Misspecification

Ramsey (1969, 1974) suggested tests for different types of misspecifications. We, however, briefly describe the most frequently used test known as regression specification error test (RESET) which is relevant for the kind of misspecifications we are considering.

He used Theil's (1965) best linear unbiased scalar (BLUS) residuals defined as

$$\tilde{\tilde{e}} = A' y$$

where $\tilde{\tilde{e}} = (\tilde{\tilde{e}}, \ldots\ldots, \tilde{\tilde{e}}_{n-k^+})'$

and A' is a $((n - k^+) \times n)$ matrix satisfying

$$A' X^+ = 0$$

$$\text{and } A' A = I_{n-k^+} \quad . \qquad \ldots (5.2.8)$$

Under the null hypothesis of no misspecification, $\tilde{\tilde{e}}$ is unbiased and has a scalar variance-covariance matrix. Ramsey (1969) observed that for misspecified models,

$$\tilde{\tilde{e}} \sim N(A'\xi, \sigma_\varepsilon^2 I_{n-k^+})$$

where ξ is a $((n - k^+) \times 1)$ nonstochastic vector whose precise definition depends upon the particular type of misspecification being considered and σ_ε^2 is the common variance of $\tilde{\tilde{e}}_t$'s $(t = 1, 2, \ldots, n - k^+)$. He also stated that under quite general conditions $A'\xi$ can be approximated by 3/

$$A'\xi \simeq \alpha_o + \alpha_1 q_1 + \alpha_2 q_2 + \cdots\cdots$$

where the q_j's $(j = 1, 2, \ldots\ldots)$ are $(n - k^+)$ dimensional vectors defined as

---

3/ If X^+ does not contain a column of unity, then

$$A'\xi \simeq \alpha_o A'i + \alpha_1 q_1 + \alpha_2 q_2 + \cdots\cdots$$

where $i = (1, 1, \ldots, 1)'$.

$$q_j = A' \hat{Y}^{(j+1)}, \qquad j = 1, 2, \ldots \qquad (5.2.9)$$

where $\hat{Y}^{(j+1)} = (\hat{y}_1^{(j+1)}, \hat{y}_2^{(j+1)}, \ldots, \hat{y}_n^{(j+1)})'$.

He then considered the following regression equation[4]

$$\tilde{e} = \alpha_o + \alpha_1 q_1 + \alpha_2 q_2 + \ldots + \alpha_p q_p + v \qquad \ldots (5.2.10)$$

where v is a $((n-k^+) \times 1)$ vector of independent disturbance terms assumed to be distributed normally with zero mean and constant variance. Clearly the null hypothesis of no misspecification reduces to testing whether all the α_i's ($\alpha = 0, 1, 2, \ldots\ldots$) are zero. This can be done by using the standard F-statistic which under the null hypothesis follows central F-distribution with $(p+1, n-k^+ - p - 1)$ degrees of freedom. Obviously the test depends upon the crucial assumption that $A'\xi$ can be approximated by q_j's ($j = 1, 2, \ldots\ldots$), an assumption which may not always be valid [see, Ramsey (1969), for details].

Ramsey and Gilbert (1972) carried out Monte Carlo studies on the power of RESET. Their main conclusion is that RESET is reasonably powerful against their respective alternatives.

It may be mentioned that RESET has been used by many researchers. Among them are Gilbert (1969), Ramsey and Zarembka (1971), Lee (1972), and Loeb (1976). Later Ramsey and Schmidt (1976) modified

---

4/ The number of q_j's to be included in the regression equation depends upon the particular problem. Ramsey, however, found that in most cases it was sufficient to use q_1, q_2 and q_3.

the RESET test procedure using OLS residuals (e^{+}_{t}'s, t = 1, 2, ..., n) instead of BLUS residuals. They suggested that e^{+} be regressed on $M_{X^+}Q$ instead of $\Lambda' Q$ where

$$M_{X^+} = [I - X^+ (X^{+\prime}X^+)^{-1} X^{+\prime}]$$

and $$Q = (\hat{Y}^{(2)}, \hat{Y}^{(3)}, \ldots, \hat{Y}^{(p+1)})$$

and the usual F-statistic used to test the joint significance of the regression coefficients of these regression equations.

5.3 Studies on Regression Models with Autocorrelated Disturbances

We shall now make a very brief review[5] of some of the studies on the problems of testing and estimation in a linear regression model with autocorrelated disturbances. It should be noted that most of the tests for autocorrelation treat the hypothesis of zero autocorrelation as the null hypothesis. If the null hypothesis is rejected, then the reestimation methods[6] are proposed which incorporate specific assumptions about the underlying error process. For a long time, the AR(1) process was the only process considered by econometricians. This was partly due to the fact that many of the data used for empirical work were annual series for which the AR(1) process seemed to be a reasonably good representation and partly because the estimation of autocorrelated models with more complicated error processes appeared to be cumbersome. With the advent of the modern computer, researchers are now increasingly

---

5/ For a detailed survey of the literature on autocorrelation, see Judge et.al. (1980). This portion of the survey draws much from this book.

6/ The term 'reestimation methods' is used in this thesis to refer to the usual two-step procedures as well as Hildreth-Lu search procedure.

using higher order AR as also other processes like the moving average (MA) or mixed autoregressive moving average (ARMA) processes. However, quite often the choice among these alternative error processes is not very simple or straightforward. In some instances theory might give indications[7] about the nature of the error process to be chosen but in many cases there may not be any a priori reason to choose a particular process.

5.3.1 Tests for Autocorrelation

There has been a growing literature on the tests for autocorrelation among the disturbances in a single equation regression model. However, most of these tests, particularly the earlier ones, consider the null hypothesis of zero autocorrelation against the alternative of AR(1) process for the errors. Relatively recently tests have been developed for error processes other than AR(1) as the alternative. In what follows we first describe some tests designed specifically for testing for AR(1) process as the alternative and then discuss some general tests suitable for testing the alternatives representing any type of autocorrelation.

(i) Tests against AR(1) alternatives: The most widely used test for autocorrelation when the alternative in an AR(1) process in the well-known Durbin-Watson (DW) test developed by Durbin and Watson (1950, 1951). The test-statistic used is given by

---

7/ As examples of how theory might help in deciding the process to be used we may cite the study by Nicholls, Pagan and Terrell (1975).

$$d = \frac{\sum_{t=2}^{n} (e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^2}$$

where e_t is the OLS residual.

Since the sampling distribution of d depends upon the data matrix X, it is impossible to tabulate the significance points that are relevant for all problems. They, however, showed that d will always lie between two other statistics d_L and d_U. Based on this, Durbin and Watson (1951) calculated the significance points d_L^* and d_U^* and suggested rejection of the null hypothesis of zero autocorrelation if $d < d_L^*$ and acceptance of it if $d > d_U^*$. The test is regarded as inconclusive if $d_L^* < d < d_U^*$. The alternative hypothesis here is one of positive autocorrelation. Subsequently, a number of studies have been made to consider this problem of inconclusiveness and propose ways of arriving at a conclusion if the DW test is inconclusive. These studies include the proposal of fitting a Beta distribution [Durbin and Watson (1951), Henshaw (1966)], using d_U^* in place of d^*, the appropriate significance point of d [Hannan (1957), Hannan and Terrell (1968)], fitting of a Beta distribution based on approximate moments [Theil and Nagar (1961)] etc.

Other suggestions include that made by Durbin and Watson (1971) to use an approximation called the '$a + bd_U$' approximation and also a technique originally suggested by Imhof (1961) and subsequently used by Koerts and Abrahamse (1969) to exactly determine d^*. Pan Jei-jian (1968) and L'Esperance, Chall and Taylor (1976) also considered the

problem and suggested techniques for determining d^*. However, it can be seen from the results of Durbin and Watson (1971) that the '$a + bd_U$' approximation is better than the other techniques in terms of computational convenience and accuracy.

A number of attempts have been made to circumvent the problem that the distribution of the DW statistic depends on the matrix X and that the OLS residuals are correlated even if the errors are not. Theil (1965, 1971, Ch. 5) suggested use of BLUS residuals for carrying out the von-Neumann ratio test which is based on the statistic

$$Q = \frac{\sum_{t=2}^{n} (e_t - e_{t-1})^2/(n-1)}{\sum_{t=1}^{n} (e_t - \bar{e})^2/n} .$$

Abrahamse and Koerts (1971) and Abrahamse and Louter (1971) developed another type of residuals which can be used in testing for autocorrelation of the type AR(1). Durbin (1970b) suggested a transformation that leads to yet another type of residuals. This transformation was subsequently modified by Sims (1975) and also used by Dent and Cassing (1978) in carrying out a Monte Carlo experiment.

So far as the power of the tests are concerned we may mention the study by Tillman (1975) in which he obtained approximate bounds on the power of the DW test. Other studies on the power of DW and the alternative tests include those of Abrahamse and Koerts (1969), Koerts and Abrahamse (1969), Abrahamse and Louter (1971), Phillips and Harvey (1974) L'Esperance and Taylor (1975) and Dubbelman, Louter and Abrahamse (1978).

The general conclusion of these studies is that the DW test is the best provided one has an accurate method of dealing with the inconclusive region.

The DW test has also been extended in other directions. Durbin (1970a) suggested a modification of the test for dealing with situations where the set of regressors include lagged dependent variables. Sargan (1964) suggested a test for AR(2) process when the null hypothesis is an AR(1) process. Monte Carlo studies on these tests have also been done [vide Maddala and Rao (1973), Kenkel (1974, 1975, 1976), Park (1975, 1976) and Spencer (1975)].

(ii) Some general tests : There is one group of tests which is based on the sample autocorrelation coefficients. These tests use the fact that a given process like AR, MA or ARMA can be described in terms of the nature of its autocorrelation functions

$$\rho_s = \frac{E(\varepsilon_t \, \varepsilon_{t-s})}{E(\varepsilon_t^2)}, \qquad \begin{array}{l} s = 1, 2, \ldots; \\ t = s+1, s+2, \ldots, n. \end{array}$$

It is well-known that for an AR process ρ_s gradually dies out with increase in the value of s and for a MA process it is zero after p, the order of the MA process.[8] One can estimate ρ_s by the sample correlation coefficient

$$r_s = \frac{\sum_{t=s+1}^{n} e_t \, e_{t-s}}{\sum_{t=1}^{n} e_t^2}$$

---

[8] For details on the behaviour of the autocorrelation functions for different processes, see Box and Jenkins (1976, Chapter 3).

where e_t is the OLS residual, and then come to a decision regarding the nature of the error process by examining the values of r_s (s=1, 2, ...). It should be pointed out, however, that this approach is based on the implicit assumption that e_t converges to ε_t in distribution as $n \to \infty$. There is an additional problem in small samples which arises because the e_t's are autocorrelated even if ε_t's are not. As has been shown by Malinvaud (1980, Ch. 13), this may give rise to substantial small-sample bias.

We may also note that, as early as 1946, Bartlett (1946) obtained an approximate expression for the sampling variance of r_s which can be used to test if ρ_s is significantly different from zero beyond a certain $s > m$, where m is the lag beyond which the population autocorrelation coefficient vanishes.

The second group of general tests for autocorrelation is the well-known likelihood ratio (LR) test. This test can be used to choose between two alternative AR or MA processes. Once this is done the order of the process is then chosen by testing a series of nested hypotheses. Suppose, for example, that the error process has been accepted to be AR. Then starting with a sufficiently high order of the process, say, AR(4), i.e.,

$$\varepsilon_t = \sum_{i=1}^{4} \rho_i \varepsilon_{t-i} + v_t$$

one can test the following sequence of null hypotheses against the specified alternatives by the LR tests :

Null hypotheses	Alternative hypotheses
$H_o^1 : \rho_4 = 0$	$H_1^1 : \rho_4 \neq 0$
$H_o^2 : \rho_3 = \rho_4 = 0$	$H_1^2 : \rho_3 \neq 0, \ \rho_4 = 0$
$H_o^3 : \rho_2 = \rho_3 = \rho_4 = 0$	$H_1^3 : \rho_2 \neq 0, \rho_3 = \rho_4 = 0$
$H_o^4 : \rho_1 = \rho_2 = \rho_3 = \rho_4 = 0$	$H_1^4 : \rho_1 \neq 0, \rho_2 = \rho_3 = \rho_4 = 0.$

It is clear that if any of the null hypotheses listed above is true then, of course, the preceding null hypotheses must be true; if, on the other hand, any of these hypotheses is false then the succeeding ones must be false. In view of this, one can continue testing the sequence of null hypotheses one by one till one of these is rejected or the final hypothesis of no autocorrelation is accepted. Obviously, this procedure cannot be generalized directly for an ARMA process. However, one can first decide on an AR or MA process and then treating that to be the null hypothesis carry out additional LR tests assuming the alternative to be an ARMA process [9].

In the third group we describe some other tests which do not have a common basis like those in the first two groups and which are used for alternatives other than AR(1) errors.[10] The test suggested by Durbin (1969)

9/ Detailed discussion of this approach and examples may be found in Fisher (1970), Anderson (1971, pp. 34-36), Kenward (1975) and Mizon (1977).

10/ The only exception, among the tests suggested here, is the one suggested by Godfrey (1978a) which can be used against an AR(1) process also.

is based upon the cumulative periodogram of least squares residuals. Geary (1970) proposed a nonparametric test which consists in counting the number of sign changes in the residuals and then testing the significance of the proportion by using the Binomial distribution. This test, however, is valid only asymptotically because the residuals are not independent even if the errors are uncorrelated. Wallis (1972) generalized the Durbin-Watson test and used it for quarterly data under the assumption that

$$\varepsilon_t = \bar{\rho}\varepsilon_{t-4} + v_t$$

where v_t is considered to be spherical. Recently Godfrey (1978a) suggested the use of the Lagrangian-multiplier technique to test the hypothesis of no autocorrelation against AR or MA process of any order. Later he (1978b) extended it for testing ARMA (p, q) against ARMA(p+r, q) or ARMA (p, q) against ARMA (p, q + r). These tests are valid asymptotically even if the X matrix contains lagged values of the dependent variable and, being based on OLS residuals, also easy to compute. There is, however, one shortcoming of these tests. In testing the hypothesis of no autocorrelation against an AR process or a MA process, they cannot distinguish between these two types of processes though they might help ascertain the order of the process because the test statistics used are the same for the two types of processes of same order.

Studies have also been made by Blattberg (1973) and Smith (1976) to compute the power of these various tests along with DW when the alternative is some non-AR(1) process. It has been found in general that the DW test performs quite satisfactorily, particularly when the alternative is AR(2) or MA(1).

5.3.2 Estimation

It is well-known that an application of OLS in an autocorrelated linear regression model gives inefficient though unbiased estimates of the regression coefficients. Therefore, the special methods of estimation developed for estimation in such a model seek to find efficient estimates of the parameters. The generalized least squares (GLS) method gives the BLUE(estimates) of the regression coefficients provided, of course, the variance-covariance matric V of the disturbances is known upto a factor of proportionality. It is, however, not usually known and hence the GLS can seldom be used in practice (see, however, below).

The methods of estimation developed in this context can be grouped into the following three categories.

(a) GLS estimation using $\hat{V}$, an estimator for V, henceforth referred to as the estimated GLS (EGLS) method of estimation.

(b) Nonlinear least squares (NLS) estimation.

(c) Maximum likelihood (ML) estimation.

We give here a brief account of these methods of estimation for each of several different assumptions made about the error process.

(i) Estimation with AR(1) process : Suppose the errors follow an AR(1) process given by

$$\varepsilon_t = \rho \varepsilon_{t-1} + v_t, \quad t = 1, 2, \ldots, n \qquad \ldots (5.3.1)$$

where $|\rho_\varepsilon| < 1$

$E(v_t) = 0$ for all t,

and $\text{Cov}(v_t, v_{t'}) = \begin{cases} 0 & \text{if } t \neq t' \\ \sigma_v^2 & \text{if } t = t'. \end{cases}$

For this case the variance-covariance matrix V depends on ρ only and hence one needs an estimate of ρ for obtaining the EGLS estimates. All the methods proposed in this context obtain the EGLS estimates in two steps and hence are often referred to as two-step procedures. While in both Cochrane-Orcutt (1949) and Prais-Winsten (1954) methods ρ is estimated, using the OLS residuals e_t, as

$$\hat{\rho} = \frac{\sum_{t=2}^{n} e_t e_{t-1}}{\sum_{t=1}^{n} e_t^2} \qquad \ldots (5.3.2)$$

in Durbin's (1960) procedure $\hat{\rho}$ is taken to be the estimate of the regression coefficient associated with y_{t-1} in the following regression equation

$$y_t = \rho y_{t-1} + (x_t' - \rho x_{t-1}')\beta + v_t$$

where x_t' is the t-th row of X. It can be shown that the estimates of ρ by both the above methods are consistent.

In the second stage of Prais-Winsten method the original model is transformed to $Y^* = X^*\beta + \varepsilon^*$, where $X^* = P_1 X$, $Y^* = P_1 Y$, $\varepsilon^* = P_1 \varepsilon$ and the (n x n) transformation matrix P_1 is such that $P_1' P_1 = (1 - \hat{\rho}^2)\, \hat{V}^{-1}$. Estimate of β is then obtained by applying OLS to the transformed model. The transformation matrix $\bar{P}_1$ used in Cochrane-Orcutt and Durbin's procedures is the same as P_1 with the first row being deleted. It is thus of order ((n-1) x n) and the number of observations after transformation reduces by one. Although this reduction may not have any effect on the estimates asymptotically, it

may result in some loss in efficiency in small samples, particularly if there is a trend in the explanatory variables and/or there is strong multicollinearity[11]. Also, strictly speaking, Cochrane-Orcutt or Durbin's estimators cannot be regarded as EGLS estimators.

The NLS estimates of β, ρ and σ_ε^2 are obtained by minimizing either $(\varepsilon_1^{*2} + \sum_{t=2}^{n} v_t^2)$ where $\varepsilon_1^* = (\sqrt{1 - \rho^2}) \varepsilon_1$ or $\sum_{t=2}^{n} v_t^2$.

The second alternative has been used, among others, by Cochrane and Orcutt (1949), Hildreth and Lu (1960) and Kmenta (1971). It may be pointed out that all these authors have loosely mentioned these NLS estimates as ML estimates. It is easy to see that to obtain the ML estimator of β, ρ and σ_ε^2 one has to maximize

$$- \frac{n}{2} \ln (\varepsilon_1^{*2} + \sum_{t=2}^{n} v_t^2) + \frac{1}{2} \ln (1 - \rho^2)$$

which has an additional term $\frac{1}{2} \ln (1 - \rho^2)$ as compared to what one would minimize to get the NLS estimates. This additional term is likely to affect the estimates unless the sample is large and ρ is not too close to one.

Now so far as the actual methods of computation for obtaining the NLS or ML estimates are concerned one has to use either an

---

11/ For details, see Kadiyala (1968), Poirier (1978b), Chipman (1979), Doran (1979), Maeshiro (1979), Spitzer (1979) and Park and Mitchell (1980).

iterative or a search procedure which is quite well-known. Of the iterative procedures, particular mention may be made of the method suggested by Cochrane and Orcutt (1949) for obtaining the NLS estimates by minimizing $\sum_{t=2}^{n} v_t^2$. This iterative process has been shown by Sargan (1964) to converge at least to a local minimum. For NLS estimation, Hildreth and Lu (1960) has suggested a search procedure. According to this procedure one has to find the residual sum of squares (RSS) for each of a number of selected values of ρ over the admissible range of -1 to +1 and the value of ρ for which the RSS is minimum is chosen as the estimate of ρ, and the estimated β corresponding to this ρ as the estimate of β. Recently specific algorithm for obtaining ML estimates have been suggested by Hildreth and Dent (1974) and by Beach and MacKinnon (1978a).

(ii) Estimation with AR(2) process : Let the error process be described as

$$\varepsilon_t = \rho_1 \varepsilon_{t-1} + \rho_2 \varepsilon_{t-2} + v_t, \qquad t = 1, 2, \ldots, n \quad \ldots(5.3.3)$$

where v_t satisfies conditions stated in (5.3.1)

and $\rho_1 + \rho_2 < 1$, $\rho_2 - \rho_1 < 1$ and $-1 < \rho_2 < 1$ (conditions for stationarity).

As in the AR(1) process, ρ_1 and ρ_2 are usually unknown and can be estimated as

$$\left.\begin{aligned} \hat{\rho}_1 &= \hat{\theta}_1 (1 - \hat{\theta}_2)/(1 - \hat{\theta}_1^2) \\ \text{and} \quad \hat{\rho}_2 &= (\hat{\theta}_2 - \hat{\theta}_1^2)/(1 - \hat{\theta}_1^2) \end{aligned}\right\} \quad \ldots (5.3.4)$$

where $$\hat{\theta}_s = \frac{\sum_{t=s+1}^{n} e_t e_{t-s}}{\sum_{t=1}^{n} e_t^2} \qquad (s = 1, 2)$$

and e_t's are OLS residuals. Alternatively, $\hat{\rho}_1$ and $\hat{\rho}_2$ can also be obtained by fitting the following equation by OLS

$$e_t = \rho_1 e_{t-1} + \rho_2 e_{t-2} + \hat{v}_t \; ; \qquad t = 3, 4, \ldots, n.$$

Then using a (n x n) transformation matrix P_2 one can, as in the previous case, obtain EGLS estimator for β as $(X' \hat{P}_2' \hat{P}_2 X)^{-1} X' \hat{P}_2' \hat{P}_2 Y$ where $\hat{P}_2$ is the estimate of P_2 from either of the above procedures.

One can also use NLS and ML methods of estimation. As in the AR(1) case, one minimizes $\varepsilon' P_2' P_2 \varepsilon$ for NLS and one maximizes

$$-\frac{n}{2} \ln \{ \varepsilon' P_2' P_2 \varepsilon \} + \frac{1}{2} \ln \left[(1 + \rho_2)^2 \{ (1-\rho_2)^2 - \rho_1^2 \} \right]$$

for ML estimation. Sometimes one minimizes $\varepsilon' \bar{P}_2' \bar{P}_2 \varepsilon$ where $\bar{P}_2$ is a $(\overline{n-2} \times n)$ transformation matrix obtained by deleting the first two rows of P_2. It is easily seen from the above that NLS and ML estimators need not be identical because of the additional term in the log-likelihood function. For both NLS and ML estimators, one has, however, to use standard iterative techniques to obtain estimates of the parameters. Beach and MacKinnon (1978b) provided an algorithm for obtaining ML estimates by incorporating the stationarity conditions which the other standard iterative methods do not. It is, therefore, expected that this estimator would do better in small samples provided the model specification is correct. One can as well use a search technique similar to the one suggested by Hildreth and Lu (1960) in the context of AR(1) error process; this will, however, be computationally more expensive now.

(iii) Estimation with AR(q), $q \geq 3$ processes : For AR processes of order 3 or more, it is very difficult to obtain the transformation matrix. For these processes, therefore, NLS and ML estimators are generally used. For such processes, a general expression for V^{-1} has been given by Wise (1955). Fuller (1976, p. 423) has defined the required transformation matrix which is very complicated. Pagan (1974) has indicated in details how the Gauss-Newton algorithm can be used to minimize $\sum_{t=q+1}^{n} v_t^2$ where q is the order of the process. Pagan and Byron (1978), on the other hand, has given a summary of the various modifications of NLS and ML estimators. Finally, mention may be made of the study by Thomas and Wallis (1971), who assumed the following type of error process :

$$\varepsilon_t = \rho \varepsilon_{t-4} + v_t \qquad \ldots \quad (5.3.5)$$

for quarterly data. It is possible to estimate ρ by replacing ε_t's by OLS residuals. NLS and ML estimators can also be appropriately obtained for this particular specification.

(iv) Estimation with MA(1) process : For this case the error process is given by

$$\varepsilon_t = v_t + \alpha_1 v_{t-1}, \quad (t = 1, 2, \ldots, n) \qquad \ldots (5.3.6)$$

where v_t satisfies, as before, the standard conditions in (5.3.1). Further $|\alpha_1| < 1$ for invertibility.

Let us also write

$$V(\varepsilon) = \sigma_\varepsilon^2 V_1 = \sigma_v^2 W_1$$

where $\sigma_\varepsilon^2 = \sigma_v^2 (1 + \alpha_1^2)$ and $W_1 = (1 + \alpha_1^2) V_1$.

For known α_1, GLS estimator of β can be obtained by minimizing $(Y - X\beta)' W_1^{-1} (Y - X\beta)$. It is also possible to derive a transformation matrix, say P_3, such that $P_3' P_3 = W_1^{-1}$; however, the structure of P_3 is complicated. Pesaran (1973) suggested a structure of P_3 which is complicated and involves the characteristic vectors of W_1[12/]. Since α_1 is usually unknown, it may be estimated from OLS residuals as

$$\hat{\alpha}_1 = \frac{1 - (1 - 4\hat{\rho}_1^2)^{\frac{1}{2}}}{2\hat{\rho}_1} \qquad \ldots \quad (5.3.7)$$

where $\hat{\rho}_1$ is $\hat{\rho}$ defined in (5.3.2). However, this estimator is not very satisfactory since $\hat{\alpha}_1$ is real only when $|\hat{\rho}_1| < 0.5$. Furthermore, the invertibility condition requires that $|\alpha_1| < 1$ and hence

$$|\rho_1| = \left| \frac{\alpha_1}{1 + \alpha_1^2} \right| < 0.5$$

where ρ_1 is the first-order autocorrelation coefficient. Although it is easy to estimate α_1 in this manner, it is usually inefficient relative to NLS estimator because it ignores the information contained in other sample autocorrelations.

12/ Very recently, Balestra (1980) has also given an exact form of P_3.

Durbin (1959) suggested approximating MA(1) with a large but finite order AR process such as

$$\varepsilon_t = \rho_1 \varepsilon_{t-1} + \rho_2 \varepsilon_{t-2} + \cdots + \rho_m \varepsilon_{t-m} + v_t .$$

Estimates of ρ_i's and hence of α_1 can be obtained by replacing ε_t's by e_t's and using OLS to estimate the equation above. Amemiya (1973c), McClave (1973), Fuller (1976) and Mentz (1977) worked further in this direction and showed that in an approximate sense, this estimator of α_1 is consistent and asymptotically efficient. Walker (1961), Hannan (1969), McClave (1974), Nelson (1974), Nicholls, Pagan and Terrell (1975) are among others who suggested other estimators.

As for NLS and ML estimators for models with MA(1) error process, one minimizes $\varepsilon' W_1^{-1} \varepsilon$ for NLS estimator and maximizes

$$-\frac{n}{2} \ln (\varepsilon' W_1^{-1} \varepsilon) - \frac{1}{2} \ln \left\{ \frac{1 - \alpha_1^{2n+2}}{1 - \alpha_1^2} \right\}$$

for ML estimator. Obviously, NLS estimates are likely to be different from ML estimates because of the additional term in the latter. However, for large n and small $|\alpha_1|$ the effect of the additional term in ML function is negligible. Both the ML and NLS estimates may be obtained by using either the standard iterative methods or the method suggested by Pesaran (1973).

(v) Estimation with MA(p), $p \geq 2$ processes : Let ε_t's follow a MA(p) process given as

$$\varepsilon_t = v_t + \sum_{j=1}^{p} \alpha_j v_{t-j} , \quad (t = 1, 2, \ldots, n) \quad \ldots (5.3.8)$$

where v_t's satisfy the conditions as in (5.3.1) and α_j's are such

that invertibility conditions [see Box and Jenkins (1976, p. 67)] hold. Following Phillips (1966) and Pagan and Nicholls (1976) we can conclude that minimizing $(v'v + \bar{v}'\bar{v})$,

where $v' = (v_1, \ldots, v_n)$

and $\bar{v}' = (v_0, v_{-1}, \ldots, v_{-(p-1)})$

with respect to

$$\beta, \alpha = (\alpha_1, \alpha_2, \ldots, \alpha_p)' \text{ and } \bar{v}$$

is equivalent to minimizing $\varepsilon' W_2^{-1} \varepsilon$ with respect to β and α where

$$V(\varepsilon) = \sigma_\varepsilon^2 V_2 = \sigma_v^2 W_2,$$

$$\text{and } W_2 = (1 + \sum_{i=1}^{p} \alpha_i^2) V_2.$$

This sum of squares, as before, can be minimized by using standard nonlinear iterative techniques. Phillips (1966), Trivedi (1970) and Hendry and Trivedi (1972) suggested minimization of $v'v$ with respect to β, α and $\bar{v}$. Under the assumption of normality of v_t's, this is equivalent to ML estimation conditional upon $\bar{v}$. The ML estimates of β and α can be found by minimizing $|W_2|^{\frac{1}{n}} (v'v + \bar{v}'\bar{v})$ with respect to β, α and $\bar{v}$. In addition to standard iterative methods, Pagan and Nicholls (1976), Osborn (1976) and Dent and Min (1978) suggested different algorithms for obtaining the estimates.

Asymptotically, there is no basis of choosing between ML estimator on the one hand, and the different NLS estimators which differ in their treatment of presample values $\bar{v}$, on the other. Under certain conditions, Pierce (1971a) showed that these estimators will be consistent,

asymptotically normally distributed and will have a covariance matrix that can be estimated by using the inverse of the matrix of second derivatives of the log-likelihood function (or the approximations being used in NLS or ML algorithms). Computationally, estimators that set $\bar{v} = 0$ are easier though they ignore some information and hence in terms of the mean square error criterion, they may be relatively poor in small samples. The ML estimator is likely to be preferred to NLS but computationally it is more difficult since it requires evaluation of $|W_2|$.

(vi) Estimation with ARMA processes : Let ε_t follow an ARMA (q, p) process given by

$$\varepsilon_t = \sum_{i=1}^{q} \rho_i \varepsilon_{t-i} + v_t + \sum_{j=1}^{p} \alpha_j v_{t-j} \quad , (t= 1, 2, \ldots, n) \quad \ldots(5.3.9)$$

where v_t's satisfy conditions in (5.3.1). The standard stationarity and invertibility conditions are assumed to hold. Since obtaining the transformation matrix even for ARMA(1, 1) process is difficult[13], it is better, as for the higher order MA processes, to minimize a simple sum of squares instead of $\varepsilon' W_3^{-1} \varepsilon$ (where $V(\varepsilon) = \sigma_\varepsilon^2 V_3 = \sigma_v^2 W_3$ and W_3 is accordingly defined in terms of α, $\rho = (\rho_1, \rho_2, \ldots, \rho_q)'$ and V_3). Following Pagan and Nicholls (1976), it can be shown that minimizing $v'v + \bar{w}' \Omega^{-1} \bar{w}$ with respect to β, ρ, α and $\bar{w}$ is equivalent to minimizing $\varepsilon' W_3^{-1} \varepsilon$ with respect to β, α and ρ, where v, $\bar{v}$ are

---

13/ For ARMA(1, 1) case, however, Tiao and Ali (1971) has given a matrix $\bar{Q}$ such that $\bar{Q}' \bar{Q} = W_3^{-1}$ and thus in this case one can simply minimize $\varepsilon' \bar{Q}' \bar{Q} \varepsilon$ to obtain the estimates.

defined as before and $\bar{w}' = (\bar{v}', \bar{\varepsilon}')$, $\bar{\varepsilon}' = (\varepsilon_0, \varepsilon_{-1}, \ldots, \varepsilon_{1-q})$, $E(\bar{w}\bar{w}') = \sigma_v^2 \Omega$ and $\bar{w}$ and v are uncorrelated. The exact structure of Ω depends on the order of the process. The expression for Ω for ARMA(1, 1) process has been given by Newbold (1974).

ML estimates under the assumption of normality are obtained by minimizing $|W_3| \frac{1}{n} (v'v + \bar{w}'\Omega^{-1}\bar{w})$ with respect to β, ρ, α and $\bar{w}$. The asymptotic properties of the ML and NLS estimates are equivalent and are given by Pierce (1971a, 1971b). In terms of small sample properties, ML estimates are likely to be preferred but computationally they are more difficult than both NLS which does not require computation of $|W_3|$ and conditional NLS which does not require computation of $|W_3|$ and $\bar{w}'\Omega^{-1}\bar{w}$.

5.4 Conclusions

The foregoing discussion clearly indicates that while significant amount of work has been done on misspecification and autocorrelation separately, not much serious attempt seems to have been made to treat misspecification (and autocorrelation due to it, if any) and autocorrelation due to reasons other than misspecification simultaneously within a single model. Since it is recognized in the literature that there can be different sources of autocorrelation, we want to argue that a proper understanding of the nature of the problem of autocorrelation should be based on a detailed study of the model where the factors

responsible for causing autocorrelation are represented separately rather than through a combined error term. We propose to adopt such an approach and reconsider the problems of testing and estimation of an autocorrelated linear regression model in the next two Chapters.

Chapter 6

Estimation and Testing in an Autocorrelated Linear Regression Model with Decomposed Error Term: The Case of Two AR(1) Components

6.1 Introduction

In the standard econometric literature estimation of a linear regression equation with autocorrelated errors is done by assuming some autocorrelation structure for the errors like that implied by an autoregressive (AR) or a moving average (MA) process. This kind of approach is not easy to justify if the error term in a linear regression equation is viewed as representing the effects of omission of variables, errors-in-observations etc., for, strictly speaking, the autocorrelation in the error term can be due to either misspecification of the equation or errors-in-observations or both. One might, therefore, be interested in examining if it is possible to detect whether an observed autocorrelation is due to one or both of these factors and how one should go about estimating the model in case both the factors of the error term are found to contribute to the autocorrelation.

The objective of the present Chapter is to make such a decomposition of the autocorrelated error term in a linear regression model in order to examine the nature of the problems arising out of it and propose methods which can be used to tackle those problems. More specifically, we consider an autocorrelated linear regression equation model in which the error term is written as the sum of two compoments, one or both of which may be autocorrelated and then examine the problems of

estimation and hypothesis testing in such a mosel. It will be seen that development of a simple and convenient method of estimation would require a knowledge about the nature of the error term since one needs the variance-covariance matrix of the error term in estimating the parameters of the model. Our approach is to get the proper variance-covariance matrix after ascertaining whether both the components of the error term are present or not and also, autocorrelated or not, if present.

It may be mentioned that a situation like the one described above may actually arise if there is misspecification in the form of omission of variables. It can then be easily shown that the error term in the misspecified model will have two additive components [see, for example, Ramsey (1969), Chaudhuri (1977, 1979) etc.] — one is the disturbance term in the true regression and the other is due to misspecification — and the autocorrelation in the error term of the misspecified equation would then be due to autocorrelation in either or both of the individual components.

We shall first point out that if the error term is broken up into its constituent elements in the manner described above, then it will not, in general, follow the same process which may be assumed to be generating the individual components. If, for example, the individual components are generated by AR(1) processes, then the sum of these two components will not be of the same type[1]; it is in fact known that the sum of two independent AR(1) processes is an ARMA(2, 1) process [vide Granger and Morris (1976), Rose (1977)]. Hence the common

1/ The same conclusion will be valid even if one of the two components is not autocorrelated.

practice of assuming an AR(1) process for the combined error term, without paying any attention to its possible components, in an autocorrelated linear regression equation becomes unsatisfactory. In fact, some of the standard estimation methods like those of Cochrane-Orcutt and Prais-Winsten which are applicable only if the error follows an AR(1) process might then be inefficient.

One might argue that since the combined error term follows an ARMA(2, 1) process when the individual components are generated by AR(1) processes, it is possible to estimate the parameters of the model by using methods available for estimating a model with ARMA(2, 1) error process [see, for example, Pierce (1971b)]. Hence, in principle, one could possibly say that the problem of estimation of such a model has been solved.

In our opinion, however, there are still valid reasons for undertaking the present type of study. It would be convenient to present our arguments by considering one of the error components as due to misspecification. Then the idea behind the decomposition of the error term and the assumption of a separate error process for each of the components is precisely to see if we can ascertain whether an autocorrelation is due partly or wholly to misspecification. Such an information is helpful for various reasons. Firstly, one can then try to remove the misspecification by, possibly, including other regressors in the model. Secondly, as has been noted by Judge et al. (1980), if autocorrelation in the error results from misspecification of the data matrix, it is likely that other assumptions of the model, namely, that expectation

of the error is zero, that the error is independent of the regressors and that the errors are homoscedastic will be violated. It is because of these reasons that Judge et al. (1980, p.171) have observed : "Because of this association between the misspecification of X and auto-correlated errors, we should exercise caution when autocorrelation is diagnosed. It may mean that the errors can be reasonably represented by some stochastic process, and experience suggests it is quite often the case, but it could also mean that X matrix should be respecified. The crucial question, and it requires further research, is the effect of incorrectly treating a misspecified equation as one with autocorrelated errors". Maddala (1977, p. 291), on the other hand, pointed out that "when serial correlation in the residuals is due to omitted variables that are themselves autocorrelated, the question of whether or not the usual procedures of 'efficient' estimation often suggested in textbooks are better than ordinary least squares is a point that needs more careful investigation". Our study is expected to throw light on the questions raised by the above-quoted authors though we have not explicitly examined the departures from standard assumptions listed by Judge et al. (see above).

It will be seen that though we start with a model whose error term is decomposed into two components with the possibilities of both being generated by AR(1) process, we are able to identify the different autocorrelated situations. We suggest using our test to first ascertain what exactly a given situation is and then estimating the model according

to the situation. This approach helps us know exactly when the standard reestimation methods or the OLS are efficient and thus clearly indicates the limitation and applicability of these methods. We think that this also answers the question raised by Maddala (1977) (see above). A straightforward use of an ARMA(2, 1) process for the combined error term would fail to provide the insight into the problem of autocorrelation.

Our test procedure as well as the estimation method developed to tackle the situation where standard methods are not applicable are based on OLS residuals and hence are computationally simple. We may note that the use of OLS residuals has been preferred to other types of more complicated residuals in a number of cases. It is well-known that Theil's BLUS residuals do not suffer from the limitation of OLS residuals. But yet econometricians usually prefer using DW test based on OLS residuals rather than von-Neumann test based upon BLUS residuals because of the computational simplicity of the former. We may also refer to a study by Godfrey (1978a) who developed a test for detecting autocorrelation of different processes, based on OLS residuals. His test is asymptotically equivalent to the LR test, and has the advantage of being based on OLS residuals.

It is well-known that for an ARMA(2, 1) process the structure of the variance-covariance matrix V would be complicated compared to the one obtained here with assumption of two independent AR(1) components and estimation of a model with an ARMA (2, 1) process is not quite simple, since it involves, among other things, the computation of the determinant

of the variance-covariance matrix V in the ML method of estimation. As contrasted to this, our proposed method of estimation even for cases where standard reestimation methods are inapplicable, is computationally much simpler.

We may finally note that a straightforward use of an ARMA (2, 1) process or some such complicated process to deal with the problem of autocorrelation totally obscures the situation. For an ARMA(2, 1) assumption, we have no way of knowing if it was, in fact, sum of two independent AR(1) components. That such information is necessary, has already been noted in the previous paragraphs. In any case there is no a priori basis for using a process like ARMA(2, 1).

In what follows we first try to develop a test procedure which would help in identifying different situations characterized by different combinations of the parameters involved in the variance-covariance matrix of the error term. It will be seen that the test suggested here cannot exhaustively distinguish all possible cases. There are a number of situations which are indistinguishable by our test. Fortunately, however, this failure does not pose any problem so far as efficient estimation is concerned.

In section 6.2 we present the model. While in section 6.3 the nature of the error process is discussed, the nature of autocorrelation of the error term is characterized in section 6.4. section 6.5 deals with the development of a large sample test for identifying the sources of autocorrelation. The estimation method is described in section 6.6 and the concluding observations are presented in section 6.7.

6.2 The Model

We consider a k-variable linear regression equation which is written in matrix notation as

$$Y = X\beta + \varepsilon^{+} \qquad \ldots (6.2.1)$$

where X is (n x k) data matrix on k regressors, $\beta = (\beta_1, \beta_2, \ldots, \beta_k)'$ is the vector of associated regression coefficients, Y is the (n x 1) vector of observations on the regressand and ε^{+} is the (n x 1) vector of disturbances.[2]

We make the following assumptions[3]:

(i) $\varepsilon_t^{+} = \varepsilon_t + Z_t$ for all $t = 1, 2, \ldots, n$ ... (6.2.2)

2/ One of the regressors in the above equation may be taken to be unity for all t incorporating thereby an intercept in the equation.

3/ If the regression equation in (6.2.1) is considered as a misspecified equation, then we show below that the assumption in (6.2.2) is indeed true.

Let us write, following Chaudhuri (1977, 1979), the true model as

$$Y = \tilde{X}\tilde{\beta} + \varepsilon$$

where $\tilde{X}$ is $(n \times \tilde{k})$ data matrix on $\tilde{k}$ $(> k)$ regressors, $\tilde{\beta}$ is the $(\tilde{k} \times 1)$ vector of associated regression coefficients and ε is the (n x 1) vector of disturbances in the true regression equation. From (6.2.1) we have

$$\varepsilon^{+} = Y - X\beta$$

$$= \varepsilon + (\tilde{X}\tilde{\beta} - X\beta) \quad \text{(using the true regression equation given above).}$$

Thus

$$\varepsilon^{+} = \varepsilon + Z$$

where $Z = (\tilde{X}\tilde{\beta} - X\beta)$,

i.e., the component in ε^{+} which is due to misspecification. Following Chaudhuri's argument that β should be redefined in order to allow

contd ...../

(ii) X is stochastic but distributed independently of both ε and Z[4/]

$$\varepsilon = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)' \quad \ldots (6.2.3)$$

and $Z = (Z_1, Z_2, \ldots, Z_n)'$

(iii) ε is independent of Z

(iv) $\varepsilon_t = \rho_\varepsilon \varepsilon_{t-1} + u_t \quad \ldots (6.2.4)$

$\ldots (6.2.5)$

where $|\rho_\varepsilon| < 1$ and u_t's are independently distributed with zero mean and constant variance σ_u^2.

Footnote 3 contd.

the included regressors to capture as much of the unexplained variation in y as possible, it can then easily be shown that

$$\operatorname{plim}_{n \to \infty} \left(\frac{1}{n} X'Z \right) = 0$$

(For details, see Chaudhuri (1977, 1979, Chapters 2 and 3) and Section 9.2 of this dissertation. There is, however, a slight change in notation : $\tilde{X}$, $\tilde{\beta}$, X and β here are X, β, X^+ and β^+ respectively there).

It can be shown that application of OLS in (6.2.1) would yield consistent estimates of the regression coefficients of the misspecified equation but the estimated variance-covariance matrix of these estimates would not be consistent. Also, as Chaudhuri (1979) has shown, the DW statistic may come out to be significantly less than 2 more often than the predetermined first kind of error. This result is valid for our model also. There is thus a risk in applying DW test for detecting autocorrelation in the error term in this situation.

4/ The assumption that X and Z are independent is needed only in the estimation procedure suggested and not in the tests proposed. We shall discuss the nature and relevance of this assumption in the concluding section.

(v) $Z_t = \rho_Z Z_{t-1} + v_t$ ... (6.2.6)

where $|\rho_Z| < 1$ and v_t's are independently distributed with zero mean and constant variance σ_v^2.

(vi) For an observed X, rank of $X = k < n$. ... (6.2.7)

(vii) $\underset{n \to \infty}{\text{plim}} \left(\frac{1}{n} X'X\right) = \Sigma_{XX}$, a positive definite matrix .. (6.2.8)

It is obvious from (6.2.5) and (6.2.6) that

$$\left.\begin{aligned} V(\varepsilon_t) = \sigma_\varepsilon^2 &= \frac{\sigma_u^2}{1-\rho_\varepsilon^2} \\ V(Z_t) = \sigma_Z^2 &= \frac{\sigma_v^2}{1-\rho_Z^2} \end{aligned}\right\} \quad \text{for all } t = 1,2,\ldots, n \quad \ldots (6.2.9)$$

6.3 Nature of the Error Process

It has been mentioned earlier that the combined error term will not, in general, follow the same process as that generating the individual components. We may now illustrate this for an AR(1) process. While the general result on this is well-known the following arguments may shed some light on the issues.

Let us assume that ε_t^+'s are given by

$$\varepsilon_t^+ = \rho\varepsilon_{t-1}^+ + w_t \quad \ldots (6.3.1)$$

where $|\rho| < 1$ and w_t's are independently distributed with zero mean and constant variance σ_w^2.

Then since $\varepsilon_t^+ = \varepsilon_t + Z_t$, we have from (6.3.1),

$$\varepsilon_t + Z_t = \rho(\varepsilon_{t-1} + Z_{t-1}) + w_t$$

or,

$$\varepsilon_t - \rho\varepsilon_{t-1} = \rho Z_{t-1} - Z_t + w_t \ . \qquad \ldots (6.3.2)$$

Let us now consider the first-order autocovariance of w_t given as

$$\begin{aligned}
\text{Cov}(w_t, w_{t-1}) &= \text{Cov}(\varepsilon_t - \rho\varepsilon_{t-1} - \rho Z_{t-1} + Z_t, \varepsilon_{t-1} - \rho\varepsilon_{t-2} - \rho Z_{t-2} + Z_{t-1}) \\
&= \rho_\varepsilon \sigma_\varepsilon^2 - \rho\sigma_\varepsilon^2 - \rho\rho_\varepsilon^2 \sigma_\varepsilon^2 + \rho^2\rho_\varepsilon \sigma_\varepsilon^2 + \rho_Z \sigma_Z^2 \\
&\quad - \rho\sigma_Z^2 - \rho\rho_Z^2 \sigma_Z^2 + \rho^2\rho_Z \sigma_Z^2 \\
&= \sigma_\varepsilon^2 (\rho_\varepsilon - \rho - \rho\rho_\varepsilon^2 + \rho^2\rho_\varepsilon) \\
&\quad + \sigma_Z^2 (\rho_Z - \rho - \rho\rho_Z^2 + \rho^2\rho_Z) \\
&= \sigma_\varepsilon^2 (1 - \rho\rho_\varepsilon)(\rho_\varepsilon - \rho) + \sigma_Z^2 (1 - \rho\rho_Z)(\rho_Z - \rho) \ .
\end{aligned}$$

$$\ldots (6.3.3)$$

Clearly, this is not, in general, equal to zero, as required for assumption (6.3.1) to be true. The covariance will be equal to zero only when

(i) $\rho = \rho_\varepsilon = \rho_Z$ or (ii) either $\sigma_Z^2 = 0$ or $\sigma_\varepsilon^2 = 0$ 5/.

