This is to certify that the

dissertation entitled

MINIMUM DISTANCE ESTIMATION FOR ARMA AND GARCH PROCESSES

presented by

Huimin Chung

has been accepted towards fulfillment
of the requirements for

Ph. D. degree in Economics

R Baillie
Major professor

Date 12. 4. 1997

# MINIMUM DISTANCE ESTIMATION FOR ARMA AND GARCH PROCESSES

By

Huimin Chung

**A DISSERTATION**

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

**DOCTOR OF PHILOSOPHY**

Department of Economics

1997

# ABSTRACT

Minimum Distance Estimation for ARMA and GARCH Processes

By

Huimin Chung

This dissertation considers the estimation of the parameters of ARMA and GARCH processes by the Minimum Distance Estimator (MDE). The estimator is obtained by minimizing a quadratic distance function between the sample autocorrelation and theoretical autocorrelation functions and has the advantage of imposing very little in terms of distributional assumptions on the innovation process. The asymptotic properties of the MDE for ARMA processes are discussed. The exact asymptotic efficiency of using a block of sample autocorrelations is derived for some low order MA and ARMA models. The MDE is surprisingly efficient for some parts of the parameter space. We also investigate the properties of the MDE when used to estimate linear GARCH models using autocorrelations of the squared process. Monte Carlo results show that the MDE performs better than Quasi-MLE with certain conditional densities which exhibit extreme departures from conditional Gaussianity. An application of the MDE to estimate a GARCH(1,1) model from high frequency exchange rate data is provided.

## ACKNOWLEDGEMENT

The completion of this dissertation would never have been possible without the help and support of numerous individuals. My greatest appreciation must go to my dissertation committee chair, Professor Richard Baillie, who introduced me to this dissertation topic, sacrificed countless hours in reading and correcting drafts, and carefully guided and inspired me through each step of my research. I have benefited greatly from his experience and wisdom. I will remain indebted to him throughout the remainder of my career.

In addition I would like to express my appreciation to Professors Robert de Jong, Peter Schmidt, and Jeff Wooldridge, whose guidance, encouragement and help have been invaluable in completing this dissertation. Many parts of this dissertation were especially strengthened by their suggestions and contributions. I also thank Professor Ching-Fan Chung for his help and encouragement in my early stage of writing the dissertation.

I would never have completed my journey through graduate school without the support of my family. Particularly crucial to this accomplishment is the unceasing support of my wife, Wen-Hua. Words would not be enough for her love and understanding. Thanks also go to my father and mother. Their encouragement and support throughout the peaks and troughs of these past few years is invaluable. I owe a special thank, too, to my younger brother Ching-Gee.

I am also grateful to many colleagues and friends at Michigan State University. I especially thank Jen-Je Su, Hailong Qian, Wen-Jen Tsay, and Yang-Seon Kim who offered many valuable comments and discussions of this dissertation.

Finally, this dissertation is dedicated to my lovely daughter, Candice, who has been giving me so much fun since she was born.

# TABLE OF CONTENTS

# LIST of TABLES

# CHAPTER 1
# INTRODUCTION

This dissertation considers the estimation of the parameters of some linear time series processes and GARCH volatility processes by the use of Minimum Distance Estimator (MDE). The estimator is obtained by minimizing a quadratic distance function between the sample autocorrelation and theoretical autocorrelation functions. The MDE is very similar to Hansen's (1982) GMM estimator. In the context of this dissertation the moment conditions correspond to sample autocorrelations. The MDE method has been previously applied to the problem of estimating the Hurst coefficient, or order of fractional integration in ARFIMA models by Tieslau, Schmidt and Baillie (1996) and Chung and Schmidt (1996). The MDE has the advantage of imposing very little in terms of distributional assumption on the innovation process. Hence in cases of extreme non-normality the MDE, which is straightforward to compute, may have some distinct advantages.

The plan of this dissertation is as follows. Chapter 2 introduces the MDE and derives the asymptotic properties of the MDE for some specific ARMA processes. We derive the exact asymptotic efficiency of the MDE relative to MLE for some specific ARMA processes. The application of the MDE to AR(p) process is first discussed. The asymptotic variance of the MDE using the first p autocorrelation is compared with that of MLE under NID innovations. The asymptotic variances of the MDEs using different sets of autocorrelations are calculated for some MA(q) and ARMA(p,q) processes for a variety of parameter values. Also the asymptotic variance of the MDE are compared with that of MLE under NID innovations.

Chapter 3 investigates the asymptotic properties of the MDE for the MA(1)-Seasonal $MA(1)_s$ model, which is applied by Box and Jenkins to model the airline passenger data and is also called "airline model", because of its widespread use in time series analysis. This airline model has been applied to model many economic time series. The asymptotic variance of the MDE are discussed for the cases that s equals to 4 and 12.

Chapter 4 deals with using the MDE based on the sample autocorrelations of the squared process to estimate the parameters of the GARCH model. The GARCH models and their extensions have been widely applied in characterizing the time dependent heteroskedasticity present in many economic and financial economics series. If the observed time series, $y_t$ follows a martingale with linear GARCH(1,1) volatility process, then $y_t^2$ has an ARMA(1,1) representation. However, despite the innovation process being serially uncorrelated, it is not independent over time. When the innovations are not i.i.d., it may not be valid to use the Bartlett's formula to calculate the weighting matrix. Nevertheless, the robust method of Domowitz and White (1982) and White (1984) can be applied to obtain robust covariance matrix estimator of the sample autocorrelation. We thus propose a Newey and West (1987) type covariance matrix estimator of the sample autocorrelations of the squared process, which is generated as a martingale GARCH(1,1). A Monte Carlo experiment is carried out to compare the performance of the MDE using the Bartlett's formula to construct the weighting matrix and the MDE using the Newey-West method to calculate the optimal weighting matrix for estimating the parameters of the GARCH(1,1) model. Many studies have found evidence that the conditional density of GARCH models for many speculative asset return series, such as exchange rates and stock prices, are non Gaussian. The Quasi-MLE (QMLE) method is usually invoked to estimate such GARCH models. A simulation experiment is performed at the end of this chapter to compare the small sample properties of the MDE and the QMLE of GARCH(1,1) models.

# CHAPTER 2

## Minimum Distance Estimation for ARMA processes

### 1. Introduction

There is a long tradition in univariate time series analysis of approximating MA and ARMA processes as infinite order AR processes, e.g. Durbin (1959, 1960). The motivation for this approach was typically in terms of the ease of estimation of AR processes by least squares and hence the avoidance of maximizing a likelihood which was highly non linear in the parameters. Before the advent of modern computer power such methods were the only realistic way of estimating ARMA models. There is still some interest in the estimation of ARMA models by approximations in terms of high order AR processes and Galbraith and Zinde Walsh (1994) consider the distance in terms of Hilbert space between the true ARMA data generating process and the AR(p) approximation. Also see Koreisha and Pukkila (1990), Saikkonen (1986), and Galbrath and Zinde Walsh (1997).

This chapter considers the estimation of the parameters of some univariate time series process by the Minimum Distance Estimator (MDE), which is closely related to Hansen's GMM estimator. In the context of this chapter the moment conditions correspond to sample autocorrelations. The MDE method has been previously applied to the problem of estimating the Hurst coefficient, or order of fractional integration in ARFIMA models by Tieslau, Schmidt and Baillie (1996) and Chung and Schmidt (1996). The MDE has the advantage of imposing very little in terms of distributional assumptions on the innovation process. Hence in cases of extreme non-normality, the MDE, which is straightforward to compute, may have some distinct advantages.

The remainder of this chapter is organized as follows. Section 2 discusses the general set up of the MDE. Section 3 then derives the exact asymptotic efficiency of the MDE relative to MLE for some specific non-seasonal stationary AR processes. For the AR(p) process the MDE using the first p autocorrelation can be asymptotically as efficient as MLE under normality. Section 4 then derive the gain in asymptotic efficiency from using higher order sample autocovariances when estimating the moving average process. Section 5 deals with ARMA (1,1) process, while section 6 considers the same approach for the higher order ARMA(p,q) process. Section 7 provides a brief conclusion.

## 2. The MDE

The minimum distance estimation method is based on the following asymptotic results for the sample autocorrelations of a stationary process. The sample autocorrelation at lag $k$ is given by

$$\hat{\rho}_k = \sum_{t=k+1}^{T} (y_t - \bar{y})(y_{t-k} - \bar{y}) / \sum_{t=1}^{T} (y_t - \bar{y})^2,$$

where $\bar{y}$ is the sample mean. Bartlett (1946) and Brockwell and Davis (1991, p.221, Theorem 7.2.1) show that if $y_t$ is a stationary process, $y_t = \sum_{j=-\infty}^{\infty} \psi_j \varepsilon_{t-j}$, $\varepsilon_t \sim$ i.i.d.$(0, \sigma^2)$, where $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$ and $E\varepsilon_t^4 < \infty$ , then

$$\sqrt{T}(\hat{\rho} - \rho) \rightarrow N(0, C), \tag{1}$$

where $\hat{\rho}' = [\hat{\rho}_1, \hat{\rho}_2, \cdots, \hat{\rho}_g]$, $\rho' = [\rho_1, \rho_2, \cdots, \rho_g]$ and $C$ is a $g \times g$ matrix with $(i, j)$th element given by

$$c_{ij} = \sum_{k=1}^{\infty}(\rho_{k+i} + \rho_{k-i} - 2\rho_i\rho_k)(\rho_{k+j} + \rho_{k-j} - 2\rho_j\rho_k).$$

Alternatively the assumption of finite fourth moment of $\epsilon_t$ can be replaced with the condition that $\sum_{j=-\infty}^{\infty}\psi_j^2|j| < \infty$.

The MDE is an alternative to the maximum likelihood estimator and has the advantage of not invoking strong distributional assumption. On defining $\lambda$ as the vector of parameter to be estimated, so that in the case of the ARMA(1,1) model $\lambda' = [\phi\ \theta]$. The MDE is obtained by minimizing the following criteria function

$$\text{Min } S = (\hat{\rho} - \rho(\lambda))'W(\hat{\rho} - \rho(\lambda)),$$

where W is a $g \times g$ symmetric, positive-definite matrix. Let $\hat{\lambda}$ denotes the value of $\lambda$ which solves the above minimization problem. Therefore, we have

$$\left.\frac{\partial S}{\partial \lambda}\right|_{\lambda=\hat{\lambda}} = 2\left[\left.\frac{\partial \rho(\lambda)}{\partial \lambda'}\right|_{\lambda=\hat{\lambda}}\right]' W(\hat{\rho} - \rho(\hat{\lambda})) = 0.$$

Let $\lambda_0$ denote the true value of $\lambda$. Using the mean value Theorem, we have $\lambda_*$ such that

$$\hat{\rho} - \rho(\hat{\lambda}) = \hat{\rho} - \rho(\lambda_0) + \left.\frac{\partial(\hat{\rho} - \rho(\lambda))}{\partial \lambda'}\right|_{\lambda=\lambda_*}(\hat{\lambda} - \lambda_0). \qquad (2)$$

For convenience we have the following definition of notation :

$$D_{\hat{\lambda}} = \left.\frac{\partial \rho(\lambda)}{\partial \lambda'}\right|_{\lambda=\hat{\lambda}}.$$

Similarly, $D_{\lambda_*}$ and $D$ denote the partial derivative of $\rho(\lambda)$ with respect to $\lambda'$ evaluated at $\lambda_*$ and $\lambda_0$, respectively.

Premultipling both sides of equation (2) with $D'_{\hat{\lambda}}W$, we have

$$D'_{\hat{\lambda}}\, W(\hat{\rho} - \rho(\hat{\lambda})) = D'_{\hat{\lambda}}\, W\ (\hat{\rho} - \rho(\lambda_0))\ \ -\ \ D'_{\hat{\lambda}}\, W\ D_{\lambda_*}(\hat{\lambda} - \lambda_0).$$

The left hand side of the above equation is the first order condition and hence, equals zero. Rearranging the above equation yields

$$\hat{\lambda} - \lambda_0 = \left[D'_{\hat{\lambda}}\ W\ D_{\lambda_*}\right]^{-1} D'_{\hat{\lambda}}\, W(\hat{\rho} - \rho(\lambda_0)).$$

Notice that $plim(\hat{\rho} - \rho(\lambda_0)) = 0$. Given that $plim\ ([D'_{\hat{\lambda}}\ W\ D_{\lambda_*}]^{-1}D'_{\hat{\lambda}}\ W)$ is finite, it follows directly that $\hat{\lambda}$ is consistent. We then have $plimD_{\lambda_*} = plimD_{\hat{\lambda}} = D$. Let $\xrightarrow{p}$ denote converge in probability. Then, it is easily verified that

$$\sqrt{T}(\hat{\lambda} - \lambda_0) \xrightarrow{p} [D'WD]^{-1}D'W\sqrt{T}\ (\ \hat{\rho} - \rho(\lambda_0)\ )$$

Using the results derived above, we obtain

$$\sqrt{T}(\hat{\lambda} - \lambda_0) \rightarrow N\ (\ 0,\ (D'WD)^{-1}D'WCW'D(D'WD)^{-1}\ )$$

On using the standard optimal weighting matrix of $W = C^{-1}$, in which case the limiting distribution of $\hat{\lambda}$ is

$$\sqrt{T}(\hat{\lambda} - \lambda_0) \rightarrow N\ (\ 0,\ (D'C^{-1}D)^{-1}\ ). \tag{3}$$

A consistent estimator of $C$ is $\hat{C}$, the $g \times g$ matrix of the sample counterpart of $C$, with the $(i, j)$th element being estimated by

$$\hat{c}_{ij} = \sum_{k=1}^{\infty}(\hat{\rho}_{k+i} + \hat{\rho}_{k-i} - 2\hat{\rho}_i\hat{\rho}_k)(\hat{\rho}_{k+j} + \hat{\rho}_{k-j} - 2\hat{\rho}_j\hat{\rho}_k). \tag{4}$$

In practical applications the MDE is obtained by solving the following minimization problem:

$$\text{Min } S = (\hat{\rho} - \rho(\lambda))'\hat{C}^{-1}(\hat{\rho} - \rho(\lambda)).$$

## 3. The MDE Applied to AR(p) Processes

The asymptotic efficiency of the MDE can be investigated for different sets of moment conditions. For the AR(1) process the MDE using the first autocorrelation is asymptotically as efficient as MLE under normality. This is easily seen from investigation of the redundancy conditions associated with the use of the first $m_1$ moments, as opposed to the use of the first $m_1 + m_2$ moments.

Let $Eg(y_t, \lambda) = 0$ be the moments condition of the GMM estimator and $\bar{g}(\lambda)$ denote the sample average of the moment condition, i.e., $\bar{g}(\lambda) = \frac{1}{T}\sum_{t=1}^{T} g(y_t, \lambda)$. The GMM and MDE estimator is obtained by minimizing a quadratic function, $\bar{g}(\lambda)'W\bar{g}(\lambda)$. Let $C$ denote the asymptotic covariance of $\sqrt{T}\bar{g}(\lambda)$. The optimum weighting matrix for the GMM estimator based on $\bar{g}(\lambda)$ is $W = C^{-1}$. For the MDE in this chapter $\bar{g}(\lambda)$ corresponds to $\hat{\rho} - \rho(\lambda)$.

To investigate whether a subset of the moment conditions is redundant, $\bar{g}(\lambda)$ is partitioned into two subsets, .i.e., $\bar{g}(\lambda) = [\bar{g}_1(\lambda) \quad \bar{g}_2(\lambda)]'$, and C is also can be partitioned as

$$C = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix},$$

where $C_{11}$ and $C_{22}$ are the asymptotic covariance of $\sqrt{T}\bar{g}_1(\lambda)$ and $\sqrt{T}\bar{g}_2(\lambda)$, respectively. The asymptotic variance of the MDE estimator using first $m_1$ moment conditions, $g_1(\lambda)$, is

$$(D_1' C_{11}^{-1} D_1)^{-1},$$

where $D_1 = E[\frac{\partial g_1(y_t,\lambda)}{\partial \lambda'}]$. Bruesch, Qian, Schmidt and Wyhowski (1997) show that if

$$D_2 - C_{12}' C_{11}^{-1} D_1 = 0, \tag{5}$$

with $D_2 = E[\frac{\partial g_2(y_t,\lambda)}{\partial \lambda'}]$, then the extra $m_2$ moment conditions are redundant and MDE based on the first $m_1$ moments is as efficient as MDE based on first $m_1 + m_2$ moments.

Based on the above result, it can be shown that for the AR(1) process, MDE using first 2 autocorrelations has no improvement in efficiency over MDE using just the first autocorrelation. The AR(1) process is

$$y_t = \phi y_{t-1} + \epsilon_t, \tag{6}$$

where $\epsilon_t$ is iid $(0, \sigma^2)$. The autocorrelation function of the AR(1) process is $\rho_h = \phi^h$ for $h = 1, 2, \cdots$. The asymptotic variance of sample autocorrelations of the AR(1) process is

$$C = (1-\phi^2)\begin{bmatrix} 1 & 2\phi & 3\phi^2 & 4\phi^3 & \cdots \\ 2\phi & 1+3\phi^2 & 2\phi+4\phi^3 & 3\phi^2+5\phi^4 & \cdots \\ 3\phi^2 & 2\phi+4\phi^3 & 1+3\phi^2+5\phi^4 & 2\phi+4\phi^3+6\phi^5 & \cdots \\ 4\phi^3 & 3\phi^2+5\phi^4 & 2\phi+4\phi^3+6\phi^5 & 1+3\phi^2+5\phi^4+7\phi^6 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}, \tag{7}$$

with the $P, Q$th element of C can be written as

$$C_{PQ} = (1-\phi^2)\sum_{j=1}^{Q}(P-Q-1+2j)\phi^{[P-Q+2(j-1)]}, \text{ for } P \geq Q.$$

Also,

$$D = -[1 \;\; 2\phi \;\; 3\phi^2 \;\; \cdots]'. \tag{8}$$

It is straightforward to show that the asymptotic variance of MDE using the first autocorrelation is $1 - \phi^2$, which is equal to the asymptotic variance of MLE under normality. A direct calculation shows that the asymptotic variance of MDE using first two autocorrelations is also equal to $1 - \phi^2$. Furthermore, it can be verified that the redundancy condition of equation (5) holds for $g \geq 2$. When $g \geq 2$, we have $D_2 = -[2\phi \;\; 3\phi^2 \;\; \cdots \;\; g\phi^{g-1}]', C_{12} = (1-\phi^2)[2\phi \;\; 3\phi^2 \;\; \cdots \;\; g\phi^{g-1}], C_{11} = (1-\phi^2)$, and $D_1 = -1$ . Hence $D_2 - C_{12}'C_{11}^{-1}D_1 = 0$.

For AR(1) process the MDE using first $g$ autocorrelations for $g \geq 2$ does not have any improvement in efficiency over the MDE using just the first autocorrelation. Alternatively, we can estimate $\phi$ from just the $h$th autocorrelation, $\hat{\rho}_h$, for $h = 3, 5, 7, \cdots$. Notice that we are not able to identify $\phi$ if only $\rho_2$ is used in the MDE, because the sign of $\phi$ can not be determined. Similarly, we can not identify the AR parameter if only $\rho_{2j}$ is used for $j = 2, 3, \cdots$. Consider the estimation of $\phi$ from $\hat{\rho}_3$. On using the results given in equations (7) and (8), we have

$$T^{1/2}(\hat{\rho}_3^{1/3} - \phi) \rightarrow N\left[\, 0, \;\frac{(1-\phi^2)(1+3\phi^2+5\phi^4)}{9\phi^4}\,\right].$$

Therefore, the estimator of $\phi$ based on $\hat{\rho}_3$ only is not asymptotically efficient since $\frac{(1+3\phi^2+5\phi^4)}{9\phi^4} > 1$ for $|\phi| < 1$. In general, for $h = 3, 5, 7\cdots$, we have the following result of the asymptotic variance of the MDE using just $\hat{\rho}_h$ $(V_{MDE}^h)$.

$$V_{MDE}^h = (1-\phi^2)\frac{\sum_{i=1}^h \phi^{2(i-1)}(2i-1)}{h^2\phi^{2(h-1)}} > (1-\phi^2) \;\; \text{for} \;\; |\phi| < 1.$$

The last inequality comes from the facts that $\sum_{i=1}^{h}(2i-1) = h^2$ and $\phi^{2(i-1)} \geq \phi^{2(h-1)}$ for $i = 1, \cdots h$. Therefore, the estimator of $\phi$ based on $\hat{\rho}_h$ only for $h = 3, 5, 7 \cdots$ is not asymptotically efficient.

Similarly, for the AR(p) process the MDE using the first p autocorrelations is asymptotically as efficient as MLE under normality, while a formal proof of this result is given below. The AR(p) process is

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \epsilon_t, \tag{9}$$

where $\epsilon_t$ is NID$(0, \sigma^2)$. The sample autocovariance of $y_t$ at lag h is defined by $\hat{\gamma}_h = T^{-1} \sum_{h+1}^{T} y_t y_{t-h}$, where $h = 0, 1, 2, \cdots$. The sample autocorrelation at lag $h$ is defined as $\hat{\rho}_h = \hat{\gamma}_h / \hat{\gamma}_0$. The vector of parameters of the AR(p) process is denoted as $\phi = [\phi_1 \ \cdots \phi_p]'$ and the true value of $\phi$ is denoted as $\phi_0$. The log-likelihood function of the AR(p) process can be written as

$$L = T^{-1} \sum_{t=p+1}^{T} l_t = -\frac{1}{2} \log 2\pi - \frac{1}{2} \log \sigma^2 - \frac{1}{2\,T\sigma^2} \sum_{t=p+1}^{T} (y_t - \phi_1 y_{t-1} - \cdots - \phi_p y_{t-p})^2.$$

Without loss of generality, assume that $\sigma^2$ is known. The MLE is asymptotically equivalent to the GMM estimator using the following population moment conditions: $E(y_t \epsilon_{t-j}) = 0$, $j = 1 \cdots p$. Let the $p \times 1$ score vector of the MLE be denoted as $\mathbf{S}$:

$$\mathbf{S} = \begin{pmatrix} T^{-1} \sum_{t=p+1}^{T} (y_t - \phi_1 y_{t-1} - \cdots - \phi_p y_{t-p}) \ y_{t-1} \\ \vdots \\ T^{-1} \sum_{t=p+1}^{T} (y_t - \phi_1 y_{t-1} - \cdots - \phi_p y_{t-p}) \ y_{t-p} \end{pmatrix}.$$

The GMM based on the above moment conditions is asymptotically as efficient as MLE under normality.

Let $\hat{\gamma} = [\hat{\gamma}_0 \ \hat{\gamma}_1 \ \cdots \ \hat{\gamma}_p \ ]'$. Note that $\mathbf{S}$ can be rewritten as

$$\mathbf{S} = \mathbf{A} \ \hat{\gamma} + \zeta$$

where $\mathbf{A}$ is a $p \times (p+1)$ matrix given by

$$\mathbf{A} = \begin{pmatrix} -\phi_1 & 1-\phi_2 & -\phi_3 & -\phi_4 & \cdots & -\phi_{p-2} & -\phi_{p-1} & -\phi_p & 0 \\ -\phi_2 & -(\phi_1+\phi_3) & 1-\phi_4 & -\phi_5 & \cdots & -\phi_{p-1} & -\phi_p & 0 & 0 \\ -\phi_3 & -(\phi_2+\phi_4) & -(\phi_1+\phi_5) & 1-\phi_6 & \cdots & -\phi_p & 0 & 0 & 0 \\ \vdots & & & & & & & & \\ -\phi_{p-2} & -(\phi_{p-3}+\phi_{p-1}) & -(\phi_{p-4}+\phi_p) & \cdots & \cdots & -\phi_1 & 1 & 0 & 0 \\ -\phi_{p-1} & -(\phi_{p-2}+\phi_p) & -\phi_{p-3} & \cdots & \cdots & -\phi_2 & -\phi_1 & 1 & 0 \\ -\phi_p & -\phi_{p-1} & -\phi_{p-2} & \cdots & \cdots & -\phi_3 & -\phi_2 & -\phi_1 & 1 \end{pmatrix}.$$

and, $\zeta$ is a $p \times 1$ vector given by

$$\zeta = T^{-1} \begin{pmatrix} -\sum_{j=2}^{p} y_j y_{j-1} + \phi_1(\sum_{j=1}^{p-1} y_j^2 + y_T^2) + \phi_2(\sum_{j=2}^{p-1} y_j y_{j-1} + y_T y_{T-1}) + \cdots + \phi_p y_T y_{T-(p-1)} \\ -\sum_{j=3}^{p} y_j y_{j-2} + \phi_1(\sum_{j=2}^{p-1} y_j y_{j-1} + y_T y_{T-1}) + \cdots + \phi_p(y_{T-1} y_{T-p+1} + y_T y_{T-p+2}) \\ \vdots \\ \phi_1 y_T y_{T-(p-1)} + \phi_2(y_{T-1} y_{T-p+1} + y_T y_{T-p+2}) + \cdots + \phi_p \sum_{j=1}^{p} y_{T-p+j}^2 \end{pmatrix}.$$

Since $p$ is fixed, it can be verified that the limiting variance of $T^{1/2}\zeta$ is equal to 0 by using the triangular inequality. Therefore, $\zeta$ is asymptotically negligible.

Now the moment conditions of MDE based on first p autocorrelations can be written as

$$\hat{\rho} - \rho(\phi) = \hat{\gamma}_0^{-1}\ \mathbf{M},$$

where

$$\mathbf{M} = T^{-1} \begin{pmatrix} \sum_{t=2}^{T} y_t y_{t-1} - \rho_1(\phi) \sum_{t=1}^{T} y_t^2 \\ \vdots \\ \sum_{t=p+1}^{T} y_t y_{t-p} - \rho_p(\phi) \sum_{t=1}^{T} y_t^2 \end{pmatrix}.$$

Similarly, we have

$$\mathbf{M} = \mathbf{B}\ \ \hat{\gamma}$$

where $\mathbf{B}$ is a $p \times (p+1)$ matrix given by

$$\mathbf{B} = \begin{pmatrix} -\rho_1 & 1 & 0 & 0 & \cdots & 0 \\ -\rho_2 & 0 & 1 & 0 & \cdots & 0 \\ -\rho_3 & 0 & 0 & 1 & \cdots & 0 \\ \vdots & & & & & \\ -\rho_p & 0 & 0 & 0 & \cdots & 1 \end{pmatrix}.$$

The MDE is asymptotically equivalent to the GMM estimator based on $\mathbf{M}$, since $\hat{\gamma}_0^{-1}$ is not a function of $\phi$. The following result demonstrates that the MDE is asymptotically as efficient as MLE under normality.

Let $\gamma$ be a $(p+1) \times 1$ vector denoted by $[\gamma_0\ \gamma_1\ \cdots\ \gamma_p\ ]'$, where $\gamma_h$ is the $h$th order autocovariance function. It is obvious that

$$\mathbf{B}\ \gamma = 0.$$

Multiplying equation (9) by $y_{t-j}$ for $j = 1, \cdots p$, and taking expectation obtains

$$\mathbf{A}\ \gamma = 0. \tag{10}$$

Hence,

$$S \ = \ A\ (\hat{\gamma} - \gamma) + \zeta,$$

and

$$M \ = \ B\ (\hat{\gamma} - \gamma).$$

The GMM estimator based on moment conditions $\mathbf{S}$ is asymptotically equivalent to the GMM estimator based on moment conditions $\mathbf{M}$ if there exists a $p \times p$, invertible matrix $\mathbf{\Psi}$ such that

$$B = \Psi A. \tag{11}$$

Let $-\rho$ denote the vector $[-\rho_1 \ -\rho_2 \ \cdots \ -\rho_p]'$. We can partition $\mathbf{B}$ as

$$\mathbf{B} = [\ -\rho \mid I_p\ ],$$

where $I_p$ is a $p \times p$ identity matrix.

Let $-\phi$ denote the vector $[-\phi_1 \ -\phi_2 \ \cdots \ -\phi_p]'$. Hence, $\mathbf{A}$ can be partitioned as

$$\mathbf{A} = [\ -\phi \mid L\ ],$$

where $L$ is a $p \times p$ matrix containing the 2th to $p$th columns of $\mathbf{A}$.

Let $\gamma_* = [\gamma_1 \ \cdots \ \gamma_p\ ]'$. Rewriting equation (10) we have

$$L\gamma_* - \phi\gamma_0 = 0.$$

Dividing the above equation by $\gamma_0$ yields

$$L\rho = \phi. \tag{12}$$

For stationary AR(p) process, $\rho$ can be uniquely determined by using the above equation, so that $L$ is invertible.

We thus argue that $\mathbf{\Psi} = L^{-1}$, i.e.,

$$\mathbf{\Psi} = \begin{pmatrix} 1-\phi_2 & -\phi_3 & -\phi_4 & \cdots & -\phi_{p-2} & -\phi_{p-1} & -\phi_p & 0 \\ -(\phi_1+\phi_3) & 1-\phi_4 & -\phi_5 & \cdots & -\phi_{p-1} & -\phi_p & 0 & 0 \\ -(\phi_2+\phi_4) & -(\phi_1+\phi_5) & 1-\phi_6 & \cdots & -\phi_p & 0 & 0 & 0 \\ \vdots & & & & & & & \\ -(\phi_{p-3}+\phi_{p-1}) & -(\phi_{p-4}+\phi_p) & \cdots & \cdots & -\phi_1 & 1 & 0 & 0 \\ -(\phi_{p-2}+\phi_p) & -\phi_{p-3} & \cdots & \cdots & -\phi_2 & -\phi_1 & 1 & 0 \\ -\phi_{p-1} & -\phi_{p-2} & \cdots & \cdots & -\phi_3 & -\phi_2 & -\phi_1 & 1 \end{pmatrix}^{-1}.$$

Also, $\mathbf{\Psi}\,\phi = \rho$, which is implied by equation (12).

Using the result given in equation (11), we now show that the GMM estimator based on moment conditions **S** is asymptotically equivalent to the GMM estimator based on moment conditions **M**.

