



National Library  
of Canada

Bibliothèque nationale  
du Canada

Canadian Theses Service Services des thèses canadiennes

Ottawa, Canada  
K1A 0N4

## CANADIAN THESES

## THÈSES CANADIENNES

### NOTICE

The quality of this microfiche is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Previously copyrighted materials (journal articles, published tests, etc.) are not filmed.

Reproduction in full or in part of this film is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30.

### AVIS

La qualité de cette microfiche dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

Les documents qui font déjà l'objet d'un droit d'auteur (articles de revue, examens publiés, etc.) ne sont pas microfilmés.

La reproduction, même partielle, de ce microfilm est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30.

**THIS DISSERTATION  
HAS BEEN MICROFILMED  
EXACTLY AS RECEIVED**

**LA THÈSE A ÉTÉ  
MICROFILMÉE TELLE QUE  
NOUS L'AVONS REÇUE**

REGRESSION ANALYSIS PROCEDURES WITH HIGHER ORDER  
MOVING AVERAGE ERRORS

by

Askar Hassan Choudhury

B. Sc. (Hons.), Jahangirnagar University, 1974

M. Sc., Jahangirnagar University, 1975

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF  
THE REQUIREMENTS FOR THE DEGREE OF  
MASTER OF SCIENCE  
in the Department  
of  
Mathematics and Statistics

© Askar Hassan Choudhury 1985

SIMON FRASER UNIVERSITY

July, 1985

All rights reserved. This thesis may not be  
reproduced in whole or in part, by photocopy  
or other means, without permission of the author.

Permission has been granted to the National Library of Canada to microfilm this thesis and to lend or sell copies of the film.

The author (copyright owner) has reserved other publication rights, and neither the thesis nor extensive extracts from it may be printed or otherwise reproduced without his/her written permission.

L'autorisation a été accordée à la Bibliothèque nationale du Canada de microfilmer cette thèse et de prêter ou de vendre des exemplaires du film.

L'auteur (titulaire du droit d'auteur) se réserve les autres droits de publication; ni la thèse ni de longs extraits de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation écrite.

ISBN 0-315-30800-1

APPROVAL

Name: Askar Choudhury  
Degree: Master of Science  
Title of Thesis: Regression Analysis Procedures with Higher Order  
Moving Average Errors.

Examining Committee:

Chairman: A.R. Freedman

---

D.M. Eaves  
Senior Supervisor

---

R.A. Lockhart

---

C. Villegas

---

K.L. Weldon  
External Examiner

Date approved: July 12, 1985

PARTIAL COPYRIGHT LICENSE

I hereby grant to Simon Fraser University the right to lend my thesis, project or extended essay (the title of which is shown below) to users of the Simon Fraser University Library, and to make partial or single copies only for such users or in response to a request from the library of any other university, or other educational institution, on its own behalf or for one of its users. I further agree that permission for multiple copying of this work for scholarly purposes may be granted by me or the Dean of Graduate Studies. It is understood that copying or publication of this work for financial gain shall not be allowed without my written permission.

Title of Thesis/Project/Extended Essay

REGRESSION ANALYSIS PROCEDURES WITH HIGHER ORDER

MOVING AVERAGE ERRORS.

Author: \_\_\_\_\_

(signature)

Askar Hassan Choudhury

(name)

October 28, 1985.

(date)

## ABSTRACT

### REGRESSION ANALYSIS PROCEDURES WITH HIGHER ORDER MOVING AVERAGE ERRORS

An easily implemented exact transformation is presented to transform the generalized regression problem with moving average errors; the transformed variables are used in Generalized Least Squares and Maximum Likelihood estimation. The MacDonald and MacKinnon Procedure is extended for higher order moving average process from the first order. A simulation experiment is conducted to observe the performance of three different procedures: a) General Procedure b) MacDonald & MacKinnon Procedure c) Phillips Procedure. In small samples, it is suggested that the General Procedure performs better. An efficient way to obtain the exact covariance determinant occurring in the likelihood function is presented.

An extension to higher order of Park and Heikes's Modified Approximate (MAPX) transformation for first order moving average process is derived. The relative efficiency of the regression coefficient estimate using this transformation to that using the exact (GLS) transformation and also to Ordinary Least Squares (OLS) is obtained numerically. The results suggested that MAPX performs as well as the exact transformation for a certain range of moving average parameters, depending on the sample size.

**Dedication**

To  
my dear  
"Abba and Amma."

## Acknowledgements

I would like to thank Dr. David Eaves, my senior supervisor, for his scholarly guidance and patient encouragement during the preparation of this thesis. Thanks are also due to Professor C. E. Villegas and Dr. R. Lockhart for their advice. Professors P. Kennedy, D. Maki, and R. A. Holmes of Economics, Dr. S. Kloster of Computing Centre, and Betty Dwyer of Mathematics & Statistics, of Simon Fraser University, for their constant cooperation throughout my research. Kathy Hammes and Sylvia Holmes were a great help to me.

The Statistics Division of the Government of Bangladesh and its scholarly Secretary Dr. A. K. M. Ghulam Rabbani provided me necessary leave and administrative support during my study at Simon Fraser University. Professors K. S. Ahmed, M. Kabir, K. C. Bhuyan, A. Sufian, A. Sobhan and M. M. Ahamed of Jahangirnagar University were always supportive to me. I am indebted to all of them.



## TABLE OF CONTENTS

Approval	(ii)
Abstract	(iii)
Dedication	(iv)
Acknowledgements	(v)
Table of Contents	(vi)
Chapter 1: Introduction	1
Chapter 2: Exact Estimator for the Higher Order Moving Average Process in Regression Errors	4
2.1 General Problem	5
2.2 Box-Jenkins Approach	6
2.2.1 Stationarity	6
2.2.2 Test for Stationarity	8
2.2.3 Stationarity Condition	8
2.2.4 Invertibility Condition	9
2.3 Model Building Strategy	10
2.3.1 Identification Problem	10
2.3.2 Estimation Problem	15
2.3.3 Diagnostic Checking	18
2.4 Regression Equation with Moving Average Errors	20
2.4.1 Regression Model	21

2.5	Different Procedures for Transformation	24
2.5.1	General Procedure	25
2.5.1.1	Transformation Matrix	26
2.5.1.2	Transformation in Recursive Form	27
2.5.2	MacDonald & MacKinnon Procedure	30
2.5.3	Phillips Procedure	33
2.6	Method of Estimation	36
2.6.1	General Procedure	36
2.6.1.1	Estimated Generalized Least Square	36
2.6.1.2	Maximum Likelihood	37
2.6.2	MacDonald & MacKinnon Procedure	38
2.6.2.1	Estimated Generalized Least Square	38
2.6.2.2	Maximum Likelihood	38
2.6.3	Phillips Procedure	39
2.6.3.1	Estimated Generalized Least Square	39
2.6.3.2	Maximum Likelihood	39

2.7 Simulation Experiment	40
2.8 Empirical Results	41
2.9 Specification Error	50
Chapter 3: Exact Determinant of Covariance Matrix	53
3.1 Determinant of Covariance Matrix $\Omega$	54
3.2 Determinant from the Transformation of General Procedure	54
3.3 Exact Determinant using $\Theta$ Matrix	55
Chapter 4: Approximate Estimator for the Higher Order Moving Average Process in Regression Errors	62
4.1 Approximate Transformations for MA process	63
4.2 Modified Approximate Transformation (MAPX) for Higher Order MA process	69
4.3 Empirical Results	73
Chapter 5: Conclusion	77
Appendix A.	81
Appendix B.	89
Appendix C.	95
Appendix E.	97
Appendix F.	103
Appendix G.	109
Appendix H.	115
Bibliography	117

CHAPTER 1

INTRODUCTION

This thesis reports on the generalized regression problem with moving average errors and the estimation of regression coefficients in the presence of higher order moving average errors.

In Chapter 2 an easily implemented exact transformation procedure is suggested, which can be used for both GLS and Maximum Likelihood. A simulation experiment which demonstrates the effectiveness of the suggested transformation procedure called General Procedure (GNL) is presented. Another procedure suggested by MacDonald and MacKinnon (MM) (1985) is extended for higher order moving average process. The simulation experiment conducted includes three different transformation procedures, a) General procedure (GNL), b) MacDonald and MacKinnon Procedure (MM) and c) Phillips Procedure (PHL). We also discuss the Univariate Time Series analysis (Box-Jenkins approach), which we have used in our regression error process.

For Maximum Likelihood the determinant can be obtained as a by-product of GNL procedure, while this is not possible for the other two procedures (discussed in Chapter 2). In Chapter 3 we also have suggested an efficient and simple way to obtain the exact determinant for the covariance matrix, which can be used in the Maximum Likelihood estimation for all three procedures, specifically for MM and PHL.

While exact transformation needs much more effort than the approximate transformation, for large samples approximate transformations can be used instead of exact transformations.

In Chapter 4 we have discussed various kinds of transformation for 1st order moving average process. We have extended the Modified Approximate transformation (MAPX) procedure, to apply to higher order moving average processes, suggested by Park and Heikes, (1983). We have computed numerically the relative efficiencies of MAPX to GLS and OLS to see its effectiveness for a certain range of moving average parameters.

C H A P T E R 2

EXACT ESTIMATOR FOR THE HIGHER ORDER MOVING AVERAGE PROCESS IN  
REGRESSION ERRORS

## 2.1 General Problem

It has become an increasing practice to use 'Univariate Time Series' analysis combined with 'Econometric' problems. In our conventional regression model, sometimes called Ordinary Least Squares (OLS), one of the assumptions about regression error is that, the variance-covariance matrix is  $\sigma^2 I_n$ , which implies that the error variance is homoscedastic and there are no correlations between the errors themselves. But problems arise when this assumption breaks down, e.g., if the homoscedastic property remain and the other one breaks down, which usually happens in Time Series problems. We then have a covariance matrix of regression error of the form  $\sigma^2 \Omega$ , where  $\Omega$  is assumed to be known or at least could be estimated. If  $\Omega$  is known then we can perform the Generalized Least Squares (GLS) which has the same properties as OLS.

In practice  $\Omega$  is unknown and some kind of restrictive assumptions are made about its structure and since we have to estimate  $\Omega$  ( where estimated  $\Omega$  will be denoted as  $V$  ), we would no longer have the properties of GLS. This would lead us to the Estimated Generalized Least Square (EGLS) method. Like GLS, EGLS is also performed in two stages, using OLS in both the stages, only with the exception that estimated values of the parameters are used instead of true value of the parameters. Since it is not, in general, an easy task to find the finite sample properties of EGLS estimators, we can use the Monte Carlo



experiment and since it is, by its nature, model specific and also depends on data sets, the findings can not be generalized, but we will have some insights about the efficiencies of the EGLS estimators. When a specific assumption about the distribution of the regression error is made, the Maximum Likelihood (ML) method of estimation becomes an alternative to EGLS and the parameters are estimated by maximizing the likelihood function.

## 2.2 Box-Jenkins Approach

If the above mentioned problem arises, i.e., if  $\Omega$  is not a diagonal matrix, we have to classify the error generating process and it can be done by using Box-Jenkins approach. We will restrict our discussion to the stationary process, since, given stationarity, any series can be well-approximated by either a moving average, autoregressive or mixture of both (see Granger, 1980, p.60).

### 2.2.1 Stationarity

One of our assumptions is that the underlying stochastic process, in our case the regression error vector, is stationary. If the covariance characteristics of the stochastic process change over time, the process is nonstationary; on the other hand, if it is fixed it is stationary. A stationary stochastic process can be classified into,

a) Strictly Stationary

b) Weakly Stationary

a) Strictly Stationary Process:

A process is said to be strictly stationary if its distributional properties are unaffected by change of time origin, i.e., if the joint probability distribution of the observations, say  $u_{t_1} u_{t_2} \dots u_{t_n}$  at a set of different time points  $t_1, t_2, \dots, t_n$  is invariant with respect to displacement of time.

b) Weakly Stationary Process:

A stochastic process is said to be weakly stationary if the moments up to some order  $k$  depends only on time differences but not on time origin. Therefore, if the series (process) is stationary with respect to mean and covariance, we call it weakly stationary of second order. This kind of weak stationary along with normality assumption makes it strictly stationary (see Box and Jenkins, 1970, p.30).

### 2.2.2 Test for Stationarity

Therefore, before proceeding to data analysis we need to check whether the data to be analysed are stationary or not. Box-Jenkins and others suggested plotting the observations against time and looking whether there is any evident trend in mean or trend in variance, and also plotting the autocorrelation function (ACF). If the ACF for different lags does not die out quickly, i.e., if they are almost .1, then it is an indication that the process is nonstationary; otherwise, it is stationary. Recently, an approach to stationarity testing has been given by Ali and Thalheimer (1983, pp.249-257). Basically, they developed four test-statistics for four different distributions: normal, logistic, double exponential, and Cauchy. They showed if the test-statistics are insignificant then the series may be considered stationary. This method is not widely practiced, which may be due to unavailability of the program.

### 2.2.3 Stationarity Condition

If an AR(p) process,

$$u_t = \phi_1 u_{t-1} + \phi_2 u_{t-2} + \dots + \phi_p u_{t-p} + \epsilon_t$$

$$\text{or, } (1 - \phi_1 \beta - \phi_2 \beta^2 - \dots - \phi_p \beta^p) u_t = \phi(\beta) u_t = \epsilon_t$$

is to be stationary, it must satisfy the condition that roots of the characteristic equation (Box and Jenkins, 1976, pp.53-54)

$$\phi(\beta) = 0 \quad (2.1)$$

should lie outside the unit circle,

i.e., the solutions  $\beta_1, \beta_2, \dots, \beta_p$  to eqn. (2.1) must all be greater than 1 in magnitude. Specifically, for AR(1) process, the equation (2.1) becomes,

$$1 - \phi_1 \beta = 0.$$

So that the solution should satisfy,

$$|\beta| = \frac{1}{|\phi_1|} > 1$$

which implies that,  $|\phi_1| < 1$ .

#### 2.2.4 Invertibility Condition .

Analogous to the stationarity condition for autoregressive process, if an MA(q) process,

$$u_t = \epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2} - \dots - \theta_q \epsilon_{t-q}$$

is to be invertible, it must satisfy the condition, that the roots of the characteristic equation,

$$\theta(\beta) = 1 - \theta_1 \beta - \theta_2 \beta^2 - \dots - \theta_q \beta^q = 0 \quad (2.2)$$

must lie outside the unit circle,

i.e., the solutions  $\beta_1, \beta_2, \dots, \beta_q$  to eqn. (2.2) must all be

greater than one (see Box-Jenkins, 1976, pp. 50-51).

For MA(1) process,

$$1 - \theta_1 \beta = 0$$

Therefore,

$$|\beta| = \frac{1}{|\theta_1|} > 1$$

which implies that  $|\theta_1| < 1$ .

### 2.3 Model Building Strategy

In model building strategy of univariate time series, Box and Jenkins (1976, pp.17-18) suggested selecting a model which has smaller number of parameters, which they called a parsimonious model. They actually proposed an iterative procedure in model selection which includes the following stages.

1. Identification of the model
2. Estimation of the parameters
3. Diagnostic checking

#### 2.3.1 Identification Problem

Box & Jenkins suggested identification of a preliminary model in this stage using ACF and PACF (partial autocorrelation function) pattern. The reason for using ACF and PACF are discussed below.

The MA(q) process can be defined as one of the form

$$u_t = \epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2} - \dots - \theta_q \epsilon_{t-q}$$

where,  $\epsilon_t$  are white noise (WN) with mean zero and constant variance  $\sigma^2$  and also zero covariances.

$$\text{Therefore, } E(u_t) = 0$$

$$\text{Var}(u_t) = E[u_t^2] = \sigma^2(1 + \theta_1^2 + \theta_2^2 + \dots + \theta_q^2)$$

$$\text{Cov}(u_t u_{t-1}) = E[u_t u_{t-1}]$$

$$= E[(\epsilon_t - \theta_1 \epsilon_{t-1} - \dots - \theta_q \epsilon_{t-q})(\epsilon_{t-1} - \theta_1 \epsilon_{t-2} - \dots - \theta_q \epsilon_{t-q-1})]$$

$$= \sigma^2[-\theta_1 + \theta_1 \theta_2 + \dots + \theta_{q-1} \theta_q]$$

$$\text{Cov}(u_t u_{t-2}) = E[u_t u_{t-2}]$$

$$= E[(\epsilon_t - \theta_1 \epsilon_{t-1} - \dots - \theta_q \epsilon_{t-q})(\epsilon_{t-2} - \theta_1 \epsilon_{t-3} - \dots - \theta_q \epsilon_{t-q-2})]$$

$$= \sigma^2[-\theta_2 + \theta_1 \theta_3 + \dots + \theta_{q-2} \theta_q]$$

and so on.

Therefore, the autocorrelation function can be written as,

$$\rho_k = \frac{-\theta_k + \theta_1 \theta_{k+1} + \dots + \theta_{q-k} \theta_q}{1 + \theta_1^2 + \theta_2^2 + \dots + \theta_q^2}, \quad k=1, 2, \dots, q$$

$$= 0, \quad k > q$$

which implies that if the process is a moving average then we will have cuts-off in correlogram (plots of ACF) for the

autocorrelation with lag greater than q, the order of MA process.