We thus find that even if both ε_t and Z_t separately follow AR(1), ε^+ would not, in general, follow the same process. As noted earlier ε^+ would in fact, follow an ARMA(2, 1) process.

5/ Both σ_ε^2 and σ_Z^2 cannot be equal to zero because that would mean there is no disturbance term in the model. If one of them (say, σ_Z^2) is equal to zero, then obviously $\rho = \rho_\varepsilon$ and hence $\text{Cov}(w_t, w_{t-1}) = 0$.

6.4 Characterization of the Nature of Autocorrelation

We may now present a method of investigating the nature of the disturbance term[6/] under fairly mild assumptions. In order to describe the method conveniently we may first enumerate the possible situations as shown in the following table.

Table 6.1: Different models for the error term

	$\rho_\varepsilon = 0$	$\rho_\varepsilon \neq 0$
$\sigma_Z^2 = 0$	Case 3*	Case 5*
$\sigma_Z^2 > 0$	Case 1 : $\rho_Z \neq 0$	Case 2 : $\rho_Z = 0$
	Case 4 : $\rho_Z = 0$	Case 6 : $\rho_Z \neq 0$ and $\rho_\varepsilon \neq \rho_Z$
		Case 7 : $\rho_Z \neq 0$ and $\rho_\varepsilon = \rho_Z = \rho$ (say)

* For cases 3 and 5, $\sigma_Z^2 = 0$ which implies $Z_t = 0$ for all t. Therefore ρ_Z is not defined for these two cases.

It is possible to further characterize the different possibilities in terms of the parameters ρ_ε, ρ_Z, σ_ε^2 and σ_Z^2. For this purpose we consider the OLS residuals e^+'s where

$$e^+ = Y - X\hat{\beta} \text{ and } \hat{\beta} \text{ is the OLS estimator of } \beta.$$

6/ In order to ensure that the composite error term has at least one random component, autocorrelated or not, we assume, without any loss of generality, that $\sigma_\varepsilon^2 > 0$.

Now,

$$\hat{\beta} = (X'X)^{-1} X'Y = \beta + (X'X)^{-1} X'\varepsilon^{+} = \beta + (X'X)^{-1} X'(\varepsilon + Z)$$

and hence

$$\underset{n\to\infty}{\text{plim}}\ \beta = \beta + \underset{n\to\infty}{\text{plim}}\ (\tfrac{1}{n} X'X)^{-1} \{\underset{n\to\infty}{\text{plim}}\ (\tfrac{1}{n} X'\varepsilon) + \underset{n\to\infty}{\text{plim}}\ (\tfrac{1}{n} X'Z)\}$$

$$= \beta \text{ by assumptions (6.2.3) and (6.2.8)} . \qquad \ldots (6.4.1)$$

Therefore, e^{+} converges in distribution to $(\varepsilon + Z)$ as $n \to \infty$.

Let us now define

$$\hat{\rho}_1 = \frac{\sum_{t=2}^{n} e_t^{+} e_{t-1}^{+}}{\sum_{t=2}^{n} e_{t-1}^{+2}} .$$

Since e^{+} converges in distribution to $(\varepsilon + Z)$,

$$\underset{n\to\infty}{\text{plim}}\ \hat{\rho}_1 = \underset{n\to\infty}{\text{plim}}\ \frac{\sum_{2}^{n} (Z_t + \varepsilon_t)(Z_{t-1} + \varepsilon_{t-1})}{\sum_{2}^{n} (Z_{t-1} + \varepsilon_{t-1})^2} .$$

Hence,

$$\underset{n\to\infty}{\text{plim}}\ \hat{\rho}_1 = \frac{\underset{n\to\infty}{\text{plim}}\ \frac{1}{n} \sum_{2}^{n} \varepsilon_t \varepsilon_{t-1} + \underset{n\to\infty}{\text{plim}}\ \frac{1}{n} \sum_{2}^{n} Z_t \varepsilon_{t-1} + \underset{n\to\infty}{\text{plim}}\ \frac{1}{n} \sum_{2}^{n} Z_{t-1} \varepsilon_t + \underset{n\to\infty}{\text{plim}}\ \frac{1}{n} \sum_{2}^{n} Z_t Z_{t-1}}{\underset{n\to\infty}{\text{plim}}\ \frac{1}{n} \sum_{2}^{n} \varepsilon_{t-1}^2 + \underset{n\to\infty}{\text{plim}}\ \frac{1}{n} \sum_{2}^{n} Z_{t-1}^2 + 2\ \underset{n\to\infty}{\text{plim}}\ \frac{1}{n} \sum_{2}^{n} \varepsilon_{t-1} Z_{t-1}} .$$

$$\ldots\ (6.4.2)$$

Assuming that u_t's in (6.2.5) and v_t's in (6.2.6) have finite fourth order moments[7], we have

$$\underset{n\to\infty}{\text{plim}}\ \frac{1}{n}\sum_{2}^{n}\varepsilon_{t-1}^2 = \sigma_\varepsilon^2$$

$$\underset{n\to\infty}{\text{plim}}\ \frac{1}{n}\sum_{2}^{n}\varepsilon_t\,\varepsilon_{t-1} = \rho_\varepsilon\,\sigma_\varepsilon^2$$

$$\underset{n\to\infty}{\text{plim}}\ \frac{1}{n}\sum_{2}^{n}Z_{t-1}^2 = \sigma_Z^2$$

$$\underset{n\to\infty}{\text{plim}}\ \frac{1}{n}\sum_{2}^{n}Z_t\,Z_{t-1} = \rho_Z\,\sigma_Z^2\ .$$

Therefore (6.4.2) reduces to

$$\underset{n\to\infty}{\text{plim}}\ \hat{\rho}_1\ (=\bar{\rho}_1,\ \text{say}) = \frac{\rho_\varepsilon\,\sigma_\varepsilon^2 + \rho_Z\sigma_Z^2}{\sigma_\varepsilon^2 + \sigma_Z^2}\ . \qquad \ldots\ (6.4.3)$$

Let us next define

$$\hat{\rho}_2 = \frac{\sum_{3}^{n} e_t^+\, e_{t-2}^+}{\sum_{2}^{n} e_t^+\, e_{t-1}^+}\ .$$

Again by similar algebraic manipulations and making the same assumptions, we get

$$\underset{n\to\infty}{\text{plim}}\ \frac{1}{n}\sum_{3}^{n}\varepsilon_t\,\varepsilon_{t-2} = \rho_\varepsilon^2\,\sigma_\varepsilon^2$$

and

$$\underset{n\to\infty}{\text{plim}}\ \frac{1}{n}\sum_{3}^{n}Z_t\,Z_{t-2} = \rho_Z^2\,\sigma_Z^2$$

7/ Vide : Goldberger (1963), pp. 149-153.

and hence

$$\underset{n \to \infty}{\text{plim}} \hat{\rho}_2 \; (= \bar{\rho}_2, \text{ say}) = \frac{\rho_\varepsilon^2 \sigma_\varepsilon^2 + \rho_Z^2 \sigma_Z^2}{\rho_\varepsilon \sigma_\varepsilon^2 + \rho_Z \sigma_Z^2} \; . \qquad \ldots (6.4.4)$$

The values of $\underset{n \to \infty}{\text{plim}} \hat{\rho}_2$ for the different cases listed earlier may now be presented in the following table.

Table 6.2 : Values of $\underset{n \to \infty}{\text{plim}} \hat{\rho}_2 \; (= \bar{\rho}_2)$ for the different cases of Table 6.1

	$\rho_\varepsilon = 0$	$\rho_\varepsilon \neq 0$
$\sigma_Z^2 = 0$	Case 3* : 0	Case 5 : ρ_ε
$\sigma_Z^2 > 0$	Case 1 : ρ_Z	Case 2 : ρ_ε
	Case 4* : 0	Case 6 : $\dfrac{\rho_\varepsilon^2 \sigma_\varepsilon^2 + \rho_Z^2 \sigma_Z^2}{\rho_\varepsilon \sigma_\varepsilon^2 + \rho_Z \sigma_Z^2}$
		Case 7 : ρ where $\rho_\varepsilon = \rho_Z = \rho \neq 0$

* Strictly speaking, for cases 3 and 4, $\bar{\rho}_2 = \frac{0}{0}$ and hence undefined. In these two cases, what is really meant by the entries in the table is that $\underset{n \to \infty}{\text{plim}} \text{ cov } (e_t^+, e_{t-2}^+) = 0$.

6.5 A Large Sample Test for Detecting the Sources of Autocorrelation

Let us define the random variable

$$\theta_t = e_t^+ - \hat{\rho}_2 \; e_{t-1}^+ \; . \qquad \ldots (6.5.1)$$

As set out in Table 6.2, $\hat{\rho}_2$ takes different values in large samples for different cases.

Now let c_s be defined as

$$c_s = \text{Cov}\,(\theta_t, \theta_{t-s}) \qquad \text{for all } s>0 .$$

Then

$$c_s = \text{Cov}\,(e_t^+, e_{t-s}^+) - \hat{\rho}_2\,\text{Cov}\,(e_t^+, e_{t-s-1}^+) + \hat{\rho}_2^2\,\text{Cov}\,(e_{t-1}^+, e_{t-s-1}^+) - \hat{\rho}_2\,\text{Cov}\,(e_{t-1}^+, e_{t-s}^+) . \quad \ldots\ (6.5.2)$$

Now in large samples

$$\text{Cov}\,(e_t^+, e_{t-s}^+) \simeq \text{Cov}\,(Z_t, Z_{t-s}) + \text{Cov}\,(\varepsilon_t, \varepsilon_{t-s}) = \rho_Z^s\,\sigma_Z^2 + \rho_\varepsilon^s\,\sigma_\varepsilon^2 . \quad \ldots\ (6.5.3)$$

Therefore from (6.5.2) we have[8/]

$$c_s = \rho_Z^s\,\sigma_Z^2 + \rho_\varepsilon^s\,\sigma_\varepsilon^2 - \bar{\rho}_2\,\rho_Z^{s+1}\,\sigma_Z^2 - \bar{\rho}_2\,\rho_\varepsilon^{s+1}\,\sigma_\varepsilon^2 + \bar{\rho}_2^2\,\rho_Z^s\,\sigma_Z^2 + \bar{\rho}_2^2\,\rho_\varepsilon^s\,\sigma_\varepsilon^2 - \bar{\rho}_2\,\rho_Z^{s-1}\,\sigma_Z^2 - \bar{\rho}_2\,\rho_\varepsilon^{s-1}\,\sigma_\varepsilon^2 . \quad \ldots\ (6.5.4)$$

In order to examine the autocovariance structures of θ_t's, we calculate c_s's for $s>0$ for each of the seven cases and present the values in the following table where the seven cases are arranged under three groups. The algebra is straightforward for all the cases excepting (1) and (6) and hence is omitted.

For case (1), since $\rho_\varepsilon = 0$ and $\bar{\rho}_2 = \rho_Z$

$$c_1 = -\rho_Z\,\sigma_\varepsilon^2 \neq 0$$

and $c_s = 0$ for all $s \geq 2$.

8/ Henceforth, without any loss of generality, we will take c_s to be plim c_s.

We, therefore, conclude that for case (1), all but the first-order autocovariance of θ_t's are zeros.

For case (6), we find from (6.5.4), that

$$c_s = \sigma_\varepsilon^2 (\rho_\varepsilon^s - \bar{\rho}_2 \rho_\varepsilon^{s+1} + \bar{\rho}_2^2 \rho_\varepsilon^s - \bar{\rho}_2 \rho_\varepsilon^{s-1}) + \sigma_Z^2 (\rho_Z^s - \bar{\rho}_2 \rho_Z^{s+1} + \bar{\rho}_2^2 \rho_Z^s - \bar{\rho}_2 \rho_Z^{s-1})$$

$$= \sigma_\varepsilon^2 M + \sigma_Z^2 N, \text{ say.}$$

As $\sigma_\varepsilon^2 > 0$, $\sigma_Z^2 > 0$ for this case, c_s will be equal to zero if either of (i) $\sigma^2 M = -\sigma_Z^2 N$ and (ii) $M = 0 = N$ (then, obviously (i) is satisfied) happens. It is easy to see by the method of contradiction that neither (a) nor (b) can happen for case (6). Hence we may conclude that for case (6), $c_s \neq 0$ for all $s \geq 1$.

Table 6.3 : Autocovariance structures of θ_t for the different cases

Group	Case	Values of ρ_ε, ρ_Z, ρ_Z^2	Autocovariance structure
I	1	$\rho_\varepsilon = 0$, $\rho_Z \neq 0$, $\sigma_Z^2 > 0$	$c_1 \neq 0$, $c_s = 0$ $\forall$ $s \geq 2$
	2	$\rho_\varepsilon \neq 0$, $\rho_Z = 0$, $\sigma_Z^2 > 0$	
II	3	$\rho_\varepsilon = 0$, $\sigma_Z^2 = 0$	$c_s = 0$ $\forall$ $s \geq 1$
	4	$\rho_\varepsilon = 0$, $\rho_Z = 0$, $\sigma_Z^2 > 0$	
	5	$\rho_\varepsilon \neq 0$, $\sigma_Z^2 = 0$	
	7	$\rho_\varepsilon = \rho_Z = \rho \neq 0$, $\sigma_Z^2 > 0$	
III	6	$0 \neq \rho_\varepsilon \neq \rho_Z \neq 0$, $\sigma_Z^2 > 0$	$c_s \neq 0$ $\forall$ $s \geq 1$

It is now clear from the above table that we may distinguish among three broad groups of cases on the basis of the first two autocovariances of θ_t's. However, the discrimination among these groups of cases on the basis of zero/non-zero values of the autocovariances of θ_t's can euqivalently be done by using autocorrelation coefficients of θ_t's. Hence the test suggested here is based on the autocorrelation coefficients of θ_t's. The basic reason for doing this is that we can then apply Bartlett's well-known test [see Box and Jenkins (1976), pp.34-36 and Malinvaud (1980) pp. 442-444].

Bartlett's test is concerned with examining whether the population autocorrelation coefficients, R_k's ($k = 1, 2, \ldots\ldots$) of the θ_t's (say) ($t = 1, 2, \ldots.$) are effectively zero beyond a certain lag. The test rests on the assumption that the θ_t's follow a stationary normal process, and uses the following approximate expression for the variance of the estimated autocorrelation coefficients r_k ($k = 1, 2, \ldots.$) of a stationary normal process :

$$V(r_k) \simeq \frac{1}{N} \sum_{j=-\infty}^{\infty} (R_j^2 + R_{j+k} R_{j-k} - 4R_k R_j R_{j-k} + 2R_j^2 R_k^2) \quad \ldots (6.5.5)$$

where N is the number of observations in the time series. Under the assumption that R_k's are zero for $|k| > q$ where q is some positive integer, (6.5.5) reduces to Bartlett's large-lag standard error formula :

$$V(r_k) \simeq \frac{1}{N}(1 + 2 \sum_{j=1}^{q} R_j^2), \quad k > q . \qquad \ldots (6.5.6)$$

Now for completely random θ_t's (6.5.6) becomes

$$V(r_k) \simeq \frac{1}{N}, \quad k > 0.$$

Then letting

$$\sigma_{r_1} = + \sqrt{V(r_1)} = \frac{1}{\sqrt{N}}$$

we first check whether r_1 is significantly different from zero. If so, we conclude that R_1 is non-zero and next examine whether $R_1 \neq 0$ but $R_k = 0$ for $k \geq 2$. To do this we replace R_1 in (6.5.6) by r_1 and calculate

$$\sigma_{r_2} = + \sqrt{\frac{1}{N}(1 + 2r_1^2)}, \quad k > 1$$

and check whether r_2 is significantly different from zero. If so, we continue further. Each r_k is compared with 1.96 times its standard error given by σ_{r_k}.

It is clear that all assumptions of the Bartlett's test[2/] are satisfied if we assume our θ_t's to be normal (this will automatically mean assumption of normality for ε_t^+'s) and hence it can be used to discriminate among our three broad groups of cases. This, however, would not solve our problem completely. To achieve complete identification we have to further distinguish among the different cases falling under the same group. It does not seem possible to distinguish between the two cases in group I which are observationally equivalent. But for group II, we can divide the four cases into two subgroups each consisting of two indistinguishable cases. While in cases (5) and (7) ε_t^+'s follow a first-order autoregressive processes, in cases (3) and (4), ε_t^+ is a random series. In fact, we test the null hypothesis

---

2/ As ε_t's and Z_t's are stationary processes and $e_t^+ \to \varepsilon_t + Z_t$ in distribution as $n \to \infty$, e_t^+'s (and hence θ_t's) would be stationary in large samples.

$$H_0 : \begin{cases} \text{either } \rho_\varepsilon = 0, & \sigma_Z^2 = 0 \\ \text{or } \rho_\varepsilon = 0, & \rho_Z = 0, \quad \sigma_Z^2 > 0 \end{cases}$$

against the alternative

$$H_1 : \begin{cases} \text{either } \rho_\varepsilon \neq 0, & \sigma_Z^2 = 0 \\ \text{or } \rho_\varepsilon = \rho_Z = \rho \neq 0 \end{cases}$$

by using the conventional Durbin-Watson and other tests.

We may now restate the final grouping of the seven exhaustive cases according to their identifiability on the basis of the tests proposed here. The broad descriptions of the situations are also given here.

Group I : (i) $\rho_\varepsilon = 0, \quad \rho_Z \neq 0, \quad \sigma_Z^2 > 0$ and (ii) $\rho_\varepsilon \neq 0, \quad \rho_Z = 0, \quad \sigma_Z^2 > 0$ — AR(1)+Random

Group II (Subgroup 1) : (i) $\rho_\varepsilon = 0, \quad \sigma_Z^2 = 0$ and (ii) $\rho_\varepsilon = 0, \quad \rho_Z = 0, \quad \sigma_Z^2 > 0$ — Random

Group II (Subgroup 2) : (i) $\rho_\varepsilon \neq 0, \quad \sigma_Z^2 = 0$ and (ii) $\rho_\varepsilon = \rho_Z = \rho \neq 0, \quad \sigma_Z^2 > 0$ — AR(1)

Group III : $0 \neq \rho_\varepsilon \neq \rho_Z \neq 0, \quad \sigma_Z^2 > 0$ — AR(1) + AR(1).

It may be noted that the failure to achieve further discrimination does not really affect the conclusion regarding the nature of autocorrelation of the error term. This may be seen by looking at the interpretations of the cases belonging to the four groups/subgroups stated above. Thus, if an observed situation is found to fall in Group I, the conclusion would be that while both the components are present, the

autocorrelation is due to only one of the two components. The situation represented by Group III, on the other hand, implies that the autocorrelation is due to both the components being autocorrelated. Cases coming under subgroup 1 of Group II indicate that there is no autocorrelation in the error term. The first case in the second subgroup of Group II would mean that the error term is autocorrelated and that is due to one component only, the other being absent.[10/] Clearly, the autocorrelated linear regression model commonly considered in the literature really deals with situations described in subgroup 2 of Group II.

6.6 Estimation

We have seen in the last section that it is not possible to know exactly which particular case out of the seven possible cases, a given set of data represents. (The only exception is case (6) under Group III.) The most that could be done by using our tests is to classify a given situation into one of the four broad groups/subgroups. However, although it is not possible to distinguish between the cases in the first three groups/subgroups, one can still consistently and efficiently estimate the regression coefficients for each of the seven cases, if the broad group/subgroup to which a particular case belongs could be identified.

We now describe the method of estimation to be applied for the different situations. For this purpose, it would be convenient to first

---

10/ In the other case of this subgroup, both the components are present but still the error can follow AR(1) process. But here $\rho_\varepsilon = \rho_Z = \rho \neq 0$, and this is unlikely to happen in observed situations.

obtain the structure of the variance-covariance matrix of ε^+ for the most general case where

$$\sigma_Z^2 > 0 \quad \text{and} \quad 0 \neq \rho_\varepsilon \neq \rho_Z \neq 0.$$

Since

$$\varepsilon_t^+ = \varepsilon_t + Z_t,$$

we have

$$\text{Cov}\,(\varepsilon_t^+, \varepsilon_{t-s}^+) = \rho_\varepsilon^s \sigma_\varepsilon^2 + \rho_Z^s \sigma_Z^2 \qquad \text{for } s = 0, 1, \ldots, t-1 \text{ and } t = 1, 2, \ldots, n.$$

Hence

$$V(\varepsilon^+) = \begin{pmatrix} \sigma_\varepsilon^2 + \sigma_Z^2 & \rho_\varepsilon \sigma_\varepsilon^2 + \rho_Z \sigma_Z^2 & \cdots & \rho_\varepsilon^{n-1}\sigma_\varepsilon^2 + \rho_Z^{n-1}\sigma_Z^2 \\ \rho_\varepsilon \sigma_\varepsilon^2 + \rho_Z \sigma_Z^2 & \sigma_\varepsilon^2 + \sigma_Z^2 & \cdots & \rho_\varepsilon^{n-2}\sigma_\varepsilon^2 + \rho_Z^{n-2}\sigma_Z^2 \\ \cdots\cdots\cdots & \cdots\cdots\cdots\cdots & \cdots & \cdots\cdots\cdots\cdots \\ \rho_\varepsilon^{n-1}\sigma_\varepsilon^2 + \rho_Z^{n-1}\sigma_Z^2 & \rho_\varepsilon^{n-2}\sigma_\varepsilon^2 + \rho_Z^{n-2}\sigma_Z^2 & \cdots & \sigma_\varepsilon^2 + \sigma_Z^2 \end{pmatrix}$$

$$= \sigma_\varepsilon^2 \begin{pmatrix} 1 + \sigma_0^2 & \rho_\varepsilon + \rho_Z \sigma_0^2 & \cdots & \rho_\varepsilon^{n-1} + \rho_Z^{n-1}\sigma_0^2 \\ \rho_\varepsilon + \rho_Z \sigma_0^2 & 1 + \sigma_0^2 & \cdots & \rho_\varepsilon^{n-2} + \rho_Z^{n-2}\sigma_0^2 \\ \cdots\cdots\cdots & \cdots\cdots\cdots\cdots & \cdots & \cdots\cdots\cdots\cdots \\ \rho_\varepsilon^{n-1} + \rho_Z^{n-1}\sigma_0^2 & \rho_\varepsilon^{n-2} + \rho_Z^{n-2}\sigma_0^2 & \cdots & 1 + \sigma_0^2 \end{pmatrix} \quad \ldots (6.6.1)$$

where $\sigma_0^2 = \sigma_Z^2 / \sigma_\varepsilon^2$.

Clearly, depending upon the particular values of the parameters involved, this matrix will assume different forms for the different cases.

Let us first consider subgroup 1 of Group II which comprises cases (3) and (4). It may be seen that the variance-covariance matrix for this group is

$$V(\varepsilon^{+}) = \sigma^{+2} \; I_n$$

where

$$\sigma^{+2} = \begin{cases} \sigma_\varepsilon^2 & \text{for case (3)} \\ \sigma_\varepsilon^2 + \sigma_Z^2 & \text{for case (4).} \end{cases}$$

It is obvious that in either case the best linear unbiased (and consistent) estimator of the regression coefficients β for such a model is given by $\hat{\beta}$, the OLS estimator, where

$$\hat{\beta} = (X'X)^{-1} X'Y.$$

The asymptotic variance-covariance matrix of $\hat{\beta}$ is

$$a.v(\hat{\beta}) = n^{-1} \sigma^{+2} \Sigma_{XX}^{-1}$$

and this can be consistently estimated by replacing σ^{+2} by

$$\hat{\sigma}^{+2} = \frac{\sum_{1}^{n} e_t^{+2}}{n-k}$$

and taking $\left(\frac{1}{n} X'X\right)^{-1}$ as estimate of Σ_{XX}^{-1}.

It may be noted that $\hat{\sigma}^{+2}$ will estimate σ_ε^2 in case (3) and $(\sigma_\varepsilon^2 + \sigma_Z^2)$ in case (4) consistently. It is therefore a consistent estimator of the disturbance variance in each case. Thus, even if discrimination between cases (3) and (4) is not possible, application of OLS will yield best estimators irrespective of the actual situation.

Now, for all the other cases, we can apply generalised least squares (GLS) taking $\sigma^2 V$ (where σ^2 and V are different for the different cases and V is assumed to be positive definite) as the form of the variance-covariance matrix of the disturbances. It is well-known that the application of GLS will yield consistent and asymptotically efficient estimate of β and a consistent estimate of the asymptotic variance-covariance matrix of this GLS estimator ($\tilde{\beta}$) which is given by

$$a.v(\tilde{\beta}) = \sigma^2 n^{-1} \lim_{n \to \infty} \left(\frac{1}{n} X' V^{-1} X \right)^{-1}$$

where σ^2 can be consistently estimated by

$$\hat{\sigma}^2 = \frac{1}{n-k} (Y - X\tilde{\beta})' V^{-1} (Y - X\tilde{\beta}).$$

The trouble, however, is that V is unknown and hence a straightforward application of GLS is not possible.

The problem can be alternatively tackled by using a result given in Theil (1971, p.399). The result[11] is that given a consistent estimator $\hat{V}$ of V, the regression coefficients, σ^2 and the asy. var ($\tilde{\beta}$) — where $\tilde{\beta}$ is GLS estimator using $\hat{V}$ — can, under certain conditions, be consistently and efficiently estimated by GLS with V replaced by $\hat{V}$. The problem then reduces to finding a consistent estimate of V for each of the possible cases.

If the test procedure suggested earlier indicates that a given situation falls under Group I, then the variance-covariance matrix of

---

11/ The details of the result and the conditions are given in Appendix 6.1. It will be seen that these conditions are satisfied in the present case.

the disturbances ε^+ is given by

$$V(\varepsilon^+) = \sigma_Z^2 \begin{pmatrix} (1 + \frac{1}{\sigma_0^2}) & \rho_Z & \rho_Z^2 & \cdots & \rho_Z^{n-1} \\ \rho_Z & (1 + \frac{1}{\sigma_0^2}) & \rho_Z & \cdots & \rho_Z^{n-2} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \rho_Z^{n-1} & \rho_Z^{n-2} & \rho_Z^{n-3} & \cdots & (1+ \frac{1}{\sigma_0^2}) \end{pmatrix} \quad \text{for case (1)}$$

$$= \sigma_Z^2 V_1$$

or,

$$V(\varepsilon^+) = \sigma_\varepsilon^2 \begin{pmatrix} (1+ \sigma_0^2) & \rho_\varepsilon & \rho_\varepsilon^2 & \cdots & \rho_\varepsilon^{n-1} \\ \rho_\varepsilon & (1+ \sigma_0^2) & \rho_\varepsilon & \cdots & \rho_\varepsilon^{n-2} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \rho_\varepsilon^{n-1} & \rho_\varepsilon^{n-2} & \rho_\varepsilon^{n-3} & \cdots & (1+ \sigma_0^2) \end{pmatrix} \quad \text{for case (2)}$$

$$= \sigma_\varepsilon^2 V_1$$

where $\sigma_0^2 = \sigma_Z^2 / \sigma_\varepsilon^2$, as before, and V_1 is defined as above according to the specific case.

We now suggest the following estimator $\hat{V}_1$ for V_1 :

$$\hat{V}_1 = \begin{pmatrix} \hat{\rho}_2/\hat{\rho}_1 & \hat{\rho}_2 & \hat{\rho}_2^2 & \cdots & \hat{\rho}_2^{n-1} \\ \hat{\rho}_2 & \hat{\rho}_2/\hat{\rho}_1 & \hat{\rho}_2 & \cdots & \hat{\rho}_2^{n-2} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \hat{\rho}_2^{n-1} & \hat{\rho}_2^{n-2} & \hat{\rho}_2^{n-3} & \cdots & \hat{\rho}_2/\hat{\rho}_1 \end{pmatrix}$$

It has already been seen that

$$\operatorname*{plim}_{n \to \infty} \hat{\rho}_2 = \begin{cases} \rho_Z & \text{for case (1)} \\ \rho_\varepsilon & \text{for case (2)} \end{cases}$$

and

$$\operatorname*{plim}_{n \to \infty} \hat{\rho}_1 = \begin{cases} \rho_Z / (1 + \frac{1}{\sigma_0^2}) & \text{for case (1)} \\ \rho_\varepsilon / (1 + \sigma_0^2) & \text{for case (2)} . \end{cases}$$

Clearly, $\hat{V}_1$ is a consistent estimator of V_1 irrespective of whether we have case (1) or case (2).

It may be noted that if a particular situation corresponds to either of cases (5) and (7) in subgroup 2 of Group II, then the combined error term actually follows AR(1) with autocorrelation coefficient ρ_ε or ρ as the case may be. For this, the well-known consistent estimator of V is given by

$$\hat{V}_2 = \begin{pmatrix} 1 & \hat{\rho}_1 & \hat{\rho}_1^2 & \cdots & \hat{\rho}_1^{n-1} \\ \hat{\rho}_1 & 1 & \hat{\rho}_1 & \cdots & \hat{\rho}_1^{n-2} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \hat{\rho}_1^{n-1} & \hat{\rho}_1^{n-2} & \cdots & \cdots & 1 \end{pmatrix}$$

since

$$\operatorname*{plim}_{n \to \infty} \hat{\rho}_1 = \begin{cases} \rho_\varepsilon & \text{for case (5)} \\ \rho & \text{for case (7)} . \end{cases}$$

Here

$$\sigma^2 = \begin{cases} \sigma_\varepsilon^2 & \text{for case (5)} \\ \sigma_\varepsilon^2 + \sigma_Z^2 & \text{for case (7)} . \end{cases}$$

We may finally consider Group III which consists of only one case viz., case (6). We have already presented the variance-covariance matrix of ε^+ for this case in (6.6.1). Now to get a consistent estimator of this variance-variance matrix, we first define

$$S_i = \frac{1}{n-i} \sum_{t=i+1}^{n} e_t^+ e_{t-i}^+ , \qquad i = 0, 1, 2, 3.$$

Then assuming that for both u_t and v_t fourth order moments are finite [vide Goldberger (1963), pp.149-153] and using the fact proved earlier that e_t^+ converges in distribution to $\varepsilon_t + Z_t$ as $n \to \infty$, it is easy to see that

$$\operatorname*{plim}_{n \to \infty} S_i = \rho_\varepsilon^i \sigma_\varepsilon^2 + \rho_Z^i \sigma_Z^2 , \qquad i = 0, 1, 2, 3.$$

We may then obtain consistent estimators of ρ_ε, ρ_Z, σ_0^2 $(= \sigma_Z^2/\sigma_\varepsilon^2)$ by solving the following equations :

$$S_0 = \sigma_\varepsilon^2 + \sigma_Z^2$$

$$S_1 = \rho_\varepsilon \sigma_\varepsilon^2 + \rho_Z \sigma_Z^2$$

$$S_2 = \rho_\varepsilon^2 \sigma_\varepsilon^2 + \rho_Z^2 \sigma_Z^2$$

$$S_3 = \rho_\varepsilon^3 \sigma_\varepsilon^2 + \rho_Z^3 \sigma_Z^2$$

and get

$$\hat{\rho}_Z = \frac{S_2 - \hat{\rho}_\varepsilon S_1}{S_1 - \hat{\rho}_\varepsilon S_0}$$

$$\hat{\sigma}_0^2 = 1 \div \left\{ \frac{S_0}{(S_1 - \hat{\rho}_\varepsilon S_0)/(\hat{\rho}_Z - \hat{\rho}_\varepsilon)} - 1 \right\}$$

where $\hat{\rho}_\varepsilon$ is a solution of the quadratic equation

$$\rho_\varepsilon^2 (S_1^2 - S_0 S_2) + \rho_\varepsilon (S_3 S_0 - S_1 S_2) + (S_2^2 - S_3 S_1) = 0$$

(see Appendix 7.1 for relevant algebra).

In case the quadratic equation yields two real roots each lying between -1 and +1, we shall choose that solution as the estimate of ρ_ε for which the residual sum of squares is minimum. Even though this general method of obtaining consistent estimates of the parameters ρ_ε, ρ_Z and σ_0^2 could, in principle, be applied in all the cases, one may not use this method for the other cases because once the relevant group is identified, the methods for obtaining consistent estimates of the relevant parameters for these cases are fairly easy and efficient, at least asymptotically.[12/]

6.7 Conclusions

We have pointed out in this Chapter that, while estimating an autocorrelated linear regression equation, the present practice of using a particularly simple structure for the variance-covariance matrix of the error term (derived on the basis of some assumed process without paying any attention to the components of the error term) may be unsatisfactory in many cases. We have been able to develop here a more reasonable and satisfactory way of dealing with the problem.

12/ Since we thus obtain consistent estimates of the parameters and since we have already assumed error to be normal for the application of Bartlett's test, it may seem that efficient estimates could as well be obtained by maximum likelihood method of estimation. Actually this has been done in the context of a more general error process of which the AR(1) is a particular case in the next Chapter.

We have decomposed the error term into two components, each following AR(1), and have suggested a large sample test to identify different situations characterized by different combinations of the parameters. It may be pointed out that if the model described here is considered as a misspecified model then in the case of Group I, subgroup 2 of Group II and Group III we are able to say definitely on the basis of our test, if there has been any misspecification. In case (6) i.e., in Group III, we can further conclude if there has been autocorrelation due to misspecification. In such cases one can therefore take remedial measures to remove misspecification by, say, including more relevant regressors (if possible), or try to estimate the model by taking into consideration the possible complications created by misspecification.

We have also provided consistent and efficient methods of estimation for those cases where existing methods fail to do so. Our treatment also brings out clearly the situations where OLS is the correct method of estimation and situations where the standard reestimation methods remain valid.

It would be worthwhile to point out here the exact role and significance of the apparently restrictive assumption that X and Z are independently distributed. It is clear that this assumption is needed only to consistently estimate the regression coefficients and the asymptotic variance-covariance matrix of these estimators. So far as the test procedure suggested here is concerned, it is enough to assume that $\underset{n\to\infty}{\text{plim}} \left(\frac{1}{n} X'Z\right) = 0$. This is a much weaker assumption and, as we have

already stated in footnote 3 of this Chapter, it gets satisfied for a misspecified model. Furthermore, if one tries to estimate the model by directly assuming an error process for ε_t^+ — an ARMA(2, 1) process in our case — then he would necessarily have to assume the independence of X and ε^+ which, given our decomposition of ε^+, would also imply independence of X and Z. Hence standard methods of estimation available for estimating models with ARMA processes are equally vitiated by the restrictiveness of this assumption.

Appendix 6.1

Theil's Theorem on Consistent and Efficient Estimation by GLS with Estimated Variance-Covariance Matrix

The following theorem due to Theil has been used in section 6.6. Theil examined the conditions of the theorem in the context of other problems. We carry out a similar examination for our problem.

Theorem [see, Theil (1971, p. 399)]

In the usual set-up for generalized least squares (GLS) for a given X, suppose that V is consistently estimated by $\hat{V}$ i.e., $\operatorname{plim}_{n\to\infty} \hat{V} = V$. Suppose also that there exists a square matrix P such that the elements of $P\varepsilon^+$ are independently and identically distributed (this automatically means $V^{-1} = P'P$; and such a P will always exist because, by assumption, V is a positive definite matrix) and $\lim_{n\to\infty} \frac{1}{n} X'V^{-1}X = Q$, $|Q| > 0$. Then, if the following conditions

$$\text{(i)} \quad \operatorname{plim}_{n\to\infty} \frac{1}{n} X'(\hat{V}^{-1} - V^{-1})X = 0$$

$$\text{(ii)} \quad \operatorname{plim}_{n\to\infty} \frac{1}{\sqrt{n}} X'(\hat{V}^{-1} - V^{-1})\varepsilon^+ = 0$$

are all satisfied, the estimator $\tilde{\tilde{\beta}}$ defined as

$$\tilde{\tilde{\beta}} = (X'\hat{V}^{-1}X)^{-1} X'\hat{V}^{-1}Y$$

is asymptotically equivalent to $\tilde{\beta}$ defined as

$$\tilde{\beta} = (X'V^{-1}X)^{-1} X'V^{-1}Y \quad \text{(where V is known)}$$

in the sense that $\sqrt{n}(\tilde{\tilde{\beta}} - \beta)$ converges in probability to zero as $n \to \infty$. Both estimators $\tilde{\beta}$ and $\tilde{\tilde{\beta}}$ are asymptotically normally distributed with mean vector β and covariance matrix $(\sigma^2/n)Q^{-1}$. If, in addition,

$$\text{(iii)} \quad \operatorname*{plim}_{n \to \infty} \frac{1}{n} \varepsilon^{+\prime} \left(\hat{V}^{-1} - V^{-1} \right) \varepsilon^{+} = 0$$

is also true, then $\hat{\sigma}^2$ and $\hat{\hat{\sigma}}^2$ defined as

$$\hat{\sigma}^2 = \frac{1}{n-k} (Y - X \tilde{\beta})' V^{-1} (Y - X \tilde{\beta})$$

and

$$\hat{\hat{\sigma}}^2 = \frac{1}{n-k} (Y - X \tilde{\tilde{\beta}})' \hat{V}^{-1} (Y - X \tilde{\tilde{\beta}})$$

are both consistent estimators of σ^2.

In what follows we check how far the conditions (i), (ii) and (iii) are satisfied for our case. Additional assumptions, if required, are also stated.

To check (i), we consider the typical, say the (h, h')-th element of the matrix $X' (\hat{V}^{-1} - V^{-1}) X$ which may be written as $X_h' (\hat{V}^{-1} - V^{-1}) X_{h'}$ where $x_h' = (x_{h1}, x_{h2}, \ldots, x_{hn})$.

Now as $\operatorname*{plim}_{n \to \infty} \hat{V} = V$, we have obviously,

$$\operatorname*{plim}_{n \to \infty} \frac{1}{n} X_h' \hat{V}^{-1} X_{h'} = \lim_{n \to \infty} \frac{1}{n} X_h' V^{-1} X_{h'}$$

which exists by the assumption that

$$\lim_{n \to \infty} \frac{1}{n} X' V^{-1} X = Q, \; |Q| > 0.$$

Thus condition (i) is checked.

To check condition (ii), let us take the typical (h-th) element of

$$\frac{1}{\sqrt{n}} X_h' (\hat{V}^{-1} - V^{-1}) \varepsilon^{+}$$

which can be written as

$$\frac{1}{\sqrt{n}}\Big[x_{h1}\{(\hat{v}^{11}-v^{11})\varepsilon_1^+ + \cdots + (\hat{v}^{1n}-v^{1n})\varepsilon_n^+\} + \cdots\cdots$$

$$\cdots + x_{hn}\{(\hat{v}^{n1}-v^{n1})\varepsilon_1^+ + \cdots + (\hat{v}^{nn}-v^{nn})\varepsilon_n^+\}\Big]$$

where $V^{-1} = ((v^{ij}))$, and $\hat{V}^{-1} = ((\hat{v}^{ij}))$.

Thus the h-th element can be written as

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} x_{hi} \sum_{j=1}^{n} (\hat{v}^{ij} - v^{ij})\varepsilon_j^+ .$$

Now, variance of ε_j^+ and different autocovariances of ε_j^+'s are finite. As $\operatorname{plim}_{n\to\infty} \hat{V} = V$,

$$\operatorname{plim}_{n\to\infty} (\hat{v}^{ij} - v^{ij}) = 0 \qquad \forall \quad i, j = 1, 2, \cdots, n$$

and

$$\operatorname{plim}_{n\to\infty} (\hat{v}^{ij} - v^{ij})(\hat{v}^{kl} - v^{kl}) = 0 \qquad \forall \quad i,j,k,l = 1,2,\ldots, n.$$

Assuming

$$\lim_{n\to\infty} \frac{1}{n} E\Big(\sum_{i=1}^{n} x_{hi}\Big)^2$$

exists, we easily find that the h-th element considered above converges to zero in probability.

As for condition (iii), because

$$\operatorname{plim}_{n\to\infty} \hat{V} = V$$

we have

$$\operatorname{plim}_{n\to\infty} \frac{1}{n} \varepsilon^{+/} \hat{V}^{-1} \varepsilon^+ = \operatorname{plim}_{n\to\infty} \frac{1}{n} \varepsilon^{+/} V^{-1} \varepsilon^+$$

provided $\operatorname{plim}_{n\to\infty} \frac{1}{n} \varepsilon^{+/} V^{-1} \varepsilon^+$ exists.

Now $V^{-1} = P'P$,

so, $\varepsilon^{+\prime} V^{-1} \varepsilon^{+} = \varepsilon^{*\prime} \varepsilon^{*}$

where $\varepsilon^{*} = P\varepsilon^{+}$.

As $V(\varepsilon^{+}) = \sigma^2 V$, we have obviously,

$$V(\varepsilon^{*}) = V(P\varepsilon^{+}) = \sigma^2 PVP' = \sigma^2 I \quad (\because V^{-1} = P'P).$$

Now

$$E\left(\frac{1}{n}\varepsilon^{+\prime} V^{-1} \varepsilon^{+}\right) = E\left(\frac{1}{n}\sum_{i=1}^{n} \varepsilon_i^{*2}\right)$$

$$= \frac{1}{n} n\sigma^2 \quad (\because \varepsilon_i^{*}\text{'s are independent with same variance } \sigma^2)$$

$$= \sigma^2 \qquad \ldots (A\ 6.1.1)$$

and

$$V\left(\frac{1}{n}\sum_{1}^{n} \varepsilon_i^{*2}\right) = \frac{1}{n}\left(\frac{1}{n}\sum_{i=1}^{n} V(\varepsilon_i^{*2})\right) = \frac{1}{n}\left(\frac{1}{n} n\upsilon\right) = \frac{\upsilon}{n} \qquad \ldots (A\ 6.1.2)$$

where

$$\upsilon = V(\varepsilon_i^{*2}) = E(\varepsilon_i^{*2} - \sigma^2)^2 = E(\varepsilon_i^{*4} - \sigma^4).$$

So υ is finite if $E(\varepsilon_i^{*4})$ is finite.

Thus with the additional assumption that the fourth order moments of ε_i^{*}'s are finite,

$$\lim \; V\left(\frac{1}{n}\sum_{i=1}^{n} \varepsilon_i^{*2}\right) = 0 \quad \text{(from A 6.1.2)} \qquad \ldots (A\ 6.1.3)$$

Hence

$$\operatorname*{plim}_{n\to\infty} \left(\frac{1}{n}\varepsilon^{+\prime}V^{-1}\varepsilon^{+}\right) = \operatorname*{plim}_{n\to\infty} \left(\frac{1}{n}\varepsilon^{*\prime}\varepsilon^{*}\right) = E\left(\frac{1}{n}\varepsilon^{*\prime}\varepsilon^{*}\right) \quad \text{(because of A 6.1.3)}$$

$$= \sigma^2 \quad (\underline{\text{vide}}\ \text{A } 6.1.1) \text{ which obviously exists.}$$

Thus for condition (iii) to be satisfied, we need an additional assumption that fourth order moments of ε_i^*'s and hence of ε_i's are finite.

Chapter 7

Estimation and Testing in an Autocorrelated Linear Regression Model with Decomposed Error Term : The Case of Two ARMA(1, 1) Components

7.1 Introduction

In the last Chapter we considered the problems of statistical inference for an autocorrelated linear regression model where the error term is decomposed into two additive components. Such a decomposition may be justified on the ground that the error term represents the effects of a number of factors all or some of which may be present in a particular situation and the autocorrelation in the composite error term may be due to autocorrelation in one or more of these components. It has been noted that the process which the composite error term would follow depends on the processes generating the different components and even if one assumes the same type of process for the individual components, the composite error term would in general follow a different process.

We have argued in the last Chapter why one should have some way of knowing if the combined error term really consists of more than one component and whether these components are random or autocorrelated. One should also have some methods for estimating the parameters of these component processes. Such tests and methods of estimation have been proposed in the last Chapter assuming that the sample size is large. The main test is based on the observed autocorrelation of different orders of a particular transformation of the OLS residuals. As for estimation,

OLS is applicable only in the simplest case. The estimation methods for other cases consist essentially in obtaining consistent estimates of the variance-covariance matrix of the composite error term and in applying GLS method utilizing these estimates. The tests and the estimation methods are, however, limited by the fact that both the components have been assumed to follow AR(1) processes.