Let $\hat{\phi}_S$ and $\hat{\phi}_M$ be the GMM estimators based on the moment conditions **S** and **M**, respectively. Under suitable regularity conditions, we have

$$\sqrt{T}(\hat{\phi}_S - \phi_0) \rightarrow N(\,0,\ [D'\Omega^{-1}D]^{-1}\,),$$

where $\Omega$ is the asymptotic variance of $\sqrt{T}\,\mathbf{S}$ and $D = \text{plim}\{\,\frac{\partial S}{\partial \phi'}\big|_{\phi=\phi_0}\}$. Similarly, we have

$$\sqrt{T}(\hat{\phi}_M - \phi_0) \rightarrow N(\,0,\ [D_M'\Omega_M^{-1}D_M]^{-1}\,),$$

where $\Omega_M$ is the asymptotic variance of $\sqrt{T}\,\mathbf{M}$ and $D_M = \text{plim}\{\,\frac{\partial M}{\partial \phi'}\big|_{\phi=\phi_0}\}$. Equation (11) implies that $T^{1/2}M = T^{1/2}\Psi S + o_p(1)$. Thus,

$$\Omega_M \;=\; \Psi\Omega\Psi'$$

and

$$D_M \;=\; \Psi D.$$

Hence, we have

$$[D_M'\Omega_M^{-1}D_M]^{-1} = [\,(\Psi D)'(\Psi\Omega_M\Psi')^{-1}\Psi D)\,]^{-1} = [\,D'\Omega^{-1}D\,]^{-1} = \Omega^{-1}.$$

The last equality comes from the fact that under the assumption of $\epsilon_t$ being i.i.d., the usual information matrix equality holds and $D = \Omega$. Thus, $[\,D'\Omega^{-1}D\,]^{-1} = \Omega^{-1}$.

The above result demonstrates that $\hat{\phi}_S$ and $\hat{\phi}_M$ are asymptotically equivalent. Obviously, $\hat{\phi}_S$ is asymptotically as efficient as MLE under normality, so is the MDE based on the first g autocorrelations.

**Example: MDE for AR(2) Process**

The first order and second order autocorrelation functions of the AR(2) process are

$$\rho_1 = \frac{\phi_1}{(1-\phi_2)},$$

and

$$\rho_2 = \phi_2 + \frac{\phi_1^2}{(1-\phi_2)}.$$

The score vector of the log-likelihood function of the AR(2) process is

$$\begin{aligned}
\mathbf{S} &= \begin{pmatrix} T^{-1}\sum_{t=p+1}^{T}(y_t - \phi_1 y_{t-1} - \phi_2 y_{t-2})\ y_{t-1} \\ T^{-1}\sum_{t=p+1}^{T}(y_t - \phi_1 y_{t-1} - \phi_2 y_{t-2})\ y_{t-2} \end{pmatrix} \\
&= \begin{pmatrix} \hat\gamma_1 - \phi_1\hat\gamma_0 - \phi_2\hat\gamma_1 \\ \hat\gamma_2 - \phi_1\hat\gamma_1 - \phi_2\hat\gamma_0 \end{pmatrix} + T^{-1}\begin{pmatrix} -y_2y_1 + \phi_1(y_1^2 + y_T^2) + \phi_2(y_Ty_{T-1}) \\ \phi_1(y_Ty_{T-1}) + \phi_2(y_T^2 + y_{T-1}^2) \end{pmatrix} \\
&= \begin{pmatrix} -\phi_1 & 1-\phi_2 & 0 \\ -\phi_2 & -\phi_1 & 1 \end{pmatrix}\begin{pmatrix} \hat\gamma_0 \\ \hat\gamma_1 \\ \hat\gamma_2 \end{pmatrix} + \zeta,
\end{aligned}$$

where,

$$\zeta = T^{-1}\begin{pmatrix} -y_2y_1 + \phi_1(y_1^2 + y_T^2) + \phi_2(y_Ty_{T-1}) \\ \phi_1(y_Ty_{T-1}) + \phi_2(y_T^2 + y_{T-1}^2) \end{pmatrix}$$

It can be verified that $\lim_{T\to\infty} Var(T^{1/2}\zeta) = 0$, so that $\zeta$ is asymptotically negligible.

The moment conditions in MDE can be expressed as

$$\begin{pmatrix} \hat\gamma_1 - \rho_1\hat\gamma_0 \\ \hat\gamma_2 - \rho_2\hat\gamma_0 \end{pmatrix} = \begin{pmatrix} -\rho_1 & 1 & 0 \\ -\rho_2 & 0 & 1 \end{pmatrix}\begin{pmatrix} \hat\gamma_0 \\ \hat\gamma_1 \\ \hat\gamma_2 \end{pmatrix}$$

For stationarity, the roots of $1 - \phi_1 B - \phi_2 B^2 = 0$ must lie outside the unit circle, which implies that $|\phi_2| < 1$. Hence, $L$ is invertible and

$$\Psi = \begin{pmatrix} 1-\phi_2 & 0 \\ -\phi_1 & 1 \end{pmatrix}^{-1} = \frac{1}{1-\phi_2}\begin{pmatrix} 1 & 0 \\ \phi_1 & 1-\phi_2 \end{pmatrix}$$

Clearly, $\Psi^{-1}$ exists and

$$\Psi\ \phi = \frac{1}{1-\phi_2}\begin{pmatrix} 1 & 0 \\ \phi_1 & 1-\phi_2 \end{pmatrix}\begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix} = \begin{pmatrix} \frac{\phi_1}{(1-\phi_2)} \\ \phi_2 + \frac{\phi_1^2}{(1-\phi_2)} \end{pmatrix} = \begin{pmatrix} \rho_1 \\ \rho_2 \end{pmatrix}.$$

## 4. MDE Applied to Moving Average Processes

### 4.1 MDE Applied to MA(1) Process

The MA(1) process is,

$$y_t = (1 - \theta L)\epsilon_t, \tag{13}$$

where $\epsilon_t$ is iid(0, $\sigma^2$) and the autocorrelation function of the MA(1) process is $\rho_1 = \frac{-\theta}{1+\theta^2}$ and $\rho_k = 0$ for $k \geq 2$. The limiting distribution of MLE of $\theta$ when $\epsilon_t$ is iid(0, $\sigma^2$) is,

$$\sqrt{T}(\hat{\theta}_{MLE} - \theta) \rightarrow N(0, 1-\theta^2).$$

To calculate the asymptotic variance of MDE, it is necessary to obtain the asymptotic variance of sample autocorrelations, i.e, C. It is easily verified that for the MA(1) process

$$C = \begin{bmatrix} 1-3\rho_1^2+4\rho_1^4 & 2\rho_1(1-\rho_1^2) & \rho_1^2 & 0 & \cdots & 0 \\ 2\rho_1(1-\rho_1^2) & 1+2\rho_1^2 & 2\rho_1 & \rho_1^2 & \ddots & \vdots \\ \rho_1^2 & 2\rho_1 & 1+2\rho_1^2 & 2\rho_1 & \ddots & 0 \\ 0 & \rho_1^2 & 2\rho_1 & 1+2\rho_1^2 & \ddots & \rho_1^2 \\ \vdots & \ddots & \ddots & \ddots & \ddots & 2\rho_1 \\ 0 & \cdots & 0 & \rho_1^2 & 2\rho_1 & 1+2\rho_1^2 \end{bmatrix},$$

and $D = [D_1 \; 0 \cdots 0]'$, where $D_1 = \frac{1-\theta^2}{(\theta^2+1)^2}$. For convenience, let the number of autocorrelations used in MDE be denoted by $g$. When $g$ is small, e.g., $g = 1$ or 2, it is very straightforward to obtain analytical results for the asymptotic variance of $\sqrt{T}(\hat{\theta} - \theta)$, i.e., $(D'C^{-1}D)^{-1}$. From direct calculation the asymptotic variance of the MDE based on only the first $g$ autocorrelations for $g$ =1, 2 and 3 are

$$V^{g=1} = \frac{1+\theta^2+4\theta^4+\theta^6+\theta^8}{(\theta^2-1)^2},$$

$$V^{g=2} = \frac{1+\theta^2+\theta^4+6\theta^6+\theta^8+\theta^{10}+\theta^{12}}{(\theta^2-1)^2(1+4\theta^2+\theta^4)},$$

and

$$V^{g=3} = \frac{1+\theta^2+\theta^4+\theta^6+8\theta^8+\theta^{10}+\theta^{12}+\theta^{14}+\theta^{16}}{(\theta^2-1)^2(1+4\theta^2+10\theta^4+4\theta^6+\theta^8)}.$$

The increase in asymptotic efficiency from using two rather than one moment is

$$(V^{g=1} - V^{g=2}) = \frac{4\theta^2(1+2\theta^2+3\theta^4+2\theta^6+\theta^8)}{(\theta^2-1)^2(1+4\theta^2+\theta^4)}.$$

Similarly,

$$(V^{g=2} - V^{g=3}) = \frac{9(\theta^4+2\theta^6+3\theta^8+4\theta^{10}+3\theta^{12}+2\theta^{14}+\theta^{16})}{(\theta^2-1)^2(1+4\theta^2+\theta^4)(1+4\theta^2+10\theta^4+4\theta^6+\theta^8)},$$

which is always positive. Therefore, adding the second order autocorrelation to the MDE improve the asymptotic efficiency, so does adding the third order autocorrelation to the MDE.

Table 1 shows that asymptotic efficiency is increased as the number of moments increases and presents the asymptotic variance of $\sqrt{T}(\hat{\theta}_{MDE} - \theta)$ for the MA(1) model for $\theta = 0.1, 0.2, 0.3, \cdots$, and 0.9. Calculations reveal that as $\theta$ increases, the asymptotic variance of the MDE increases. Table 1 only reports positive values of $\theta$ since the results are symmetric around zero. If the absolute values of the MA(1) coefficients are the same, the asymptotic variance of the MDE is the same.

Durbin (1959) has suggested estimating MA(q) models by their AR approximation and shows that if the number of AR coefficients is large enough, this estimator is asymptotically as efficient as MLE. Analogously, it is necessary to have more autocorrelations for the MDE in order to achieve relative efficiency for large value of $\theta$. Galbraith and Zinde-Walsh (1994) suggest estimate the MA models based on minimizing the Hilbert distance between the MA model and its AR approximation and have similar conclusions.

In many cases it is surprising to note that a remarkably small number of autocorrelations are necessary for the MDE to be asymptotically efficient to two decimal places when compared to that of the MLE under normality. For example, when $g = 5$ and $\theta \in (0, 0.4)$ the asymptotic variance of MDE is very close to that of MLE. The result implies that if enough autocorrelations are used, the MDE can be asymptotically as efficient as MLE. Apart one extreme case when the moving average parameter, $\theta$ is .9, the MDE is seen to be remarkably efficient.

While the result in Table 1 indicates that the MDE appears to be as efficient as MLE given that $g$ is large enough, we now provide a theoretical justification of

Table 1: Asymptotic Variance of MDE of MA(1) Processes: $y_t = (1 - \theta L)\epsilon_t$

| | $\theta$ | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | 0.10 | 0.20 | 0.30 | 0.40 | 0.50 | 0.60 | 0.70 | 0.80 | 0.90 |
| g= 1 | 1.031 | 1.135 | 1.356 | 1.796 | 2.701 | 4.741 | 10.095 | 28.614 | 149.482 |
| 2 | 0.991 | 0.973 | 0.973 | 1.030 | 1.217 | 1.705 | 3.046 | 7.710 | 37.999 |
| 3 | 0.990 | 0.961 | 0.919 | 0.885 | 0.899 | 1.041 | 1.541 | 3.394 | 15.526 |
| 5 | 0.990 | 0.960 | 0.910 | 0.842 | 0.767 | 0.717 | 0.776 | 1.247 | 4.693 |
| 10 | 0.990 | 0.960 | 0.910 | 0.840 | 0.750 | 0.641 | 0.526 | 0.472 | 0.934 |
| 20 | 0.990 | 0.960 | 0.910 | 0.840 | 0.750 | 0.640 | 0.510 | 0.363 | 0.280 |
| $V_{MLE}$ | 0.990 | 0.960 | 0.910 | 0.840 | 0.750 | 0.640 | 0.510 | 0.360 | 0.190 |

Note: g=1 means only the first autocorrelation are used in MDE, while g=2 represents the case that first 2 autocorrelations are used in MDE, etc. $V_{MLE}$ represents the asymptotic variance of $\sqrt{T}(\hat{\theta}_{MLE} - \theta)$.

this result. The basic idea is to rewrite the score of the log-likelihood function as a weighted sum of the sample autocovariances at various lags. It is found that the weight for each sample autocovariance decreases exponentially. The lower bound of the asymptotic variance of the MDE is equal to the asymptotic variance of MLE under normality. Given that $g$ is large enough, it will be shown later that the upper and lower bounds of the asymptotic variance of MDE are the same when T $\rightarrow \infty$. Therefore, if the number of autocorrelations used in the MDE is large enough, the MDE for the MA(1) process is asymptotically as efficient as MLE under normality.

First, note that the log-likelihood function of the MA(1) model is

$$L = T^{-1}\sum_{t=1}^{T} l_t = -\frac{1}{2}\log 2\pi - \frac{1}{2}\log\sigma^2 - \frac{1}{2T\sigma^2}\sum_{t=2}^{T}\left(\sum_{j=0}^{t-1}\theta^j\ y_{t-j}\right)^2 .$$

Without loss of generality we assume $\sigma^2$ is known. The score of the log-likelihood function is

$$\begin{aligned}
S &= T^{-1}\sum_{t=2}^{T}(\sum_{j=0}^{t-1}\theta^j\ y_{t-j})\ (\sum_{i=1}^{t-1} i\ \theta^{i-1}\ y_{t-i}) \\
&= T^{-1}\sum_{t=2}^{T}\sum_{j=0}^{t-1}\theta^j\ y_{t-j} j\theta^{j-1} y_{t-j} \\
&\quad + T^{-1}\sum_{h=1}^{T-1}\sum_{t=2}^{T}\left[\sum_{j=0}^{t-1}\theta^j\ y_{t-j}\ (j+h)\theta^{j+h-1} y_{t-j-h} + \sum_{j=0}^{t-1}\theta^{j+h}\ y_{t-j-h}\ j\theta^{j-1} y_{t-j}\right] \\
&= T^{-1}\sum_{t=2}^{T}\sum_{j=0}^{t-1} j\theta^{2j-1} y_{t-j}^2 + T^{-1}\sum_{h=1}^{T-1}\sum_{t=2}^{T}\sum_{j=0}^{t-1}(2j+h)\theta^{2j+h-1} y_{t-j-h} y_{t-j} \\
&= \sum_{j=0}^{T-1} j\ \theta^{2j-1}\ \hat{\gamma}_0\ +\ \sum_{h=1}^{T-1}[\theta^{h-1}\sum_{j=0}^{T-1}(2j+h)\ \theta^{2j}\ \hat{\gamma}_h] - \zeta_1
\end{aligned}$$

where

$$\zeta_1 = T^{-1}\sum_{j=1}^{T-1} j\theta^{2j-1}\sum_{\tau=T-j+1}^{T} y_\tau^2\ +\ \sum_{h=1}^{T-1}\theta^{h-1}\sum_{j=1}^{T-1}(2j+h)\theta^{2j}\sum_{\tau=T-j+1}^{T} y_\tau y_{\tau-h},$$

and the sample autocovariance of $y_t$ at lag $h$ is defined by $\hat{\gamma}_h = T^{-1}\sum_{t=h+1}^{T} y_t y_{t-h}$, where $h = 0, 1, 2, \cdots$.

The score of the log-likelihood function can be rewritten as

$$\begin{aligned}
S &= \sum_{j=0}^{T-1} j\ \theta^{2j-1}\ \hat{\gamma}_0\ +\ \sum_{h=1}^{T-1}[\theta^{h-1}\sum_{m=0}^{T-1}(h+2m)\ \theta^{2m}\ \hat{\gamma}_h] - \zeta_1 \\
&= \sum_{i=0}^{T-1} w_i\ \hat{\gamma}_i - \zeta_1,
\end{aligned}$$

where $w_0 = \sum_{j=0}^{T-1} j\ \theta^{2j-1}$, and $w_h = \theta^{h-1}\sum_{m=0}^{T-1}(h+2m)\ \theta^{2m}$ for $h = 1, 2, \cdots$. The variance of $T^{1/2}\zeta_1$ approaches zero as $T \to \infty$, while a formal proof of this result will be given later. Hence, $\zeta_1$ is asymptotically negligible.

When $T$ is very large, $w_0$ is approximately equal to $w_0^*$, where

$$w_0^* = \frac{\theta}{(1-\theta^2)^2}.$$

Similarly, for $h \geq 1$ $w_h$ is approximately equal to $w_h^*$, where

$$w_h^* = \theta^{h-1}\left[\frac{h-(h-2)\theta^2}{(1-\theta^2)^2}\right], \quad \text{for } h = 1, 2, \cdots. \tag{14}$$

Hence, given that $T$ is very large, we can rewrite $S$ as

$$S = \sum_{i=0}^{T-1} w_i^* \, \hat{\gamma}_i - \zeta_1. \tag{15}$$

This result demonstrates that the score of the MLE of MA(1) process is approximately equal to a weighted sum of the sample autocovariance at various lags. The weight ($w_i^*$) for each sample autocovariance decreases exponentially.

Let $S_{(g)}$ be defined as the summation of only the first $g+1$ terms of $S$, so that

$$S_{(g)} = \sum_{i=0}^{g} w_i^* \, \hat{\gamma}_i. \tag{16}$$

Now the moment conditions of MDE based on the first g autocorrelations can be written as

$$\hat{\rho} - \rho(\theta) = \hat{\gamma}_0^{-1} \, \mathbf{M},$$

where

$$\mathbf{M} = \begin{bmatrix} \hat{\gamma}_1 - \rho_1(\theta)\hat{\gamma}_0 \\ \hat{\gamma}_2 \\ \vdots \\ \hat{\gamma}_g \end{bmatrix}.$$

Let $\mathbf{R}$ be a $g \times g$ matrix given by

$$\mathbf{R} = \begin{bmatrix} w_1^* & w_2^* & w_3^* & \cdots & & w_g^* \\ 0 & 1 & 0 & \cdots & & 0 \\ 0 & 0 & 1 & \cdots & & 0 \\ \vdots & \vdots & & & & \vdots \\ 0 & 0 & & & & 0 \\ 0 & 0 & 0 & \cdots & 0 & 1 \end{bmatrix}.$$

Notice that the MDE is asymptotically equivalent to the GMM estimator based on moment conditions $\mathbf{RM}$, because the GMM estimator minimizing $\mathbf{M}'\Omega^{-1}\mathbf{M}$ also minimizes $(\mathbf{RM})'(\mathbf{R}\Omega\mathbf{R}')^{-1}\mathbf{RM}$, where $\Omega$ is the asymptotic variance of $\sqrt{T}\,\mathbf{M}$.

The first element of the vector $\mathbf{R} \times \mathbf{M}$ is equal to $w_1^* \times (-\rho_1)\ \hat{\gamma}_0 + \sum_{i=1}^{g} w_i^* \hat{\gamma}_i$. Using the result given in equation (14) yields

$$w_1^* = \frac{1+\theta^2}{(1-\theta^2)^2}.$$

Because $\rho_1 = \frac{-\theta}{1+\theta^2}$, we have

$$w_1^* \times (-\rho_1) = \frac{\theta}{(1-\theta^2)^2} = w_0^*.$$

It follows directly that the 1st element of $\mathbf{R} \times \mathbf{M}$ is equal to $S_{(g)}$. Hence,

$$\mathbf{R} \times \mathbf{M} = \begin{bmatrix} w_1^* \times (-\rho_1)\ \hat{\gamma}_0 + \sum_{i=1}^{g} w_i^* \hat{\gamma}_i \\ \hat{\gamma}_2 \\ \vdots \\ \hat{\gamma}_g \end{bmatrix} = \begin{bmatrix} S_{(g)} \\ \hat{\gamma}_2 \\ \vdots \\ \hat{\gamma}_g \end{bmatrix}.$$

For convenience let the asymptotic variances of GMM estimator based on $\mathbf{RM}$ and $S_{(g)}$ be denoted as $V_{RM}$ and $V_{Sg}$, respectively. Also $V_{MLE}$ denotes the asymptotic variance of MLE. Since $S_{(g)}$ is only a subset of moment conditions of $\mathbf{RM}$, we have $V_{MLE} \leq V_{RM} \leq V_{Sg}$.

It will be demonstrated below that when $g$ is very large, the asymptotic variances of $S_{(g)}$ and $S$ are the same, so that $S_{(g)}$ and $S$ are asymptotically equivalent. If g is very large, $V_{Sg} = V_{MLE}$. It also implies that $V_{RM} = V_{MLE}$. Therefore, the MDE is asymptotically as efficient as MLE given that the number of autocorrelations is large enough.

The following result demonstrates that when $g$ is very large, the asymptotic variances of $S_{(g)}$ and $S$ are the same. By the Cauchy-Schwarz inequality, it is sufficient to show that

$$\lim_{g\to\infty} \limsup_{T\to\infty} \left(E|\ T^{1/2}(S - S_{(g)})\ |^2\right)^{1/2} = 0.$$

First note that

$$T^{1/2}(S - S_{(g)}) = T^{1/2}[\ \sum_{h=g+1}^{T-1} (w_h^*\, T^{-1} \sum_{t=h+1}^{T} y_t y_{t-h})\ ] - T^{1/2}\zeta_1,$$

where $\lim_{T\to\infty} Var(\ T^{1/2}\ \zeta_1\ ) = 0$. Using the triangular inequality gives

$$\begin{aligned}
& \lim_{g\to\infty} \limsup_{T\to\infty} \left(E|\ T^{1/2}(S - S_{(g)})\ |^2\right)^{1/2} \\
= \ & \lim_{g\to\infty} \limsup_{T\to\infty} \left(\ E\ |\ \sum_{h=g+1}^{T-1} (w_h^*\, T^{-1/2} \sum_{t=h+1}^{T} y_t y_{t-h}) - T^{-1/2}\ \zeta_1\ |^2\ \right)^{1/2} \\
\leq \ & \lim_{g\to\infty} \limsup_{T\to\infty} \|\ \sum_{h=g+1}^{T-1}\ w_h^*\, T^{-1/2} \sum_{t=h+1}^{T} y_t y_{t-h}\|_2 \\
\leq \ & \lim_{g\to\infty} \limsup_{T\to\infty} \sum_{h=g+1}^{T-1}\ w_h^*\ \|\ T^{-1/2} \sum_{t=h+1}^{T} y_t y_{t-h}\|_2 \\
\leq \ & c_1 \lim_{g\to\infty} \limsup_{T\to\infty} \sum_{h=g+1}^{T-1}\ w_h^* \\
\leq \ & c_2 \lim_{g\to\infty} exp(-c_3\ g) \\
= \ & 0.
\end{aligned}$$

By using the result given in Brockwell and Davis (1991, p.227) it follows directly that $c_1$ is a finite positive constant which does not depend on $h$. Note that $c_2$ and $c_3$ are some positive constants which are functions of $\theta$.

The following result demonstrates that $\lim_{T\to\infty} Var(\, T^{1/2}\, \zeta_1 \,) = 0$.

$$
\begin{aligned}
& \lim_{T\to\infty} (E \mid T^{1/2}\, \zeta_1 \mid^2)^{1/2} \\
= & \lim_{T\to\infty} (\, E \mid T^{-1/2} \{ \sum_{j=1}^{T-1} j\theta^{2j-1} \sum_{\tau=T-j+1}^{T} y_\tau^2 \\
& \quad + \sum_{h=1}^{T-1} \theta^{h-1} \sum_{j=1}^{T-1} (2j+h)\theta^{2j} \sum_{\tau=T-j+1}^{T} y_\tau y_{\tau-h} \} \mid^2 )^{1/2} \\
= & \lim_{T\to\infty} (\, E \mid T^{-1/2} \{ \sum_{j=1}^{T-1} (2j+1)\theta^{2j} \sum_{\tau=T-j+1}^{T} y_\tau^2 + \sum_{h=1}^{T-1} \theta^{h-1} \sum_{j=1}^{T-1} 2j\theta^{2j} \sum_{\tau=T-j+1}^{T} y_\tau y_{\tau-h} \\
& \quad + \sum_{h=1}^{T-1} \theta^{h-1} h \sum_{j=1}^{T-1} \theta^{2j} \sum_{\tau=T-j+1}^{T} y_\tau y_{\tau-h} \mid^2 )^{1/2} \\
\leq & \lim_{T\to\infty} T^{-1/2} \{ \sum_{j=1}^{T-1} (2j+1)\theta^{2j} \sum_{\tau=T-j+1}^{T} \| y_\tau^2 \|_2 + \sum_{h=1}^{T-1} \theta^{h-1} \sum_{j=1}^{T-1} 2j\theta^{2j} \sum_{\tau=T-j+1}^{T} \| y_\tau y_{\tau-h} \|_2 \\
& \quad + \sum_{h=1}^{T-1} \theta^{h-1} h \sum_{j=1}^{T-1} \theta^{2j} \sum_{\tau=T-j+1}^{T} \| y_\tau y_{\tau-h} \|_2 \} \\
\leq & \lim_{T\to\infty} T^{-1/2} \{ \sum_{j=1}^{T-1} (2j+1)\theta^{2j} j\, a_1 + \sum_{h=1}^{T-1} \theta^{h-1} \sum_{j=1}^{T-1} 2j\theta^{2j} j\, a_1 + \sum_{h=1}^{T-1} \theta^{h-1} h \sum_{j=1}^{T-1} \theta^{2j} j\, a_1 \} \\
\leq & \lim_{T\to\infty} T^{-1/2}\, a_1 (\, b_1 + \sum_{h=1}^{T-1} \theta^{h-1}\, b_2 + \sum_{h=1}^{T-1} \theta^{h-1} h\, b_3 \,) \\
\leq & \lim_{T\to\infty} T^{-1/2}\, a_1 (\, b_1 + \frac{b_2}{1-\theta} + \frac{b_3}{(1-\theta^2)^2} \,) \\
= & \; 0,
\end{aligned}
$$

where $a_1 = \| y_\tau^2 \|_2 = (E y_\tau^4)^{1/2}$, and $\| y_\tau y_{\tau-h} \|_2 = (E y_\tau^2 y_{\tau-h}^2)^{1/2} \leq (E y_\tau^4 E y_{\tau-h}^4)^{1/4} = a_1$.

Also, it can be verified that $b_1$, $b_2$ and $b_3$ are finite constants such that

$$
\begin{aligned}
b_1 &= \lim_{T\to\infty} \sum_{j=1}^{T-1} j\theta^{2j-1} j = \frac{\theta + \theta^3}{(1-\theta^2)^3}, \\
b_2 &= \lim_{T\to\infty} \sum_{j=1}^{T-1} 2j\theta^{2j} j = \frac{2\theta^2(1+\theta^2)}{(1-\theta^2)^3},
\end{aligned}
$$

and

$$b_3 = \lim_{T\to\infty} \sum_{j=1}^{T-1} \theta^{2j} j = \frac{\theta^2}{(1-\theta^2)^2}.$$

### 4.2 MDE Applied to MA(2) Process

The asymptotic variance of the MDE for estimating the parameters of an MA(2) process can be investigated in an analogous manner. The MA(2) process is defined as,

$$y_t = (1 - \theta_1 L - \theta_2 L^2)\epsilon_t = (1 - \delta_1 L)(1 - \delta_2 L)\epsilon_t,$$

where $\epsilon_t$ is iid(0, $\sigma^2$) and the second expression is in terms of the two invertible roots of the MA polynomial. Let the maximum likelihood estimator of the MA(2) model be denoted as $\hat{\lambda}_{MLE}$, where $\lambda$ corresponds to $[\theta_1\ \theta_2]'$. Then, we have $\sqrt{T}(\hat{\lambda}_{MLE} - \lambda_0) \to N(0, V_{MLE})$, where

$$V_{MLE} = \begin{bmatrix} 1-\theta_2^2 & \theta_1(1-\theta_2) \\ \theta_1(1-\theta_2) & 1-\theta_2^2 \end{bmatrix}.$$

Hence, the asymptotic variances of MLE of $\theta_1$ and $\theta_2$ depend only on the parameter $\theta_2$. The autocorrelation function is

$$\rho_1 = \frac{-\theta_1(1-\theta_2)}{1+\theta_1^2+\theta_2^2},$$
$$\rho_2 = \frac{-\theta_2}{1+\theta_1^2+\theta_2^2},$$

and

$$\rho_k = 0 \text{ for } k \geq 3.$$

In the appendix, we provide the analytical results of the asymptotic variance of the sample autocorrelations calculated by the Bartlett's formula.

For the MA(2) process with i.i.d. innovations, and given a set of parameter values the asymptotic covariance matrix C, of the sample autocorrelation is calculated from Bartlett's formula, while the matrix of partial derivatives, D, is analytically straightforward to calculate. The asymptotic variance of $\sqrt{T}(\hat{\lambda}_{MDE} - \lambda_0)$ for the MA(2) model are then the diagonal elements of the matrix $(D'C^{-1}D)^{-1}$, and are reported in Tables 2 and 3, for four different values of $\delta_1$ and eight values of $\delta_2$ which give a total of thirty two points of the parameter space. The theoretical asymptotic variances of the parameter estimates from the MDE are calculated from the use of g autocorrelations, where g = 2,3,5,10,15 and 20. As the number of autocorrelations, g, increases, the asymptotic variance of the MDE parameter estimates decreases and approaches that of the MLE. The first panel of Table 2 presents the asymptotic variance of MDE of MA(2) processes with $\delta_1$ fixed at 0.1, whereas in the second panel $\delta_1 = 0.3$. Table 3 presents the results for $\delta_1 = 0.5$ and 0.7. When $\delta_2$ increases, the number of autocorrelations used in the MDE has to be increased in order for the MDE to be asymptotically as efficient as MLE under normality. If both $\delta_1$ and $\delta_2$ are not too large, the asymptotic variance of the MDE using only 5 autocorrelations is very close to that of MLE under normality.