Similarly, the AR process can be defined as one satisfying

$$u_t = \phi_1 u_{t-1} + \phi_2 u_{t-2} + \dots + \phi_p u_{t-p} + \epsilon_t$$

so that  $\gamma_0 = V(u_t) = E[u_t^2]$

$$= E[u_t(\phi_1 u_{t-1} + \phi_2 u_{t-2} + \dots + \phi_p u_{t-p} + \epsilon_t)]$$

$$= \phi_1 \gamma_1 + \phi_2 \gamma_2 + \dots + \phi_p \gamma_p + \sigma^2$$

where,  $\gamma_1, \gamma_2, \dots, \gamma_p$  are the autocovariances with lag(1), lag(2), ..., lag(p) and  $\phi_1, \phi_2, \dots, \phi_p$  are the autoregressive parameters.

$$\gamma_1 = \text{Cov}(u_{t-1}, u_t)$$

$$= E[(u_{t-1}(\phi_1 u_{t-1} + \phi_2 u_{t-2} + \dots + \phi_p u_{t-p} + \epsilon_t))]$$

$$= \phi_1 \gamma_0 + \phi_2 \gamma_1 + \dots + \phi_p \gamma_{p-1}$$

$$\gamma_2 = \text{Cov}(u_{t-2}, u_t)$$

$$= \phi_1 \gamma_1 + \phi_2 \gamma_0 + \dots + \phi_p \gamma_{p-2}$$

⋮

$$\gamma_p = \phi_1 \gamma_{p-1} + \phi_2 \gamma_{p-2} + \dots + \phi_p \gamma_0$$

and for  $k > p$

$$\gamma_k = \phi_1 \gamma_{k-1} + \phi_2 \gamma_{k-2} + \dots + \phi_p \gamma_{k-p}$$

Therefore, the autocorrelation functions for AR(p) process becomes (known as Yule-Walker equations),

$$\rho_1 = \phi_1 + \phi_2 \rho_1 + \dots + \phi_p \rho_{p-1}$$

$$\rho_2 = \phi_1 \rho_1 + \phi_2 + \dots + \phi_p \rho_{p-2}$$

⋮  
⋮  
⋮

$$\rho_p = \phi_1 \rho_{p-1} + \phi_2 \rho_{p-2} + \dots + \phi_p$$

and for  $k > p$

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \dots + \phi_p \rho_{k-p},$$

which implies that unlike MA process AR process does not have cuts-off in ACF for  $k > p$ . Therefore, we can conclude that if the ACF cuts off after a certain point the process can be thought of as an MA process. But, if it does not cut-off, rather dieing out slowly, it can be thought of as an AR process. More confirmation can be drawn from PACF.

Box and Jenkins suggested that the partial autocorrelation can be approximated by Yule-Walker estimates of the successive autoregressive process. This is discussed in detail by Pindyck and Rubinfeld (1981, pp.524-526). For the pth order autoregressive process PACF has a cut-off after lag p. Box and Jenkins (1976, p.70) also show that for MA process PACF does not cut-off after lag q, rather dieing out slowly, as opposite to AR process.



Since we do not know the autocorrelation and partial autocorrelation in practice, we have to estimate them from the observation. But the estimated autocorrelation and partial autocorrelation will not necessarily be exactly zero. Rather, it will be approximately zero for  $k > q$  or  $p$  for MA process or AR process respectively.

Therefore we need to find the standard error of autocorrelation and partial autocorrelation estimates. Using Bartlett's approximation (see Box and Jenkins, 1976, pp.34-35) the variance of  $r_k$  (which is the estimated autocorrelation for  $\rho_k$ ) is,

$$\text{Var}(r_k) \approx [1 + 2 \sum_{i=1}^q \rho_i^2] / n, \quad k > q.$$

For the autoregressive process we can use the result from Quenouille (see Box and Jenkins, 1976, p.65), that the variance of partial autocorrelation coefficient for lag period greater than the order  $p$  of the process can be approximated as,

$$\text{Var}[\hat{\phi}_{kk}] \approx 1/n, \quad k > p$$

where  $n$  is the number of observations.

Recently, some other methods were proposed for identification of the model. One of them is Corner method proposed by Beguin, Gourierouse and Monfort (1980, pp.423-436) and another one is proposed by Pukkila (1982, pp.81-103). He

suggested that, since, specifically for mixed model it is difficult to get an idea about the order of the process using ACF and PACF, one should estimate the parsimonious models ARMA(p,q) beginning with one parameter model (either p=1, or q=1) and test whether the parameters are significantly different from zero and also test whether the estimated residuals behaves like white noise and ~~if one~~ of the above tests fails for all possible low order models, then proceed for higher order model. Thus, the models that could be checked are ARMA(0,1), ARMA(1,0), ARMA(1,1), ARMA(0,2) and so on. This seems to us to be reasonable, because, as Box-Jenkins and other authors agree, the models we have in practice have a small number of parameters. Since we have computer packages available to estimate the parameters quickly, we can use this technique too.

After tentatively identifying a model we can proceed to the next stage.

### 2.3.2 Estimation problem

Among the estimation methods frequently used for univariate time series analysis are the method of moments, the method of back forecasting, the method of least squares and the method of maximum likelihood(ML). The method of moments estimate for the first order moving average process is,

$$\theta = \frac{-1 \pm \sqrt{1-4r_1^2}}{2r_1}$$

where,  $r_1$ , is the estimated autocorrelation coefficient of lag 1. From the above we see that there are two possible solutions, but only one of them will satisfy the invertibility condition. Box and Jenkins (1976,p.188) show that only one of the multiple moment solutions for any MA order will satisfy the invertibility condition. According to Judge, et. al. (1980,p.198) these estimates are inefficient relative to the Nonlinear Least Square (NLS) estimator. Most authors suggest using NLS or ML and to do this we need an initial estimate of the parameters.

Box-Jenkins suggest using the estimate obtained by the method of moments as an initial value of the parameters for NLS or ML.

We can obtain the preliminary estimates for AR process by solving the Yule-Walker equations discussed previously. Specifically, for AR(2) process these estimates are,

$$\hat{\phi}_1 = \frac{r_1(1-r_2)}{1-r_1^2}$$

$$\hat{\phi}_2 = \frac{r_2-r_1^2}{1-r_1^2}$$

In this paper, we will focus our attention upon the MA process.

A great deal of discussion has been given on nonlinear estimation technique in Box and Jenkins (1976,pp.231-242). Most of the computer packages use nonlinear least squares and not the maximum likelihood, because of the complexity in estimating the determinant of the covariance matrix for MA process.

Box and Jenkins (1976,p.213) suggested an approximation to the ML method by disregarding the determinant, because the determinant  $|\Omega|$  is dominated by the exponent of the likelihood function. Mcleod(1977, pp.531-534) proposed a method by substituting an approximation of the determinant term in the likelihood function, claiming a closer approximation to the exact ML estimator. Osborn (1976,pp.75-87) uses the technique to calculate the exact covariance determinant mentioned by Box and Jenkins(1970,pp.270-272):  $|\Omega|=|R'R|$ , where,  $R$  is a matrix of order  $(n+q) \times q$ , so that  $R'R$  is of order  $q \times q$  and  $R$  can be calculated recursively. If  $n$  is big enough this technique might make problems in computer space and time in forming the  $R$  matrix recursively and also in multiplication of them. As a modification of this we will introduce a more efficient method to obtain the exact covariance determinant.

Another approach was given by Phadke and Kedem (1978,pp.511-519) for the moving average process. They get the transformation matrix by decomposing the covariance matrix  $\Omega$  using the Cholesky decomposition and they also use the decomposed matrix to get the determinant. Later it was extended for ARMA model by Craig(1979,pp.59-65). Both of them use a library algorithm to decompose  $\Omega$ .

In this paper later, we will show how the determinant can be easily obtained from the transformation procedure (used in Chapter 2 for GNL) without any further effort.

### 2.3.3 Diagnostic Checking

In this stage we test the model which we chose tentatively, as to whether it appears to agree with the data.

The best way to investigate the adequacy of model fitting, is to observe its performance outside the sample period. That is, the whole sample is divided into two sets, one set is used for estimating the model and the other set is to check how well the model fits. But, most of the time insufficient amount of data prevents us from doing this. So, we use the same set of data for both the purposes.

Among the tests, we check whether the estimated parameters are significantly different from zero or not. Then we go for residuals checking, which should be WN, that is, to see that the residuals are as a whole uncorrelated among themselves. One of the test statistics suggested by Box and Pierce (1970) is,

$$Q_1(r) = n \sum_{k=1}^m r_k^2$$

where,

$$r_k = \frac{\sum_{t=1}^{n-k} \hat{\epsilon}_t \hat{\epsilon}_{t-k}}{\sum_{t=1}^n \hat{\epsilon}_t^2}$$

and they showed that this statistic is asymptotically distributed as  $\chi^2$  with  $(m - p - q)$  d.f., where,  $p$  and  $q$  are the order of the AR and MA process respectively and  $m$  is the highest lag period for autocorrelation (i.e., the time displacement are  $1, 2, \dots, m$ ) considered.

Later, Ljung and Box (1978, pp.297-303) conclude that the test statistic,

$$Q_2(r) = n(n+2) \sum_{k=1}^m \frac{r_k^2}{n-k}$$

has better statistical properties than the above.

Pukkila (1982, pp.81-103) said that the above two statistics are not sensitive to slight departures from WN for reasonable sample size, thus, he proposed another test statistic under null hypothesis of WN,

$$Q_3(r) = \sum_{k=2}^m \frac{\{r_k - \hat{\phi}_{kk} + E(\hat{\phi}_{kk})\}^2}{\text{Var}(r_k - \hat{\phi}_{kk})}$$

where,  $\hat{\phi}_{kk}$  is the estimated partial autocorrelation at lag k and  $m=2\sqrt{n}$ , which he says reasonable for  $50 \leq n \leq 100$ . But, most computer packages use the Box-Pierce test for its simplicity.

Box and Jenkins also suggested overfitting the model, i.e., after identifying a model one has to select some other model around the identified one. If two models are identified to be selected then choose the one which has the smaller number of parameters.

#### 2.4 Regression Equation with Moving Average Errors:

It can be observed that, recently, considerable attention has been given to the regression model with ARMA errors. Though the error process can be any of the three processes mentioned before, most researchers in practice assume the process to be autoregressive and most of the time a conclusion is drawn using the Durbin-Watson(DW) statistic proposed by Durbin and Watson(1950,1951). It is to be mentioned that the DW-statistic is not valid for the error process other than AR(1); see Koutsoyiannis(1977,pp.212,216). Harvey (1981,p.209-210) also expressed the same view. Therefore the DW-statistic is sometimes misleading. In the case where the lag dependent variable appears as an independent variable, Durbin(1970) suggested another test-statistic. Wallis (1972) developed a test-statistic for the seasonal fourth order autocorrelation in the error term of a regression equation estimated from quarterly data, generalizing the DW-statistic. Since most computer packages provide a test and estimation technique for the AR(1) process, most researchers assume that the underlying process of regression error is AR(1), as mentioned by Harvey(1981,p.189), though there is no reason why the other processes should not be entertained equally. Under these circumstances, researchers became interested in exploring the other areas and some related works are Phillips(1966), Aigner(1971), Pagan(1973), Pierce(1971), Pagan(1974), Nicholls,Pagan and Terrel(1975) and Pagan & Nicholls(1976).

The model we considered for this paper is the regression model with MA errors. Until recently, little work have been done on higher order MA process. Almost all the Monte Carlo or similar kinds of numerical comparisons were done on the MA(1) process. We will attempt a Monte Carlo comparison for three different procedures discussed below for the MA(2) process.

Most of the time transformations developed for MA(1) are difficult, sometimes impossible to generalize for higher order; e.g., Balestra(1980), Pesaran(1973), which is also mentioned by Judge(1980,p.196). But the transformation we are proposing in this paper does not need the inversion of a matrix nor even the transformation matrix and can easily be implemented for higher order MA process.

#### 2.4.1 Regression Model

Let us consider a regression model,

$$Y = X\beta + U$$

(2.3)

Where,  $Y$  is a response vector of dimension  $n \times 1$

$X$  is a nonstochastic design matrix of dimension  $n \times k$   
with rank  $k$ ,  $k < n$

$U$  is a random vector of dimension  $n \times 1$

$\beta$  is a parameter vector of dimension  $k \times 1$



i.e.,

$$X = \begin{bmatrix} X_{11} & X_{21} & X_{31} & \dots & X_{k1} \\ X_{12} & X_{22} & X_{32} & \dots & X_{k2} \\ X_{13} & X_{23} & X_{33} & \dots & X_{k3} \\ X_{14} & X_{24} & X_{34} & \dots & X_{k4} \\ X_{15} & X_{25} & X_{35} & \dots & X_{k5} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ \vdots & \vdots & \vdots & \dots & \vdots \\ \vdots & \vdots & \vdots & \dots & \vdots \\ X_{1n} & X_{2n} & X_{3n} & \dots & X_{kn} \end{bmatrix}$$

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \\ \vdots \\ Y_n \end{bmatrix} \quad u = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \\ u_5 \\ \vdots \\ u_n \end{bmatrix} \quad \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \vdots \\ \beta_k \end{bmatrix}$$

and assume the disturbance term  $U$  follows a second order moving average process,

$$u_t = \epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2} \quad (2.4)$$

The random variable  $\epsilon_t$  is assumed to be independent with zero mean and constant variance, i.e.,  $E(\epsilon_t) = 0$ ,  $\text{Var}(\epsilon_t) = \sigma^2$ .

Therefore, the random vector  $u_t$  is thus characterized by,

$$E(u_t) = 0 \text{ and } E(u_t u_t') = \sigma^2 \Omega.$$

$\Omega$  is a five-diagonal matrix, with  $1 + \theta_1^2 + \theta_2^2$  in the main diagonal,  $-\theta_1(1 - \theta_2)$  in the second diagonal above and below the main diagonal and  $-\theta_2$  in the third diagonal above and below the second diagonal. This follows since,

$$E(u_t u_t) = E(u_t^2) = (1 + \theta_1^2 + \theta_2^2) \sigma^2$$

$$E(u_t u_{t-1}) = E[(\epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2})(\epsilon_{t-1} - \theta_1 \epsilon_{t-2} - \theta_2 \epsilon_{t-3})]$$

$$= -\theta_1 E(\epsilon_{t-1})^2 + \theta_1 \theta_2 E(\epsilon_{t-2})^2 + 0 + 0 + \dots$$

$$= -\theta_1 \sigma^2 + \theta_1 \theta_2 \sigma^2, \text{ (since } \epsilon_t \text{ is homoscedastic),}$$

$$= -\theta_1(1 - \theta_2) \sigma^2$$

$$E(u_t u_{t-2}) = -\theta_2 \sigma^2$$

$$E(u_t u_{t-3}) = 0$$

$$E(u_t u_{t-n}) = 0$$

## 2.5 Different Procedures<sup>1</sup> for Transformation

It is well known that an analytical expression for the transformation to transform the generalized regression problem into a simple (OLS) regression problem is generally available for the regression model with autoregressive disturbances. See J. Wise (1955)<sup>2</sup> for  $\Omega^{-1}$  and Fuller (1976, p.423) for this transformation. However, an analytical expression for such a transformation in the case of a regression model with moving average disturbances is available only for the first order, i.e., for MA(1) disturbances Balestra (1980). Pesaran (1973) also found the transformation matrix for the first order moving average process, and that involves more complexity than Balestra's method. Both the above procedures have a limitation in the sense that they are not readily extended for higher order MA process.

-----  
<sup>1</sup>By different procedures we mean the different ways of transformation of a generalized regression problem into a simple regression problem; in other words, the procedure is a way of writing down the model equations (2.3) and (2.4) so that the resulting model becomes an OLS problem.

<sup>2</sup>This inverse matrix is also given in Kendall, Stuart and Ord (1983, p.543)

The estimation of regression coefficients when the regression model has moving average process disturbances can be handled in several ways. We will discuss the following three procedures:

- a) General Procedure
- b) MacDonal and MacKinnon Procedure
- c) Phillips Procedure

### 2.5.1 General Procedure

To retain the BLUE (Best Linear Unbiased Estimate) properties of regression coefficient with white noise error, we need to transform the original observations so that the regression residuals after transformation becomes white noise (W.N.).

Let  $T$  be a non-singular matrix such that,

$$T\Omega T' = I_n$$

$$\text{or, } T^{-1}T\Omega T' = T^{-1}$$

$$\text{or, } I\Omega T'(T')^{-1} = T^{-1}(T')^{-1}$$

$$\text{or, } I\Omega I = T^{-1}(T')^{-1}$$

$$\text{or, } \Omega = (T'T)^{-1}$$

$$\text{or, } \Omega^{-1} = T'T$$

The transformed model is then,

$$TY = TX\beta + TU$$

and the covariance matrix of transformed residual is,

$$\begin{aligned} E[TU(TU)'] &= E[TUU'T'] \\ &= T E(UU') T' \\ &= T \sigma^2 \Omega T' \\ &= \sigma^2 T\Omega T' \\ &= \sigma^2 I_n \end{aligned}$$

The Ordinary Least Square (OLS) estimate of regression co-efficient on transformed observation is,

$$\begin{aligned} \beta &= [(TX)'(TX)]^{-1} (TX)'TY \\ &= [X'T'TX]^{-1} X'\Omega^{-1}Y \\ &= [X'\Omega^{-1}X]^{-1} X'\Omega^{-1}Y \end{aligned}$$

called OLS estimate (see Goldberger, 1963, p.232-234).