In the present Chapter we consider the same problem but now we consider more general process for the components of the error term. More specifically, we would now assume each of the error components to follow mixed autoregressive moving average (ARMA(1, 1)) of which AR(1) is obviously a special case. We may here refer to a result due to Granger and Morris (1976) — see also Rose (1977) — which proves that the sum of two independent ARMA(1, 1) process is an ARMA(r, s) process where $r \leq 2$ and $s \leq 2$. However, the order of the resulting process being not uniquely specified one cannot directly use methods available for estimating a model with ARMA error process [see, for example, Pierce (1971b)]. This then provides a justification for adopting the approach suggested here in addition to the arguments given in the last Chapter.

It appears that the approach tried in the last Chapter are no longer applicable under the assumption of ARMA(1, 1) process for each of the components. Hence we propose to use the maximum likelihood (ML) method of estimation and the likelihood ratio (LR) test to discriminate among different situations. It may be mentioned that the standard linear regression model with errors following an ARMA(1,1) process is obviously a special case of our model. Discussions on the problems of testing and

estimation of such models can be found in Pierce (1971a, 1971b), Tiao and Ali (1971), Pagan (1973), Godfrey (1978b), Judge et al. (1980, Ch.5), etc.

The format of this Chapter is as follows : In sections 7.2, 7.3 and 7.4 we describe the model, the nature of the error process and the estimation method respectively, while the asymptotic properties of the estimators are described in section 7.5. In section 7.6 we suggest a consistent initial estimator of the parameters. Likelihood ratio tests discriminating among different situations are described in section 7.7. The conclusions are given in section 7.8.

7.2 The Model

We consider, as before, a k-regressor linear regression model written, in matrix form, as

$$Y = X\beta + \varepsilon^{+} \qquad \ldots\ldots (7.2.1)$$

where X is a (n x k) matrix of observations on k-regressors and β and ε^{+} are (k x 1) and (n x 1) vectors of associated regression coefficients and disturbances respectively. One of the regressors takes the value 1 for all the observations incorporating thereby an intercept term in the model.

We make the following assumptions (cf. (6.2.2)-(6.2.4), (6.2.7) and (6.2.8)):

(i) ε_t^{+} has two additive components, ε_t and Z_t :

$$\varepsilon_t^{+} = \varepsilon_t + Z_t , \qquad t = 1, 2, \ldots\ldots, n. \qquad \ldots (7.2.2)$$

(ii) X is stochastic and distributed independently of both ε and Z .... (7.2.3)

where

$$\varepsilon = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)'$$

and $$Z = (Z_1, Z_2, \ldots, Z_n)'.$$

(iii) ε is independent of Z . .... (7.2.4)

(iv) $\underset{n \to \infty}{\text{plim}} \left(\frac{1}{n} X'X \right) = \Sigma_{XX}$, a positive definite matrix. .... (7.2.5)

(v) For an observed X, rank of $X = k (< n)$. .... (7.2.6)

(vi) Both ε_t and Z_t follow ARMA(1, 1) processes given by

$$\varepsilon_t - \rho_\varepsilon \varepsilon_{t-1} = u_t - \rho_u u_{t-1} \quad \ldots\ (7.2.7)$$

and $$Z_t - \rho_Z Z_{t-1} = v_t - \rho_v v_{t-1} \quad \ldots\ (7.2.8)$$

where u_t's and v_t's are independently distributed with zero mean and variance σ_u^2 and σ_v^2 respectively. The stationarity and invertibility conditions are assumed to hold [vide Box and Jenkins (1976), p. 76] i.e.,

$$-1 < \rho_\varepsilon, \ \rho_u, \ \rho_Z, \ \rho_v < 1 .$$

Other assumptions, needed for proving particular results, will be mentioned in appropriate places.

7.3 Nature of the Error Process

It is fairly well-known that, given our assumption about the processes generating ε_t and Z_t, it is not, in general, true that ε_t^+ will follow ARMA(1, 1)[1]. We shall show below that, barring highly improbable coincidences of parameter values, the combined error term will be ARMA(1,1 when it is indecomposable in the sense that ε_t^+ has only one component which follows ARMA(1, 1). Statistical inference of such cases have been considered by authors like Pierce (1971a, 1971b). Obviously the assumption that the error process follows ARMA(1, 1) may not be a safe one in many cases and one should examine whether the error is decomposable or not.

To prove our assertion, we assume that ε_t^+'s follow an ARMA(1,1) process, viz.,

$$\varepsilon_t^+ - \rho\varepsilon_{t-1}^+ = w_t - \rho_w w_{t-1} \qquad \ldots (7.3.1)$$

where w_t's are distributed independently with zero mean and same variance σ_w^2 and $-1 < \rho, \rho_w < 1$. We obviously have

$$V(\varepsilon_t^+) = \sigma_\varepsilon^2 + \sigma_Z^2 \quad \text{for all } t = 1, 2, \ldots, n. \qquad \ldots (7.3.2)$$

The autocovariances of different lags of ε_t^+'s are given by [vide Box and Jenkins (1976), p. 76]

1/ The relevant theorem may be quoted here from Granger and Morris (1976). $\varepsilon_t \sim$ ARMA (p_1, q_1) and $Z_t \sim$ ARMA (p_2, q_2) and they are independent, then $\varepsilon_t^+ = \varepsilon_t + Z_t \sim$ ARMA(m, n) where $m \leq p_1 + p_2$ and $n \leq \max (p_1 + q_2, p_2 + q_1)$.

$$\text{Cov}\,(\varepsilon_t^+, \varepsilon_{t-s}^+) = \rho^{s-1}\ \upsilon_\varepsilon^+\ \sigma_\varepsilon^{+2}\ , \quad s \geq 1 \qquad \ldots\ (7.3.3)$$

where

$$\upsilon_\varepsilon^+ = \frac{(1 - \rho\rho_w)\,(\rho - \rho_w)}{1 + \rho_w^2 - 2\rho\rho_w}$$

and

$$\sigma_\varepsilon^{+2} = V(\varepsilon_t^+) \qquad \text{for all } t = 1, 2, \ldots\ldots, n\,.$$

Now, in order that ε_t^+'s actually follow ARMA(1, 1), the different autocovariances should conform to the pattern given in (7.3.3). If, however, we write the autocovariances of ε_t^+'s in terms of those of ε_t's and Z_t's, we have, using the assumptions of the model, the following relations

$$\text{Cov}\,(\varepsilon_t^+, \varepsilon_{t-s}^+) = \text{Cov}\,(\varepsilon_t, \varepsilon_{t-s}) + \text{Cov}\,(Z_t, Z_{t-s}), \quad s \geq 1$$

i.e., $$\text{Cov}\,(\varepsilon_t^+, \varepsilon_{t-s}^+) = \rho_\varepsilon^{s-1}\ \upsilon_\varepsilon\ \sigma_\varepsilon^2 + \rho_Z^{s-1}\ \upsilon_Z\ \sigma_Z^2\ , \quad s \geq 1 \qquad \ldots\ (7.3.4)$$

where

$$\upsilon_\varepsilon = \frac{(1 - \rho_\varepsilon\ \rho_u)\,(\rho_\varepsilon - \rho_u)}{1 + \rho_u^2 - 2\rho_\varepsilon\ \rho_u}$$

and

$$\upsilon_Z = \frac{(1 - \rho_Z\ \rho_v)\,(\rho_Z - \rho_v)}{1 + \rho_v^2 - 2\rho_Z\ \rho_v}\ . \qquad \ldots\ (7.3.5)$$

From the above expressions it is clear that (7.3.3) and (7.3.4) are not equal, in general. A necessary and sufficient condition for these two to be equal for all $s \geq 1$ is

$$\rho_u = \rho_v\ (= \rho_w) \quad \text{and} \quad \rho_\varepsilon = \rho_Z\ (= \rho)$$

which is very unlikely to be true in any practical situation. We shall

overlook this possibility in the rest of this section. The other sufficient condition where the expressions (7.3.3) and (7.3.4) are equal is

$$\sigma_Z^2 = 0 \quad (\text{or} \quad \sigma_\varepsilon^2 = 0) \text{ 2/}$$

i.e., either of ε_t's and Z_t's have zero values for all the observations. In other words, this refers to the case where the error term ε_t^+ consists of one component only. Similar observations seem to be true for ARMA processes of higher order.

Now, the error process is not ARMA(1, 1) if both the components are AR(1) or MA(1) or if one is AR(1) and the other MA(1) or if one is MA(1) and the other is random. The only possible case where the error term follows ARMA(1, 1) is : one component follows AR(1) while the other is random and in this case the ARMA process must obey some restrictions. Thus, once we admit the possibility that the error term can be the sum of two components, the assumption that the composite error follows ARMA(1, 1) appears to be highly restrictive; if one of the components follows ARMA(1, 1), the other must then be absent. 3/

7.4 Estimation

We suggest maximum likelihood method of estimation of the model described in section 7.2. It may be seen from the assumptions stated

---

2/ Obviously, both σ_ε^2 and σ_Z^2 cannot be zero.

3/ In the context of misspecific model, this is impossible unless we assume there is no misspecification in the model.

in that section that

$$E(\varepsilon^+) = 0$$

and

$$V(\varepsilon^+)(=V, \text{say}) = \begin{pmatrix} \sigma_\varepsilon^2 + \sigma_Z^2 & \upsilon_\varepsilon \sigma_\varepsilon^2 + \upsilon_Z \sigma_Z^2 & \cdots & \rho_\varepsilon^{n-2} \upsilon_\varepsilon \sigma_\varepsilon^2 + \rho_Z^{n-2} \upsilon_Z \sigma_Z^2 \\ \upsilon_\varepsilon \sigma_\varepsilon^2 + \upsilon_Z \sigma_Z^2 & \sigma_\varepsilon^2 + \sigma_Z^2 & \cdots & \rho_\varepsilon^{n-3} \upsilon_\varepsilon \sigma_\varepsilon^2 + \rho_Z^{n-3} \upsilon_Z \sigma_Z^2 \\ \cdots\cdots\cdots & \cdots\cdots\cdots & \cdots & \cdots\cdots\cdots \\ \rho_\varepsilon^{n-2} \upsilon_\varepsilon \sigma_\varepsilon^2 + \rho_Z^{n-2} \upsilon_Z \sigma_Z^2 & \rho_\varepsilon^{n-3} \upsilon_\varepsilon \sigma_\varepsilon^2 + \rho_Z^{n-3} \upsilon_Z \sigma_Z^2 & \cdots & \sigma_\varepsilon^2 + \sigma_Z^2 \end{pmatrix}$$

For the purpose of estimation we make the following additional assumptions:

Assumption 1 : For a given X, Y is distributed normally with mean $X\beta$ and variance-covariance matrix V.

Assumption 2 : The matrix $V = V(\theta)$

[where $\theta = (\theta_1, \theta_2, \ldots, \theta_6)' = (\sigma_\varepsilon^2, \sigma_Z^2, \rho_\varepsilon, \rho_Z, \rho_u, \rho_v)'$,

$\theta \in \Omega$, the set of admissible values of θ]

is positive definite. Also, the elements of V are twice differentiable functions of the parameters in θ.

The log-likelihood equation of y_t's, for a given X, is given by

$$L = \text{constant} - \frac{1}{2} \ln |V| - \frac{1}{2}(Y - X\beta)' V^{-1} (Y - X\beta). \qquad \ldots (7.4.1)$$

Let η be the vector of all the parameters given by $\eta = (\beta', \theta')'$. Then our problem is to estimate the parameter vector η on the basis of a single (n x 1) vector of sample observations Y[1]. We can use any of

---

[1] It may be noted, and will also be seen in section 7.6, that there is an identification problem which makes the information matrix singular [vide Rothenberg (1971)]. However, as has been shown in section 7.6, one can reparametrize the model and then follow the procedure suggested in this and the next sections. This is what we have, in fact, done.

the following two methods for obtaining ML estimates of the parameters from (7.4.1). The first method for obtaining the estimates can be described as follows.

For a given $\theta_o \in \Lambda$, $\hat{\beta}(\theta_o)$, the ML estimator of β for the given θ_o, is obtained as

$$\hat{\beta}(\theta_o) = (X' V_o^{-1} X)^{-1} X' V_o^{-1} Y \qquad \cdots \quad (7.4.2)$$

where $V_o = V(\theta_o)$.

Then the maximum value of the log-likelihood function for a given X, excepting the constant term, at θ_o, is given by[5]

$$\hat{L}(\theta_o) = -\frac{1}{2} \ln |V_o| - \frac{1}{2} (Y - X\hat{\beta}(\theta_o))' V_o^{-1} (Y - X\hat{\beta}(\theta_o)).$$

The ML estimate of θ, say $\hat{\theta}$, is obtained by searching over the range of admissible values of the parameters[6] in θ and then choosing that value of θ as the estimate for which the value of $\hat{L}(\theta_o)$ is the largest. Under very general assumptions,[7] $\hat{\beta}_{ML} = \hat{\beta}(\hat{\theta})$ and $\hat{\theta}$ would be the ML estimates of β and θ respectively.

5/ For convenience, henceforth we shall not always write 'for a given X' in relevant contexts.

6/ Of the six parameters in θ, each of ρ_ε, ρ_Z, ρ_u and ρ_v lies between -1 and +1. Therefore searching over admissible values of these parameters is not much of a problem. However, since σ_ε^2 and σ_Z^2 lie between 0 and ∞, we search over σ_ε^{*2} and σ_Z^{*2} (which lie between 0 and 1) instead of over σ_ε^2 and σ_Z^2, respectively, taking

$$\sigma_\varepsilon^2 = \sigma_\varepsilon^{*2}/(1 - \sigma_\varepsilon^{*2})$$

and $\sigma_Z^2 = \sigma_Z^{*2}/(1 - \sigma_Z^{*2})$

7/ See, for example, Dhrymes (1970, Chapter 3).

The second procedure of obtaining the ML estimate is due to Magnus (1978). It consists of the following steps :

(a) For a given value θ_o of $\theta \in \Omega$, calculate

$$e_o^+ = Y - X\hat{\beta}(\theta_o)$$

where $\hat{\beta}(\theta_o)$ is obtained from (7.4.2).

(b) Solve for the six parameters in θ from six nonlinear equations obtained from

$$\operatorname{tr}\left(\frac{\partial V^{-1}}{\partial \theta_h} V\right)_{\theta = \hat{\theta}} = e_o^{+\prime}\left(\frac{\partial V^{-1}}{\partial \theta_h}\right)_{\theta = \hat{\theta}} e_o^+ \quad ; \; h = 1,2,\ldots, 6.$$

(c) From the solution of θ obtained from (b), the steps (a) and (b) are repeated with new V's until convergence.

Magnus (1978) pointed out that under very general conditions, this procedure converges to the unique ML estimate in view of the results proved earlier by Oberhofer and Kmenta (1974).

The first method is very laborious as it involves searching over different combinations of values of the six parameters. The approach of Magnus, on the other hand, is iterative in nature. Convergence, though guaranteed under fairly general conditions, may take time if the initial value is not well chosen. A second difficulty with the method suggested by Magnus is that the equations in θ are highly nonlinear.

In order to avoid these problems, a simple approach for obtaining ML estimates would be to use the scoring method of estimation [vide Rao (1974), pp.366-367]. According to this method the estimates are obtained from repeated adjustments of the form

$$\hat{\eta}_1 = \hat{\eta}_0 - \left(\frac{\partial L}{\partial \eta}\right)_{\eta=\hat{\eta}_0} \left(E\left(\frac{\partial^2 L}{\partial \eta \partial \eta'}\right)\right)^{-1}_{\eta=\hat{\eta}_0}$$

where $\hat{\eta}_0$ and $\hat{\eta}_1$ are initial and adjusted estimates of η respectively. This method has the advantage that if $\hat{\eta}_0$ is a consistent estimate of η, then $\hat{\eta}_1$ will have the same asymptotic properties as ML estimate [vide Rothenberg and Leenders (1964); see also Amemiya (1973b, p. 933)].

Though this method would require a consistent initial estimate and involve the inversion of the information matrix, we suggest its use in view of the stated property and computational simplicity.[8] In section 7.6, we discuss the procedure for obtaining a consistent initial estimate for η. Before doing so, it would be convenient to keep in mind the asymptotic properties of the ML estimator which are stated in the next section.

7.5 Asymptotic Properties of the ML Estimator

In this section we study the asymptotic properties of the ML estimator in the context of our problem. In order to obtain the asymptotic properties of ML estimator $\hat{\eta}$ of η obtained from a single vector of observations on y whose variance-covariance matrix V increases in size with n, we have to make some additional assumptions. Though the assumptions and the theorem are given in Magnus (1978) we state them here for the sake of convenience and completeness.

---

[8] Though we would not be gaining anything asymptotically if we continue the iteration beyond the second stage, in small samples we may continue doing so until convergence is attained to get efficient estimates.

Assumption 3 : Every element of $\frac{1}{n} X' V^{-1} X$ converges as $n \to \infty$ to a finite function of θ, uniformly for θ in any compact set.

Assumption 4 : Every diagonal element of $\frac{1}{n^2} X'\left(\frac{\partial V^{-1}}{\partial \theta_h}\right) V \left(\frac{\partial V^{-1}}{\partial \theta_h}\right) X$ converges as $n \to \infty$ to zero, uniformly for θ in any compact set $(h = 1, 2, \ldots, 6)$.

Assumption 5 : $\frac{1}{n} \operatorname{tr}\left(\left(\frac{\partial V^{-1}}{\partial \theta_h}\right) V \left(\frac{\partial V^{-1}}{\partial \theta_{h'}}\right) V\right)$ converges as $n \to \infty$ to a finite function of θ uniformly for θ in any compact set $(h, h' = 1, 2, \ldots, 6)$.

Assumption 6 : $\frac{1}{n^2} \operatorname{tr}\left(\left(\frac{\partial^2 V^{-1}}{\partial \theta_h \partial \theta_{h'}}\right) V\right)^2$ converges as $n \to \infty$ to zero uniformly for θ in any compact set $(h, h' = 1, 2, \ldots, 6)$.

Assumption 7 : Let γ_i denote the $(n \times 1)$ i-th column vector of V^{-1} and γ be the $(n^2 \times 1)$ column vector $\begin{pmatrix} \gamma_1 \\ \vdots \\ \gamma_n \end{pmatrix}$. The six vectors $\frac{\partial \gamma}{\partial \theta_h}$ $(h = 1, 2, \ldots, 6)$ are linearly independent.

Now differentiating (7.4.1) with respect to the parameters, we have

$$\left.\begin{aligned}
&\frac{\partial L}{\partial \beta} = (Y - X\beta)\, V^{-1} X \\
&\frac{\partial L}{\partial \theta_h} = \operatorname{tr}(V - \varepsilon^+ \varepsilon^{+\prime})\, \frac{\partial V^{-1}}{\partial \theta_h}; \qquad h = 1, 2, \ldots, 6 \\
&\frac{\partial^2 L}{\partial \beta \partial \beta'} = - X' V^{-1} X \\
&\frac{\partial^2 L}{\partial \beta \partial \theta_h} = (Y - X\beta)\, \frac{\partial V^{-1}}{\partial \theta_h} X \quad ; \qquad h = 1, 2, \ldots, 6 \\
&\frac{\partial^2 L}{\partial \theta_h \partial \theta_{h'}} = \frac{1}{2} \sum_{h,h'} \{ v_{hh'} - \varepsilon_h^+ \varepsilon_{h'}^+ \}\, \frac{\partial^2 v_{hh'}^{-1}}{\partial \theta_h \partial \theta_{h'}} \\
&\qquad - \frac{1}{2} \operatorname{tr}\left(\frac{\partial V^{-1}}{\partial \theta_h} V \frac{\partial V^{-1}}{\partial \theta_{h'}} V\right); \qquad h, h' = 1, 2, \ldots, 6
\end{aligned}\right\} \quad \ldots(7.5.1)$$

where $V = ((v_{hh'}))$ and $V^{-1} = ((v^{-1}_{hh'}))$.

Thus for a given X, using (7.2.1) and the fact that $E(\varepsilon_h^+ \varepsilon_{h'}^+) = v_{hh'}$, we have from (7.5.1),

$$E\left(\frac{\partial^2 L}{\partial\beta\partial\beta'}\right) = -X'V^{-1}X$$

$$E\left(\frac{\partial^2 L}{\partial\beta\partial\theta_h}\right) = 0 \quad ; \qquad h = 1, 2, \ldots, 6$$

and

$$E\left(\frac{\partial^2 L}{\partial\theta_h\partial\theta_{h'}}\right) = -\frac{1}{2}\,\mathrm{tr}\left(\frac{\partial V^{-1}}{\partial\theta_h} V \frac{\partial V^{-1}}{\partial\theta_{h'}} V\right) ; \qquad h, h' = 1, 2, \ldots, 6.$$

Therefore, for a given X, the information matrix of the log-likelihood function defined as

$$I(\eta) = -E\left(\frac{\partial^2 L}{\partial\eta\partial\eta'}\right)$$

is given by

$$\ldots\ I(\eta) = \begin{pmatrix} X'V^{-1}X & 0 \\ 0 & \frac{1}{2}D \end{pmatrix}$$

where $D = ((d_{hh'}))$, $\qquad d_{hh'} = \mathrm{tr}\left(\frac{\partial V^{-1}}{\partial\theta_h} V \frac{\partial V^{-1}}{\partial\theta_{h'}} V\right)$

and D in positive definite under (7.2.6) and Assumptions 1 and 2 and 7.

Theorem : Under (7.2.6) and Assumptions 1-7, ML estimator $\hat{\eta}$ of η is weakly consistent, asymptotically normally distributed and asymptotically efficient in the maximum probability sense of Weiss and Wolfowitz (1967). Further $\sqrt{n}(\hat{\eta} - \eta)$ has the asymptotic distribution $N[0, \lim_{n\to\infty} n(I(\eta))^{-1}]$ 9/.

9/ For proof, see Magnus (1978), Theorem 8, pp. 295-296.

7.6 A Consistent Estimator of η

It has been pointed out earlier that in order to successfully apply the scoring method for obtaining ML estimates, we need a consistent initial estimate of η. It is well-known [see, for example, Johnston (1972), Theil (1971)] that β can be consistently estimated by applying OLS to the model. The problem, however, arises with obtaining a consistent estimate for θ.

Let $e^+ = Y - X\hat{\beta}$ be the OLS residual where $\hat{\beta}$ is the OLS estimator of β.

Now,

$$\begin{aligned} e^+ &= Y - X\hat{\beta} \\ &= X(\beta - \hat{\beta}) + \varepsilon^+ \qquad \text{(from 7.2.1)} \\ &= X(\beta - \hat{\beta}) + \varepsilon + Z \qquad \text{(from 7.2.2).} \end{aligned}$$

Since $\underset{n \to \infty}{\text{plim}}\ \hat{\beta} = \beta$, as already stated, e^+ converges in distribution to $(\varepsilon + Z)$ as $n \to \infty$.

Let us now define S_i to be the i-th autocovariance of e_t^+'s i.e.,

$$S_i = \frac{1}{n-i} \sum_{t=i+1}^{n} e_t^+ e_{t-i}^+ \quad ; \quad i = 0, 1, 2, 3, 4. \qquad \ldots (7.6.1)$$

As $e_t^+ \to \varepsilon_t + Z_t$ (as $n \to \infty$) is distribution, we have from the above relations using (7.2.4),

$$\underset{n \to \infty}{\text{plim}}\ S_i = \underset{n \to \infty}{\text{plim}}\ \frac{1}{n-i} \sum_{t=i+1}^{n} \varepsilon_t \varepsilon_{t-i} + \underset{n \to \infty}{\text{plim}}\ \frac{1}{n-i} \sum_{t=i+1}^{n} Z_t Z_{t-i} .$$

It is known that under very general conditions i.e., when the dependence between distant values of ε_t's and also of Z_t's are not too strong so that

$$\lim_{s\to\infty} \xi_{\varepsilon s} = 0, \qquad \lim_{s\to\infty} \xi_{Zs} = 0, \qquad \lim_{s\to\infty} (\delta_{\varepsilon s} - \sigma_\varepsilon^4) = 0,$$

$$\lim_{s\to\infty} (\delta_{Zs} - \sigma_Z^4) = 0, \qquad \lim_{s\to\infty} (\omega_{\varepsilon s} - \xi_{\varepsilon 1}^2) = 0,$$

$$\lim_{s\to\infty} (\omega_{Zs} - \xi_{Z1}^2) = 0,$$

and when the fourth order moments of ε_t's and Z_t's exist [vide Goldberger (1963), pp.142-149],

[where $\xi_{\varepsilon s} = E(\varepsilon_t\ \varepsilon_{t+s})$, $\xi_{Zs} = E(Z_t\ Z_{t+s})$,

$\delta_{\varepsilon s} = E(\varepsilon_t\ \varepsilon_{t+s})^2$, $\delta_{Zs} = E(Z_t\ Z_{t+s})^2$,

$\omega_{\varepsilon s} = E(\varepsilon_t\ \varepsilon_{t+1}\ \varepsilon_{t+s}\ \varepsilon_{t+s+1})$ and $\omega_{Z_s} = E(Z_t Z_{t+1} Z_{t+s} Z_{t+s+1})$]

$$\operatorname*{plim}_{n\to\infty} \frac{1}{n-i} \sum_{t=i+1}^{n} \varepsilon_t\ \varepsilon_{t-i}$$

and $$\operatorname*{plim}_{n\to\infty} \frac{1}{n-i} \sum_{t=i+1}^{n} Z_t\ Z_{t-i},$$

are consistent estimates of the corresponding population moments[10].

Hence from (7.6.1), we have

$$\operatorname*{plim}_{n\to\infty} S_o = \sigma_\varepsilon^2 + \sigma_Z^2$$

and

$$\operatorname*{plim}_{n\to\infty} S_i = \rho_\varepsilon^{i-1}\, \upsilon_\varepsilon\, \sigma_\varepsilon^2 + \rho_Z^{i-1}\, \upsilon_Z\, \sigma_Z^2, \quad i = 1,2,3,4. \qquad \ldots (7.6.2)$$

---

[10] For the result to be true, ε_t's and Z_t's have to be stationary processes which they are assumed to be in our model.

It can be seen from the algebraic solutions given in Appendix 7.1 that the above equations would yield consistent estimates of $\tilde{\theta}$

[where $\tilde{\theta} = (\sigma^2, \rho_\varepsilon, \rho_Z, \tilde{\upsilon}_\varepsilon, \tilde{\upsilon}_Z)'$]

in terms of S_0, S_1, S_2, S_3 and S_4 where

$$\sigma^2 = (\sigma_\varepsilon^2 + \sigma_Z^2), \quad \tilde{\upsilon}_\varepsilon = \upsilon_\varepsilon \sigma_\varepsilon^2 \quad \text{and} \quad \tilde{\upsilon}_Z = \upsilon_Z \sigma_Z^2$$

and not of the six individual parameters in θ. It is obvious that addition of one more equation, say,

$$\operatorname*{plim}_{n \to \infty} S_5 = \rho_\varepsilon^4 \upsilon_\varepsilon \sigma_\varepsilon^2 + \rho_Z^4 \upsilon_Z \sigma_Z^2$$

would not be of any further help with regard to obtaining estimates of all the six parameters in θ. For the purpose of estimation, we can, however, reparametrize the model in terms of the parameters for which consistent initial estimates are available.[11/] It can be seen that V can be specified completely in terms of the five parameters in $\tilde{\theta}$:

$$V = \begin{pmatrix} \sigma^2 & \tilde{\upsilon}_\varepsilon + \tilde{\upsilon}_Z & \cdots & \rho_\varepsilon^{n-2} \tilde{\upsilon}_\varepsilon + \rho_Z^{n-2} \tilde{\upsilon}_Z \\ \tilde{\upsilon}_\varepsilon + \tilde{\upsilon}_Z & \sigma^2 & \cdots & \rho_\varepsilon^{n-3} \tilde{\upsilon}_\varepsilon + \rho_Z^{n-3} \tilde{\upsilon}_Z \\ \cdots\cdots & \cdots\cdots\cdots & \cdots & \cdots\cdots\cdots\cdots \\ \rho_\varepsilon^{n-2} \tilde{\upsilon}_\varepsilon + \rho_Z^{n-2} \tilde{\upsilon}_Z & \rho_\varepsilon^{n-3} \tilde{\upsilon}_\varepsilon + \rho_Z^{n-3} \tilde{\upsilon}_Z & \cdots & \sigma^2 \end{pmatrix}$$

Thus we have an identification problem in the model if we are interested in all the six individual parameters in θ. It, however, does not pose any

---

11/ In fact, Assumption 7 rules out such cases, as, for example, one where

$$V = (\sigma_\varepsilon^2 + \sigma_Z^2) I_n.$$

In other words, if Assumption 7 is to be true for all V's for the different cases under ARMA - ARMA set up, such reparameterization is necessary.

problem so far as estimation of β is concerned. It will be seen in the next section that this identification problem poses some difficulties with regard to testing certain hypotheses of interest. So far as the estimation of this model is concerned, the parameters that we would be interested in estimating are contained in the vector

$$\eta = (\beta', \tilde{\theta}')' .$$

It is obvious that the assumptions and results similar to that of sections 7.4 and 7.5 stated in terms of θ and η remain valid also for $\tilde{\theta}$ and η.

As we have already mentioned we use the scoring method to obtain ML estimates of η. It may be noted that for successful application of the scoring method, we require derivatives of the type $\frac{\partial V^{-1}}{\partial \tilde{\theta}_h}$ $(h = 1,2,\ldots,5)$. But since V has a complicated structure, it is difficult to get algebraic form of V^{-1} and hence expressions for $\frac{\partial V^{-1}}{\partial \tilde{\theta}_h}$ $(h = 1, 2, \ldots, 5)$ cannot be directly obtained. We may, however, use the relationship

$$\frac{\partial V^{-1}}{\partial \tilde{\theta}_h} = - V^{-1} \frac{\partial V}{\partial \tilde{\theta}_h} V^{-1} \qquad [\text{See, T. Sawa (1978), p. 171}]$$

to overcome this problem. Clearly expressions for $\frac{\partial V}{\partial \tilde{\theta}_h}$ $(h = 1,2,\ldots,5)$ are easily available.

7.7 Likelihood Ratio Tests for Discriminating the Different Situations

In the previous section we described the procedure for obtaining the consistent initial estimates for β and $\tilde{\theta}$ so that by using the scoring method we can obtain the ML estimate of η. We also noted that there is an identification problem which makes it impossible to obtain

the individual estimates of the six parameters in θ. If we are interested only in the ML estimate of β, non-availability of consistent estimate of θ would pose no problem since V could in any case be consistently obtained by using a consistent estimate of $\mathfrak{F}$. However, in addition to being interested in ML estimate of β, we are also interested in ML estimate of θ because we would then be able to test different hypotheses in order to know how exactly a given autocorrelation situation has arisen. Thus, for example, we may be interested in knowing if an observed autocorrelation is due to both the components following MA(1) error process. Such hypotheses may be of interest to us for various reasons.[12/]

Now in doing this, we first need to check if the different hypotheses specified in terms of the original parameters in θ can as well be stated in terms of the parameters in $\mathfrak{F}$. One can easily see that relevant hypotheses formulated in terms of the parameters in θ cannot all be equivalently stated in terms of those in $\mathfrak{F}$. For instance, to test whether the error process is really MA(1) - MA(1), the relevant hypotheses is

$$H_o : \rho_\varepsilon = \rho_Z = 0$$

in terms of the parameters in θ, and this hypothesis remains the same

---

12/ One may be interested in knowing the exact nature of the autocorrelation of the error process because this helps in specifying the model more correctly. Such knowledge can be helpful in other ways also. If, for example, the exact nature of autocorrelation is known, some standard tests or methods of estimation can be readily used for further studies. The algebraic solutions of the consistent initial estimates will become simpler. Forecasts etc., can be made more efficiently if the variance-covariance matrix of the error is known more precisely. From the point of view of misspecification, acceptance of hypotheses of the type which specify that autocorrelation is due to one component only may lead to the conclusion that the model might not have been misspecified. If the tests reveal the presence of both the components, one might try to remove the error of misspecification from the model by possibly including some more relevant regressors.

in terms of those in $\mathfrak{F}$. But the corresponding hypothesis for the AR(1) - MA(1) case

$$H_o : \rho_u = 0 \quad (\text{i.e., } \upsilon_\varepsilon = \rho_\varepsilon) \text{ and } \rho_Z = 0$$

in terms of original parameters cannot obviously be equivalently stated in terms of the redefined parameters in $\mathfrak{F}$. We, however, argue below that this difficulty can be tackled to some extent.

Since ML estimate of η is available, we suggest likelihood ratio (LR) tests for examining different hypotheses[13]. The likelihood ratio test statistic defined as

$$LR(\cdot) = -2\left[L(H_N) - L(H_A)\right]$$

follows a limiting chi-square distribution with p degrees of freedom where $L(H_N)$ and $L(H_A)$ are the maximum values of the log-likelihood function under the null (H_N) and alternative (H_A) hypotheses and p is the number of additional independent restrictions on the parameters imposed by the null hypothesis.

We can now set up different hypotheses regarding the parameters ρ_ε, ρ_Z, $\tilde{\upsilon}_\varepsilon$ and $\tilde{\upsilon}_Z$ which would correspond to different patterns of autocorrelation of the disturbances and test them against suitable alternatives. Consider first testing the hypothesis of zero autocorrelation

---

13/ One can, in principle, use Wald's test and lagrangian-multiplier (LM) test for testing different relevant hypotheses. However, all the three tests viz., LR-test, Wald's test and LM-test are asymptotically equivalent (see Moran (1970) for details). Also, joint confidence region for all the parameters or for any subset of parameters can be obtained from the estimated variance-covariance matrix of the ML estimator and then used to decide upon acceptance or rejection of different hypotheses.

of the disturbances i.e.,

$$H_o^1 : \rho_\varepsilon = \rho_Z = \tilde{\upsilon}_\varepsilon = \tilde{u}_Z = 0$$

against H_1 which puts no restriction on $\tilde{\theta}$.[14] The relevant test statistic is

$$LR(1) = -2\left[L(H_o^1) - L(H_1)\right] \sim \chi_4^2 \text{ under } H_o^1.$$

If H_o^1 is accepted we conclude that there is no autocorrelation in the disturbances. However, it will not be possible to conclude whether one or both of the components are present (though non-autocorrelated) in the model. Such information is, however, of not much interest from the point of view of knowing the nature of autocorrelation. In both the cases, OLS is obviously the best estimator for the parameters.

Depending on prior information one may be interested in testing whether any of the following two situations represents the actual nature of autocorrelation :

(a) one of the error components follows ARMA(1, 1) while the other follows MA(1); and

(b) one of the error components follows ARMA(1, 1) while the other follows AR(1).[15]

In terms of the parameters in $\tilde{\theta}$, the hypothesis corresponding to situation (a) is

$$H_o^2 : \rho_Z = 0 .$$

14/ In terms of the original parameters in θ, H_o^1 will be :

$$\rho_\varepsilon = \rho_Z = \rho_u = \rho_v = 0.$$

15/ Without any loss of generality we assume that ε_t's follow ARMA(1, 1) while Z_t's follow MA(1) (or AR(1)).

The LR test statistic for testing H_o^2 against H_1 is given by

$$LR(2) = -2\left[L(H_o^2) - L(H_1)\right] \sim \chi_1^2 \quad \text{under } H_o^2.$$

It is to be noted that the hypothesis corresponding to situation (b) cannot be formulated in terms of the parameters in $\tilde{\theta}$ because (b) refers to the case where

$$\rho_v = 0 \quad \text{i.e.,} \quad \upsilon_Z = \rho_Z \quad \text{(see equation(7.3.5))}.$$

But after reparametrization what is available to us is $\tilde{\upsilon}_Z$ $(=\upsilon_Z \sigma_Z^2)$ but not υ_Z. Similarly cases like AR(1) - AR(1) or MA(1) - AR(1) cannot be formulated in terms of $\tilde{\theta}$ and hence cannot be tested against H_1. We shall come back to this point a little later.

One may also be interested in testing the case where both the components follow MA(1) error process.[16/] Denoting this null hypothesis by

$$H_o^3 : \rho_\varepsilon = 0, \quad \rho_Z = 0,$$

we have the relevant test statistic as

$$LR(3) = -2\left[L(H_o^3) - L(H_1)\right] \sim \chi_2^2 \quad \text{under } H_o^3.$$

16/ It should be noted that so far as the estimation of the parameters is concerned, we cannot uniquely estimate $\tilde{\upsilon}_\varepsilon$ and $\tilde{\upsilon}_Z$ separately for the MA(1) - MA(1) case though their sum $\tilde{\upsilon}$ ($= \tilde{\upsilon}_\varepsilon + \tilde{\upsilon}_Z$) can be uniquely estimated. Hence the ML estimates of the parameters of the original model are not uniquely determined in this case. It is, however, clear from the principle of LR test that this does not create any problem is using the LR test for identification of this case represented by the hypothesis H_o^3. It may be further noted in this context that H_o^3 and H_o^5 are not identical though the structure of the variance-covariance matrix is observationally the same for the two cases. This is because H_o^5 imposes an additional restriction i.e., $\tilde{\upsilon}_Z = 0$ which enables one to determine the parameters uniquely for the case given by H_o^5.

Finally, it might be of interest to test if the autocorrelation in the disturbances is due to one component only (say, ε_t's, without any loss of generality) and that ε_t's follow ARMA(1)or MA(1) process (note that ε_t's following AR(1) cannot be tested because of the identification problem)17/. The hypotheses corresponding to the above situations are

$$H_o^4 : \rho_Z = 0, \quad \tilde{\upsilon}_Z = 0 \quad \text{(when } \varepsilon_t\text{'s follow ARMA(1))}$$

and

$$H_o^5 : \rho_Z = 0, \quad \tilde{\upsilon}_Z = 0, \quad \rho_\varepsilon = 0 \quad \text{(when } \varepsilon_t\text{'s follow MA(1))}.$$

The corresponding LR-test statistics are

$$LR(4) = -2\left[L(H_o^4) - L(H_1)\right] \sim \chi_2^2 \text{ under } H_o^4$$

$$LR(5) = -2\left[L(H_o^5) - L(H_1)\right] \sim \chi_3^2 \text{ under } H_o^5 .$$

Now, acceptence of H_o^4 or H_o^5 will enable us to infer that the observed autocorrelation in the model is due to one component only while the other component is either present (but is non-autocorrelated) or absent. Strictly speaking, this latter case refers to standard linear model with ARMA(1,1) autocorrelation process in the disturbances. Now, from the point of view of the nature of autocorrelation the presence (though non-autocorrelated) or absence of the other component makes no difference. Thus we can conclude that the observed autocorrelation is due to autocorrelation in one component only if H_o^4 or H_o^5 is accepted.

17/ For both H_o^4 and H_o^5, $\tilde{\upsilon}_Z = 0$.

From (7.3.5), $\tilde{\upsilon}_Z = -\rho_v \sigma_Z^2 / (1 + \rho_v^2)$ ($\because \rho_Z = 0$ for these cases).
Thus $\tilde{\upsilon}_Z = 0 \implies$ either $\rho_v = 0$ or $\sigma_Z^2 = 0$.
If $\sigma_Z^2 = 0$, then $\rho_v = 0$ automatically.

Instead of testing a specific hypothesis that corresponds to a particular type of autocorrelation in the disturbances, we may be interested in a series of nested hypotheses so as to arrive at the correct form of the autocorrelation process generating the disturbances.[18] This can be regarded as a possible strategy to determine the exact nature of autocorrelation. To do this, we can test one or both of the following sequences of null hypotheses

$$H_o^2 : \rho_Z = 0 \qquad H_o^2 : \rho_Z = 0$$

$$H_o^3 : \rho_Z = \rho_\varepsilon = 0 \qquad H_o^4 : \rho_Z = \tilde{\upsilon}_Z = 0$$

$$H_o^5 : \rho_Z = \rho_\varepsilon = \tilde{\upsilon}_Z = 0 \qquad H_o^5 : \rho_Z = \tilde{\upsilon}_Z = \rho_\varepsilon = 0$$

$$H_o^1 : \rho_Z = \rho_\varepsilon = \tilde{\upsilon}_Z = \tilde{\upsilon}_\varepsilon = 0 \qquad H_o^1 : \rho_Z = \tilde{\upsilon}_Z = \rho_\varepsilon = \tilde{\upsilon}_\varepsilon = 0$$

against the sequences of alternatives

$$H_1^2 : \rho_Z \neq 0 \qquad H_1^2 : \rho_Z \neq 0$$

$$H_1^3 : \rho_Z = 0, \ \rho_\varepsilon \neq 0 \qquad H_1^4 : \rho_Z = 0, \ \tilde{\upsilon}_Z \neq 0$$

$$H_1^5 : \rho_Z = \rho_\varepsilon = 0, \ \tilde{\upsilon}_Z \neq 0 \qquad H_1^5 : \rho_Z = \tilde{\upsilon}_Z = 0, \ \rho_\varepsilon \neq 0$$

$$H_1^1 : \rho_Z = \rho_\varepsilon = \tilde{\upsilon}_Z = 0, \ \tilde{\upsilon}_\varepsilon \neq 0 \qquad H_1^1 : \rho_Z = \tilde{\upsilon}_Z = \rho_\varepsilon = 0, \ \tilde{\upsilon}_\varepsilon \neq 0.$$

All the appropriate LR-test statistics will follow χ^2 with 1 degree of freedom under the corresponding null hypotheses. If any of the null hypothesis is true, the preceding ones must be true; if any one is false, the succeeding ones must be false. Thus we continue testing the sequence of null hypotheses until one is rejected or the final null hypothesis of

18/ See, in this context, a discussion by Mizon (1977) on the sequence of testing such nested hypotheses.

no autocorrelation is accepted and accordingly infer about the nature of autocorrelation of the disturbances.

Now we come back to case (b) where we wanted to test if the autocorrelation in the disturbances was generated by an error process where one of the components follows ARMA(1, 1) while the other follows AR(1). There we found that specification of this hypothesis in terms of the redefined parameters was not possible and therefore such hypotheses like the one represented by (b) or, say, both the error processes following AR(1) etc., cannot be tested against suitable alternatives directly. Conceding that due to an identification problem it is not possible for us to estimate individually all the six parameters in θ and hence certain hypotheses identifying certain autocorrelation situations which could otherwise be formulated in terms of the parameters in θ, could not be tested against suitable alternatives, we suggest in order to overcome this problem that we treat ARMA(1, 1) - AR(1) situation as the unrestricted hypothesis and test different hypotheses describing different autocorrelation situations against this alternative.

Now, under the assumption of disturbances following ARMA(1,1)- AR(1) error process, the parameters involved in the model are given by the vector

$$\tilde{\eta} = (\beta', \tilde{\theta}')'$$

where $\tilde{\theta} = (\sigma_\varepsilon^2, \sigma_Z^2, \rho_u, \rho_\varepsilon, \rho_Z)'$. 19/

However, as we can specify V in terms of υ_ε instead of ρ_u and since a consistent initial estimator of ρ_u has to be obtained via υ_ε for our

19/ For this set-up we now onwards assume that ε_t's follow ARMA(1, 1) and Z_t's follow AR(1).

method (see Appendix 7.1), we obtain ML estimate of

$$\underset{\sim}{\underset{\sim}{\eta}} = (\beta', \underset{\sim}{\underset{\sim}{\theta}}')'$$

where $\underset{\sim}{\underset{\sim}{\theta}} = (\sigma_\varepsilon^2, \sigma_Z^2, \upsilon_\varepsilon, \rho_\varepsilon, \rho_Z)'$.

It will be seen later on that different hypotheses characterizing different autocorrelation situations in this framework can all be specified in terms of the parameters in $\underset{\sim}{\underset{\sim}{\theta}}$. Obviously assumptions and results similar to those in sections 7.2, 7.3, 7.4, 7.5 and 7.6 are now valid in terms of $\underset{\sim}{\underset{\sim}{\theta}}$ and $\underset{\sim}{\underset{\sim}{\eta}}$ instead of θ and η respectively. For this situation $\rho_v = 0$ i.e. $\upsilon_Z = \rho_Z$. The estimation method is, as before, the ML method. The consistent initial estimates of the parameters for this case are given in Appendix 7.1. It should be noted that under such an assumption of the error process, it is possible for us to separately estimate all the five parameters involved in $\underset{\sim}{\underset{\sim}{\theta}}$. As before we may be be interested in testing different specific hypotheses characterizing different types of autocorrelation against the alternative given by ARMA(1,1) - AR(1) disturbances. Without going into details we write down different hypotheses that might interest us from the point of view of the nature of autocorrelation.

$H_o^6 : \rho_u = 0$ i.e., both ε_t and Z_t follow AR(1) process.

$H_o^7 : \rho_\varepsilon = 0$ i.e., ε_t follows MA(1) while Z_t follows AR(1).

$H_o^8 : \rho_u = 0, \ \rho_\varepsilon = 0$ (or, $\rho_Z = 0$) i.e., one of the components follows AR(1) while the other is non-autocorrelated.

H_o^9 : $\rho_u = 0$, $\rho_\varepsilon = 0$ (or, $\rho_Z = 0$), $\sigma_\varepsilon^2 = 0$ (or, $\sigma_Z^2 = 0$) i.e., one of the components follows AR(1) while the other is absent.

H_o^{10} : $\rho_\varepsilon = 0$, $\rho_Z = 0$: ε_t follows MA(1) while Z_t is non-autocorrelated.

H_o^{11} : $\rho_\varepsilon = 0$, $\rho_Z = 0$, $\sigma_Z^2 = 0$: ε_t follows MA(1) and Z_t is absent.