## 5 MDE Applied to ARMA(1,1) Process

In this section, we discuss the asymptotic properties of MDE of the ARMA(1,1)

Table 2: Asymptotic Variance of $\sqrt{T}(\hat{\theta}_1 - \theta_1)$ and $\sqrt{T}(\hat{\theta}_2 - \theta_2)$ of MA(2) processes, $y_t = (1 - \theta_1 L - \theta_2 L^2)\epsilon_t = (1 - \delta_1 L)(1 - \delta_2 L)\epsilon_t$ with $\delta_1 = 0.10$ and $0.30$

| | $\delta_1$ | 0.10 | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | $\delta_2$ | 0.20 | 0.40 | 0.60 | 0.80 | -0.20 | -0.40 | -0.60 | -0.80 |
| | | Asymptotic variance of $\sqrt{T}(\hat{\theta}_1 - \theta_1)$ | | | | | | | |
| g= | 2 | 1.00 | 1.10 | 2.45 | 27.00 | 1.00 | 1.04 | 1.53 | 9.35 |
| | 3 | 1.00 | 1.00 | 1.16 | 5.18 | 1.00 | 1.00 | 1.11 | 3.13 |
| | 5 | 1.00 | 1.00 | 1.00 | 1.41 | 1.00 | 1.00 | 1.01 | 1.34 |
| | 10 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.01 |
| | 15 | 1.00 | 1.00 | 1.00 | 0.99 | 1.00 | 1.00 | 1.00 | 1.00 |
| | 20 | 1.00 | 1.00 | 1.00 | 0.99 | 1.00 | 1.00 | 1.00 | 0.99 |
| | $V_{MLE}$ | 1.00 | 1.00 | 1.00 | 0.99 | 1.00 | 1.00 | 1.00 | 0.99 |
| | | Asymptotic variance of $\sqrt{T}(\hat{\theta}_2 - \theta_2)$ | | | | | | | |
| g= | 2 | 1.38 | 2.09 | 3.14 | 3.78 | 1.04 | 1.36 | 2.09 | 3.73 |
| | 3 | 1.04 | 1.25 | 1.70 | 2.22 | 1.00 | 1.09 | 1.41 | 2.22 |
| | 5 | 1.00 | 1.02 | 1.14 | 1.45 | 1.00 | 1.00 | 1.08 | 1.43 |
| | 10 | 1.00 | 1.00 | 1.00 | 1.08 | 1.00 | 1.00 | 1.00 | 1.07 |
| | 15 | 1.00 | 1.00 | 1.00 | 1.01 | 1.00 | 1.00 | 1.00 | 1.01 |
| | 20 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| | $V_{MLE}$ | 1.00 | 1.00 | 1.00 | 0.99 | 1.00 | 1.00 | 1.00 | 0.99 |
| | $\delta_1$ | 0.30 | | | | | | | |
| | $\delta_2$ | 0.20 | 0.40 | 0.60 | 0.80 | -0.20 | -0.40 | -0.60 | -0.80 |
| | | Asymptotic variance of $\sqrt{T}(\hat{\theta}_1 - \theta_1)$ | | | | | | | |
| g= | 2 | 1.06 | 1.65 | 8.12 | 117.59 | 1.02 | 1.06 | 1.40 | 5.43 |
| | 3 | 1.00 | 1.01 | 1.54 | 15.07 | 1.00 | 0.99 | 1.07 | 2.54 |
| | 5 | 1.00 | 0.99 | 0.98 | 1.77 | 1.00 | 0.99 | 0.98 | 1.30 |
| | 10 | 1.00 | 0.99 | 0.97 | 0.95 | 1.00 | 0.99 | 0.97 | 0.97 |
| | 15 | 1.00 | 0.99 | 0.97 | 0.94 | 1.00 | 0.99 | 0.97 | 0.95 |
| | 20 | 1.00 | 0.99 | 0.97 | 0.94 | 1.00 | 0.99 | 0.97 | 0.94 |
| | $V_{MLE}$ | 1.00 | 0.99 | 0.97 | 0.94 | 1.00 | 0.99 | 0.97 | 0.94 |
| | | Asymptotic variance of $\sqrt{T}(\hat{\theta}_2 - \theta_2)$ | | | | | | | |
| g= | 2 | 2.11 | 3.26 | 4.00 | 1.71 | 1.05 | 1.08 | 1.50 | 3.08 |
| | 3 | 1.23 | 1.63 | 2.15 | 1.70 | 1.02 | 1.06 | 1.31 | 2.27 |
| | 5 | 1.01 | 1.06 | 1.26 | 1.42 | 1.00 | 0.99 | 1.03 | 1.39 |
| | 10 | 1.00 | 0.99 | 0.98 | 1.05 | 1.00 | 0.99 | 0.97 | 1.02 |
| | 15 | 1.00 | 0.99 | 0.97 | 0.97 | 1.00 | 0.99 | 0.97 | 0.96 |
| | 20 | 1.00 | 0.99 | 0.97 | 0.95 | 1.00 | 0.99 | 0.97 | 0.95 |
| | $V_{MLE}$ | 1.00 | 0.99 | 0.97 | 0.94 | 1.00 | 0.99 | 0.97 | 0.94 |

Table 3: Asymptotic Variance of $\sqrt{T}(\hat{\theta_1} - \theta_1)$ and $\sqrt{T}(\hat{\theta_2} - \theta_2)$ of MA(2) processes, $y_t = (1 - \theta_1 L - \theta_2 L^2)\epsilon_t = (1 - \delta_1 L)(1 - \delta_2 L)\epsilon_t$ with $\delta_1$ = 0.50 and 0.70

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | $\delta_1$ | 0.50 | | | | | | | |
| | $\delta_2$ | 0.20 | 0.40 | 0.60 | 0.80 | -0.20 | -0.40 | -0.60 | -0.80 |
| | | Asymptotic variance of $\sqrt{T}(\hat{\theta_1} - \theta_1)$ | | | | | | | |
| g= | 2 | 1.86 | 6.59 | 48.76 | 714.85 | 1.14 | 1.19 | 1.52 | 4.65 |
| | 3 | 1.04 | 1.33 | 5.04 | 80.51 | 1.02 | 0.98 | 1.01 | 2.22 |
| | 5 | 0.99 | 0.96 | 0.99 | 4.63 | 0.99 | 0.96 | 0.93 | 1.21 |
| | 10 | 0.99 | 0.96 | 0.91 | 0.85 | 0.99 | 0.96 | 0.91 | 0.88 |
| | 15 | 0.99 | 0.96 | 0.91 | 0.84 | 0.99 | 0.96 | 0.91 | 0.85 |
| | 20 | 0.99 | 0.96 | 0.91 | 0.84 | 0.99 | 0.96 | 0.91 | 0.84 |
| | $V_{MLE}$ | 0.99 | 0.96 | 0.91 | 0.84 | 0.99 | 0.96 | 0.91 | 0.84 |
| | | Asymptotic variance of $\sqrt{T}(\hat{\theta_2} - \theta_2)$ | | | | | | | |
| g= | 2 | 3.20 | 4.15 | 2.38 | 53.73 | 1.39 | 1.18 | 1.41 | 3.03 |
| | 3 | 1.65 | 2.16 | 2.16 | 4.19 | 1.16 | 1.16 | 1.39 | 2.66 |
| | 5 | 1.09 | 1.23 | 1.44 | 1.15 | 1.01 | 0.98 | 0.99 | 1.36 |
| | 10 | 0.99 | 0.96 | 0.95 | 1.00 | 0.99 | 0.96 | 0.91 | 0.91 |
| | 15 | 0.99 | 0.96 | 0.91 | 0.89 | 0.99 | 0.96 | 0.91 | 0.85 |
| | 20 | 0.99 | 0.96 | 0.91 | 0.85 | 0.99 | 0.96 | 0.91 | 0.84 |
| | $V_{MLE}$ | 0.99 | 0.96 | 0.91 | 0.84 | 0.99 | 0.96 | 0.91 | 0.84 |
| | $\delta_1$ | 0.70 | | | | | | | |
| | $\delta_2$ | 0.20 | 0.40 | 0.60 | 0.80 | -0.20 | -0.40 | -0.60 | -0.80 |
| | | Asymptotic variance of $\sqrt{T}(\hat{\theta_1} - \theta_1)$ | | | | | | | |
| g= | 2 | 12.64 | 63.91 | 478.11 | 7168.56 | 2.39 | 2.12 | 2.42 | 5.41 |
| | 3 | 2.29 | 7.08 | 49.72 | 816.60 | 1.38 | 1.27 | 1.18 | 2.01 |
| | 5 | 1.04 | 1.11 | 2.59 | 42.59 | 1.05 | 1.00 | 0.92 | 1.12 |
| | 10 | 0.98 | 0.92 | 0.83 | 0.89 | 0.98 | 0.93 | 0.83 | 0.76 |
| | 15 | 0.98 | 0.92 | 0.83 | 0.70 | 0.98 | 0.92 | 0.82 | 0.69 |
| | 20 | 0.98 | 0.92 | 0.82 | 0.70 | 0.98 | 0.92 | 0.82 | 0.69 |
| | $V_{MLE}$ | 0.98 | 0.92 | 0.82 | 0.69 | 0.98 | 0.92 | 0.82 | 0.69 |
| | | Asymptotic variance of $\sqrt{T}(\hat{\theta_2} - \theta_2)$ | | | | | | | |
| g= | 2 | 3.81 | 2.08 | 33.90 | 1872.12 | 2.28 | 1.84 | 2.04 | 4.16 |
| | 3 | 2.14 | 2.00 | 2.58 | 187.19 | 1.64 | 1.64 | 2.01 | 4.05 |
| | 5 | 1.33 | 1.44 | 1.27 | 6.98 | 1.17 | 1.10 | 1.09 | 1.56 |
| | 10 | 1.01 | 0.98 | 0.98 | 0.86 | 0.99 | 0.93 | 0.83 | 0.76 |
| | 15 | 0.98 | 0.93 | 0.85 | 0.79 | 0.98 | 0.92 | 0.82 | 0.70 |
| | 20 | 0.98 | 0.92 | 0.83 | 0.73 | 0.98 | 0.92 | 0.82 | 0.69 |
| | $V_{MLE}$ | 0.98 | 0.92 | 0.82 | 0.69 | 0.98 | 0.92 | 0.82 | 0.69 |

process, which is

$$(1-\phi L)y_t = (1-\theta L)\epsilon_t,$$

where $\epsilon_t$ is i.i.d(0, $\sigma^2$). Let $\rho_k$ denote the $k$th order autocorrelation of $y_t$. We then have

$$\rho_1 = \frac{(1-\phi\theta)(\phi-\theta)}{1+\theta^2-2\phi\theta},$$

and

$$\rho_k = \rho_{k-1}\phi \quad \text{for} \quad k \geq 2.$$

Let $\lambda = [\ \phi \quad \theta\ ]'$ and $\hat{\lambda}_{MLE}$ denote the maximum likelihood estimator of ARMA(1,1) model. Then, we have $\sqrt{T}(\hat{\lambda}_{MLE} - \lambda_0) \rightarrow N(0, V_{MLE})$, where

$$V_{MLE} = \frac{1-\phi\theta}{(\phi-\theta)^2}\begin{bmatrix} (1-\phi^2)(1-\phi\theta) & (1-\phi^2)(1-\theta^2) \\ (1-\phi^2)(1-\theta^2) & (1-\theta^2)(1-\phi\theta) \end{bmatrix}.$$

Given a set of parameter values the matrix C is calculated from Bartlett's formula, while the matrix of partial derivatives, D, is analytically straightforward to calculate. The asymptotic variance of $\sqrt{T}(\hat{\lambda}_{MDE} - \lambda)$ for the ARMA(1,1) model are then the diagonal elements of the matrix $(D'C^{-1}D)^{-1}$, and are reported in Tables 4 and 5 , for four different values of $\phi$ and nine values of $\theta$ which give a total of thirty six points of the parameter space. Identification requires that at least 2 autocorrelations are used in MDE. We present the cases that first 2, 3, 5, 8, 10, 15 and 20 autocorrelations are used in MDE.

In general, as the number of autocorrelations used in the MDE increases, the asymptotic variance of the MDE decreases. The MDE appears to be asymptotically as efficient as MLE under normality when the number of autocorrelations is large enough.

The first portion of Table 4 presents results for ARMA(1,1) models with the AR coefficient being set to -0.5 and $\theta$ being varied from 0.1 to 0.9 with steps of 0.1. The second part of Table 4 reports cases whose AR coefficients are fixed at 0.8, while MA coefficients are set to 0.1, 0.2, 0.25, 0.3, 0.35, 0.4, 0.45, 0.5, and 0.6. Calculations reveal that if $\theta$ is less than 0.3, MDE using first 5 autocorrelations can be asymptotically as efficient as MLE under normality, while the asymptotic variance of MDE using first 10 autocorrelations is very close to that of MLE when $\theta$ lies between 0.4 and 0.6. Furthermore, this result seems not to be affected by the value of $\phi$.

Some further investigation is presented in Table 5, which shows the results for ARMA(1,1) models whose AR coefficients are fixed at 0.6 and 0.3. In Part I of Table 5, the MA coefficients are set to both positive and negative values, i.e., 0.2, 0.3, 0.4, 0.8, 0.9, -0.2, -0.4, -0.6, and -0.8. The results in Table 5 are very similar to those in Table 4. If the absolute value of $\theta$ of the ARMA(1,1) process is small, MDE using first 5 autocorrelations can be as efficient as MLE. In sum, when the absolute value of MA coefficient is less than 0.3, asymptotic variance of MDE using first 5 autocorrelations is very close to that of MLE under normality. If $|\theta| \in (0.3, 0.7)$, asymptotic variance of MDE using first 10 autocorrelations is very close to that of MLE. When $|\theta|$ is close to one, g should be higher than 20.

Table 4: Asymptotic Variance of $\sqrt{T}(\hat{\lambda}_{MDE} - \lambda)$ of ARMA(1,1) processes, $(1 - \phi L)y_t = (1 - \theta L)\epsilon_t$ with $\phi$=-0.5 and 0.8

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | $\phi$ | -0.5 | | | | | | | | |
| | $\theta$ | 0.10 | 0.20 | 0.30 | 0.40 | 0.50 | 0.60 | 0.70 | 0.80 | 0.90 |
| | | Asymptotic variance of $\sqrt{T}(\hat{\phi} - \phi)$ | | | | | | | | |
| g= | 2 | 2.36 | 2.02 | 1.83 | 1.72 | 1.65 | 1.61 | 1.58 | 1.57 | 1.56 |
| | 3 | 2.30 | 1.87 | 1.59 | 1.42 | 1.32 | 1.26 | 1.22 | 1.19 | 1.18 |
| | 5 | 2.30 | 1.85 | 1.55 | 1.34 | 1.19 | 1.09 | 1.03 | 1.00 | 0.98 |
| | 8 | 2.30 | 1.85 | 1.55 | 1.33 | 1.17 | 1.05 | 0.97 | 0.92 | 0.89 |
| | 10 | 2.30 | 1.85 | 1.55 | 1.33 | 1.17 | 1.05 | 0.96 | 0.90 | 0.86 |
| | 20 | 2.30 | 1.85 | 1.55 | 1.33 | 1.17 | 1.05 | 0.95 | 0.87 | 0.82 |
| | $V_{MLE}$ | 2.30 | 1.85 | 1.55 | 1.33 | 1.17 | 1.05 | 0.95 | 0.87 | 0.80 |
| | | Asymptotic variance of $\sqrt{T}(\hat{\theta} - \theta)$ | | | | | | | | |
| g= | 2 | 3.16 | 2.79 | 2.76 | 3.07 | 3.94 | 6.01 | 11.48 | 30.18 | 151.32 |
| | 3 | 3.03 | 2.40 | 2.01 | 1.81 | 1.85 | 2.26 | 3.57 | 8.25 | 38.61 |
| | 5 | 3.03 | 2.37 | 1.88 | 1.51 | 1.25 | 1.12 | 1.24 | 2.10 | 8.15 |
| | 8 | 3.03 | 2.37 | 1.88 | 1.49 | 1.17 | 0.91 | 0.75 | 0.81 | 2.18 |
| | 10 | 3.03 | 2.37 | 1.88 | 1.49 | 1.17 | 0.90 | 0.68 | 0.60 | 1.23 |
| | 20 | 3.03 | 2.37 | 1.88 | 1.49 | 1.17 | 0.89 | 0.65 | 0.42 | 0.32 |
| | $V_{MLE}$ | 3.03 | 2.37 | 1.88 | 1.49 | 1.17 | 0.89 | 0.65 | 0.42 | 0.20 |
| | $\phi$ | 0.8 | | | | | | | | |
| | $\theta$ | 0.10 | 0.20 | 0.25 | 0.30 | 0.35 | 0.40 | 0.45 | 0.50 | 0.60 |
| | | Asymptotic variance of $\sqrt{T}(\hat{\phi} - \phi)$ | | | | | | | | |
| g= | 2 | 0.63 | 0.78 | 0.90 | 1.09 | 1.39 | 1.88 | 2.72 | 4.25 | 13.95 |
| | 3 | 0.62 | 0.71 | 0.78 | 0.87 | 1.00 | 1.20 | 1.53 | 2.11 | 5.53 |
| | 5 | 0.62 | 0.71 | 0.76 | 0.83 | 0.92 | 1.05 | 1.23 | 1.51 | 2.95 |
| | 8 | 0.62 | 0.71 | 0.76 | 0.83 | 0.92 | 1.04 | 1.20 | 1.44 | 2.48 |
| | 10 | 0.62 | 0.71 | 0.76 | 0.83 | 0.92 | 1.04 | 1.20 | 1.44 | 2.44 |
| | 20 | 0.62 | 0.71 | 0.76 | 0.83 | 0.92 | 1.04 | 1.20 | 1.44 | 2.43 |
| | $V_{MLE}$ | 0.62 | 0.71 | 0.76 | 0.83 | 0.92 | 1.04 | 1.20 | 1.44 | 2.43 |
| | | Asymptotic variance of $\sqrt{T}(\hat{\theta} - \theta)$ | | | | | | | | |
| g= | 2 | 1.78 | 2.22 | 2.58 | 3.08 | 3.81 | 4.90 | 6.58 | 9.34 | 23.73 |
| | 3 | 1.71 | 1.91 | 2.05 | 2.24 | 2.52 | 2.94 | 3.57 | 4.61 | 9.84 |
| | 5 | 1.71 | 1.88 | 1.98 | 2.11 | 2.26 | 2.46 | 2.74 | 3.18 | 5.30 |
| | 8 | 1.71 | 1.88 | 1.98 | 2.10 | 2.25 | 2.43 | 2.67 | 3.01 | 4.42 |
| | 10 | 1.71 | 1.88 | 1.98 | 2.10 | 2.25 | 2.43 | 2.67 | 3.00 | 4.34 |
| | 20 | 1.71 | 1.88 | 1.98 | 2.10 | 2.25 | 2.43 | 2.67 | 3.00 | 4.33 |
| | $V_{MLE}$ | 1.71 | 1.88 | 1.98 | 2.10 | 2.25 | 2.43 | 2.67 | 3.00 | 4.33 |

Table 5: Asymptotic Variance of $\sqrt{T}(\hat{\lambda}_{MDE} - \lambda)$ of ARMA(1,1) processes, $(1 - \phi L)y_t = (1 - \theta L)\epsilon_t$ with $\phi$=0.6 and 0.4

| | $\phi$ | 0.6 | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | $\theta$ | 0.20 | 0.30 | 0.40 | 0.80 | 0.90 | -0.20 | -0.40 | -0.60 | -0.80 |
| | | Asymptotic variance of $\sqrt{T}(\hat{\phi} - \phi)$ | | | | | | | | |
| g= | 2 | 3.57 | 6.85 | 18.48 | 75.83 | 52.75 | 1.35 | 1.20 | 1.15 | 1.13 |
| | 3 | 3.13 | 5.07 | 11.11 | 27.97 | 19.00 | 1.26 | 1.04 | 0.95 | 0.91 |
| | 5 | 3.10 | 4.79 | 9.35 | 11.14 | 7.12 | 1.25 | 0.99 | 0.85 | 0.79 |
| | 8 | 3.10 | 4.78 | 9.24 | 6.35 | 3.65 | 1.25 | 0.98 | 0.83 | 0.74 |
| | 10 | 3.10 | 4.78 | 9.24 | 5.35 | 2.86 | 1.25 | 0.98 | 0.82 | 0.73 |
| | 20 | 3.10 | 4.78 | 9.24 | 4.36 | 1.75 | 1.25 | 0.98 | 0.82 | 0.72 |
| | $V_{MLE}$ | 3.10 | 4.78 | 9.24 | 4.33 | 1.50 | 1.25 | 0.98 | 0.82 | 0.72 |
| | | Asymptotic variance of $\sqrt{T}(\hat{\theta} - \theta)$ | | | | | | | | |
| g= | 2 | 5.47 | 9.87 | 23.88 | 98.68 | 201.91 | 2.22 | 2.68 | 5.65 | 29.75 |
| | 3 | 4.71 | 7.25 | 14.59 | 30.09 | 53.40 | 1.91 | 1.57 | 2.10 | 8.11 |
| | 5 | 4.65 | 6.81 | 12.27 | 9.14 | 12.08 | 1.88 | 1.31 | 1.03 | 2.05 |
| | 8 | 4.65 | 6.80 | 12.13 | 4.18 | 3.56 | 1.88 | 1.29 | 0.84 | 0.78 |
| | 10 | 4.65 | 6.80 | 12.13 | 3.28 | 2.12 | 1.88 | 1.29 | 0.83 | 0.58 |
| | 20 | 4.65 | 6.80 | 12.13 | 2.46 | 0.65 | 1.88 | 1.29 | 0.82 | 0.41 |
| | $V_{MLE}$ | 4.65 | 6.80 | 12.13 | 2.43 | 0.45 | 1.88 | 1.29 | 0.82 | 0.40 |

| | $\phi$ | 0.40 | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | $\theta$ | 0.10 | 0.20 | 0.60 | 0.70 | 0.80 | -0.20 | -0.40 | -0.60 | -0.80 |
| | | Asymptotic variance of $\sqrt{T}(\hat{\phi} - \phi)$ | | | | | | | | |
| g= | 2 | 8.93 | 20.88 | 50.14 | 30.39 | 23.56 | 3.02 | 2.40 | 2.18 | 2.10 |
| | 3 | 8.61 | 18.01 | 23.83 | 13.51 | 10.09 | 2.75 | 1.91 | 1.62 | 1.51 |
| | 5 | 8.60 | 17.78 | 14.36 | 7.09 | 4.86 | 2.72 | 1.78 | 1.36 | 1.21 |
| | 8 | 8.60 | 17.77 | 12.36 | 5.30 | 3.20 | 2.72 | 1.77 | 1.30 | 1.09 |
| | 10 | 8.60 | 17.77 | 12.17 | 5.00 | 2.83 | 2.72 | 1.77 | 1.29 | 1.05 |
| | 20 | 8.60 | 17.77 | 12.13 | 4.84 | 2.44 | 2.72 | 1.77 | 1.29 | 1.02 |
| | $V_{MLE}$ | 8.60 | 17.77 | 12.13 | 4.84 | 2.43 | 2.72 | 1.77 | 1.29 | 1.02 |
| | | Asymptotic variance of $\sqrt{T}(\hat{\theta} - \theta)$ | | | | | | | | |
| g= | 2 | 10.55 | 23.89 | 49.22 | 36.81 | 50.36 | 3.66 | 3.60 | 6.47 | 30.70 |
| | 3 | 10.15 | 20.59 | 20.70 | 12.84 | 14.80 | 3.15 | 2.14 | 2.46 | 8.43 |
| | 5 | 10.14 | 20.31 | 11.29 | 5.14 | 4.25 | 3.11 | 1.79 | 1.23 | 2.16 |
| | 8 | 10.14 | 20.31 | 9.44 | 3.34 | 1.85 | 3.11 | 1.77 | 1.01 | 0.84 |
| | 10 | 10.14 | 20.31 | 9.28 | 3.07 | 1.42 | 3.11 | 1.77 | 0.99 | 0.62 |
| | 20 | 10.14 | 20.31 | 9.24 | 2.94 | 1.05 | 3.11 | 1.77 | 0.98 | 0.44 |
| | $V_{MLE}$ | 10.14 | 20.31 | 9.24 | 2.94 | 1.04 | 3.11 | 1.77 | 0.98 | 0.44 |

## 6. MDE Applied to Higher Order ARMA(p,q) Processes

This section provides further investigation of the application of MDE to some higher order ARMA(p,q) processes. The results for the ARMA(2,1) process are presented in Table 6 and 7. The ARMA(2,1) process is

$$(1-\phi_1 L-\phi_2 L^2)y_t=(1-\theta L)\epsilon_t,$$

where $\epsilon_t$ is NID(0, $\sigma^2$). To calculate the asymptotic variance of MDE of ARMA(2,1) processes, the first step is to obtain explicit representations of autocorrelation functions in term of $\phi_1$, $\phi_2$, and $\theta$. The autocorrelation functions of ARMA(2,1) process are

$$\rho_1=\frac{\theta(1-\phi_2^2)-\phi_1(1-\theta\phi_1+\theta^2)}{(\phi_2-1)(1-\theta\phi_1+\theta^2)+\phi_1\theta(1+\phi_2)},$$

and

$$\rho_k=\phi_1\rho_{k-1}+\phi_2\rho_{k-2}\ \text{ for } k\geq 2.$$

Letting $\hat{\lambda}_{MLE}$ denote the maximum likelihood estimator of ARMA(2,1) model, $\sqrt{T}(\hat{\lambda}_{MLE}-\lambda)\rightarrow N(0,V_{MLE})$, where

$$V_{MLE}=\begin{bmatrix} \frac{1-\phi_2}{(1+\phi_2)[(1-\phi_2)^2-\phi_1^2]} & \frac{\phi_1}{(1+\phi_2)[(1-\phi_2)^2-\phi_1^2]} & -\frac{1}{1-\phi_1\theta-\phi_2\theta^2} \\ \frac{\phi_1}{(1+\phi_2)[(1-\phi_2)^2-\phi_1^2]} & \frac{1-\phi_2}{(1+\phi_2)[(1-\phi_2)^2-\phi_1^2]} & \frac{-\theta}{1-\phi_1\theta-\phi_2\theta^2} \\ -\frac{1}{1-\phi_1\theta-\phi_2\theta^2} & \frac{-\theta}{1-\phi_1\theta-\phi_2\theta^2} & \frac{1}{1-\theta^2} \end{bmatrix}^{-1},$$

and $\lambda'=[\phi_1\ \phi_2\ \theta]$. Table 6 presents the numerical calculation results of the asymptotic variance of the MDE and MLE for ARMA(2,1) process. The AR coefficients, $\phi_1$ and $\phi_2$, are equal to -1.00 and -0.16, respectively, whereas $\theta$ is varied from 0.1 to 0.90 by steps of 0.1. Alternatively, we can rewrite the ARMA(2,1) process as

$$(1 - \alpha_1 L)(1 - \alpha_2 L)y_t = (1 - \theta L)\epsilon_t.$$

Thus, the ARMA(2,1) process with $\phi_1 = -1.00$ and $\phi_2 = -0.16$ corresponds to $\alpha_1 = -0.2$ and $\alpha_2 = -0.8$. Clearly, identification requires at least 3 autocorrelations to be used in MDE. The asymptotic variance of MDE using first 3, 5, 8, 10, and 20 autocorrelation are reported. Table 7 presents the results for the cases that $\phi_1$ and $\phi_2$ are equal to 1.00 and -0.21, while $\theta$ is varied from -0.1 to -0.9 by steps of -0.1.

In sum, the results given in Table 6 and 7 indicate that the efficiency loss in MDE appears to diminish as g increases. In particular, the results of the MDE for ARMA(2,1) processes are very similar to those of ARMA(1,1) and MA(1) processes. When the absolute value of $\theta$ of the ARMA(2,1) process is small, MDE using first 5 autocorrelations is as efficient as MLE under normality. Furthermore, given that $\phi_1$ and $\phi_2$ are the same, the higher the value of $\theta$, the higher the number of autocorrelations are needed to guarantee that the asymptotic variance of the MDE is close to that of MLE under normality.

We also investigate the asymptotic variance of MDE for the ARMA(1,2) process. The ARMA(1,2) process is

$$(1 - \phi L)y_t = (1 - \theta_1 L - \theta_2 L^2)\epsilon_t,$$

where $\epsilon_t$ is i.i.d(0, $\sigma^2$). An alternative representation of the ARMA(1,2) process is

$$(1 - \phi L)y_t = (1 - \delta_1 L)(1 - \delta_2 L)\epsilon_t.$$

The autocorrelation function of the ARMA(1,2) process is

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

$$\rho_2 = \phi\rho_1 - \frac{\theta_2(\phi^2-1)}{\theta_1(\phi-\theta_1)+\theta_2(\phi^2-\theta_1\phi-\theta_2)+\phi[\theta_1+\theta_2(\phi-\theta_1)]-1},$$

and

$$\rho_k = \phi\rho_{k-1} \text{ for } k \geq 3.$$

We also compare the asymptotic variance of MDE to that of MLE. For the MLE of ARMA(1,2) processes, we have the following result of the asymptotic variance of $\sqrt{T}(\hat{\lambda}_{MLE}-\lambda)$, while a detail derivation is presented in the appendix.

$$V_{MLE} = \begin{bmatrix} \frac{1}{1-\phi^2} & \frac{-1}{1-\phi\theta_1-\phi^2\theta_2} & \frac{-\phi}{1-\phi\theta_1-\phi^2\theta_2} \\ \frac{-1}{1-\phi\theta_1-\phi^2\theta_2} & \frac{1-\theta_2}{(1+\theta_2)[(1-\theta_2)^2-\theta_1^2]} & \frac{\theta_1}{(1+\theta_2)[(1-\theta_2)^2-\theta_1^2]} \\ \frac{-\phi}{1-\phi\theta_1-\phi^2\theta_2} & \frac{\theta_1}{(1+\theta_2)[(1-\theta_2)^2-\theta_1^2]} & \frac{1-\theta_2}{(1+\theta_2)[(1-\theta_2)^2-\theta_1^2]} \end{bmatrix}^{-1}.$$

The calculation results of the asymptotic variance of MDE of MA(2) processes might give some implications for ARMA(1,2) processes. In particular, it is expected that given that $\phi$, $\delta_2$ and number of autocorrelations (g) used in MDE are fixed, the higher the value of $\delta_1$, the higher the asymptotic variance of MDE for ARMA(1,2) processes. Table 8 presents the ARMA(1,2) processes whose $\delta_2$ =-0.35 and $\phi$ =-0.60, while $\delta_1$ are equals to 0.10, 0.20, 0.30, 0.40, 0.50, 0.70, 0.80 and 0.90. Table 9 presents the results of the cases that $\delta_2$ =0.1 and $\phi$ =-0.60, while $\delta_1$ are equals to 0.20, 0.30, 0.40, 0.50, 0.60, 0.70, 0.80 and 0.90. The results in Table 8 and 9 reveal that the asymptotic variance of the MDE is very close to that of MLE under normality if g is large enough.