### 2.5.1.1 Transformation Matrix

To get the transformation matrix we can proceed as follows:  
There exists a uniquely defined positive upper triangular matrix S (i.e., upper triangular with positive diagonal elements)

such that,  $S'S = \Omega$

$$\text{or, } S = (S')^{-1} \Omega$$

Since,  $\Omega$  is a  $(2q+1)$ -diagonal symmetric band matrix,  $S$  is an upper triangular matrix with non-zero elements in the main diagonal and also the  $q$  diagonals immediately above the main diagonal are non-zero and all other elements are zero;  $S'$  is a similar kind of lower triangular matrix and therefore,  $(S')^{-1}$  is also a lower triangular matrix.

From the relationship,

$$S = (S')^{-1} \Omega$$

assuming we have  $S$  and  $\Omega$ , we can solve recursively for the elements of the matrix  $(S')^{-1}$ , which is our required transformation matrix.

Again we know that,  $S S^{-1} = I$ , therefore, solving for the elements of  $S^{-1}$  from the equation system we could get the transformation matrix, which would be more efficient than the above computation. However, we did not use either of these two transformation matrices in the simulation experiment, because the recursive transformation (described below) is simpler and easier to implement.

#### 2.5.1.2 Transformation in Recursive Form

The transformation in a recursive way is so simple that it avoids the inversion of  $\Omega$  and even the transformation matrix. Then all we need to obtain  $\sqrt{\Omega}$  which we denote as  $S$ , are the

values  $S_{ij}$ . For practical purposes the transformation can be shown in simple recursive way as below;

$$\bar{Y}_1 = Y_1 / s_{11}$$

$$\bar{Y}_2 = (Y_2 - s_{12} \bar{Y}_1) / s_{22}$$

$$\bar{Y}_3 = (Y_3 - s_{23} \bar{Y}_2 - s_{13} \bar{Y}_1) / s_{33}$$

$$\bar{Y}_4 = (Y_4 - s_{34} \bar{Y}_3 - s_{24} \bar{Y}_2 - s_{14} \bar{Y}_1) / s_{44}$$

$$\bar{Y}_5 = (Y_5 - s_{45} \bar{Y}_4 - s_{35} \bar{Y}_3 - s_{25} \bar{Y}_2 - s_{15} \bar{Y}_1) / s_{55}$$

.

.

$$\bar{Y}_{q+1} = (Y_{q+1} - s_{q,q+1} \bar{Y}_q - s_{q-1,q+1} \bar{Y}_{q-1} - \dots - s_{1,q+1} \bar{Y}_1) / s_{q+1,q+1}$$

.

.

$$\bar{Y}_n = (Y_n - s_{n-1,n} \bar{Y}_{n-1} - s_{n-2,n} \bar{Y}_{n-2} - \dots - s_{n-q,n} \bar{Y}_{n-q}) / s_{n,n}$$

where,  $\bar{Y}$  and  $Y$  are the transformed and original observations respectively;  $q$  is the order of MA process and  $n$  is the number of observations. Similarly, we can get the transformed variable for each column of the X-matrix.

The square root matrix  $S$  of  $\Omega$  could easily be obtained from the relationship,

$$\Omega = S'S$$

$$\text{Since, } \Omega_{ii} = S_{1i}^2 + S_{2i}^2 + S_{3i}^2 + \dots + S_{ii}^2, \quad i=j$$

$$\text{and } \Omega_{ij} = S_{1i} S_{1j} + S_{2i} S_{2j} + S_{3i} S_{3j} + \dots + S_{ii} S_{ij}, \quad i < j$$

we have,

$$S_{11} = \sqrt{\Omega_{11}} \quad \text{and} \quad S_{1j} = \Omega_{1j} / S_{11} \quad , i=1$$

$$S_{ii} = \sqrt{(\Omega_{ii} - \sum_{k=1}^{i-1} S_{ki}^2)} \quad , i > 1$$

$$S_{ij} = (\Omega_{ij} - \sum_{k=1}^{i-1} S_{ki} S_{kj}) / S_{ii} \quad , i > 1 \quad \text{and} \quad j > i$$

$S_{ij} = 0$ ,  $i > j$  and  $j > i+q$ , where  $q$  is the order of MA process.  
Specifically, for MA(2) the  $S_{ij}$ 's are,

$$S_{11} = \sqrt{\Omega_{11}} \quad , i=1$$

$$S_{22} = \sqrt{(\Omega_{11} - S_{12}^2)} \quad , i=2$$

$$S_{i,i} = \sqrt{(\Omega_{11} - S_{i-2,i}^2 - S_{i-1,i}^2)} \quad , i = 3, 4, 5, \dots, n$$

$$S_{12} = \Omega_{12} / S_{11} \quad , i=2$$

$$S_{i-1,i} = (\Omega_{12} - S_{i-2,i-1} S_{i-2,i}) / S_{i-1,i-1} \quad , i = 3, 4, 5, \dots, n$$

$$S_{i-2,i} = \Omega_{13} / S_{i-2,i-2} \quad , i = 3, 4, 5, \dots, n$$



In particular, for second order moving average process the recursive transformation can be shown as follows:

$$Y_1 = Y_1 / S_{11}$$

$$Y_2 = (Y_2 - S_{12} Y_1) / S_{22}$$

$$Y_i = (Y_i - S_{i-1,i} Y_{i-1} - S_{i-2,i} Y_{i-2}) / S_{ii} \quad , i = 3, 4, 5, \dots, n.$$

One can use a library subroutine to calculate the elements of  $S_{ij}$  (square root matrix of  $\Omega$ ) to get the transformation. But as we observe from the above algebra, a self contained program can be written to get the transformation without using library subroutine and the computer program is given in Appendix H.1 for transformation in the case of MA(2) process. The above model equations may then be analysed as described in Section 2.6.1.

### 2.5.2 MacDonald & MacKinnon Procedure

MacDonald and MacKinnon (1985) propose a very simple procedure to deal with MA(1) residuals in regression equation by GLS.

Without using the matrix M for transformation in Osborn's paper (1978) for univariate time series model, which is tedious to form and also needs computer space, one can use the recursive form of transformation, as MacDonald and MacKinnon did in their paper for MA(1) and we will extend MacDonald and MacKinnon procedure for higher order MA process.

The model to be considered here is same as previous one,

$$Y_t = X_t \beta + U_t \tag{2.5}$$

$$\text{where, } U_t = \epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2} - \dots - \theta_q \epsilon_{t-q}. \tag{2.6}$$

We can combine the above two equations into  $Y = X\beta + Z'\eta + \epsilon$ ; with some algebra one can verify that this represents (2.5) and (2.6) when the transformed variables  $Y$ ,  $X$  and  $Z$  are defined and calculated recursively as follows:

$$Y_t = y_t + \theta_1 Y_{t-1} + \theta_2 Y_{t-2} + \dots + \theta_q Y_{t-q}$$

$$X_t = x_t + \theta_1 X_{t-1} + \theta_2 X_{t-2} + \dots + \theta_q X_{t-q}$$

$$Z_t = \theta Z_{t-1}, \quad \text{where,}$$

$$\Theta = \begin{bmatrix} \theta_1 & 1 & 0 & 0 & \dots & 0 \\ \theta_2 & 0 & 1 & 0 & \dots & 0 \\ \theta_3 & 0 & 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \theta_{q-1} & 0 & 0 & 0 & \dots & 1 \\ \theta_q & 0 & 0 & 0 & \dots & 0 \end{bmatrix} \quad q \times q$$

$$= \begin{bmatrix} \theta_{q-1} & I \\ \theta_q & 0 \end{bmatrix}$$

where,  $\theta_{q-1}$  is a column vector of  $\theta_1, \theta_2, \theta_3, \dots, \theta_{q-1}$ ,  
 I is a identity matrix of order  $(q-1) \times (q-1)$ ,  $\theta_q$  is a scalar and  
 '0' is a row vector of zeroes of  $(q-1)$  elements;

$$Z_0 = \begin{bmatrix} -1 \\ 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad \eta = \begin{bmatrix} \epsilon_0 \\ \epsilon_{-1} \\ \epsilon_{-2} \\ \vdots \\ \epsilon_{-q+1} \end{bmatrix}$$

$qx1$                        $(qx1)$

$Y_0, Y_{-1}, \dots, Y_{-q+1}$  and  $X_0, X_{-1}, \dots, X_{-q+1}$  being zeroes.

In particular for MA(2) process the above procedure can be described as below,

The model is,

$$Y_t = X_t \beta + U_t$$

$U_t = \epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2}$  and the combination of the above two equations is,

$$Y = X\beta + Z'\eta + \epsilon$$

where,

$$Y_t = y_t + \theta_1 Y_{t-1} + \theta_2 Y_{t-2}$$

$$X_t = x_t + \theta_1 X_{t-1} + \theta_2 X_{t-2}$$

$$Z_t = \Theta Z_{t-1}$$

$$Y_0 = Y_{-1} = 0$$

$$X_0 = X_{-1} = 0$$

$$Z_0 = \begin{bmatrix} -1 \\ 0 \end{bmatrix} \quad \eta = \begin{bmatrix} \epsilon_0 \\ \epsilon_{-1} \end{bmatrix} \quad \Theta = \begin{bmatrix} \theta_1 & 1 \\ \theta_2 & 0 \end{bmatrix}$$

These model equations may then be analysed as described in Section 2.6.2.

### 2.5.3 Phillips Procedure

This procedure was first introduced by Phillips (1966) and applied by Trivedi (1970), further studied by Pagan & Nicholls (1976).

The regression error  $u_t$ , which follows MA process,

$$u_t = \epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2} - \dots - \theta_q \epsilon_{t-q}$$

can be written as,

$$U = M\epsilon - M\tilde{\epsilon}$$

$$\text{where, } U' = [u_1 \ u_2 \ u_3 \ \dots \ u_n]$$

$$\epsilon' = [\epsilon_1 \ \epsilon_2 \ \epsilon_3 \ \dots \ \epsilon_n]$$

$$\tilde{\epsilon}' = [\epsilon_0 \ \epsilon_{-1} \ \epsilon_{-2} \ \dots \ \epsilon_{-q+1}]$$

$$\begin{matrix}
 M = \\
 (nxn)
 \end{matrix}
 \left[ \begin{array}{cccccccc}
 1 & 0 & 0 & 0 & \dots & 0 & 0 \\
 -\theta_1 & 1 & 0 & 0 & \dots & 0 & 0 \\
 -\theta_2 & -\theta_1 & 1 & 0 & \dots & 0 & 0 \\
 \dots & \dots & \dots & \dots & \dots & \dots & \dots \\
 \dots & \dots & \dots & \dots & \dots & \dots & \dots \\
 -\theta_q & & & & & & \\
 0 & & & & & & \\
 \vdots & & & & & & \\
 \vdots & & & & & & \\
 0 & \dots & \dots & -\theta_q & \dots & -\theta_1 & 1
 \end{array} \right]$$

$$\begin{matrix}
 M = \\
 nxq
 \end{matrix}
 \left[ \begin{array}{cccccc}
 \theta_1 & \theta_2 & \theta_3 & \dots & \theta_{q-1} & \theta_q \\
 \theta_2 & \theta_3 & \theta_4 & \dots & \theta_q & 0 \\
 \vdots & & & & & \vdots \\
 \vdots & & & & & \vdots \\
 \vdots & & & & & \vdots \\
 \theta_q & & & & & \\
 0 & \dots & \dots & \dots & \dots & 0 \\
 \vdots & \dots & \dots & \dots & \dots & \\
 \vdots & \dots & \dots & \dots & \dots & \\
 0 & \dots & \dots & \dots & \dots & 0
 \end{array} \right]$$

Pagan and Nicholls establish a theorem, that minimizing  $e'e + \tilde{e}'\tilde{e}$  with respect to  $\beta$ ,  $\theta$  and  $\tilde{e}$  is equivalent to minimizing  $U'\Omega^{-1}U$  with respect to  $\beta$  and  $\theta$ , therefore NLS or ML can be applied by calculating the errors in the sum of squares  $e'e + \tilde{e}'\tilde{e}$  recursively as follows:

$$\epsilon_{-q+1} = \tilde{\epsilon}_{-q+1}$$

...

...

$$\epsilon_{-1} = \tilde{\epsilon}_{-1}$$

$$\epsilon_0 = \tilde{\epsilon}_0$$

$$\epsilon_1 = y_1 - x_1' \beta + \sum_{j=1}^q \theta_j \tilde{\epsilon}_{1-j}$$

$$\epsilon_t = y_t - x_t' \beta + \sum_{j=1}^{t-1} \theta_j \epsilon_{t-j} + \sum_{j=t}^q \theta_j \tilde{\epsilon}_{t-j}, \quad t=2, 3, \dots, q$$

$$\epsilon_t = y_t - x_t' \beta + \sum_{j=1}^q \theta_j \epsilon_{t-j}, \quad t=q+1, q+2, \dots, n$$

To calculate the above residuals one needs to have starting values for the vector  $\tilde{\epsilon}$ . One possibility is to set  $\tilde{\epsilon} = 0$  and the other way is to estimate  $\tilde{\epsilon}$ . To estimate  $\tilde{\epsilon}$  we can use back forecasting method discussed by Box & Jenkins (1970, p.213-215), which will be more efficient than setting  $\tilde{\epsilon} = 0$ . Osborn's (1978) method using least squares can also be used to estimate  $\tilde{\epsilon}$ .

In particular for MA(2) process, the successive residuals are defined by,

$$\epsilon_{-1} = \tilde{\epsilon}_{-1}$$

$$\epsilon_0 = \tilde{\epsilon}_0$$

$$\epsilon_1 = y_1 - \beta_0 - x_1 \beta + \theta_1 \tilde{\epsilon}_0 + \theta_2 \tilde{\epsilon}_{-1}$$

$$\epsilon_2 = y_2 - \beta_0 - x_2 \beta + \theta_1 \epsilon_1 + \theta_2 \tilde{\epsilon}_0$$

$$\epsilon_3 = y_3 - \beta_0 - x_3 \beta + \theta_1 \epsilon_2 + \theta_2 \epsilon_1$$

⋮

$$\epsilon_n = y_n - \beta_0 - x_n \beta + \theta_1 \epsilon_{n-1} + \theta_2 \epsilon_{n-2}$$

These model equations may then be analysed as described in Section 2.6.3.

## 2.6 Method of Estimation

As a method of estimation we considered Estimated Generalized Least Squares (EGLS) and Maximum Likelihood (ML). Details of these two methods for three different Procedures is given below.

### 2.6.1 General Procedure

#### 2.6.1.1 Estimated Generalized Least Square (EGLS):

Here, we transform the original observations replacing  $\theta$  by its estimate,  $\hat{\theta}$ . One possible estimator of  $\theta$  is,

$$\hat{\theta} = \frac{1 - \sqrt{(1 - 4\hat{\rho}_1^2)}}{2\hat{\rho}_1}$$

where,  $\hat{\rho}_1$  is the autocorrelation coefficient of OLS errors for one period lag. This estimator is used by MacDonald and MacKinnon, Judge and others for MA(1). For MA(2) a similar kind of estimate can be obtained by solving the nonlinear equations,

$$\rho_1 = \frac{-\theta_1(1-\theta_2)}{1+\theta_1^2+\theta_2^2}$$

$$\rho_2 = \frac{-\theta_2}{1+\theta_1^2+\theta_2^2}$$

for  $\theta_1$  and  $\theta_2$ , but the estimated values for  $\hat{\theta}_1$  and  $\hat{\theta}_2$  can be read off directly from 'Chart C' at the end of the book by Box and Jenkins (1970, p.519). According to Judge (1980, p.198) these estimates are inefficient relative to the NLS estimator and also it is very difficult to get the estimates for  $\theta$ 's as the order of the moving average process increases. Therefore, it could be suggested that, since there are many computer packages (e.g., MINITAB) which can easily give us the NLS estimate of  $\theta$ 's we can use these packages to get the estimates of  $\theta$ 's to use in the transformation.

Therefore, the EGLS estimate of regression coefficient is  $\hat{\beta} = (\hat{X}'\hat{X})^{-1}\hat{X}'\hat{Y}$  where,  $\hat{X} = TX$ ,  $\hat{Y} = TY$  and  $T$  is the estimated transformation matrix  $T = T(\hat{\theta})$ .

#### 2.6.1.2 Maximum Likelihood:

Under the assumption of normality, log-likelihood function deleting the constant can be written as,

$$L(\beta, \theta, \sigma^2) = -\frac{n}{2} \ln \sigma^2 - \frac{1}{2} \ln |V| - \left[ \frac{1}{2\sigma^2} (\hat{Y} - \hat{X}\beta)' (\hat{Y} - \hat{X}\beta) \right].$$

Where  $V$  is defined as before the estimated covariance matrix  $\Omega$ . Now replacing  $\sigma^2$  by its ML estimator,

$$\hat{\sigma}^2 = \frac{(\hat{Y} - \hat{X}\hat{\beta})' (\hat{Y} - \hat{X}\hat{\beta})}{n},$$



after some simplifications the log-likelihood function becomes,  

$$L(\beta, \theta) = - \frac{n}{2} \ln[ |V|^{(1/n)} (\mathbf{Y} - \mathbf{X}\beta)' (\mathbf{Y} - \mathbf{X}\beta) ].$$

Therefore maximizing the above likelihood is equivalent to minimizing  $|V|^{(1/n)} (\mathbf{Y} - \mathbf{X}\beta)' (\mathbf{Y} - \mathbf{X}\beta)$  with respect to  $\beta$  and  $\theta$ . A computer program e.g., FORTRAN subroutine from NAG, for minimizing the above objective function, can give us the ML estimate of  $\beta$  and  $\theta$ , where  $\beta$  and  $\theta$  are the vectors of regression and MA parameters respectively.