(In all the hypotheses in this set-up, the restriction $\rho_u = 0$ will imply $\upsilon_\varepsilon = \rho_\varepsilon$ in terms of the parameters in $\tilde{\tilde{\theta}}$.)

Any of these hypotheses can be tested against the alternative H_1^I where no restriction is imposed on the parameters and the following will be the appropriate LR test statistics :

$$LR(6) = -2\left[L(H_o^6) - L(H_1^I)\right] \sim \chi_1^2 \text{ under } H_o^6$$

$$LR(7) = -2\left[L(H_o^7) - L(H_1^I)\right] \sim \chi_1^2 \text{ under } H_o^7$$

$$LR(8) = -2\left[L(H_o^8) - L(H_1^I)\right] \sim \chi_2^2 \text{ under } H_o^8$$

$$LR(9) = -2\left[L(H_o^9) - L(H_1^I)\right] \sim \chi_3^2 \text{ under } H_o^9$$

$$LR(10) = -2\left[L(H_o^{10}) - L(H_1^I)\right] \sim \chi_2^2 \text{ under } H_o^{10}$$

$$LR(11) = -2\left[L(H_o^{11}) - L(H_1^I)\right] \sim \chi_3^2 \text{ under } H_o^{11}.$$

As before, one can as well think of testing a series of nested hypotheses to ascertain the correct nature of autocorrelation of the disturbances. The following three sets of nested hypotheses can be specified.

$H_o^6 : \rho_u = 0$

$H_o^7 : \rho_\varepsilon = 0$

$H_o^7 : \rho_\varepsilon = 0$

$H_o^8 : \rho_u = \rho_\varepsilon = 0$

(or, $\rho_u = \rho_Z = 0$)

$H_o^8 : \rho_\varepsilon = \rho_u = 0$

$H_o^9 : \rho_\varepsilon = \rho_u = \sigma_\varepsilon^2 = 0$

$H_o^{10} : \rho_\varepsilon = \rho_Z = 0$

$H_o^{11} : \rho_\varepsilon = \rho_Z = \sigma_Z^2 = 0$

$H_o^9 : \rho_u = \rho_\varepsilon = \sigma_\varepsilon^2 = 0$

(or, $\rho_u = \rho_Z = \sigma_Z^2 = 0$)

$H_o^{12} : \rho_\varepsilon = \rho_u = \sigma_\varepsilon^2 = \rho_Z = 0$

$H_o^{12} : \rho_\varepsilon = \rho_Z = \sigma_Z^2 = \rho_u = 0$.

$H_o^{12} : \rho_u = \rho_\varepsilon = \sigma_\varepsilon^2 = \rho_Z = 0$

(or, $\rho_u = \rho_Z = \sigma_Z^2 = \rho_\varepsilon = 0$)

These can be tested against the alternatives

$H_1^6 : \rho_u \neq 0$

$H_1^7 : \rho_\varepsilon \neq 0$

$H_1^7 : \rho_\varepsilon \neq 0$

$H_1^8 : \rho_u = 0, \rho_\varepsilon \neq 0$

(or, $\rho_u = 0, \rho_Z \neq 0$)

$H_1^8 : \rho_\varepsilon = 0, \rho_u \neq 0$

$H_1^9 : \rho_\varepsilon = \rho_u = 0, \sigma_\varepsilon^2 \neq 0$

$H_1^{10} : \rho_\varepsilon = 0, \rho_Z \neq 0$

$H_1^{11} : \rho_\varepsilon = \rho_Z = 0, \sigma_Z^2 \neq 0$

$H_1^9 : \rho_u = \rho_\varepsilon = 0, \sigma_\varepsilon^2 \neq 0$

(or, $\rho_u = \rho_Z = 0, \sigma_Z^2 \neq 0$)

$H_1^{12} : \rho_\varepsilon = \rho_u = \sigma_\varepsilon^2 = 0, \rho_Z \neq 0$

$H_1^{12} : \rho_\varepsilon = \rho_Z = \sigma_Z^2 = 0, \rho_\varepsilon \neq 0$.

$H_1^{12} : \rho_u = \rho_\varepsilon = \sigma_\varepsilon^2 = 0, \rho_Z \neq 0$.

(or, $\rho_u = \rho_Z = \sigma_Z^2 = 0, \rho_\varepsilon \neq 0$)

All the appropriate LR-test statistics would follow χ_1^2 under corresponding null hypotheses. It should be noted that here we are able to identify all possible types of autocorrelations within the set-up of our model including those cases where there is only one component in the

disturbances. In terms of misspecification, this means we are now able to distinguish (by LR-tests) between situations like (i) the autocorrelation in the disturbances is due to both misspecification and errors-in-observations and (ii) the autocorrelation in the disturbances is due to autocorrelation in one component (anyone of misspecification and errors-in-observation) only while the other is absent and so on. It should also be noted that situations like the one where one of the components follows MA(1) while the other is present but non-autocorrelated and where one of the components follows MA(1) while the other is absent — which could not be tested in the ARMA(1, 1) - ARMA(1, 1) set up — can now be discriminated. Though maximum likelihood values under the two situations will remain the same, the degree(s) of freedom, as stated earlier, will now change.

To sum up, if we start with a model where the two-component disturbance term follows a process of the type ARMA(1, 1) - ARMA(1, 1), we can test different hypotheses of interest like ARMA(1,1)- MA(1), MA(1) - MA(1) and so on. If, however, the hypothesis of ARMA(1,1)- MA(1) is rejected in favour of ARMA(1,1) - ARMA(1,1), we are unable to test ARMA(1,1) - AR(1) against ARMA(1,1) - ARMA(1,1). To overcome this we test hypotheses that can be treated as cases under ARMA(1,1) - AR(1) e.g., AR(1) - AR(1), MA(1) - AR(1) etc., against ARMA(1,1) - AR(1) or suitable alternatives. If, however, all these are rejected in favour of ARMA(1, 1) - AR(1), we cannot conclude if the actual situation is ARMA(1, 1) - ARMA(1,1) or ARMA(1,1) - AR(1). Before concluding we note

that to test these hypotheses we will require maximum likelihood values under the relevant hypotheses. In Appendix 7.1 we give consistent initial estimators for all the cases.

7.8 Conclusions

In this Chapter we have considered the problem of estimation and testing in an autocorrelated linear regression model with the error term decomposed into two components each following an ARMA(1, 1) process. We have suggested the maximum likelihood method for estimation of such a model. Under the standard assumptions the ML method of estimation obviously provides consistent and efficient estimates of both the regression coefficients and of the variance-covariance matrix of these estimates. Likelihood ratio tests have also been suggested for identifying the exact nature of the autocorrelation structure of the error term.

Appendix 7.1

Algebraic Solutions of the Parameters Under Different Assumptions Regarding the Error Components

To start with we present the relevant equations obtainable from (7.6.2), under different assumptions about the error components and get the algebraic solutions for the parameters.

ARMA(1,1) - ARMA(1,1)

$$S_0 = \sigma^2 \qquad \ldots (A\ 7.1.1)$$

$$S_1 = \tilde{\upsilon}_\varepsilon + \tilde{\upsilon}_Z \qquad \ldots (A\ 7.1.2)$$

$$S_2 = \rho_\varepsilon\ \tilde{\upsilon}_\varepsilon + \rho_Z\ \tilde{\upsilon}_Z \qquad \ldots (A\ 7.1.3)$$

$$S_3 = \rho_\varepsilon^2\ \tilde{\upsilon}_\varepsilon + \rho_Z^2\ \tilde{\upsilon}_Z \qquad \ldots (A\ 7.1.4)$$

$$S_4 = \rho_\varepsilon^3\ \tilde{\upsilon}_\varepsilon + \rho_Z^3\ \tilde{\upsilon}_Z . \qquad \ldots (A\ 7.1.5)$$

Solution of σ^2 is obviously given by $\hat{\sigma}^2 = S_0 .$ $\ldots (A\ 7.1.6)$

From (A 7.1.2)

$$\tilde{\upsilon}_\varepsilon = S_1 - \tilde{\upsilon}_Z . \qquad \ldots (A\ 7.1.7)$$

Substituting $\tilde{\upsilon}_\varepsilon$ fror (A 7.1.7) in (A 7.1.3), (A 7.1.4) and (A 7.1.5), we have

$$\tilde{\upsilon}_Z (\rho_Z - \rho_\varepsilon) = S_2 - \rho_\varepsilon\ S_1 \qquad \ldots (A\ 7.1.8)$$

$$\tilde{\upsilon}_Z (\rho_Z^2 - \rho_\varepsilon^2) = S_3 - \rho_\varepsilon^2\ S_1 \qquad \ldots (A\ 7.1.9)$$

$$\tilde{\upsilon}_Z (\rho_Z^3 - \rho_\varepsilon^3) = S_4 - \rho_\varepsilon^3\ S_1 . \qquad \ldots (A\ 7.1.10)$$

Now, replacing the expression of $\tilde{\upsilon}_Z(\rho_Z - \rho_\varepsilon)$ from (A 7.1.8) in (A 7.1.9) and (A 7.1.10), we have

$$\rho_Z + \rho_\varepsilon = \frac{S_3 - \rho_\varepsilon^2 S_1}{S_2 - \rho_\varepsilon S_1} \quad \ldots \text{(A 7.1.11)}$$

and

$$\rho_Z^2 + \rho_\varepsilon^2 + \rho_Z\rho_\varepsilon = \frac{S_4 - \rho_\varepsilon^3 S_1}{S_2 - \rho_\varepsilon S_1} \cdot \quad \ldots \text{(A 7.1.12)}$$

From (A 7.1.11),

$$\rho_Z = \frac{S_3 - \rho_\varepsilon^2 S_1}{S_2 - \rho_\varepsilon S_1} - \rho_\varepsilon$$

$$= \frac{S_3 - \rho_\varepsilon S_2}{S_2 - \rho_\varepsilon S_1} \cdot \quad \ldots \text{(A 7.1.13)}$$

Putting the value of ρ_Z from (A 7.1.13) into (A 7.1.12), we find

$$\left(\frac{S_3 - \rho_\varepsilon S_2}{S_2 - \rho_\varepsilon S_1}\right)^2 + \rho_\varepsilon^2 + \rho_\varepsilon\left(\frac{S_3 - \rho_\varepsilon S_2}{S_2 - \rho_\varepsilon S_1}\right) = \frac{S_4 - \rho_\varepsilon^3 S_1}{S_2 - \rho_\varepsilon S_1} \cdot$$

After algebraic simplification we have the solution for ρ_ε, say $\hat{\rho}_\varepsilon$, as a root of the following quadratic equation

$$\rho_\varepsilon^2(S_2^2 - S_1S_3) + \rho_\varepsilon(S_4S_1 - S_3S_2) + (S_3^2 - S_4S_2) = 0 \cdot \quad \ldots \text{(A 7.1.14)}$$

Solutions of ρ_Z, $\tilde{\upsilon}_Z$ and $\tilde{\upsilon}_\varepsilon$ are obtained from (A 7.1.13), (A 7.1.8) and (A 7.1.7) as

$$\hat{\rho}_Z = \frac{S_3 - \hat{\rho}_\varepsilon S_2}{S_2 - \hat{\rho}_\varepsilon S_1}$$

$$\hat{\tilde{\upsilon}}_Z = \frac{S_2 - \hat{\rho}_\varepsilon S_1}{\hat{\rho}_Z - \hat{\rho}_\varepsilon}$$

$$\hat{\tilde{\upsilon}}_\varepsilon = S_1 - \hat{\tilde{\upsilon}}_Z \cdot$$

In case we have two real solutions of ρ_ε from (A 7.1.14) and both of them lie between -1 and +1 ($-1 < \rho_\varepsilon < 1$ from stationary condition), we choose the root for which (using corresponding values of $\tilde{\upsilon}_\varepsilon$, $\tilde{\upsilon}_Z$ and ρ_Z) the value of the likelihood function is greater.

ARMA(1, 1) - AR(1)

$$S_0 = \sigma_\varepsilon^2 + \sigma_Z^2 \qquad \ldots (A\ 7.1.15)$$

$$S_1 = \upsilon_\varepsilon \sigma_\varepsilon^2 + \rho_Z \sigma_Z^2 \qquad \ldots (A\ 7.1.16)$$

$$S_2 = \rho_\varepsilon \upsilon_\varepsilon \sigma_\varepsilon^2 + \rho_Z^2 \sigma_Z^2 \qquad \ldots (A\ 7.1.17)$$

$$S_3 = \rho_\varepsilon^2 \upsilon_\varepsilon \sigma_\varepsilon^2 + \rho_Z^3 \sigma_Z^2 \qquad \ldots (A\ 7.1.18)$$

$$S_4 = \rho_\varepsilon^3 \upsilon_\varepsilon \sigma_\varepsilon^2 + \rho_Z^4 \sigma_Z^2 . \qquad \ldots (A\ 7.1.19)$$

From (A 7.1.16),

$$\upsilon_\varepsilon \sigma_\varepsilon^2 = S_1 - \rho_Z \sigma_Z^2 . \qquad \ldots (A\ 7.1.20)$$

Substituting $\upsilon_\varepsilon \sigma_\varepsilon^2$ from (7.1.20) in (A 7.1.17), (A 7.1.18) and (A 7.1.19), we have

$$\rho_Z \sigma_Z^2 (\rho_Z - \rho_\varepsilon) = S_2 - \rho_\varepsilon S_1 \qquad \ldots (A\ 7.1.21)$$

$$\rho_Z \sigma_Z^2 (\rho_Z^2 - \rho_\varepsilon^2) = S_3 - \rho_\varepsilon^2 S_1 \qquad \ldots (A\ 7.1.22)$$

and

$$\rho_Z \sigma_Z^2 (\rho_Z^3 - \rho_\varepsilon^3) = S_4 - \rho_\varepsilon^3 S_1 . \qquad \ldots (A\ 7.1.23)$$

Replacing the expression of $\rho_Z \sigma_Z^2 (\rho_Z - \rho_\varepsilon)$ from (A 7.1.21) in (A 7.1.22) and (A 7.1.23), we find

$$\rho_Z = \frac{S_3 - \rho_\varepsilon^2 S_1}{S_2 - \rho_\varepsilon S_1} - \rho_\varepsilon$$

$$= \frac{S_3 - \rho_\varepsilon S_2}{S_2 - \rho_\varepsilon S_1} . \qquad \ldots (A\ 7.1.24)$$

Substituting ρ_Z from (A 7.1.24) in (A 7.1.23), we have

$$\rho_\varepsilon^2 (S_2^2 - S_1 S_3) + \rho_\varepsilon (S_4 S_1 - S_3 S_2) + (S_3^2 - S_4 S_2) = 0 . \quad \ldots (A\ 7.1.25)$$

Solution of ρ_ε, say $\hat{\rho}_\varepsilon$ is obtained from (A 7.1.25) and those of ρ_Z, σ_Z^2, σ_ε^2 and υ_ε are obtained from (A 7.1.24), (A 7.1.21), (A 7.1.15) and (A 7.1.20) as

$$\hat{\rho}_Z = \frac{S_3 - \hat{\rho}_\varepsilon S_2}{S_2 - \hat{\rho}_\varepsilon S_1}$$

$$\hat{\sigma}_Z^2 = \frac{S_2 - \hat{\rho}_\varepsilon S_1}{\hat{\rho}_Z (\hat{\rho}_Z - \hat{\rho}_\varepsilon)}$$

$$\hat{\sigma}_\varepsilon^2 = S_0 - \hat{\sigma}_Z^2$$

and

$$\hat{\upsilon}_\varepsilon = \frac{S_1 - \hat{\rho}_Z \hat{\sigma}_Z^2}{\hat{\sigma}_\varepsilon^2} .$$

As in the earlier case, in the case of more than one real solution of ρ_ε lying within the range -1 and +1, we choose that root corresponding to which the likelihood value is greater.

ARMA(1, 1) - MA(1)

$$S_0 = \sigma^2 \quad \ldots \quad (A\ 7.1.26)$$

$$S_1 = \tilde{\upsilon}_\varepsilon + \tilde{\upsilon}_Z \quad \ldots \quad (A\ 7.1.27)$$

$$S_2 = \rho_\varepsilon \tilde{\upsilon}_\varepsilon \quad \ldots \quad (A\ 7.1.28)$$

$$S_3 = \rho_\varepsilon^2 \tilde{\upsilon}_\varepsilon . \quad \ldots \quad (A\ 7.1.29)$$

The estimates of ρ_ε, $\tilde{\upsilon}_\varepsilon$, $\tilde{\upsilon}_Z$ and σ^2 are obtained from the above relations as

$$\hat{\rho}_\varepsilon = \frac{S_3}{S_2}$$

$$\hat{\tilde{\upsilon}}_\varepsilon = \frac{S_2}{\hat{\rho}_\varepsilon}$$

$$\hat{\tilde{\upsilon}}_Z = S_1 - \hat{\tilde{\upsilon}}_\varepsilon$$

and

$$\hat{\sigma}^2 = S_0 .$$

MA(1) - MA(1)

Solutions of σ^2 and $\tilde{\upsilon}$ $(= \tilde{\upsilon}_\varepsilon + \tilde{\upsilon}_Z)$ are given as

$$\hat{\sigma}^2 = S_0 \quad \text{and} \quad \hat{\tilde{\upsilon}} = S_1 .$$

As stated in footnote 16, even though solutions of $\tilde{\upsilon}_\varepsilon$ and $\tilde{\upsilon}_Z$ are not obtainable separately, the maximum value of the likelihood function in terms of the parameters σ^2 and $\tilde{\upsilon}$ is obviously obtained from $\hat{\sigma}^2$ and $\hat{\tilde{\upsilon}}$.

AR(1) - AR(1)

$$S_0 = \sigma_\varepsilon^2 + \sigma_Z^2 \qquad \ldots (\Lambda\ 7.1.30)$$

$$S_1 = \rho_\varepsilon \sigma_\varepsilon^2 + \rho_Z \sigma_Z^2 \qquad \ldots (\Lambda\ 7.1.31)$$

$$S_2 = \rho_\varepsilon^2 \sigma_\varepsilon^2 + \rho_Z^2 \sigma_Z^2 \qquad \ldots (\Lambda\ 7.1.32)$$

$$S_3 = \rho_\varepsilon^3 \sigma_\varepsilon^2 + \rho_Z^3 \sigma_Z^2 . \qquad \ldots (\Lambda\ 7.1.33)$$

From (Λ 7.1.30),

$$\sigma_\varepsilon^2 = S_0 - \sigma_Z^2 . \qquad \ldots (\Lambda\ 7.1.34)$$

Substituting the value of σ_ε^2 in (Λ 7.1.31), (Λ 7.1.32) and (Λ 7.1.33), we find

$$S_1 = \rho_\varepsilon S_0 + \sigma_Z^2 (\rho_Z - \rho_\varepsilon) \qquad \ldots (\Lambda\ 7.1.35)$$

$$S_2 = \rho_\varepsilon^2 S_0 + \sigma_Z^2 (\rho_Z^2 - \rho_\varepsilon^2) \qquad \ldots (\Lambda\ 7.1.36)$$

$$S_3 = \rho_\varepsilon^3 S_0 + \sigma_Z^2 (\rho_Z^3 - \rho_\varepsilon^3) . \qquad \ldots (\Lambda\ 7.1.37)$$

After substituting the value of $\sigma_Z^2 (\rho_Z - \rho_\varepsilon)$ from (Λ 7.1.35) in (Λ 7.1.36) and (Λ 7.1.37) and then simplifying, we have as in the first two cases the solutions of ρ_ε as the roots of the following quadratic equation :

$$\rho_\varepsilon^2 (S_1^2 - S_0 S_2) + \rho_\varepsilon (S_3 S_0 - S_1 S_2) + (S_2^2 - S_3 S_1) = 0 .$$

The solutions of ρ_Z, σ_Z^2 and σ_ε^2 are given as

$$\hat{\rho}_Z = \frac{S_2 - \hat{\rho}_\varepsilon S_1}{S_1 - \hat{\rho}_\varepsilon S_0}$$

$$\hat{\sigma}_Z^2 = \frac{S_1 - \hat{\rho}_\varepsilon S_0}{\hat{\rho}_Z - \hat{\rho}_\varepsilon}$$

and $\hat{\sigma}_\varepsilon^2 = S_0 - \hat{\sigma}_Z^2 .$

MA(1) - No autocorrelation (or absent)

The solutions of σ^2 and $\tilde{\upsilon}_\varepsilon$ are given as

$$\hat{\sigma}^2 = S_0$$

and

$$\hat{\tilde{\upsilon}}_\varepsilon = S_1 .$$

MA(1) - absent

The solutions of σ_ε^2 and υ_ε are given from $S_0 = \sigma_\varepsilon^2$ and $S_1 = \upsilon_\varepsilon \sigma_\varepsilon^2$ as

$$\hat{\sigma}_\varepsilon^2 = S_0$$

and

$$\hat{\upsilon}_\varepsilon = S_1 / \hat{\sigma}_\varepsilon^2 .$$

AR(1) - No autocorrelation

The solutions of ρ_ε, σ_ε^2 and σ_Z^2 are obtained from the following equations

$$S_0 = \sigma_\varepsilon^2 + \sigma_Z^2$$

$$S_1 = \rho_\varepsilon \sigma_\varepsilon^2$$

$$S_2 = \rho_\varepsilon^2 \sigma_\varepsilon^2$$

as $\hat{\rho}_\varepsilon = S_2 / S_1$, $\hat{\sigma}_\varepsilon^2 = S_1 / \hat{\rho}_\varepsilon$ and $\hat{\sigma}_Z^2 = S_0 - \hat{\sigma}_\varepsilon^2$.

AR(1) - absent

The solutions of σ_ε^2 and ρ_ε are given as

$$\hat{\sigma}_\varepsilon^2 = S_0$$

and

$$\hat{\rho}_\varepsilon = S_1 / \hat{\sigma}_\varepsilon^2 .$$

PART III

A MONTE CARLO STUDY ON SOME ESTIMATORS IN AUTOCORRELATED MODELS WITH MISSPECIFICATION

Chapter 8

Background of the Study and Survey of Literature

8.1 Introduction

The interest in this part of the thesis centres on small-sample properties of various estimators (generally used in the context of autocorrelated linear regression models with AR(1) error process) in autocorrelated linear regression models where the problem of misspecification is present. It has been noted in Chapter 5 that until recently most of the reestimation methods (also called two-step procedures) e.g., Cochrane-Orcutt (CO), Prais-Winsten (PW), Durbin's two-step(D2)and Hildreth and Lu's (HL) search procedures, assume the error term to follow the AR(1) process. It is known that under standard conditions [see Theil (1971), Ch. 8, pp.405-407, Magnus (1978)] all these conventional estimators possess the same asymptotic distribution. More recently, researchers have suggested the use of ML method of estimation for such autocorrelated linear regression models. In fact, Hildreth and Dent (1974) and Beach and MacKinnon (1978a) have developed algorithms for obtaining ML estimates of autocorrelated linear regression models with AR(1) error process. Asymptotically, ML estimator is equivalent to the conventional estimators if the first-order autocorrelation coefficient of the error process is not too large. It is also known that in autocorrelated linear regression models, OLS is not, in general, efficient. Since asymptotically all these methods,excepting the OLS, are equivalent, a choice among these alternative estimators must be based on their small-sample properties. However,

the small-sample properties are difficult to investigate analytically, and so one has to use Monte Carlo studies for this purpose. Some small-sample studies have been conducted on this problem. These are discussed in the following section. It appears that these studies are not adequate; for they fail to give any clear-cut indication regarding the choice among the alternative estimators.

It may also be noted in this context that all such small-sample studies on . autocorrelated linear regression model with AR(1) error process, take no note of the possibility that the working model may have been misspecified[1/] and that there may be a component, apart from the one which is the disturbance term of the true regression model, in the error term which is due to misspecification and which might or might not contribute to autocorrelation. We have argued in Chapter 6 why such decomposition of the error term provides valuable insight into the nature of autocorrelation and helps in ascertaining if the available methods of estimation are appropriate in a particular situation and if not, in evolving suitable alternative methods. We also quoted there from Maddala (1977, p. 291) to argue why there should be further researches in order to ascertain how different standard 'efficient' methods viz., CO, PW, D2 and HL perform in relation to OLS in presence of misspecification which might or might not give rise to autocorrelation. One such study in the context of large samples has been done by Chaudhuri (1977, 1979, Chapters 2 and 3) some of whose results have been mentioned in Chapter 5.

1/ As in Part II of this dissertation, we mean by the term 'misspecification' situations where either some relevant regressors have been left out or some functions of some of the included regressors have been omitted.

She proved that excepting for the case where the excluded regressors bear strictly linear relationship to the included ones, the conventional two-step estimators would yield inconsistent estimates of the properly defined parameters of the misspecified model. However, no Monte Carlo study on the small-sample performance of conventional estimators in autocorrelated models with misspecification has so far been undertaken.

Our objective in this part of the dissertation is precisely to report the results of such a Monte Carlo study conducted by us. In other words, the question we investigate is : How do the conventional estimators (which obviously disregard misspecification) perform in small samples in the presence of misspecification giving rise to possible autocorrelation in the errors ? Since our primary interest was to examine the performance of the conventional two-step methods and the OLS, we left out the ML method of estimation from the purview of our study.[2/] Furthermore, though the methods of estimation suggested by us in Chapters 6 and 7 are relevant for the type of models being considered here, we could not actually examine the small-sample properties of our methods because the Monte Carlo experiment was completed before these methods were developed.

In the remaining sections of this Chapter, we make a brief survey of the earlier small-sample studies on autocorrelated linear regression models most of which assume AR(1) error process. This, it is hoped, will highlight the distinctive features of our present work. The next Chapter (i.e., Chapter 9) is devoted to an examination of the unbiasedness and

---

2/ Actually, one explanation of why the ML method of estimation has not been considered is that it has come to be used for such models only relatively recently.

consistency of the estimators obtained by some standard methods of estimation in presence of misspecification. These results (some of which were obtained by previous researchers) are presented for the sake of completeness and convenience in appraising the small-sample results. In Chapters 10 and 11 we present the results of our Monte Carlo study; Chapter 10 describes the design of the experiment and Chapter 11 gives the numerical results.

8.2 Monte Carlo Studies on Autocorrelated Linear Regression Models

Most of the studies on the performance of different methods of estimation in small samples have been done on the assumption that the disturbances follow an AR(1) process. Particular mention may be made of the studies by Hildreth and Lu (1969), Rao and Griliches (1969), Hildreth and Dent (1974), Beach and MacKinnon (1978a) and Spitzer (1979).

Rao and Griliches (1969) were the first to undertake a Monte Carlo study on this problem. They considered a one-regressor model where the error was assumed to be generated by an AR(1) process. Their model was

$$y_t = \beta x_t + \varepsilon_t \; ; \qquad t = 1, 2, \ldots, n$$

$$x_t = \eta x_{t-1} + v_t$$

and $$\varepsilon_t = \rho \varepsilon_{t-1} + u_t$$

where $|\eta| < 1$, $|\rho| < 1$ and each of u_t's and v_t's are uncorrelated and distributed with zero mean and same variance σ_u^2 and σ_v^2 respectively; also x and ε are assumed to be independent. They considered five

different methods viz., Ordinary Least Squares (OLS), Prais-Winsten (PW), Cochrane-Orcutt (CO), Durbin's two-step (D2) and a nonlinear[3/] method. The relative efficiency of these methods were assessed by means of mean square error (mse) criterion. They fixed the value of β at 1.0 and obtained the estimates corresponding to the values of ρ and η at intervals of 0.1 and 0.2 respectively in the range $-1 < \rho, \eta < 1$; σ_u^2 and σ_v^2 were adjusted so as to make the square of the true correlation coefficient between x and y to be equal to 0.9. For any given ρ and η, 50 samples each of size 20 were drawn from independent x and ε series. Their main conclusions were the following :

(i) No method is better than the others over the entire range of parameter values.

(ii) For samples of size 20, there is a gain in efficiency in the two-step procedures relative to OLS if $|\rho| > 0.3$. For values of $|\rho| < 0.3$, there may be a little loss in efficiency in using the two-step procedures.

(iii) The estimate of ρ in Durbin's two-step method is probably better than the other estimates and a mixed two-stage estimator where $\hat{\rho}$ obtained by Durbin's procedure is used to estimate the regression coefficient by the Prais-Winsten method should be preferred.

(iv) The nonlinear method does not appear to do better than the others.

3/ In the nonlinear method, least squares is applied to the following equation

$$y_t = \rho y_{t-1} + \beta x_t - \beta \rho x_{t-1} + u_t$$

with the imposition of the nonlinear constraint $\widehat{\beta\rho} = \hat{\beta}\hat{\rho}$.

The later Monte Carlo studies on the linear autocorrelated regression model with AR(1) error process have concentrated mainly on the performance of ML estimator[4/] which is, strictly speaking, different from the conventional two-step estimators (see Judge et al. (1980) and the next Chapter for details). No major Monte Carlo study appears to have been carried out to compare in detail the small-sample performance of all the conventional two-step estimators and the ML estimator. However, Beach and MacKinnon (1978a) have considered a model somewhat similar to that of Rao and Griliches and have compared the ML estimator vis-a-vis the CO estimator in small samples. They examined the following model :

$$y_t = \alpha + \beta x_t + \varepsilon_t$$

$$\varepsilon_t = \rho \varepsilon_{t-1} + u_t$$

$$u_t \sim N(0, 0.0036)$$

where $|\rho| < 1$ and u_t's are independent.

$$x_t = \exp(0.04t) + v_t$$

$$v_t \sim N(0, 0.00009)$$

and v_t's are independent.

They fixed the true values of both α and β at 1.0. Each experiment was replicated 200 times and two different sample sizes, 20 and 50, were considered. Three different values of ρ — 0.6, 0.8 and 0.99 —

4/ The ML method has recently been suggested by Hildreth and Dent (1974) and Beach and MacKinnon (1978a) for estimation of such models. They have also suggested algorithms for obtaining the ML estimates. Beach and MacKinnon have in their algorithm incorporated the stationary condition.

were used. The ML and CO procedures were compared in respect of bias and root mean square error (rmse).

Their main conclusions can be summarized as follows :

(i) Judging by the rmse criterion, the superiority of ML estimator over CO is often dramatic for α ; for β the superiority is always substantial.

(ii) By the criterion of number of replications for which the ML estimates were closer to the true values than that of CO, ML estimators of α and β always did better than CO and the difference was usually significant.

(iii) The ML estimation procedure was found to be computationally quite efficient. More specifically, it was found to be computationally no more expensive than the CO procedure.

They also considered nontrended x-values where x_t's were assumed to be independent and having distribution N(0, 0.0625). Sample size 20 was considered for this case. The values of ρ considered remained same. As for the estimate of β virtually no difference was observed between the ML and CO estimators. However, ML estimator was found to be better for α.

Spitzer (1979) compared the ML estimator and the CO iterative estimator[5/] (the latter ignores the first observation and the Jacobian term of the ML equation) in small samples. He also observed the superiority of ML method over the CO iterative procedure.

5/ For the CO iterative estimator, estimates of the regression coefficients obtained at the second step is used to estimate the errors. These are used to obtain another estimate of the autocorrelation coefficient and then of the regression coefficients and so on till convergence is attained. Sargan (1964) proved that this iterative procedure will converge at least to a local minima.

Since the estimate of ρ affects that of β, some work has been done to study the small-sample properties of the estimator of ρ obtained by different methods. Maddala (1977, p.280), for example, has reported some empirical results where the D2 procedure has been found to yield estimate of ρ substantially different from those obtained by the ML method. Experiments by Hildreth and Lu (1969) and Beach and MacKinnon (1978a), on the other hand, indicate substantial small-sample bias in the ML estimate of ρ. Hildreth and Dent (1974) suggested adjustment to the ML estimator that reduces the bias but increases the sampling variance. The adjusted ML estimator of ρ has generally been found to be the best when $\rho > 0.4$. But for values of $\rho < 0.4$, the Theil and Nagar (1961) estimator 6/ ρ^* seems to be the best. However, whether or not this leads to improved estimates of the regression coefficients needs further investigation.

There is no significant small-sample study where the disturbances are assumed to follow autoregressive process of order greater than one. Though the estimators given by the standard methods proposed for such models are asymptotically equivalent, there is little evidence to suggest the superiority of any particular method over the others in small samples. Nevertheless, the ML method of estimation suggested by Beach and MacKinnon (1978b), for models with AR(2) error process being based on all the observations and having incorporated the stationarity conditions, may do better than the other methods.

6/ Theil and Nagar (1961) estimator of ρ is defined as

$$\rho^* = \frac{n^2(1-d/2) + k^2}{n^2 - k^2}$$

where d is the value of the DW statistic,
n is the number of observations,
and k is the number of regressors in the regression equation.

There have also been very few small-sample studies with MA processes for the disturbances. Excepting for the study by Hendry and Trivedi (1972), the other available studies e.g., Nelson (1974), Dent and Min (1978) were carried out in the context of pure time series models and hence the results are not directly relevant for the autocorrelated linear regression models. However, here again the ML method of estimation that uses all the observations may do better than the other methods provided one is willing to undertake the additional computations.7/

8.3 Conclusions

We may reiterate that the small-sample studies done so far in the context of autocorrelated linear regression models with AR(1) error process disregard the possible presence of a misspecification component in the disturbances which might or might not contribute to autocorrelation. There appears to have been no work to study the effect of misspecification (and therefore of consequent autocorrelation) on the conventional reestimation methods in an autocorrelated linear regression model in the context of small samples. We shall present the results of such a study in Chapters 10 and 11. In Chapter 9 we present some results, mainly based on large samples, bearing on this problem which should be useful for comparative purposes.

7/ Some notable references of Monte Carlo studies done in the past on different econometric problems are : Wagner (1958), Basmann (1960), Nagar (1960), Quandt (1965), Summers (1965), Cragg (1967), Kmenta and Gilbert (1970), Schmidt (1971), Byron (1972), Ramsey and Gilbert (1972), Smith and Hall (1972), Yancy et al. (1972), Spitzer (1978) etc.

Chapter 9

Unbiasedness and Consistency of Some Estimators in an Autocorrelated Linear Regression Model in the Context of Misspecification

9.1 Introduction

We have stated in the last Chapter that the main purpose in this Part of the dissertation is to investigate the small-sample performance of different conventional reestimation methods used in estimating the parameters of an autocorrelated linear regression model in situations complicated by the presence of misspecification which might give rise to autocorrelation. In this Chapter we examine, for convenience of comparison, the unbiasedness and consistency properties of these estimators when the model has been misspecified. Some of the conventional reestimation methods used for the linear model where the error follows AR(1) are also briefly described for the sake of convenience since we shall compare the small-sample performance of these methods in the case of our misspecified model. Chaudhuri (1977, 1979) examined these properties for some of these estimators. We shall first summarize some of her findings that are relevant in the context of our present study and then examine a few other methods not explicitly analysed by her.

While in section 9.2 we describe the model which has been misspecified and whose properties we are going to study, in section 9.3 we examine unbiasedness and consistency properties of some estimators. Conclusions are given in section 9.4.

9.2 The Model

Following Chaudhuri, we specify the true model as one with k regressors and write this in matrix notation as

$$Y = X\beta + \varepsilon \qquad \ldots (9.2.1)$$

where $Y = (y_1, y_2, \ldots\ldots, y_n)'$ is a (nx1) vector of observations on the dependent variable,

X is a (nxk) matrix of observations on k stochastic regressors given by

$$X = \begin{pmatrix} x_{11} & x_{21} & \ldots & x_{k1} \\ x_{12} & x_{22} & \ldots & x_{k2} \\ \ldots & \ldots & \ldots & \ldots \\ x_{1n} & x_{2n} & \ldots & x_{kn} \end{pmatrix}$$

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

and

$$\varepsilon = (\varepsilon_1, \varepsilon_2, \ldots\ldots, \varepsilon_n)'$$

are (kx1) and (nx1) vectors of the regression coefficients and disturbances respectively. One of the regressors may be unity for all t incorporating thereby an intercept term in the regression equation. We make standard assumptions about ε_t's and X :

ε_t's are stationary with $E(\varepsilon_t) = 0$ for all $t = 1, 2, \ldots\ldots, n$.

$V(\varepsilon_t) = \sigma_\varepsilon^2 V_\varepsilon$ where V_ε is a positive definite matrix.

X is assumed to be independent of ε.

$\underset{n \to \infty}{\text{plim}} \left(\frac{1}{n} X'X \right) = \Sigma_{XX}$, a positive definite matrix.

Let us now suppose that m regressors have been omitted from the true model. Let us, without any loss of generality, assume that the last m regressors have been left out. As in Chapter 5, we define the misspecified model as

$$Y = X^{+}\beta^{+} + \varepsilon^{+} \qquad \ldots (9.2.2)$$

where X^{+} is the submatrix formed by the first k-m ($= k^{+}$, say) columns of X,

$\beta^{+} = (\beta_1^{+}, \beta_2^{+}, \ldots, \beta_{k^{+}}^{+})'$ is the $(k^{+} \times 1)$ vector of associated regression coefficients,

and ε^{+} is the $(n \times 1)$ vector of disturbances associated with the misspecified model.

It should be noted that β^{+} is not the sub-vector containing the first k^{+} elements of β. Since estimating and testing the significance of individual regression coefficients of the misspecified equation seem to be of practical importance, we argue, following Chaudhuri, that the regression coefficients associated with the included regressors should be redefined allowing them to capture as much of the partial influence of the omitted regressors on Y as possible; or in other words, to enable the regression function

$$Y = X^{+}\beta^{+}$$

to approximate as closely as possible the systematic component of Y i.e., $X\beta$.

Minimizing

$$E\{(Y - X^{+}\beta^{+})'(Y - X^{+}\beta^{+})\}$$

with respect to β^{+}, we can find the relation between β^{+} and β as

$$\beta^+ = P\beta \qquad \ldots \quad (9.2.3)$$

where 1/

$$P = \{E(X^{+\prime}X^+)\}^{-1} E(X^{+\prime}X) . \qquad \ldots \quad (9.2.4)$$

Now,

$$\varepsilon^+ = Y - X^+\beta^+ \qquad \text{(from (9.2.2))}$$

$$= \varepsilon + (X - X^+P)\beta \qquad \text{(using (9.2.1) and (9.2.3))}.$$

Thus

$$\varepsilon^+ = \varepsilon + Z \qquad \ldots \quad (9.2.5)$$

where 2/

$$Z = (X - X^+P)\beta. \qquad \ldots \quad (9.2.6)$$

We now make a further assumption that

$$\operatorname*{plim}_{n\to\infty} \left(\frac{1}{n} X^{+\prime}X^+ \right) = \Sigma_{X^+X^+} \quad \text{which exists and is non-singular.}$$

1/ It can easily be seen that minimization of

$$E\{(Y - X^+\beta^+)'(Y - X^+\beta^+)\}$$

is equivalent to minimizing

$$E\{(X\beta - X^+\beta^+)'(X\beta - X^+\beta^+)\} + nv\sigma_\varepsilon^2$$

where v is the t-th diagonal element of V_ξ for all t=1,2,..., n.

Differentiating the above expression with respect to β^+ and then equating the partial derivatives to zero, we easily get

$$\beta^+ = P\beta.$$

This solution of β^+ corresponds to the plane of best fit, on the average, in finite samples.

2/ Since P can be written as

$$P = (I \mid P_{21})$$

where $P_{21} = \{E(X^{+\prime}X^+)\}^{-1} E(X^{+\prime}\bar{X})$, and

$\bar{X}$ is the submatrix formed by the last m columns of X,

we can write

$$\beta^+ = \tilde{\beta} + P_{21}\tilde{\tilde{\beta}}$$

where $\beta' = (\tilde{\beta}' \mid \tilde{\tilde{\beta}}')$

and hence Z can as well be written as

$$Z = (\bar{X} - X^+P_{21})\tilde{\tilde{\beta}} .$$

9.3 Examination on the Properties of Unbiasedness and Consistency of Some Estimators

Having described the model in the previous section, we now study the properties of unbiasedness and consistency of some conventional estimators (e.g., OLS, GLS, etc.) in the context of our model. Before doing so, we may note that, in general,

$$E(Z) \neq 0$$

and hence, in general,

$$E(\varepsilon^{+}) \neq 0.$$

However, if the excluded regressors bear a strictly linear relationship with the included ones, then from (9.2.6) it easily follows that

$$E(Z) = 0$$

so that

$$E(\varepsilon^{+}) = 0.$$

We have also quite generally[3/]

$$\operatorname*{plim}_{n \to \infty} \left(\frac{1}{n} X^{+\prime} Z \right) = 0. \qquad \ldots \quad (9.3.1)$$

Let us now proceed to examine the aforesaid properties of some of the usual estimators (including the OLS and the GLS estimators) in the context of the misspecified model described in (9.2.2) to (9.2.6).

3/ Vide Chaudhuri (1979, pp.159-161) for a proof. However, the result can easily be verified from (9.2.6) and the fact that

$$\operatorname*{plim}_{n \to \infty} \left(\frac{1}{n} X'X \right) = \Sigma_{XX}$$

which implies that

$$\lim_{n \to \infty} E\left(\frac{1}{n} X'X \right) = \Sigma_{XX}$$

[vide Goldberger (1963, pp. 118-119)].

I. Ordinary Least Squares (OLS) Method

Let $\hat{\beta}^+_{OLS}$ denote the OLS estimator of β^+. Clearly,

$$\hat{\beta}^+_{OLS} = (X^{+\prime}X^+)^{-1} X^{+\prime} Y$$

$$= \beta^+ + (X^{+\prime}X^+)^{-1} X^{+\prime}\varepsilon^+ \qquad \text{(using 9.2.2)}$$

$$= \beta^+ + (X^{+\prime}X^+)^{-1} X^{+\prime}\varepsilon + (X^{+\prime}X^+)^{-1} X^{+\prime} Z \quad \text{(using 9.2.5)} \qquad \ldots\ldots\ldots\ldots (9.3.2)$$

Therefore, in general,

$$E(\hat{\beta}^+_{OLS}) \neq \beta^+$$

since

$$E(Z \mid X^+) \neq 0.$$

Now from (9.3.2),

$$\operatorname*{plim}_{n\to\infty} \hat{\beta}^+_{OLS} = \beta^+ + \operatorname*{plim}_{n\to\infty} \left(\frac{1}{n} X^{+\prime}X^+\right)^{-1} \operatorname*{plim}_{n\to\infty} \left(\frac{1}{n} X^{+\prime}\varepsilon\right)$$

$$+ \operatorname*{plim}_{n\to\infty} \left(\frac{1}{n} X^{+\prime}X^+\right)^{-1} \operatorname*{plim}_{n\to\infty} \left(\frac{1}{n} X^{+\prime}Z\right) \quad \ldots (9.3.3).$$

Since from (9.3.1),

$$\operatorname*{plim}_{n\to\infty} \left(\frac{1}{n} X^{+\prime}Z\right) = 0$$

and since we have assumed X and ε to be independent (so that X^+ and ε are independent) and ε has mean zero so that

$$\operatorname*{plim}_{n\to\infty} \left(\frac{1}{n} X^{+\prime}\varepsilon\right) = 0,$$

we have from (9.3.3),

$$\operatorname*{plim}_{n\to\infty} \hat{\beta}^+_{OLS} = \beta^+.$$

We thus find that if the disturbances in the true regression are auto-correlated, then the OLS estimator of the regression coefficients in

the misspecified model are biased but consistent. It is obvious from (9.3.3) that even if we assume Z_t's to be autocorrelated, this result holds.

II. Generalized Least Squares (GLS) Method

Let $\sigma^2 V$ denote the variance-covariance matrix of ε^+ where V is assumed to be positive definite and known.[4/] The GLS estimator of β^+, say $\hat{\beta}^+_{GLS}$, is then given as

$$\hat{\beta}^+_{GLS} = (X^{+\prime} V^{-1} X^+)^{-1} X^{+\prime} V^{-1} Y$$

$$= \beta^+ + (X^{+\prime} V^{-1} X^+)^{-1} X^{+\prime} V^{-1} \varepsilon + (X^{+\prime} V^{-1} X^+)^{-1} X^{+\prime} V^{-1} Z \quad \text{(using (9.2.2) and (9.2.5))} \qquad \ldots \quad (9.3.4)$$

Thus, in general

$$E(\hat{\beta}^+_{GLS}) \neq \beta^+$$

since

$$E(Z \mid X^+) \neq 0 \quad \text{necessarily.}$$

Now from (9.2.6) and (9.2.3),

$$Z = X\beta - X^+ P\beta \quad = X\beta - X^+\beta^+$$

and hence we have from (9.3.4),

$$\hat{\beta}^+_{GLS} = (X^{+\prime} V^{-1} X^+)^{-1} X^+ V^{-1} \varepsilon + (X^{+\prime} V^{-1} X^+)^{-1} X^{+\prime} V^{-1} X\beta \qquad \ldots (9.3.5)$$

4/ If X is non-stochastic,

$$V = V_\varepsilon \quad \text{and} \quad \sigma^2 = \sigma^2_\varepsilon .$$

These equalities do not hold if X is stochastic, as assumed here.