## 7. Concluding Remarks

This chapter investigates the asymptotic properties of the minimum distance estimator for a variety of ARMA processes. The results show that as the number of autocorrelations used in the MDE increases, the asymptotic variance of the MDE decreases. Numerical calculation results show that if the number of autocorrelations used in the MDE is large enough, the MDE appears to be asymptotically as efficient as MLE under normality. A formal proof of this result is provided for the MA(1) process. Interestingly, for the MA(1) and ARMA(1,1) models, if the absolute value of the moving average coefficient is not too large, the asymptotic variance of MDE based on first 3 to 5 autocorrelations is very close to that of MLE under normality. Furthermore, the higher the value of the moving average coefficient, the higher the number of autocorrelations is needed for the asymptotic variance of MDE to be close to that of MLE. The MDE is surprisingly efficient for some parts of the parameter space.

Table 6: Asymptotic Variance of $\sqrt{T}(\hat{\lambda}_{MDE} - \lambda)$ of ARMA(2,1) processes: $(1 - \phi_1 L - \phi_2 L^2)y_t = (1 - \theta L)\epsilon_t$ ; $(1 - \alpha_1 L)(1 - \alpha_2 L)y_t = (1 - \theta L)\epsilon_t$ with $\alpha_1$ =-0.20 and $\alpha_2$ = -0.80. ( $\phi_1$ =-1.00, and $\phi_2$ =-0.16 )

| | | $\theta$ | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | 0.10 | 0.20 | 0.30 | 0.40 | 0.50 | 0.60 | 0.70 | 0.80 | 0.90 |
| | | Asymptotic variance of $\sqrt{T}(\hat{\phi}_1 - \phi_1)$ | | | | | | | | |
| g= | 3 | 16.80 | 10.10 | 7.28 | 5.88 | 5.12 | 4.68 | 4.43 | 4.29 | 4.22 |
| | 5 | 16.19 | 8.81 | 5.54 | 3.88 | 2.98 | 2.49 | 2.22 | 2.07 | 2.00 |
| | 8 | 16.19 | 8.80 | 5.50 | 3.76 | 2.75 | 2.13 | 1.77 | 1.58 | 1.49 |
| | 10 | 16.19 | 8.80 | 5.50 | 3.76 | 2.74 | 2.10 | 1.70 | 1.47 | 1.37 |
| | 20 | 16.19 | 8.80 | 5.50 | 3.76 | 2.74 | 2.09 | 1.66 | 1.36 | 1.18 |
| | $V_{MLE}$ | 16.19 | 8.80 | 5.50 | 3.76 | 2.74 | 2.09 | 1.66 | 1.36 | 1.14 |
| | | Asymptotic variance of $\sqrt{T}(\hat{\phi}_2 - \phi_2)$ | | | | | | | | |
| g= | 3 | 13.38 | 8.44 | 6.31 | 5.24 | 4.64 | 4.30 | 4.09 | 3.98 | 3.92 |
| | 5 | 12.91 | 7.41 | 4.87 | 3.53 | 2.79 | 2.37 | 2.14 | 2.01 | 1.95 |
| | 8 | 12.91 | 7.40 | 4.83 | 3.43 | 2.58 | 2.05 | 1.73 | 1.55 | 1.47 |
| | 10 | 12.91 | 7.40 | 4.83 | 3.42 | 2.57 | 2.02 | 1.66 | 1.45 | 1.36 |
| | 20 | 12.91 | 7.40 | 4.83 | 3.42 | 2.57 | 2.01 | 1.62 | 1.35 | 1.18 |
| | $V_{MLE}$ | 12.91 | 7.40 | 4.83 | 3.42 | 2.57 | 2.01 | 1.62 | 1.34 | 1.13 |
| | | Asymptotic variance of $\sqrt{T}(\hat{\theta} - \theta)$ | | | | | | | | |
| g= | 3 | 17.15 | 10.27 | 7.56 | 6.59 | 6.82 | 8.61 | 14.02 | 32.88 | 154.38 |
| | 5 | 16.48 | 8.74 | 5.25 | 3.46 | 2.53 | 2.17 | 2.41 | 4.16 | 16.30 |
| | 8 | 16.48 | 8.73 | 5.20 | 3.30 | 2.16 | 1.47 | 1.12 | 1.20 | 3.27 |
| | 10 | 16.48 | 8.73 | 5.20 | 3.29 | 2.15 | 1.42 | 0.96 | 0.81 | 1.69 |
| | 20 | 16.48 | 8.73 | 5.20 | 3.29 | 2.15 | 1.40 | 0.89 | 0.52 | 0.37 |
| | $V_{MLE}$ | 16.48 | 8.73 | 5.20 | 3.29 | 2.15 | 1.40 | 0.89 | 0.51 | 0.22 |

Table 7: Asymptotic Variance of $\sqrt{T}(\hat{\lambda}_{MDE} - \lambda)$ of ARMA(2,1) processes: $(1 - \phi_1 L - \phi_2 L^2)y_t = (1 - \theta L)\epsilon_t$ ; $(1 - \alpha_1 L)(1 - \alpha_2 L)y_t = (1 - \theta L)\epsilon_t$ with $\alpha_1$ =0.30 and $\alpha_2$ = 0.70. ( $\phi_1$ =1.00, and $\phi_2$ =-0.21 )

| | | $\theta$ | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | -0.10 | -0.20 | -0.30 | -0.40 | -0.50 | -0.60 | -0.70 | -0.80 | -0.90 |
| | | Asymptotic variance of $\sqrt{T}(\hat{\phi_1} - \phi_1)$ | | | | | | | | |
| g= | 3 | 11.72 | 7.80 | 5.95 | 4.98 | 4.43 | 4.10 | 3.91 | 3.80 | 3.75 |
| | 5 | 11.31 | 6.84 | 4.59 | 3.35 | 2.65 | 2.25 | 2.03 | 1.91 | 1.85 |
| | 8 | 11.31 | 6.84 | 4.55 | 3.25 | 2.45 | 1.95 | 1.65 | 1.48 | 1.41 |
| | 10 | 11.31 | 6.84 | 4.55 | 3.25 | 2.44 | 1.92 | 1.58 | 1.39 | 1.30 |
| | 20 | 11.31 | 6.84 | 4.55 | 3.25 | 2.44 | 1.91 | 1.55 | 1.29 | 1.14 |
| | $V_{MLE}$ | 11.31 | 6.84 | 4.55 | 3.25 | 2.44 | 1.91 | 1.55 | 1.29 | 1.10 |
| | | Asymptotic variance of $\sqrt{T}(\hat{\phi_2} - \phi_2)$ | | | | | | | | |
| g= | 3 | 8.84 | 6.25 | 4.99 | 4.31 | 3.92 | 3.68 | 3.54 | 3.46 | 3.42 |
| | 5 | 8.54 | 5.52 | 3.91 | 2.99 | 2.44 | 2.12 | 1.94 | 1.84 | 1.79 |
| | 8 | 8.54 | 5.52 | 3.89 | 2.90 | 2.27 | 1.86 | 1.60 | 1.46 | 1.39 |
| | 10 | 8.54 | 5.52 | 3.89 | 2.90 | 2.27 | 1.83 | 1.54 | 1.37 | 1.29 |
| | 20 | 8.54 | 5.52 | 3.89 | 2.90 | 2.26 | 1.83 | 1.51 | 1.28 | 1.14 |
| | $V_{MLE}$ | 8.54 | 5.52 | 3.89 | 2.90 | 2.26 | 1.83 | 1.51 | 1.28 | 1.10 |
| | | Asymptotic variance of $\sqrt{T}(\hat{\theta} - \theta)$ | | | | | | | | |
| g= | 3 | 12.22 | 8.14 | 6.40 | 5.85 | 6.27 | 8.15 | 13.60 | 32.45 | 153.90 |
| | 5 | 11.74 | 6.93 | 4.44 | 3.06 | 2.32 | 2.04 | 2.32 | 4.09 | 16.23 |
| | 8 | 11.74 | 6.92 | 4.40 | 2.91 | 1.98 | 1.38 | 1.07 | 1.17 | 3.25 |
| | 10 | 11.74 | 6.92 | 4.40 | 2.91 | 1.96 | 1.33 | 0.92 | 0.79 | 1.67 |
| | 20 | 11.74 | 6.92 | 4.40 | 2.91 | 1.96 | 1.31 | 0.85 | 0.50 | 0.36 |
| | $V_{MLE}$ | 11.74 | 6.92 | 4.40 | 2.91 | 1.96 | 1.31 | 0.85 | 0.49 | 0.22 |

Table 8: Asymptotic Variance of $\sqrt{T}(\hat{\lambda}_{MDE} - \lambda)$ of ARMA(1,2) processes: $(1 - \phi L)y_t = (1 - \theta_1 L - \theta_2 L^2)\epsilon_t$ ; $(1 - \phi L)y_t = (1 - \delta_1 L)(1 - \delta_2 L)\epsilon_t$ with $\delta_2$ =-0.35 and $\phi$ =-0.60

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | $\theta_1=$ | -0.25 | -0.15 | -0.05 | 0.05 | 0.15 | 0.35 | 0.45 | 0.55 |
| | $\theta_2=$ | 0.03 | 0.07 | 0.10 | 0.14 | 0.17 | 0.24 | 0.28 | 0.32 |
| | | $\delta_1$ | | | | | | | |
| | | 0.10 | 0.20 | 0.30 | 0.40 | 0.50 | 0.70 | 0.80 | 0.90 |
| | | Asymptotic variance of $\sqrt{T}(\hat{\phi} - \phi)$ | | | | | | | |
| g= | 3 | 19.39 | 14.39 | 11.80 | 10.39 | 9.60 | 8.89 | 8.74 | 8.67 |
| | 5 | 14.83 | 12.61 | 11.02 | 9.85 | 9.00 | 8.07 | 7.87 | 7.77 |
| | 8 | 14.66 | 12.53 | 10.99 | 9.83 | 8.93 | 7.74 | 7.43 | 7.28 |
| | 10 | 14.65 | 12.53 | 10.99 | 9.83 | 8.93 | 7.67 | 7.30 | 7.11 |
| | 20 | 14.65 | 12.53 | 10.99 | 9.83 | 8.93 | 7.63 | 7.15 | 6.82 |
| | $V_{MLE}$ | 14.65 | 12.53 | 10.99 | 9.83 | 8.93 | 7.63 | 7.14 | 6.74 |
| | | Asymptotic variance of $\sqrt{T}(\hat{\theta}_1 - \theta_1)$ | | | | | | | |
| g= | 3 | 20.87 | 16.00 | 13.52 | 12.16 | 11.35 | 10.34 | 10.78 | 24.50 |
| | 5 | 16.05 | 14.01 | 12.58 | 11.54 | 10.80 | 9.82 | 9.41 | 10.67 |
| | 8 | 15.87 | 13.92 | 12.54 | 11.52 | 10.76 | 9.74 | 9.38 | 9.13 |
| | 10 | 15.87 | 13.92 | 12.54 | 11.52 | 10.75 | 9.71 | 9.36 | 9.05 |
| | 20 | 15.87 | 13.92 | 12.54 | 11.52 | 10.75 | 9.69 | 9.31 | 9.03 |
| | $V_{MLE}$ | 15.87 | 13.92 | 12.54 | 11.52 | 10.75 | 9.69 | 9.31 | 9.01 |
| | | Asymptotic variance of $\sqrt{T}(\hat{\theta}_2 - \theta_2)$ | | | | | | | |
| g= | 3 | 4.35 | 4.53 | 4.94 | 5.67 | 6.80 | 10.77 | 14.56 | 24.59 |
| | 5 | 3.46 | 4.07 | 4.66 | 5.27 | 5.97 | 8.24 | 10.28 | 14.49 |
| | 8 | 3.42 | 4.04 | 4.65 | 5.24 | 5.84 | 7.28 | 8.51 | 10.73 |
| | 10 | 3.42 | 4.04 | 4.65 | 5.24 | 5.83 | 7.10 | 8.06 | 9.79 |
| | 20 | 3.42 | 4.04 | 4.65 | 5.24 | 5.83 | 7.00 | 7.60 | 8.47 |
| | $V_{MLE}$ | 3.42 | 4.04 | 4.65 | 5.24 | 5.83 | 7.00 | 7.58 | 8.17 |

Table 9: Asymptotic Variance of $\sqrt{T}(\hat{\lambda}_{MDE} - \lambda)$ of ARMA(1,2) processes: $(1 - \phi L)y_t = (1 - \theta_1 L - \theta_2 L^2)\epsilon_t$ ; $(1 - \phi L)y_t = (1 - \delta_1 L)(1 - \delta_2 L)\epsilon_t$ with $\delta_2 = 0.10$ and $\phi$ =-0.60

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | $\theta_1=$ | 0.30 | 0.40 | 0.50 | 0.60 | 0.70 | 0.80 | 0.90 | 1.00 |
| | $\theta_2=$ | -0.02 | -0.03 | -0.04 | -0.05 | -0.06 | -0.07 | -0.08 | -0.09 |
| | | $\delta_1$ | | | | | | | |
| | | 0.20 | 0.30 | 0.40 | 0.50 | 0.60 | 0.70 | 0.80 | 0.90 |
| | | Asymptotic variance of $\sqrt{T}(\hat{\phi} - \phi)$ | | | | | | | |
| g= | 3 | 3.33 | 3.14 | 3.03 | 2.96 | 2.91 | 2.88 | 2.87 | 2.86 |
| | 5 | 2.88 | 2.54 | 2.31 | 2.16 | 2.06 | 2.00 | 1.97 | 1.95 |
| | 8 | 2.88 | 2.52 | 2.26 | 2.06 | 1.91 | 1.81 | 1.75 | 1.72 |
| | 10 | 2.88 | 2.52 | 2.26 | 2.05 | 1.89 | 1.77 | 1.70 | 1.66 |
| | 20 | 2.88 | 2.52 | 2.26 | 2.05 | 1.88 | 1.75 | 1.64 | 1.57 |
| | $V_{MLE}$ | 2.88 | 2.52 | 2.26 | 2.05 | 1.88 | 1.75 | 1.64 | 1.55 |
| | | Asymptotic variance of $\sqrt{T}(\hat{\theta}_1 - \theta_1)$ | | | | | | | |
| g= | 3 | 4.21 | 3.90 | 3.56 | 3.16 | 2.89 | 4.19 | 17.23 | 167.31 |
| | 5 | 3.83 | 3.47 | 3.21 | 3.01 | 2.83 | 2.64 | 2.78 | 10.35 |
| | 8 | 3.82 | 3.45 | 3.16 | 2.94 | 2.77 | 2.62 | 2.49 | 2.95 |
| | 10 | 3.82 | 3.45 | 3.16 | 2.93 | 2.76 | 2.61 | 2.49 | 2.51 |
| | 20 | 3.82 | 3.45 | 3.16 | 2.93 | 2.75 | 2.60 | 2.48 | 2.37 |
| | $V_{MLE}$ | 3.82 | 3.45 | 3.16 | 2.93 | 2.75 | 2.60 | 2.47 | 2.36 |
| | | Asymptotic variance of $\sqrt{T}(\hat{\theta}_2 - \theta_2)$ | | | | | | | |
| g= | 3 | 3.63 | 4.37 | 5.38 | 6.64 | 8.15 | 9.72 | 10.79 | 8.79 |
| | 5 | 2.64 | 2.63 | 2.70 | 2.89 | 3.23 | 3.71 | 4.27 | 4.52 |
| | 8 | 2.63 | 2.57 | 2.52 | 2.49 | 2.52 | 2.65 | 2.93 | 3.27 |
| | 10 | 2.63 | 2.57 | 2.52 | 2.47 | 2.45 | 2.48 | 2.65 | 2.95 |
| | 20 | 2.63 | 2.57 | 2.52 | 2.47 | 2.43 | 2.39 | 2.36 | 2.44 |
| | $V_{MLE}$ | 2.63 | 2.57 | 2.52 | 2.47 | 2.43 | 2.39 | 2.35 | 2.31 |

# APPENDIX

# APPENDIX

## A1 Asymptotic variance of sample autocorrelations of MA(2) process

For MA(2) process, we have $\rho_k = 0$ for $k >= 3$. Similarly, by using the Bartlett's formula we derive the following result for the asymptotic variance of $\sqrt{T}(\hat{\rho} - \rho)$ :

$$C_{MA(2)} = \begin{bmatrix} c_{11} & c_{12} & c_{13} & c_{14} & c_{15} & 0 & \cdots & \cdots & 0 \\ c_{21} & c_{22} & c_{23} & c_{24} & c_{25} & c_{15} & \ddots & & \vdots \\ c_{31} & c_{32} & c_{33} & c_{34} & c_{35} & c_{25} & \ddots & \ddots & \vdots \\ c_{41} & c_{42} & c_{43} & c_{33} & c_{45} & c_{35} & \ddots & \ddots & 0 \\ c_{51} & c_{52} & c_{53} & c_{54} & c_{33} & c_{45} & \ddots & \ddots & c_{15} \\ 0 & c_{51} & c_{52} & c_{53} & c_{54} & c_{33} & \ddots & \ddots & c_{25} \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & c_{35} \\ \vdots & & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & c_{45} \\ 0 & \cdots & \cdots & 0 & c_{51} & c_{52} & c_{53} & c_{54} & c_{33} \end{bmatrix},$$

where

$$\begin{aligned}
c_{11} &= (1 + \rho_2 - 2\rho_1^2) \\
c_{12} &= c_{21} = (1 + \rho_2 - 2\rho_1^2)(\rho_1 - 2\rho_1\rho_2) + (\rho_1 - 2\rho_1\rho_2)(1 - 2\rho_2^2) + \rho_1\rho_2 \\
c_{13} &= c_{31} = (1 + \rho_2 - 2\rho_1^2)\rho_2 + \rho_1(\rho_1 - 2\rho_1\rho_2) + \rho_2 \\
c_{14} &= c_{41} = (\rho_1 - 2\rho_1\rho_2)\rho_2 + \rho_1\rho_2 \\
c_{51} &= c_{51} = \rho_2^2 \\
c_{22} &= (\rho_1 - 2\rho_1\rho_2)^2 + (1 - 2\rho_2^2)^2 + \rho_2^2 + \rho_1^2 \\
c_{23} &= c_{32} = \rho_1(2 + 2\rho_2 - 4\rho_2^2) \\
c_{24} &= c_{42} = \rho_2(1 - 2\rho_2^2) + \rho_1^2) + \rho_2^2)
\end{aligned}$$

$$
\begin{aligned}
c_{25} &= c_{52} = 2\rho_1\rho_2 \\
c_{33} &= 1 + 2\rho_1^2 + 2\rho_2^2 \\
c_{34} &= c_{43} = 2\rho_1\rho_2 + 2\rho_1 \\
c_{35} &= c_{53} = \rho_1^2 + 2\rho_2 \\
c_{44} &= c_{33} = 1 + 2\rho_1^2 + 2\rho_2^2 \\
c_{45} &= c_{54} = 2\rho_1(1 + \rho_2) \\
c_{55} &= c_{33} = 1 + 2\rho_1^2 + 2\rho_2^2.
\end{aligned}
$$

Recall that the autocorrelation functions of the MA(2) process are $\rho_1 = -\theta_1(1 - \theta_2)/(1 + \theta_1^2 + \theta_2^2)$, $\rho_2 = -\theta_2/(1 + \theta_1^2 + \theta_2^2)$, and $\rho_k = 0$ for $k \geq 3$. Hence,

$$
D = \begin{bmatrix}
\frac{(\theta_2-1)(1-\theta_1^2+\theta_2^2)}{(1+\theta_1^2+\theta_2^2)^2} & \frac{\theta_1(1+\theta_1^2+2\theta_2-\theta_2^2)}{(1+\theta_1^2+\theta_2^2)^2} \\
\frac{2\theta_1\theta_2}{(1+\theta_1^2+\theta_2^2)^2} & \frac{-1-\theta_1^2+\theta_2^2}{(1+\theta_1^2+\theta_2^2)^2} \\
0 & 0 \\
\vdots & \vdots \\
0 & 0
\end{bmatrix}.
$$

## A2 Asymptotic variance of MLE of ARMA(2,1) processes

To derive the asymptotic variance of MLE for ARMA(2,1) process, we apply a result presented by Brockwell and Davis (1991, p.258). The ARMA(p,q) process can be written as

$$
(1 - \phi_1 L - \cdots \phi_p L^p) y_t = (1 - \theta_1 L - \cdots - \theta_q L^q)\epsilon_t,
$$

where $\epsilon_t$ is i.i.d. $(0, \sigma^2)$. Let $\lambda$ denote the vector of parameters of the ARMA(p,q) model, i.e., $\lambda' = [\phi_1 \cdots \phi_p \ \theta_1 \cdots \theta_q]$, and $\hat{\lambda}_{MLE}$ denote the maximum likelihood estimator of an ARMA(p,q) model. Brockwell and Davis (1991, p.258) show that $\sqrt{T}(\hat{\lambda}_{MLE} - \lambda) \rightarrow N(0, V_{MLE})$ with

$$V_{MLE} = \sigma^2 \begin{bmatrix} EU_tU_t' & EU_tV_t' \\ EV_tU_t' & EV_tV_t' \end{bmatrix}^{-1},$$

where $U_t = (u_t \cdots u_{t+1-p})'$, $V_t = (v_t \cdots v_{t+1-q})'$ and $u_t, v_t$ are AR processes defined by

$$\begin{aligned} (1 - \phi_1 L - \cdots \phi_p L^p) u_t &= \epsilon_t, \quad \text{and} \\ (1 - \theta_1 L - \cdots - \theta_q L^q) v_t &= -\epsilon_t. \end{aligned}$$

For the case of ARMA(2,1) process, we have $U_t = (u_t \ u_{t-1})'$ and $V_t = v_t$. Denote the autocovariance of $u_t$ at lag k as $\gamma_k^u$ and the autocovariance of $v_t$ at lag k as $\gamma_k^v$. Hence, we have

$$V_{MLE} = \sigma^2 \begin{bmatrix} \gamma_0^u & \gamma_1^u & E(u_t v_t) \\ \gamma_1^u & \gamma_0^u & E(u_{t-1} v_t) \\ E(u_t v_t) & E(u_{t-1} v_t) & \gamma_0^v \end{bmatrix}^{-1}. \tag{17}$$

Since $u_t$ follows an AR(2) process and $v_t$ follows an AR(1) process, we have the following results of the autocovariance functions of $u_t$ and $v_t$.

$$\begin{aligned} \gamma_0^u &= \frac{1 - \phi_2}{(1 + \phi_2)[(1 - \phi_2)^2 - \phi_1^2]} \sigma^2, \\ \gamma_1^u &= \frac{\phi_1}{(1 + \phi_2)[(1 - \phi_2)^2 - \phi_1^2]} \sigma^2, \quad \text{and} \\ \gamma_0^v &= \frac{1}{1 - \theta^2} \sigma^2. \end{aligned}$$

In order to derive results for $E(u_t v_t)$ and $E(u_{t-1} v_t)$, we assume that $(1 - \phi_1 L - \phi_2 L^2) = 0$ has two different roots and $(1 - \phi_1 L - \phi_2 L^2)$ can be factorized as $(1 -$

$\delta_1 L)(1-\delta_2 L)$, where $\phi_1 = \delta_1 + \delta_2$ and $\phi_2 = -\delta_1\delta_2$. Since

$$(1-\phi_1 L - \phi_2 L^2)^{-1} = \frac{1}{\delta_1 - \delta_2}\left(\frac{\delta_1}{1-\delta_1 L} - \frac{\delta_2}{1-\delta_2 L}\right),$$

we have the following results for $E(u_t v_t)$:

$$\begin{aligned}
E(u_t v_t) &= E\left[(1-\phi_1 L - \phi_2 L^2)^{-1}\epsilon_t (1-\theta L)^{-1}(-\epsilon_t)\right] \\
&= -E\left[\frac{\delta_1}{\delta_1-\delta_2}(1+\delta_1 L + \delta_1^2 L^2 + \cdots) - \frac{\delta_2}{\delta_1-\delta_2}\ (1+\delta_2 L + \delta_2^2 L^2 + \cdots)\right]\epsilon_t \times \\
&\quad (1+\theta L + \theta^2 L^2 + \cdots)\epsilon_t \\
&= \frac{-\sigma^2}{\delta_1-\delta_2}\left[\frac{\delta_1}{1-\delta_1\theta} - \frac{\delta_2}{1-\delta_2\theta}\right] \\
&= \frac{-1}{1-(\delta_1+\delta_2)\theta + \delta_1\delta_2\theta^2}\,\sigma^2 \\
&= \frac{-1}{1-\phi_1\theta - \phi_2\theta^2}\,\sigma^2.
\end{aligned}$$

Similarly,

$$\begin{aligned}
E(u_{t-1} v_t) &= E\left[(1-\phi_1 L - \phi_2 L^2)^{-1}\epsilon_{t-1} (1-\theta L)^{-1}(-\epsilon_t)\right] \\
&= -E\left[\frac{\delta_1}{\delta_1-\delta_2}(L+\delta_1 L^2 + \cdots) - \frac{\delta_2}{\delta_1-\delta_2}(L + \delta_2 L^2 + \cdots)\right]\epsilon_t \times \\
&\quad (1+\theta L + \theta^2 L^2 + \cdots)\epsilon_t \\
&= \frac{-\theta}{\delta_1-\delta_2}\left[\frac{\delta_1}{1-\delta_1\theta} - \frac{\delta_2}{1-\delta_2\theta}\right]\sigma^2 \\
&= \frac{-\theta}{1-(\delta_1+\delta_2)\theta + \delta_1\delta_2\theta^2}\,\sigma^2 \\
&= \frac{-\theta}{1-\phi_1\theta - \phi_2\theta^2}\,\sigma^2.
\end{aligned}$$

Substituting the above results into (17) yields

$$V_{MLE} = \begin{bmatrix} \frac{1-\phi_2}{(1+\phi_2)[(1-\phi_2)^2-\phi_1^2]} & \frac{\phi_1}{(1+\phi_2)[(1-\phi_2)^2-\phi_1^2]} & -\frac{1}{1-\phi_1\theta-\phi_2\theta^2} \\ \frac{\phi_1}{(1+\phi_2)[(1-\phi_2)^2-\phi_1^2]} & \frac{1-\phi_2}{(1+\phi_2)[(1-\phi_2)^2-\phi_1^2]} & \frac{-\theta}{1-\phi_1\theta-\phi_2\theta^2} \\ -\frac{1}{1-\phi_1\theta-\phi_2\theta^2} & \frac{-\theta}{1-\phi_1\theta-\phi_2\theta^2} & \frac{1}{1-\theta^2} \end{bmatrix}^{-1}.$$

**Asymptotic variance of MLE of ARMA(1,2) processes**

To obtain the asymptotic variance of MLE of ARMA(1,2) processes, we can apply similar method we use above. Notice that for the ARMA(1,2) processes, $u_t$ and $v_t$ are defined as $(1-\phi L)u_t = \epsilon_t$, and $(1-\theta_1 L - \theta_2 L^2)v_t = -\epsilon_t$. For the ARMA(1,2) process, we have $U_t = u_t$ and $V_t = (v_t \; v_{t-1})'$. The asymptotic variance of MLE is

$$V_{MLE} = \sigma^2 \begin{bmatrix} \gamma_0^u & E(u_t v_t) & E(u_t v_{t-1}) \\ E(u_t v_t) & \gamma_0^v & \gamma_1^v \\ E(u_t v_{t-1} & \gamma_1^v & \gamma_0^v \end{bmatrix}^{-1}.$$

Since $u_t$ is a AR(1) process and $v_t$ is a AR(2) process, we have

$$\begin{aligned} \gamma_0^u &= \frac{1}{1-\phi^2}\,\sigma^2 \\ \gamma_0^v &= \frac{1-\theta_2}{(1+\theta_2)[(1-\theta_2)^2-\theta_1^2]}\,\sigma^2 \\ \gamma_1^v &= \frac{\theta_1}{(1+\theta_2)[(1-\theta_2)^2-\theta_1^2]}\,\sigma^2 \end{aligned}$$

By using the same method in previous appendix, we obtain $E(u_t v_t) = \frac{-1}{1-\phi\theta_1-\phi^2\theta_2}\sigma^2$ and $E(u_{t-1}v_t)\frac{-\phi}{1-\phi\theta_1-\phi^2\theta_2}\sigma^2$. Thus,

$$V_{MLE} = \begin{bmatrix} \frac{1}{1-\phi^2} & \frac{-1}{1-\phi\theta_1-\phi^2\theta_2} & \frac{-\phi}{1-\phi\theta_1-\phi^2\theta_2} \\ \frac{-1}{1-\phi\theta_1-\phi^2\theta_2} & \frac{1-\theta_2}{(1+\theta_2)[(1-\theta_2)^2-\theta_1^2]} & \frac{\theta_1}{(1+\theta_2)[(1-\theta_2)^2-\theta_1^2]} \\ \frac{-\phi}{1-\phi\theta_1-\phi^2\theta_2} & \frac{\theta_1}{(1+\theta_2)[(1-\theta_2)^2-\theta_1^2]} & \frac{1-\theta_2}{(1+\theta_2)[(1-\theta_2)^2-\theta_1^2]} \end{bmatrix}^{-1}.$$

# Chapter 3

# MDE for Seasonal ARMA Processes

## 1. Introduction

The previous chapter discussed the asymptotic properties of the MDE for a variety of ARMA models. Many economic time series exhibit periodic behavior. For example, monthly observations that are 12 periods apart might behave very similarly. As noted by Hylleberg (1992), the seasonality observed in the economic data might be caused by changes of weather, the calendar, and the production and consumption decision made by economic agents. In this chapter we discuss the asymptotic properties of MDE of seasonal ARMA models. Following the notation of Box and Jenkins (1976), the general multiplicative model is

$$\phi_p(L)\Phi_P(L^s)\nabla^d\nabla_s^D y_t = \theta_q(L)\Theta_Q(L^s)\epsilon_t$$

where $\epsilon_t$ is i.i.d. $(0,\sigma^2)$, $\phi_p(L) = 1 - \phi_1 L - \cdots - \phi_p L^p$, $\Phi_P(L^s) = 1 - \Phi_1 L^s - \cdots - \Phi_P L^{Ps}$, $\nabla = 1 - L$, $\nabla_s = 1 - L^s$, $\theta_q(L) = 1 - \theta_1 L - \cdots - \theta_q L^q$, and, $\Theta_Q(L^s) = 1 - \Theta_1 L^s - \cdots - \Theta_Q L^{Qs}$. As noted by Box and Jenkins (1976), the general multiplicative model is said to be order $(p,d,q) \times (P,D,Q)_s$. For most applications, $s$ is equal to either 4 or 12. The multiplicative model is an appropriate model for describing seasonal pattern observed in many data series. An example of the application of the general multiplicative model is the "airline model" provided by Box and Jenkins (1976). International airline passengers data that are 12 months apart behave very similarly. Box and Jenkins modeled the differenced and seasonal differenced monthly data of international airline passengers by the MA(1)-Seasonal $MA(1)_{12}$ model. Similarly, Hillmer and Tiao (1982) fit a MA(1)-$SMA(1)_s$ model to the regular and seasonal differenced monthly data of US unemployment males aged 16 to 19 from January 1965 to August 1979. This airline model has been applied to model many economic time series. See for example Abraham and Ledolter (1983): chap. 6; Granger and Newbold (1986): chap. 3; and Franses (1996): chap. 3.