## 2.6.2 MacDonalD & MacKinnon Procedure

### 2.6.2.1 Estimated Generalized Least Square (EGLS):

We will use the same estimator of  $\theta$ 's as we did in the General Procedure, and using the transformation technique discussed in Section 2.5.2 we have the transformed model,  

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{Z}'\eta + \epsilon.$$

Now by applying Ordinary Least Squares on transformed variables we can obtain the estimates of  $\beta$  and  $\eta$ .

### 2.6.2.2 Maximum Likelihood:

Under the assumption of normality, log-likelihood function after deleting constant term can be written as,

$$L(\beta, \theta, \eta) = - \frac{n}{2} \ln[ |V|^{(1/n)} (\mathbf{Y} - \mathbf{X}\beta - \mathbf{Z}'\eta)' (\mathbf{Y} - \mathbf{X}\beta - \mathbf{Z}'\eta) ]$$

where  $\beta$ ,  $\theta$  and  $\eta$  are vectors of parameters.

Therefore, maximizing the above log-likelihood is equivalent to minimizing  $|V|^{(1/n)} (\mathbf{Y} - \mathbf{X}\beta - \mathbf{Z}'\eta)' (\mathbf{Y} - \mathbf{X}\beta - \mathbf{Z}'\eta)$  with respect to  $\beta$ ,  $\theta$  and  $\eta$ .

### 2.6.3 Phillips Procedure

#### 2.6.3.1 Estimated Generalized Least Square (EGLS) (or NLS):

Since the sum of squares of errors becomes  $\epsilon' \epsilon + \tilde{\epsilon}' \tilde{\epsilon}$  and we can calculate them recursively as shown in Section 2.5.3, we can minimize the above sum of squares with respect to  $\beta$ ,  $\theta$  and  $\tilde{\epsilon}$  using a minimization program to obtain the EGLS (or NLS) estimates of the parameters.

#### 2.6.3.2 Maximum Likelihood :

Under the assumption that  $\epsilon_t$  are normally distributed, the log-likelihood function can be written as,

$$L(\beta, \theta, \tilde{\epsilon}) = -\frac{n}{2} \ln(2\pi\sigma^2) - \frac{1}{2} \ln|V| - \frac{\epsilon' \epsilon + \tilde{\epsilon}' \tilde{\epsilon}}{2\sigma^2}$$

Now, replacing  $\sigma^2$  by its ML estimator

$$\hat{\sigma}^2 = \frac{\epsilon' \epsilon + \tilde{\epsilon}' \tilde{\epsilon}}{n},$$

the log-likelihood function can be written as

$$L(\beta, \theta, \tilde{\epsilon}) = -\frac{n}{2} \ln[|V|^{(1/n)} (\epsilon' \epsilon + \tilde{\epsilon}' \tilde{\epsilon})].$$

Therefore, maximizing the above log-likelihood is equivalent to minimizing  $|V|^{(1/n)} (\epsilon' \epsilon + \tilde{\epsilon}' \tilde{\epsilon})$  with respect to  $\beta$ ,  $\theta$  and  $\tilde{\epsilon}$ .

## 2.7 SIMULATION EXPERIMENT :

The simulation experiment is carried out for small samples as well as for moderate samples, with MA(2) error process in a regression model, to compare the performance of three different procedures. For small sample we choose size 10, and 50 as a moderate size.

In this experiment for simplicity the regression model considered is,

$$Y_t = \beta_0 + \beta_1 X_t + U_t; \text{ where, } U_t = \epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2}$$

In the above model we specified the value of the regression parameters  $\beta_0 = 0$  and  $\beta_1 = 5$ .

For each of the sample sizes both the moving average parameters, namely  $\theta_1$  and  $\theta_2$  are made to vary as below;

we choose  $\theta_2$  as  $-0.50$  and  $-0.80$ . For each  $\theta_2$ ,  $\theta_1$  takes values  $\pm 1.45$  and  $\pm 1.15$ . Therefore we have eight different pairs of  $\theta$ 's and for each pair of  $\theta$ 's we generate 100 samples (repetitions) of same size.

The different stages involved in this experiment are discussed below:

### a) Stage-I :

This stage involves computing of  $u_t$  from  $\epsilon_t$  a random variable with mean zero and unit variance from a normal population using a specified pair of  $\theta$ 's.  $\epsilon_t$  is obtained by using the MINITAB computer package. Then using the specified values of regression parameters and a set of  $X_t$  adding with  $u_t$

we can obtain the response vector  $Y_t$  for each pair of  $\theta$ 's. Thus we are introducing MA(2) error in our regression model.

**b) Stage-II :**

In this stage we regress  $Y$  on  $X$  (OLS) to estimate the regression error  $\hat{u}$ , which will be used to identify and estimate MA parameters. To identify the order of the process we use the identification criteria discussed in Chapter-2 and estimate the  $\theta$ 's from regression errors using BMDQ2T computer package.

**c) Stage- III :**

Now, we apply three procedures, which were discussed in Section 2.5 and obtain the estimated values of the regression coefficients for each pair of  $\theta$ 's and for two different sample sizes mentioned above.

## **2.8 Empirical Results :**

Here, we have reported the results found by the simulation experiment. We will discuss the performance of different procedures through efficiency in two ways, efficiency due to variance of regression coefficient and efficiency due to computational time (i.e., *cpu* time).

From Appendix A.1 to Appendix A.8, it can be inferred that the regression coefficients are virtually unbiased regardless of the procedure and of the  $\theta$  values considered. But it can be noticed that as the sample size increases the amount of bias decreases though the amount of bias is very small even for

sample size 10.

Concerning efficiency (due to variance), we use the relative efficiency of regression coefficients of a particular procedure with respect to the OLS (i.e., estimated variance of regression coefficient obtained by OLS divided by the estimated variance of regression coefficient obtained by using a particular procedure from 100 repetitions).

For the sample size 50, all three procedures are more or less the same in efficiency gain, though GNL (General) performs better in most of the cases and all three procedures perform better than OLS.

When sample size is 10 (Appendix A.1 to Appendix A.4), the efficiency gain by GNL (General) is higher over MM (MacDonald and MacKinnon) for both EGLS and ML and for all the pairs of  $\theta$ 's considered, but, GNL(ML) is slightly less efficient than PHL(ML) (Phillips) in two cases, equal in two cases and better in four cases, though the difference is very small. It is also observed that PHL(ML) gives regression estimates identical with GNL(ML) in more than 95% of the samples. It is hard to justify the comparison of PHL(NLS) with EGLS of other two, because PHL(NLS) has several iterations, whereas GNL and MM need only two OLS regression for EGLS. But it can be noticed that PHL(NLS) does better than GNL(EGLS) and MM(EGLS) when  $\theta_1$  takes positive value and does poorly than GNL(EGLS) when  $\theta_1$  takes negative value. For the same sample size ( $n=10$ ) efficiency gain is higher by ML than by EGLS for both the procedures (GNL and MM) except the cases

(at Appendix A.2)  $\theta_1 = -1.15$ ,  $\theta_2 = -.50$  and (at Appendix A.4) when  $\theta_1 = -1.15$  and  $\theta_2 = -.80$ , where OLS performs better than MM; the reason may be that most of the time  $\theta_2$  lies on the boundary of the invertibility region. In this context it can be mentioned that MacDonald and MacKinnon in their paper (1985, for MA-1) also found that OLS performs better than EGLS when  $\theta$  is negative and MM(EGLS) performs very badly when  $\theta = -.80$  and  $n = 100$ .

Another interesting behaviour is that, when sign of  $\theta$ 's are different i.e., when  $\theta_1$  is positive and  $\theta_2$  negative (we consider only negative  $\theta_2$  in this paper), the efficiency gain with intercept and slope are same, though the tendency is higher for intercept. But, when the sign of  $\theta_1$  is negative, efficiency with slope is always higher than with intercept and the difference is much more as the sample size increases.

Since the results obtained form a complicated structure we also use covariance analysis to summarize our results. In this summarization we shall use terms like "significance" rather loosely. This is not a pretense of probability sampling or of genuine inference, but only a means of summarizing the extent to which the pattern emerging from our simulations conforms to the additive ANOCOVA model described. The dependent variables considered for the model below are:

a) Relative efficiency for intercept, b) Relative efficiency for slope, c) Bias for intercept and d) Bias for slope (absolute bias is used). The independent variables are sample size, value of  $\theta_2$ , sign of  $\theta_1$ , magnitude of  $\theta_1$  and six different methods as

a categorical variable, namely, GNL(EGLS), GNL(ML), MM(EGLS), MM(ML), PHL(NLS) and PHL(ML). In more detail, in Appendices B.1 and B.2 the additive model,

$$Y = \text{constant} + T_i + (\text{SMSZ} \times \text{coefficient}) + (\text{TH2} \times \text{coefficient}) + (\text{STH1} \times \text{coefficient}) + (\text{MTH1} \times \text{coefficient}) + \text{error}$$

was imposed upon our simulated results, as a mechanical device for organizing and presenting our results. Here,  $\sum T_i = 0$ , so that  $T_i$  is the "effect" of using method number as below,

$$T_1 = \text{GNL(EGLS)}$$

$$T_2 = \text{GNL(ML)}$$

$$T_3 = \text{MM(EGLS)}$$

$$T_4 = \text{MM(ML)}$$

$$T_5 = \text{PHL(NLS)}$$

$$T_6 = \text{PHL(ML)}$$

and also,

SMSZ = sample size

TH2 = specific value of  $\theta_2$

STH1 = sign of  $\theta_1$

MTH1 = magnitude of  $\theta_1$

In Appendices B.3 to B.6, similar but separate analyses were performed, first for all the simulated samples of size 10 and then for all the simulated samples of size 50.

When bias is concerned from the ANOCOVA table (Appendix B.2), we see that, for all the six methods the value of the estimated effects of method on bias are so small that they can be ignored. The significance levels suggest that the

coefficients are not significantly different from zero for both intercept and slope. But for sample size (SMSZ, which is an independent variable), significance levels indicate that sample size has significant effect on bias, though it is very very small. The sign of the coefficient is negative, indicating the inverse relationship with the dependent variable, so that as the sample size increases bias decreases, which supports our previous discussions, which were based on informal examination without using ANOCOVA. Again the significance level for sign of  $\theta_1$  (for intercept) tells us that the sign of  $\theta_1$  has influence on the amount of bias. But, since all the estimated coefficients for bias as dependent variable are so small we can ignore their effects.

In the case of efficiency gain for both intercept and slope by different procedures, it can be observe that (Appendix B.3) the estimated coefficient .25640 for GNL(ML) and .25265 for PHL(ML) (for intercept), and (for slope) the estimated coefficient .56810 for GNL(ML) and .49023 for PHL(ML) with significant t-statistic puts the GNL(ML) in the top rank for the sample size 10. This can again be verified by examining Appendix A.1 to Appendix A.4, where for four pairs of  $\theta$ 's GNL(ML) does better than PHL(ML), for two pairs of  $\theta$ 's they are same and for two pairs of  $\theta$ 's GNL(ML) could not perform better than PHL(ML), though the amount of relative efficiency is almost the same. Therefore, considering all the eight pairs of  $\theta$ 's GNL(ML) is marginally superior than PHL(ML).



For the same sample size, both GNL(ML) and PHL(ML) performs better than MM(ML) for all pairs of  $\theta$ 's. Again considering EGLS, GNL performs better than MM for both the sample sizes and for all pairs of  $\theta$ 's considered (Appendices A.1 to A.8).

The significance levels (Appendix B.1) for sample size (SMSZ) in both intercept and slope indicates the certainty of sample size's effect on efficiency gain, though the estimated values of the coefficients are very small, but the positive sign indicates that as sample size increases efficiency gain also increases. The magnitude of  $\theta_1$  (MTH1) also has a significance level which shows strong suggestion of positive effect upon efficiency gain and the value of the estimated coefficient is as high as 23.982 (Appendix B.1) for slope efficiency ( also high for intercept efficiency ) with positive sign. This indicates that for high values of  $\theta_1$  efficiency gain is higher for both slope and intercept, which is very effective when sample size gets large (see Appendix B.3 and Appendix B.5).

Sign of  $\theta_1$  (STH1) also has a definite effect for intercept efficiency gain (looking through significance level) and the value of the coefficient is 2.657 (Appendix B.1) with positive sign. This suggests that, as long as the sign for  $\theta_1$  is positive, efficiency gain is higher than for negative and this is true for both the sample sizes (Appendix B.3 and Appendix B.5) and for sample size 50 the estimated coefficient is as high as 4.905. It can be further confirmed by looking through the Rel. Efficiency table (Appendix A.1 to Appendix A.4), for sample

size 10 (except for one case  $\theta_2 = -.50$ ,  $\theta_1 = \pm 1.45$  in GNL(ML)) that efficiency gain with intercept is higher when  $\theta_1$  is positive. This can also be seen when  $n=50$  (Appendix A.5 to Appendix A.8). It is observed that for all pairs of  $\theta$ 's and for each method, efficiency gain with intercept is higher for positive  $\theta_1$  and the difference is more as the magnitude of  $\theta_1$  goes up. This is also true for slope efficiency i.e., as the magnitude of  $\theta_1$  increases the amount of relative efficiency increases (Appendix A.5 to Appendix A.8,  $n=50$ ) in every case and for all methods. This is further confirmed by the ANOCOVA table (Appendix B.5,  $n=50$ ), where with very much favourable significance level and an estimated coefficient of 47.128 the positive sign tells us that as  $\theta_1$  increases in magnitude efficiency gain also increases. But with slope efficiency for sign of  $\theta_1$  (Appendix B.1), significance level is not favourable and the estimated value of the coefficient has a negative sign. Therefore with a non-significant significance level and a small estimated coefficient (as  $-.765$ ), it appears that the sign of  $\theta_1$  virtually has no effect on efficiency gain with slope.

For the value of  $\theta_2$  (TH2), which appears as an independent variable in the model described above, the significance level (Appendix B.1, for the combined samples) confirms the effect on slope efficiency but not on intercept efficiency, though the estimated value of the coefficients is high for both with positive sign. Again, for sample size 10 (Appendix B.3) the estimated coefficients of the TH2 variable are very small and

the signs are negative, whereas for sample size 50 (Appendix B.5) the estimated coefficients are high with positive sign, for slope it is as high as 23.146 and is confirmed through significance level. Therefore, it is tempting to draw the conclusion with sample size 50, that as  $\theta_2$  increases (decreases in magnitude) the efficiency gain is higher, but if we look to the results in Appendix A.5 to Appendix A.8, specifically for slope, when  $\theta_1 = \pm 1.15$ , then as  $\theta_2$  goes from  $-0.50$  to  $-0.80$  efficiency increases (Appendix A.6 and Appendix A.8) and for  $\theta_1 = \pm 1.45$  efficiency decreases for the same change of  $\theta_2$  (Appendix A.5 and Appendix A.7). Therefore, it is difficult to draw any conclusion only with two values of  $\theta_2$  and at the same time ignoring  $\theta_1$ .

If we consider computing efficiency, i.e., the *cpu* time required for the computations, using the method EGLS for GNL and MM procedure, it seems that both procedures are more or less the same in *cpu* time requirement and they need a fraction of a second. Since ML or NLS needs several iterations and therefore more *cpu* time, we recorded the *cpu* time (in IBM-3081) for different procedures in Appendix A.1 to Appendix A.8. For sample size 10 GNL(ML) takes always less time than the others and in general all methods take less time when  $\theta_1$  is negative. If we consider rank of taking less *cpu* time, it is GNL(ML), PHL(ML) and MM(ML) respectively for both sample sizes. It is also observed that as sample size increases the time requirement is also increased, but the proportion of time requirement more or

less remains same among the three procedures.

Now, if we bring PHL(NLS) in consideration, we see that for sample size 10 GNL(ML) is always in the first rank (in respect of taking less time) and PHL(ML) is in the second position except the case  $\theta_1 = -1.15$  and  $\theta_2 = -.80$ , where PHL(NLS) takes the second position and MM(ML) is either in 3rd or 4th position. But, for sample size 50 PHL(NLS) is in the first position in six sets, GNL(ML) is in the second position in six sets and first in other two, whereas PHL(ML) takes the third position in six sets. Therefore, PHL(NLS) may have considerable attention for computing efficiency in large sample, but, considering efficiency due to variance simultaneously, GNL(ML) should be preferred over the others. However, MM(ML) did not perform very well in computing efficiency.

Therefore, from the above discussions, it appears that considering all factors GNL performs better than all others. Of course this kind of conclusion has limitations in the sense that the results may be sensitive to the particular model chosen and also the X values(given) and the MA parameters considered. In conclusion, considering all aspects, this simulated experiment is able to make a suggestion that GNL is to be preferred than the others for higher order MA process in regression errors.