We will assume that

$$\operatorname*{plim}_{n\to\infty} \left(\frac{1}{n} X^{+\prime} V^{-1} X \right) = C \quad \text{(exists)} \qquad \ldots \quad (9.3.6)$$

$$\lim_{n\to\infty} E\left(\frac{1}{n} X^{+\prime} V^{-1} V_{\varepsilon} V^{-1} X^{+} \right) = D \quad \text{(exists and positive definite)} \qquad \ldots \quad (9.3.7)$$

$$\operatorname*{plim}_{n\to\infty} \left(\frac{1}{n} X^{+\prime} V^{-1} X^{+} \right) = E \quad \text{(exists and non-singular)} \ldots (9.3.8)$$

Now noting that

$$E\left(\frac{1}{n} X^{+\prime} V^{-1} \varepsilon \right) = 0$$

and

$$\lim_{n\to\infty} V\left(\frac{1}{n} X^{+\prime} V^{-1} \varepsilon \right) = 0 \quad \text{(from (9.3.7))}$$

it follows from Chebychev's inequality that

$$\operatorname*{plim}_{n\to\infty} \left(\frac{1}{n} X^{+\prime} V^{-1} \varepsilon \right) = 0 \,. \qquad \ldots (9.3.9)$$

Thus from (9.3.5),

$$\operatorname*{plim}_{n\to\infty} \hat{\beta}^{+}_{GLS} = \operatorname*{plim}_{n\to\infty} \left(\frac{1}{n} X^{+\prime} V^{-1} X^{+} \right)^{-1} \operatorname*{plim}_{n\to\infty} \left(\frac{1}{n} X^{+\prime} V^{-1} X \right) \beta \quad \text{(using (9.3.9))}$$

$$= E^{-1} C \beta$$

$$\neq P_{\infty} \beta \quad , \text{ in general,}$$

where $P_{\infty} = \lim_{n\to\infty} P = \lim_{n\to\infty} \left[\{ E(X^{+\prime} X^{+}) \}^{-1} E(X^{+\prime} X) \right]$.

Thus it is seen that an application of GLS in a misspecified model with known variance-covariance matrix of the disturbances will usually yield both biased and inconsistent estimates of the regression

coefficients of the misspecified model.[5] Even if $V_\varepsilon = I$, the above result still remains true except that now D has to be defined as

$$D = \lim_{n\to\infty} E\left(\frac{1}{n} X^{+\prime} V^{-1} V^{-1} X^{+}\right).$$

If, however, the regressions of the excluded regressors on the included ones are strictly linear then

$$E(Z \mid X^{+}) = 0 \text{ and hence } E(Z) = 0.$$

Obviously we then have,

$$E(\hat{\beta}^{+}_{GLS}) = \beta^{+}.$$

5/ We may, however, define the coefficient vector β^{+} in a different manner, by minimizing

$$E\{(Y - X^{+}\beta^{+})' V^{-1} (Y - X^{+}\beta^{+})\}$$

instead of minimizing

$$E\{(Y - X^{+}\beta^{+})'(Y - X^{+}\beta^{+})\}.$$

We then get

$$\beta^{+} = \bar{P}\beta$$

where $\bar{P} = \{E(X^{+\prime}V^{-1}X^{+})\}^{-1} E(X^{+\prime}V^{-1}X)$.

With this definition of β^{+}, and hence of $Z(=(X - X^{+}\bar{P})\beta)$, it can be shown in a fashion similar to that used in proving the result in (9.3.1) that

$$\operatorname{plim}_{n\to\infty} \left(\frac{1}{n} X^{+\prime} V^{-1} Z\right) = 0.$$

GLS would then yield consistent estimate of β^{+} thus defined. However, since in our Monte Carlo study we do not need the explicit definition of P (or $\bar{P}$) and β^{+} and since the point as to what should be the exact definition of β^{+} needs further research, we continue defining β^{+} as we have done so far following Chaudhuri.

We also get

$$E\left(\frac{1}{n} X^{+\prime} V^{-1} \varepsilon^{+}\right) = 0$$

and

$$\lim_{n \to \infty} V\left(\frac{1}{n} X^{+\prime} V^{-1} \varepsilon^{+}\right) = 0 \quad \text{(under the assumption in (9.3.8).}$$

and hence by Chebychev's inequality,

$$\operatorname{plim}_{n \to \infty} \left(\frac{1}{n} X^{+\prime} V^{-1} \varepsilon^{+}\right) = 0.$$

Then we easily find from (9.3.4) that

$$\operatorname{plim}_{n \to \infty} \left(\hat{\beta}^{+}_{GLS}\right) = \beta^{+}.$$

So under the assumption that

$$E(Z \mid X^{+}) = 0$$

GLS will give both unbiased and consistent estimates of β^{+}.

III. Estimated Generalized Least Squares (EGLS) Method

The above results are derived under the assumption that V is known a priori. Typically, however, V will not be known beforehand and hence the GLS estimate specified above can not be obtained.

Let us now examine the properties of the estimator of the regression coefficients when we work with an estimated V, say $\hat{V}$. Under the assumption that $\hat{V}$ is a consistent estimator of V, we have

$$\operatorname{plim}_{n \to \infty} \left(\frac{1}{n} X^{+\prime} \hat{V}^{-1} X\right) = \operatorname{plim}_{n \to \infty} \left(\frac{1}{n} X^{+\prime} V^{-1} X\right) = C \quad \text{(exists by (9.3.6))}, \ \ldots \ (9.3.10)$$

Now since $\lim_{n \to \infty} E\left(\frac{1}{n} \sum_{t=1}^{n} x_{ht}^{2}\right)$ exists by our assumption that

$$\operatorname{plim}_{n \to \infty} \left(\frac{1}{n} X' X\right) = \Sigma_{XX}, \quad \text{a p. d. matrix,}$$

where x_{ht} is the t-th element of the h-th row of X', it can be shown that

$$\operatorname*{plim}_{n\to\infty} \left(\frac{1}{\sqrt{n}} X^{+/} (\hat{V}^{-1} - V^{-1})\varepsilon\right) = 0 \qquad \ldots (9.3.11)$$

[vide Chaudhuri (1979, pp. 321-323)].

We further note that

$$E\left(\frac{1}{\sqrt{n}} X^{+/} V^{-1} \varepsilon\right) = 0$$

and

$$\lim_{n\to\infty} V\left(\frac{1}{\sqrt{n}} X^{+/} V^{-1} \varepsilon\right) = \sigma_\varepsilon^2 D .$$

Then by Chebyshev's inequality

$$\operatorname*{plim}_{n\to\infty} \left(\frac{1}{n} X^{+/} V^{-1} \varepsilon\right) = 0$$

and therefore from (9.3.11)

$$\operatorname*{plim}_{n\to\infty} \left(\frac{1}{n} X^{+/} \hat{V}^{-1} \varepsilon\right) = 0 . \qquad \ldots (9.3.12).$$

Now, the EGLS estimator for β^+ which is defined as

$$\hat{\beta}^+_{EGLS} = (X^{+/} \hat{V}^{-1} X^+)^{-1} X^{+/} \hat{V}^{-1} Y$$

can easily be reduced to (as in the case of GLS)

$$\hat{\beta}^+_{EGLS} = (X^{+/} \hat{V}^{-1} X^+)^{-1} X^{+/} \hat{V}^{-1} \varepsilon + (X^{+/} \hat{V}^{-1} X^+)^{-1} X^{+/} \hat{V}^{-1} X\beta . \qquad \ldots (9.3.13)$$

Thus using the assumptions (9.3.6) -(9.3.8) and the results in (9.3.10) and (9.3.12), it now follows from (9.3.13) that

$$\operatorname*{plim}_{n\to\infty} \hat{\beta}^+_{EGLS} = E^{-1} C\beta \neq P_\infty \beta \text{ , in general.}$$

It is thus seen that even if we have a consistent estimator of V, β^+ is not, in general, consistently estimated by the EGLS method.

Let us now consider the special case where $E(Z \mid X^+) = 0$. In addition to assumptions (9.3.6), (9.3.7) and (9.3.8), we assume that

$$\lim_{n\to\infty} E\left(\frac{1}{n} X^{+/} V^{-1} V_Z V^{-1} X^+\right) = F \text{ (positive definite)} \quad \ldots(9.3.14)$$

where $V_Z = V(Z) = E(ZZ')$ (excepting the constant of multiplicity).

Since

$$X^{+/} \hat{V}^{-1} Z = X^{+/} \hat{V}^{-1}(X - X^+ P)\beta$$

$$= (X^{+/} \hat{V}^{-1} X - X^{+/} \hat{V}^{-1} X^+ P)\beta$$

we have, by using (9.3.6), (9.3.8) and (9.3.10)

$$\operatorname*{plim}_{n\to\infty} \left(\frac{1}{n} X^{+/} \hat{V}^{-1} Z\right) = \operatorname*{plim}_{n\to\infty} \left(\frac{1}{n} X^{+/} V^{-1} Z\right). \quad \ldots (9.3.15)$$

Again,

$$\operatorname*{plim}_{n\to\infty} \left(\frac{1}{n} X^{+/} V^{-1} Z\right) = 0 \quad \ldots (9.3.16)$$

because

$$E\left(\frac{1}{n} X^{+/} V^{-1} Z\right) = 0$$

and

$$\lim_{n\to\infty} V\left(\frac{1}{n} X^{+/} V^{-1} Z\right) = 0 \quad \text{(using (9.3.14))}.$$

Therefore, we have from (9.3.15), by using (9.3.16)

$$\operatorname*{plim}_{n\to\infty} \left(\frac{1}{n} X^{+/} \hat{V}^{-1} Z\right) = 0. \quad \ldots (9.3.17)$$

Now,

$$\hat{\beta}^+_{EGLS} = (X^{+/} \hat{V}^{-1} X^+)^{-1} X^{+/} \hat{V}^{-1} Y$$

$$= \beta^+ + (X^{+/} \hat{V}^{-1} X^+)^{-1} X^{+/} \hat{V}^{-1} \varepsilon + (X^{+/} \hat{V}^{-1} X^+)^{-1} X^{+/} \hat{V}^{-1} Z.$$

Using (9.3.17) and $\operatorname{plim}_{n\to\infty} \left(\frac{1}{n} X^{+} \hat{V}^{-1} \varepsilon \right) = 0$ (i.e., (9.3.12), we thus have

$$\operatorname{plim}_{n\to\infty} \hat{\beta}^{+}_{EGLS} = \beta^{+}.$$

Thus if we have a consistent estimator for V and if $E(Z \mid X^{+}) = 0$, then EGLS utilizing $\hat{V}$ in place of V consistently estimates β^{+}.

IV. The Two-Step Reestimation Methods

In all the two-step methods and in the search procedure suggested by Hildreth and Lu (1960), the disturbance is assumed to follow AR(1) process which means that the variance-covariance matrix V, excepting the constant of multiplicity, has the following structure

$$V = \begin{pmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{n-1} \\ \rho & 1 & \rho & \cdots & \rho^{n-2} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \cdots & 1 \end{pmatrix}$$

where ρ is the first-order autocorrelation coefficient of the AR(1) process. However, as we have already pointed out (quoting Chaudhuri) in Chapter 5, if there is any misspecification, then irrespective of whether the true disturbances are autocorrelated or not, the disturbances of the misspecified equation will not, in general, follow an AR(1) process. Of course, one does not know beforehand whether the model has been misspedified or not, and therefore one usually reestimates the regression coefficients β^{+} by using standard reestimation methods under the assumption that the disturbances of the misspecified equation follow an AR(1) process. For the sake of completeness, we give below a description

of different two-step and search methods in the context of a misspecified regression equation where the error term is assumed to follow an AR(1) process.

(a) Cochrane and Orcutt (CO) method : The first step of this method suggested by Cochrane and Orcutt (1949) is to estimate ρ from

$$\hat{\rho} = \frac{\sum_{t=2}^{n} e_t^+ e_{t-1}^+}{\sum_{t=1}^{n} e_t^{+2}} \qquad \ldots \quad (9.3.18)$$

where $e_t^+ = Y_t - x_t' \hat{\beta}_{OLS}^+$ and $x_t' = (x_{1t}, x_{2t}, \ldots, x_{k+t})$, $t = 1,2,\ldots,n$.

In the second step OLS is applied on the new variables defined as

$$\left.\begin{aligned} x_{it}^* &= x_{it} - \rho\, x_{i(t-1)} \\ \text{and } y_t^* &= y_t - \rho\, y_{t-1} \end{aligned}\right\} \quad \text{for } i = 1, 2, \ldots, k^+ \text{ and } t = 2, 3, \ldots, n$$

i.e., $\hat{\beta}_{CO}^+$, the estimator of β^+ by CO method, is obtained as

$$\hat{\beta}_{CO}^+ = (X^{*\prime} X^*)^{-1} X^{*\prime} Y^* \qquad \ldots \quad (9.3.19)$$

where

$$X^* = \begin{pmatrix} x_{12}^* & x_{22}^* & \ldots & x_{k+2}^* \\ x_{13}^* & x_{23}^* & \ldots & x_{k+3}^* \\ \ldots & \ldots & \ldots & \ldots \\ x_{1n}^* & x_{2n}^* & \ldots & x_{k+n}^* \end{pmatrix}$$

and

$$Y^* = (y_2^*, y_3^*, \ldots, y_n^*)'.$$

It is obvious from our discussion on the estimated GLS method (and it has been explicitly shown by Chaudhuri (1979)) that except for the special case where $E(Z \mid X^+) = 0$, $\hat{\beta}^+_{CO}$ will not, in general, consistently estimate β^+.

(b) Durbin's two-stop (D 2) process : Durbin (1960) suggested taking the OLS estimate of the regression coefficient of y_{t-1} in the following regression equation

$$y_t = \rho y_{t-1} + \beta_1^+ (x_{1t} - \rho x_{1(t-1)}) + \beta_2^+ (x_{2t} - \rho x_{2(t-1)})$$

$$+ \cdots\cdots + \beta_{k^+} (x_{k^+t} - \rho x_{k^+(t-1)}) + w_t \qquad (t = 2, 3, \ldots, n)$$

where $w_t = \varepsilon_t^+ - \rho \varepsilon_{t-1}^+$ and w_t's are uncorrelated with zero mean and variance σ_w^2,

as the estimate of ρ. Let $\hat{\hat{\rho}}$ denote this estimate. Then the Durbin's estimator of β^+, say $\hat{\beta}^+_{D2}$, is given by (9.3.19) with $\hat{\rho}$ replaced by $\hat{\hat{\rho}}$. As in the case of the CO estimator, $\hat{\beta}^+_{D2}$ will also not, in general, consistently estimate β^+, excepting for the special case where $E(Z \mid X^+) = 0$. Furthermore, as Chaudhuri (1977, 1979) has shown $\hat{\hat{\rho}}$ will not, in general, consistently estimate ρ.

(c) Prais and Winsten (PW) method : Since both CO and D2 methods ignore the first observation, neither of these methods would, strictly speaking, provide estimators which are identical with those obtained by the EGLS procedure with V replaced by $\hat{V}$. The matrix of the transformation used by these two methods is, in fact, a $(n-1) \times n$ matrix given by

$$\begin{pmatrix} -\rho & 1 & 0 & \cdots & 0 & 0 \\ 0 & -\rho & 1 & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & -\rho & 1 \end{pmatrix}.$$

Prais and Winsten (1954) [vide Kadiyala (1968)] suggested the following transformation matrix of order n x n given by

$$\begin{pmatrix} \sqrt{1-\rho^2} & 0 & 0 & \cdots & 0 & 0 \\ -\rho & 1 & 0 & \cdots & 0 & 0 \\ 0 & -\rho & 1 & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & -\rho & 1 \end{pmatrix}.$$

It is well-known that the use of this transformation matrix will make the two-step methods exactly equivalent to the EGLS procedure. Now in the PW method ρ is first estimated by $\hat{\rho}$ using (9.3.18) and then the regression coefficients are estimated as

$$\hat{\beta}^{+}_{PW} = (X^{**\prime} X^{**})^{-1} X^{**\prime} Y^{**} \qquad \ldots (9.3.20)$$

where

$$X^{**} = \begin{pmatrix} \sqrt{1-\hat{\rho}^2}\, x_{11} & \sqrt{1-\hat{\rho}^2}\, x_{21} & \cdots & \sqrt{1-\hat{\rho}^2}\, x_{k+1} \\ x^{*}_{12} & x^{*}_{22} & \cdots & x^{*}_{k+2} \\ \cdots & \cdots & \cdots & \cdots \\ x^{*}_{1n} & x^{*}_{2n} & \cdots & x^{*}_{k+n} \end{pmatrix}$$

$$Y^{**} = (\sqrt{1-\hat{\rho}^2}\, y_1, \; y^{*}_2, \; \ldots\ldots, \; y^{*}_n)'$$

and x^{*}_{ij}'s $(i = 1,2,..,k^{+};\ j = 2,3,.,n)$ and y^{*}_{i}'s $(i = 2,3,.., n)$

are defined as in the CO method. For this method also, Chaudhuri (1977,1979) showed that excepting for the special case where

$$E(Z \mid X^+) = 0,$$

$\hat{\beta}^+_{PW}$ does not, in general, consistently estimate β^+. All the above methods (a) - (c) can, in principle, be iterated in small samples.

(d) Hildreth and Lu (HL) search procedure : Hildreth and Lu suggested (1960) a search procedure which can be used to estimate β^+. According to this method, one selects a number of values of ρ over the interval -1 and + 1 and for each value of ρ, β^+ is estimated by (9.3.19) and the corresponding residual sum of squares is calculated. The required estimate of ρ and hence of β^+ are those corresponding to which the residual sum of squares is the smallest. Naturally, in order to economise both time and money, we first search over a wider intervals of ρ and then over smaller intervals around the proper value of ρ. As this is GLS method with ρ known, it follows from our discussion that excepting for the case with $E(Z \mid X^+) = 0$, $\hat{\beta}^+_{HL}$, the estimate of β^+ obtained by this method will not be consistent.

V. Maximum Likelihood (ML) Method

The estimators obtained by the above methods (a) to (d) are often referred to as ML estimators under the assumption of normality [see, e.g., Cochrane and Orcutt (1949), Hildreth and Lu (1960), and Kmenta (1971)]. However, since CO, D2 and HL estimators are based upon the joint density of n-1 observations and not on all the n observations, these methods can be considered as equivalent to the ML procedure conditional upon the

first observation. PW method is based on all the n observations; yet it is not, strictly speaking, equivalent to the ML method. This is because, under the assumption of normality, the log-likelihood function based upon n observations and with the assumption that

$$E(\varepsilon^+ \mid X^+) = 0$$

is for a given X^+, 6/

$$L = -\frac{n}{2} \ln \sigma_W^2 + \frac{1}{2} \ln (1 - \rho^2) - \frac{1}{2\sigma_W^2}(Y^{**} - X^{**}\beta^+)'(Y^{**} - X^{**}\beta^+) \quad \ldots\ldots \quad (9.3.21)$$

where $\sigma_W^2 = \sigma^2(1 - \rho^2)$.

It may be seen from (9.3.21) that PW method is not fully equivalent to the ML method of estimation because the latter contains an additional term $\frac{1}{2}\ln(1 - \rho^2)$. In large samples, however, it is expected that this additional term will be dominated by the other terms if ρ is not too close to 1. Since $\frac{1}{2}\ln(1 - \rho^2)$ is independent of n, we thus find that asymptotically the PW method of estimation is equivalent to the ML method in the special case. It is also clear that the first observation will not asymptotically have much weightage and hence we can conclude that other reestimation methods like CO, D2 and HL are also asymptotically equivalent to the ML method of estimation in the special case. 7/

6/ This, in fact, corresponds to the special case where

$$E(Z \mid X^+) = 0.$$

In the general case

$$E(Z \mid X^+) \neq 0$$

and the ML estimates can not obviously be obtained.

7/ Hildreth and Dent (1974) and Beach and MacKinnon (1978a) have provided algorithms for obtaining ML estimates of the parameters.

9.4 Conclusions

In this Chapter we have examined the unbiasedness and consistency properties of the OLS, the GLS, the EGLS estimators in the context of a misspecified regression model. We have also briefly described the different conventional reestimation methods (also called two-step procedures) that are often used to efficiently estimate the parameters of an autocorrelated linear regression model where the error term follows an AR(1) process; the results regarding the consistency property of these estimators (which clearly are not the most efficient estimators because of improper assumption of the error process) in the presence of misspecification are also stated. However, we do not know how these reestimation methods perform in the presence of misspecification in small samples. We intend to study the performance of these methods along with the OLS in such cases. The results of such a Monte Carlo study are given in the next two chapters.

Chapter 10

Plan of the Monte Carlo Experiment

10.1 Introduction

In the last Chapter we discussed unbiasedness and consistency of the OLS, GLS and EGLS methods of estimation of regression coefficients in an autocorrelated linear regression model in the presence of misspecification. We also outlined the different standard reestimation methods viz., Cochrane-Orcutt (CO), Prais-Winsten (PW), Durbin's two-step (D2) and Hildreth-Lu (HL), used for efficient estimation of these parameters when the error term follows AR(1) process.[1/] Further, we quoted Chaudhuri's (1977, 1979) results regarding the consistency of the estimators given by these methods in the presence of misspecification.

In the present Chapter we shall describe the model and the design of the Monte Carlo experiment conducted by us in order to study the small-sample performance of different conventional reestimation methods in the presence of misspecification which may or may not contribute to the autocorrelation of the disturbance term of the misspecified equation.[2/] In this study we have considered two types of situations with regard to the

---

1/ It has already been stated in Chapters 5 and 6 that the error term of the misspecified equation does not, in general, follow AR(1) process even if each component does so. So, in the presence of misspecification, these methods do not remain most efficient any longer.

2/ Our study intends to examine the consequences of using the conventional reestimation methods which ignore the possibility of misspecification when, in fact, there is misspecification in small samples.

autocorrelation in the disturbances. As stated in previous Chapters, the disturbance term in a misspecified model has two components — one due to misspecification and the other, the disturbance term associated with the true regression equation. We consider two situations:(1)where one of them follows AR(1) process while the other component is present but non-autocorrelated, and (2) where both the components follow AR(1) process.

In section 10.2 we present the model and the assumptions. The design of the experiment including the choice of parameter values are described in section 10.3. While the method of generation of data is described in section 10.4, the criteria used for assessing the comparitive performance of the methods are stated in section 10.5. Conclusions are given in section 10.6.

10.2 The Model

For the purpose of the present study, we consider a two-regressor linear regression equation[3]

$$y_t = \beta_1 x_{1t} + \beta_2 x_{2t} + \varepsilon_t \quad ; \qquad t = 1, 2, \ldots, n \quad \ldots(10.2.1)$$

as the true relationship. It is assumed that both x_1 and x_2 are uncorrelated with ε. The misspecified equation, on the other hand, is taken to be the one where the second regressor x_2 has been omitted. Then following Chaudhuri and the approach adopted in Chapter 9 of this dissertation we redefine the regression coefficient[4] associated with x_1 so as to enable x_1 to capture as

---

3/ All the regressors are considered to be measured from their respective means and hence we do not explicitly include a constant term.

4/ Although we have followed Chaudhuri in presenting the model, the results of our experiment are independent of whether β^+ is redefined or not. Hence, Gupta and Massoumi's (1979) comment about the redefinition is not relevant here.

much of the effect of x_2 on y as possible. The misspecified model then can be written, following section 9.2 [cf. equations (9.2.3) to (9.2.6) and footnote 2 of Chapter 9], as

$$y_t = \beta^+ x_{1t} + \varepsilon_t^+ \quad \ldots \quad (10.2.2)$$

where

$$\beta^+ = \beta_1 + P_{21}\,\beta_2 \quad \ldots \quad (10.2.3)$$

$$\varepsilon_t^+ = \varepsilon_t + Z_t \quad \ldots \quad (10.2.4)$$

$$Z_t = \beta_2 W_t \quad \ldots \quad (10.2.5)$$

$$W_t = x_{2t} - P_{21} x_{1t} \quad \ldots \quad (10.2.6)$$

$$P_{21} = \{ E(x_1' x_1)\}^{-1} E(x_1' x_2) \quad \ldots \quad (10.2.7)$$

$$x_1' = (x_{11}, x_{12}, \ldots, x_{1n})$$

and $x_2' = (x_{21}, x_{22}, \ldots, x_{2n})$.

It is now assumed that ε and W (hence ε and Z) and also x_1 and W are uncorrelated.[5] Furthermore ε_t, W_t (and hence Z_t) and also x_{1t} are considered to have been generated by AR(1) processes given by

$$W_t = \delta W_{t-1} + u_t, \qquad |\delta| < 1 \quad \ldots \quad (10.2.8)$$

$$\varepsilon_t = \xi \varepsilon_{t-1} + v_t, \qquad |\xi| < 1 \quad \ldots \quad (10.2.9)$$

and $$x_{1t} = \eta x_{1t-1} + \theta_t, \qquad |\eta| < 1 \quad \ldots \quad (10.2.10)$$

where u_t's, v_t's and θ_t's are serially uncorrelated and distributed with zero mean and variances σ_u^2, σ_v^2 and σ_θ^2 respectively.

5/ The uncorrelatedness between ε and W, however, follows from the absence of correlation between ε and either x_1 or x_2.

We shall consider the following two types of situations :

Situation I : The autocorrelation in the disturbances ε_t^+ of the misspecified model is due to either misspecification (Z_t) or other reasons (ε_t) though both the components are present.[6/]

Situation II : The autocorrelation in ε_t^+ is due to autocorrelation in both ε_t and Z_t.[7/]

10.3 Design of the Experiment

After having described the model in (10.2.1) to (10.2.10) we now discuss the different aspects of the experiment including the generation of data, choice of parameter values etc.

10.3.1 Structure of the Variance-Covariance Matrix

We may mention that in estimating a regression model one usually tests for autocorrelation among the disturbances unless there are a priori reasons to believe that there is no such autocorrelation. If the null hypothesis of zero autocorrelation is found to be tenable one applies OLS to obtain best linear unbiased estimates of the parameters. If, however, there is indication of significant autocorrelation, then the common practice, at least until recently, has been to assume that the disturbances

---

6/ Obviously, the autocorrelation coefficient for the other process is then zero.

7/ For the sake of computational convenience, η in (10.2.10) was put equal to zero for Situation II. This relaxation was done partly because none of the methods of estimation was found to be particularly sensitive to η for Situation I.

(i.e., ε_t^+'s in our model) follow an AR(1) process i.e.,

$$\varepsilon_t^+ = \rho_o \varepsilon_{t-1}^+ + m_t \qquad \ldots (10.3.1)$$

where $|\rho_o| < 1$ and m_t's are serially uncorrelated and distributed with zero mean and constant variance σ_m^2, and then to use one or the other reestimation method to obtain efficient estimates of the parameters. An application of these methods require the use of the variance-covariance matrix of the disturbances. Under the assumed error process in (10.3.1), the variance-covariance matrix is given by

$$V_o = \frac{\sigma_m^2}{1-\rho_o^2} \begin{pmatrix} 1 & \rho_o & \rho_o^2 & \ldots & \rho_o^{n-1} \\ \rho_o & 1 & \rho_o & \ldots & \rho_o^{n-2} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ \rho_o^{n-1} & \rho_o^{n-2} & \rho_o^{n-3} & \ldots & \end{pmatrix} \qquad \ldots (10.3.2)$$

However, as we have already noted in section 6.3 of this dissertation, ε_t^+ defined as $\varepsilon_t + Z_t$ cannot, in general, follow an AR(1) process, even if both ε_t and Z_t are generated by AR(1) processes or if one of them follows an AR(1) process while the other is present but non-autocorrelated. It is easy to see that, when W_t's (hence Z_t's) and ε_t's satisfy (10.2.8) and (10.2.9) respectively, the first order autocorrelation coefficient of ε_t^+ — denoted by ρ_o for the sake of simplicity — is given by

$$\rho_o = \frac{\xi\sigma_\varepsilon^2 + \delta\sigma_Z^2}{\sigma_\varepsilon^2 + \sigma_Z^2} . \qquad \ldots (10.3.3)$$

The variance-covariance matrix given in (10.3.2) with ρ_o defined in (10.3.3) is therefore, under the assumed conditions, not the proper matrix.

The proper variance-covariance matrix (V) is, in fact, given by

$$V = \begin{pmatrix} \sigma_\varepsilon^2 + \sigma_Z^2 & \xi\sigma_\varepsilon^2 + \delta\sigma_Z^2 & \cdots & \xi^{n-1}\sigma_\varepsilon^2 + \delta^{n-1}\sigma_Z^2 \\ \xi\sigma_\varepsilon^2 + \delta\sigma_Z^2 & \sigma_\varepsilon^2 + \sigma_Z^2 & \cdots & \xi^{n-2}\sigma_\varepsilon^2 + \delta^{n-2}\sigma_Z^2 \\ \cdots\cdots & \cdots\cdots & \cdots & \cdots\cdots\cdots \\ \xi^{n-1}\sigma_\varepsilon^2+\delta^{n-1}\sigma_Z^2 & \xi^{n-2}\sigma_\varepsilon^2+\delta^{n-2}\sigma_Z^2 & \cdots & \sigma_\varepsilon^2+\sigma_Z^2 \end{pmatrix} \quad \ldots(10.3.4)$$

The effect of working with a wrong variance-covariance matrix is expected to be reflected in the efficiency of the estimates given by the methods.

10.3.2 Choice of Parameter Values

We are interested in this study to compare the performance of five different methods of estimation viz., ordinary least squares (OLS), Cochrane-Orcutt (CO), Prais-Winsten (PW) Durbin's two-step (D2) and Hildreth and Lu (HL) search procedure.[8/]

In order to explain the choice of the values of the parameters involved, it would be convenient to first present the expressions for the different estimators. It should be noted that excepting for the OLS, all the other methods are either GLS or EGLS or approximate EGLS methods of estimation. So denoting by $\hat{\beta}^+_{OLS}$ and $\hat{\beta}^+_{GLS}$ the OLS and the GLS estimators of β^+ respectively, we have

$$\hat{\beta}^+_{OLS} = \left(\sum_{t=1}^{n} x_{1t}^2 \right)^{-1} \sum_{t=1}^{n} x_{1t}\, y_{1t} \qquad \ldots (10.3.5)$$

$$\hat{\beta}^+_{GLS} = (x_1' V_o^{-1} x_1)^{-1} x_1' V_o^{-1} Y \qquad \ldots (10.3.6)$$

where $Y = (y_1, y_2, \ldots, y_n)'$.

8/ As already mentioned in section 8.2, we do not consider for the purpose of this study either the ML method or the methods developed by us in Chapters 6 and 7.

and

$$V(\hat{\beta}^{+}_{OLS}) = E\left[\frac{1}{\left(\sum_{t=1}^{n} x_{1t}^2\right)} \left(\sum_{t=1}^{n} x_{1t}\ e_t^{+}\right)\right]^2$$

which, as shown in Appendix 10.1, can be approximated for a given large sample as

$$V(\hat{\beta}^{+}_{OLS}) \simeq \frac{1}{\sum_{t=1}^{n} x_{1t}^2}\left[\sigma_\varepsilon^2\ \frac{1+\eta\xi}{1-\eta\xi} + \beta_2^2\,\sigma_W^2\ \frac{1+\eta\delta}{1-\eta\delta}\right]. \quad \ldots (10.3.7)$$

Also,

$$V(\hat{\beta}^{+}_{GLS}) = (x_1' V_o^{-1} x_1)^{-1}\ x_1' V_o^{-1} V V_o^{-1} x_1\ (x_1' V_o^{-1} x_1)^{-1} \quad \ldots (10.3.8)$$

which cannot be easily approximated by simple expressions even for a large sample because of difficulty in evaluating the expression

$$(x_1' V_o^{-1} V V_o^{-1} x_1).$$

In obtaining estimates of the regression coefficients by these methods one also needs an estimatorof ρ_o. The estimator of ρ_o used in some of these methods is based on OLS residuals and is given by

$$\hat{\rho}_o = \frac{\sum_{t=2}^{n} e_t^{+}\ e_{t-1}^{+}}{\sum_{t=1}^{n} e_t^{+2}}$$

where

$$e_t^{+} = y_t - \hat{\beta}^{+}_{OLS}\ x_{1t}.$$

Again, as shown in Appendix 10.1, we have

$$E(\hat{\rho}_o) \simeq \frac{\frac{\sigma_\varepsilon^2}{\beta_2^2 \sigma_W^2}\left[\xi(n-1) - \frac{2(\eta+\xi)}{1-\eta\xi} + \eta\,\frac{1+\eta\xi}{1-\eta\xi}\right] + \left[\delta\,(n-1) - \frac{2(\eta+\delta)}{1-\eta\delta} + \eta\,\frac{1+\eta\delta}{1-\eta\delta}\right]}{\frac{\sigma_\varepsilon^2}{\beta_2^2 \sigma_W^2}\left[n - \frac{1+\eta\xi}{1-\eta\xi}\right] + \left[n - \frac{1+\eta\delta}{1-\eta\delta}\right]} \quad \ldots\ldots (10.3.9)$$

and hence $\hat{\rho}_o$ is a biased estimator of ρ_o, the magnitude of bias being dependent on n and the following parameters: η, δ, ξ, $\sigma_\varepsilon^2 / \beta_2^2 \sigma_W^2$.

Let us now discuss the actual choice of values of the parameters involved in the model being considered here. We may mention, first of all, that in our study we considered samples of size n=15 only.[9/] Also, the values of β_2 and β^+ were fixed at 2.0 and 3.0 respectively. It may next be noted that of the 16 parameters, viz., ρ_o, β_1, β_2, β^+, σ_ε^2, σ_θ^2, σ_u^2, σ_v^2, σ_W^2, $\sigma_{x_1}^2$, σ_Z^2, σ_m^2, η, δ, ξ and P_{21}, some are not independent. It is easy to see that given the values of δ, σ_W^2, $\sigma_{x_1}^2$, η, ξ, σ_ε^2 and β_2, one can find values of $\sigma_Z^2 = \beta_2^2 \sigma_W^2$, ρ_o (which is dependent on δ, ξ, σ_Z^2 and σ_ε^2), $\sigma_u^2 = \sigma_W^2 (1-\delta^2)$, $\sigma_\theta^2 = \sigma_{x_1}^2 (1-\eta^2)$, $\sigma_v^2 = \sigma_\varepsilon^2 (1-\xi^2)$ and $\sigma_m^2 = (1-\rho_o^2)(\sigma_\varepsilon^2 + \sigma_Z^2)$.[10/]

9/ Sample size of as low as 15 lead to enormous volume of computations and results to be analysed and hence samples of any bigger size were not considered.

10/ It may be seen that we do not explicitly require values of P_{21} and β_1. But, as can be seen from the relation in (10.2.3), it follows that $\beta_1 + 2P_{21} = 3$ and hence there is a restriction on the values that these parameters can assume. In fact, P_{21} will have a fixed value for a given population though it need not be explicitly mentioned for our study.

To identify further the set of independent parameters whose values have to be fixed at different levels, we examine the expressions for

$$V_o,\ V,\ E(\hat{\rho}_o),\ V(\hat{\beta}^{+}_{OLS}) \text{ and } V(\hat{\beta}^{+}_{GLS}),$$

since the estimators (excepting the OLS) involve V_o and $\hat{\rho}_o$ and their true sampling variancesinvolve V also. It may be noticed from (10.3.7)-(10.3.9) that the expressions for

$$E(\hat{\rho}_o),\ V(\hat{\beta}^{+}_{OLS}) \text{ and } V(\hat{\beta}^{+}_{GLS})$$

involve three ratios of variances viz.,

$$\sigma_\varepsilon^2 / \sigma_{x_1}^2,\quad \beta_2^2 \sigma_W^2 / \sigma_{x_1}^2 \text{ and } \sigma_\varepsilon^2 / \beta_2^2 \sigma_W^2$$

rather than absolute values of

$$\sigma_\varepsilon^2,\ \beta_2^2 \sigma_W^2 \text{ and } \sigma_{x_1}^2 \left(\simeq \frac{1}{n} \sum_{t=1}^{n} x_{1t}^2 \right).$$ 11/

So we decided to fix $\sigma_{x_1}^2$ at 1.0 throughout our study. With these chosen values of some of the parameters, we now find that for the two different types of autocorrelated situations proposed to be considered here, the

---

11/ We have already noted that even in large samples it is quite cumbersome to approximate the expression for $V(\hat{\beta}^{+}_{GLS})$. We may, however, see from Appendix 10.1 that

$$(x_1' V_o^{-1} x_1)^{-1} \simeq \frac{(\sigma_\varepsilon^2 + \beta_2^2 \sigma_W^2)}{\sum_{t=1}^{n} x_{1t}^2} \cdot \frac{(1 - \rho_o^2)}{(1 + \rho_o^2 - 2\eta \rho_o)}$$

which involves the ratios of variances viz.,

$$\sigma_\varepsilon^2 / \sigma_{x_1}^2 \text{ and } \beta_2^2 \sigma_W^2 / \sigma_{x_1}^2$$

(where $\sigma_{x_1}^2 \simeq \frac{1}{n} \sum_{t=1}^{n} x_{1t}^2$),

apart from n, η and ρ_o. Looking at $(x_1' V_o^{-1} V V_o^{-1} x_1)$ which involves V_o and V (which in turn involve $\rho_o, \delta, \xi, \sigma_\varepsilon^2$ and $\beta_2^2 \sigma_W^2$), we can come to the stated conclusion.

set of parameters whose values should be allowed to vary over their possible ranges are the following :

Situation I : δ or ξ, η, σ_ε^2, σ_W^2

and

Situation II : δ , ξ, σ_ε^2, σ_W^2 .

It is known from the stationarity conditions that each of the parameters δ, ξ and η lies in the interval -1 to +1 since negative autocorrelation is unlikely to occur. We choose three values for each of these parameters i.e., 0.2, 0.6 and 0.95 in this interval. Problem, however, arises in fixing the values of σ_ε^2 and σ_W^2 since, by definition, these parameters can take any non-negative values. Therefore, instead of choosing values of these variances directly, we decided to base this choice on the relationships, presented in Appendix 10.2, between these variances and the total, multiple and partial correlation coefficients relevant for the model. This way of obtaining values of σ_ε^2 and σ_W^2 has an additional advantage in that the effect of misspecification on the properties of the different estimators can then be studied by alternating values of the correlation coefficients (e.g., very high $R^2_{y.x_1x_2}$ and low $r^2_{yx_1}$ means that here has been serious misspecification). In fact, effect of misspecification on the estimators have been studied in this way.

Now the different correlation coefficients which need to be considered for this purpose are the (squared) multiple correlation coefficient (R^2), the simple (squared) correlation coefficient between y and x_1 (r^2), the (squared) partial correlation coefficient between y and x_2 when x_1 is held fixed ($r^2_{y2.1}$) and the (squared) correlation coefficient between x_1 and x_2 (r^2_{12}). It is known, however, that the three correlations

R^2, r^2 and $r^2_{y2.1}$ are connected by the relationship

$$r^2_{y2.1} = \frac{R^2 - r^2}{1 - r^2}$$

and hence it is enough to choose any two of the three. We decided to fix values for R^2 and r^2 which automatically determines the values for $r^2_{y2.1}$. Also, since we need only two relations to determine values of σ^2_W and σ^2_ε, we decided to ignore r^2_{12}. [12/] We, therefore, obtain the values of σ^2_W and σ^2_ε by using the following two relations derived in Appendix 10.2:

$$R^2 = \frac{1}{1 + \dfrac{1}{\dfrac{\beta^{+2}\sigma^2_{x_1}}{\sigma^2_\varepsilon} + \dfrac{\beta_2^2 \sigma^2_W}{\sigma^2_\varepsilon}}} \qquad \ldots (10.3.10)$$

and

$$r^2 = \frac{1}{1 + \dfrac{\beta_2^2 \sigma^2_W}{\beta^{+2}\sigma^2_{x_1}} + \dfrac{\sigma^2_\varepsilon}{\beta^{+2}\sigma^2_{x_1}}} \qquad \ldots (10.3.11)$$

Then by substituting the values of $\beta^+ = 3.0$, $\beta_2 = 2.0$ and $\sigma^2_{x_1} = 1.0$, we easily find that

$$\sigma^2_\varepsilon = 9(1 - R^2)/r^2 \qquad \ldots (10.3.12)$$

and

$$\sigma^2_W = 9(R^2 - r^2)/4r^2 \quad . \qquad \ldots (10.3.13)$$

We considered three different values of R^2 and for each of these values of R^2, three values of r^2 were taken. The nine pairs of values are as follows :

12/ See, however, the next paragraph of this section.

values of R^2	values of r^2		
0.4	0.1	0.25	0.35
0.7	0.1	0.35	0.65
0.95	0.1	0.65	0.90

The values of σ^2_ε and σ^2_W were worked out by using the relationships mentioned in (10.3.12) and (10.3.13).

We may finally note that we have not explicitly considered variations in r^2_{12}. This is because in our formulation we have

$$x_{2t} = P_{21}\, x_{1t} + W_t \qquad (t = 1, 2, \ldots, 15) \text{(cf. (10.2.6))}$$

and the regression coefficient associated with the included regressor x_1 in the misspecified equation has been redefined so as to enable it to capture as much influence of x_2 on y as possible. Therefore, r^2_{12} is not of much importance in fixing the values of the parameters. One can see, however, that our formulation does allow for variation in r^2_{12} also. It is shown in Appendix 10.2 that

$$r^2_{12} = \frac{1}{1 + \dfrac{\sigma^2_W}{P^2_{21}\,\sigma^2_{x_1}}} \quad .$$

Then, since $\sigma^2_{x_1} = 1.0$ and as noted in footnote 9, P_{21} has a fixed value (though not explicitly mentioned), a change in the value of σ^2_W effected through variations in R^2 and r^2 implies change in the value of r^2_{12} also.

10.4 Generation of Data

It is clear from the formulation of the model and the discussions so far that we need to generate only three series of data — those for x_1, ε and W. Data for y can then be generated from (10.2.2), (10.2.4) and (10.2.5). This was done in the following manner :

(a) The first observation on x_1 i.e., x_{11} was generated from a normal population with zero mean and unit variance. This was done by drawing one observation from a N(0,1) population.[13/]

(b) The other observations on x_1 were obtained by using (10.2.10). At each stage a N(0, 1) value was drawn and then multiplied by $\sigma_\theta (= +\sqrt{\sigma_\theta^2})$. This product was then added to η times the previous value of x_1.

(c) The values of W_t and ε_t (t = 1, 2, ..., 15) were generated in a manner exactly similar to those of x_{1t}'s using (10.2.8) and (10.2.9) respectively.

(d) y_t's were then obtained by using (10.2.2), (10.2.4) and (10.2.5) and the series of values on x_1, ε and W already generated.

It may, however, be noted that in case of Situation I, one of ε_t and W_t (in fact, it was assumed to be ε_t) is non-autocorrelated and then its values were obtained by drawing the required number of random normal

---

13/ The observations from a N(0, 1) population were generated with a Borroughs 6700 computer system at the Regional Computer Centre, Jadavpur University, Calcutta, India. A description of the method of generation of the tape is given in Appendix 10.3. About 2,00,000 random normal deviates were generated, stored in a magnetic tape and then used as and when required. All the required computations were done with this computer system.

deviates and multiplying these values by the corresponding standard deviation. In Situation II, $\eta = 0$ and hence x_{1t}'s are independent N(0, 1) variates and thus obtained directly from the tape mentioned earlier.

10.5 Criteria Used For Assessing the Comparative Performance of the Methods

For the purpose of examining the performance of the different methods of estimation, 50 samples, each of size 15, were generated for each set of parameter values considered and the computations made for the different methods of estimation viz., OLS, CO, PW, D2 and HL. We describe below the different computations made with emphasis on criteria used in appraising the five methods of estimation :

(a) For each of the 50 samples, we computed the following :

(i) Estimate of the regression coefficient β^+ by each of the five methods[14/] denoted $\hat{\beta}^+_{ij}$ (i= 1,2,...., 5; j = 1, 2, ..., 50).

(ii) Sampling variances of the estimators $\hat{\beta}^+_{ij}$'s for each of the five methods using the conventional formulae for sampling variance.

(iii) Standard errors of $\hat{\beta}^+_{ij}$'s (i = 1, 2, ..., 5; j = 1, 2, ..., 50), i.e., the square roots of sampling variances in (ii).

(b) These 50 values obtained in each of a(i)– a(iii) for the five

---

14/ In case of HL, we restricted ρ_o to $-0.99 \leq \rho_o \leq 0.99$.

methods were then used to compute the following summary statistics :

(i) Estimate of bias defined as $\frac{1}{50} \sum_{j=1}^{50} (\hat{\beta}^{+}_{ij} - \beta^{+})$, $i = 1, 2, \ldots, 5$.

(ii) Estimate of mean square error (mse) defined as $\frac{1}{50} \sum_{j=1}^{50} (\hat{\beta}^{+}_{ij} - \beta^{+})^2$, $i = 1, 2, \ldots, 5$.

(iii) Estimate of variance defined as $\frac{1}{49} \sum_{j=1}^{50} (\hat{\beta}^{+}_{ij} - \bar{\hat{\beta}}^{+}_{i})^2$, $i = 1, 2, \ldots, 5$,

where $\bar{\hat{\beta}}^{+}_{i} = \frac{1}{50} \sum_{j=1}^{50} \hat{\beta}^{+}_{ij}$ $(i = 1, 2, \ldots, 5)$.