This chapter presents the properties of MDE of MA(1)-$SMA(1)_s$ processes. The MA(1)-Seasonal $MA(1)_s$ process is

$$y_t = (1 - \theta L)(1 - \Theta L^s)\epsilon_t, \tag{18}$$

where $\epsilon_t$ is i.i.d. $(0, \sigma^2)$. The autocorrelation function for the seasonal MA(1)-$SMA(1)_s$ process is

$$\begin{aligned} \rho_1 &= -\theta/(1+\theta^2), \\ \rho_2 &= \rho_3 = \cdots = \rho_{s-2} = 0, \\ \rho_{s-1} &= \theta\Theta/(1+\theta^2)(1+\Theta^2), \\ \rho_s &= -\Theta/(1+\Theta^2), \\ \rho_{s+1} &= \theta\Theta/(1+\theta^2)(1+\Theta^2), \end{aligned}$$

and

$$\rho_k = 0 \quad \text{for } k \geq s+2.$$

The first order autocorrelation function of MA(1)-SMA(1)$_s$ process is the same as simple MA(1) process and is not affected by the presence of the seasonal MA factor.

Also we have the following result of the maximum likelihood estimator of MA(1)-SMA(1)$_s$ processes. Similar to the notation defined in chapter 2, we denote $\hat{\lambda}_{MLE}$ as the ML estimator, where $\lambda' = [\theta \;\; \Theta]$. Without loss of generality, we assume that $\sigma^2$ is known. The limiting distribution of MLE is $\sqrt{T}(\hat{\lambda}_{MLE} - \lambda) \rightarrow N(0, V_{MLE})$ , where

$$V_{MLE} = \begin{bmatrix} (1-\theta^2)^{-1} & \theta^{s-1}(1-\theta^s\Theta)^{-1} \\ \theta^{s-1}(1-\theta^s\Theta)^{-1} & (1-\Theta^2)^{-1} \end{bmatrix}^{-1}, \tag{19}$$

or,

$$V_{MLE} = \frac{1}{(1-\theta^s\Theta)^2 - \theta^{2s-2}} \begin{bmatrix} (1-\theta^2)(1-\theta^s\Theta)^2 & -\theta^{s-1}(1-\theta^2)(1-\Theta^2) \\ -\theta^{s-1}(1-\theta^2)(1-\Theta^2) & (1-\Theta^2)(1-\theta^s\Theta)^2 \end{bmatrix}. \tag{20}$$

If $\theta$ is small enough or $s$ is large, the asymptotic variance of $\sqrt{T}(\hat{\theta}_{MLE} - \theta)$ will be very close to $1 - \theta^2$ and $\sqrt{T}(\hat{\Theta}_{MLE} - \Theta)$ will be very close to $1 - \Theta^2$.

## 2. Asymptotic Variance of MDE for MA(1)-SMA(1)$_4$ Processes

A seasonal MA(1)-SMA(1)$_4$ process is

$$y_t = (1 - \theta L)(1 - \Theta L^4)\epsilon_t, \tag{21}$$

Given that $s$ equals to 4, we then have the following results for the autocorrelation function of the seasonal MA(1)-SMA(1)$_4$ process: $\rho_1 = -\theta/(1+\theta^2)$, $\rho_2 = 0$, $\rho_3 = \theta\Theta/(1+\theta^2)(1+\Theta^2)$, $\rho_4 = -\Theta/(1+\Theta^2)$, $\rho_5 = \theta\Theta/(1+\theta^2)(1+\Theta^2)$, and $\rho_k =$

0 for $k \geq 6$. The second order autocorrelation function of MA(1)-SMA$(1)_4$ process equals to zero, while its first order autocorrelation function is not affected by the seasonal MA coefficient. Similarly, the autocorrelation at lag 4 is not affected by the MA coefficient.

Given a set of parameter values, the asymptotic variance of the MDE for MA(1)-SMA$(1)_4$ process can be numerically calculated by the same method described in chapter 2. Table 10 presents the asymptotic variance of MDE for some MA(1)-SMA$(1)_4$ models. The nonseasonal moving average coefficient is fixed at 0.15 for all cases, while the seasonal MA coefficient ($\Theta$) is varied from 0.1 to 0.9 by steps of 0.1. In each case, the asymptotic variance of MDE using the first 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 16, 24, and 32 autocorrelations are reported. Given that $\theta$ are all equal to 0.15, we find that when the seasonal MA coefficient is small, e.g., $\Theta = 0.1$ and 0.2, the MDE using first 8 autocorrelations can be asymptotically as efficient as MLE under normality. As the number of autocorrelations used in MDE increases, the asymptotic variance of $\sqrt{T}(\hat{\Theta} - \Theta)$ and $\sqrt{T}(\hat{\theta} - \theta)$ decrease. Hence, if the number of autocorrelations is large enough, MDE can be as efficient as MLE. Given that $\theta = 0.15$, if the absolute value of the seasonal MA coefficient of MA(1)-SMA$(1)_4$ process is less than 0.2, MDE using first 8 autocorrelations can be as efficient as MLE.

We also investigate the asymptotic variance of MDE for MA(1)-SMA$(1)_4$ process for the cases of negative $\Theta$, e.g., $\Theta = -0.1, -0.2, \cdots, -0.9$. If $\theta$ is not very large and $|\Theta|$ are the same, the asymptotic variances of MDEs are almost the same. In general, the results of negative $\Theta$ is very similar to those of positive $\Theta$.

Obviously, identification requires at least two autocorrelations be used in MDE. A very simple estimator is the MDE based on the autocorrelation at lag 1 and 3. However, if only the first 2 autocorrelations are used in the MDE, we are not able to identify the model. When only first 3 autocorrelations are used in MDE, asymptotic variance of $\sqrt{T}(\hat{\Theta}_{MDE} - \Theta)$ is very large comparing with the asymptotic variance of MDE using the first 4 autocorrelations. Consider the example that $\theta = 0.15$ and $\Theta = 0.6$ given in Table 10, the asymptotic variance of $\sqrt{T}(\hat{\Theta}_{MDE} - \Theta)$ of MDE using the first 3 autocorrelations (g=3) is 321.77, while that of MDE using the first 4 autocorrelations (g=4) is 4.88. For other cases we also find that the asymptotic variance of $\sqrt{T}(\hat{\Theta}_{MDE} - \Theta)$ reduced a lot when g is increased from 3 to 4. This result suggests that MDE using $\rho_1$ and $\rho_3$ is not an appropriate estimator for MA(1)-SMA$(1)_4$ models. Furthermore, this result also implies that the autocorrelations at lag 4, 8, and 12 are very important for MDE estimation in MA(1)-SMA$(1)_4$ models when $\Theta$ is large. For the case that $\theta = 0.15$ and $\Theta = 0.8$, asymptotic variances of MDEs using first 7 autocorrelations and first 8 autocorrelations are 27.42 and 7.87, while asymptotic variance of MDE using first 9 ACF is 7.71. A similar result is found when we compared the asymptotic variances of MDEs using first 11 and first 12 autocorrelations. This result demonstrates that the 4th, 8th, and 12th order autocorrelations are important moments for MDE estimation in MA(1)-SMA$(1)_4$ models. Besides, neglecting the autocorrelations at lag of multiples of 4 can cause a large efficiency loss. This is particularly important to the asymptotic variance of the MDE estimator of seasonal MA parameter.

Given that MA coefficients are the same, it is found that the higher the seasonal MA parameter, the higher the number of autocorrelations is needed to guarantee that MDE is efficient. For the cases of large seasonal MA parameters, we find that the number of autocorrelations should be higher than 32.

Table 11 reports results of the MA(1)-SMA$(1)_4$ models whose moving average coefficients all equal to 0.35, while Table 12 and 13 present asymptotic variance of MDE for the MA(1)-SMA$(1)_4$ models whose moving average coefficients equal to 0.55 and -0.25, respectively. The seasonal MA coefficient ($\Theta$) is varied from 0.1 to 0.9 by steps of 0.1. The asymptotic variance of MDE using first 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 16, 24, and 32 autocorrelations are reported. Given that $\theta = 0.35$ and $\Theta = 0.20$, the asymptotic variance of $\sqrt{T}(\hat{\Theta} - \Theta)$ of MDE using first 12 autocorrelations is very close to that of MLE, whereas the asymptotic variance of $\sqrt{T}(\hat{\theta} - \theta)$ of MDE using only first 8 autocorrelations can be as efficient as MLE.

The results in Table 12 show that given that $\theta = 0.55$, if $\Theta \leq 0.3$, the asymptotic variance of $\sqrt{T}(\hat{\theta} - \theta)$ for MDE using first 12 autocorrelations is very close to that of MLE. All the cases reported in Table 13 have their MA coefficients equal to -0.25. If both $|\theta|$ and $|\Theta|$ are small, the asymptotic variance of MDE using first 8 autocorrelations is very close to that of MLE.

In general, for the MA(1)-SMA$(1)_4$ process the efficiency loss in MDE appears to diminish as the number of autocorrelations used in MDE increases. The MDE using a small number of autocorrelations is surprisingly efficient for a subset of the parameter space. Calculations also reveal that the 4th, 8th, and 12th order autocorrelations are important moments of MDE in estimating MA(1)-SMA$(1)_4$ models.

### 3. Asymptotic Variance of MDE for MA(1)-SMA$(1)_{12}$ Processes

In this section, we present the results for the asymptotic variance of MDE of MA(1)-Seasonal MA$(1)_{12}$ process. The MA(1)-Seasonal MA$(1)_{12}$ process is

$$y_t = (1 - \theta L)(1 - \Theta L^{12})\epsilon_t, \tag{22}$$

where $\epsilon_t$ is i.i.d. $(0\ ,\sigma^2)$. For the seasonal MA(1)-SMA$(1)_{12}$ process we have $\rho_1 = -\theta/(1+\theta^2)$, $\rho_{11} = \theta\Theta/(1+\theta^2)(1+\Theta^2)$, $\rho_{12} = -\Theta/(1+\Theta^2)$, $\rho_{13} = \theta\Theta/(1+\theta^2)(1+\Theta^2)$, and $\rho_k = 0$ for $k = 2, 3, \cdots 10$, and $k \geq 14$. The first order autocorrelation function of MA(1)-SMA$(1)_{12}$ process is the same as that of MA(1) process and is not affected by the presence of the seasonal MA factor.

Equation (19) provides a general result of the asymptotic variance of MLE of MA(1)-SMA$(1)_s$ process. Setting $s = 12$ yields the following result of the maximum likelihood estimator of seasonal MA(1)-SMA$(1)_{12}$ processes: $\sqrt{T}(\hat{\lambda}_{MLE} - \lambda) \rightarrow N(0, V_{MLE})$, where

$$V_{MLE} = \begin{bmatrix} (1-\theta^2)^{-1} & \theta^{11}(1-\theta^{12}\Theta)^{-1} \\ \theta^{11}(1-\theta^{12}\Theta)^{-1} & (1-\Theta^2)^{-1} \end{bmatrix}^{-1},$$

and $\lambda' = [\theta \quad \Theta]$. Since the $s$ equals 12 in this case, the off-diagonal elements of the asymptotic variance matrix will be very close to zero given that $|\theta|$ is not close to unity.

It is interesting to investigate the asymptotic properties of MDE of MA(1)-SMA$(1)_{12}$ process. The asymptotic variance of $\sqrt{T}(\hat{\theta} - \theta)$ and $\sqrt{T}(\hat{\Theta} - \Theta)$ of MDE of MA(1)-

Table 10: Asymptotic Variance of $\sqrt{T}(\hat{\theta}-\theta)$ and $\sqrt{T}(\hat{\Theta}-\Theta)$ of MA(1)-Seasonal MA(1)$_4$ models, $y_t = (1-\theta L)(1-\Theta L^4)\epsilon_t$, with $\theta = 0.15$

| | | Θ | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | 0.10 | 0.20 | 0.30 | 0.40 | 0.50 | 0.60 | 0.70 | 0.80 | 0.90 |
| | | Asymptotic variance of $\sqrt{T}(\hat{\theta}-\theta)$ | | | | | | | | |
| g = | 3 | 1.01 | 1.08 | 1.17 | 1.26 | 1.35 | 1.42 | 1.48 | 1.51 | 1.53 |
| | 4 | 0.99 | 1.02 | 1.08 | 1.15 | 1.23 | 1.30 | 1.36 | 1.40 | 1.42 |
| | 5 | 0.98 | 0.98 | 1.00 | 1.03 | 1.08 | 1.13 | 1.19 | 1.23 | 1.25 |
| | 6 | 0.98 | 0.98 | 1.00 | 1.02 | 1.07 | 1.12 | 1.18 | 1.22 | 1.24 |
| | 7 | 0.98 | 0.98 | 0.99 | 1.00 | 1.04 | 1.08 | 1.13 | 1.18 | 1.20 |
| | 8 | 0.98 | 0.98 | 0.99 | 1.00 | 1.04 | 1.08 | 1.13 | 1.18 | 1.20 |
| | 9 | 0.98 | 0.98 | 0.98 | 0.99 | 1.00 | 1.04 | 1.08 | 1.12 | 1.15 |
| | 11 | 0.98 | 0.98 | 0.98 | 0.98 | 0.99 | 1.02 | 1.06 | 1.10 | 1.13 |
| | 12 | 0.98 | 0.98 | 0.98 | 0.98 | 0.99 | 1.02 | 1.06 | 1.10 | 1.13 |
| | 16 | 0.98 | 0.98 | 0.98 | 0.98 | 0.98 | 0.99 | 1.02 | 1.06 | 1.09 |
| | 24 | 0.98 | 0.98 | 0.98 | 0.98 | 0.98 | 0.98 | 0.99 | 1.01 | 1.05 |
| | 32 | 0.98 | 0.98 | 0.98 | 0.98 | 0.98 | 0.98 | 0.98 | 0.99 | 1.02 |
| | $V_{MLE}$ | 0.98 | 0.98 | 0.98 | 0.98 | 0.98 | 0.98 | 0.98 | 0.98 | 0.98 |
| | | Asymptotic variance of $\sqrt{T}(\hat{\Theta}-\Theta)$ | | | | | | | | |
| g= | 3 | 46.91 | 54.57 | 70.75 | 103.02 | 169.90 | 321.77 | 723.84 | 2123.5 | 11293 |
| | 4 | 1.05 | 1.16 | 1.39 | 1.84 | 2.78 | 4.88 | 10.40 | 29.50 | 154.13 |
| | 5 | 1.03 | 1.14 | 1.36 | 1.80 | 2.70 | 4.74 | 10.10 | 28.62 | 149.54 |
| | 6 | 1.03 | 1.14 | 1.36 | 1.80 | 2.70 | 4.74 | 10.10 | 28.61 | 149.48 |
| | 7 | 1.03 | 1.13 | 1.35 | 1.77 | 2.64 | 4.59 | 9.72 | 27.42 | 142.88 |
| | 8 | 0.99 | 0.98 | 0.98 | 1.04 | 1.23 | 1.73 | 3.10 | 7.87 | 38.80 |
| | 9 | 0.99 | 0.97 | 0.97 | 1.03 | 1.22 | 1.71 | 3.05 | 7.71 | 38.01 |
| | 11 | 0.99 | 0.97 | 0.97 | 1.03 | 1.21 | 1.68 | 2.99 | 7.54 | 37.08 |
| | 12 | 0.99 | 0.96 | 0.92 | 0.89 | 0.90 | 1.05 | 1.56 | 3.44 | 15.77 |
| | 16 | 0.99 | 0.96 | 0.91 | 0.85 | 0.80 | 0.82 | 1.02 | 1.93 | 8.07 |
| | 24 | 0.99 | 0.96 | 0.91 | 0.84 | 0.76 | 0.68 | 0.66 | 0.91 | 3.07 |
| | 32 | 0.99 | 0.96 | 0.91 | 0.84 | 0.75 | 0.65 | 0.56 | 0.60 | 1.55 |
| | $V_{MLE}$ | 0.99 | 0.96 | 0.91 | 0.84 | 0.75 | 0.64 | 0.51 | 0.36 | 0.19 |

Note: g=3 indicates that $\rho = [\rho_1\ \rho_2\ \rho_3]'$. Hence, g=3 represents that first 3 autocorrelations are used in MDE, while g=4 represents that first 4 autocorrelations are used in MDE, and etc. $V_{MLE}$ is the asymptotic variance of $\sqrt{T}(\hat{\theta}_{MLE}-\theta)$.

Table 11: Asymptotic Variance of $\sqrt{T}(\hat{\theta}-\theta)$ and $\sqrt{T}(\hat{\Theta}-\Theta)$ of MA(1)-Seasonal $MA(1)_4$ processes, $y_t = (1-\theta L)(1-\Theta L^4)\epsilon_t : \theta = 0.35$

| | | $\Theta$ | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | 0.10 | 0.20 | 0.30 | 0.40 | 0.50 | 0.60 | 0.70 | 0.80 | 0.90 |
| | | Asymptotic variance of $\sqrt{T}(\hat{\theta}-\theta)$ | | | | | | | | |
| g= | 3 | 1.08 | 1.20 | 1.33 | 1.47 | 1.59 | 1.68 | 1.75 | 1.80 | 1.83 |
| | 4 | 0.90 | 0.95 | 1.03 | 1.12 | 1.22 | 1.31 | 1.39 | 1.44 | 1.47 |
| | 5 | 0.89 | 0.91 | 0.94 | 1.00 | 1.07 | 1.14 | 1.20 | 1.25 | 1.28 |
| | 6 | 0.88 | 0.89 | 0.92 | 0.96 | 1.03 | 1.10 | 1.16 | 1.21 | 1.24 |
| | 7 | 0.88 | 0.88 | 0.89 | 0.92 | 0.96 | 1.02 | 1.08 | 1.13 | 1.17 |
| | 8 | 0.88 | 0.88 | 0.89 | 0.91 | 0.95 | 1.01 | 1.07 | 1.13 | 1.16 |
| | 9 | 0.88 | 0.88 | 0.88 | 0.90 | 0.93 | 0.97 | 1.03 | 1.08 | 1.11 |
| | 10 | 0.88 | 0.88 | 0.88 | 0.89 | 0.92 | 0.96 | 1.01 | 1.06 | 1.10 |
| | 11 | 0.88 | 0.88 | 0.88 | 0.88 | 0.90 | 0.93 | 0.97 | 1.02 | 1.06 |
| | 12 | 0.88 | 0.88 | 0.88 | 0.88 | 0.90 | 0.93 | 0.97 | 1.02 | 1.06 |
| | 16 | 0.88 | 0.88 | 0.88 | 0.88 | 0.88 | 0.90 | 0.93 | 0.97 | 1.01 |
| | 24 | 0.88 | 0.88 | 0.88 | 0.88 | 0.88 | 0.88 | 0.89 | 0.92 | 0.96 |
| | 32 | 0.88 | 0.88 | 0.88 | 0.88 | 0.88 | 0.88 | 0.88 | 0.90 | 0.93 |
| | $V_{MLE}$ | 0.88 | 0.88 | 0.88 | 0.88 | 0.88 | 0.88 | 0.88 | 0.88 | 0.88 |
| | | Asymptotic variance of $\sqrt{T}(\hat{\Theta}-\Theta)$ | | | | | | | | |
| g= | 3 | 9.37 | 10.82 | 13.85 | 19.82 | 32.10 | 59.78 | 132.73 | 385.95 | 2042.56 |
| | 4 | 1.16 | 1.28 | 1.53 | 2.04 | 3.08 | 5.42 | 11.58 | 32.88 | 171.91 |
| | 5 | 1.05 | 1.15 | 1.37 | 1.82 | 2.73 | 4.79 | 10.20 | 28.91 | 151.02 |
| | 6 | 1.03 | 1.14 | 1.36 | 1.80 | 2.71 | 4.75 | 10.12 | 28.67 | 149.78 |
| | 7 | 1.03 | 1.13 | 1.31 | 1.68 | 2.43 | 4.09 | 8.41 | 23.27 | 120.09 |
| | 8 | 1.00 | 0.99 | 1.01 | 1.08 | 1.30 | 1.85 | 3.33 | 8.48 | 41.94 |
| | 9 | 0.99 | 0.98 | 0.98 | 1.04 | 1.23 | 1.72 | 3.07 | 7.78 | 38.36 |
| | 10 | 0.99 | 0.98 | 0.98 | 1.03 | 1.22 | 1.71 | 3.05 | 7.72 | 38.07 |
| | 11 | 0.99 | 0.97 | 0.97 | 1.02 | 1.18 | 1.61 | 2.80 | 6.93 | 33.76 |
| | 12 | 0.99 | 0.96 | 0.93 | 0.90 | 0.93 | 1.09 | 1.64 | 3.64 | 16.75 |
| | 16 | 0.99 | 0.96 | 0.91 | 0.86 | 0.81 | 0.83 | 1.05 | 2.01 | 8.47 |
| | 24 | 0.99 | 0.96 | 0.91 | 0.84 | 0.76 | 0.68 | 0.67 | 0.94 | 3.17 |
| | 32 | 0.99 | 0.96 | 0.91 | 0.84 | 0.75 | 0.65 | 0.56 | 0.61 | 1.59 |
| | $V_{MLE}$ | 0.99 | 0.96 | 0.91 | 0.84 | 0.75 | 0.64 | 0.51 | 0.36 | 0.19 |

Table 12: Asymptotic Variance of $\sqrt{T}(\hat{\theta}-\theta)$ and $\sqrt{T}(\hat{\Theta}-\Theta)$ of MA(1)-Seasonal $MA(1)_4$ processes, $y_t = (1-\theta L)(1-\Theta L^4)\epsilon_t$ : $\theta = 0.55$

| | | $\Theta$ | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | 0.10 | 0.20 | 0.30 | 0.40 | 0.50 | 0.60 | 0.70 | 0.80 | 0.90 |
| | | Asymptotic variance of $\sqrt{T}(\hat{\theta}-\theta)$ | | | | | | | | |
| g= | 3 | 1.64 | 1.91 | 2.19 | 2.46 | 2.69 | 2.87 | 3.01 | 3.09 | 3.14 |
| | 4 | 0.87 | 0.98 | 1.13 | 1.29 | 1.46 | 1.61 | 1.73 | 1.81 | 1.85 |
| | 5 | 0.78 | 0.84 | 0.93 | 1.04 | 1.16 | 1.27 | 1.37 | 1.45 | 1.49 |
| | 6 | 0.74 | 0.78 | 0.85 | 0.93 | 1.04 | 1.15 | 1.25 | 1.33 | 1.37 |
| | 7 | 0.72 | 0.73 | 0.76 | 0.82 | 0.89 | 0.99 | 1.09 | 1.16 | 1.21 |
| | 8 | 0.72 | 0.73 | 0.75 | 0.79 | 0.85 | 0.94 | 1.03 | 1.10 | 1.14 |
| | 9 | 0.71 | 0.72 | 0.73 | 0.76 | 0.81 | 0.89 | 0.96 | 1.03 | 1.08 |
| | 10 | 0.71 | 0.72 | 0.72 | 0.75 | 0.79 | 0.85 | 0.93 | 1.00 | 1.04 |
| | 11 | 0.71 | 0.71 | 0.71 | 0.72 | 0.75 | 0.79 | 0.86 | 0.92 | 0.97 |
| | 12 | 0.71 | 0.71 | 0.71 | 0.72 | 0.74 | 0.78 | 0.85 | 0.91 | 0.96 |
| | 16 | 0.71 | 0.71 | 0.71 | 0.71 | 0.72 | 0.73 | 0.77 | 0.83 | 0.88 |
| | 24 | 0.71 | 0.71 | 0.71 | 0.71 | 0.71 | 0.71 | 0.72 | 0.76 | 0.80 |
| | 32 | 0.71 | 0.71 | 0.71 | 0.71 | 0.71 | 0.71 | 0.71 | 0.73 | 0.77 |
| | $V_{MLE}$ | 0.71 | 0.71 | 0.71 | 0.71 | 0.71 | 0.71 | 0.71 | 0.71 | 0.71 |
| | | Asymptotic variance of $\sqrt{T}(\hat{\Theta}-\Theta)$ | | | | | | | | |
| g= | 3 | 4.88 | 5.59 | 7.06 | 9.93 | 15.78 | 28.88 | 63.24 | 182.17 | 959.13 |
| | 4 | 1.31 | 1.44 | 1.73 | 2.30 | 3.48 | 6.13 | 13.09 | 37.19 | 194.53 |
| | 5 | 1.11 | 1.21 | 1.44 | 1.90 | 2.85 | 4.99 | 10.61 | 30.06 | 157.02 |
| | 6 | 1.06 | 1.17 | 1.39 | 1.84 | 2.76 | 4.84 | 10.29 | 29.16 | 152.30 |
| | 7 | 1.06 | 1.15 | 1.31 | 1.61 | 2.21 | 3.54 | 7.00 | 18.87 | 96.03 |
| | 8 | 1.03 | 1.03 | 1.06 | 1.16 | 1.41 | 2.01 | 3.66 | 9.35 | 46.33 |
| | 9 | 1.01 | 1.00 | 1.01 | 1.08 | 1.27 | 1.79 | 3.20 | 8.10 | 39.96 |
| | 10 | 1.01 | 1.00 | 1.00 | 1.06 | 1.25 | 1.74 | 3.11 | 7.87 | 38.81 |
| | 11 | 1.01 | 0.99 | 0.99 | 1.03 | 1.17 | 1.54 | 2.59 | 6.24 | 29.92 |
| | 12 | 1.01 | 0.98 | 0.95 | 0.93 | 0.97 | 1.16 | 1.76 | 3.95 | 18.22 |
| | 16 | 1.01 | 0.98 | 0.93 | 0.88 | 0.84 | 0.87 | 1.11 | 2.15 | 9.10 |
| | 24 | 1.01 | 0.98 | 0.93 | 0.86 | 0.77 | 0.69 | 0.69 | 0.98 | 3.35 |
| | 32 | 1.01 | 0.98 | 0.93 | 0.85 | 0.76 | 0.66 | 0.57 | 0.63 | 1.66 |
| | $V_{MLE}$ | 1.01 | 0.98 | 0.93 | 0.85 | 0.76 | 0.65 | 0.52 | 0.36 | 0.19 |

Table 13: Asymptotic Variance of $\sqrt{T}(\hat{\theta}-\theta)$ and $\sqrt{T}(\hat{\Theta}-\Theta)$ of MA(1)-Seasonal $MA(1)_4$ processes,$y_t = (1-\theta L)(1-\Theta L^4)\epsilon_t$ : $\theta$ =-0.25

| | | $\Theta$ | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | 0.10 | 0.20 | 0.30 | 0.40 | 0.50 | 0.60 | 0.70 | 0.80 | 0.90 |
| | | Asymptotic variance of $\sqrt{T}(\hat{\theta}-\theta)$ | | | | | | | | |
| g= | 3 | 1.02 | 1.11 | 1.21 | 1.32 | 1.42 | 1.50 | 1.56 | 1.60 | 1.62 |
| | 4 | 0.95 | 0.99 | 1.05 | 1.13 | 1.22 | 1.30 | 1.36 | 1.41 | 1.43 |
| | 5 | 0.94 | 0.95 | 0.98 | 1.02 | 1.07 | 1.13 | 1.19 | 1.23 | 1.26 |
| | 6 | 0.94 | 0.94 | 0.96 | 1.00 | 1.05 | 1.11 | 1.17 | 1.21 | 1.24 |
| | 7 | 0.94 | 0.94 | 0.95 | 0.97 | 1.00 | 1.06 | 1.11 | 1.16 | 1.19 |
| | 8 | 0.94 | 0.94 | 0.95 | 0.97 | 1.00 | 1.05 | 1.11 | 1.15 | 1.18 |
| | 9 | 0.94 | 0.94 | 0.94 | 0.95 | 0.97 | 1.01 | 1.06 | 1.10 | 1.13 |
| | 10 | 0.94 | 0.94 | 0.94 | 0.95 | 0.97 | 1.00 | 1.05 | 1.10 | 1.13 |
| | 11 | 0.94 | 0.94 | 0.94 | 0.94 | 0.95 | 0.98 | 1.02 | 1.07 | 1.10 |
| | 12 | 0.94 | 0.94 | 0.94 | 0.94 | 0.95 | 0.98 | 1.02 | 1.07 | 1.10 |
| | 16 | 0.94 | 0.94 | 0.94 | 0.94 | 0.94 | 0.95 | 0.98 | 1.02 | 1.06 |
| | 24 | 0.94 | 0.94 | 0.94 | 0.94 | 0.94 | 0.94 | 0.95 | 0.97 | 1.01 |
| | 32 | 0.94 | 0.94 | 0.94 | 0.94 | 0.94 | 0.94 | 0.94 | 0.95 | 0.99 |
| | $V_{MLE}$ | 0.94 | 0.94 | 0.94 | 0.94 | 0.94 | 0.94 | 0.94 | 0.94 | 0.94 |
| | | Asymptotic variance of $\sqrt{T}(\hat{\Theta}-\Theta)$ | | | | | | | | |
| g= | 3 | 17.28 | 20.04 | 25.83 | 37.33 | 61.06 | 114.79 | 256.74 | 750.29 | 3981.56 |
| | 4 | 1.10 | 1.21 | 1.45 | 1.93 | 2.91 | 5.11 | 10.91 | 30.95 | 161.77 |
| | 5 | 1.03 | 1.14 | 1.36 | 1.80 | 2.71 | 4.76 | 10.12 | 28.69 | 149.90 |
| | 6 | 1.03 | 1.14 | 1.36 | 1.80 | 2.70 | 4.74 | 10.10 | 28.62 | 149.53 |
| | 7 | 1.03 | 1.13 | 1.33 | 1.73 | 2.54 | 4.37 | 9.13 | 25.55 | 132.61 |
| | 8 | 0.99 | 0.98 | 0.99 | 1.06 | 1.26 | 1.78 | 3.20 | 8.13 | 40.14 |
| | 9 | 0.99 | 0.97 | 0.97 | 1.03 | 1.22 | 1.71 | 3.05 | 7.73 | 38.09 |
| | 10 | 0.99 | 0.97 | 0.97 | 1.03 | 1.22 | 1.71 | 3.05 | 7.71 | 38.01 |
| | 11 | 0.99 | 0.97 | 0.97 | 1.02 | 1.19 | 1.65 | 2.91 | 7.27 | 35.61 |
| | 12 | 0.99 | 0.96 | 0.92 | 0.89 | 0.91 | 1.07 | 1.59 | 3.53 | 16.18 |
| | 16 | 0.99 | 0.96 | 0.91 | 0.85 | 0.81 | 0.82 | 1.03 | 1.96 | 8.23 |
| | 24 | 0.99 | 0.96 | 0.91 | 0.84 | 0.76 | 0.68 | 0.66 | 0.92 | 3.11 |
| | 32 | 0.99 | 0.96 | 0.91 | 0.84 | 0.75 | 0.65 | 0.56 | 0.60 | 1.57 |
| | $V_{MLE}$ | 0.99 | 0.96 | 0.91 | 0.84 | 0.75 | 0.64 | 0.51 | 0.36 | 0.19 |

$SMA(1)_{12}$ processes are reported in Table 14-16. In each table, the MA(1) coefficients are all the same, while the seasonal MA coefficients ($\Theta$) are varied from 0.1 to 0.9 by steps of 0.1. The moving average coefficient is fixed at 0.25 for all cases in Table 14, whereas in Table 15 and Table 16 the MA coefficients are 0.45 and 0.65 respectively. In each case, the asymptotic variance of MDE using first 11, 12, 13, 22, 23, 24, 25, 26, 35, 36, 37, and 48 autocorrelations are reported. Note that we are not able to identify the parameters if the first 10 autocorrelations are used in MDE. The asymptotic variance of $\sqrt{T}(\hat{\theta}_{MDE} - \theta)$ of MDE using first 25 autocorrelations is very close to that of MLE when both of $\theta$ and $\Theta$ are small.