## 2.9 Specification Error:

In a variety of specification errors, one kind is about the assumption of regression error, which costs on efficiency of regression estimates very much. If the assumed process is not the true process, the variance of the estimated regression coefficient will be biased, i.e., if the true process is (let us say), MA(2) and the assumed process is MA(1) or AR(1), the regression coefficient will be inefficient. G. S. Watson (1955) found the analytical expression of the lower bound to the efficiency of the regression estimates for the special case  $X'X = I$  and in continuation to that paper Watson and Hannan (1956) apply that lower bound for various choice of true error process and assumed process.

On an experimental basis, we did a similar kind of experiment to see how good the identification criteria discussed in Section 2.3.1 (Box-Jenkins approach) works for small samples such as 10. For large samples and even for moderate size as 50 Box-Jenkins identification criteria works very well. But for small sample using ACF (autocorrelation function) and PACF (partial autocorrelation function) it is sometimes hard to detect the order and the model of the process.

Therefore, after identifying the model using the criteria discussed in Section 2.3.1, we apply GNL(ML) procedure for correcting the regression errors when it is MA(1) and parallel to that for AR(1) we use Beach and MacKinnon's (1978) procedure

of Maximum Likelihood and we also apply correction for MA(2), since we know the true process is MA(2). We use only two pairs of MA parameters,

a)  $\theta_2 = -.50$  and  $\theta_1 = 1.15$

b)  $\theta_2 = -.50$  and  $\theta_1 = -1.15$

and the relative efficiencies of assumed (identified) process to the true process is given below:

	$\theta_2 = -.50$			
	$\theta_1 = 1.15$		$\theta_1 = -1.15$	
	$\beta_0$	$\beta_1$	$\beta_0$	$\beta_1$
	n1 = 30		n2 = 21	
REAR1	.9318	.9143	.9908	.9580
	n3 = 51		n4 = 36	
REMA1	.8588	.8413	.9918	.9851

From the above results we see that relative efficiency<sup>3</sup>

<sup>3</sup> Where, the relative efficiency REAR1 is defined as the estimated variance of true process MA(2) divided by the estimated variance of identified process AR(1) for the identified number of samples n1 and n2 for two pairs of  $\theta$ 's. Similarly, REMA1 is the relative efficiency defined as the estimated variance of true MA(2) process divided by the estimated variance of identified MA(1) process for the identified number of samples n3 and n4.

does not fall very much with respect to the true process, which tells us that the Box-jenkins identification criteria does not give us very poor result though it identifies the model wrongly for the number of cases mentioned in the table (i.e. the number of repetitions). But, since the results obtained only used a few cases (where it identifies the model other than MA(2)), these findings can not be very reliable and may not be justified for other cases. But the truth is that a wrongly identified model can give us a better estimate of the regression coefficient, though this is not always true.

*C H A P T E R 3***EXACT DETERMINANT OF COVARIANCE MATRIX**



### 3.1 Determinant of Covariance Matrix $\Omega$

For Maximum Likelihood method the determinant of  $\Omega$  plays an important role. Judge(1980,p.205) mentioned that evaluation of  $|\Omega|$  is a headache in the case of ML.

There have been a number of approaches for exact maximum likelihood and also for approximate ML, but the latter one is asymptotically the former one. The approximation can arise from the approximation in transformation or from the approximation in determinant. Box and Jenkins(1976,p.213) for univariate time series suggested an approximation of the latter kind by disregarding the the determinant. McLeod(1978) proposed an approximation of the determinant term and claims a closer approximation to the exact ML. Ansley(1976,p.59) discussed the fact that approximation by disregarding the determinant can lead to inferior estimates; he gave references to different Monte Carlo works.

### 3.2 Determinant from the Transformation of General Procedure

Ansley use the determinant proposed by Phadke and Kedem(1978) using library subroutine: It is,

$$\begin{aligned} |\Omega| &= |S'S| \\ &= |S'| |S| \\ &= (\text{product of diagonal elements in } S)^2 . \end{aligned}$$

This can be obtained from our square root matrix  $S$  discussed in Chapter 2, when we are using General Procedure, because the elements of  $S$  are already obtained for transformation. We just need to use the diagonal elements of  $S$ . Therefore, we are getting the determinant as a by-product. But, if one is not using General Procedure (GNL), then it is better to apply the technique discussed below to estimate the determinant of  $\Omega$ .

### 3.3 Exact Determinant Using $\Theta$ Matrix

We know from Osborn(1976,pp.76-77) that the determinant of  $\Omega$  can be written as  $|R'R|=|\Omega|$

where,

$$R = \begin{bmatrix} 0 & 0 & 0 & \dots & 0 & 1 \\ 0 & 0 & 0 & \dots & 1 & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots & \vdots \\ 0 & 0 & 1 & \dots & 0 & 0 \\ 0 & 1 & 0 & \dots & 0 & 0 \\ 1 & 0 & 0 & \dots & 0 & 0 \\ \theta_1 & \theta_2 & \theta_3 & \dots & \theta_{q-1} & \theta_q \\ \theta_1^2 + \theta_2 & \theta_1 \theta_2 + \theta_3 & \theta_1 \theta_3 + \theta_4 & \dots & \theta_1 \theta_{q-1} + \theta_q & \theta_1 \theta_q \\ \text{and so on} \end{bmatrix}$$

More precisely, the rows of  $R^4$  after  $q$  rows can be found as follows:

1. The  $(q+1, j)$ th element is the sum of  $\theta_1$  multiplied by  $(q, j)$ th element,  $\theta_2$  multiplied by  $(q-1, j)$ th element, ..... ,  $\theta_q$  multiplied by  $(1, j)$ th element.

2. The  $(q+2, j)$ th element is the sum of  $\theta_1$  multiplied by  $(q+1, j)$ th element,  $\theta_2$  multiplied by  $(q, j)$ th element, ..... ,  $\theta_q$  multiplied by  $(2, j)$ th element.

.

.

.

n. The  $(q+n, j)$ th element is the sum of  $\theta_1$  multiplied by  $(q+n-1, j)$ th element,  $\theta_2$  multiplied by  $(q+n-2, j)$ th element, ..... ,  $\theta_q$  multiplied by  $(q+n-q, j)$ th element.

In compact form  $R$  can be written as,

$$R = \begin{bmatrix} I & & & \\ & q \times q & & \\ & & A' & \\ & & & n \times q \end{bmatrix}$$

$(n+q) \times q$

---

<sup>4</sup>Instead of  $R$ , Osborn(1976) and Box & Jenkins(1970) use the notation  $X$ .

where,

$$\Gamma = \begin{bmatrix} 0 & 0 & 0 & \dots & 0 & 1 \\ 0 & 0 & 0 & \dots & 1 & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \dots & \vdots & \vdots \\ 1 & 0 & 0 & \dots & 0 & 0 \end{bmatrix}$$

$$\text{and } A = [ A_1 \quad A_2 \quad A_3 \quad \dots \quad A_k \quad \dots \quad A_n ]$$

where,  $A_i$ 's are column vectors, such that,

$$\begin{aligned} A_1 &= \Theta A_0 \\ A_2 &= \Theta A_1 \\ \vdots & \\ \vdots & \\ A_k &= \Theta A_{k-1} \\ \vdots & \\ A_n &= \Theta A_{n-1} \end{aligned} \quad A_0 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ \vdots \\ 0 \\ 0 \end{bmatrix}$$

$q \times 1$

$$\Theta = \begin{bmatrix} \theta_1 & 1 & 0 & 0 & \dots & 0 \\ \theta_2 & 0 & 1 & 0 & \dots & 0 \\ \theta_3 & 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \theta_{q-1} & 0 & 0 & 0 & \dots & 1 \\ \theta_q & 0 & 0 & 0 & \dots & 0 \end{bmatrix}$$

The above  $\Theta$  matrix is defined earlier (Section 2.5.2).

Now,  $R'R$  can be written as,

$$R'R = \begin{bmatrix} I' & A \end{bmatrix} \begin{bmatrix} I \\ A' \end{bmatrix} \quad (n+q) \times q$$

$$= [ I'I + AA' ]$$

$$= [ I + AA' ]$$

$q \times q$

Where,

$$AA' = A_1 A_1' + A_2 A_2' + \dots + A_n A_n' = \sum A_i A_i'$$

$$\begin{aligned}
\text{or, } AA' &= \theta A_0 A_0' \theta' + \theta A_1 A_1' \theta' + \dots + \theta A_{n-1} A_{n-1}' \theta' \\
&= \theta A_0 A_0' \theta' + \theta^2 A_0 A_0' (\theta^2)' + \dots + \theta^n A_0 A_0' (\theta^n)' \\
&= \sum_{i=1}^n \theta^i A_0 A_0' (\theta^i)'
\end{aligned}$$

Therefore,

$$\begin{aligned}
|R'R| &= |\Gamma'\Gamma + AA'| \\
&= |I + \sum_{i=1}^n \theta^i A_0 A_0' (\theta^i)'| \\
&= |I + \sum_{i=1}^n \theta^i M_1 (\theta^i)'| \tag{3.1}
\end{aligned}$$

where,  $M_1 = A_0 A_0' = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & 0 & \dots & 0 \end{bmatrix}$

we can also use ,

$$|R'R| = \left| I + \sum_{i=1}^n A_i A_i' \right| \quad (3.2)$$

where  $|R'R|$  is of order  $qxq$ .

Thus, we are reducing our work to dealing with a  $qxq$  matrix rather than a big matrix  $(n+q)qxq$ , and usually the order  $q$  of the MA process is very smaller relative to  $n$ . Therefore, one can use either (3.1) or (3.2) above to get the determinant of covariance matrix. The computer program in FORTRAN to calculate the determinant is given in Appendix H.2. It is also simple in the sense that here we do not need to form the complicated matrix  $R$ .

It can be noticed that for the 1st order MA process our determinant becomes,

$$|R'R| = \left| 1 + \sum_{i=1}^n \theta^i M1 (\theta^i)' \right|$$

$$= \left| 1 + \sum_{i=1}^n \theta^i (\theta^i)' \right|$$

$$= \left| 1 + \sum_{i=1}^n \theta^{2i} \right|.$$

Since  $\theta$  is a scalar and  $M_1$  is 1,

$$|R'R| = 1 + \theta^2 + \theta^4 + \dots + \theta^{2n}$$

$$= \frac{1 - \theta^{2(n+1)}}{1 - \theta^2}$$

which was reported by Box and Jenkins (1976, p.272) and also by Balestra (1980, p.381).



*C H A P T E R 4*

APPROXIMATE ESTIMATOR FOR THE HIGHER ORDER MOVING AVERAGE  
PROCESS IN REGRESSION ERRORS

#### 4.1 Approximate Transformations for MA Process:

Along with the exact transformations, researchers also get themselves involved with approximate transformations. It is obvious that we can get better results using exact transformation than using approximate, but the latter is asymptotically the former. Approximate transformations are used for computational simplicity. Balestra(1980.pp.390-394) proposed an approximate transformation matrix  $T^*$  of dimension  $(n-1) \times n$  for MA(1) and showed that the transformation can be carried out recursively as,

$$\tilde{\epsilon}_i = \sum_{j=1}^i c^{i-j} \epsilon_j, \quad i \geq 2,$$

where,  $c$  is the MA parameter;

Thus, Balestra is losing the first observation. With that in mind later Park and Heikes(1983) propose another approximate transformation  $P$  augmenting  $T^*$  by a first row consisting of 1 in the 1st column and zeroes in others for the same order of MA process, where  $P$  is of order  $n \times n$ .

In both the papers, they considered a simple regression equation with constant term only and they derived the analytical expression for the variance of the estimated constant using their transformation, which is reproduced here below.

**Balestra**For  $c < 1$ :

OLS:

$$\sigma^2 \left[ \frac{(1-c)^2}{n} + \frac{2c}{n^2} \right]$$

APX:

$$\sigma^2 \left[ \frac{(1-c)^2}{n-1-2c^2r_1+c^4r_2} + \frac{(1-c)^2c^4(r_1-c^2r_2)^2}{(n-1-2c^2r_1+c^4r_2)^2} \right]$$

where  $r_1 = (1-c^{n-1})/(1-c)$ , and  $r_2 = (1-c^{2n-2})/(1-c^2)$ .

AITKEN (GLS):

$$\sigma^2 \frac{(1-c)^2}{n} \left[ 1 - \frac{2c(1-c)^n}{n(1-c)(1+c^{n+1})} \right]^{-1}$$

For  $c=1$ :

$$\text{OLS: } \sigma^2 \frac{2}{n^2}$$

$$\text{APX: } \sigma^2 \frac{3n(n^2-1)(3n+10)}{[(2n+1)(n+1)n-6]^2}$$

$$\text{AITKEN (GLS): } \sigma^2 \frac{n}{n(n+1)(n+2)}$$

**Park and Heikes**

For  $c \leq 1$ :

$$\sigma^2 \left[ \frac{(1-c^2)(1-c)^2}{H} + \frac{c^2(1-c)^2(1-c^n)^2(1-c^{n+1})^2}{H^2} \right]$$

where  $H = n(1-c^2) - c(1-c^n)(2+c-c^{n+1})$

For  $c = 1$ :

$$\sigma^2 \left[ \frac{(9n+3)(n+2)}{n(2n+1)^2(n+1)} \right]$$

where, APX stands for Approximate Transformation by Balestra (hereafter BL) and MAPX is the Modified Approximate Transformation by Parks and Heikes.

On the basis of numerical computations of relative efficiencies, for six sample sizes and six values of  $c$ , BL finds that:

- (a) OLS performs better than APX when  $c$  is low.
- (b) APX performs extremely well for  $c$  around 0.5, but does very poorly for high values of  $c$ , even in larger sample sizes.

On the other hand, Park and Heikes using their transformation observe that:

- (a) For  $c$  around 0.5, MAPX performs extremely well and better than APX for  $c \leq 0.5$ .
- (b) Like APX, MAPX does not perform well for high values of  $c$ ,



They found the variance of the estimate of constant for the same 'intercept only' model is,

$$\sigma^2 \left[ \frac{(1-c)^3 (1+c^2)^2}{A} + \frac{c^4 (1+c^2) (1-c)^2 (1-c^{n-1})^2 (1-c^{n+1})^2}{A^2} \right]$$

where,  $A = (1-c)(1+c^2)(n+nc^2-2c) - c^2(1+c)(1-c^{n-1})(c^2-c^{n+1}+2)$

For  $c=1$ , the variance is:

$$\sigma^2 \frac{6\{2n(4n^2-1)+3(n^2-1)^2+6\}}{\{n(4n^2-1)+3\}^2}$$

They also compute numerically the relative efficiency for APX, MAPX and FMAPX (Further Modified Approximate transformation) to the GLS for eight different sample sizes and six values of  $c$  in Appendix C.1, which shows that FMAPX is a better approximation in the sense that, it performs better than the previous two for all values of  $c$  and the sample sizes considered. It can be found that FMAPX performs as well as GLS for  $c$  less than 0.7. In Appendix C.2 (reproduced here), they showed the relative efficiency of APX and MAPX to FMAPX.

The approximation proposed by Park and Heikes is also discussed by Pollock (1979, pp.203-207) for MA(1) process.

Both the previous two approximations i.e., APX and MAPX uses the covariance matrix whose determinant is 1 as shown by Balestra (1980, p.390). Therefore, ML estimate for these is

equivalent to the least squares. For the FMAPX the determinant can be found as  $1+c^2$ . Therefore, ML estimate can be obtained by maximizing the log-likelihood,

$$L(\beta, c, \sigma^2) = \frac{n}{2} \ln \sigma^2 - \frac{1}{2} \ln (1+c^2) - \frac{1}{2\sigma^2} (\hat{Y} - \hat{X}\beta)' (\hat{Y} - \hat{X}\beta)$$

using the ML estimate of  $\sigma^2$ , the concentrated log-likelihood function becomes,

$$L(\beta, c) = \frac{n}{2} \ln [1+c^2]^{(1/n)} (\hat{Y} - \hat{X}\beta)' (\hat{Y} - \hat{X}\beta)$$

where,  $\hat{X}$  and  $\hat{Y}$  are the transformed variables.





Therefore, we can see that only the first four elements in the left upper corner of  $V_0$  are approximated with the corresponding elements of exact  $V$  matrix; elsewhere the elements are identical.

The above transformation matrix  $T$  can be obtained starting with 1's in the main diagonal and all zeroes above the main diagonal. Then the (2,1)th element is the multiplication of  $\theta_1$  with the (1,1)th element; (3,1)th element is the sum of multiplication of  $\theta_1$  with (2,1) and  $\theta_2$  with (2,2); (3,2)th element is the multiplication of  $\theta_1$  with (2,2); ....; (k,1)th element is the sum of multiplication of  $\theta_1$  with (k-1,1),  $\theta_2$  with (k-1,2), .....,  $\theta_{k-1}$  with (k-1,k-1)th element; similarly (k,2)th element is the sum of multiplication of  $\theta_1$  with (k-1,2),  $\theta_2$  with (k-1,3), .....,  $\theta_{k-2}$  with (k-1,k-1); and so on.