(c) Absolute deviation of each $\hat{\beta}^{+}_{ij}$ from β^{+} i.e.,

$$A_{ij} = \left| \hat{\beta}^{+}_{ij} - \beta^{+} \right| \quad (i = 1, 2, \ldots, 5 \text{ and } j = 1, 2, \ldots, 50)$$

was calculated. These 50 values for each of the methods were then used to rank the 5 methods according to the size of the deviation. The method with the least absolute deviation was given rank 1, the next smallest one rank 2, and so on. Thus we found how many times (out of 50 samples), a particular method stood first, second and so on. These were used to calculate the average rank of each method ($R_1, R_2, \ldots, R_5$) defined as

$$R_i = \frac{\sum_{k=1}^{5} k\, p_{ik}}{50}; \qquad i = 1, 2, \ldots, 5,$$

where p_{ik} is the number of samples where the i-th method occupied the k-th rank. These average ranks were used solely for descriptive purposes. For tests of significance we

relied mainly on pairwise comparisons of the methods. See (d) below.

(d) Using the A_{ij}'s ($i = 1, 2, \ldots, 5$; $j = 1, 2, \ldots, 50$) calculated in (c), we found out, how many times (out of 50 samples), any particular method gave estimates closer to β^+ than any other method. These counts for different pairs of methods were then used to carry out (non-parametric) sign test to infer about the comparative performance of the methods.

(e) For each of the 50 samples, the 95 per cent confidence interval for the regression coefficient β^+ were calculated for each of the five different methods. The number of times (out of 50 samples), the actual value of β^+ i.e., 3.0 fell outside the confidence intervals were then counted.

(f) For each method the average of 50 sampling variances computed by the usual formulae — see a(ii) above — was compared with the estimate of variance (considered as the true sampling variance) obtained in b(iii) to find if the sampling variance formula for the method overestimates, underestimates or correctly estimates the true sampling variance.[15]

10.6 Conclusions

In this Chapter we have set out in details the different aspects of the Monte Carlo experiment conducted by us. The results of the analysis are being reported in the next Chapter.

15/ Gupta and Maasoumi (1979, p.379), however, observed that "the OLS formula for the estimated variance matrix of the regression coefficients is more likely to underestimate the appropriate criterion of estimator reliability which is the Mean Square Errors Matrix".

Appendix 10.1

Derivation of Expressions of $V(\hat{\beta}^{+}_{OLS})$, $(x_1' V_o^{-1} x_1)^{-1}$ and $E(\hat{\rho}_o)$ for a Given Sample

Here we derive the expressions throwing light on the influence of different parameters on $V(\hat{\beta}^{+}_{OLS})$, $V(\hat{\beta}^{+}_{GLS})$ and $E(\hat{\rho}_o)$.

(i) Derivation of the relation in (10.3.7)

Proof

$$\left(\sum_{t=1}^{n} x_{1t}\varepsilon_t^{+} \right)^2 = \left(\sum_{t=1}^{n} x_{1t}\,\varepsilon_t + \sum_{t=1}^{n} x_{1t}\,Z_t \right)^2 .$$

Thus

$$E\left(\sum_{t=1}^{n} x_{1t}\varepsilon_t^{+} \right)^2 = E\left(\sum_{t=1}^{n} x_{1t}\,\varepsilon_t \right)^2 + E\left(\sum_{t=1}^{n} x_{1t}Z_t \right)^2 + 2E\left(\sum_{t=1}^{n} x_{1t}\,\varepsilon_t \right)\left(\sum_{t=1}^{n} x_{1t}\,Z_t \right).$$

Now, carrying out algebraic manipulations similar to those in Rao and Griliches (1969, pp.269–270) and noting that each of $|\delta|$, $|\eta|$ and $|\xi|$ are less than 1, we have, for fairly large samples

$$E\left(\sum_{t=1}^{n} x_{1t}\,\varepsilon_t \right)^2 \simeq \sigma_\varepsilon^2 \left(\frac{1+\eta\xi}{1-\eta\xi} \right) \sum_{t=1}^{n} x_{1t}^2$$

and

$$E\left(\sum_{t=1}^{n} x_{1t}\,Z_t \right)^2 \simeq \beta_2^2\,\sigma_W^2 \left(\frac{1+\eta\delta}{1-\eta\delta} \right) \sum_{t=1}^{n} x_{1t}^2 .$$

Since ε and Z are uncorrelated and $E(\varepsilon) = E(Z) = 0$, we have

$$E\left(\sum_{t=1}^{n} x_{1t}\,\varepsilon_t\right)\left(\sum_{t=1}^{n} x_{1t}\,Z_t\right) = 0 .$$

Thus from (10.3.5), $V(\hat{\beta}^{+}_{OLS})$ can, in large samples, be approximated as

$$V(\hat{\beta}^{+}_{OLS}) \simeq \frac{1}{\sum_{t=1}^{n} x_{1t}^2}\left[\sigma_\varepsilon^2 \frac{1+\eta\xi}{1-\eta\xi} + \beta_2^2\,\sigma_W^2\,\frac{1+\eta\delta}{1-\eta\delta}\right].$$

(ii) Approximate expression for $(x_1' V_o^{-1} x_1)^{-1}$

$$x_1' V_o^{-1} x_1 = \frac{1}{\sigma_m^2}(x_{11}\ x_{12}\ \cdots\ x_{1n})\begin{pmatrix} 1 & -\rho_o & 0 & \cdots & 0 & 0 \\ -\rho_o & (1+\rho_o^2) & -\rho_o & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & -\rho_o & 1 \end{pmatrix}\begin{pmatrix} x_{11} \\ x_{12} \\ \cdots \\ x_{1n} \end{pmatrix}$$

Now, carrying out algebraic manipulations similar to those in Rao and Griliches (1969, p. 254), we can approximate $x_1' V_o^{-1} x_1$ for large samples as

$$x_1' V_o^{-1} x_1 \simeq (1 + \rho_o^2 - 2\eta\rho_o)\frac{\sum_{t=1}^{n} x_{1t}^2}{\sigma_m^2} .$$

Thus

$$(x_1' V_o^{-1} x_1)^{-1} \simeq \frac{(\sigma_\varepsilon^2 + \beta_2^2\,\sigma_W^2)}{\sum_{t=1}^{n} x_{1t}^2}\ \frac{(1-\rho_o^2)}{(1+\rho_o^2-2\eta\rho_o)}$$

since

$$\sigma_m^2 = (1-\rho_o^2)(\sigma_\varepsilon^2 + \beta_2^2\ \sigma_W^2).$$

(iii) Derivation of the expression for $E(\hat{\rho}_o)$ in (10.3.9)

We know

$$\hat{\rho}_o = \frac{\sum_{t=2}^{n} e_t^+ e_{t-1}^+}{\sum_{t=1}^{n} e_t^{+2}}$$

where $e_t^+ = y_t - \hat{\beta}^+_{OLS}\, x_{1t}$.

The first term in the Taylor expansion of $E(\hat{\rho}_o)$ gives terms up to the order of $(1/n)$ — [vide Rao and Griliches (1969), p.256] —

$$E(\hat{\rho}_o) \simeq \frac{E\left(\sum_{t=2}^{n} e_t^+ e_{t-1}^+\right)}{E\left(\sum_{t=1}^{n} e_t^{+2}\right)}.$$

Now,

$$e_t^+ = \beta^+ x_{1t} + \varepsilon_t^+ - \left(\frac{\sum_{t=1}^{n} x_{1t} y_t}{\sum_{t=1}^{n} x_{1t}^2}\right) x_{1t} \quad \text{(from (10.2.2))}.$$

Thus

$$e_t^+ = \varepsilon_t^+ - x_{1t} \frac{\sum_{t=1}^{n} x_{1t}\, \varepsilon_t^+}{\sum_{t=1}^{n} x_{1t}^2} \quad \text{(from (10.2.2))}$$

and therefore,

$$\sum_{t=2}^{n} e_t^+ e_{t-1}^+ = \sum_{t=2}^{n} \varepsilon_t^+ \varepsilon_{t-1}^+ - \frac{\sum_{t=2}^{n} x_{1t} \varepsilon_{t-1}^+ \sum_{t=1}^{n} x_{1t} \varepsilon_t^+}{\sum_{t=1}^{n} x_{1t}^2} - \frac{\sum_{t=2}^{n} x_{1\overline{t-1}} \varepsilon_t^+ \sum_{t=1}^{n} x_{1t} \varepsilon_t^+}{\sum_{t=1}^{n} x_{1t}^2}$$

$$+ \frac{\sum_{t=2}^{n} x_{1t} x_{1\overline{t-1}}}{\left(\sum_{t=1}^{n} x_{1t}^2\right)^2} \left(\sum_{t=1}^{n} x_{1t} \varepsilon_t^+\right)^2 . \qquad \ldots (A\ 10.1.1)$$

Since ε and Z are uncorrelated and $E(\varepsilon) = E(Z) = 0$, we have from (A 10.1.1),

$$E\left(\sum_{t=2}^{n} e_t^+ e_{t-1}^+\right) = E\left[\sum_{t=2}^{n} \varepsilon_t \varepsilon_{t-1} - \frac{\sum_{t=2}^{n} x_{1t} \varepsilon_{t-1} \sum_{t=1}^{n} x_{1t} \varepsilon_t}{\sum_{t=1}^{n} x_{1t}^2}\right.$$

$$\left. - \frac{\sum_{t=2}^{n} x_{1\overline{t-1}} \varepsilon_t \sum_{t=1}^{n} x_{1t} \varepsilon_t}{\sum_{t=1}^{n} x_{1t}^2} + \frac{\sum_{t=2}^{n} x_{1t} x_{1\overline{t-1}} \left(\sum_{t=1}^{n} x_{1t} \varepsilon_t\right)^2}{\left(\sum_{t=1}^{n} x_{1t}^2\right)^2}\right]$$

$$+ E\left[\sum_{t=2}^{n} Z_t Z_{t-1} - \frac{\sum_{t=2}^{n} x_{1t} Z_{t-1} \sum_{t=1}^{n} x_{1t} Z_t}{\sum_{t=1}^{n} x_{1t}^2}\right.$$

$$- \frac{\sum_{t=2}^{n} x_{1\overline{t-1}} Z_t \sum_{t=1}^{n} x_{1t} Z_t}{\sum_{t=1}^{n} x_{1t}^2}$$

$$\left. + \frac{\sum_{t=2}^{n} x_{1t} x_{1\overline{t-1}} \left(\sum_{t=1}^{n} x_{1t} Z_t\right)^2}{\left(\sum_{t=1}^{n} x_{1t}^2\right)^2}\right] \qquad \ldots (A\ 10.1.2)$$

We simplify (A 10.1.2) in a manner similar to Rao and Griliches (1969, pp. 270-272) and find that the two expressions within the curly brackets in (A 10.1.2) are approximately equal to

$$\sigma_\varepsilon^2 \left[\xi(n-1) - \frac{2(\eta+\xi)}{1-\eta\xi} + \eta\,\frac{1+\eta\xi}{1-\eta\xi}\right] \qquad \ldots \quad (A\ 10.1.3)$$

and

$$\beta_2^2\,\sigma_W^2 \left[\delta(n-1) - \frac{2(\eta+\delta)}{1-\eta\delta} + \eta\,\frac{1+\eta\delta}{1-\eta\delta}\right] \qquad \ldots \quad (A\ 10.1.4)$$

respectively.

Thus we have from (A 10.1.2) - (A 10.1.4),

$$E\left(\sum_{t=2}^{n} e_t^+ e_{t-1}^+\right) \simeq \sigma_\varepsilon^2 \left[\xi(n-1) - \frac{2(\eta+\xi)}{1-\eta\xi} + \eta\,\frac{1+\eta\xi}{1-\eta\xi}\right] + \beta_2^2\,\sigma_W^2 \left[\delta(n-1) - \frac{2(\eta+\delta)}{1-\eta\delta} + \eta\,\frac{1+\eta\delta}{1-\eta\delta}\right]. \qquad \ldots \quad (A\ 10.1.5)$$

Now it can easily be seen from the previous results that

$$E\left(\sum_{t=1}^{n} e_t^{+2}\right) \simeq \sigma_\varepsilon^2 \left[n - \frac{1+\eta\xi}{1-\eta\xi}\right] + \beta_2^2\,\sigma_W^2 \left[n - \frac{1+\eta\delta}{1-\eta\delta}\right] \qquad \ldots \quad (A\ 10.1.6)$$

since $E\left(\sum_{t=1}^{n} \varepsilon_t^{+2}\right) = n\left(\sigma_\varepsilon^2 + \beta_2^2\,\sigma_W^2\right)$.

Thus from (A 10.1.5) and (A 10.1.6) we find

$$E(\hat{\rho}_o) \simeq \frac{\sigma_\varepsilon^2 \left[\xi(n-1) - \frac{2(\eta+\xi)}{1-\eta\xi} + \eta\frac{1+\eta\xi}{1-\eta\xi}\right] + \beta_2^2\,\sigma_W^2 \left[\delta(n-1) - \frac{2(\eta+\delta)}{1-\eta\delta} + \eta\frac{1+\eta\delta}{1-\eta\delta}\right]}{\sigma_\varepsilon^2 \left[n - \frac{1+\eta\xi}{1-\eta\xi}\right] + \beta_2^2\,\sigma_W^2 \left[n - \frac{1+\eta\delta}{1-\eta\delta}\right]}$$

$$= \frac{\frac{\sigma_\varepsilon^2}{\beta_2^2\,\sigma_W^2}\left[\xi(n-1) - \frac{2(\eta+\xi)}{1-\eta\xi} + \eta\frac{1+\eta\xi}{1-\eta\xi}\right] + \left[\delta(n-1) - \frac{2(\eta+\delta)}{1-\eta\delta} + \eta\,\frac{1+\eta\delta}{1-\eta\delta}\right]}{\frac{\sigma_\varepsilon^2}{\beta_2^2\,\sigma_W^2}\left[n - \frac{1+\eta\xi}{1-\eta\xi}\right] + \left[n - \frac{1+\eta\delta}{1-\eta\delta}\right]}$$

Appendix 10.2

Derivation of Expressions for Different Correlation Coefficients

We know[1] true

$$R^2 = 1 - \frac{V(\varepsilon)}{V(y)} .$$

Since

$$V(y) = V(\beta^{+} x_1 + \beta_2 W + \varepsilon)$$

$$= \beta^{+2} \sigma^2_{x_1} + \beta_2^2 \sigma_W^2 + \sigma_\varepsilon^2 ,$$

we have

$$R^2 = 1 - \frac{\sigma^2_\varepsilon}{\beta^{+2} \sigma^2_{x_1} + \beta_2^2 \sigma_W^2 + \sigma_\varepsilon^2}$$

$$= \frac{\beta^{+2} \sigma^2_{x_1} + \beta_2^2 \sigma_W^2}{\beta^{+2} \sigma^2_{x_1} + \beta_2^2 \sigma_W^2 + \sigma_\varepsilon^2} .$$

Thus

$$R^2 = \frac{1}{1 + \dfrac{\sigma^2_\varepsilon}{\beta^{+2} \sigma^2_{x_1} + \beta_2^2 \sigma_W^2}}$$

$$= \frac{1}{1 + \dfrac{1}{\dfrac{\beta^{+2} \sigma^2_{x_1}}{\sigma^2_\varepsilon} + \dfrac{\beta_2^2 \sigma_W^2}{\sigma^2_\varepsilon}}} . \qquad \ldots\ldots \text{(A 10.2.1)}$$

1/ In this Appendix the symbols y, ε, x_1 etc., denote the variables and not vectors of observations on them.

Now,

$$r^2 = \{ \mathrm{Cov}(x_1, y)\}^2 / V(x_1) V(y)$$

$$= \frac{\beta^{+2} \sigma^4_{x_1}}{\sigma^2_{x_1} (\beta^{+2} \sigma^2_{x_1} + \beta_2^2 \sigma^2_W + \sigma^2_\varepsilon)}$$

$$= \frac{1}{1 + \dfrac{\beta_2^2 \sigma^2_W}{\beta^{+2} \sigma^2_{x_1}} + \dfrac{\sigma^2_\varepsilon}{\beta^{+2} \sigma^2_{x_1}}} \quad \ldots \quad (A\ 10.2.2)$$

Now substituting values of β^+, $\sigma^2_{x_1}$ and β_2 i.e., $\beta^+ = 3.0$; $\sigma^2_{x_1} = 1.0$ and $\beta_2 = 2.0$ in(A 10.2.1) and (A 10.2.2), we find

$$\left.\begin{aligned} \sigma^2_\varepsilon - 4p\, \sigma^2_W - 9p = 0 \\ \text{and} \\ 4\sigma^2_W + \sigma^2_\varepsilon - 9q = 0 \end{aligned}\right\} \quad \ldots (A\ 10.2.3)$$

where

$$p = \frac{1}{R^2} - 1$$

and

$$q = \frac{1}{r^2} - 1 .$$

Solving the two equations in (A 10.2.3), we find

$$\sigma^2_\varepsilon = \frac{9 (1 - R^2)}{r^2}$$

and

$$\sigma^2_W = \frac{9(R^2 - r^2)}{4r^2} .$$

Lastly

$$r_{12}^2 = \frac{\{\text{Cov}\ (x_1, x_2)\}^2}{V(x_1)\ V(x_2)} \quad .$$

Now

$$\text{Cov}\ (x_1, x_2) = \text{Cov}\ (x_1, W + x_1\ P_{21}) \qquad (\text{since } x_{2t} - x_{1t}P_{21} = W_t \text{ from (10. 2.6)})$$

$$= P_{21}\ \sigma_{x_1}^2$$

and

$$V(x_2) = P_{21}^2\ \sigma_{x_1}^2 + \sigma_W^2 \quad .$$

Thus

$$r_{12}^2 = \frac{P_{21}^2\ \sigma_{x_1}^4}{\sigma_{x_1}^2\ (P_{21}^2\ \sigma_{x_1}^2 + \sigma_W^2)}$$

$$= \frac{P_{21}^2\ \sigma_{x_1}^2}{P_{21}^2\ \sigma_{x_1}^2 + \sigma_W^2}$$

$$= \frac{1}{1 + \dfrac{\sigma_W^2}{P_{21}^2\ \sigma_{x_1}^2}} \quad .$$

Appendix 10.3

Generation of Random Normal Deviates

There are several methods available for generating random normal deviates [see, for example, Wold (1954), Hastings et al.(1955), Box and Muller (1958) (vide Bard (1974, pp. 316-317)) and Tocher (1963)]. We used the following method for generating 2,00,000 random normal deviates used in our study.

Random three-digited numbers in the range 001 to 999 were read using library function available in the computer system. These numbers were treated as fractions (cumulative probabilities) by prefixing a decimal point before them. These cumulative probabilities were converted into random normal deviates using Hastings's approximation. Hastings et al. (1955) suggested the following approximation for computing normits i.e., normal deviates corresponding to assigned tail probabilities.

Function :

$$1 = \frac{1}{\sqrt{2\pi}} \int_{X(1)}^{\infty} e^{-\frac{1}{2}r^2} dr .$$

Range :

$$0 < 1 \leq 0.5.$$

Approximation :

$$g = \sqrt{\ln\left(\frac{1}{l^2}\right)}$$

$$X^*(l) = g - \left[\frac{a_0 + a_1 g + a_2 g^2}{1 + b_1 g + b_2 g^2 + b_3 g^3}\right]$$

where $a_0 = 2.515517$

$a_1 = 0.802853$

$a_2 = 0.010328$

$b_1 = 1.432788$

$b_2 = 0.189269$

$b_3 = 0.001308$.

Though the error margin of this approximation is very small, care was taken to reduce the error margin further by using standard correction for reduction of error margin of such approximations. In effect, we have replaced the normal distribution by a discrete distribution. The approximation, however, seems to be quite satisfactory for our sample size.

Chapter 11

Results of the Monte Carlo Experiment

11.1 Introduction

We shall now present the results of our Monte Carlo experiment and try to assess the relative performance of the five methods of estimation viz., OLS, CO, PW, D2 and HL. We have already stated that we are considering two types of autocorrelated situations : the autocorrelation is due to one of the components of ε_t and Z_t, the other being random, and (ii) the autocorrelation is due to both ε_t and Z_t. In either case, each error component, where it is autocorrelated, is assumed to follow AR(1) error process. We have also noted in the previous Chapter that in the former case we have assumed that the x_1-series is also generated by an AR(1) process, but in the latter case, we have assumed that x_1 is a random series.

We have already explained our choice of parameter values. We consider as many as 81 parametric combinations (3 values for each of R^2, r^2, δ and η or ξ) for both the situations considered. For each of the 81 combinations of parameter values we have drawn 50 samples each of size 15. The variation in parameter values across the 81 cases is expected to throw light on the effect of the individual parameters on the relative performance of the estimators being considered. As mentioned in the previous Chapter, the analysis of the results will be done in terms of

(i) absolute deviation (ii) bias (iii) sampling variance and (iv) mean square error (mse) of the estimated values $\hat{\beta}^+$. [1/, 2/] We have also calculated the 95 per cent confidence intervals for β^+ in order to examine the interval estimates. It should also be mentioned at the outset that we did look at the results obtained from individual samples separately for each parametric combination. We, however, present here only the summary statistics and discuss the broad conclusions in terms of those since the individual samples do not seem to give much additional statistical information beyond what is contained in the summary statistics. The results of our experiment for the two situations are discussed separately in sections 11.2 and 11.3. In the last section we present the concluding observations of our study.

11.2 Situation I : Z_t Follows AR(1), but e_t is Random

11.2.1 Comparison of Absolute Deviations

Let us first examine the performance of the different methods in terms of the absolute deviations $|\hat{\beta}^+ - \beta^+|$. Since for each parametric combination we have 50 samples, there will be 50 absolute deviations for

1/ While ranking the five methods by absolute bias and mse criteria, we left out those cases (i.e., parametric combinations) where the range of the 5 absolute bias/mse figures for the five methods was at most 0.01. This was done for both the situations.

2/ In case of ties (i.e., agreement upto the second place of decimal) all the tied cases were given the lower rank in ranking the five methods by absolute bias and mse criteria in both the situations. Thus, for example, the rank (position) of the 5 absolute bias figures could be like : 1, 1, 3, 4, 5.

each of the five methods. Hence a pairwise comparison of the different methods can be done by finding out the number of samples, out of 50, where the absolute deviation of each method is less than that for each other method. The results of such comparisons for each of the ten pairs of methods are presented in Table 11.1. It is clear from this table that PW method is undoubtedly better than the other methods. It has come out to be superior to OLS, CO, D2 and HL in 48, 52, 55 and 59 cases respectively. Next to PW are CO and D2 methods which are quite close to each other. Both CO and D2 beat the OLS, PW and HL in almost equal number of cases. As between CO and D2, CO beats D2 36 times and is beaten by it in 39 cases. It seems from this table that HL is the worst and the OLS is somewhat better than the HL.

Table 11.1 : Number of cases out of 81 where each method beats* each other method in absolute deviations $|\hat{\beta}^{+} - \beta^{+}|$ in 50 samples : Situation I— both Z_t and x_{1t} follow AR(1), but ε_t is random.

method beating	method beaten				
	OLS	CO	PW	D2	HL
(1)	(2)	(3)	(4)	(5)	(6)
Ordinary Least Squares (OLS)	-	30	24	32	38
Cochrane-Orcutt (CO)	43	-	23	36	57
Prais-Winsten (PW)	48	52	-	55	59
Durbin's two-step (D2)	40	39	22	-	57
Hildreth-Lu (HL)	35	19	16	18	-

* A method is considered as beating another method if the absolute deviation $|\hat{\beta}^{+} - \beta^{+}|$ for the method is less than that for the other at least 26 times in the 50 samples.

The results in Table 11.1 are based on simple counting. The sign test was used to examine whether the absolute deviations for any method tended to be smaller than those for any other method in the 50 samples for any particular parametric combination. Results of such pairwise comparisons are presented in Table 11.2. Thus, for example, the figure in the first row and the second column of this table indicates that in 5 out of the 81 cases, the absolute deviations for the OLS method are exceeded by those for the CO significantly at the 5 per cent level. The same table shows that CO beats OLS in the same sense in 11 out of 81 cases. The PW method is found to be superior at the 5 per cent level of significance. While, on the one hand, it beats the OLS, CO, D2 and HL in 16, 7, 9 and 11 cases respectively, it is, on the other hand, beaten by these methods in only 2, 1, 3 and 3 cases respectively at the 5 per cent level of significance. At the same level of significance, CO seems to occupy the second best position. It has beaten the OLS, D2 and HL in 11, 4 and 15 cases respectively and has been beated by these methods in 5, 1 and 0 cases only. Following CO are the OLS and D2. Though the OLS and D2 seem to be quite close to each other, the OLS is slightly better than D2. It is clear from the table that HL is the worst among the five methods. Furthermore, it can be seen from this table that similar conclusions hold for the relative positions of the methods at the 1 per cent level of significance also.

Table 11.2 : Number of cases out of 81 where each method beats* each other method significantly by sign test based on absolute deviations $|\hat{\beta}^{+} - \beta^{+}|$ in 50 samples : Situation I — both Z_t and x_{1t} follow AR(1), but ϵ_t is random.

method beating	method beaten				
	OLS	CO	PW	D2	HL
(1)	(2)	(3)	(4)	(5)	(6)
(a) at 5 per cent level of significance					
Ordinary Least Squares (OLS)	-	5	2	7	7
Cochrane-Orcutt (CO)	11	-	1	4	15
Prais-Winsten (PW)	16	7	-	9	11
Durbin's two-step (D2)	8	1	3	-	6
Hildreth-Lu (HL)	7	0	3	0	-
(b) at 1 per cent level of significance					
Ordinary Least Squares (OLS)	-	2	2	3	4
Cochrane-Orcutt (CO)	6	-	0	1	7
Prais-Winsten (PW)	8	3	-	3	6
Durbin's two-step (D2)	6	0	1	-	1
Hildreth-Lu (HL)	4	0	1	0	-

* A method is considered as beating another method at 5 per cent and 1 per cent levels of significance if the absolute deviation $|\hat{\beta}^{+} - \beta^{+}|$ for the method is less than that for the other in at least 32 and 34 times in the 50 samples respectively.

A comparison of the different methods was also done in terms of the averages over 50 samples of ranks of the five estimators by size of absolute deviations. The number of times out of the 81 cases, a particular method obtained the first position (i.e., showed the lowest average rank among the methods), the second position etc., are presented in Table 11.3.

Table 11.3 : Frequency distribution of relative positions* of different methods in the 81 cases according to averages over 50 samples of ranks based on absolute deviations $|\hat{\beta}^{+} - \beta^{+}|$: Situation I — both Z_t and x_{1t} follow AR(1), but ε_t is random.

position	method				
(1)	OLS (2)	CO (3)	PW (4)	D2 (5)	HL (6)
first	16	10	36	15	6
second	14	25	22	16	7
third	15	26	8	20	13
fourth	9	16	6	25	19
fifth	27	4	9	5	36
total	81	81	81	81	81

* In case of ties (i.e., more than one method having exactly the same average rank), the lower rank was given to all the tied methods.

It can be seen that out of the 81 cases, the PW method secured the first position 36 times. All the other methods were far behind so far as occupying the first position is concerned. CO appears to be the second best method, followed closely by D2 although in terms of the frequency of the first position, D2 has a slight edge over CO.

11.2.2 Comparison of Bias

Let us now consider the bias in the estimation of β^+ for the different methods. We calculate the bias as the average of $(\hat{\beta}^+ - \beta^+)$ values over the 50 samples for each method and for each parametric combination. The figures for the bias thus obtained are presented in Appendix Table A11.1. Here we discuss only the relative performance of the methods in respect of the absolute value of the bias by considering the number of parametric combinations in which the magnitude of absolute bias for a particular method is the lowest (1st position), second lowest (2nd position) etc. These figures are given in Table 11.4. It may be pointed out that in 2 out of the 81 cases the difference in absolute bias among the five methods turned out to be negligible[3/] and hence these cases were left out of consideration. From this table it appears that OLS is the best so far as occupying the first position is concerned. It has stood first as often as 32 times. However, from Appendix Table A 11.1 we may find that the difference in the figures of absolute bias of OLS vis-a-vis other methods are not always very wide and that OLS has a tendency to be the best among the five methods or being somewhat close to the best

3/ As stated in footnote 1, in these cases the range of the 5 bias figures was at most 0.01 (vide Appendix Table A 11.1).

method when δ is low (e.g., $\delta = 0.2$). Thus the conclusion on the basis of absolute bias may not be very clear. If we take first and second positions together, PW seems to be slightly better than the other two-setp methods. In respect of occupying the last position, PW seems to be much better than OLS. On the whole the performance of OLS seems to be somewhat erratic. Also, PW does not come out to be distinctly superior to the other two-step methods; in fact, PW, CO and D2 seem to be broadly similar though CO seems to be somewhat worse since it tends to occupy the third position in a large proportion of cases. The performance of HL also seems to be quite comparable to the other two-step methods. It has occupied first, third and firth positions 21, 19 and 22 times respectively. We may therefore say that the ranking of the methods by absolute bias is somewhat unclear. OLS (and to some extent HL also) cannot be easily placed in relation to CO, PW and D2; they occupy the extreme ranks (1 and 5) more often than any of these methods.

Table 11.4 : Frequency distribution of relative positions* of different methods in the 81 cases** according to absolute bias of $\hat{\beta}^+$: Situation I — both Z_t and x_{1t} follow AR(1), but ε_t is random.

position	method				
	OLS	CO	PW	D2	HL
(1)	(2)	(3)	(4)	(5)	(6)
first	32	11	13	18	21
second	8	16	28	19	8
third	6	28	17	17	19
fourth	13	17	12	20	9
fifth	20	7	9	5	22
total	79	79	79	79	79

* See footnote 2.

** As stated in footnote 1, we left out 2 cases (out of 81) where the difference in absolute bias figures for any two methods was at most 0.01.

11.2.3 Comparison of Actual and Estimated (Average) Sampling Variances

One might be interested in looking at the suitability of the conventional sampling variance estimator $\hat{V}(\hat{\beta}^+)$ for the five methods. In the presence of misspecification of the type considered here, the usual estimator of the sampling variance may overestimate, underestimate or correctly estimate the true sampling variance of $\hat{\beta}^+$, when the regressors are stochastic [Chaudhuri (1979)]. To see this, we calculated the estimates of the sampling variances of each of the five estimators by using the respective formula for each of the 50 samples. We also estimated the true sampling variance of each estimator on the basis of the 50 values of $\hat{\beta}^+$ for the 50 samples (see b(iii) of Chapter 10). This latter estimate considered as the "actual sampling variance" of $\hat{\beta}^+$ was obtained and compared with the average of the 50 estimates of sampling variance computed from the 50 samples by conventional formula for each of the 81 parametric combinations and for each estimator. These figures are given in Appendix Table A 11.2 . It is possible to find out, on the basis of these figures, the number of cases in which the actual sampling variance of any of the estimators is overestimated/underestimated/correctly estimated[4/] by the corresponding conventional sampling variance estimator for the particular method. Such summary results are presented in

---

4/ Since our "actual sampling variance" is also estimated on the basis of the 50 values of $\hat{\beta}^+$ for the 50 samples, the terms "overestimation/underestimation" are, strictly speaking, not valid. However, we are using these terms here in a somewhat loose sense.

Table 11.5. This table also gives some idea about the magnitude of over-estimation/underestimation. In the last column of the table we have presented an average (over the 81 cases) of the percentage difference between the conventional estimate of sampling variance and the actual sampling variance. It is clear from this table that for all the methods underestimation is significantly more frequent than overestimation. The counts in columns (2) and (3) suggest that OLS is the best, though only moderately superior to CO, PW, and D2, and HL is somewhat behind all others. The percentages in column (4) indicate that the conventional formula underestimates the true sampling variance by 17 to 20 per cent (approx.) for all the methods excepting for HL where the downward bias is about 25 per cent. In this latter respect again the OLS seems to be the best, though CO, PW and D2 are very close to it. It should be noted that the averaging over 81 parametric combinations blurs the picture to some extent.

Table 11.5: Number of cases out of 81* where the actual sampling variance is overestimated/underestimated by the conventional formula and the average magnitude of overestimation/underestimation, for each method of estimation**: Situation I — both Z_t and x_{1t} follow AR(1), but ε_t is random.

method	number of times the actual sampling variance is		average percentage of
	over-estimated	under-estimated	over (+) estimation/ under (−) estimation
(1)*	(2)	(3)	(4)
Ordinary Least Squares (OLS)	22	56	− 17.05
Cochrane-Orcutt (CO)	13	65	− 18.76
Prais-Winsten (PW)	12	68	− 19.56
Durbin's two-step (D2)	12	67	− 19.10
Hildreth - Lu (HL)	7	73	− 25.05

* Ties (i.e., agreement upto the second place of decimal of the two variance figures — vide Appendix Table A 11.2) are left out.

** For each parametric combination the actual sampling variance has been estimated from the 50 $\hat{\beta}^+$ values for the 50 samples. This has been compared with the average of the 50 sampling variances based on conventional formulation.

11.2.4 Performance of Confidence Intervals

In order to examine how well β^+ is estimated by means of confidence intervals constructed by conventional methods, we calculated the 95 per cent confidence interval for β^+ for the 50 samples for each of the 81 cases and for each of the five methods. The number of times these intervals failed to include the true value of β^+ was then counted. These numbers have been presented in Table 11.6. It may be seen that these counts do not, in general, indicate any marked difference on the performance of the confidence intervals for the different methods, nor do they show any systematic effect of the variation in the individual parameter values. We may, however, note that for high values of η and δ, particularly the latter, the intervals miss the true β^+ much more frequently than in the other cases. Thus, for example, when $\eta = 0.95$ and $\delta = 0.95$, the number of times the calculated intervals fail to include the true value of β^+ is, in general, higher (and much higher than the expected number) for all the methods irrespective of the values of R^2 and r^2. This, however, does not help us much in assessing the relative performance of the methods which, in fact, is our main concern. It is also not a very surprising observation. It only tells us that the higher the degree of autocorrelation in the x_1-series and in the autocorrelated component of the error term, the more pronounced will be the consequences for the interval estimates of β^+.

Table 11.6 : Number of samples where the true parameter value β^+ is not covered by the 95 per cent confidence interval, for each parametric combination and method of estimation : Situation I — both Z_t and x_{1t} follow AR(1), but ε_t is random.

srl. no.	values of correlation and auto-correlation coefficients				number of samples where β^+ is not covered				
	R^2	r^2	η	δ	OLS	CO	PW	D2	HL
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
1	0.40	0.10	0.20	0.20	2	2	3	2	3
2				0.60	4	4	4	4	5
3				0.95	4	3	2	3	3
4			0.60	0.20	2	1	2	1	1
5				0.60	2	4	4	4	4
6				0.95	4	6	5	5	5
7			0.95	0.20	2	6	4	7	6
8				0.60	8	10	8	11	10
9				0.95	9	12	12	12	11
10	0.40	0.25	0.20	0.20	1	5	5	5	5
11				0.60	4	6	5	6	6
12				0.95	2	4	4	4	5
13			0.60	0.20	2	3	3	4	4
14				0.60	1	0	0	0	3
15				0.95	3	4	7	5	8
16			0.95	0.20	1	3	2	3	4
17				0.60	6	7	6	7	7
18				0.95	5	8	9	6	8

contd...../-

Table 11.6 (contd.)

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
19	0.40	0.35	0.20	0.20	3	6	6	7	8
20				0.60	2	4	4	5	5
21				0.95	1	2	2	2	2
22			0.60	0.20	2	3	5	3	3
23				0.60	7	9	9	8	9
24				0.95	5	7	6	8	10
25			0.95	0.20	3	7	7	7	7
26				0.60	4	4	4	4	4
27				0.95	7	10	11	10	10
28	0.70	0.10	0.20	0.20	1	5	5	6	6
29				0.60	4	5	5	5	5
30				0.95	4	3	4	4	5
31			0.60	0.20	5	7	7	7	7
32				0.60	6	3	3	3	3
33				0.95	10	4	5	4	5
34			0.95	0.20	4	3	4	5	5
35				0.60	9	8	6	8	9
36				0.95	24	13	14	12	12
37	0.70	0.35	0.20	0.20	0	2	2	2	2
38				0.60	3	2	3	2	3
39				0.95	1	4	5	6	6
40			0.60	0.20	6	7	7	7	7
41				0.60	2	4	5	5	5
42				0.95	8	6	6	8	10
43			0.95	0.20	2	6	6	6	7
44				0.60	12	9	9	8	8
45				0.95	16	11	13	13	11

contd...../-

Table 11.6 (contd.)

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
46	0.70	0.65	0.20	0.20	1	2	1	2	5
47				0.60	4	3	6	3	3
48				0.95	2	2	2	3	2
49			0.60	0.20	2	2	1	4	5
50.				0.60	4	6	7	7	7
51				0.95	1	3	2	3	4
52			0.95	0.20	2	5	4	5	6
53				0.60	4	4	4	4	4
54				0.95	13	12	12	11	12
55	0.95	0.10	0.20	0.20	5	6	6	6	6
56				0.60	5	5	6	5	7
57				0.95	4	6	6	6	6
58			0.60	0.20	3	3	4	4	5
59				0.60	5	4	4	4	4
60				0.95	10	2	2	1	1
61			0.95	0.20	7	6	7	8	6
62				0.60	14	8	6	8	9
63				0.95	28	12	17	14	13

contd..../-

Table 11.6 (contd.)

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
64	0.95	0.65	0.20	0.20	2	1	2	1	2
65				0.60	1	3	3	3	3
66				0.95	8	6	7	6	5
67			0.60	0.20	4	5	5	5	5
68				0.60	4	4	3	4	4
69				0.95	13	4	5	3	2
70			0.95	0.20	8	4	4	4	4
71				0.60	13	9	7	9	12
72				0.95	25	15	16	15	15
73	0.95	0.90	0.20	0.20	3	4	4	4	4
74				0.60	4	4	4	4	4
75				0.95	5	7	7	7	7
76			0.60	0.20	6	5	4	6	6
77				0.60	5	7	9	8	7
78				0.95	5	2	4	3	3
79			0.95	0.20	0	1	0	1	1
80				0.60	21	16	18	15	17
81				0.95	17	13	13	12	13

We may mention that the performance of the confidence intervals depends on the bias of the point estimator and the bias of the sampling variance estimator. The results presented so far could lead one to expect some of the methods like PW to be superior to others like HL. These expectations do not appear to be fulfilled. We may also mention that in order to see if the bias of the different estimators were significant so that confidence intervals could be valid, we computed the ratio of absolute bias to the square root of actual sampling variance defined in subsection 11.2.3. Following Cochran (1977), we considered the effect of bias on the accuracy of the estimate as negligible if the bias was less than one tenth of the standard deviation of the estimate. For all the methods we found that in many cases the bias was not negligible.

11.2.5 Comparison of Mean Square Error

Finally, we shall analyse the mean square errors (mse) of the estimates of β^+ by the different methods. Considering the importance of mse as a criterion for judging the efficiency of an estimator, we have presented in Table 11.7 the mse of $\hat{\beta}^+$ obtained by the five methods for each of the 81 parametric combinations.

Table 11.7 : Mean square error of $\hat{\beta}^{+}$ estimated by different methods for the 81 parametric combinations : Situation I — both Z_t and x_{1t} follow AR(1), but ε_t is random.

values of auto-correlation coefficients		values of correlation coefficients								
		$R^2 = 0.40$			$R^2 = 0.70$			$R^2 = 0.95$		
		r^2=.10	r^2=.25	r^2=.35	r^2=.10	r^2=.35	r^2=.65	r^2=.10	r^2=.65	r^2= .90
η	δ									
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
		(a) Ordinary Least Squares (OLS)								
0.20	0.20	5.32	2.32	1.47	5.62	1.40	12.42	6.95	0.27	3.07
	0.60	6.48	2.40	1.64	4.90	1.49	0.39	9.36	0.39	0.06
	0.95	5.86	2.29	1.22	6.14	0.74	0.28	4.53	0.42	0.08
0.60	0.20	6.10	2.56	1.30	8.57	1.79	0.35	9.24	0.57	0.09
	0.60	7.68	1.67	2.16	10.35	1.92	0.42	12.16	0.72	0.12
	0.95	10.92	2.88	1.50	13.48	3.29	0.37	14.80	0.66	0.17
0.95	0.20	14.09	5.36	3.61	12.23	2.38	0.63	26.04	0.96	0.26
	0.60	25.26	5.80	3.27	23.54	6.82	1.10	32.22	1.96	0.48
	0.95	23.82	7.42	4.04	83.35	7.37	1.44	74.52	3.01	1.08
		(b) Cochrane-Orcutt (CO)								
0.20	0.20	5.61	2.43	1.61	7.15	1.61	39.01	7.51	0.27	0.07
	0.60	6.65	2.62	1.74	6.11	1.41	0.36	6.43	0.32	0.06
	0.95	5.29	2.52	1.24	4.26	1.13	0.36	1.33	0.17	0.07
0.60	0.20	7.15	2.77	1.18	11.19	2.18	0.45	7.89	0.59	0.12
	0.60	7.00	2.25	2.18	8.49	1.90	0.45	8.71	0.53	0.11
	0.95	9.11	3.05	1.82	6.90	2.22	0.35	2.78	0.22	0.09
0.95	0.20	19.09	7.75	4.04	13.55	3.54	1.19	24.52	0.93	0.18
	0.60	28.21	6.73	3.46	22.29	5.47	1.15	30.41	2.00	0.37
	0.95	27.31	7.44	4.07	43.13	6.46	1.88	28.91	1.15	1.17

contd...../-

Table 11.7 (contd.)

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
				(c) Prais-Winsten (PW)						
0.20	0.20	5.83	2.40	1.70	6.38	1.44	15.28	6.52	0.29	0.07
	0.60	6.35	2.24	1.59	5.04	1.37	0.37	5.83	0.29	0.05
	0.95	5.14	2.40	1.24	4.21	1.08	0.33	1.25	0.18	0.06
0.60	0.20	6.48	2.62	1.33	10.11	1.88	0.40	7.97	0.60	0.11
	0.60	6.50	1.76	2.38	8.37	1.75	0.42	7.81	0.50	0.11
	0.95	8.57	3.05	1.56	6.50	2.05	0.35	2.94	0.24	0.09
0.95	0.20	14.00	6.92	4.04	12.80	2.63	0.80	25.97	0.87	0.14
	0.60	22.86	5.55	3.29	18.07	5.89	1.03	23.95	1.62	0.33
	0.95	22.19	7.22	3.97	42.69	6.13	1.39	31.64	1.08	1.01
				(d) Durbin's two-step (D_2)						
0.20	0.20	5.96	2.63	1.60	7.43	1.66	39.84	7.65	0.28	0.08
	0.60	7.19	2.71	1.76	6.33	1.34	0.37	6.34	0.32	0.06
	0.95	5.12	2.56	1.26	4.44	1.21	0.40	1.33	0.15	0.07
0.6	0.20	7.20	2.75	1.17	11.35	2.29	0.47	8.11	0.63	0.13
	0.60	7.21	2.24	2.27	8.60	2.09	0.46	8.62	0.49	0.12
	0.95	8.40	3.29	2.19	6.60	2.17	0.38	0.94	0.19	0.09
0.95	0.20	20.01	8.16	3.88	13.71	4.18	1.24	26.33	0.94	0.19
	0.60	29.40	6.33	3.74	22.81	4.93	1.16	30.45	2.06	0.36
	0.95	27.06	7.15	4.29	33.28	6.40	1.88	17.14	1.15	1.12
				(e) Hildreth-Lu (HL)						
0.20	0.20	6.67	2.80	1.79	7.65	1.64	57.85	7.92	0.30	0.08
	0.60	8.99	2.74	1.77	6.65	1.34	0.38	6.35	0.32	0.06
	0.95	5.27	2.66	1.26	4.55	1.24	0.42	1.35	0.13	0.07
0.60	0.20	7.48	2.82	1.20	11.85	2.42	0.54	10.43	0.68	0.14
	0.60	7.30	2.40	2.28	9.30	2.39	0.46	8.69	0.53	0.12
	0.95	8.29	3.43	2.57	6.62	2.21	0.46	0.92	0.18	0.09
0.95	0.20	22.43	23.06	4.07	21.54	5.29	1.40	25.53	1.02	0.19
	0.60	31.18	7.08	4.18	36.73	4.72	1.22	36.29	3.45	0.79
	0.95	27.48	8.51	4.12	29.15	6.69	1.98	18.87	1.36	1.12

A number of observations regarding the sensitivity of the mse with respect to variation in the parameter values can be made on the basis of this table. One can see, first of all, that for given values if R^2, r^2 and δ, the mse's increase with increase in the value of η for all the five methods. No such conclusion can be drawn with respect to variation in the values of δ when other parameters are kept fixed. In other words, whereas increase in the degree of autocorrelation in the x_1-series adversely affects the mse, the same cannot be said about the effect of the degree of autocorrelation in the error term due to misspecification.

So far as the effects of the correlation coefficients are concerned, it is noticed that the mse's decrease sharply when the value of r^2 only is increased keeping other parameters fixed. This is, of course, to be expected. It only tells us that as the correlation between y and x_1 increases the efficiency of the estimator $\hat{\beta}^+$ also increases. In fact, a higher value of r^2 (when R^2 is held fixed) can also be taken to imply that the effect of misspecification is diminished, and in such situations we find that mse values also decrease. Thus we can conclude the lower the effect of misspecification, the lower is the mse.