In general, if the number of autocorrelations used in MDE is large enough, MDE appears to be asymptotically as efficient as MLE. Given that MA coefficients are the same, it is found that the higher the seasonal parameter, the higher the number of autocorrelations is needed to guarantee that MDE is efficient.

The results in Table 15 are very similar to those in Table 14. The asymptotic variance of MDE is very close to that of MLE given that the number of autocorrelations used in MDE is large enough. One set of the parameter values are set very close to those found in the airline model of Box, Jenkins, and Reinsel (1994). Given that $\theta = 0.45$, if g=25, asymptotic variance of $\sqrt{T}(\hat{\theta}_{MDE} - \theta)$ is very close to that of MLE when $\Theta$ is small, e.g., $\Theta$ =0.1, 0.2, and 0.3. But g needs to be higher than 25 in order to guarantee that the asymptotic variance of seasonal MA estimators is very close to that of MLE for the seasonal models that $\Theta = 0.2$, and 0.3. Table 16 reports the results for the seasonal models which all have $\theta = 0.65$, while $\Theta$ changes for all cases. When $\Theta = 0.1$, MDE using first 25 autocorrelations can be as efficient as MLE.

For the MA(1)-SMA(1)$_{12}$ model MDE using first 24 autocorrelations shows a lot of efficiency improvement than the MDE using first 23 autocorrelations. This result demonstrates that the 12th, 24th, and 36th order autocorrelation might be very important for MDE in estimating the seasonal models with 12 periods. If $\Theta$ is very large, adding autocorrelations at higher lags such as $\rho_{48}$, $\rho_{60}$, $\rho_{72}$, $\cdots$, might give more efficiency gain than adding other autocorrelations.

Table 17 provides a comparison of the asymptotic variances of MDE using different sets of autocorrelations. It is found that the MDE using first 1-13 autocorrelations and autocorrelations at lag 23, 24, 25, 35, 36, and 37 has some advantage over the other methods. In Table 17 method 1 indicates the case that the MDE using autocorrelations at lag 1, 11, 12, 13, 23, 24, 25, 35, 36, and 37, whereas method 2 represents the case of MDE using autocorrelations at lag 1-13, 23, 24, 25, 35, 36, and 37. The results in Table 13 show that method 2 is very useful in reducing the asymptotic variance of the MA coefficient. Consider the case that $\theta = 0.30$ and $\Theta = 0.1$. It is clearly that the asymptotic variance of $\theta$ significantly improved by using method 2.

## 4. Estimation Results of MDE for Airline Model

The MDE method is applied to estimate the airline passenger data, which is also analyzed by Box and Jenkins (1976). The total number of observations (T) of this data set after the regular and seasonal difference is 131. We present the estimation

Table 14: Asymptotic Variance of $\sqrt{T}(\hat{\theta}-\theta)$ and $\sqrt{T}(\hat{\Theta}-\Theta)$ of MA(1)-Seasonal MA$(1)_{12}$ processes,$y_t = (1-\theta L)(1-\Theta L^{12})\epsilon_t$ : $\theta = 0.25$

| | | $\Theta$ | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | 0.10 | 0.20 | 0.30 | 0.40 | 0.50 | 0.60 | 0.70 | 0.80 | 0.90 |
| | | Asymptotic variance of $\sqrt{T}(\hat{\theta}-\theta)$ | | | | | | | | |
| g= | 11 | 0.96 | 1.01 | 1.08 | 1.16 | 1.24 | 1.30 | 1.35 | 1.38 | 1.40 |
| | 12 | 0.95 | 0.98 | 1.03 | 1.10 | 1.17 | 1.24 | 1.29 | 1.33 | 1.34 |
| | 13 | 0.94 | 0.95 | 0.97 | 1.00 | 1.05 | 1.10 | 1.14 | 1.18 | 1.20 |
| | 22 | 0.94 | 0.94 | 0.95 | 0.98 | 1.01 | 1.06 | 1.11 | 1.14 | 1.17 |
| | 23 | 0.94 | 0.94 | 0.95 | 0.96 | 0.99 | 1.04 | 1.08 | 1.12 | 1.14 |
| | 24 | 0.94 | 0.94 | 0.95 | 0.96 | 0.99 | 1.04 | 1.08 | 1.12 | 1.14 |
| | 25 | 0.94 | 0.94 | 0.94 | 0.95 | 0.97 | 1.00 | 1.04 | 1.07 | 1.10 |
| | 26 | 0.94 | 0.94 | 0.94 | 0.94 | 0.96 | 0.99 | 1.03 | 1.06 | 1.09 |
| | 35 | 0.94 | 0.94 | 0.94 | 0.94 | 0.95 | 0.97 | 1.01 | 1.05 | 1.07 |
| | 36 | 0.94 | 0.94 | 0.94 | 0.94 | 0.95 | 0.97 | 1.01 | 1.05 | 1.07 |
| | 37 | 0.94 | 0.94 | 0.94 | 0.94 | 0.94 | 0.96 | 0.99 | 1.02 | 1.05 |
| | 48 | 0.94 | 0.94 | 0.94 | 0.94 | 0.94 | 0.95 | 0.97 | 1.01 | 1.04 |
| | $V_{MLE}$ | 0.94 | 0.94 | 0.94 | 0.94 | 0.94 | 0.94 | 0.94 | 0.94 | 0.94 |
| | | Asymptotic variance of $\sqrt{T}(\hat{\Theta}-\Theta)$ | | | | | | | | |
| g= | 11 | 16.80 | 19.52 | 25.28 | 36.80 | 60.72 | 115.13 | 259.30 | 761.46 | 4051.71 |
| | 12 | 1.10 | 1.21 | 1.45 | 1.93 | 2.91 | 5.12 | 10.92 | 30.99 | 161.98 |
| | 13 | 1.03 | 1.14 | 1.36 | 1.80 | 2.71 | 4.76 | 10.14 | 28.75 | 150.18 |
| | 22 | 1.03 | 1.14 | 1.36 | 1.80 | 2.70 | 4.74 | 10.09 | 28.61 | 149.48 |
| | 23 | 1.03 | 1.12 | 1.32 | 1.72 | 2.53 | 4.35 | 9.09 | 25.45 | 132.14 |
| | 24 | 0.99 | 0.98 | 0.99 | 1.06 | 1.26 | 1.78 | 3.20 | 8.13 | 40.16 |
| | 25 | 0.99 | 0.97 | 0.97 | 1.03 | 1.22 | 1.71 | 3.06 | 7.74 | 38.13 |
| | 26 | 0.99 | 0.97 | 0.97 | 1.03 | 1.22 | 1.71 | 3.05 | 7.71 | 38.01 |
| | 35 | 0.99 | 0.97 | 0.97 | 1.02 | 1.19 | 1.65 | 2.90 | 7.24 | 35.48 |
| | 36 | 0.99 | 0.96 | 0.92 | 0.89 | 0.91 | 1.07 | 1.59 | 3.53 | 16.19 |
| | 37 | 0.99 | 0.96 | 0.92 | 0.89 | 0.90 | 1.04 | 1.54 | 3.40 | 15.57 |
| | 48 | 0.99 | 0.96 | 0.91 | 0.85 | 0.81 | 0.82 | 1.03 | 1.96 | 8.23 |
| | $V_{MLE}$ | 0.99 | 0.96 | 0.91 | 0.84 | 0.75 | 0.64 | 0.51 | 0.36 | 0.19 |

Note: g=11 indicates that $\rho = [\ \rho_1\ \rho_2 \cdots\ \rho_3]'$. Hence, g=11 represents that first 11 autocorrelations are used in MDE, while g=12 represents that the first 12 autocorrelations are used in MDE, and etc. $V_{MLE}$ is the asymptotic variance of $\sqrt{T}(\hat{\theta}_{MLE}-\theta)$.

Table 15: Asymptotic Variance of $\sqrt{T}(\hat{\theta}-\theta)$ and $\sqrt{T}(\hat{\Theta}-\Theta)$ of MA(1)-Seasonal $MA(1)_{12}$ processes,$y_t = (1-\theta L)(1-\Theta L^{12})\epsilon_t$ : $\theta = 0.45$

| | | $\Theta$ | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | 0.10 | 0.20 | 0.30 | 0.40 | 0.50 | 0.60 | 0.70 | 0.80 | 0.90 |
| | | Asymptotic variance of $\sqrt{T}(\hat{\theta}-\theta)$ | | | | | | | | |
| g= | 11 | 0.81 | 0.86 | 0.92 | 0.99 | 1.05 | 1.11 | 1.15 | 1.18 | 1.19 |
| | 12 | 0.81 | 0.84 | 0.89 | 0.95 | 1.02 | 1.07 | 1.12 | 1.15 | 1.16 |
| | 13 | 0.80 | 0.82 | 0.85 | 0.89 | 0.94 | 0.99 | 1.04 | 1.07 | 1.08 |
| | 22 | 0.80 | 0.80 | 0.81 | 0.83 | 0.86 | 0.90 | 0.94 | 0.97 | 0.99 |
| | 23 | 0.80 | 0.80 | 0.81 | 0.82 | 0.85 | 0.89 | 0.93 | 0.96 | 0.98 |
| | 24 | 0.80 | 0.80 | 0.81 | 0.82 | 0.85 | 0.89 | 0.93 | 0.96 | 0.98 |
| | 25 | 0.80 | 0.80 | 0.80 | 0.81 | 0.83 | 0.87 | 0.90 | 0.93 | 0.95 |
| | 26 | 0.80 | 0.80 | 0.80 | 0.81 | 0.82 | 0.85 | 0.88 | 0.92 | 0.94 |
| | 35 | 0.80 | 0.80 | 0.80 | 0.80 | 0.81 | 0.83 | 0.86 | 0.89 | 0.92 |
| | 36 | 0.80 | 0.80 | 0.80 | 0.80 | 0.81 | 0.83 | 0.86 | 0.89 | 0.91 |
| | 37 | 0.80 | 0.80 | 0.80 | 0.80 | 0.81 | 0.82 | 0.85 | 0.88 | 0.90 |
| | 48 | 0.80 | 0.80 | 0.80 | 0.80 | 0.80 | 0.81 | 0.83 | 0.86 | 0.88 |
| | $V_{MLE}$ | 0.80 | 0.80 | 0.80 | 0.80 | 0.80 | 0.80 | 0.80 | 0.80 | 0.80 |
| | | Asymptotic variance of $\sqrt{T}(\hat{\Theta}-\Theta)$ | | | | | | | | |
| g= | 11 | 5.17 | 5.96 | 7.64 | 10.98 | 17.93 | 33.71 | 75.48 | 220.88 | 1173.21 |
| | 12 | 1.23 | 1.36 | 1.64 | 2.18 | 3.29 | 5.80 | 12.39 | 35.18 | 183.93 |
| | 13 | 1.07 | 1.18 | 1.41 | 1.86 | 2.81 | 4.94 | 10.52 | 29.83 | 155.89 |
| | 22 | 1.03 | 1.14 | 1.36 | 1.80 | 2.70 | 4.74 | 10.09 | 28.61 | 149.48 |
| | 23 | 1.02 | 1.10 | 1.26 | 1.57 | 2.21 | 3.64 | 7.36 | 20.16 | 103.50 |
| | 24 | 1.00 | 1.00 | 1.02 | 1.11 | 1.34 | 1.92 | 3.48 | 8.88 | 43.96 |
| | 25 | 0.99 | 0.98 | 0.98 | 1.05 | 1.24 | 1.75 | 3.13 | 7.94 | 39.16 |
| | 26 | 0.99 | 0.97 | 0.98 | 1.03 | 1.22 | 1.71 | 3.06 | 7.76 | 38.23 |
| | 35 | 0.99 | 0.97 | 0.96 | 1.00 | 1.14 | 1.52 | 2.60 | 6.35 | 30.70 |
| | 36 | 0.99 | 0.96 | 0.93 | 0.91 | 0.94 | 1.11 | 1.69 | 3.77 | 17.35 |
| | 37 | 0.99 | 0.96 | 0.92 | 0.89 | 0.91 | 1.06 | 1.57 | 3.47 | 15.89 |
| | 48 | 0.99 | 0.96 | 0.91 | 0.86 | 0.81 | 0.84 | 1.07 | 2.06 | 8.71 |
| | $V_{MLE}$ | 0.99 | 0.96 | 0.91 | 0.84 | 0.75 | 0.64 | 0.51 | 0.36 | 0.19 |

Table 16: Asymptotic Variance of $\sqrt{T}(\hat{\theta} - \theta)$ and $\sqrt{T}(\hat{\Theta} - \Theta)$ of MA(1)-Seasonal $MA(1)_{12}$ processes, $y_t = (1 - \theta L)(1 - \Theta L^{12})\epsilon_t$ : $\theta = 0.65$

| | | $\Theta$ | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | 0.10 | 0.20 | 0.30 | 0.40 | 0.50 | 0.60 | 0.70 | 0.80 | 0.90 |
| | | Asymptotic variance of $\sqrt{T}(\hat{\theta} - \theta)$ | | | | | | | | |
| g= | 11 | 0.60 | 0.65 | 0.70 | 0.76 | 0.81 | 0.86 | 0.89 | 0.91 | 0.93 |
| | 12 | 0.60 | 0.63 | 0.68 | 0.73 | 0.79 | 0.83 | 0.87 | 0.89 | 0.90 |
| | 13 | 0.59 | 0.62 | 0.66 | 0.70 | 0.75 | 0.80 | 0.83 | 0.86 | 0.87 |
| | 22 | 0.58 | 0.58 | 0.59 | 0.61 | 0.64 | 0.67 | 0.70 | 0.73 | 0.74 |
| | 23 | 0.58 | 0.58 | 0.59 | 0.60 | 0.63 | 0.66 | 0.69 | 0.72 | 0.73 |
| | 24 | 0.58 | 0.58 | 0.59 | 0.60 | 0.62 | 0.66 | 0.69 | 0.71 | 0.73 |
| | 25 | 0.58 | 0.58 | 0.58 | 0.60 | 0.62 | 0.65 | 0.68 | 0.71 | 0.72 |
| | 26 | 0.58 | 0.58 | 0.58 | 0.59 | 0.61 | 0.64 | 0.67 | 0.69 | 0.71 |
| | 35 | 0.58 | 0.58 | 0.58 | 0.58 | 0.59 | 0.61 | 0.63 | 0.66 | 0.68 |
| | 36 | 0.58 | 0.58 | 0.58 | 0.58 | 0.59 | 0.61 | 0.63 | 0.66 | 0.68 |
| | 37 | 0.58 | 0.58 | 0.58 | 0.58 | 0.59 | 0.60 | 0.63 | 0.65 | 0.67 |
| | 48 | 0.58 | 0.58 | 0.58 | 0.58 | 0.58 | 0.59 | 0.61 | 0.63 | 0.65 |
| | $V_{MLE}$ | 0.58 | 0.58 | 0.58 | 0.58 | 0.58 | 0.58 | 0.58 | 0.58 | 0.58 |
| | | Asymptotic variance of $\sqrt{T}(\hat{\Theta} - \Theta)$ | | | | | | | | |
| g= | 11 | 2.48 | 2.82 | 3.54 | 4.98 | 7.97 | 14.72 | 32.55 | 94.50 | 499.92 |
| | 12 | 1.37 | 1.51 | 1.81 | 2.40 | 3.63 | 6.38 | 13.63 | 38.69 | 202.28 |
| | 13 | 1.15 | 1.27 | 1.52 | 2.02 | 3.05 | 5.35 | 11.42 | 32.40 | 169.36 |
| | 22 | 1.03 | 1.14 | 1.36 | 1.80 | 2.70 | 4.74 | 10.10 | 28.62 | 149.49 |
| | 23 | 1.01 | 1.06 | 1.17 | 1.38 | 1.84 | 2.86 | 5.54 | 14.77 | 74.77 |
| | 24 | 1.00 | 1.01 | 1.05 | 1.15 | 1.41 | 2.04 | 3.71 | 9.50 | 47.11 |
| | 25 | 1.00 | 0.99 | 1.00 | 1.08 | 1.29 | 1.84 | 3.31 | 8.41 | 41.57 |
| | 26 | 0.99 | 0.98 | 0.99 | 1.05 | 1.25 | 1.76 | 3.16 | 8.00 | 39.50 |
| | 35 | 0.99 | 0.97 | 0.95 | 0.97 | 1.06 | 1.36 | 2.22 | 5.26 | 24.96 |
| | 36 | 0.99 | 0.96 | 0.93 | 0.92 | 0.96 | 1.15 | 1.76 | 3.96 | 18.31 |
| | 37 | 0.99 | 0.96 | 0.92 | 0.90 | 0.92 | 1.09 | 1.63 | 3.62 | 16.64 |
| | 48 | 0.99 | 0.96 | 0.91 | 0.86 | 0.82 | 0.86 | 1.11 | 2.14 | 9.10 |
| | $V_{MLE}$ | 0.99 | 0.96 | 0.91 | 0.84 | 0.75 | 0.64 | 0.51 | 0.36 | 0.19 |

Table 17: Asymptotic Variance of $\sqrt{T}(\hat{\lambda}_{MDE} - \lambda)$ of Seasonal MA(1) models, $y_t = (1 - \theta L)(1 - \Theta L^{12})e_t$

| | | | | |
|---|---|---|---|---|
| | $\theta =$ | 0.10 | 0.30 | 0.37 | 0.30 |
| | $\Theta =$ | 0.30 | 0.10 | 0.57 | 0.70 |
| 1-13* | | 1.01 0.00 | 0.91 0.00 | 1.03 0.07 | 1.13 0.07 |
| | | 0.00 1.36 | 0.00 1.04 | 0.07 4.01 | 0.07 10.19 |
| 1-25 | | 0.99 0.00 | 0.91 0.00 | 0.92 0.02 | 1.01 0.02 |
| | | 0.00 0.97 | 0.00 0.99 | 0.02 1.52 | 0.02 3.07 |
| 1-37 | | 0.99 0.00 | 0.91 0.00 | 0.88 0.01 | 0.96 0.01 |
| | | 0.00 0.92 | 0.00 0.99 | 0.01 0.98 | 0.01 1.55 |
| Method 1† | | 1.03 0.00 | 1.35 0.00 | 1.65 0.00 | 1.42 0.00 |
| | | 0.00 0.92 | 0.00 1.00 | 0.00 1.00 | 0.00 1.56 |
| Method 2‡ | | 0.99 0.00 | 0.97 0.04 | 0.97 0.04 | 1.03 0.03 |
| | | 0.00 0.92 | 0.04 1.00 | 0.04 1.00 | 0.03 1.56 |
| $V_{MLE}$ | | 0.99 -0.00 | 0.91 -0.00 | 0.86 -0.00 | 0.91 -0.00 |
| | | -0.00 0.91 | -0.00 0.99 | -0.00 0.68 | -0.00 0.51 |

* 1-13 represents the case that first 13 autocorrelations are used in MDE, while 1-25 represents MDE using first 25 autocorrelations.

† Method 1 represents the case of MDE using autocorrelations at lag 1, 11, 12, 13, 23, 24, 25, 35, 36, and 37.

‡ Method 2 represents the case of MDE using autocorrelations at lag 1-13, 23, 24, 25, 35, 36, and 37.

result of the MDE using the first 48 autocorrelations. The MD (minimum distance) estimates are $\hat{\theta}_{MDE} = 0.399$ and $\hat{\Theta}_{MDE} = 0.523$ with standard error 0.0893 and 0.0982, respectively. The estimation results of approximate MLE are also reported as a comparison. The MD estimates of airline model are very close to those of MLE. The asymptotic variance of $\hat{\lambda}_{MDE}$ is calculated as $(1/T)(D'_{\hat{\lambda}}\hat{C}^{-1}D_{\hat{\lambda}})^{-1}$, where $D_{\hat{\lambda}}$ denotes the partial derivative of $\rho(\lambda)$ with respect to $\lambda'$ evaluated at $\hat{\lambda}_{MDE}$ and $\hat{C}$ is calculated by using equation (4). The standard deviation is calculated by taking squared root of the diagonal elements of the variance matrix. We also calculate the residuals of the MA(1)-SMA$(1)_{12}$ model. By calculating the ACF of the residuals, we find that the Box-Pierce LM test from 25 degree of freedom is 22.29.

## 5. Concluding Remarks

This chapter discusses the properties of the MDE for MA(1)-seasonal MA$(1)_s$ processes, where $s$ is equal to either 4 or 12. This airline model has been applied to model many economic time series data. Although the main focus of the estimation of the seasonal ARMA models is the MLE method, the MDE is an attracting alternative to the estimation of seasonal ARMA models. The MDE has the advantage of imposing very little in terms of distributional assumptions on the innovation process. Also, the MDE is relatively simple to compute.

Calculations reveal that if the number of autocorrelations used in MDE is large enough, the MDE for the airline model appears to be asymptotically as efficient as MLE under normality. It is also found that when the MA parameter is fixed, the

Table 18: Estimation Results of MDE of Airline Model: $y_t = (1 - \theta L)(1 - \Theta L^{12})e_t$

| $\hat{\theta}_{MDE}$ | $\hat{\Theta}_{MDE}$ |
|---|---|
| 0.399 | 0.523 |
| (0.089)* | (0.098) |

The acf of the residuals is:

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| -0.0077 | 0.0242 | -0.1331 | -0.0787 | 0.0853 | 0.0740 | -0.0433 | -0.0025 | 0.1376 |
| 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 |
| -0.0545 | -0.0027 | -0.1226 | 0.0004 | 0.0220 | 0.0782 | -0.1197 | 0.0588 | 0.0145 |
| 19 | 20 | 21 | 22 | 23 | 24 | 25 | | |
| -0.0922 | -0.0937 | -0.0150 | -0.0280 | 0.2156 | -0.0287 | -0.0626 | | |

The Box-Pierce LM test from 25 df, Q(25) : 22.29

| $\hat{\theta}_{MLE}$ | $\hat{\Theta}_{MLE}$ |
|---|---|
| 0.377 | 0.572 |
| (0.085) | (0.070) |

The Maximized value of the log likelihood :-385.576.

The acf of the residuals is:

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| -0.045 | 0.042 | -0.117 | -0.165 | 0.038 | 0.061 | -0.067 | -0.043 | 0.108 |
| 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 |
| -0.127 | -0.012 | 0.017 | 0.006 | 0.060 | 0.089 | -0.173 | -0.019 | -0.002 |
| 19 | 20 | 21 | 22 | 23 | 24 | 25 | | |
| -0.104 | -0.070 | -0.010 | -0.068 | 0.177 | 0.017 | 0.019 | | |

The Box-Pierce LM test from 25 df, Q(25) : 21.040

*Standard deviation is in the parenthesis.

higher the value of the seasonal MA parameter, the higher the number of autocorrelations is needed to guarantee that MDE is efficient. For the MDE of MA(1)-SMA$(1)_s$ processes, adding autocorrelations at lags of multiples of $s$, such as $\rho_s$, $\rho_{2s}$, $\rho_{3s}$, $\cdots$, to the MDE might have a larger improvement in efficiency than adding other autocorrelations.

# APPENDIX

# APPENDIX

## Asymptotic Variance of MLE of MA(1)-SMA(1)$_s$ Processes

The log-likelihood function of the MA(1)-SMA(1)$_s$ process is

$$L = -\frac{T}{2}\log(2\pi) - \frac{T}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{t=1}^{T}[(1-\theta L)^{-1}(1-\Theta L^s)^{-1}y_t]^2,$$

where $T$ denotes the total number of observations. The first order conditions of the maximization of log-likelihood function of the MA(1)-SMA(1)$_s$ process are

$$\frac{\partial L}{\partial \theta} = -\frac{1}{\sigma^2}\sum_{t=1}^{T}\epsilon_t[(1-\theta L)^{-2}(1-\Theta L^s)^{-1}y_{t-1}],$$

$$\frac{\partial L}{\partial \Theta} = -\frac{1}{\sigma^2}\sum_{t=1}^{T}\epsilon_t[(1-\theta L)^{-1}(1-\Theta L^s)^{-2}y_{t-s}],$$

and

$$\frac{\partial L}{\partial \sigma^2} = \frac{1}{2\sigma^4}\sum_{t=1}^{T}\epsilon_t^2 - \frac{T}{2\sigma^2}.$$

The second order conditions are

$$\frac{\partial^2 L}{\partial \theta^2} = -\frac{1}{\sigma^2}\sum_{t=1}^{T}\left([(1-\theta L)^{-2}(1-\Theta L^s)^{-1}y_{t-1}]^2 + 2\epsilon_t[(1-\theta L)^{-3}(1-\Theta L^s)^{-1}y_{t-2}]\right)$$

$$\frac{\partial^2 L}{\partial \Theta^2} = -\frac{1}{\sigma^2}\sum_{t=1}^{T}\Big([(1-\theta L)^{-1}(1-\Theta L^s)^{-2}y_{t-s}]^2 + 2\,\epsilon_t\,[(1-\theta L)^{-1}(1-\Theta L^s)^{-3}y_{t-2s}]\Big)$$

$$\frac{\partial^2 L}{\partial \theta \partial \Theta} = -\frac{1}{\sigma^2}\left\{\sum_{t=1}^{T}[(1-\theta L)^{-1}(1-\Theta L^s)^{-2}y_{t-s}]\,[(1-\theta L)^{-2}(1-\Theta L^s)^{-1}y_{t-1}] + \sum_{t=1}^{T}\epsilon_t(1-\theta L)^{-2}(1-\Theta L^s)^{-2}y_{t-1-s}\right\}$$

$$\frac{\partial^2 L}{\partial \sigma^4} = \frac{-1}{\sigma^6}\sum_{t=1}^{T}\epsilon_t^2 + \frac{T}{2\sigma^4}$$

$$\frac{\partial^2 L}{\partial \sigma^2 \partial \theta} = \frac{1}{\sigma^2}\sum_{t=1}^{T}\epsilon_t[(1-\theta L)^{-2}(1-\Theta L^s)^{-1}y_{t-1}]$$

and

$$\frac{\partial^2 L}{\partial\sigma^2\partial\Theta} = \frac{1}{\sigma^2}\sum_{t=1}^{T}\epsilon_t(1-\theta L)^{-1}(1-\Theta L^s)^{-2}y_{t-s}$$

Using the result that $\epsilon_t = (1-\theta L)^{-1}(1-\Theta L^s)^{-1}y_t$, we have

$$\begin{aligned}
\frac{\partial^2 L}{\partial\theta^2} &= -\frac{1}{\sigma^2}\left\{\sum_{t=1}^{T}[(1-\theta L)^{-1}\epsilon_{t-1}]^2 + 2\sum_{t=1}^{T}\epsilon_t(1-\theta L)^{-2}\epsilon_{t-2}\right\} \\
\frac{\partial^2 L}{\partial\Theta^2} &= -\frac{1}{\sigma^2}\left\{\sum_{t=1}^{T}\left[(1-\Theta L^s)^{-1}\epsilon_{t-s}\right]^2 + 2\sum_{t=1}^{T}\epsilon_t(1-\Theta L^s)^{-2}\epsilon_{t-2s}\right\} \\
\frac{\partial^2 L}{\partial\theta\partial\Theta} &= -\frac{1}{\sigma^2}\sum_{t=1}^{T}\Big([(1-\Theta L^s)^{-1}\epsilon_{t-s}][(1-\theta L)^{-1}\epsilon_{t-1}] \\
&\quad +\epsilon_t\ (1-\theta L)^{-1}(1-\Theta L^s)^{-1}\epsilon_{t-1-s}\Big)
\end{aligned}$$

Because $E(\epsilon_t y_{t-j}) = 0$ for $j \geq 1$, we have $E(\frac{\partial^2 L}{\partial\sigma^2\partial\theta}) = 0$ and $E(\frac{\partial^2 L}{\partial\sigma^2\partial\Theta}) = 0$. To get the results for $E(\frac{\partial^2 L}{\partial\theta^2})$, $E(\frac{\partial^2 L}{\partial\Theta^2})$, and $E(\frac{\partial^2 L}{\partial\theta\partial\Theta})$, we define $u_t$ and $v_t$ as $(1-\theta L)u_t = \epsilon_t$, and $(1-\Theta L^s)v_t = \epsilon_t$. Hence, $u_t$ follows an AR(1) process and $v_t$ follows an AR(s) process.