But, more simply, we can obtain the transformation matrix after obtaining the 1st column, because, all the diagonal elements of a diagonal on and below the main diagonal are the same. Therefore, first element of 1st row will construct main diagonal, first element of 2nd row will construct 2nd diagonal, first element of 3rd row will construct 3rd diagonal and so on. It is easy to obtain the first column recursively as follows:

a) 1st element is 1.

b) 2nd element is  $\theta_1(a)$ .

c) 3rd element is  $\theta_1(b) + \theta_2(a)$ .

d) 4th element is  $\theta_1(c) + \theta_2(b) + \theta_3(a)$ .

.

.

.

q) qth element is  $\theta_1(q-1) + \theta_2(q-2) + \dots + \theta_{q-1}(a)$ .

q+1) (q+1)th element is  $\theta_1(q) + \theta_2(q-1) + \dots + \theta_q(a)$ .

q+2) (q+2)th element is  $\theta_1(q+1) + \theta_2(q) + \dots + \theta_q(b)$ .

.

.

n) nth element is  $\theta_1(n-1) + \theta_2(n-2) + \dots + \theta_q(n-q)$ .

If we want to form the transformation matrix, the above procedure is convenient, but, if our objective is only to transform the variable then the former procedure seems more convenient in respect of computer space, because, all we need to store is the previous row, to form the present row.

But, for practical purposes, we can use the recursive transformation discussed below.

Since our purpose is to transform the variable, the following super-simple recursive procedure can be adopted.

$$\hat{Y}_1 = Y_1$$

$$\hat{Y}_2 = Y_2 + \theta_1 \hat{Y}_1$$

$$\hat{Y}_3 = Y_3 + \theta_1 \hat{Y}_2 + \theta_2 \hat{Y}_1$$

.

.

$$\hat{Y}_q = Y_q + \theta_1 \hat{Y}_{q-1} + \dots + \theta_{q-1} \hat{Y}_1$$

$$\hat{Y}_{q+1} = Y_{q+1} + \theta_1 \hat{Y}_q + \dots + \theta_q \hat{Y}_1$$

.

.

$$\hat{Y}_n = Y_n + \theta_1 \hat{Y}_{n-1} + \dots + \theta_q \hat{Y}_{n-q}$$

Thus, for example, for MA(2) process the transformation can be carried out as,

$$\hat{Y}_1 = Y_1$$

$$\hat{Y}_2 = Y_2 + \theta_1 \hat{Y}_1$$

$$\hat{Y}_3 = Y_3 + \theta_1 \hat{Y}_2 + \theta_2 \hat{Y}_1$$

$$\hat{Y}_4 = Y_4 + \theta_1 \hat{Y}_3 + \theta_2 \hat{Y}_2$$

.

.

$$\hat{Y}_n = Y_n + \theta_1 \hat{Y}_{n-1} + \theta_2 \hat{Y}_{n-2}$$

Therefore, if  $\theta$  is unknown, using the estimated value of  $\theta$ 's we can transform the variables using our above recursive transformation procedure to perform two-stage EGLS.

### 4.3 Empirical Results:

We have computed the relative efficiencies of MAPX and GLS to OLS <sup>5</sup> and also the relative efficiency of GLS to MAPX<sup>6</sup> for the model considered by Balestra and hence we can not generalize the results obtained, since the performance of the approximate estimator also depends on the independent variables. But from this simple model we can at least show some of the criteria of an approximate estimator. We have computed the relative efficiency numerically (presented in Appendix E.1 to Appendix E.6, Appendix F.1 to Appendix F.6 and Appendix G.1 to Appendix G.6) for the pairs of  $\theta$ 's as below:

$$-1.90 \leq \theta_1 \leq 1.90$$

$$-0.95 \leq \theta_2 \leq 0.95$$

both with an increment of 0.10 and also,

$$\theta_2 + \theta_1 \leq 0.95$$

$$\theta_2 - \theta_1 \leq 0.95$$

The expression for the variances of the estimated constant are given below:

$$\text{OLS: } \sigma^2 (X'X)^{-1} X'VX (X'X)^{-1} \quad (\text{see Johnston p.246})$$

<sup>5</sup>i.e., the variance of regression coefficient obtained by MAPX is divided by the variance obtained by OLS, and the variance of GLS divided by the variance of OLS.

<sup>6</sup>i.e., the variance of GLS divided by the variance of MAPX.

$$\sigma^2 \sum_{i=1}^n \sum_{j=1}^n v_{ij}$$

$$= \frac{\quad}{n^2} \quad , n \text{ is the sample size.}$$

$$\text{MAPX: } \sigma^2 (\mathbf{X}'\mathbf{V}_0^{-1}\mathbf{X})^{-1} \mathbf{X}'\mathbf{V}_0^{-1}\mathbf{V} \mathbf{V}_0^{-1}\mathbf{X} (\mathbf{X}'\mathbf{V}_0^{-1}\mathbf{X})^{-1}$$

$$= \sigma^2 (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'(\mathbf{T}\mathbf{V}\mathbf{T}')\mathbf{X} (\mathbf{X}'\mathbf{X})^{-1}$$

where,  $\mathbf{X}$  is the transformed observations using the approximate transformation and  $\mathbf{V}_0^{-1}$  is the inverse of the approximate  $\mathbf{V}_0$  matrix which is defined above.

$$\text{GLS: } \sigma^2 (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}$$

$$= \sigma^2 (\mathbf{X}'\mathbf{X})^{-1}$$

where,  $\mathbf{X}$  is the exact transformation using the transformation matrix  $\mathbf{S}$  discussed in Chapter 2.

From the results presented in Appendix F.1 to Appendix F.6 for six different sample sizes, we observe that,

a) MAPX does extremely well with respect to OLS when  $\theta_1$  and  $\theta_2$  both are positive and sample size is greater than 10; for sample size 10 or less its performance is not very good.

b) For the region of  $\theta_2$  in between -0.55 and 0.0 (inclusive), with positive  $\theta_1$  its performance is very good for all sample sizes considered and as sample size increases the performance is excellent, specifically for higher values of  $\theta_1$ . For negative  $\theta_1$  it does not perform very well except for the couple of points in the region ,

$$-.60 \leq \theta_1 \leq 0.0$$

$$-.35 \leq \theta_2 \leq 0.35$$

c) Again its performance can be appreciated in the (tables, i.e. Appendix F.1 to Appendix F.6) lower triangle of the reactangle bounded by the region,

$$0.0 \leq \theta_1 \leq 1.90$$

$$-.95 \leq \theta_2 < -.55$$

for sample sizes greater than 10 and for size 10 or less the performance is reasonable.

Compared with GLS, the approximate estimator does well (middle of the Appendix G.1 to Appendix G.6) and as the sample size increases the performance increases too. If we compare Appendix G.1 and Appendix G.6, we see that in the middle of tables the number of ones (1.000) increases in substantial amount. Again comparing Appendix G.5(n=50) and Appendix G.6(n=100) the improvement of approximate estimator can be observed very clearly, specifically the lower part of the table, i.e., when both  $\theta$ 's are positive. Therefore, for negative  $\theta$ 's it is not doing very well.

As we said before, for the region,

$$-.30 \leq \theta_1 \leq 0.30$$

$$-.25 \leq \theta_2 \leq 0.25$$

(i.e., in the middle of the tables) there is no reason not to consider approximate estimator rather than exact (GLS) on the ground of computational simplicity and the above region of  $\theta$ 's becomes increases as sample size gets larger.

Thus, we can say that the approximate estimator does as well as exact for small values of  $\theta$ 's and can also be considered for large  $\theta$ 's when sample size is larger.

From Appendix E.1 to Appendix E.6, it is observed that OLS does worse with respect to GLS in the bottom line of the tables (i.e., for the highest values of  $\theta_1$ , for each  $\theta_2$ ) but does well in the middle of tables, where both  $\theta$ 's are near to zero, which means there is virtually no moving average effect and hence GLS merges to OLS. Another important aspect is that, as both the  $\theta$ 's tends to zero from both ends efficiency gain by GLS over OLS diminishes.

As we discussed in the previous Chapter, for high values of  $\theta_1$  (for intercept) efficiency gain over OLS is higher than for low values of  $\theta_1$ ; this can be reconfirmed from Appendix E.1 to Appendix E.6 and also it is to be observed that when the sign of  $\theta_1$  is positive (with high Value of  $\theta_1$ ) efficiency gain is much more than with negative  $\theta_1$ , which verifies our previous result.

CHAPTER 5

CONCLUSION



It is apparent from the results of simulation experiment from Appendix A.1 to Appendix A.8 that the General Procedure (GNL) performs better than the other two procedure for both sample sizes considered. It performs excellently for sample size 10. We have compared the relative efficiency of the estimated regression coefficient for a simple regression model with an intercept and a slope coefficient. The relative efficiency is defined as the estimated variance of the regression coefficient by OLS divided by the estimated variance of regression coefficient by three different procedures obtained from 100 repetitions. The results obtained are reported in Appendix A.1 to Appendix A.8, which are further analysed using analysis of covariance and the findings are reported in Appendix B.1 to Appendix B.6. It can be observed from Appendix A.1 to Appendix A.4, for sample size 10, that the General procedure (GNL) performs very well over the others. Which can easily be understood from Appendix B.3 (sample size 10) that the proposed transformation GNL(ML) does better than the others.

The efficiency due to computational time, i.e., the time required by the central processing unit (*cpu*) in the computer is reported in Appendix A.1 to Appendix A.8 and it is observed that GNL(ML) takes always less time than the other two for both the sample sizes considered.

In Chapter 3 we have shown the determinant for covariance matrix can be obtained as a by-product when General Procedure (GNL) is in consideration. We also have shown in details how an exact determinant can be obtained. It can be observed that the dimension of the matrix for which the determinant is to be calculated is reduced from  $(n+q) \times q$  to  $q \times q$ , which reduces our work and makes the computational effort much simpler. Because, the order of the moving average process  $q$  is very much smaller than the number of observations  $n$ , we have in practice. Therefore, one can use this exact determinant for any of the three procedures discussed in Chapter 2. Specifically, it is important to use this determinant when one is using MacDonald and MacKinnon Procedure or Phillips Procedure. Because, they do not have any determinant to be obtain as a by-product like General Procedure.

In Chapter 4, we have discussed the efficiency of proposed approximate estimator relative to the GLS and OLS. We have computed the relative efficiency for six different sample sizes. The results are tabulated from Appendix E.1 to Appendix E.6 for relative efficiency of GLS to OLS, from Appendix F.1 to Appendix F.6 for relative efficiency of MAPX to OLS and from Appendix G.1 to Appendix G.6 for relative efficiency of GLS to MAPX. It is to be observed from Appendix G.1 to Appendix G.6 that approximate estimator performs as well as exact estimator in the middle of the tables, i.e., when both the moving average parameters are small in magnitude. As sample size increases, i.e., as we move

from Appendix G.1 to Appendix G.6 the efficiency of the approximate estimator increases. Therefore, in that region of the MA parameters, i.e., in the middle of the tables the approximate transformation may be considered instead of exact transformation on the ground of computational simplicity.

## APPENDIX A.1

REL. EFFICIENCY AND BIAS OF $B_0$ AND $B_1$ ESTIMATES WITH RESPECT TO OLS									
n = 10	EGLS				ML				
	Bo		B1		Bo		B1		
	Eff.	Bias	Eff.	Bias	Eff.	Bias	Eff.	Bias	
GNL:	1.663	0.104	1.548	-.056	1.734	0.038	1.616	-.021	
MM:	1.449	0.068	1.316	-.033	1.607	-.006	1.469	.007	
PHL:(NLS)	1.770	0.124	1.582	-.072	1.709	0.028	1.592	-.017	

$\theta_1 = 1.45$	CPU TIME:( in milliseconds)			
	GNL(ml)	MM(ml)	PHL(ml)	PHL(nls)
$\theta_2 = -.50$	280.3	526.0	397.9	749.0

REL. EFFICIENCY AND BIAS OF $B_0$ AND $B_1$ ESTIMATES WITH RESPECT TO OLS									
n = 10	EGLS				ML				
	Bo		B1		Bo		B1		
	Eff.	Bias	Eff.	Bias	Eff.	Bias	Eff.	Bias	
GNL:	1.562	-.169	2.162	0.034	1.763	-.115	2.821	0.006	
MM:	1.260	-.195	1.585	0.012	1.344	-.133	1.641	-.022	
PHL:(NLS)	1.535	-.089	1.912	-.003	1.748	-.128	2.734	0.011	

$\theta_1 = -1.45$	CPU TIME:( in milliseconds)			
	GNL(ml)	MM(ml)	PHL(ml)	PHL(nls)
$\theta_2 = -.50$	166.3	519.0	288.7	557.0

## APPENDIX A.2

REL. EFFICIENCY AND BIAS OF $B_0$ AND $B_1$ ESTIMATES WITH RESPECT TO OLS									
n = 10	EGLS					ML			
	$B_0$		$B_1$			$B_0$		$B_1$	
	Eff.	Bias	Eff.	Bias	Eff.	Bias	Eff.	Bias	
	GNL:	1.914	0.030	1.783	-.025	2.218	0.023	2.103	-.022
MM:	1.688	0.019	1.510	-.020	2.188	-.039	2.004	0.012	
PHL:(NLS)	2.325	0.036	2.173	-.028	2.234	0.022	2.116	-.022	

$\theta_1 = 1.15$	CPU TIME:( in milliseconds)			
	GNL(ml)	MM(ml)	PHL(ml)	PHL(nls)
$\theta_2 = -.50$	280.8	520.0	383.0	545.0

REL. EFFICIENCY AND BIAS OF $B_0$ AND $B_1$ ESTIMATES WITH RESPECT TO OLS									
n = 10	EGLS					ML			
	$B_0$		$B_1$			$B_0$		$B_1$	
	Eff.	Bias	Eff.	Bias	Eff.	Bias	Eff.	Bias	
	GNL:	1.525	-.178	1.863	0.043	1.520	-.139	2.136	0.024
MM:	1.140	-.157	1.170	-.006	1.196	-.138	1.133	-.007	
PHL:(NLS)	1.329	-.096	1.560	0.001	1.544	-.131	2.168	0.022	

$\theta_1 = -1.15$	CPU TIME:( in milliseconds)			
	GNL(ml)	MM(ml)	PHL(ml)	PHL(nls)
$\theta_2 = -.50$	166.3	541.0	295.0	434.0

## APPENDIX A.3

REL. EFFICIENCY AND BIAS OF $B_0$ AND $B_1$ ESTIMATES WITH RESPECT TO OLS								
n = 10	EGLS				ML			
	$B_0$		$B_1$		$B_0$		$B_1$	
	Eff.	Bias	Eff.	Bias	Eff.	Bias	Eff.	Bias
GNL:	2.113	0.080	1.958	-.050	3.217	0.023	2.933	-.020
MM:	1.941	0.065	1.772	-.044	2.604	0.018	2.359	-.018
PHL:(NLS)	2.940	0.071	2.621	-.046	3.217	0.023	2.933	-.020

$\theta_1 = 1.45$	CPU TIME:( in milliseconds)			
	GNL(ml)	MM(ml)	PHL(ml)	PHL(nls)
$\theta_2 = -.80$	280.5	491.0	380.5	430.0

REL. EFFICIENCY AND BIAS OF $B_0$ AND $B_1$ ESTIMATES WITH RESPECT TO OLS								
n = 10	EGLS				ML			
	$B_0$		$B_1$		$B_0$		$B_1$	
	Eff.	Bias	Eff.	Bias	Eff.	Bias	Eff.	Bias
GNL:	1.528	-.173	2.422	0.027	1.707	-.144	3.927	0.019
MM:	1.161	-.139	1.488	-.027	1.398	-.116	1.525	-.036
PHL:(NLS)	1.506	-.074	1.959	-.020	1.702	-.136	3.449	0.009

$\theta_1 = -1.45$	CPU TIME:( in milliseconds)			
	GNL(ml)	MM(ml)	PHL(ml)	PHL(nls)
$\theta_2 = -.80$	156.3	455.0	298.0	415.0

## APPENDIX A.4

REL. EFFICIENCY AND BIAS OF $B_0$ AND $B_1$ ESTIMATES WITH RESPECT TO OLS								
n = 10	EGLS				ML			
	$B_0$		$B_1$		$B_0$		$B_1$	
	Eff.	Bias	Eff.	Bias	Eff.	Bias	Eff.	Bias
GNL:	2.246	-.006	2.214	-.013	3.042	-.009	3.136	-.012
MM:	1.740	0.003	1.535	-.021	2.169	0.017	1.979	-.030
PHL:(NLS)	2.763	0.018	2.829	-.023	3.042	-.009	3.136	-.012

$\theta_1 = 1.15$	CPU TIME:( in milliseconds)			
	GNL(ml)	MM(ml)	PHL(ml)	PHL(nls)
$\theta_2 = -.80$	266.8	547.0	374.7	506.0

REL. EFFICIENCY AND BIAS OF $B_0$ AND $B_1$ ESTIMATES WITH RESPECT TO OLS								
n = 10	EGLS				ML			
	$B_0$		$B_1$		$B_0$		$B_1$	
	Eff.	Bias	Eff.	Bias	Eff.	Bias	Eff.	Bias
GNL:	1.364	-.169	1.747	0.033	1.427	-.191	1.979	0.046
MM:	1.002	-.116	0.824	-.044	0.983	-.119	0.702	-.054
PHL:(NLS)	1.282	-.155	1.613	0.024	1.402	-.186	1.900	0.042