We may also note that in most of the cases where $\delta = 0.2$, the mse values for OLS are smaller than those for the other methods. This is, of course, quite understandable. Leaving aside parametric combinations where the difference between mse's of any two of the methods did not exceed 0.01 at the most (vide footnote 1), we studied the relative positions (ranks) according to mse of different methods in the remaining 78 cases in order to get an overall picture of the comparative performance of the

five methods. The summary results are presented in Table 11.8. From this table we observe that PW has occupied the first position 34 times, the second position 28 times and so on. Thus the superiority of PW over all the other methods is once again established. HL seems to come last among the five methods : it secured the last position in the ranking as many as 43 times. Against this, OLS stood first 30 times, but it stood last 26 times. Thus, even though OLS follows PW in terms of the frequency of occupying the first position (in fact, most of these cases are where $\delta = 0.2$), when we consider the frequencies of the other positions, we find that the performance of OLS is quite variable. Looking at the second position, we find that the performances of CO and D2 are about the same, but CO stood third 40 times whereas the corresponding figure for D2 is only 15. Thus here also we can conclude that PW is the best followed by CO and D2. HL is the worst. The relative position of OLS fluctuates widely : it occupies ranks 1 and 5 with high frequency. Thus, it appears that, as compared to the other methods, OLS is more sensitive to the values of the parameter δ.

Table 11.8 : Frequency distribution of relative positions* of different methods in the 81 cases** according to the mean square error of $\hat{\beta}^+$: Situation I — both Z_t and x_{1t} follow AR(1), but ε_t is random.

position	method				
	OLS	CO	PW	D2	HL
(1)	(2)	(3)	(4)	(5)	(6)
first	30	6	34	7	8
second	13	18	28	19	4
third	3	40	7	15	9
fourth	6	13	7	34	14
fifth	26	1	2	3	43
total	78	78	78	78	78

* See footnote 2.

** As stated in footnote 1, we left out 3 cases (out of 81) where the difference in mse values for any two methods was at most 0.01.

11.3 Situation II : Both ε_t and Z_t Follow AR(1) Process

We shall now consider the second situation i.e., where autocorrelation in the error term (ε_t^+) of the misspecified model is due to both misspecification (Z_t) and other reasons (ε_t). As stated earlier, we have assumed both the error components to follow AR(1) process. Also, we now give up the assumption (made in Situation I) that the x_1-series follows an AR(1) process; this is done since the effect of η (the autocorrelation coefficient of the x_1-series) was not found to be quite important so far as the relative performance of the different methods is concerned in Situation I. We assume that x_1 is a random process and this reduces the bulk of the computations. The presentation of the results in this section follows the same arrangement as made in the preceding section.

11.3.1 Comparison of Absolute Deviations

As in Situation I, we first carry out our analysis with the criterion of absolute deviation. The results are presented in Tables 11.9 — 11.11. Table 11.9 gives the actual number of cases (out of 81) where each method is found to be superior to each other method on the basis of absolute deviations in the 50 samples. It is obvious from this table that OLS is far inferior to all other methods. OLS beats CO, PW, D2 and HL in 14, 10, 16 and 15 cases (out of 81 cases) respectively; it is, on the other hand, beaten by these methods in 65, 68, 63 and 61 cases respectively. The superiority of PW among all the five methods is also clear. It beats these methods more frequently than is beaten by them. As regards CO and D2, there is very little difference. Thus, D2 beats CO 38 times whereas it is

beaten by CO 31 times. The relative performance of HL in this situation is much better compared to Situation I though it is still behind CO and D2. Whereas HL has been beaten by CO and D2 42 and 41 times respectively, it has beaten them as often as 34 and 33 times, respectively. The difference between HL, on the one hand, and D2 and CO, on the other, was much greater in Situation I than in the present situation.

Table 11.9 : Number of cases out of 81 where each method beats* each other method in absolute deviations $|\hat{\beta}^{+} - \beta^{+}|$ in 50 samples : Situation II - both ε_t and Z_t follow AR(1), but x_{1t} is random.

method beating	method beaten				
	OLS	CO	PW	D2	HL
(1)	(2)	(3)	(4)	(5)	(6)
Ordinary Least Squares (OLS)	-	14	10	16	15
Cochrane-Orcutt (CO)	65	-	28	31	42
Prais-Winsten (PW)	68	38	-	40	46
Durbin's two-step (D2)	63	38	27	-	41
Hildreth-Lu (HL)	61	34	24	33	-

* A method is considered as beating another method if the absolute deviation $|\hat{\beta}^{+} - \beta^{+}|$ for the method is less than that for the other at least 26 times in the 50 samples.

The ranking of the methods thus obtained remains more or less unaltered when sign test is applied to examine the significance of the difference in the number of times one method is superior to another.

We observe from Table 11.10 that PW is still the best although the margin of its superiority over the other methods has slightly fallen as compared to Situation I. On careful examination, it appears that HL, D2 and CO are about equally good in Situation II while OLS has the worst performance. The relative performance of OLS is much worse than in the previous situation. In Situation II, it has significantly beaten CO, PW, D2 and HL only 2, 0, 0 and 2 times (at the 5 per cent level of significance) whereas it has been beaten by these methods as often as 39, 41, 38 and 37 times respectively (at the same level of significance).

Table 11.10 : Number of cases out of 81 where each method beats* each other method significantly by sign test based on absolute deviations $|\hat{\beta}^{+} - \beta^{+}|$ in 50 samples : Situation II — both ε_t and Z_t follow AR(1), but x_{1t} is random.

method beating	method beaten				
	OLS	CO	PW	D2	HL
(1)	(2)	(3)	(4)	(5)	(6)
(a) at 5 per cent level of significance					
Ordinary Least Squares (OLS)	-	2	0	0	2
Cochrane-Orcutt (CO)	39	-	1	1	2
Prais-Winsten (PW)	41	6	-	4	7
Durbin's two-step (D2)	38	4	2	-	5
Hildreth-Lu (HL)	37	2	0	4	-
(b) at 1 per cent level of significance					
Ordinary Least Squares (OLS)	-	1	0	0	0
Cochrane-Orcutt (CO)	31	-	0	0	1
Prais-Winsten (PW)	35	2	-	0	1
Durbin's two-step (D2)	30	0	0	-	0
Hildreth-Lu (HL)	29	0	0	1	-

* A method is considered as beating another method at the 5 per cent and 1 per cent levels of significance, if the absolute deviation $|\hat{\beta}^{+} - \beta^{+}|$ for the method is less than that for the other in at least 32 and 34 times in the 50 samples respectively.

The positions of different methods remain almost the same by the criterion of frequency distribution based on average ranks given in Table 11.11. The superiority of PW is again clear; it is followed by D2, CO and HL which are fairly close among themselves. The OLS occupies the last position as in the preceding tables. HL is far superior to OLS. Whereas HL has occupied the first, second and last positions, 12, 24 and 14 times respectively, the corresponding counts for OLS are 7, 7 and 57.

Table 11.11 : Frequency distribution of relative positions* of different methods in the 81 cases according to averages over 50 samples of ranks based on absolute deviations $|\hat{\beta}^{+} - \beta^{+}|$: Situation II — both ε_t and Z_t follow AR(1), but x_{1t} is random.

position	method				
	OLS	CO	PW	D2	HL
(1)	(2)	(3)	(4)	(5)	(6)
first	7	17	31	17	12
second	7	15	20	19	24
third	5	31	10	19	13
fourth	5	18	17	20	18
fifth	57	0	3	6	14
total	81	81	81	81	81

* In case of ties (i.e., more than one method having exactly the same average rank), the lower rank was given to all the tied methods.

11.3.2 Comparison of Bias

The relative performance of the different estimators $\hat{\beta}^{+}$ in respect of absolute bias can be studied from Table 11.12. As before, this is a summary table and the details (i.e., bias of different estimators of β^{+} for each of the 81 cases) are given in Appendix Table A 11.3. It may be noted that like the previous situation, there were six parametric combinations for which the differences in bias across the five methods were almost negligible.5/ These 6 cases were left out in preparing the table. Most of these cases had high values of R^2 and r^2 e.g., $R^2 = 0.95$ and $r^2 = 0.90$. In several other cases, the absolute biases for some of these methods (were small and) coincided upto the second place of decimal. In such cases of ties, all the tied cases were given the lower rank (vide footnote 2). As in Situation I, the picture (as emerges from this table) regarding the relative performance of the five methods by this criterion does not seem to be quite clear. Among the conventional reestimation (i.e., CO, PW, D2 and HL) methods, it appears that PW is somewhat better than the others. CO, D2 and HL are quite close among themselves though so far as the frequency of occupying the first position is concerned, HL beats CO and D2. Whereas HL has occupied the first position 23 times, the corresponding figuresfor CO and D2 are 15 and 16 respectively. OLS can not easily be placed in relation to these methods since it occupies the extreme ranks (1 and 5) more often than any of these methods. Unlike other methods, the distribution of OLS from this table appears to be

5/ As noted in footnote 1, in each of these cases the range of the 5 absolute bias figures was at most 0.01 (vide Appendix Table A 11.3).

U-shaped and this OLS seems to be somewhat more sensitive to the parameter values than the others. On the whole, there is no sharp variation among the reestimation methods in respect of the absolute bias of the estimator and the OLS is somewhat fluctuating.

Table 11.12 : Frequency distribution of relative positions* of different methods in the 81 cases** according to absolute bias of $\hat{\beta}^{+}$: Situation II — both ε_t and Z_t follow AR(1), but x_{1t} is random.

position	method				
	OLS	CO	PW	D2	HL
(1)	(2)	(3)	(4)	(5)	(6)
first	27	15	24	16	23
second	8	16	20	20	13
third	3	28	10	17	15
fourth	6	10	13	15	12
fifth	31	6	8	7	12
total	75	75	75	75	75

* See footnote 2.

** As stated in footnote 1, we left out 6 cases (out of 81) where the difference in absolute bias figures for <u>any</u> two methods was at most 0.01.

11.3.3 <u>Comparison of Actual and Estimated (Average) Sampling Variances</u>

The performance of the five different methods in this respect is summarized in Table 11.13. The picture here is considerably different from that found for the previous situation. As in Situation I, the two variances are exactly equal (agreement upto the second place of decimal)

only in a small number of cases, though the numbers are slightly higher in the present situation for the two-step methods. However, all other findings are appreciably different. For OLS, the conventional formula seems to be overestimating the actual variance, by about 8 per cent, but this may not be statistically significant. For all other methods, the conventional formulae underestimate the actual variance in a significant way, but the extent of underestimation is much lower now compared to Situation I, being around 6 to 9 per cent. It may be recalled that in Situation I, we found significant tendency to underestimate, for all the five methods, and the extent of underestimation was about 17 to 20 per cent for the four methods other than HL and about 25 per cent for HL.

Table 11.13 : Number of cases out of 81* where the actual sampling variance is overestimated/underestimated by the conventional formula and the average magnitude of overestimation/underestimation, for each method of estimation**: Situation II — both ε_t and Z_t follow AR(1), but x_{1t} is random.

method	number of times the actual sampling variance is		average percentage of over (+) estimation/ under (-) estimation
	over-estimated	under-estimated	
(1)	(2)	(3)	(4)
Ordinary Least Squares (OLS)	43	34	+ 8.05
Cochrane-Orcutt (CO)	28	49	- 8.42
Prais-Winsten (PW)	25	48	- 5.73
Durbin's two-step (D2)	22	49	- 8.19
Hildreth-Lu (HL)	20	51	- 8.92

* Ties (i.e., agreement upto the second place of decimal of the two variance figures — (vide Appendix Table A 11.4) are left out.

** For each parametric combination the actual sampling variance has been estimated from the 50 $\hat{\beta}^*$ values for the 50 samples. This has been compared with the average of the 50 sampling variances based on conventional formula.

11.3.4 Performance of Confidence Intervals

As in the previous situation, the number of samples (out of 50) where the 95 per cent confidence interval failed to include the true value of β^+ is presented in Table 11.14 for each of the methods and for each of the 81 parametric combinations. Our first observation on this table is that the performance of the different methods is much more satisfactory here than in the earlier situation. There are very few cases where the intervals fail to cover the true value in much more than 5 per cent of the samples. Unlike the previous situation, even for cases with high values of δ and ξ, the performance of the conventional confidence interval is quite good. As in the previous situation, no method appears to be appreciably better than the others by this criterion. OLS, which has been found to be relatively poor compared to the other methods in respect of absolute deviations $|\hat{\beta}^+ - \beta^+|$ does not seem to be appreciably different so far as robustness of confidence intervals is concerned. As in the previous situation, we found that for all the methods, the bias was not negligible in many cases.

Table 11.14 : Number of samples where the true parameter value β^+ is not covered by the 95 per cent confidence interval, for each parametric combination and each method of estimation : Situation II — both ε_t and Z_t follow AR(1), but x_{1t} is random.

srl. no.	values of correlation and auto-correlation coefficients				number of samples where β^+ is not covered				
	R^2	r^2	δ	ξ	OLS	CO	PW	D2	HL
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
1	0.40	0.10	0.20	0.20	3	3	2	3	4
2				0.60	2	5	5	5	5
3				0.95	2	5	5	5	6
4			0.60	0.20	2	3	4	3	3
5				0.60	2	3	3	3	3
6				0.95	4	1	2	2	2
7			0.95	0.20	0	3	3	3	3
8				0.60	4	4	3	5	5
9				0.95	4	5	5	6	6
10	0.40	0.25	0.20	0.20	4	5	4	6	5
11				0.60	4	4	3	4	3
12				0.95	4	7	4	6	6
13			0.60	0.20	4	6	7	6	6
14				0.60	2	4	4	5	5
15				0.95	2	4	3	3	4
16			0.95	0.20	4	4	5	5	5
17				0.60	1	5	5	6	7
18				0.95	5	0	1	3	2

contd...../-

Table 11.14 (contd.)

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
19	0.40	0.35	0.20	0.20	4	4	4	4	3
20				0.60	2	3	3	3	3
21				0.95	2	3	2	3	3
22			0.60	0.20	2	2	2	1	1
23				0.60	4	4	3	5	5
24				0.95	1	1	2	1	1
25			0.95	0.20	4	5	5	5	5
26				0.60	4	3	2	4	3
27				0.95	2	2	2	2	2
28	0.70	0.10	0.20	0.20	1	2	2	2	3
29				0.60	2	1	3	1	3
30				0.95	2	3	3	3	3
31			0.60	0.20	3	4	4	3	3
32				0.60	3	6	7	7	6
33				0.95	2	2	2	2	2
34			0.95	0.20	3	5	5	4	4
35				0.60	1	2	3	4	4
36				0.95	1	3	2	4	3
37	0.70	0.35	0.20	0.20	5	5	5	5	4
38				0.60	0	3	3	3	3
39				0.95	3	3	2	3	3
40			0.60	0.20	7	6	5	7	8
41				0.60	8	5	5	7	6
42				0.95	2	3	3	3	3
43			0.95	0.20	1	4	4	4	4
44				0.60	0	3	2	3	3
45				0.95	2	2	2	2	1

contd...../

Table 11.14 (contd.)

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
46	0.70	0.65	0.20	0.20	0	1	1	1	1
47				0.60	2	2	2	2	2
48				0.95	1	4	4	5	5
49			0.60	0.20	0	1	1	3	3
50				0.60	3	4	3	4	4
51				0.95	5	1	2	1	1
52			0.95	0.20	3	2	3	2	2
53				0.60	5	1	1	1	1
54				0.95	3	0	1	1	0
55	0.95	0.10	0.20	0.20	2	4	5	6	7
56				0.60	3	2	3	2	2
57				0.95	1	4	4	4	3
58			0.60	0.20	3	5	4	5	5
59				0.60	3	3	2	2	3
60				0.95	2	2	2	2	2
61			0.95	0.20	6	3	2	3	3
62				0.60	3	2	2	2	2
63				0.95	3	3	0	4	5

contd...../

Table 11.14 (contd.)

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
64	0.95	0.65	0.20	0.20	2	2	3	3	3
65				0.60	1	3	2	3	3
66				0.95	3	2	3	2	2
67			0.60	0.20	2	2	2	2	3
68				0.60	2	4	3	3	3
69				0.95	1	2	2	2	2
70			0.95	0.20	2	4	2	4	4
71				0.60	5	4	3	3	5
72				0.95	3	2	2	1	1
73	0.95	0.90	0.20	0.20	4	3	5	4	3
74				0.60	1	3	1	3	3
75				0.95	5	6	4	7	7
76			0.60	0.20	2	5	4	3	4
77				0.60	5	8	7	8	8
78				0.95	1	2	1	3	3
79			0.95	0.20	6	8	8	7	8
80				0.60	1	5	5	5	5
81				0.95	1	3	1	3	3

11.3.5 Comparisons of Mean Square Error

A comparative analysis of the performance of different methods in terms of mse of $\hat{\beta}^+$ again brings out clearly the superiority of the PW method. This is seen from Table 11.16 which gives the frequency distribution of the relative positions (ranks) of the different methods according to mse of $\hat{\beta}^+$ for 78 parametric combinations.[6] The mse values for each of the five estimators and for each of the 81 parametric combinations are given in Table 11.15.

The sensitivity of the mse values with respect to the parameters can be observed from Table 11.15. One can see, first of all, that other parameters remaining unchanged, mse mostly decreases with increase in the value of either δ or ξ for the two-step and search methods. It may also be observed that for given R^2, higher values of r^2 (δ and ξ remaining constant) result in lower mse's. This really means that the smaller the effect of misspecification, the lower are the mse values i.e., as the effect of misspecification decreases, the mse also decreases. OLS seems to be the worst among the methods particularly when the values of either δ or ξ or both δ and ξ are large. In fact, mse values for OLS are often very high compared to other methods. Sometimes it is as high as two/three times that of the others.

---

[6] The three other cases were left out because in each of these cases the range of 5 mse figures was at most 0.01, and thus for these cases the mse values could be considered to be virtually the same for the different methods (vide footnote 1).

Table 11.15 : Mean Square Error of $\hat{\beta}^{+}$ estimated by different methods for the 81 parametric combinations : Situation II — both ε_t and Z_t follow AR(1), but x_{1t} is random.

values of autocorrelation coefficients		values of correlation coefficients								
		$R^2 = 0.40$			$R^2 = 0.70$			$R^2 = 0.95$		
δ	ξ	r^2=.10	r^2=.25	r^2=.35	r^2=.10	r^2=.35	r^2=.65	r^2=.10	r^2=.65	r^2=.90
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
				(a) Ordinary Least Squares (OLS)						
0.20	0.20	4.67	2.76	1.26	5.72	1.49	0.30	4.92	0.30	0.06
	0.60	6.44	2.34	1.48	4.78	0.95	0.29	6.31	0.30	0.05
	0.95	5.56	2.63	1.21	5.31	1.17	0.21	5.15	0.36	0.13
0.60	0.20	5.38	2.57	0.96	6.40	1.59	0.27	5.51	0.29	0.07
	0.60	5.19	2.16	1.87	8.43	2.26	0.35	5.89	0.29	0.08
	0.95	10.25	1.87	1.14	7.10	1.29	0.30	6.71	0.19	0.05
0.95	0.20	5.87	3.35	1.39	6.91	1.06	0.37	6.56	0.31	0.07
	0.60	5.96	1.99	1.10	5.36	1.05	0.39	6.70	0.34	0.06
	0.95	7.50	2.46	1.26	4.76	1.89	0.47	4.62	0.30	0.07
				(b) Cochrane-Orcutt (CO)						
0.20	0.20	5.72	2.80	1.18	7.12	1.37	0.33	6.09	0.31	0.06
	0.60	5.86	1.32	0.89	4.98	0.81	0.13	6.38	0.32	0.05
	0.95	2.12	0.89	0.18	4.49	0.74	0.06	5.96	0.35	0.07
0.60	0.20	5.99	2.48	1.06	5.57	1.38	0.28	3.75	0.24	0.07
	0.60	3.02	1.61	0.87	4.79	1.24	0.20	4.61	0.19	0.07
	0.95	1.89	0.35	0.12	2.31	0.47	0.06	2.89	0.13	0.02
0.95	0.20	4.21	2.49	1.47	3.68	1.02	0.24	0.74	0.07	0.05
	0.60	2.11	0.94	0.66	1.84	0.25	0.15	0.63	0.07	0.02
	0.95	0.56	0.14	0.07	0.42	0.14	0.03	0.48	0.02	0.01

contd...../

Table 11.15 (contd.)

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
					(c) Prais-Winsten (PW)					
0.20	0.20	5.38	2.65	1.23	6.39	1.45	0.33	5.49	0.33	0.05
	0.60	5.28	1.54	0.87	4.36	0.72	0.13	6.48	0.29	0.04
	0.95	1.98	0.89	0.19	4.00	0.74	0.06	5.57	0.34	0.06
0.60	0.20	6.08	2.27	0.99	5.02	1.29	0.27	3.52	0.23	0.07
	0.60	2.90	1.26	0.96	4.63	1.26	0.18	4.24	0.20	0.07
	0.95	2.01	0.31	0.10	2.21	0.36	0.06	2.99	0.13	0.02
0.95	0.20	4.26	2.40	1.34	3.61	0.80	0.26	0.87	0.06	0.04
	0.60	1.85	0.87	0.53	1.69	0.20	0.21	0.65	0.08	0.03
	0.95	0.56	0.15	0.08	0.38	0.16	0.03	0.48	0.03	0.01
					(d) Durbin's two-step (D2)					
0.20	0.20	5.85	2.82	1.17	7.27	1.32	0.34	6.25	0.31	0.06
	0.60	5.85	1.28	0.91	5.15	0.90	0.12	0.45	0.32	0.05
	0.95	2.03	0.83	0.18	4.52	0.75	0.06	6.04	0.35	0.05
0.60	0.20	6.04	2.57	1.07	5.58	1.40	0.30	3.66	0.23	0.07
	0.60	2.80	1.70	0.79	4.59	1.18	0.19	4.59	0.20	0.06
	0.95	1.53	0.37	0.11	2.04	0.35	0.06	2.96	0.13	0.02
0.95	0.20	4.21	2.46	1.53	3.27	1.02	0.24	0.69	0.07	0.05
	0.60	1.89	0.92	0.75	1.87	0.27	0.14	0.58	0.07	0.02
	0.95	0.59	0.13	0.07	0.38	0.08	0.01	0.42	0.02	0.01
					(e) Hildreth-Lu (HL)					
0.20	0.20	5.97	2.84	1.17	7.67	1.26	0.34	6.68	0.33	0.06
	0.60	5.95	1.26	0.88	5.37	0.87	0.11	6.49	0.33	0.05
	0.95	2.10	0.79	0.18	4.59	0.74	0.06	6.48	0.35	0.05
0.60	0.20	6.23	2.87	1.10	5.71	1.40	0.33	3.54	0.23	0.07
	0.60	2.65	1.81	0.72	4.67	1.21	0.18	4.62	0.20	0.05
	0.95	1.55	0.36	0.11	2.06	0.36	0.06	2.91	0.13	0.02
0.95	0.20	4.27	2.52	1.61	3.03	1.05	0.23	0.68	0.07	0.05
	0.60	1.67	0.91	0.67	1.87	0.28	0.13	0.59	0.07	0.02
	0.95	0.59	0.12	0.07	0.38	0.08	0.01	0.43	0.02	0.01

From Table 11.16, we observe that PW has occupied the first position the highest number (37) of times followed by HL (31 times), D2 (24 times), CO (16 times) and OLS (9 times). It may be noted that in case of ties i.e., agreement upto the second decimal place, we followed the rule given in footnote 2 in ranking the methods. OLS is clearly the worst of the methods as it occupies the 5th position 64 times out of 78. Thus, from this table, we observe that PW is still somewhat better than the other methods (although its superiority over other methods is somewhat reduced in this situation compared to the previous one) and OLS is the worst among the methods. There is not much to choose between CO, D2 and HL, although closer scrutiny shows that CO is slightly interior to both D2 and HL. The worst performance of OLS may be understood from the result for Situation I where as δ rises, OLS, in general, falls in the ranking among the five methods. In the case considered for Sitiation II, the autocorrelation of the ε_t^+ series is generally somewhat high as both the components ε_t and Z_t are autocorrelated excepting for those cases where both δ and ξ are low e.g., $\delta = 0.2$ and $\xi = 0.2$.

Table 11.16 : Frequency distribution of relative positions* of different methods in the 81 cases** according to the mean square error of $\hat{\beta}^+$: Situation II — both ε_t and Z_t follow AR(1), but x_{1t} is random.

position	method				
	OLS	CO	PW	D2	HL
(1)	(2)	(3)	(4)	(5)	(6)
first	9	16	37	24	31
second	4	20	10	24	15
third	1	27	14	15	8
fourth	0	15	17	15	12
fifth	64	0	0	0	12
total	78	78	78	78	78

* See footnote 2.

** As stated in footnote 1, we left out 3 cases (out of 81) where the difference in mse values for any two methods was at most 0.01.

11.4 Conclusions

In this concluding section we discuss the main results of our Monte Carlo experiment designed to investigate the small-sample properties of some commonly used estimators in a linear regression model with autocorrelated errors. The error term has been considered as consisting of two components — one due to misspecification and the other due to factors other than misspecification — and the autocorrelation is assumed to be caused by one or both the components. It has been seen in Chapter 9 of our dissertation that in this kind of a situation all the conventional reestimation methods are, in general, inconsistent; the OLS estimator on the other hand is, in general, consistent.[7/] Besides, there are questions of efficiency and valid estimation of sampling errors. The special situation where these methods give consistent though inefficient estimates includes the cases being considered here.

We carried out the present experiment to investigate how different conventional reestimation methods and the OLS perform in small samples when the error term is decomposed as stated above meaning thereby that the model has been misspecified which might or might not give rise to autocorrelation. Since the standard treatment of the problem of autocorrelation among disturbances does not consider decomposed error terms, it would be instructive to examine the properties of these estimators, particularly in the small samples for which analytical results cannot be obtained.

It might have been noticed that the emphasis and focus of attention of this study has been on the relative performance of some conventional methods of estimation viz., the OLS, CO, PW, D2 and HL. In the standard

7/ See, in this context, footnote 2 of Chapter 5 and footnote 5 of Chapter 9.

autocorrelated linear regression model some of these methods are recommended from the point of view of efficiency. In the models considered by us, however, all these methods have been shown to be not fully efficient (vide Chapter 10). The grounds for using any of the conventional methods, therefore, become rather shaky. However, they are likely to be used in practice because the efficient methods (see next paragraph) are somewhat difficult and their small-sample performance is yet to be investigated. In any case, the choice among these methods and the OLS, for our kind of models, has to be made on the basis of their small-sample performance. This is what we have tried to examine in the Monte Carlo study.

It may also be pointed out that we have not considered for the purpose of this experiment either the standard ML method of estimation (with AR(1) error process) or the methods of estimation developed by us in Part II of this dissertation. This is because the present experiment was completed before the ML method came to be frequently used in such situations[8/] and before the work reported in Part II of this thesis was taken up. Admittedly, it would have been better if the small-sample properties of these methods were also studied. However, it appears that the standard ML method of estimation (with AR(1) error process) will not be fully efficient in small samples since it suffers from the same deficiency as the reestimation methods i.e., the error ε_t^+ does not, in general, follow AR(1) process under the assumed conditions.

8/ The ML method of estimation, though a standard method of estimation in statistics, has come to be used for estimation of autocorrelated linear regression models relatively recently.

We may now summarise the major findings of our Monte Carlo experiment.

(i) For low values of δ (e.g., $\delta = 0.2$) OLS appears to be as good as the other methods; in fact, sometimes it seems to be better than the others. This is quite understandable. However, this observation is true only for Situation I. For Situation II this observation is somewhat true only when both δ and ξ (i.e., the autocorrelation coefficient of ε_t series) are low (e.g., $\delta = 0.2$ and $\xi = 0.2$). For other values of δ and ξ, the autocorrelation in the ε_t^+ series is generally somewhat high even if one of δ and ξ has low value.

(ii) For the other combinations of parametric cases, judging by absolute deviation of estimates $\hat{\beta}^+$ for β^+, bias, absolute bias and mse, we find that for Situation I PW is superior to the remaining methods. Next to PW are CO and D2 in that order. The difference between CO and D2 are, however, small. HL is better than OLS but much worse than even CO and D2. As for Situation II, the superiority of PW is retained though its superiority is now slightly less compared to that in Situation I. The interesting feature here is that HL is almost as good as CO and D2, sometimes even slightly better than CO. The performance of OLS is the worst and this is more pronounced in this situation. It may, however, be noted that this observation is very clear with respect to absolute deviation and mse. The ranking is not so clear with respect to bias and absolute bias though it appears to be somewhat similar.

(iii) Confidence interval estimates do not throw much light on the relative performance of the methods in both the situations. However, for

Situation II, the proportion of cases in which the interval estimates leave out the true value of β^+ is not much higher than the stated probability value (α). This is unlike Situation I where for high values of δ and η, the proportion of such cases is much higher than α.

(iv) For both the situations, the mse values decrease with decrease in the degree of misspecification (as given by the difference of R^2 and r^2 with a fixed R^2).

(v) For all the methods, the incidence of underestimation of the actual sampling variance of $\hat{\beta}^+$ by the conventional formulae for sampling variance associated with these methods is quite large for Situation I. While for all the methods excepting for HL the extent of underestimation is about 17-20 per cent, for HL it is about 25 per cent. OLS seems to be somewhat better in this respect. However, for Situation II, the extent of underestimation is about half of that for Situation I for the two-step methods. For OLS, there is now overestimation by about 8 per cent.

Apart from the deficiency in performance in respect of confidence intervals, the Prais-Winsten (PW) method seems to be a fairly satisfactory procedure for both the situations, though the standard ML method (with AR(1) error process) or the procedures developed in Part II of this thesis are expected to be more efficient. CO and D2 may be regarded as the second best methods in both the situations. The HL is inferior in Situation I but quite comparable to these methods in Situation II. OLS is the worst among all the methods for both the situations — more particularly in Situation II; however, for small values of δ in Situation I, OLS seems to be as good as the others, in general.

Appendix 11

Table A 11.1 : Bias in the estimation of β^+ by the different methods for the 81 parametric combinations : Situation I— both Z_t and x_{1t} follow AR(1), but ε_t is random.

values of autocorrelation coefficients		values of correlation coefficients								
		R^2=0.40			R^2=0.70			R^2=0.95		
		r^2=.10	r^2=.25	r^2=.35	r^2=.10	r^2=.35	r^2=.65	r^2=.10	r^2=.65	r^2=.90
η	δ	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
					(a) Ordinary Least Squares (OLS)							
0.20	0.20	0.07	-0.08	-0.18	-0.17	0.08	0.38	0.03	0.10	-0.07
	0.60	-0.10	-0.15	0.10	-0.38	0.08	-0.04	0.16	-0.03	0.04
	0.95	-0.12	-0.20	0.04	-0.16	0.11	0.03	0.18	0.13	0.09
0.60	0.20	0.07	-0.19	0.22	0.02	0.01	0.19	0.05	-0.04	-0.03
	0.60	0.27	0.12	-0.08	0.03	0.04	0.01	0.06	0.05	0.05
	0.95	-0.84	0.50	0.02	-0.16	-0.15	0.06	-0.44	0.06	0.08
0.95	0.20	0.02	-0.13	-0.23	-0.78	0.04	-0.10	-0.32	0.01	-0.01
	0.60	-0.07	0.14	-0.42	-0.93	0.82	-0.06	0.64	0.21	-0.10
	0.95	-0.06	0.32	0.09	-1.47	-0.17	-0.05	0.56	-0.31	-0.15
					(b) Cochrane-Orcutt (CO)							
0.20	0.20	0.07	0.06	-0.24	-0.24	0.11	-0.39	0.10	0.16	-0.08
	0.60	0.02	-0.07	0.11	-0.62	0.23	-0.02	0.02	-0.05	0.01
	0.95	-0.02	-0.20	-0.04	-0.20	0.16	0.05	-0.22	0.09	0.01
0.60	0.20	0.10	-0.22	0.21	-0.16	-0.05	0.17	0.22	-0.10	-0.04
	0.60	0.27	0.14	-0.03	0.27	0.05	-0.08	0.22	0.04	0.02
	0.95	-0.61	0.41	-0.02	0.11	-0.16	0.10	-0.08	-0.04	0.02
0.95	0.20	0.02	-0.09	-0.40	-0.91	0.00	-0.05	-0.68	-0.08	-0.02
	0.60	0.85	0.06	-0.40	-1.06	0.66	-0.16	0.45	0.25	-0.03
	0.95	-0.17	0.58	-0.03	-0.60	-0.08	-0.12	0.50	-0.10	-0.23

contd...../-

Table A 11.1 (contd.)

(1)	(3)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
				(c) Prais-Winsten (PW)						
0.20	0.20	0.11	-0.01	-0.20	-0.12	0.12	0.36	-0.03	0.13	-0.07
	0.60	-0.12	-0.07	0.11	-0.62	0.21	-0.02	0.08	-0.03	0.01
	0.95	-0.08	-0.25	0.03	-0.22	0.15	0.02	-0.23	-0.08	0.03
0.60	0.20	0.09	-0.25	0.18	-0.11	0.00	0.19	0.12	-0.09	-0.04
	0.60	0.21	0.11	-0.02	0.38	-0.05	-0.02	0.15	0.05	0.05
	0.95	-0.72	0.45	-0.03	0.14	-0.17	0.05	-0.13	-0.04	0.01
0.95	0.20	0.06	-0.08	-0.27	-0.91	-0.01	-0.06	-0.45	-0.02	-0.02
	0.60	0.31	0.16	-0.38	-0.85	0.82	-0.07	0.54	0.21	-0.05
	0.95	0.01	0.42	0.07	-0.43	-0.14	-0.05	0.77	-0.12	-0.20
				(d) Durbin's two-step (D2)						
0.20	0.20	0.13	0.03	-0.24	-0.25	0.10	-0.21	0.08	0.16	-0.08
	0.60	0.06	-0.07	0.11	-0.62	0.27	-0.02	0.00	-0.06	0.00
	0.95	0.03	-0.17	0.02	-0.15	0.16	0.06	-0.23	0.08	0.00
0.60	0.20	0.11	-0.22	0.21	-0.20	-0.05	0.15	0.25	-0.12	-0.05
	0.60	0.29	0.13	-0.01	0.31	0.00	-0.09	0.26	0.03	0.02
	0.95	-0.57	0.39	-0.13	0.19	-0.14	0.09	-0.08	-0.06	0.01
0.95	0.20	-0.02	0.03	-0.40	-0.94	0.03	-0.05	-0.60	-0.07	-0.03
	0.60	0.93	0.12	-0.36	-0.56	0.56	-0.16	0.27	0.26	0.01
	0.95	-0.11	0.50	-0.07	-0.11	-0.06	-0.11	0.38	-0.06	-0.22
				(e) Hildreth-Lu (HL)						
0.20	0.20	0.17	0.05	-0.24	-0.23	0.10	-0.24	0.04	0.17	-0.08
	0.60	0.11	-0.05	0.11	-0.70	0.30	-0.01	-0.05	-0.05	0.00
	0.95	-0.08	-0.20	0.03	-0.17	0.18	0.07	-0.23	0.04	-0.02
0.60	0.20	0.10	-0.19	0.23	-0.30	-0.05	0.18	0.36	-0.14	-0.05
	0.60	0.23	0.14	-0.01	0.30	-0.03	-0.09	0.32	0.00	0.02
	0.95	-0.43	0.41	-0.18	0.25	-0.12	0.08	-0.13	-0.05	0.00
0.95	0.20	-0.01	0.48	-0.45	-1.25	0.10	-0.04	-0.76	-0.08	-0.01
	0.60	1.05	0.05	-0.33	-0.17	0.38	-0.21	0.37	0.29	0.10
	0.95	-0.14	0.58	-0.04	-0.07	0.06	-0.12	0.79	0.12	-0.21

Table A 11.2 : Actual sampling variance of $\hat{\beta}^{+}$ estimated from the 50 $\hat{\beta}^{+}$ values for the 50 samples and the average of 50 sampling variances based on conventional formula for each parametric combination and method of estimation : Situation I — both Z_t and x_{1t} follow AR(1), but ε_t is random.

srl. no.	values of correlation and autocorrelation efficients				OLS		CO		PW		D2		HL	
	R^2	r^2	η	δ	actual variance	estimated variance (average)	actual variance	estimated variance (average)	actual variance	estimated variance (average)	actual variance	estimated variance (average)	actual variance	estimated variance (average)
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
1	0.40	0.10	0.20	0.20	5.42	6.13	5.72	6.18	5.94	5.63	6.06	6.13	6.78	6.04
2				0.60	6.61	6.75	6.79	6.63	6.46	5.99	7.33	6.52	9.16	6.34
3				0.95	5.97	5.58	5.40	4.82	5.23	4.53	5.22	4.77	5.37	4.78
4			0.60	0.20	6.21	6.54	7.28	5.90	6.60	5.70	7.33	5.85	7.62	5.84
5				0.60	7.76	6.50	7.07	6.47	6.59	6.10	7.27	6.40	7.40	6.49
6				0.95	10.42	6.20	8.91	6.38	8.21	5.86	8.24	6.41	8.27	6.44
7			0.95	0.20	14.38	15.98	19.48	16.86	14.28	16.15	20.41	16.79	22.89	17.16
8				0.60	25.77	19.48	28.05	26.05	23.22	22.49	29.12	26.38	30.70	26.68
9				0.95	28.90	12.44	27.84	16.28	22.65	14.43	27.60	16.53	28.02	16.77

contd...../-

Table A 11.2 (contd.)

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
10	0.40	0.25	0.20	0.20	2.36	2.02	2.48	2.03	2.44	1.86	2.68	2.01	2.86	2.00
11				0.60	2.43	1.80	2.67	1.76	2.28	1.65	2.76	1.76	2.79	1.75
12				0.95	2.30	2.18	2.53	1.89	2.38	1.79	2.58	1.87	2.68	1.85
13			0.60	0.20	2.57	2.22	2.78	2.17	2.61	2.08	2.76	2.18	2.84	2.18
14				0.60	1.69	2.11	2.27	2.27	1.78	2.04	2.27	2.27	2.43	2.30
15				0.95	2.69	2.36	2.94	2.56	2.91	2.30	3.21	2.57	3.33	2.58
16			0.95	0.20	5.45	8.42	7.90	8.10	7.06	8.05	8.32	8.54	23.30	9.18
17				0.60	5.90	4.51	6.86	4.73	5.64	4.33	6.45	4.82	7.22	4.67
18				0.95	7.47	5.44	7.25	5.82	7.19	5.31	7.04	5.75	8.34	5.98

contd...../-

Table A 11.2 (contd.)

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
19	0.40	0.35	0.20	0.20	1.46	1.28	1.59	1.19	1.69	1.12	1.57	1.19	1.77	1.18
20				0.60	1.66	1.23	1.76	1.18	1.61	1.12	1.78	1.17	1.79	1.17
21				0.95	1.24	1.32	1.26	1.19	1.26	1.12	1.28	1.19	1.29	1.19
22			0.60	0.20	1.28	1.23	1.16	1.23	1.33	1.14	1.15	1.21	1.17	1.21
23				0.60	2.20	1.61	2.22	1.74	2.43	1.53	2.31	1.75	2.33	1.72
24				0.95	1.53	1.63	1.86	1.58	1.59	1.50	2.22	1.59	2.59	1.58
25			0.95	0.20	3.63	3.20	3.96	3.56	4.05	3.17	3.80	3.50	3.95	3.55
26				0.60	3.16	3.52	3.37	4.01	3.21	3.55	3.69	4.11	4.16	4.17
27				0.95	4.11	2.61	4.15	2.89	4.05	2.52	4.37	2.93	4.20	2.88

contd..... /-

Table A 11.2 (contd.)

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
28	0.70	0.10	0.20	0.20	5.71	5.61	7.24	5.55	6.49	5.11	7.52	5.52	7.75	5.47
29				0.60	4.86	5.31	5.85	4.32	4.75	3.99	6.07	4.24	6.28	4.19
30				0.95	6.24	5.71	4.30	3.16	4.25	3.02	4.51	3.06	4.61	3.00
31			0.60	0.20	8.75	6.41	11.39	6.87	10.30	6.07	11.54	6.78	12.00	6.87
32				0.60	10.56	7.72	8.59	7.45	8.39	7.12	8.68	7.48	9.40	7.42
33				0.95	13.73	6.74	7.03	5.09	6.61	4.48	6.70	4.92	6.69	4.87
34			0.95	0.20	11.86	11.22	12.97	14.81	12.22	12.48	13.09	15.09	20.38	15.50
35				0.60	23.14	11.57	21.61	17.52	17.70	15.20	22.95	18.92	37.45	19.40
36				0.95	82.86	11.43	43.64	17.08	43.37	15.15	33.95	16.99	29.74	17.58

contd..... /-

Table A 11.2 (contd.)

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
37	0.70	0.35	0.20	0.20	1.42	1.45	1.63	1.41	1.46	1.35	1.68	1.40	1.67	1.41
38				0.60	1.51	1.26	1.39	1.05	1.35	1.01	1.29	1.04	1.27	1.02
39				0.95	0.74	1.20	1.13	0.79	1.07	0.74	1.21	0.77	1.24	0.77
40			0.60	0.20	1.83	1.18	2.23	1.22	1.91	1.15	2.34	1.23	2.47	1.22
41				0.60	1.96	1.40	1.93	1.45	1.78	1.34	2.13	1.41	2.44	1.40
42				0.95	3.33	1.53	2.24	1.30	2.06	1.24	2.19	1.27	2.24	1.26
43			0.95	0.20	2.43	3.61	3.61	4.28	2.68	3.63	4.26	4.47	5.39	4.43
44				0.60	6.27	2.53	5.14	3.45	5.33	3.08	4.71	3.56	4.67	3.65
45				0.95	7.50	2.40	6.59	3.82	6.24	3.05	6.53	3.70	6.82	3.95

contd...../-

Table A 11.2 (contd.)

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
46	0.70	0.65	0.20	0.20	12.52	21.09	39.66	43.18	15.47	17.79	40.61	41.40	58.97	43.68
47				0.60	0.39	0.34	0.37	0.31	0.38	0.29	0.38	0.31	0.38	0.31
48				0.95	0.29	0.36	0.36	0.33	0.34	0.31	0.40	0.32	0.42	0.32
49			0.60	0.20	0.32	0.33	0.43	0.35	0.37	0.32	0.45	0.35	0.52	0.36
50				0.60	0.43	0.34	0.45	0.36	0.43	0.31	0.46	0.36	0.46	0.36
51				0.95	0.37	0.36	0.35	0.37	0.36	0.32	0.38	0.37	0.46	0.37
52			0.95	0.20	0.63	0.83	1.21	0.89	0.81	0.77	1.26	0.89	1.43	0.90
53				0.60	1.12	1.37	1.15	1.45	1.05	1.32	1.15	1.44	1.20	1.48
54				0.95	1.47	0.96	1.90	1.37	1.41	1.03	1.90	1.39	2.00	1.42

contd..... /-

Table A 11.2 (contd.)

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
55	0.95	0.10	0.20	0.20	7.09	5.78	7.66	5.44	6.65	5.11	7.80	5.40	8.08	5.41
56				0.60	9.52	5.75	6.56	4.23	5.94	4.04	6.47	4.12	6.48	4.05
57				0.95	4.59	4.86	1.30	0.93	1.23	0.98	1.31	0.87	1.32	0.85
58			0.60	0.20	9.42	7.18	8.00	8.07	8.12	7.25	8.21	8.09	10.51	7.98
59				0.60	12.41	7.03	8.84	6.34	7.95	5.87	8.73	6.26	8.77	6.28
60				0.95	14.91	6.23	2.84	1.72	2.99	1.77	0.95	1.53	0.92	1.52
61			0.95	0.20	26.47	14.74	24.55	19.93	26.29	16.65	26.49	20.42	25.46	20.27
62				0.60	32.46	17.58	30.82	24.33	24.14	22.54	30.99	25.05	36.89	24.87
63				0.95	75.72	13.94	29.25	8.50	31.68	8.23	17.35	7.56	18.62	8.16

contd.../

Table A 11.2 (contd.)

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
64	0.95	0.65	0.20	0.20	0.27	0.33	0.25	0.31	0.28	0.30	0.26	0.31	0.27	0.31
65				0.60	0.39	0.39	0.32	0.28	0.29	0.27	0.32	0.28	0.33	0.28
66				0.95	0.41	0.30	0.16	0.11	0.18	0.11	0.14	0.11	0.13	0.10
67			0.60	0.20	0.58	0.46	0.59	0.53	0.61	0.46	0.63	0.53	0.67	0.54
68				0.60	0.73	0.43	0.54	0.43	0.51	0.41	0.50	0.43	0.54	0.42
69				0.95	0.67	0.32	0.23	0.17	0.25	0.16	0.19	0.16	0.18	0.16
70			0.95	0.20	0.98	0.93	0.94	1.20	0.89	1.07	0.95	1.20	1.04	1.23
71				0.60	1.96	0.87	1.97	1.17	1.61	0.99	2.03	1.18	3.44	1.37
72				0.95	2.97	0.57	1.16	0.52	1.09	0.47	1.17	0.52	1.37	0.60

contd...../

Table A 11.2 (contd.)