Finally, the following results is obtained by using the result that $E(\epsilon_t\epsilon_{t-j}) = 0$ for $j \geq 1$.

$$\begin{aligned}
-E\left(\frac{\partial^2 L}{\partial\theta^2}\right) &= \frac{1}{\sigma^2}\sum_{t=1}^{T}Eu_{t-1}^2 = \frac{T}{1-\theta^2}, \\
-E\left(\frac{\partial^2 L}{\partial\Theta^2}\right) &= \frac{1}{\sigma^2}\sum_{t=1}^{T}Ev_{t-s}^2 = \frac{T}{1-\Theta^2} \\
-E\left(\frac{\partial^2 L}{\partial\theta\partial\Theta}\right) &= \frac{1}{\sigma^2}\sum_{t=1}^{T}E(u_{t-1}v_{t-s}) \\
&= \frac{1}{\sigma^2}\sum_{t=1}^{T}E[(1+\theta L+\theta^2L^2+\cdots)\epsilon_{t-1}(1+\Theta L^s+\Theta^2L^{2s}+\cdots)\epsilon_{t-s}] \\
&= \frac{1}{\sigma^2}T(\theta^{s-1}+\theta^{2s-1}\Theta^s+\theta^{3s-1}\Theta^{2s}+\cdots)\sigma^2 \\
&= T\theta^{s-1}(1+\theta^s\Theta^s+\theta^{2s}\Theta^{2s}+\cdots)
\end{aligned}$$

$$
\begin{aligned}
&= \frac{T\theta^{s-1}}{1-\Theta\theta} \\
-E\left(\frac{\partial^2 L}{\partial\sigma^4}\right) &= \frac{T\sigma^2}{\sigma^6} - \frac{T}{2\sigma^4} = \frac{T}{2\sigma^4}.
\end{aligned}
$$

We then have $\sqrt{T}(\hat{\lambda}_{MLE} - \lambda) \to N(0, V_{MLE})$, where

$$
V_{MLE} = \begin{bmatrix} \frac{1}{1-\theta^2} & \frac{\theta^{s-1}}{1-\Theta\theta} & 0 \\ \frac{\theta^{s-1}}{1-\Theta\theta} & \frac{1}{1-\Theta^2} & 0 \\ 0 & 0 & \frac{1}{2\sigma^4} \end{bmatrix}^{-1} .
$$

# CHAPTER 4

## Minimum Distance Estimation for GARCH Models

### 1 Introduction

The use of GARCH processes and their extensions to represent the time dependent heteroskedasticity present in many economic and financial economics series is now a widespread econometric procedure. Bollerslev, Chou and Kroner (1992) describe many of the applications of the methodology. Inference in the GARCH class of models is usually based on MLE or QMLE assuming a Gaussian conditional density and Bollerslev, Engle and Nelson (1994) and Hansen and Lee (1994) discuss many of the inferential issues involved in likelihood based procedures. For example, one of the most important models in empirical work is when the observed time series, $y_t$ follows a martingale with linear GARCH(1,1) volatility process, so that

$$y_t = \sigma_t u_t, \tag{23}$$

where $u_t$ is i.i.d. N(0,1) and $\sigma_t$ is a positive, time varying and measurable function with respect to the information set which is available at time t-1, and

$$\sigma_t^2 = \omega + \alpha y_{t-1}^2 + \beta \sigma_{t-1}^2, \tag{24}$$

where $\omega > 0, \alpha \geq 0$, $\beta \geq 0$, and $\alpha + \beta < 1$. It is then straightforward to maximize the Gaussian conditional density and obtain robust standard errors from the QMLE method of Bollerslev and Wooldridge (1992).

This chapter is concerned with using the Minimum Distance Estimator (MDE) based on the sample autocorrelations of the squared process to estimate the parameters of the GARCH model. For example, if $y_t$ follows the above martingale - GARCH(1,1) process, then $y_t^2$ has the ARMA(1,1) representation of

$$y_t^2 = \omega + (\alpha + \beta) y_{t-1}^2 + v_t - \beta v_{t-1}. \tag{25}$$

where $v_t = y_t^2 - \sigma_t^2$ and the expectation and variance conditional on information available at time t-1 are $E_{t-1} v_t = 0$ and $Var_{t-1} v_t = 2\sigma_t^4$. Hence despite the innovation process, $v_t$, being serially uncorrelated, it is not independent over time.

The MDE of the GARCH process parameters are very simple to compute from the first g sample autocorrelations of the squares of a realization of the process. The MDE is particularly attractive to use as an estimator in situations where the true underlying data generating process has extreme non-normality. When estimating a GARCH model with data exhibiting extreme kurtosis, the maximization of a Gaussian density and the subsequent use of QMLE to obtain robust standard errors, is not necessarily going to realize asyptotically efficient parameter estimates. Monte Carlo evidence presented in this chapter provides evidence that with certain conditional densities, and over certain regions of the parameter space, the MDE can compare very favorably with QMLE in terms of parameter estimation bias and mean squared error. In cases where difficulty is experienced in estimating GARCH models from extreme non Gaussian densities, the MDE can be recommended as an attractive alternative which can avoid problems of convergence.

The remainder of this chapter is organized as follows. Section 2 defines the MDE for the GARCH model. Section 3 illustrates the application of the MDE procedure to estimating the parameters of the ARMA(1,1) process with i.i.d. innovations. Monte Carlo results are presented. Section 4 then applies the MDE to estimating the parameters of the GARCH(1,1) process. The asymptotic efficiency of the MDE is found for various points in the parameter space for the conditional normal densities. Section 5 considers the MDE applied to estimate the parameters of the GARCH model for hourly exchange rate data. Section 6 provides a brief conclusion.

## 2. The MDE of GARCH Models

It is noted in chapter 2 that when the innovations in the ARMA process are i.i.d., the asymptotic variance of the sample autocorrelation is given by the Bartlett's formula. Hence, a consistent estimator of $C$ is $\hat{C}$, with $(i,j)$th element given by

$$\hat{c}_{ij} = \sum_{k=1}^{\infty}(\hat{\rho}_{k+i} + \hat{\rho}_{k-i} - 2\hat{\rho}_i\hat{\rho}_k)(\hat{\rho}_{k+j} + \hat{\rho}_{k-j} - 2\hat{\rho}_j\hat{\rho}_k) \tag{26}$$

In practical application the MDE is obtained by solving the following minimization problem:

$$\text{Min } S = (\hat{\rho} - \rho(\lambda))'\hat{C}^{-1}(\hat{\rho} - \rho(\lambda)). \tag{27}$$

While the computation of the optimal weighting matrix is straightforward in the case of i.i.d. innovations, the estimation of GARCH process parameters to be discussed later requires estimation of an ARMA process with non i.i.d. innovations. When the innovations are not i.i.d., the robust covariance matrix estimator of Domowitz and White (1982) and White (1984) can be implemented. On letting

$\tilde{\gamma}_k = T^{-1}\sum_{t=g+1}^{T}(x_t - \mu)(x_{t-k} - \mu)$ and $\tilde{\rho}_k = \tilde{\gamma}_k/\tilde{\gamma}_0$ , where $\bar{x} = T^{-1}\sum_{t=1}^{T} x_t$ and $E(x_t) = \mu$. For the model described in equation (25), $x_t$ corresponds to $y_t^2$. The robust covariance matrix estimator of the sample autocorrelation can be obtained by first noting that

$$\sqrt{T}(\tilde{\rho} - \rho) = T^{-1/2}\,(1/\tilde{\gamma}_0)\sum_{t=g+1}^{T} Z_t,$$

where $Z_t$ is a $g \times 1$ vector defined by

$$Z_t = \begin{pmatrix} (x_t - \mu)(x_{t-1} - \mu) - \rho_1(\lambda)(x_t - \mu)^2 \\ \vdots \\ (x_t - \mu)(x_{t-g} - \mu) - \rho_g(\lambda)(x_t - \mu)^2 \end{pmatrix}.$$

Clearly, $E(Z_t) = 0$ and under suitable regularity conditions, $T^{-1/2}\sum_{t=1}^{T} Z_t \rightarrow N(0, V_z)$, where $V_z = \sum_{j=-\infty}^{\infty}\Gamma_j$, and $\Gamma_j = E(Z_t Z_{t-j}')$. Since $\sqrt{T}(\tilde{\rho} - \rho) = \tilde{\gamma}_0^{-1}\,T^{-1/2}\sum_{t=1}^{T} Z_t$ and $\tilde{\gamma}_0 \rightarrow \gamma_0$, it then follows that,

$$\sqrt{T}(\tilde{\rho} - \rho) \;\rightarrow\; N(\,0,\;\; \gamma_0^{-2}\,V_z\,),$$

In many practical applications $V_z$ can be consistently estimated by the Newey and West (1987) procedure by using,

$$\hat{V}_z = \hat{\Gamma}_0 + \sum_{j=1}^{q}\,(1 - \frac{j}{1+q})(\hat{\Gamma}_j + \hat{\Gamma}_j') \tag{28}$$

where $\hat{\Gamma}_j$ is a covariance matrix estimator at lag $j$, $\hat{\Gamma}_j = \frac{1}{T}\sum_{t=j+1}^{T} Z_t^*\; Z_{t-j}^{*\prime}$ with

$$Z_t^* = \begin{pmatrix} (x_t - \bar{x})(x_{t-1} - \bar{x}) - \rho_1(\hat{\lambda})(x_t - \bar{x})^2 \\ \vdots \\ (x_t - \bar{x})(x_{t-g} - \bar{x}) - \rho_g(\hat{\lambda})(x_t - \bar{x})^2 \end{pmatrix},$$

and $\hat{\lambda}$ is a consistent estimator of $\lambda$. The value of $q$ in equation (28) can be determined by a data dependent automatic rule provided by Newey and West (1994).

The $\text{MDE}_{NW}$ denoted by $\hat{\lambda}$, can be first obtained by setting the weighting matrix equal to the identity matrix. Hence the optimal weighting matrix for the MDE can be consistently estimated by

$$\hat{C}_{NW} = (\frac{1}{\hat{\gamma_0}})^2 \hat{V_z}, \tag{29}$$

where $\hat{\gamma_0} = T^{-1} \sum_{t=1}^{T} (x_t - \bar{x})^2$ and $\hat{C}_{NW}$ is a consistent estimator of C, given that $\hat{\gamma_0}$ and $\hat{V_z}$ consistently estimate $\gamma_0$ and $V_z$, respectively. The resulting estimate of C, i.e. $\hat{C}_{NW}$ can then be used in the quadratic form,

$$\text{Min } S = (\hat{\rho} - \rho(\lambda))' \hat{C}_{NW}^{-1} (\hat{\rho} - \rho(\lambda)) \tag{30}$$

to obtain the feasible MDE of the parameter vector $\lambda$ in the case of non i.i.d. innovations.

## 3. Simulation Results of MDE for ARMA(1,1) with i.i.d. Innovations

In the next section, MDE of the parameters of the GARCH(1,1) process from its ARMA(1,1) representation in the squared variable will be considered. The main additional complication of this model is the presence of non i.i.d. innovations. Before considering the non i.i.d. case, it is convenient to first consider the MDE applied to estimating the parameters in a classic time series setting of an ARMA process with i.i.d. innovations. In the ARMA(1,1) process,

$$y_t - \mu = \phi(y_{t-1} - \mu) + \epsilon_t - \theta \epsilon_{t-1}, \tag{31}$$

where $\epsilon_t$ is i.i.d.(0, $\sigma^2$), the vector of structural parameters, neglecting $\sigma^2$, is $\lambda =$

$[\phi \;\; \theta]'$.

The asymptotic variance of MDE for the ARMA(1,1) model are reported in chapter 2. As the number of autocorrelations, g, increases, the asymptotic variance of the MDE parameter estimates decreases and approaches to that of the MLE. If the absolute value of $\theta$ is not close to one, the MDE is seen to be remarkably efficient. In many cases, it is surprising to note that a remarkably small number of autocorrelations is necessary for the MDE to be as asymptotically efficient to two decimal places as that of the MLE under normality.

While our results in chapter 2 show that the MDE appears to be asymptotically efficient, it is also necessary to investigate its small sample performance. Table 19 presents some simulation results based on 1,000 replications, to evaluate the bias and MSE of the MDE applied to estimating the parameters of the ARMA(1,1) model, for the two points in the parameter space of $\phi = 0.8$, $\theta = 0.4$ and $\phi = 0.3$, $\theta = 0.6$. The total number of observations (T) is equal to 500 and there are 1,000 replications for each design. The parameters of the ARMA(1,1) process are estimated by:

(i) MDE using Bartlett's method to calculate the weighting matrix, where the number of autocorrelations used in computing the MDE are either 2, 5, 10, 20, or 30.

(ii) MDE using a weighting matrix estimated by the Newey and West (1987) method where the number of autocorrelations used in computing the MDE are again either 2, 5, 10, 20, or 30.

(iii) MLE assuming Gaussian disturbances.

Table 20 presents the simulation results for the cases that $\phi = 0.5$, $\theta = 0.2$ and

$\phi$ = 0.1, $\theta$ =0.3, while g is set to either 2, 5, 10, 15, or 20. From the results in Table 19 and 20, it is apparent that both the MDEs have a very small parameter estimation bias for both designs. There are no obvious departures from randomness and consequently no clear biases in either direction.

The RMSE of the parameter estimates from the sample size of T = 500 are close, but not quite efficient when compared with the MLE. The MDE based on only g = 2 and particularly g = 5 autocorrelations are remarkably efficient for both the parameter designs considered. The RMSE of the MLE presented in the table are calculated from the 1000 replications. Alternatively these RMSE could be compared with those derived analytically from the theoetical limiting distribution using the formula given in chapter 2. For example, in the case of $\phi$ = 0.8 and $\theta$ = 0.4 the theoretical RMSE are .0456 and .0697 compared with .0494 and .0743 for the RMSE of the estimates of $\phi$ and $\theta$ respectively. From Table 19 it appears that the MDE based on Bartlett's formula performs slightly better than the MDE using the Newey and West method of estimating the C matrix in equation (30). When the d.g.p. is $\phi$ = 0.8 and $\theta$ = 0.4, the MDE based on the first 5 autocorrelations performs better than that using more than 10 autocorrelations. For the case that $\phi$ = 0.3 and $\theta$ = 0.6, the MDE using 10 autocorrelations performs better than when more autocorrelations are used.

Another interesting feature of Table 19 and 20 is that the MDE based on g=5 autocorrelations, for the design of $\phi$ = 0.8 and $\theta$ = 0.4, have lower RMSE than the MDE based on g = 2, 10, 20, and 30 autocorrelations. Althought large sample theory suggests that it is better to include as many moments as possible in the estimation procedure, the simulation results show that it seems not to be true at the sample size analyzed here. This apparent trade-off occurs between the information used in the estimator as defined by the number of moments being used, and the quality of the objective function, as measured by the precision of the estimated weighting matrix. A similar deterioration in the quality of the GMM estimation method has been reported in cases where the number of identifying moment restrictions is increased beyond a certain level. This result is also noted by Chung and Schmidt (1996) and Andersen and Sorensen (1996).

## 4. Simulation Results of MDE for GARCH(1,1) Process

As previously noted, the innovation sequence $v_t$ in the ARMA(1,1) model for $y_t^2$ is clearly not i.i.d. and hence estimation of the parameters by MDE will require formula (30) where the Newey West method is used to estimate the weighting matrix, or equivalently the covariance matrix of the sample autocorrelations of a realization of the process. In this section, we first discuss the autocorrelation function of the squared GARCH(1,1) process and then present some simulation results of the MDE using the two different methods of estimating the weighting matrix.

The autocorrelation function of $y_t^2$, given the data generating process of a GARCH(1,1) process, has been derived by Bollerslev (1988) and Ding and Granger (1996), who also derive the autocorrelation function of the IGARCH process. We now briefly summarize the autocorrelation function of the squared process. In particular, the derivation

Table 19: Simulated mean and root mean square error (RMSE) of MDE and MLE for ARMA(1,1) model: $y_t = \phi y_{t-1} + e_t - \theta e_{t-1}$, $e_t \sim iid\ N(0,1)$. T= 500 and total number of replications =1000.

| $\phi = 0.8$ | $\theta = 0.4$ | $\hat{\phi}$ | | $\hat{\theta}$ | |
|---|---|---|---|---|---|
| | g | mean | RMSE | mean | RMSE |
| $MDE_B$ | 2 | 0.7865 | 0.0662 | 0.3911 | 0.1045 |
| | 5 | 0.7890 | 0.0516 | 0.3780 | 0.0776 |
| | 10 | 0.7973 | 0.0534 | 0.3750 | 0.0829 |
| | 20 | 0.8109 | 0.0573 | 0.3709 | 0.0896 |
| | 30 | 0.8215 | 0.0633 | 0.3653 | 0.0976 |
| $MDE_{NW}$ | 2 | 0.7865 | 0.0662 | 0.3911 | 0.1045 |
| | 5 | 0.7825 | 0.0546 | 0.3797 | 0.0782 |
| | 10 | 0.7813 | 0.0575 | 0.3774 | 0.0830 |
| | 20 | 0.7785 | 0.0624 | 0.3768 | 0.0901 |
| | 30 | 0.7756 | 0.0655 | 0.3748 | 0.0975 |
| MLE | | 0.7917 | 0.0494 | 0.3905 | 0.0743 |

| $\phi = 0.3$ | $\theta = 0.6$ | $\hat{\phi}$ | | $\hat{\theta}$ | |
|---|---|---|---|---|---|
| | g | mean | RMSE | mean | RMSE |
| $MDE_B$ | 2 | 0.3133 | 0.1499 | 0.6230 | 0.1540 |
| | 5 | 0.3068 | 0.1308 | 0.6075 | 0.1139 |
| | 10 | 0.2939 | 0.1309 | 0.5974 | 0.1070 |
| | 20 | 0.2836 | 0.1488 | 0.5953 | 0.1190 |
| | 30 | 0.2768 | 0.1612 | 0.5981 | 0.1311 |
| $MDE_{NW}$ | 2 | 0.3122 | 0.1504 | 0.6211 | 0.1540 |
| | 5 | 0.3091 | 0.1327 | 0.6089 | 0.1162 |
| | 10 | 0.2987 | 0.1340 | 0.5981 | 0.1127 |
| | 20 | 0.2932 | 0.1430 | 0.5939 | 0.1221 |
| | 30 | 0.2899 | 0.1502 | 0.5913 | 0.1332 |
| MLE | | 0.2945 | 0.1149 | 0.5977 | 0.0971 |

Note: The MDE using Bartlett's Formula to calculate the weighting matrix is denoted as $MDE_B$, whereas $MDE_{NW}$ denotes the cases that asymptotic variance of sample ACF is calculated by Newey and West method. The number of autocorrelation used in MDE is denoted as g.

Table 20: Simulated mean and root mean square error (RMSE) of MDE and MLE for ARMA(1,1) model: $y_t = \phi y_{t-1} + e_t - \theta e_{t-1}$, $e_t \sim iid\ N(0,1)$. T= 500 and total number of replications =1000.

| $\phi = 0.5$ | $\theta = 0.2$ | $\hat{\phi}$ | | $\hat{\theta}$ | |
|---|---|---|---|---|---|
| | g | mean | RMSE | mean | RMSE |
| $MDE_B$ | 2 | 0.4927 | 0.1356 | 0.1982 | 0.1526 |
| | 5 | 0.4848 | 0.1295 | 0.1818 | 0.1393 |
| | 10 | 0.4886 | 0.1394 | 0.1792 | 0.1496 |
| | 15 | 0.4933 | 0.1429 | 0.1785 | 0.1523 |
| | 20 | 0.4973 | 0.1454 | 0.1766 | 0.1568 |
| $MDE_{NW}$ | 2 | 0.4927 | 0.1356 | 0.1982 | 0.1526 |
| | 5 | 0.4815 | 0.1310 | 0.1826 | 0.1415 |
| | 10 | 0.4828 | 0.1347 | 0.1845 | 0.1473 |
| | 15 | 0.4842 | 0.1433 | 0.1876 | 0.1548 |
| | 20 | 0.4854 | 0.1430 | 0.1904 | 0.1578 |
| MLE | | 0.4941 | 0.1225 | 0.1951 | 0.1350 |

| $\phi = 0.1$ | $\theta = 0.3$ | $\hat{\phi}$ | | $\hat{\theta}$ | |
|---|---|---|---|---|---|
| | g | mean | RMSE | mean | RMSE |
| $MDE_B$ | 2 | 0.1312 | 0.2624 | 0.3381 | 0.2673 |
| | 5 | 0.1049 | 0.2288 | 0.3042 | 0.2189 |
| | 10 | 0.1005 | 0.2405 | 0.3030 | 0.2290 |
| | 15 | 0.0915 | 0.2528 | 0.2978 | 0.2402 |
| | 20 | 0.0914 | 0.2639 | 0.3008 | 0.2485 |
| $MDE_{NW}$ | 2 | 0.1312 | 0.2624 | 0.3384 | 0.2681 |
| | 5 | 0.0965 | 0.2347 | 0.2952 | 0.2284 |
| | 10 | 0.0945 | 0.2491 | 0.2915 | 0.2402 |
| | 15 | 0.0886 | 0.2549 | 0.2858 | 0.2453 |
| | 20 | 0.0889 | 0.2640 | 0.2867 | 0.2513 |
| MLE | | 0.0984 | 0.2264 | 0.2978 | 0.2202 |

of the autocorrelation function here does not rely on any distributional assumption. In general, only assumption 1 and 2 given below are required.

**Assumption 1** $u_t$ is i.i.d.(0,1) with $E(u_t^4) = \eta < \infty$.

Recall that the ARMA(1,1) representation of $y_t^2$ is

$$y_t^2 = \omega + (\alpha + \beta)y_{t-1}^2 + v_t - \beta v_{t-1}.$$

where $v_t = y_t^2 - \sigma_t^2$. Let $E_{t-1}(\cdot)$ denote mathematical expectation of the process conditional on the information available at time t-1. Clearly, $E_{t-1}v_t = 0$ and the conditional variance of $v_t$ is $E_{t-1}(v_t^2) = (\eta - 1)\sigma_t^4$. [1]

Since $E(y_t^4) = E(u_t^4)E(\sigma_t^4) = \eta E(\sigma_t^4)$, it follows that the fourth moment of $y_t$ is finite given that $E(\sigma_t^4)$ is finite. Let $\sigma^2$ denote the unconditional expection of $y_t^2$, so that $\sigma^2 = E(y_t^2) = E(\sigma_t^2)$. Given that $\alpha + \beta < 1$ , we have $\omega = (1 - \alpha - \beta)\sigma^2$. Taking the square of equation (24) yields

$$\begin{aligned} E(\sigma_t^4) &= E[(1 - \alpha - \beta)\sigma^2 + \alpha y_{t-1}^2 + \beta\sigma_{t-1}^2]^2 \\ &= \sigma^4[1 - (\alpha + \beta)^2] + E[\alpha^2 y_{t-1}^4 + \beta^2\sigma_{t-1}^4 + 2\alpha\beta y_{t-1}^2\sigma_{t-1}^2] \end{aligned}$$

Therefore,

$$E(\sigma_t^4) = \sigma^4[1 - (\alpha + \beta)^2] + (\eta\alpha^2 + \beta^2 + 2\alpha\beta)E(\sigma_{t-1}^4). \tag{32}$$

It follows that if $\eta\alpha^2 + 2\alpha\beta + \beta^2 < 1$ , $E(\sigma_t^4)$ exists, so does the fourth moment of $y_t$. Under the normality assumption of $u_t$, we have $\eta = 3$ and the condition for the fourth moment of $y_t$ being finite corresponds to $3\alpha^2 + 2\alpha\beta + \beta^2 < 1$, which is

[1] $Var_{t-1}(v_t) = E[(y_t^2 - \sigma_t^2)^2 \,|\Omega_{t-1}] = E[(y_t^4 - 2y_t^2\sigma_t^2 + \sigma_t^4) \,|\Omega_{t-1}] = (\eta - 1)\sigma_t^4$.

given in Bollerslev (1986). A very general result of the necessary conditions for the existence of the $y_t^{2k}$ moments for $k = 1, 2, \cdots$ is given in Terasvirta (1997).

**Assumption 2** $\eta\alpha^2 + 2\alpha\beta + \beta^2 \; < \; 1.$

If $\eta \geq 1$, assumption 2 also implies that $\alpha + \beta < 1$. Under assumptions 1 and 2 we have

$$E(\sigma_t^4) = \frac{\sigma^4 \left[1 - (\alpha + \beta)^2\right]}{1 - (\eta\alpha^2 + \beta^2 + 2\alpha\beta)}. \tag{33}$$

The following results of the autocorrelation and autocovariance functions of $y_t^2$ are based on the above two assumptions. First note that $\gamma_0 = E(y_t^2 - \sigma^2)^2 = E(y_t^4 - 2y_t^2\sigma^2 + \sigma^4)$, so that

$$\gamma_0 = \eta E(\sigma_t^4) - \sigma^4 \tag{34}$$

Using the result given in equation (33) yields

$$\gamma_0 = \frac{(\eta - 1)(1 - \beta^2 - 2\alpha\beta)}{1 - (\eta\alpha^2 + \beta^2 + 2\alpha\beta)} \sigma^4.$$

Rearranging the ARMA(1,1) equation of $y_t^2$ yields

$$y_t^2 - \sigma^2 = (\alpha + \beta)(y_{t-1}^2 - \sigma^2) + v_t - \beta v_{t-1}.$$

Therefore,

$$\gamma_1 = (\alpha + \beta)\gamma_0 + E[(y_{t-1}^2 - \sigma^2)v_t] - \beta E[(y_{t-1}^2 - \sigma^2)v_{t-1}].$$

Using the result that $E[\ (y_{t-j}^2 - \sigma^2)\ v_t\ ] = 0$ for $j \geq 1$, we obtain

$$\gamma_1 \;\; = \;\; (\alpha + \beta)\gamma_0 - \beta\ [\ \eta E(\sigma_t^4) - E(\sigma_t^4)\ ] \tag{35}$$

and

$$\gamma_k = (\alpha+\beta)\gamma_{k-1}, \quad \text{for} \quad k \geq 2.$$

Substituting the result given in equation (33) into the above equation we have

$$\gamma_1 = \frac{(\eta-1)\alpha(1-\alpha\beta-\beta^2)}{1-(\eta\alpha^2+\beta^2+2\alpha\beta)}\,\sigma^4. \tag{36}$$

It follows that under assumptions 1 and 2, the autocorrelation function of $y_t^2$ from the GARCH(1,1) process is

$$\rho_1 = (\alpha + \frac{\alpha^2\beta}{1-2\alpha\beta-\beta^2}), \tag{37}$$

and

$$\rho_k = (\alpha + \frac{\alpha^2\beta}{1-2\alpha\beta-\beta^2})(\alpha+\beta)^{k-1} \quad \text{for } k \geq 2. \tag{38}$$

Therefore, similar to the autocorrelation function of the standard ARMA(1,1) process, the autocorrelation function of $y_t^2$ from the GARCH(1,1) process decreases exponentially. For the standard ARMA model: $(1-\phi L)y_t = (1-\theta L)\epsilon_t$, where $\epsilon_t$ is white noise, we have $\rho_1 = \frac{(1-\phi\theta)(\phi-\theta)}{1+\theta^2-2\phi\theta}$, and $\rho_k = \rho_{k-1}\phi$ for $k \geq 2$. By substituting $\phi = \alpha+\beta$ and $\theta = \beta$ into the above equation, we have the same results of the autocorrelation functions as given in equations (37) and (38), given that $E(y_t^4)$ is finite.

In sum, the results given in equations (37) and (38) do not rely on any specific distributional assumptions. In particular, if the standardized innovation ($u_t$) of GARCH(1,1) process is iid(0, $\sigma^2$) with finite fourth moment and $E(y_t^4)$ is finite, the autocorrelation functions of the squared process are defined by equations (37) and (38).

In much high frequency data it is unclear if the unconditional fourth moment of $y_t$ exists. In the appendix we present some results of the autocorrelation function of

$y_t^2$ when $E(y_t^4)$ does not exist. [2]

In order to assess the relative small sample performance of the MDE applied to estimating the GARCH(1,1) model, a detailed Monte Carlo study was carried out. The data generating process was a GARCH(1,1) model with NID(0,1) standardized innovations. The parameter values are required to satisfy the constrain that $3\alpha^2 + 2\alpha\beta + \beta^2 < 1$, so that the fourth moment of $y_t$ is finite. The total number of observations, T, was set at 1,000. In each of the 1,000 replications 10,000 initial values were generated to avoid startup problems. For each replication, the following statistical procedures were calculated:

(i) Equation (25) was estimated by MDE through minimizing the criteria function given in equation (27). The weighting matrix was formed from $W = \hat{C}^{-1}$ where the $(i, j)$th element of $\hat{C}$ was computed from equation (26). Hence this simple version of the MDE neglects the non i.i.d. nature of the disturbances and is known to be sub optimal. However, this version of the MDE is extremely easy to compute and there is some interest in knowing its properties in this non standard situation.

(ii) Equation (25) was again estimated by MDE through using the criteria function given by equation (30). The implementation of this method requires the weighting matrix, W to be set to $\hat{C}_{NW}^{-1}$ and is most easily done by initially estimating $\hat{\lambda}$ through using an arbitrary weighting matrix such as the identity matrix and the resulting new estimate used to provide a new update of $\hat{C}_{NW}$. For the various repli-

[2]Under the assumption of normality of $u_t$, Ding and Granger (1996) also show that if the process starts at a very long time ago and $3\alpha^2 + 2\alpha\beta + \beta^2 \geq 1$ but $\alpha + \beta \leq 1$, then the autocorrelation function of $y_t^2$ can be approximated by, $\rho_k = [\alpha + (1/3)\beta](\alpha + \beta)^{k-1}$. They also show that for the integrated GARCH model, $\rho_k = (1/3)[1 + 2\alpha](1 + 2\alpha^2)^{-k/2}$.

cations and the parameter values considered in this study, the number of iterations required to estimate the weighting matrix was quite small, and generally less than four. Although this practical version of the MDE requires the use of iterative procedures, the total number of computer operations is still much less than that of MLE.

(iii) Equations (23) and (24) were estimated by assuming a Gaussian density.

## 4.1 Monte Carlo results of MDE of GARCH(1,1) with NID innovations

Table 21 reports the results of the simulation study and are based on 1,000 replications for each quantity. Both versions of the MDE are calculated from using the first 2, 5, 10, 20,30, and 40 autocorrelations.

The parameter values are set very close to the estimation results of hourly exchange rate data reported in Baillie and Bollerslev (1991). The MDE using weighting matrix calculated by Bartlett's formula is denoted as $MDE_B$, whereas the MDE using weighting matrix calculated by Newey and West method is denoted as $MDE_{NW}$. The only difference in these two methods is the way of calculating the optimal weighting matrix.