$\theta_1 = -1.15$	CPU TIME:( in milliseconds)			
	GNL(ml)	MM(ml)	PHL(ml)	PHL(nls)
$\theta_2 = -.80$	153.5	472.0	416.0	255.0

## APPENDIX A.5

REL. EFFICIENCY AND BIAS OF $B_0$ AND $B_1$ ESTIMATES WITH RESPECT TO OLS								
n = 50	EGLS				ML			
	Bo		B1		Bo		B1	
	Eff.	Bias	Eff.	Bias	Eff.	Bias	Eff.	Bias
GNL:	13.086	-.022	12.652	0.011	25.057	-.012	23.850	0.007
MM:	13.008	-.021	12.592	0.011	27.229	-.007	27.350	0.004
PHL:(NLS)	22.786	0.017	22.416	-.007	31.123	-.008	30.664	0.005

$\theta_1 = 1.45$	CPU TIME:( in milliseconds)			
	GNL(ml)	MM(ml)	PHL(ml)	PHL(nls)
$\theta_2 = -.50$	1807.0	3675.0	2581.0	2653.0

REL. EFFICIENCY AND BIAS OF $B_0$ AND $B_1$ ESTIMATES WITH RESPECT TO OLS								
n = 50	EGLS				ML			
	Bo		B1		Bo		B1	
	Eff.	Bias	Eff.	Bias	Eff.	Bias	Eff.	Bias
GNL:	1.841	0.005	12.632	0.014	1.861	0.030	47.952	0.003
MM:	1.768	-.000	11.627	0.014	1.819	0.025	47.362	0.004
PHL:(NLS)	1.470	0.096	39.547	0.004	1.880	0.029	47.878	0.004

$\theta_1 = -1.45$	CPU TIME:( in milliseconds)			
	GNL(ml)	MM(ml)	PHL(ml)	PHL(nls)
$\theta_2 = -.50$	1037.0	2188.0	1369.0	2400.0



## APPENDIX A.6

REL. EFFICIENCY AND BIAS OF $B_0$ AND $B_1$ ESTIMATES WITH RESPECT TO OLS									
n = 50	EGLS				ML				
	Bo		B1		Bo		B1		
	Eff.	Bias	Eff.	Bias	Eff.	Bias	Eff.	Bias	
GNL:	5.302	-.032	4.881	0.017	4.579	-.020	4.358	0.011	
MM:	5.178	-.032	4.768	0.018	4.386	-.023	4.189	0.013	
PHL:(NLS)	3.938	-.018	3.800	0.011	4.416	-.026	4.257	0.014	

$\theta_1 = 1.15$	CPU TIME:( in milliseconds)			
	GNL(ml)	MM(ml)	PHL(ml)	PHL(nls)
$\theta_2 = -.50$	1037.0	1841.0	1211.0	917.0

REL. EFFICIENCY AND BIAS OF $B_0$ AND $B_1$ ESTIMATES WITH RESPECT TO OLS									
n = 50	EGLS				ML				
	Bo		B1		Bo		B1		
	Eff.	Bias	Eff.	Bias	Eff.	Bias	Eff.	Bias	
GNL:	1.477	-.012	3.830	0.020	1.480	-.010	4.863	0.020	
MM:	1.442	-.013	3.609	0.016	1.466	-.012	4.740	0.015	
PHL:(NLS)	1.442	-.008	4.464	0.021	1.480	-.010	4.863	0.020	

$\theta_1 = -1.15$	CPU TIME:( in milliseconds)			
	GNL(ml)	MM(ml)	PHL(ml)	PHL(nls)
$\theta_2 = -.50$	377.0	800.0	505.0	353.0

## APPENDIX A.7

REL. EFFICIENCY AND BIAS OF $B_0$ AND $B_1$ ESTIMATES WITH RESPECT TO OLS								
n = 50	EGLS				ML			
	Bo		B1		Bo		B1	
	Eff.	Bias	Eff.	Bias	Eff.	Bias	Eff.	Bias
GNL:	10.167	-.005	9.311	0.004	13.175	-.007	12.823	0.005
MM:	9.609	-.005	8.985	0.004	13.596	-.004	14.060	0.004
PHL:(NLS)	13.488	-.007	13.128	0.005	14.746	-.006	14.732	0.004

$\theta_1 = 1.45$	CPU TIME:( in milliseconds)			
	GNL(ml)	MM(ml)	PHL(ml)	PHL(nls)
$\theta_2 = -.80$	1002.0	1485.0	1224.0	476.0

REL. EFFICIENCY AND BIAS OF $B_0$ AND $B_1$ ESTIMATES WITH RESPECT TO OLS								
n = 50	EGLS				ML			
	Bo		B1		Bo		B1	
	Eff.	Bias	Eff.	Bias	Eff.	Bias	Eff.	Bias
GNL:	1.816	0.003	9.526	0.016	1.703	0.008	12.167	0.014
MM:	1.784	0.006	8.482	0.010	1.618	0.011	10.734	0.009
PHL:(NLS)	1.675	0.015	12.463	0.011	1.692	0.010	12.559	0.013

$\theta_1 = -1.45$	CPU TIME:( in milliseconds)			
	GNL(ml)	MM(ml)	PHL(ml)	PHL(nls)
$\theta_2 = -.80$	451.0	999.0	806.0	216.4

## APPENDIX A.8

REL. EFFICIENCY AND BIAS OF $B_0$ AND $B_1$ ESTIMATES WITH RESPECT TO OLS								
n = 50	EGLS				ML			
	Bo		B1		Bo		B1	
	Eff.	Bias	Eff.	Bias	Eff.	Bias	Eff.	Bias
GNL:	5.829	-.016	5.710	0.011	7.667	0.001	7.755	0.003
MM:	5.746	-.018	5.448	0.012	6.824	0.010	6.816	-.002
PHL:(NLS)	5.620	0.014	5.547	-.003	7.167	0.004	7.101	0.001

$\theta_1 = 1.15$	CPU TIME:( in milliseconds)			
	GNL(ml)	MM(ml)	PHL(ml)	PHL(nls)
$\theta_2 = -.80$	716.0	1576.0	1310.0	371.0

REL. EFFICIENCY AND BIAS OF $B_0$ AND $B_1$ ESTIMATES WITH RESPECT TO OLS								
n = 50	EGLS				ML			
	Bo		B1		Bo		B1	
	Eff.	Bias	Eff.	Bias	Eff.	Bias	Eff.	Bias
GNL:	1.448	-.015	5.382	0.024	1.446	0.023	9.125	0.006
MM:	1.415	-.018	4.934	0.021	1.379	0.019	8.270	0.003
PHL:(NLS)	1.444	0.029	8.487	0.002	1.436	0.025	8.960	0.005

$\theta_1 = -1.15$	CPU TIME:( in milliseconds)			
	GNL(ml)	MM(ml)	PHL(ml)	PHL(nls)
$\theta_2 = -.80$	440.0	1125.0	927.0	223.0

APPENDIX B.1  
(ANOCOVA)  
(combined samples)

ANALYSIS OF VARIANCE OF REL. EFFICIENCY OF B <sub>0</sub>					
SOURCE	DF	SUM SQRS	MEAN SQR	F-STAT	SIGNIF
REGRESSION	9	1477.8	164.20	8.6434	.0000
FIRST 5 VARS	5	40.081	8.0162	.42198	.8322
ERROR	86	1633.7	18.997		
TOTAL	95	3111.5			

  

VARIABLE	COEFF	STD ERROR	T-STAT	SIGNIF
CONSTANT	-10.645	4.3843	-2.4279	.0173
101.T1	-.731	.9947	-.7349	.4644
102.T2	.439	.9947	.4410	.6603
103.T3	-.953	.9947	-.9580	.3408
104.T4	.327	.9947	.3285	.7433
105.T5	.460	.9947	.4621	.9632
106.T6	.873	.9947	.8772	.3828
5.SMSZ	.117	.2224	-1 5.2580	.0000
6.TH2	2.743	2.9656	.9248	.3577
7.STH1	2.657	.4448	5.9720	.0000
8.MTH1	10.061	2.9656	3.3927	.0010

ANALYSIS OF VARIANCE OF REL. EFFICIENCY OF B <sub>1</sub>					
SOURCE	DF	SUM SQRS	MEAN SQR	F-STAT	SIGNIF
REGRESSION	9	4680.1	520.01	9.7655	.0000
FIRST 5 VARS	5	360.51	72.101	1.3540	.2497
ERROR	86	4579.5	53.249		
TOTAL	95	9259.5			

  

VARIABLE	COEFF	STD ERROR	T-STAT	SIGNIF
CONSTANT	-24.740	7.3404	-3.3704	.0011
101.T1	-2.402	1.6654	-1.4422	.1529
102.T2	1.593	1.6654	.9568	.3413
103.T3	-2.900	1.6654	-1.7415	.0852
104.T4	1.143	1.6654	.6866	.4942
105.T5	.503	1.6654	.3022	.7632
106.T6	2.062	1.6654	1.2382	.2190
5.SMSZ	.268	.3724	-1 7.2033	.0000
6.TH2	10.931	4.9651	2.2015	.0304
7.STH1	-.765	.7448	-1.0277	.3070
8.MTH1	23.982	4.9651	4.8300	.0000

APPENDIX B.2  
(ANOCOVA)  
(combined samples)

ANALYSIS OF VARIANCE OF BIAS OF Bo						
SOURCE	DF	SUM SQRS	MEAN SQR	F-STAT	SIGNIF	
REGRESSION	9	.20377	.22641 -1	17.225	.0000	
FIRST 5 VARS	5	.37730 -2	.75461 -3	.57410	.7196	
ERROR	86	.11304	.13144 -2			
TOTAL	95	.31681				

  

VARIABLE	COEFF	STD ERROR	T-STAT	SIGNIF
CONSTANT	.10420	.36469 -1	2.8573	.0054
101.T1	.11240 -1	.82740 -2	1.3584	.1779
102.T2	-.27542 -2	.82740 -2	-.3329	.7400
103.T3	.23271 -2	.82740 -2	.2813	.7792
104.T4	-.89417 -2	.82740 -2	-1.0807	.2829
105.T5	.17833 -2	.82740 -2	.2155	.8299
106.T6	-.36542 -2	.82740 -2	-.4416	.6599
5.SMSZ	-.18224 -2	.18501 -3	-9.8501	.0000
6.TH2	.24625 -1	.24668 -1	.9982	.3210
7.STH1	-.27137 -1	.37003 -2	-7.3340	.0000
8.MTH1	.14528 -1	.24668 -1	.5889	.5575

ANALYSIS OF VARIANCE OF BIAS OF B1						
SOURCE	DF	SUM SQRS	MEAN SQR	F-STAT	SIGNIF	
REGRESSION	9	.71467 -2	.79408 -3	5.8421	.0000	
FIRST 5 VARS	5	.13378 -2	.26756 -3	1.9684	.0914	
ERROR	86	.11690 -1	.13592 -3			
TOTAL	95	.18836 -1				

  

VARIABLE	COEFF	STD ERROR	T-STAT	SIGNIF
CONSTANT	.31487 -1	.11728 -1	2.6849	.0087
101.T1	.72198 -2	.26607 -2	2.7135	.0080
102.T2	-.26365 -2	.26607 -2	-.99088	.3245
103.T3	.18635 -2	.26607 -2	.70039	.4856
104.T4	-.25740 -2	.26607 -2	-.96739	.3361
105.T5	-.73958 -4	.26607 -2	-.27796 -1	.9779
106.T6	-.37990 -2	.26607 -2	-1.4278	.1570
5.SMSZ	-.38547 -3	.59495 -4	-6.4790	.0000
6.TH2	-.50764 -2	.79327 -2	-.63993	.5239
7.STH1	-.26771 -3	.11899 -2	-.22498	.8225
8.MTH1	-.43403 -2	.79327 -2	-.54714	.5857

## APPENDIX B.3

(ANOCOVA)  
(Sample Size = 10)

ANALYSIS OF VARIANCE OF REL. EFFICIENCY OF B <sub>0</sub>					
SOURCE	DF	SUM SQRS	MEAN SQR	F-STAT	SIGNIF
REGRESSION	8	11.817	1.4771	13.462	.0000
FIRST 5 VARS	5	2.6112	.52224	4.7597	.0017
ERROR	39	4.2792	.10972		
TOTAL	47	16.096			

  

VARIABLE	COEFF	STD ERROR	T-STAT	SIGNIF
CONSTANT	.98063	.46573	2.1056	.0417
101.T1	-.82729	.10691	-.7738	.4437
102.T2	.25640	.10691	2.3983	.0214
103.T3	-.39948	.10691	-3.7367	.0006
104.T4	-.13598	.10691	-1.2719	.2109
105.T5	.10915	.10691	1.0209	.3136
106.T6	.25265	.10691	2.3632	.0232
6.TH2	-1.04600	.31874	-3.2816	.0022
7.STH1	.40844	.47811	-.85428	.0000
8.MTH1	.12431	.31874	.38999	.6987

ANALYSIS OF VARIANCE OF REL. EFFICIENCY OF B <sub>1</sub>					
SOURCE	DF	SUM SQRS	MEAN SQR	F-STAT	SIGNIF
REGRESSION	8	11.726	1.4658	6.4164	.0000
FIRST 5 VARS	5	8.8933	1.7787	7.7860	.0000
ERROR	39	8.9093	.22844		
TOTAL	47	20.635			

  

VARIABLE	COEFF	STD ERROR	T-STAT	SIGNIF
CONSTANT	.93514	.57201	.13916	.8900
101.T1	-.51146	.15426	-.33156	.7420
102.T2	.56810	.15426	3.6828	.0007
103.T3	-.61327	.15426	-3.9756	.0003
104.T4	-.41177	.15426	-2.6693	.0110
105.T5	.17854	.15426	.11574	.9085
106.T6	.49023	.15426	3.17790	.0029
6.TH2	-1.2837	.45991	-2.7913	.0081
7.STH1	.79104	.68987	-.11467	.2585
8.MTH1	.83486	.45991	1.8153	.0772

## APPENDIX B.4

(ANOCOVA)  
(Sample Size = 10)

ANALYSIS OF VARIANCE OF BIAS OF Bo						
SOURCE	DF	SUM SQRS	MEAN SQR	F-STAT	SIGNIF	
REGRESSION	8	.14085	.17607 -1	17.945	.0000	
FIRST 5 VARS	5	.78468 -2	.15694 -2	1.5995	.1831	
ERROR	39	.38265 -1	.98115 -3			
TOTAL	47	.17912				

  

VARIABLE	COEFF	STD ERROR	T-STAT	SIGNIF
CONSTANT	.55708 -1	.44041 -1	1.2649	.2134
101.T1	.24692 -1	.10110 -1	2.4424	.0192
102.T2	-.34208 -2	.10110 -1	-.33838	.7369
103.T3	.64667 -2	.10110 -1	.63966	.5261
104.T4	-.15633 -1	.10110 -1	-1.5464	.1301
105.T5	-.60583 -2	.10110 -1	-.59927	.5525
106.T6	-.60458 -2	.10110 -1	-.59803	.5533
6.TH2	.19778 -1	.30141 -1	.65618	.5156
7.STH1	-.52287 -1	.45211 -2	-11.565	.0000
8.MTH1	.35389 -1	.30141 -1	1.1741	.2475

ANALYSIS OF VARIANCE OF BIAS OF B1						
SOURCE	DF	SUM SQRS	MEAN SQR	F-STAT	SIGNIF	
REGRESSION	8	.19637 -2	.24547 -3	1.0281	.4319	
FIRST 5 VARS	5	.12370 -2	.24739 -3	1.0362	.4102	
ERROR	39	.93114 -2	.23875 -3			
TOTAL	47	.11275 -1				

  

VARIABLE	COEFF	STD ERROR	T-STAT	SIGNIF
CONSTANT	.27111 -2	.21725 -1	.1248	.9013
101.T1	.97229 -2	.49870 -2	1.9496	.0584
102.T2	-.40646 -2	.49870 -2	-.8150	.4200
103.T3	.51042 -3	.49870 -2	.1024	.9190
104.T4	-.20146 -2	.49870 -2	-.4040	.6884
105.T5	.18354 -2	.49870 -2	.3680	.7148
106.T6	-.59896 -2	.49870 -2	-1.2010	.2370
6.TH2	-.22931 -1	.14868 -1	-1.5422	.1311
7.STH1	.15896 -2	.22303 -2	.7127	.4803
8.MTH1	.59028 -2	.14868 -1	.3970	.6935

## APPENDIX B.5

(ANOCOVA)  
(Sample Size = 50)

ANALYSIS OF VARIANCE OF REL. EFFICIENCY OF B <sub>0</sub>					
SOURCE	DF	SUM SQRS	MEAN SQR	F-STAT	SIGNIF
REGRESSION	8	1692.1	211.51	9.3942	.0000
FIRST 5 VARS	5	59.265	11.853	.52644	.7548
ERROR	39	878.10	22.515		
TOTAL	47	2570.2			

  

VARIABLE	COEFF	STD ERROR	T-STAT	SIGNIF
CONSTANT	-15.253	6.6715	-2.2863	.0278
101.T1	-1.379	1.5315	-.9007	.3733
102.T2	.621	1.5315	.4054	.6874
103.T3	-1.506	1.5315	-.9836	.3314
104.T4	.790	1.5315	.5155	.6091
105.T5	-.172	1.5315	-.1123	.9911
106.T6	1.492	1.5315	.9745	.3358
6.TH2	6.531	4.5659	1.4304	.1606
7.STH1	4.905	.6849	7.1618	.0000
8.MTH1	19.999	4.5659	4.3800	.0001