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
73	0.95	0.90	0.20	0.20	0.07	0.07	0.07	0.07	0.06	0.06	0.07	0.07	0.07	0.06
74				0.60	0.06	0.07	0.06	0.06	0.05	0.06	0.07	0.06	0.06	0.06
75				0.95	0.07	0.06	0.07	0.05	0.06	0.05	0.07	0.05	0.07	0.05
76			0.60	0.20	0.09	0.09	0.12	0.10	0.11	0.09	0.13	0.10	0.14	0.10
77				0.60	0.12	0.08	0.11	0.08	0.11	0.07	0.12	0.08	0.12	0.08
78				0.95	0.17	0.08	0.09	0.07	0.09	0.06	0.09	0.07	0.09	0.06
79			0.95	0.20	0.27	0.30	0.19	0.29	0.14	0.21	0.19	0.28	0.20	0.28
80				0.60	0.43	0.14	0.37	0.19	0.34	0.16	0.37	0.20	0.80	0.24
81				0.95	1.08	0.20	1.14	0.24	0.99	0.22	1.09	0.25	1.09	0.26

Table A 11.3 : Bias in the estimation of β^+ by the different methods for the 81 parametric combinations : Situation II - both ϵ_t and Z_t follow AR(1), but x_{1t} is random.

values of autocorrelation coefficients		values of correlation coefficients								
		R^2=0.40			R^2=0.70			R^2=0.90		
		r^2=.10	r^2=.25	r^2=.35	r^2=.10	r^2=.35	r^2=.65	r^2=.10	r^2=.65	r^2=.90
δ	ξ									
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
					(a) Ordinary Least Squares (OLS)					
0.20	0.20	0.18	-0.05	-0.08	-0.12	0.37	-0.05	0.19	0.11	-0.06
	0.60	0.38	-0.22	-0.12	-0.17	0.01	-0.07	-0.07	-0.07	0.01
	0.95	0.17	-0.18	0.03	0.37	0.01	0.08	-0.30	-0.05	0.08
0.60	0.20	-0.05	-0.15	0.24	-0.40	-0.28	0.02	-0.12	-0.02	0.01
	0.60	-0.12	-0.18	-0.48	0.28	0.09	0.13	0.08	0.02	0.02
	0.95	-0.24	0.17	-0.17	0.14	-0.11	-0.03	0.11	0.03	0.02
0.95	0.20	0.68	0.34	0.00	-0.03	0.09	0.08	-0.11	-0.07	-0.03
	0.60	0.94	-0.09	0.13	0.11	0.22	-0.02	0.00	0.07	-0.05
	0.95	0.57	0.00	0.31	0.12	-0.29	0.08	0.13	-0.01	0.03
					(b) Cochrane-Orcutt (CO)					
0.20	0.20	0.22	0.04	-0.10	0.04	0.40	-0.03	0.35	0.16	-0.07
	0.60	0.23	-0.11	-0.09	-0.26	-0.02	-0.11	0.17	-0.09	-0.04
	0.95	-0.26	-0.05	0.06	0.41	-0.02	0.01	-0.42	-0.02	0.03
0.60	0.20	-0.26	-0.22	0.29	-0.48	-0.11	-0.01	-0.08	-0.04	0.01
	0.60	-0.14	-0.20	-0.28	0.17	-0.05	0.09	0.11	0.02	0.00
	0.95	-0.21	0.14	0.00	0.13	-0.13	-0.04	0.23	0.06	0.03
0.95	0.20	0.61	0.42	0.02	-0.41	0.23	0.06	-0.04	-0.09	0.03
	0.60	0.45	-0.05	0.09	-0.25	0.02	-0.01	0.03	0.06	0.00
	0.95	-0.05	0.02	-0.02	-0.15	-0.06	0.02	0.14	0.01	0.01

contd...../-

Table A.11.3 (contd.)

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
					(c) Prais-Winsten (PW)					
0.20	0.20	0.13	-0.01	-0.04	-0.03	0.37	-0.03	0.34	0.13	-0.06
	0.60	0.25	-0.08	-0.15	-0.22	0.02	-0.09	0.03	-0.06	-0.03
	0.95	-0.20	-0.05	0.06	0.36	-0.02	0.01	-0.45	-0.04	0.05
0.60	0.20	-0.18	-0.15	0.24	-0.33	-0.11	0.00	0.01	-0.05	0.00
	0.60	-0.15	-0.15	-0.31	0.18	-0.03	0.12	0.18	0.02	0.01
	0.95	-0.17	0.15	0.00	0.13	-0.10	-0.04	0.20	0.05	0.02
0.95	0.20	0.58	0.43	0.01	-0.41	0.21	0.07	-0.07	-0.07	0.01
	0.60	0.42	-0.05	0.12	-0.25	-0.01	-0.05	0.03	0.07	-0.01
	0.95	-0.07	0.02	-0.02	-0.12	-0.06	0.01	0.18	0.01	0.01
					(d) Durbin's two-step (D2)					
0.20	0.20	0.17	0.05	-0.09	0.04	0.39	-0.03	0.36	0.17	-0.07
	0.60	0.22	-0.14	-0.10	-0.28	-0.05	-0.12	0.19	-0.10	-0.04
	0.95	-0.32	-0.04	0.06	0.40	-0.02	0.00	-0.43	-0.02	0.04
0.60	0.20	-0.27	-0.24	0.28	-0.49	-0.10	-0.02	-0.01	-0.04	0.01
	0.60	-0.12	-0.17	-0.25	0.15	-0.07	0.10	0.09	0.02	0.01
	0.95	-0.12	0.16	0.02	0.11	-0.12	-0.03	0.23	0.07	0.03
0.95	0.20	0.62	0.40	0.01	-0.39	0.23	0.05	-0.04	-0.09	0.04
	0.60	0.38	-0.05	0.10	-0.32	0.00	-0.01	0.02	0.06	0.00
	0.95	-0.06	0.03	-0.02	-0.11	-0.04	0.00	0.16	0.01	0.01
					(e) Hildreth-Lu (HL)					
0.20	0.20	0.23	0.04	-0.09	0.06	0.38	-0.03	0.42	0.17	-0.07
	0.60	0.23	-0.14	-0.10	-0.34	-0.04	-0.11	0.18	-0.10	0.04
	0.95	-0.37	-0.04	0.06	0.39	-0.03	0.00	-0.41	-0.02	0.04
0.60	0.20	-0.28	-0.23	0.25	-0.46	-0.09	-0.05	-0.11	-0.03	0.00
	0.60	-0.09	-0.15	-0.22	0.12	-0.07	0.10	0.11	0.01	0.01
	0.95	-0.11	0.17	0.03	0.07	-0.15	-0.03	0.22	0.06	0.03
0.95	0.20	0.56	0.40	-0.02	-0.45	0.25	0.05	-0.04	-0.10	0.04
	0.60	0.32	-0.06	0.09	-0.34	0.00	-0.01	0.03	0.05	0.00
	0.95	-0.04	0.02	-0.02	-0.11	-0.05	-0.01	0.16	0.00	0.01

Table A 11.4 : Actual sampling variance of $\hat{\beta}_+$ estimated from the 50 $\hat{\beta}_+$ values for the 50 samples and the average of 50 sampling variances based on conventional formula for each parametric combination and method of estimation : Situation II — both ε_t and Z_t follow AR(1), but x_{1t} is random.

srl. no.	values of correlation and autocorrelation coefficients				OLS		CO		PW		D2		HL	
	R^2	r^2	δ	ξ	actual variance	estimated variance (average)	actual variance	estimated variance (average)	actual variance	estimated variance (average)	actual variance	estimated variance (average)	actual variance	estimated variance (average)
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
1	0.40	0.10	0.20	0.20	4.74	6.05	5.79	5.53	5.47	5.23	5.94	5.45	6.04	5.45
2				0.60	6.42	6.16	5.93	3.73	5.32	3.78	5.92	3.67	6.02	3.65
3				0.95	5.64	4.76	2.09	2.28	1.98	2.14	1.97	2.19	2.00	2.12
4			0.60	0.20	5.48	6.28	6.05	5.06	6.17	4.87	6.09	5.03	6.28	4.96
5				0.60	5.28	6.12	3.07	3.56	2.93	3.49	2.84	3.47	2.70	3.37
6				0.95	10.41	8.45	1.89	1.44	2.02	1.47	1.55	1.26	1.57	1.22
7			0.95	0.20	5.82	5.36	3.91	3.64	3.99	3.44	3.90	3.53	4.03	3.47
8				0.60	5.19	6.05	1.95	1.91	1.71	1.94	1.78	1.81	1.60	1.76
9				0.95	7.33	7.35	0.56	0.48	0.57	0.51	0.60	0.35	0.60	0.36

cont...../-

Table A 11.4 (contd.)

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
10	0.40	0.25	0.20	0.20	2.81	2.06	2.85	1.94	2.70	1.79	2.88	1.93	2.89	1.92
11				0.60	2.33	1.88	1.33	1.10	1.57	1.04	1.29	1.06	1.27	1.06
12				0.95	2.65	1.89	0.90	0.59	0.91	0.58	0.84	0.56	0.80	0.53
13			0.60	0.20	2.60	2.17	2.48	1.86	2.29	1.75	2.57	1.83	2.87	1.83
14				0.60	2.17	2.45	1.60	1.21	1.26	1.26	1.71	1.16	1.82	1.13
15				0.95	1.88	2.16	0.34	0.35	0.30	0.38	0.35	0.33	0.34	0.32
16			0.95	0.20	3.30	2.35	2.36	1.64	2.26	1.60	2.34	1.61	2.41	1.59
17				0.60	2.02	1.97	0.95	0.86	0.89	0.82	0.94	0.84	0.92	0.81
18				0.95	2.51	2.61	0.14	0.17	0.15	0.19	0.14	0.14	0.12	0.14

contd..... /-

Table A 11.4 (contd.)

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
19	0.40	0.35	0.20	0.20	1.28	1.08	1.19	0.94	1.26	0.94	1.18	0.93	1.19	0.93
20				0.60	1.49	1.24	0.90	0.72	0.86	0.68	0.91	0.70	0.89	0.69
21				0.95	1.24	1.19	0.18	0.20	0.19	0.20	0.18	0.18	0.18	0.18
22			0.60	0.20	0.93	1.29	1.00	1.24	0.95	1.13	1.01	1.22	1.06	1.22
23				0.60	1.67	1.40	0.81	0.71	0.88	0.71	0.75	0.69	0.69	0.67
24				0.95	1.13	1.15	0.12	0.14	0.11	0.15	0.11	0.12	0.11	0.12
25			0.95	0.20	1.42	1.36	1.50	1.14	1.36	1.08	1.56	1.13	1.64	1.13
26				0.60	1.11	1.14	0.67	0.59	0.52	0.59	0.76	0.56	0.68	0.55
27				0.95	1.18	1.70	0.07	0.08	0.08	0.10	0.07	0.07	0.07	0.07

contd..... /-

Table A 11.4 (contd)

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
28	0.70	0.10	0.20	0.20	5.82	6.52	7.26	6.06	6.52	5.57	7.42	5.93	7.82	5.97
29				0.60	4.85	5.29	5.01	4.76	4.40	4.39	5.17	4.73	5.37	4.67
30				0.95	5.28	6.00	4.41	4.25	3.95	3.98	4.44	4.16	4.53	4.13
31			0.60	0.20	6.37	5.39	5.45	4.06	5.02	3.69	5.45	3.99	5.61	3.96
32				0.60	8.52	5.61	4.86	3.11	4.70	2.90	4.66	2.95	4.75	2.90
33				0.95	7.23	6.29	2.35	2.24	2.24	2.14	2.07	2.11	2.10	2.08
34			0.95	0.20	7.05	5.52	3.58	2.67	3.51	2.63	3.18	2.52	2.89	2.46
35				0.60	5.46	6.05	1.81	1.33	1.66	1.30	1.80	1.20	1.79	1.20
36				0.95	4.84	8.76	0.41	0.44	0.38	0.46	0.37	0.38	0.37	0.39

contd..... /-

Table A 11.4 (contd.)

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
37	0.70	0.35	0.20	0.20	1.38	1.38	1.23	1.29	1.34	1.21	1.19	1.28	1.14	1.28
38				0.60	0.97	1.32	0.82	0.95	0.73	0.92	0.92	0.94	0.88	0.93
39				0.95	1.20	1.31	0.76	0.71	0.75	0.67	0.76	0.70	0.75	0.69
40			0.60	0.20	1.55	1.28	1.39	0.88	1.31	0.84	1.41	0.87	1.42	0.86
41				0.60	2.30	1.41	1.26	0.80	1.28	0.79	1.20	0.78	1.23	0.75
42				0.95	1.30	1.27	0.46	0.50	0.36	0.51	0.34	0.46	0.35	0.45
43			0.95	0.20	1.07	1.54	0.98	0.78	0.77	0.76	0.98	0.76	1.01	0.76
44				0.60	1.02	1.61	0.26	0.35	0.20	0.34	0.27	0.32	0.29	0.32
45				0.95	1.84	1.24	0.14	0.11	0.16	0.12	0.08	0.08	0.08	0.08

contd..... /-

Table A 11.4 (contd.)

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
46	0.70	0.65	0.20	0.20	0.31	0.35	0.33	0.31	0.34	0.30	0.35	0.30	0.34	0.30
47				0.60	0.29	0.35	0.12	0.18	0.12	0.18	0.11	0.17	0.10	0.17
48				0.95	0.21	0.32	0.06	0.07	0.06	0.06	0.06	0.06	0.06	0.06
49			0.60	0.20	0.28	0.32	0.29	0.30	0.27	0.27	0.31	0.29	0.33	0.29
50				0.60	0.34	0.34	0.20	0.19	0.17	0.18	0.19	0.18	0.18	0.17
51				0.95	0.30	0.27	0.06	0.05	0.06	0.05	0.06	0.05	0.06	0.05
52			0.95	0.20	0.37	0.41	0.25	0.31	0.26	0.29	0.24	0.31	0.24	0.31
53				0.60	0.40	0.37	0.15	0.19	0.21	0.18	0.14	0.18	0.13	0.17
54				0.95	0.48	0.39	0.03	0.03	0.03	0.03	0.01	0.02	0.01	0.02

contd..../-

Table A 11.4 (contd.)

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
55	0.95	0.10	0.20	0.20	4.98	5.45	6.09	4.79	5.48	4.48	6.24	4.71	6.64	4.67
56				0.60	6.44	5.89	6.48	5.30	6.61	5.04	6.54	5.30	6.59	5.23
57				0.95	5.16	6.67	5.89	5.47	5.48	5.24	5.98	5.42	6.45	5.35
58			0.60	0.20	5.60	6.38	3.83	3.20	3.60	2.99	3.72	3.08	3.60	3.08
59				0.60	6.00	6.32	4.69	3.46	4.30	3.35	4.68	3.38	4.70	3.35
60				0.95	6.83	6.26	2.89	3.35	3.01	3.28	2.97	3.24	2.91	3.15
61			0.95	0.20	6.68	6.37	0.75	0.67	0.88	0.71	0.70	0.58	0.69	0.57
62				0.60	6.84	6.38	0.64	0.55	0.67	0.58	0.59	0.47	0.60	0.46
63				0.95	4.69	6.40	0.47	0.48	0.46	0.56	0.40	0.38	0.41	0.39

contd..... /-

Table A 11.4 (contd.)

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
64	0.95	0.65	0.20	0.20	0.29	0.33	0.29	0.30	0.32	0.29	0.29	0.30	0.30	0.29
65				0.60	0.30	0.39	0.32	0.34	0.29	0.32	0.31	0.34	0.33	0.34
66				0.95	0.36	0.37	0.35	0.30	0.35	0.27	0.36	0.29	0.36	0.29
67			0.60	0.20	0.30	0.44	0.24	0.27	0.23	0.25	0.23	0.26	0.24	0.26
68				0.60	0.30	0.41	0.20	0.21	0.20	0.21	0.20	0.21	0.20	0.20
69				0.95	0.19	0.37	0.13	0.16	0.13	0.15	0.13	0.15	0.13	0.15
70			0.95	0.20	0.31	0.44	0.07	0.08	0.06	0.08	0.06	0.07	0.06	0.07
71				0.60	0.34	0.33	0.07	0.05	0.07	0.05	0.06	0.05	0.06	0.05
72				0.95	0.30	0.38	0.03	0.02	0.03	0.02	0.02	0.02	0.02	0.02

contd. ..... /-

Table A 11.4 (contd.)

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
73	0.95	0.90	0.20	0.20	0.06	0.07	0.05	0.06	0.05	0.05	0.05	0.05	0.05	0.05
74				0.60	0.05	0.08	0.05	0.06	0.04	0.05	0.05	0.05	0.05	0.05
75				0.95	0.12	0.08	0.07	0.04	0.06	0.04	0.05	0.04	0.05	0.04
76			0.60	0.20	0.07	0.07	0.07	0.05	0.07	0.05	0.07	0.05	0.07	0.05
77				0.60	0.09	0.07	0.07	0.04	0.07	0.04	0.06	0.04	0.06	0.04
78				0.95	0.05	0.08	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02
79			0.95	0.20	0.07	0.07	0.05	0.04	0.04	0.04	0.05	0.04	0.05	0.04
80				0.60	0.06	0.09	0.02	0.02	0.03	0.02	0.02	0.02	0.02	0.02
81				0.95	0.07	0.07	0.01	0.01	0.01	0.01	0.01	0.00	0.01	0.00

References

1. Abrahamse, A.P.J., and J. Koerts (1969) : "A Comparison of the Power of the Durbin-Watson Test and the BLUS Test," Journal of the American Statistical Association, 64, pp. 938-948.

2. _____ and _____(1971) : "New Estimates of Disturbances in Regression Analysis," Journal of the American Statistical Association, 66, pp.71-74.

3. _____ and A. S. Louter (1971) : "On a New Test for Autocorrelation in Least Squares Regression," Biometrika,58, pp. 53-60.

4. Amemiya, T. (1973a) : "Regression Analysis When the Dependent Variable is Truncated Normal," Econometrica, 41, pp. 997-1016.

5. _____ (1973b) : "Regression Analysis When the Variance of of the Dependent Variable is Proportional to the Square of Its Expectation," Journal of the American Ststistical Association, 68, pp. 928-934.

6. _____ (1973c) : "Generalized Least Squares with an Estimated Autocovariance Matrix," Econometrica, 41, pp.723-732.

7. Anderson, T. W. (1971) : The Statistical Analysis of Time Series, London, John Wiley.

8. Balestra, P. (1980) : "A Note on the Exact Transformation Associated With the First-Order Moving Average Process," Journal of Econometrics, 14, pp. 381-394.

9. Bard, Y. (1974) : Nonlinear Parameter Estimation, New York: Academic Press.

10. Bartlett, M. S. (1946) : "On the Theoretical Specification of Sampling Properties of Autocorrelated Time Series", Journal of the Royal Statistical Society, Series B, 8, pp.27-41.

11. Basmann, R. L. (1960) : "On Finite Sample Distributions of Generalised Classical Linear Indentifiability Test Statistics", Journal of the American Statistical Association, 55, pp.650-659.

12. Beach, C. M., and J. G. MacKinnon (1978a) : "A Maximum Likelihood Procedure for Regression with Autocorrelated Errors", Econometrica, 46, pp.51-58.

13. _____ and _____(1978b) : "Full Maximum Likelihood Estimation of Second-Order Autoregressive Error Models", Journal of Econometrics, 7, pp.187-198.

14. Blattberg, R. C. (1973) : "Evaluation of the Power of the Durbin-Watson Statistic for Non-First Order Serial Correlation Alternatives," The Review of Economics and Statistics, 55, pp. 508-515.

15. Box, G. E. P., and D. R. Cox (1964) : "An Analysis of Transformations," Journal of the Royal Statistical Society, Series B, 26, pp. 211-243.

16. _____ and G. M. Jenkins (1976) : Time Series Analysis, Forecasting and Control, Revised Edition, San Francisco: Holden-Day.

17. _____ and M. E. Muller (1958) : "A Note on the Generation of Random Normal Deviates," Annals of Mathematical Statistics, 29, pp. 601-611.

18. _____ and P. W. Tidwell (1962) : "Transformation of the Independent Variables", Technometrics, 4, pp.531-550.

19. Byron, R. P. (1972) : "Testing for Misspecification in Econometric Systems Using Full Information", International Economic Review, 13, pp. 745-756.

20. Chang, H. S. (1977) : "Functional Form and the Demand for Meat in the United States", Review of Economics and Statistics, 59, pp.355-359.

21. Chaudhuri (Mukherjee), M. (1977) : "Autocorrelated Disturbances in the Light of Specification Analysis", Journal of Econometrics, 5, pp. 301-313.

22. _____ (1979) : Some Problems on Econometric Regression Analysis, Unpublished Ph. D. thesis, Indian Statistical Institute, Calcutta, India.

23. Chipman, J. S. (1979) : "Efficiency of Least Squares Estimation of Linear Trend When Residuals Are Auto-correlated", Econometrica, 47, pp.115-128.

24. Cochrane, D., and G. H. Orcutt (1949) : "Application of Least Squares Regressions to Relationships Containing Autocorrelated Error Terms," Journal of the American Statistical Association, 44, pp.32-61.

25. Cochran, W. C. (1977) : Sampling Technique, 2nd edition, New Delhi : Wiley Eastern.

26. Coondoo, D., and N. Sarkar (1979) : "Regression Analysis with Heteroscedastic Disturbances", Technical Report No.ERU/9/79, Indian Statistical Institute, Calcutta, India.

27. Cox, D., and D. V. Hinkley (1974) : Theoretical Statistics, London: Chapman and Hill.

28. Cox, D. R., and E. J. Snell (1968) : "A General Definition of Residuals," Journal of the Royal Statistical Society, Series B, 30, pp.248-265.

29. Cragg, J. G. (1967) : "On the Relative Small Sample Properties of Several Structural Equation Estimators," Econometrica, 35, pp. 89-100.

30. _____ (1971) : "Some Statistical Models for Limited Dependent Variables with Applications to the Demand for Durable Goods", Econometrica, 39, pp. 828-844.

31. Dent, W., and A. S. Min (1978) : "A Monte Carlo Study of Autoregressive Integrated Moving Average Processes", Journal of Econometrics, 7, pp.23-55.

32. _____ and S. Cassing (1978) : "On Durbin's and Sims' Residuals in Autocorrelation Tests", Econometrica, 46, pp. 1489-1492.

33. Dhrymes, P. J. (1970) : Econometrics : Statistical Foundations and Applications, New York: Harper and Row.

34. _____ and M. Kurz (1964) : "Technology and Scale in Electricity Generation", Econometrica, 32, pp.287-315.

35. Doran, H. E. (1979) : "Omission of the Weighted First Observation in an Autocorrelated Regression Model : A Discussion on Loss of Efficiency", Working Papers in Econometrics and Applied Statistics No. 6, University of New England, Armidale, Australia.

36. Draper, N. R., and D. R. Cox (1969) : "On Distributions and Their Transformation to Normality", Journal of the Royal Statistical Society, Series B, 31, pp.472-476.

37. Dubbelman, C., A. S. Louter and A. P. J. Abrahamse (1978) : "On Typical Characteristics of Economic Time Series and the Relative Qualities of Five Autocorrelated Tests", Journal of Econometrics, 8, pp. 295-306.

38. Durbin, J. (1959) : "Efficient Estimation of Parameters in Moving Average Models", Biometrika, 46, pp.306-316.

39. _____ (1960) : "Estimation of Parameters in Time-Series Regression Models", Journal of the Royal Statistical Society, Series B, 22, pp. 139-153.

40. _____ (1969) : "Tests for Serial Correlation in Regression Analysis Based on the Periodogram of Least Squares Residuals", Biometrika, 56, pp.1-15.

41. _____ (1970a) : "Testing for Serial Correlation in Least-Squares Regression When Some of the Regressors Are Lagged Dependent Variables", Econometrica, 38, pp. 410-421.

42. _____ (1970b) : "An Alternative to the Bounds Test for Testing for Serial Correlation in Least Squares Regression", Econometrica, 38, pp. 422-429.

43. _____ and G. S. Watson (1950) : "Testing for Serial Correlation in Least Squares Regression I", Biometrika, 37, pp.409-428.

44. _____ and _____ (1951) : "Testing for Serial Correlation in Least Squares Regression II ", Biometrika, 38, pp. 159-178.

45. _____ and _____ (1971) : "Testing for Serial Correlation in Least Squares Regression III," Biometrika, 58, pp.1-42.

46. Egy, D., and K. Lahiri (1978) : Simultaneous Estimation and Tests for Functional Form and Heteroscedasticity, Paper presented at the Chicago Meetings of the Econometric Society.

47. Feigl, P., and M. Zelen (1965) : "Estimation of Exponential Survival Probabilities with Concomitant Information", Biometrics, 21, pp. 826-838.

48. Feller, W. (1972) : An Introduction to Probability Theory and Its Applications, Vol. 1, New Delhi, India: Wiley Eastern.

49. Fisher, G. R. (1970) : "Quarterly Dividend Behaviour" in The Econometric Study of the United Kingdom (eds. K. Hilton and D. F. Heathfield), London: Macmillan, pp. 149-184.

50. Fuller, W. A. (1976) : Introduction to Statistical Time Series, New York: Wiley.

51. Geary, R. C. (1970) : "Relative Efficiency of Count of Sign Changes for Assessing Residual Autoregression in Least Squares Regression", Biometrika, 57, pp. 123-127.

52. Gilbert, R.F. (1969) : The Demand for Money : an Analysis of Specification Error, Unpublished Ph. D. Thesis, Michigan State University, East Lansing, Michigan.

53. Godfrey, L. G. (1978a) : "Testing Against General Autoregressive and Moving Average Error Models When the Regressors Include Lagged Dependent Variables", Econometrica, 46, pp. 1293-1302.

54. _____ (1978b) : "Testing for Higher Order Serial Correlation in Regression Equations When the Regressors Include Lagged Dependent Variables", Econometrica, 46, pp. 1303-1310.

55. Goldberger, A.S. (1963) : Econometric Theory, John Wiley and Sons.

56. Goldfeld, S. M., and R. E. Quandt (1972) : Nonlinear Methods of Econometrics, Amsterdam: North-Holland.

57. Granger, C.W.J., and M. J. Morris (1976) : "Time Series Modelling and Interpretation", Journal of the Royal Statistical Society, Series A, 139, pp. 246-257.

58. Gupta, Y. P., and E. Maasoumi (1979) : "Omitted Variables, Variability of Estimated Parameters and the Appearance of Autocorrelated Disturbances", Journal of Econometrics, 9, pp. 379-385.

59. Hannan, E. J. (1957) : "Testing for Serial Correlation in Least Squares Regression", Biometrika, 44, pp. 57-66.

60. _____ (1969) : "The Estimation of Mixed Moving Average Autoregressive Systems", Biometrika, 56, pp. 579-593.

61. _____ and R. D. Terrell (1968) : "Testing for Serial Correlation After Least Squares Regression", Econometrica, 36, pp. 133-150.

62. Hastings, C., J. T. Hayward and J. P. Wong (1955) : Approximations for Digital Computers, Princeton University Press.

63. Hausman, J. A., and D. A. White (1972) : "Social Experimentation, Truncated Distributions and Efficient Estimation," Econometrica, 45, pp. 919-938.

64. Hendry, D. F., and P. K. Trivedi (1972) : "Maximum Likelihood Estimation of Difference Equations with Moving Average Errors : A Simulation Study", Review of Economic Studies, 39, pp. 117-146.

65. Henshaw, R. C. (1966) : "Testing Single-Equation Least-Squares Regression Models for Autocorrelated Disturbances", Econometrica, 34, pp. 646-660.

66. Hildreth, C. and J. Y. Lu (1960) : Demand Relations with Autocorrelated Disturbances, Michigan State University Agricultural Experiment Station Technical Bulletin 276, East Lansing, Mich.

67. _____ and _____ (1969) : A Monte Carlo Study of the Regression Model with Autoregressive Disturbances, Rand Memorandom RM 5728PR.

68. _____ and W. Dent (1974) : "An Adjusted Maximum Likelihood Estimator" in Econometrics and Economic Theory : Essays in Honor of Jan Tinbergen (ed. W. Sellekaert), London: Macmillan, pp.3-25.

69. Imhof, J. P. (1961) : "Computing the Distribution of Quadratic Forms in Normal Variables", Biometrika, 48, pp. 419-426.

70. Jennrich, R. I. (1969) : "Asymptotic Properties of Non-linear Least Squares Estimators", Annals of Mathematical Statistics, 40, pp.633-643.

71. Johnston, J. (1972) : Econometric Methods, 2nd edition, McGraw-Hill Book Company.

72. Judge, G. G. W. E. Griffiths, R. C. Hill and T. C. Lee (1980) : The Theory and Practice of Econometrics, New York: John Wiley and Sons.

73. Kadiyala, K. R. (1968) : "A Transformation Used to Circumvent the Problem of Autocorrelation", Econometrica, 36, pp. 93-96.

74. Kau, J. B., and C. F. Lee (1976) : "The Functional Form in Estimating the Density Gradient : An Empirical Investigation", Journal of the American Statistical Association, 71, pp. 326-327.

75. Kendall, M. G., and A. Stuart (1968) : The Advanced Theory of Statistics, Vol. 3, Design and Analysis, and Time Series, 2nd. ed., London: Griffin.

76. Kenkel, J. L. (1974) : "Some Small Sample Properties of Durbin's Tests for Serial Correlation in Regression Models Containing Lagged Dependent Variables", Econometrica, 42, pp. 763-769.

77. _____ (1975) : "Small Sample Tests for Serial Correlation in Models Containing Lagged Dependent Variables", Review of Economics and Statistics, 57, pp. 383-386.

78. _____ (1976) : "Comment on the Small Sample Power of Durbin's h test," Journal of the American Statistical Association, 71, pp.96-97.

79. Kenward, L. R. (1975) : "Autocorrelation and Dynamic Methodology with an Application to Wage Determination Models", Journal of Econometrics, 3, pp. 179-187.

80. Kloek, T. (1975) : "Note on a Large Sample Result in Specification Analysis", Econometrica, 43, pp. 933-937.

81. Kmenta, J. (1971) : Elements of Econometrics, New York: Macmillan.

82. _____ and R. F. Gilbert (1970) : "Small Sample Properties of Alternative Estimators of Seemingly Unrelated Regressions," Journal of the American Statistical Association, 65, pp.186-197.

83. Koerts, J. and A. P. J. Abrahamse (1969) : On the Theory and Application of the General Linear Model, Rotterdam: Rotterdam University Press.

84. _____ and _____(1970) : "The Correlation Coefficient in the General Linear Model", European Economic Review, 1, pp. 401-427.

85. Lahiri, K., and D. Egy (1981) : "Joint Estimation and Testing for Functional Form and Heteroscedasticity," Journal of Econometrics, 15, pp. 299-308.

86. Lee, K. W. (1972) : An International Study of Manufacturing Production Functions, Unpublished Ph. D. Thesis, Mechigan State University, East Lansing, Michigan.

87. L'Esperance, W. L., and D. Taylor (1975) : "The Power of Four Tests of Autocorrelation in the Linear Regression Model," , Journal of Econometrics, 3, pp. 1-21.

88. _____, D. Chall and D. Taylor (1976) : "An Algorithm for Determining the Distribution Function of the Durbin-Watson Statistic," Econometrica, 44, pp.1325-1346.

89. Loeb, P. D. (1976) : "Specification Error Tests and Investment Functions", Econometrica, 44, pp.185-194.

90. McClave, J. T. (1973) : "On the Bias of Autoregressive Approximations to Moving Averages", Biometrika, 60, pp. 599-605.

91. _____ (1974) : "A Comparison of Moving Average Estimation Procedures", Communication in Statistics, 3, pp. 865-883.

92. Maddala, G. S. (1977) : Econometrics, New York: McGraw-Hill.

93. _____ and A. S. Rao (1973) : "Tests for Serial Correlation in Regression Models with Lagged Dependent Variables and Serially Correlated Errors", Econometrica, 41, pp. 761-774.

94. Maeshiro, A. (1979) : "On the Retention of the First Observations in Serial Correlation Adjustment of Regression Models", International Economic Review, 20, pp. 259-265.

95. Magnus, J. R. (1978) : "Maximum Likelihood Estimation of the GLS Model with Unknown Parameters in the Disturbance Covariance Matrix", Journal of Econometrics, 7, pp. 281-312.

96. Malinvaud, E. (1980) : Statistical Methods in Econometrics, Third Revised Edition, Amsterdam : North Holland.

97. Mentz, R. P. (1977) : "Estimation in the First-Order Moving Average Model Through the Finite Autoregressive Approximation : Some Asymptotic Results", Journal of Econometrics, 6, pp. 225-236.

98. Mizon, G. E. (1977) : Inferential Procedures in Nonlinear Models : an Application in a UK Industrial Cross Section Study of Factor Substitution and Returns to Scale, Econometrica, 45, pp. 1221-1242.

99. Moran, P.A.P. (1970) : "On Asymptotically Optimal Tests of Composite Hypotheses", Biometrika, 57, pp. 47-55.

100. Mukerji, V. (1963) : "A Generalized SMAC Function with Constant Ratios of Elasticities of Substitution", Review of Economic Studies, 30, pp.233-236.

101. Nagar, A. L. (1960) : "A Monte Carlo Study of Alternative Simultaneous Equation Estimators", Econometrica, 28, pp. 573-590.

102. Nelson, C. R. (1974) : "The First-Order Moving Average Process", Journal of Econometrics, 2, pp. 121-141.

103. Newbold, P. (1974) : "The Exact Likelihood for a Mixed ARMA Process", Biometrika, 61, pp. 423-426.

104. Nicholls, D. F., A. R. Ragan, and R. D. Terrell (1975) : "The Estimation and Use of Models with Moving Average Disturbance Term : A Survey", International Economic Review, 16, pp.113-134.

105. Oberhofer, W., and J. Kmenta (1974) : "A General Procedure for Obtaining Maximum Likelihood Estimates in Generalised Regression Models", Econometrica, 42, pp.579-590.

106. O'Hara, J. B., and R. C. McClelland (1964) : Statistics in Accounting and Business, New York : Holt, Rinehart and Winston.

107. Osbern, D. R., (1976) : "Maximum Likelihood Estimation of Moving Average Processes", Annals of Economics and Social Measurement, 3, pp. 75-87.

108. Oslen, D. R. (1978) : "Note on the Uniqueness of the Maximum Likelihood Estimator for the Tobit Model", Econometrica, 46, pp. 1211-1215.

109. Pagan, A. (1973) : "Efficient Estimation of Models with Composite Disturbance Terms, Journal of Econometrics, 1, pp. 329-340.

110. _____(1974) : "A Generalized Approach to the Treatment of Autocorrelation", Australian Economic Papers, 13, pp. 267-280.

111. Pagan, A. R., and D. P. Nicholls (1976) : "Exact Maximum Likelihood Estimation of Regression Models with Finite Order Moving Average Errors", Review of Economic Studies, 43, pp. 383-388.

112. _____ and R. P. Byron (1978) : "A Synthetic Approach to the Estimation of Models with Autocorrelated Disturbance Terms" in Stability and Inflation (eds. A. R. Bergstrom, et al.), New York: Wiley.

113. Pan Jei-jian (1968) : "Distributions of the Noncircular Serial Correlation Coefficients," Selected Translations in Mathematical Statistics, 7, Institute of Mathematical Statistics and the American Mathematical Society.

114. Park, R. E., and B. M. Mitchell (1980) : "Estimating the Autocorrelated Error Model with Trended Data", Journal of Econometrics, 13, pp. 185-202.

115. Park, S. (1975) : "On the Small Sample Power of Durbin's h test," Journal of the American Statistical Association, 70, pp.60-63.

116. _____ (1976) : "Rejoinder", Journal of the American Statistical Association, 71, pp. 97-98.

117. Pesaran, M. H. (1973) : "Exact Maximum Likelihood Estimation of a Regression Equation with a First-Order Moving-Average Error", Review of Economic Studies, 40, pp. 529-536.

118. Phillips, A. W. (1966) : "The Estimation of Systems of Difference Equations with Moving Average Disturbances", Econometric Society Meeting, San Francisco, reprinted in Stability and Inflation (eds., A. E. Bergstrom et al.), New York: Wiley.

119. Phillips, G.D.A., and A. C. Harvey (1974) : "A Simple Test for Serial Correlation in Regression Analysis", Journal of the American Statistical Association, 69, pp. 935-939.

120. Pierce, D. A. (1971a) : "Distribution of Residual Autocorrelations in the Regression Model with Autoregressive-Moving Average Errors", Journal of the Royal Statistical Society, Series B, 33, pp. 140-146.

121. _____ (1971b) : "Least Squares Estimation in the Regression Model with Autoregressive-Moving Average Errors", Biometrika, 58, pp. 299-312.

122. _____ (1972) : "Least Squares Estimation in Dynamic-Dirturbance Time Series Models", Biometrika, 59, pp. 73-78.

123. Poirier, D. J. (1978a) : "Box-Cox Transformation in Limited Dependent Variable Models", Journal of the American Statistical Association, 73, pp. 284-289.

124. _____ (1978b) : "The Effect of the First Observation in Regression Models with First-Order Autoregressive Disturbances", Applied Statistics, 27, pp. 67-68.

125. Prais, S. J., and C. B. Winsten (1954) : "Trend Estimators and Serial Correlation", Chicago: Cowles Commission Discussion Paper No. 383.

126. Quandt, R. E. (1965) : "On Certain Small Sample Properties of k-Class Estimators", International Economic Review, 6, pp. 92-104.

127. Ramsey, J. B. (1969) : "Tests for Specification Errors in Classical Linear Least Squares Regression Analysis", Journal of the Royal Statistical Society, Series B, 31, pp. 350-371.

128. _____ (1974) : "Classical Model Selection Through Specification Error Tests" in Frontiers in Econometrics (ed. P. Zarembka), New York: Academic Press, pp. 13-47.

129. _____ and P. Schmidt (1976) : "Some Further Results on the Use of OLS and BLUS Residuals in Specification Error Tests", Journal of the American Statistical Association, 71, pp. 389-390.

130. _____ and P. Zarembka (1971) : "Specification Error Tests and Alternative Functional Form of Aggregate Production Function", Journal of the American Statistical Association, 66, pp. 471-477.

131. _____ and R. Gilbert (1972) : "Some Small Sample Properties of Tests for Specification Error", Journal of the American Statistical Association, 67, pp. 180-186.

132. Rao, C. R. (1974) : Linear Statistical Inference and Its Applications, 2nd Edition, New Delhi, India : Wiley Eastern.

133. Rao, P., and Z. Griliches (1969) : "Small Sample Properties of Several Two-Stage Regression Methods in The Context of Auto-correlated Errors", Journal of the American Statistical Association, 64, pp. 253-272.

134. Rose, D. E. (1977) : "Forecasting Aggregates of Independent ARIMA Processes", Journal of Econometrics, 5, pp. 323-345.

135. Rothenberg, T.J.(1971) : "Identification in Parametric Models", Econometrica, 39, pp. 577-591.

136. _____ and C. T. Leenders (1964) : "Efficient Estimation of Simultaneous Equation Systems, Econometrica, 32, pp.57-76.

137. Rutemiller, H., and D. A. Bowers (1968) : "Estimation in a Heteroscedastic Regression Model", Journal of the American Statistical Association, 63, pp. 552-557.

138. Sargan, J. D. (1964) : "Wages and Prices in the United Kingdom : A Study in Econometric Methodology" in Econometric Analysis for National Economic Planning (eds., P. E. Hart, G. Mills, and J. K. Whitaker), London : Butterworths, pp. 25-54.

139. Sarin, N. E., and K. J. White (1978) : "Estimation and Testing for Functional Form and Autocorrelation : A Simultaneous Approach", Journal of Econometrics, 8, pp. 1-12.

140. Sawa, T. (1978) : "The Exact Moments of the Least Squares Estimator for the Autoregressive Model", Journal of Econometrics, 8, pp.159-172.

141. Schlesselman, J. (1971) : "Power Families : A Note on the Box and Cox Transformation", Journal of the Royal Statistical Society, Series B, 33, pp.307-311.

142. Schmidt, P. J. (1971) : "Estimation of a Distributed Lag Model with Second-Order Autoregressive Disturbances : A Monte Carlo Experiment", International Economic Review, 12, pp. 372-380.

143. Schmidt, P. (1974) : "A Note on Theil's Minimum Standard Error Criterion When the Disturbances Are Autocorrelated", Review of Economics and Statistics, 56, pp. 122-123.

144. Sims, C. A. (1975) : "A Note on Exact Tests for Serial Correlation", Journal of the American Statistical Association, 70, pp. 162-165.

145. Smith, V. K. (1976) : "The Estimated Power of Several Tests for Autocorrelation with Non-First-Order Alternatives", Journal of the American Statistical Association, 71, pp.879-883.

146. Smith, V. K., and T. W. Hall (1972) : "A Comparison of Maximum Likelihood Versus BLUE Estimators", Review of Economics and and Statistics, 54, pp. 186-190.

147. Spencer, B. G. (1975) : "The Small Sample Bias of Durbin's Tests for Serial Correlation When One of the Regressors Is the Lagged Dependent Variable and the Null Hypothesis Is True", Journal of Econometrics, 3, pp. 249-254.

148. Spitzer, J. J. (1976) : "The Demand for Money, the Liquidity Trap, and Functional Forms", International Economic Review, 17, pp. 220-227.

149. _____ (1977a) : "A Simultaneous Equations System of Money Demand and Supply Using Generalized Functional Forms", Journal of Econometrics, 5, pp.117-128.

150. _____ (1977b) : "Coefficient Variance Estimates and Hypothesis Testing in Models Utilizing the Box-Cox Transformation", unpublished, Department of Economics, State University of New York, Brockport.

151. _____ (1978) : "A Monte Carlo Investigation of the Box-Cox Transformation in Small Samples", Journal of the American Statistical Association, 73, pp. 488-495.

152. ____ (1979) : "Small-Sample Properties of Nonlinear Least Squares and Maximum Likelihood Estimators in the Context of Autocorrelated Errors," Journal of the American Statistical Association, 74, pp.41-47.

153. Summers, R. (1965) : "A Capital Intensive Approach to the Small Sample Properties of Various Simultaneous Equation Estimators", Econometrica, 33, pp. 1-41.

154. Theil, H. (1957) : "Specification Errors and the Estimation of Economic Relationships", Review of the International Statistical Institute, 25, pp. 41-51.

155. ____ (1965) : "The Analysis of Disturbances in Regression Analysis", Journal of the American Statistical Association, 60, pp. 1067-1079.

156. ____ (1968) : "A Simplification of the BLUS Procedure for Analysing Regression Disturbances", Journal of the American Statistical Association, 63, pp. 242-251.

157. ____ (1971) : Principles of Econometrics, New York: Wiley.

158. ____ and A. L. Nagar (1961) : "Testing the Independence of Regression Disturbances", Journal of the American Statistical Association, 56, pp.793-806.

159. Thomas, J. J., and K. F. Wallis (1971) : "Seasonal Variation in Regression Analysis", Journal of the Royal Statistical Society, Series A, 134, pp. 67-72.

160. Tiao, G. C., and M. M. Ali (1971) : "Analysis of Correlated Random Effects : Linear Model with Two Random Components", Biometrika, 58, pp. 37-51.

161. Tillman, J. A. (1975) : "The Power of the Durbin-Watson Test", Econometrica, 43, pp. 959-973.

162. Tintner, G., and G. Kadekodi (1971) : "Note on the Transformation of Variables in Simultaneous Equations Systems", Journal of Indian Society for Agricultural Statistics, 23, pp. 163-173.

163. Tobin, J. (1958) : "Estimation of Relationships for Limited Dependent Variables", Econometrica, 26, pp. 24-36.

164. Tocher, K. D. (1963) : The Art of Simulation, The English University Press Ltd.

165. Trivedi, P. K. (1970) : "Inventory Behaviour in U. K. Manufacturing, 1956-67", Review of Economic Studies, 37, pp. 517-536.

166. Tukey, J. W. (1957) : "On the Comparative Anatomy of Transformations", Annals of Mathematical Statistics, 28, pp. 602-632.

167. Turner, M. E., R. J. Monroe and H. L. Lucas (1961) : "Generalized Asymptotic Regression and Non-Linear Path Analysis", Biometrics, 17, pp.120-143.

168. Wagner, H. (1958) : "A Monte Carlo Study of Estimates of Simultaneous Linear Structural Equations", Econometrica, 26, pp. 117-133.

169. Wald, A. (1949) : "Note on the Consistency of Maximum Likelihood Estimate", Annals of Mathematical Statistics, 20, pp. 595-601.

170. Walker, A. M. (1961) : "Large Sample Estimation of Parameters for Moving Average Models", Biometrika, 48, pp. 343-357.

171. Wallis, K. F. (1972) : "Testing for Fourth-Order Autocorrelation in Quarterly Regression Equations", Econometrica, 40, pp. 617-636.

172. Weiss, L., and J. Wolfowitz (1967) : "Maximum Probability Estimators", Annals of the Institute of Statistical Mathematics, 19. pp. 193-206.

173. White, K. J. (1972) : "Estimation of the Liquidity Trap With a Generalized Functional Form", Econometrica, 40, pp. 193-199.

174. Wise, J. (1955) : "Autocorrelation Function and the Spectoral Density Function", Biometrika, 42, pp. 151-159.

175. Wold, H. (1954) : Random Normal Deviates, Tracts for Computers, XXV, Cambridge University Press.

176. Yancy, T. A., M. Bock and G. G. Judge (1972) : "Some Finite Sample Results for Theil's Mixed Regression Estimator", Journal of the American Statistical Association, 67, pp. 176-179.

177. Zarembka, P. (1968) : "Functional Form in the Demand for Money", Journal of the American Statistical Association, 63, pp. 502-511.

178. _____ (1974) : "Transformation of Variables in Econometrics" in Frontiers in Econometrics (ed. P. Zarembka), New York : Academic Press, pp. 81-104.