ANALYSIS OF VARIANCE OF REL. EFFICIENCY OF B <sub>1</sub>					
SOURCE	DF	SUM SQRS	MEAN SQR	F-STAT	SIGNIF
REGRESSION	8	3724.3	465.53	6.5982	.0000
FIRST 5 VARS	5	622.52	124.50	1.7646	.1430
ERROR	39	2751.6	70.555		
TOTAL	47	6475.9			

  

VARIABLE	COEFF	STD ERROR	T-STAT	SIGNIF
CONSTANT	-33.479	11.810	-2.8348	.0072
101.T1	-4.752	2.710	-1.7530	.0875
102.T2	2.619	2.711	.9660	.3400
103.T3	-5.187	2.711	-1.9134	.0631
104.T4	2.699	2.711	.9953	.3257
105.T5	.989	2.711	.3647	.7173
106.T6	3.634	2.711	1.3404	.1879
6.TH2	23.146	8.083	2.8636	.0067
7.STH1	-1.610	1.212	-1.3278	.1920
8.MTH1	47.128	8.083	5.8309	.0000



## APPENDIX B.6

(ANOCOVA)  
(Sample Size = 50)

ANALYSIS OF VARIANCE OF BIAS OF B <sub>0</sub>						
SOURCE	DF	SUM SQRS	MEAN SQR	F-STAT	SIGNIF	
REGRESSION	8	.20657 -2	.25821 -3	1.2442	.3003	
FIRST 5 VARS	5	.89468 -3	.17894 -3	.86219	.5151	
ERROR	39	.80940 -2	.20754 -3			
TOTAL	47	.10160 -1				

  

VARIABLE	COEFF	STD ERROR	T-STAT	SIGNIF	
CONSTANT	.43353 -1	.20255 -1	2.1403	.0386	
101.T1	-.22125 -2	.46496 -2	-.4759	.6368	
102.T2	-.20875 -2	.46496 -2	-.4490	.6559	
103.T3	-.18125 -2	.46496 -2	-.3898	.6988	
104.T4	-.22500 -2	.46496 -2	-.4839	.6312	
105.T5	.96250 -2	.46496 -2	2.0701	.0451	
106.T6	-.12625 -2	.46496 -2	-.2715	.7874	
6.TH2	.29472 -1	.13862 -1	2.1261	.0399	
7.STH1	-.19875 -2	.20794 -2	-.9558	.3450	
8.MTH1	-.63333 -2	.13862 -1	-.4569	.6503	

ANALYSIS OF VARIANCE OF BIAS OF B <sub>1</sub>						
SOURCE	DF	SUM SQRS	MEAN SQR	F-STAT	SIGNIF	
REGRESSION	8	.10259 -2	.12824 -3	6.0295	.0000	
FIRST 5 VARS	5	.40314 -3	.80627 -4	3.7909	.0068	
ERROR	39	.82947 -3	.21269 -4			
TOTAL	47	.18554 -2				

  

VARIABLE	COEFF	STD ERROR	T-STAT	SIGNIF	
CONSTANT	.37135 -1	.64842 -2	5.7270	.0000	
101.T1	.47167 -2	.14884 -2	3.1689	.0030	
102.T2	-.12083 -2	.14884 -2	-.8118	.4218	
103.T3	.32167 -2	.14884 -2	2.1611	.0369	
104.T4	-.31333 -2	.14884 -2	-2.1051	.0418	
105.T5	-.19833 -2	.14884 -2	-1.3325	.1904	
106.T6	-.16083 -2	.14884 -2	-1.0805	.2865	
6.TH2	.12778 -1	.44377 -2	2.8794	.0064	
7.STH1	-.21250 -2	.66565 -3	-3.1924	.0028	
8.MTH1	-.14583 -1	.44377 -2	-3.2862	.0022	

## APPENDIX C.1

The ratios of the variance of AITKEN to the variances of APX, MAPX and FMAPX. For a given sample size, the first row is for APX, the second for MAPX, and the third for FMAPX.

n/c	.10	.30	.50	.70	.90	.99
3	0.712	0.812	0.921	0.988	0.944	0.895
	1.000	0.998	0.981	0.926	0.834	0.789
	1.000	1.000	0.999	0.993	0.971	0.955
5	0.833	0.902	0.971	0.970	0.786	0.678
	1.000	0.998	0.981	0.899	0.715	0.626
	1.000	1.000	0.999	0.985	0.921	0.872
10	0.918	0.956	0.989	0.959	0.614	0.423
	1.000	0.999	0.989	0.914	0.583	0.411
	1.000	1.000	0.999	0.983	0.817	0.663
15	0.946	0.971	0.993	0.968	0.570	0.313
	1.000	0.999	0.993	0.938	0.550	0.308
	1.000	1.000	1.000	0.987	0.776	0.527
20	0.960	0.979	0.995	0.976	0.569	0.251
	1.000	1.000	0.994	0.953	0.553	0.249
	1.000	1.000	1.000	0.990	0.769	0.440
30	0.973	0.986	0.997	0.984	0.612	0.185
	1.000	1.000	0.996	0.969	0.600	0.184
	1.000	1.000	1.000	0.993	0.794	0.335
50	0.984	0.992	0.998	0.990	0.714	0.129
	1.000	1.000	0.998	0.981	0.704	0.128
	1.000	1.000	1.000	0.996	0.857	0.240
100	0.992	0.996	0.999	0.995	0.841	0.089
	1.000	1.000	0.999	0.991	0.835	0.089
	1.000	1.000	1.000	0.998	0.927	0.168

## APPENDIX C.2

Ratios of the variance of FMAPX to the variances of APX and MAPX. The first row is for APX, and the second for MAPX.

n/c	.10	.30	.50	.70	.90	.99
3	0.712	0.812	0.921	0.996	0.973	0.938
	1.000	0.998	0.981	0.932	0.859	0.826
5	0.833	0.902	0.972	0.985	0.853	0.777
	1.000	0.998	0.982	0.913	0.776	0.718
10	0.918	0.956	0.990	0.976	0.752	0.638
	1.000	0.999	0.990	0.930	0.715	0.619
15	0.946	0.971	0.994	0.981	0.735	0.593
	1.000	0.999	0.993	0.950	0.709	0.584
20	0.960	0.979	0.995	0.985	0.740	0.572
	1.000	1.000	0.995	0.962	0.720	0.566
30	0.973	0.986	0.997	0.990	0.771	0.551
	1.000	1.000	0.997	0.975	0.755	0.548
50	0.984	0.992	0.998	0.994	0.831	0.536
	1.000	1.000	0.998	0.985	0.821	0.535
100	0.992	0.996	0.999	0.997	0.908	0.527
	1.000	1.000	0.999	0.993	0.901	0.526









































## APPENDIX H.1

```

C**** This FORTRAN program is to be use to obtain the transformed
C**** variables, to perform the EGLS and can also be use as a sub
C**** -routine for ML, when the Regression Model has MA(2) Errors.
      REAL*8 TH(2),YT(10),XTO(10),XT(10),S(10,10)
      REAL*8 YST(10),XST1(10),XST2(10),A11,A12,A13
C**** TH's are the two MA prameters reading from MTS file.
C**** YT,XTO,XT are the dependent variable and two independent
C**** variable, one of them may be column of 1's for intercept.
C**** S is the square root matrix of covariance matrix V.
C**** YST,XST1,XST2 are the transformed variables of YT,XTO,XT.
      INTEGER N,M,I,L,J,K
      N=10
C**** This N is for sample size,to be changed,when necessary.
      READ (3,60) TH
      READ (5,60) YT
      READ (2,58) XTO
      READ (4,60) XT
      A11=1.0+TH(1)*TH(1)+TH(2)*TH(2)
      A12=-TH(1)*(1.0-TH(2))
      A13=-TH(2)
      S(1,1)=A11** .5
      YST(1)=YT(1)/S(1,1)
      XST1(1)=XTO(1)/S(1,1)
      XST2(1)=XT(1)/S(1,1)
      S(1,2)=A12/S(1,1)
      S(2,2)=(A11-S(1,2)*S(1,2))** .5
      YST(2)=(YT(2)-S(1,2)*YST(1))/S(2,2)
      XST1(2)=(XTO(2)-S(1,2)*XST1(1))/S(2,2)
      XST2(2)=(XT(2)-S(1,2)*XST2(1))/S(2,2)
      K=1
      L=2
C**** The values given to K and L are fixed, not to be changed.
      DO 10 J=3,N
      S(K,J)=A13/S(K,K)
      S(L,J)=(A12-S(L-1,J-1)*S(L-1,J))/S(L,L)
      S(J,J)=(A11-S(K,J)*S(K,J)-S(L,J)*S(L,J))** .5
      YST(J)=(YT(J)-S(L,J)*YST(J-1)-S(K,J)*YST(J-2))/S(J,J)
      XST1(J)=(XTO(J)-S(L,J)*XST1(J-1)-S(K,J)*XST1(J-2))/S(J,J)
      XST2(J)=(XT(J)-S(L,J)*XST2(J-1)-S(K,J)*XST2(J-2))/S(J,J)
      K=K+1
      L=L+1
10    CONTINUE
      WRITE (6,10025) YST
      WRITE (7,10025) XST1
      WRITE (8,10025) XST2
C**** These transformed variables may be used for EGLS.
10025 FORMAT(F16.8)
58    FORMAT(F3.0)
60    FORMAT(F9.5)
      STOP
      END

```

## APPENDIX H.2

```

C*** This program will calculate the exact determinant of
C*** covariance matrix  $\Omega$  for MA(2) error process using  $\Theta$  matrix.

REAL*8 A1(50,3),SA1,SA2,SA3,DET,THETA(2),Z1(50),Z2(50)
READ (5,100) THETA

C*** Where THETA is for moving average parameters.

MM=50
C*** Where MM is the sample size, change when necessary along
C*** with the dimensions specified above.
Z1(1)=-1*THETA(1)
Z2(1)=-1*THETA(2)
A1(1,1)=Z1(1)*Z1(1)
A1(1,2)=Z2(1)*Z2(1)
A1(1,3)=Z1(1)*Z2(1)
SA1=A1(1,1)
SA2=A1(1,2)
SA3=A1(1,3)
DO 115 I=2,MM
Z1(I)=THETA(1)*Z1(I-1)+Z2(I-1)
Z2(I)=THETA(2)*Z1(I-1)
A1(I,1)=Z1(I)*Z1(I)
A1(I,2)=Z2(I)*Z2(I)
A1(I,3)=Z1(I)*Z2(I)
SA1=A1(I,1)+SA1
SA2=A1(I,2)+SA2
115 SA3=A1(I,3)+SA3

DET=(1+SA1)*(1+SA2)-(SA3)*(SA3)

WRITE(6,121) DET
121 FORMAT(F10.3)
100 FORMAT(F12.8)
STOP
END

```

## B I B L I O G R A P H Y

- Aigner, D.j. (1971). A compendium of estimation of the autoregressive moving average model from time series data, *International Economic Review*, 12, 348-371.
- Ali, M.M. and Thalheimer, R. (1983). Stationary tests in time series model, *Journal of Forecasting*, vol.2, 249-257.
- Ansley, C.F. (1979). An algorithm for the exact likelihood of a mixed autoregressive-moving average process, *Biometrika*, 66,1, 59-65.
- Balestra, P. (1980). A note on the exact transformation associated with the first-order moving average process, *Journal of Econometrics*, 14, 381-394.
- Beach, C.M. and MacKinnon, J.G.(1978). A maximum likelihood procedure for regression with autocorrelated errors, *Econometrica*, 46, 51-58.
- Beguín, J.-M., Gourieroux, C. and Monfort, A. (1980). Identification of a mixed autoregressive-moving average process: The corner method, *Time Series*. Amsterdam: North-Holland publishing company, 423-436.
- BMDQ2T, ( ). *BMDQ2T TSPACK*, Health sciences computing facility, University of California, LA, CA 90024.
- Box, G.E.P. and Jenkins, G.M. (1970). *Time Series Analysis Forecasting and Control*, San Francisco: Holden Day.
- Box, G.E.P. and Jenkins, G.M. (1976). *Time Series Analysis Forecasting and Control*, San Francisco: Holden Day.
- Box, G.E.P. and Pierce, D.A.(1970). Distribution of residual autocorrelation in autoregressive-integrated moving average time series models, *Journal of American Statistical Association*, 65, 1509-1526.
- Choudhury, A.H. and Chaudhury, M.M. (1984). A note on approximate estimator for the first-order moving average process, *Journal of Econometrics*, (submitted).
- Durbin, J. (1970). Testing for serial correlation in least squares regression when some of the regressors are lagged dependent variables. *Econometrica*, 38, 410.
- Durbin, J. and Watson, G.S.(1950). Testing for serial correlation in Least Squares Regression I, *Biometrika*, 37, 409-428.



- Durbin, J. and Watson, G.S. (1951). Testing for serial correlation in Least Squares Regression II, *Biometrika*, 38, 159-178.
- Fuller, W.A. (1976). *Introduction to Statistical Time Series*, New York: Wiley.
- Goldberger, A.S. (1963). *Econometric Theory*, New York: Wiley.
- Granger, C.W.J. (1980). *Forecasting in business and economics*, New York: Academic Press.
- Harvey, A.C. (1981). *The Econometric Analysis of Time Series*, London: Philip Allan.
- Johnston, J. (1972). *Econometric Methods*. New York: Wiley, MacGraw-Hill.
- Judge, G.G., Griffiths, W.E., Hill, R.C. and Lee, T.-C. (1980). *The Theory and Practice of Econometrics*, New York: Wiley.
- Kendall, Sir.M., Stuart, A. and Ord, J.K. (1983), vol.3., *The Advanced Theory of Statistics*, London: Griffin.
- Ljung, G.M. and Box, G.E.P. (1978). On a measure of lack of fit in time series model, *Biometrika*, 65, 297-303.
- MacDonald, G.M. and MacKinnon, J.G. (1985). Convenient methods for estimation of linear regression models with MA(1) errors, *Canadian Journal of Economics*, XVIII, no.1, 106.
- McLeod, A.I. (1977). Improved Box-Jenkins estimators, *Biometrika*, 64, 531-534.
- MINITAB, (1982). *MINITAB Reference Manual*, Statistics department, 215 pond laboratory, The Pennsylvania State University, pa. 16802.
- NAG, (1982). *NAG Library Manual*, Numerical Algorithms Group, Mayfield House, 256 Banbury Road, Oxford, Ox2 7DE, UK.
- Nicholls, D.F., Pagan, A.R. and Terrel, R.D. (1975), The estimation and use of models with moving average disturbance terms: A Survey, *International Economic Review*, 16, 113-134.
- Osborn, D.R. (1976). Maximum Likelihood estimation of moving average process, *Annals of Economic and Social measurement*, 3, 75-87.
- Pagan, A. (1973). Efficient estimation of models with composite disturbance terms, *Journal of Econometrics*, 1, 329-340.

- Pagan, A. (1974). A Generalized approach to the treatment of autocorrelation, *Australian Economic papers*, 13, 267-280.
- Pagan, A.R. and Nicholls, D.F. (1976). Exact Maximum Likelihood estimation of regression models with finite order moving average errors, *Review of Economic Studies*, 43, 383-388.
- Park, C.Y. and Heikes, R.G. (1983). A note on Balestra's (1980) approximate estimator for the first-order moving average process, *Journal of Econometrics*, 21, 387-388.
- Pesaran, M.H. (1973). Exact maximum likelihood estimation of a regression equation with a first-order moving average errors, *Review of Economic Studies*, 40, 529-536.
- Phadke, M.S. and Kedem, G. (1978). Computation of the exact likelihood function of multivariate moving average models, *Biometrika*, 65, 3, 511-519.
- Phillips, A.W. (1966). The estimation of systems of difference equations with moving average disturbances, *Econometric Society Meeting, San Francisco*, reprinted in A.E. Bergstrom, et al., eds., *Stability and Inflation*, New York: Wiley.
- Pierce, D.A. (1971). Distribution of residual autocorrelations in the regression model with autoregressive-moving average errors, *Journal of the Royal Statistical Society, series B*, 33, 140-146.
- Pindyck, R.S. and Rubinfeld, D.L. (1981). *Econometric models and economic forecasts*. New York: McGraw-Hill.
- Pollock, D.S.G. (1979). *The algebra of econometrics*, New York: Wiley.
- Pukkila, T.M. (1982). On the identification of ARMA(p,q) models, *Time Series Analysis: Theory and practice I*, Amsterdam: North-Holland publishing company, 81-103.
- Trivedi, P.L. (1970). Inventory Behavior in U.K. Manufacturing, 1956-67, *Review of Economic Studies*, 37, 517-527.
- Wallis, K.F. (1972). Testing for fourth-order autocorrelation in quarterly regression equations, *Econometrica*, 40, 617-636.
- Watson, G.S. (1955). Serial correlation in regression analysis I, *Biometrika*, 42, 327-341.
- Watson, G.S. and Hannan, E.J. (1956). Serial correlation in regression analysis II, *Biometrika*, 43, 436-445.
- Wise, J. (1955). Autocorrelation function and the spectral density function, *Biometrika*, 42, 151-159.