We find that the standard deviation of MDE using the Bartlett weighting matrix is higher than that of MDE using the Newey and West covariance matrix estimator. It might be caused by the use of suboptimal weighting matrix. This result is expected since the errors in the ARMA(1,1) process for the squared observation are not i.i.d.

For the cases of $MDE_{NW}$, when $\alpha$=0.2 and $\beta = 0.6$, the root mean square error (RMSE) of MDEs using 20 and 30 autocorrelations are lower than those of MDE using 5 and 10 autocorrelations. The improvement in efficiency from g=5 to g=10 is quiet large. Similarly, when g is increased from 10 to 20, the RMSE is decreased for both $\hat{\alpha}_{MDE}$ and $\hat{\beta}_{MDE}$. However, there is little change in the RMSE when the number of autocorrelations are more than 20.

Table 22 presents the results for two sets of parameter values: $\alpha = 0.1$, $\beta = 0.8$ and $\alpha = 0.05$, $\beta = 0.92$. This simulation design corresponds to the cases that $\alpha + \beta$ is close to 1, but the fourth moment of $y_t$ still exists. These cases represent the GARCH model with strong persistent in volatility. These two sets of parameter values are very close to the estimation results of the exchange rates of U.S. dollar versus the British pound and Deutschmark for a total number of 1,245 observations reported in Bollerslev (1987). The Monte Carlo results indicate that the optimal number of autocorrelations at the sample size of 1000 is around 30 or 40. Hence, it suggests when $\alpha + \beta$ is close to 1, a larger number of autocorrelations is required to be used in the MDE. Also, both $\hat{\alpha}_{MDE}$ and $\hat{\beta}_{MDE}$ have small downward biases, which diminish as the number of autocorrelations increases. Table 23 presents the simulation results for the case of T=5,000. When $\alpha = 0.2$ and $\beta = 0.6$, the RMSE of $\hat{\alpha}_{MDE}$ using the first 5 autocorrelations is almost the same as that of using either 10 or 20 autocorrelations. The RMSE of $\hat{\beta}^{(20)}_{MDE}$ (the MDE using the first 20 autocorrelations) appears to be lower than that of $\hat{\beta}^{(10)}_{MDE}$. The RMSE of MLE is apparently lower than that of MDE for both $\alpha$ and $\beta$.

For comparison, we also present simulation results of MLE of $\alpha$ and $\beta$ in Table 21 and 22. The small sample properties of MLE for the GARCH models have been investigated by Lumsdaine (1995) and Bollerslev and Wooldridge (1992). Bollerslev, Engle, and Nelson (1994, p.2983) provide a summary of the small sample properties of MLE for the GARCH(1,1) process. In particular, Monte Carlo evidence suggests that the ML estimate of $\alpha + \beta$ is downward biased. This bias comes from a downward bias in $\hat{\beta}$, whereas $\hat{\alpha}$ is upward biased. Some results for the case of T=200 are presented in Table 1 of Bollerslev and Wooldridge (1992). Our results of the MLE are consistent to previous Monte Carlo studies of MLE of GARCH(1,1) process. When the sample size is equal to 1000, $\hat{\beta}_{MLE}$ is slightly downward biased. Clearly, the bias diminishes as the sample size increases. The simulation results reported in Table 23 show that this bias in $\hat{\beta}_{MLE}$ disappears when total number of observations is equal to 5,000.

The asymptotic standard error of MDE of GARCH(1,1) process can be approximated in a simulation environment. The approach utilizes the fact that D can be determined analytically, while the variance of sample autocorrelations given a certain sample size may be estimated from a set of arbitrage large simulated samples.

The data generating process is the usual GARCH(1,1) with the standardized innovations being NID(0,1), while the total number of observations is equal to 1,000. The GARCH(1,1) data series is simulated 50,000 times. For each replication the sample autocorrelations up to 40 lags are calculated. The variance of the sample autocorrelation is estimated and denoted as $V^*$. Then the asymptotic variance of MDE when T=1000 is calculated as $( D'V^{*-1}D )^{-1}$.

Table 24 reports the simulation results of the standard deviation of the MDE when T=1,000. The number of moments are set to 2, 5, 10, 20, 30, 40. We present results

Table 21: Simulated mean and root mean square error (RMSE) of various forms of MDE and MLE for GARCH(1,1) process: $\epsilon_t = \sigma_t u_t$, $u_t \sim i.i.d.\ N(0,1)$ and $\sigma_t^2 = \omega + \alpha\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2$. T=1000 and total number of replications = 1000.

| $\alpha = 0.2$ | $\beta = 0.6$ | | | | |
|---|---|---|---|---|---|
| | | $\hat{\alpha}$ | | $\hat{\beta}$ | |
| | g | mean | RMSE | mean | RMSE |
| | 5 | 0.1968 | 0.0703 | 0.5711 | 0.1582 |
| | 10 | 0.2059 | 0.0683 | 0.5599 | 0.1467 |
| $MDE_B$ | 20 | 0.2123 | 0.0731 | 0.5566 | 0.1495 |
| | 30 | 0.2164 | 0.0753 | 0.5550 | 0.1538 |
| | 40 | 0.2208 | 0.0786 | 0.5532 | 0.1555 |
| | 5 | 0.1727 | 0.0687 | 0.5741 | 0.1689 |
| | 10 | 0.1745 | 0.0590 | 0.5521 | 0.1462 |
| $MDE_{NW}$ | 20 | 0.1730 | 0.0567 | 0.5604 | 0.1283 |
| | 30 | 0.1715 | 0.0585 | 0.5779 | 0.1207 |
| | 40 | 0.1732 | 0.0573 | 0.5832 | 0.1190 |
| MLE | | 0.2015 | 0.0433 | 0.5843 | 0.0914 |

| $\alpha = 0.15$ | $\beta = 0.7$ | | | | |
|---|---|---|---|---|---|
| | | $\hat{\alpha}$ | | $\hat{\beta}$ | |
| | g | mean | RMSE | mean | RMSE |
| | 5 | 0.1460 | 0.0634 | 0.6714 | 0.1657 |
| | 10 | 0.1597 | 0.0568 | 0.6560 | 0.1411 |
| $MDE_B$ | 20 | 0.1655 | 0.0602 | 0.6553 | 0.1363 |
| | 30 | 0.1703 | 0.0642 | 0.6517 | 0.1417 |
| | 40 | 0.1738 | 0.0676 | 0.6512 | 0.1424 |
| | 5 | 0.1284 | 0.0596 | 0.6694 | 0.1737 |
| | 10 | 0.1316 | 0.0474 | 0.6573 | 0.1380 |
| $MDE_{NW}$ | 20 | 0.1319 | 0.0453 | 0.6699 | 0.1232 |
| | 30 | 0.1326 | 0.0441 | 0.6845 | 0.1118 |
| | 40 | 0.1327 | 0.0445 | 0.6949 | 0.1102 |
| MLE | | 0.1496 | 0.0359 | 0.6853 | 0.0884 |

Note: $\omega$ is equal to $(1 - \alpha - \beta) * 0.02$. The MDE using Bartlett's Formula to calculate the weighting matrix is denoted as $MDE_B$, whereas $MDE_{NW}$ denotes the cases that asymptotic variance of sample autocorrelation is calculated by Newey and West method. The number of autocorrelation used in MDE is denoted as g.

Table 22: Simulated mean and root mean square error (RMSE) of various forms of MDE and MLE for GARCH(1,1) process: $\epsilon_t = \sigma_t u_t$, $u_t \sim i.i.d.\ N(0,1)$ and $\sigma_t^2 = \omega + \alpha\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2$. T=1000 and total number of replications = 1000.

| $\alpha = 0.1$ | $\beta = 0.8$ | | | | |
|---|---|---|---|---|---|
| | | $\hat{\alpha}$ | | $\hat{\beta}$ | |
| | g | mean | RMSE | mean | RMSE |
| | 5 | 0.0972 | 0.0572 | 0.7562 | 0.1910 |
| | 10 | 0.1103 | 0.0454 | 0.7542 | 0.1419 |
| $MDE_B$ | 20 | 0.1170 | 0.0457 | 0.7515 | 0.1227 |
| | 30 | 0.1202 | 0.0481 | 0.7508 | 0.1213 |
| | 40 | 0.1232 | 0.0508 | 0.7479 | 0.1296 |
| | 5 | 0.0847 | 0.0536 | 0.7514 | 0.1936 |
| | 10 | 0.0902 | 0.0379 | 0.7549 | 0.1345 |
| $MDE_{NW}$ | 20 | 0.0939 | 0.0331 | 0.7600 | 0.1137 |
| | 30 | 0.0955 | 0.0316 | 0.7687 | 0.1025 |
| | 40 | 0.0968 | 0.0325 | 0.7737 | 0.1058 |
| MLE | | 0.1032 | 0.0292 | 0.7763 | 0.0803 |

| $\alpha = 0.05$ | $\beta = 0.92$ | | | | |
|---|---|---|---|---|---|
| | | $\hat{\alpha}$ | | $\hat{\beta}$ | |
| | g | mean | RMSE | mean | RMSE |
| | 5 | 0.0461 | 0.0444 | 0.8322 | 0.2349 |
| | 10 | 0.0504 | 0.0366 | 0.8705 | 0.1670 |
| $MDE_B$ | 20 | 0.0581 | 0.0301 | 0.8834 | 0.1149 |
| | 30 | 0.0622 | 0.0290 | 0.8856 | 0.0939 |
| | 40 | 0.0655 | 0.0306 | 0.8828 | 0.0983 |
| | 5 | 0.0392 | 0.0430 | 0.8307 | 0.2417 |
| | 10 | 0.0390 | 0.0355 | 0.8777 | 0.1558 |
| $MDE_{NW}$ | 20 | 0.0428 | 0.0275 | 0.8969 | 0.1041 |
| | 30 | 0.0479 | 0.0222 | 0.8961 | 0.0802 |
| | 40 | 0.0505 | 0.0208 | 0.8932 | 0.0827 |
| MLE | | 0.0503 | 0.0180 | 0.9024 | 0.0636 |

Table 23: Simulated mean and root mean square error (RMSE) of various forms of MDE and MLE for GARCH(1,1) model: $y_t = \sigma_t u_t$, $u_t \sim$ *i.i.d.* N(0,1) and $\sigma_t^2 = \omega + \alpha y_{t-1}^2 + \beta\sigma_{t-1}^2$. T=5000 and total number of replications =1000.

| $\alpha = 0.2$ | $\beta = 0.6$ | $\hat{\alpha}$ | | $\hat{\beta}$ | |
|---|---|---|---|---|---|
| | g | mean | RMSE | mean | RMSE |
| | 5 | 0.1993 | 0.0398 | 0.5889 | 0.0792 |
| | 10 | 0.2018 | 0.0390 | 0.5862 | 0.0714 |
| $MDE_B$ | 20 | 0.2031 | 0.0394 | 0.5859 | 0.0712 |
| | 30 | 0.2042 | 0.0398 | 0.5859 | 0.0711 |
| | 40 | 0.2054 | 0.0403 | 0.5855 | 0.0716 |
| | 5 | 0.1858 | 0.0340 | 0.5863 | 0.0716 |
| | 10 | 0.1809 | 0.0337 | 0.5837 | 0.0635 |
| $MDE_{NW}$ | 20 | 0.1771 | 0.0349 | 0.5874 | 0.0596 |
| | 30 | 0.1753 | 0.0362 | 0.5947 | 0.0586 |
| | 40 | 0.1744 | 0.0371 | 0.6016 | 0.0574 |
| MLE | | 0.2002 | 0.0184 | 0.5962 | 0.0373 |

| $\alpha = 0.2$ | $\beta = 0.5$ | $\hat{\alpha}$ | | $\hat{\beta}$ | |
|---|---|---|---|---|---|
| | g | mean | RMSE | mean | RMSE |
| | 5 | 0.1994 | 0.0371 | 0.4924 | 0.0766 |
| | 10 | 0.2005 | 0.0371 | 0.4908 | 0.0736 |
| $MDE_B$ | 20 | 0.2014 | 0.0372 | 0.4911 | 0.0738 |
| | 30 | 0.2023 | 0.0376 | 0.4911 | 0.0743 |
| | 40 | 0.2032 | 0.0381 | 0.4909 | 0.0749 |
| | 5 | 0.1878 | 0.0347 | 0.4856 | 0.0729 |
| | 10 | 0.1837 | 0.0321 | 0.4802 | 0.0690 |
| $MDE_{NW}$ | 20 | 0.1779 | 0.0342 | 0.4883 | 0.0650 |
| | 30 | 0.1757 | 0.0359 | 0.4971 | 0.0655 |
| | 40 | 0.1751 | 0.0369 | 0.5026 | 0.0675 |
| MLE | | 0.1996 | 0.0200 | 0.4994 | 0.0439 |

Note: The MDE using Bartlett's Formula to calculate the weighting matrix is denoted as $MDE_B$, whereas $MDE_{NW}$ denotes the cases that asymptotic variance of sample autocorrelation is calculated by Newey and West method. The number of autocorrelation used in MDE is denoted as g.

of three sets of parameter values : $\alpha = 0.2$, $\beta = 0.5$, $\alpha = 0.2$, $\beta = 0.6$ , and $\alpha = 0.15$, $\beta = 0.7$. In general, as the number of autocorrelations used in MDE increased the asymptotic variance of MDE decreases. When g is increased from 2 to 5, the standard deviation of MDE reduced a lot. It is found that for the cases that $\beta = 0.5$ and 0.6, the asymptotic variance of the MDE using the first 20 autocorrelations is very close to that of the MDE using the first 30 autocorrelations. For the third case with $\beta$ =0.7, the improvement in efficiency is also very small when the number of autocorrelation is increased from 20 to 30. Hence, for the three cases examined here the incorporation of more than 30 autocorrelation in MDE is not likely to be very beneficial.

### 4.2 Monte Carlo results for GARCH(1,1) with leptokurtic errors

Many empirical applications of the GARCH models report that the assumption of conditional normality for the standardized innovation is usually not valid. Bollerslev (1987) provides evidence showing that the simple GARCH(1,1)-t model fit many of the speculative asset return series better than the GARCH models with conditional normal errors. Similarly, Nelson (1991) uses the generalized error distribution (GED) as the density function of the MLE in estimating the Exponential GARCH (EGARCH) model for stock returns data. Baillie and Myers (1991) find that the GARCH model with a conditional student t density provides a better description of commodity price changes than the GARCH model with conditional normality. However, the true conditional density is usually not known for many economic and financial time series. Alternatively, the Quasi-MLE (QMLE) method can be applied

Table 24: Simulated results of asymptotic standard deviation of MDE of GARCH(1,1) process for T=1000.

| | $\alpha =$ 0.2 | 0.2 | 0.15 |
| --- | --- | --- | --- |
| | $\beta =$ 0.5 | 0.6 | 0.7 |
| standard deviation of $\hat{\alpha}_{MDE}$ | | | |
| g= 2 | 0.0711 | 0.0918 | 0.0979 |
| 5 | 0.0591 | 0.0601 | 0.0527 |
| 10 | 0.0582 | 0.0560 | 0.0452 |
| 20 | 0.0582 | 0.0557 | 0.0442 |
| 30 | 0.0581 | 0.0557 | 0.0442 |
| 40 | 0.0581 | 0.0557 | 0.0442 |
| standard deviation of $\hat{\beta}_{MDE}$ | | | |
| g= 2 | 0.2821 | 0.3238 | 0.3913 |
| 5 | 0.1456 | 0.1328 | 0.1348 |
| 10 | 0.1358 | 0.1077 | 0.0936 |
| 20 | 0.1356 | 0.1059 | 0.0876 |
| 30 | 0.1356 | 0.1058 | 0.0873 |
| 40 | 0.1356 | 0.1058 | 0.0873 |

to estimate the GARCH models. Bollerslev and Wooldridge (1992) investigate the properties of the QMLE and related test statistics, when a normal log-likelihood is maximized but the assumption of normality is violated and show that if the first two conditional moments are correctly specified, the QMLE is generally consistent and has a limiting normal distribution.

To compare the performance of the MDE and the QMLE for the GARCH model of data series exhibiting extreme departures from conditional Gaussianity, a simple simulation study very similar to those in section 4.1 is performed. The true data generating process (DGP) is the GARCH(1,1) process with innovations being either standardized t or chi-square distributed. The total number of observations is equal to 1,000 and there are 1,000 replications for each design.

Table 25 and 26 provide simulation results of the MDE and the QMLE for the GARCH(1,1) with the standardized innovations being a standardized t distribution with degree of freedom 5 ($t_{\nu=5}$), i.e. $u_t = \xi_t/\sqrt{5/3}$, where $\xi_t$ are i.i.d. $t_{\nu=5}$ variate. Provided that $\nu > 2$, the student $t$ variable has the population mean zero and variance given by $\nu/(\nu-2)$. If $\nu > 4$, the population fourth moment of a t variable is $3\nu/[(\nu-2)(\nu-4)]$. The variance and kurtosis of the $t_{\nu=5}$ are 5/3 and 9, respectively. The first portion of Table 25 reports the results of the MDE and QMLE when $\alpha = 0.1$ and $\beta = 0.6$. In general, when the number of autocorrelations used in MDE increases, the RMSE of MDE decreases. However, the decrease in RMSE is not significant when g is increased from 30 to 40. It is also found that the MDE of $\alpha$ when $g \geq 20$ ( $\hat{\alpha}_{MDE}$ ) has smaller RMSE than the QMLE of $\alpha$ ( $\hat{\alpha}_{MLE}$), while the RMSE of $\hat{\beta}_{MDE}$ and

Table 25: Simulated mean and root mean square error (RMSE) of various forms of MDE and QMLE for GARCH(1,1) process: $y_t = \sigma_t u_t$, $\sigma_t^2 = \omega + \alpha y_{t-1}^2 + \beta \sigma_{t-1}^2$. where $u_t$ is i.i.d. standardized $t_{v=5}$. T=1000 and total number of replications = 1000.

| $\alpha = 0.1$ | $\beta = 0.6$ | | | | |
|---|---|---|---|---|---|
| | | $\hat{\alpha}$ | | $\hat{\beta}$ | |
| | g | mean | RMSE | mean | RMSE |
| | 5 | 0.1001 | 0.0773 | 0.5234 | 0.3139 |
| | 10 | 0.1076 | 0.0762 | 0.5135 | 0.2810 |
| $MDE_B$ | 20 | 0.1106 | 0.0780 | 0.5079 | 0.2815 |
| | 30 | 0.1133 | 0.0809 | 0.5060 | 0.2843 |
| | 40 | 0.1153 | 0.0834 | 0.5107 | 0.2822 |
| | 5 | 0.0840 | 0.0601 | 0.4909 | 0.3272 |
| | 10 | 0.0885 | 0.0543 | 0.4907 | 0.2745 |
| $MDE_{NW}$ | 20 | 0.0925 | 0.0549 | 0.5207 | 0.2530 |
| | 30 | 0.0959 | 0.0567 | 0.5420 | 0.2444 |
| | 40 | 0.0992 | 0.0575 | 0.5463 | 0.2438 |
| QMLE | | 0.1101 | 0.0652 | 0.5419 | 0.2369 |

| $\alpha = 0.2$ | $\beta = 0.6$ | | | | |
|---|---|---|---|---|---|
| | | $\hat{\alpha}$ | | $\hat{\beta}$ | |
| | g | mean | RMSE | mean | RMSE |
| | 5 | 0.1838 | 0.1160 | 0.5450 | 0.2525 |
| | 10 | 0.1996 | 0.1157 | 0.5292 | 0.2303 |
| $MDE_B$ | 20 | 0.2078 | 0.1223 | 0.5234 | 0.2302 |
| | 30 | 0.2123 | 0.1265 | 0.5236 | 0.2326 |
| | 40 | 0.2165 | 0.1287 | 0.5215 | 0.2341 |
| | 5 | 0.1488 | 0.0933 | 0.5584 | 0.2327 |
| | 10 | 0.1555 | 0.0834 | 0.5502 | 0.1963 |
| $MDE_{NW}$ | 20 | 0.1600 | 0.0814 | 0.5679 | 0.1740 |
| | 30 | 0.1633 | 0.0813 | 0.5792 | 0.1729 |
| | 40 | 0.1658 | 0.0817 | 0.5796 | 0.1763 |
| QMLE | | 0.2044 | 0.0780 | 0.5719 | 0.1305 |

Table 26: Simulated mean and root mean square error (RMSE) of various forms of MDE and QMLE for GARCH(1,1) process: $y_t = \sigma_t u_t$, $\sigma_t^2 = \omega + \alpha y_{t-1}^2 + \beta\sigma_{t-1}^2$, where $u_t$ is i.i.d. standardized $t_{v=5}$. T=1000 and total number of replications = 1000.

| $\alpha = 0.1$ | $\beta = 0.5$ | | | | |
|---|---|---|---|---|---|
| | | $\hat{\alpha}$ | | $\hat{\beta}$ | |
| | g | mean | RMSE | mean | RMSE |
| | 5 | 0.0930 | 0.0743 | 0.4631 | 0.3026 |
| | 10 | 0.0979 | 0.0746 | 0.4496 | 0.2834 |
| $MDE_B$ | 20 | 0.1011 | 0.0767 | 0.4428 | 0.2808 |
| | 30 | 0.1026 | 0.0791 | 0.4473 | 0.2829 |
| | 40 | 0.1045 | 0.0806 | 0.4453 | 0.2839 |
| | 5 | 0.0808 | 0.0585 | 0.4113 | 0.3167 |
| | 10 | 0.0851 | 0.0559 | 0.4101 | 0.2821 |
| $MDE_{NW}$ | 20 | 0.0893 | 0.0535 | 0.4498 | 0.2662 |
| | 30 | 0.0923 | 0.0539 | 0.4682 | 0.2609 |
| | 40 | 0.0957 | 0.0557 | 0.4759 | 0.2601 |
| QMLE | | 0.1041 | 0.0625 | 0.4484 | 0.2562 |

| $\alpha = 0.1$ | $\beta = 0.8$ | | | | |
|---|---|---|---|---|---|
| | | $\hat{\alpha}$ | | $\hat{\beta}$ | |
| | g | mean | RMSE | mean | RMSE |
| | 5 | 0.0916 | 0.0870 | 0.7321 | 0.2675 |
| | 10 | 0.1109 | 0.0825 | 0.7301 | 0.2143 |
| $MDE_B$ | 20 | 0.1198 | 0.0860 | 0.7248 | 0.2052 |
| | 30 | 0.1237 | 0.0901 | 0.7260 | 0.2006 |
| | 40 | 0.1270 | 0.0933 | 0.7239 | 0.2018 |
| | 5 | 0.0670 | 0.0680 | 0.7529 | 0.2493 |
| | 10 | 0.0754 | 0.0522 | 0.7694 | 0.1698 |
| $MDE_{NW}$ | 20 | 0.0858 | 0.0486 | 0.7642 | 0.1456 |
| | 30 | 0.0905 | 0.0471 | 0.7680 | 0.1343 |
| | 40 | 0.0922 | 0.0521 | 0.7729 | 0.1346 |
| QMLE | | 0.1083 | 0.0586 | 0.7635 | 0.1242 |

Table 27: Simulated mean and root mean square error (RMSE) of various forms of MDE and QMLE for GARCH(1,1) model: $y_t = \sigma_t u_t, \sigma_t^2 = \omega + \alpha y_{t-1}^2 + \beta \sigma_{t-1}^2$, where $u_t$ is i.i.d. standardized chi-square variable with degree of freedom 1, i.e. $u_t = (\xi_t - 1)/\sqrt{2}$, where $\xi_t$ is i.i.d. $\chi^2_{(1)}$ variates. $T$ = 1000 and total number of replications = 1000.

| $\alpha = 0.1$ | $\beta = 0.55$ | | | | |
|---|---|---|---|---|---|
| | | $\hat{\alpha}$ | | $\hat{\beta}$ | |
| | g | mean | RMSE | mean | RMSE |
| | 5 | 0.0882 | 0.0807 | 0.4845 | 0.3350 |
| | 10 | 0.0950 | 0.0803 | 0.4680 | 0.3114 |
| $MDE_B$ | 20 | 0.0979 | 0.0822 | 0.4675 | 0.3073 |
| | 30 | 0.1007 | 0.0839 | 0.4624 | 0.3071 |
| | 40 | 0.1034 | 0.0858 | 0.4585 | 0.3112 |
| | 5 | 0.0702 | 0.0650 | 0.4002 | 0.3572 |
| | 10 | 0.0775 | 0.0612 | 0.4315 | 0.2948 |
| $MDE_{NW}$ | 20 | 0.0839 | 0.0601 | 0.4808 | 0.2687 |
| | 30 | 0.0891 | 0.0610 | 0.5016 | 0.2649 |
| | 40 | 0.0932 | 0.0620 | 0.5097 | 0.2603 |
| QMLE | | 0.1157 | 0.0949 | 0.4858 | 0.2705 |

| $\alpha = 0.1$ | $\beta = 0.65$ | | | | |
|---|---|---|---|---|---|
| | | $\hat{\alpha}$ | | $\hat{\beta}$ | |
| | g | mean | RMSE | mean | RMSE |
| | 5 | 0.0881 | 0.0911 | 0.5606 | 0.3462 |
| | 10 | 0.0980 | 0.0907 | 0.5514 | 0.3127 |
| $MDE_B$ | 20 | 0.1033 | 0.0912 | 0.5454 | 0.3056 |
| | 30 | 0.1055 | 0.0931 | 0.5443 | 0.3037 |
| | 40 | 0.1079 | 0.0950 | 0.5416 | 0.3045 |
| | 5 | 0.0677 | 0.0715 | 0.4970 | 0.3704 |
| | 10 | 0.0750 | 0.0646 | 0.5274 | 0.3025 |
| $MDE_{NW}$ | 20 | 0.0832 | 0.0615 | 0.5693 | 0.2561 |
| | 30 | 0.0882 | 0.0615 | 0.5879 | 0.2452 |
| | 40 | 0.0923 | 0.0621 | 0.5970 | 0.2399 |
| QMLE | | 0.1166 | 0.0928 | 0.5816 | 0.2502 |

Note: The QMLE represents the cases that the simulated data series are estimated by usual MLE method while assuming normality of $u_t$.

Table 28: Simulated mean and root mean square error (RMSE) of various forms of MDE and QMLE for GARCH(1,1) model: $y_t = \sigma_t u_t, \sigma_t^2 = \omega + \alpha y_{t-1}^2 + \beta \sigma_{t-1}^2$, where $u_t$ is i.i.d. standardized chi-square with degree of freedom 2, i.e. $u_t = (\xi_t - 2)/2$, where $\xi_t$ is i.i.d. $\chi^2_{(2)}$ variates. $T$ = 1000 and total number of replications = 1000.

| $\alpha = 0.1$ | $\beta = 0.55$ | | | | |
|---|---|---|---|---|---|
| | | $\hat{\alpha}$ | | $\hat{\beta}$ | |
| | g | mean | RMSE | mean | RMSE |
| | 5 | 0.0939 | 0.0764 | 0.4729 | 0.3193 |
| | 10 | 0.1002 | 0.0759 | 0.4633 | 0.2914 |
| $MDE_B$ | 20 | 0.1031 | 0.0775 | 0.4572 | 0.2873 |
| | 30 | 0.1058 | 0.0797 | 0.4571 | 0.2904 |
| | 40 | 0.1080 | 0.0818 | 0.4565 | 0.2959 |
| | 5 | 0.0765 | 0.0623 | 0.3845 | 0.3602 |
| | 10 | 0.0842 | 0.0568 | 0.3957 | 0.3011 |
| $MDE_{NW}$ | 20 | 0.0925 | 0.0565 | 0.4613 | 0.2689 |
| | 30 | 0.0988 | 0.0580 | 0.4966 | 0.2539 |
| | 40 | 0.1041 | 0.0594 | 0.5080 | 0.2452 |
| QMLE | | 0.1099 | 0.0673 | 0.4691 | 0.2656 |

| $\alpha = 0.15$ | $\beta = 0.65$ | | | | |
|---|---|---|---|---|---|
| | | $\hat{\alpha}$ | | $\hat{\beta}$ | |
| | g | mean | RMSE | mean | RMSE |
| | 5 | 0.1308 | 0.0988 | 0.5859 | 0.2805 |
| | 10 | 0.1482 | 0.0988 | 0.5720 | 0.2471 |
| $MDE_B$ | 20 | 0.1554 | 0.1058 | 0.5630 | 0.2494 |
| | 30 | 0.1592 | 0.1102 | 0.5629 | 0.2515 |
| | 40 | 0.1623 | 0.1137 | 0.5625 | 0.2539 |
| | 5 | 0.1007 | 0.0832 | 0.5831 | 0.2796 |
| | 10 | 0.1114 | 0.0685 | 0.5689 | 0.2232 |
| $MDE_{NW}$ | 20 | 0.1190 | 0.0633 | 0.6121 | 0.1828 |
| | 30 | 0.1239 | 0.0624 | 0.6258 | 0.1774 |
| | 40 | 0.1288 | 0.0642 | 0.6328 | 0.1719 |
| QMLE | | 0.1626 | 0.0773 | 0.5998 | 0.1792 |

Note: The QMLE represents the cases that the simulated data series are estimated by usual MLE method while assuming normality of $u_t$.

$\hat{\beta}_{MLE}$ are very close. However, when $\alpha = 0.2$ and $\beta = 0.6$, the QMLE appears to have smaller RMSE than the MDE for both $\alpha$ and $\beta$.

Table 27 presents simulation results of MDE, where the innovations in the GARCH(1,1) process is assumed to have a standardized chi-square distribution with degree of freedom 1 ($\chi^2_{(1)}$), i.e. $u_t = (\xi_t - 1)/\sqrt{2}$, where $\xi$ is iid $\chi^2_{(1)}$. As given in Evans, Hastings, and Peacock (1993 p. 45), a chi-square variable with degree of freedom $\nu$ has skewness $= \frac{2^{1.5}}{\sqrt{\nu}}$ and Kurtosis $= 3 + \frac{12}{\nu}$, so that the error distribution of $\chi^2_{(1)}$ variable is asymmetric and has the coefficients of skewness and kurtosis equal to $2\sqrt{2}$ and 15, respectively. The first part of Table 27 presents the results for the case that $\alpha = 0.1$ and $\beta = 0.55$, while in the second part $\alpha$ and $\beta$ are set to 0.1 and 0.65, respectively. It is found that for both of the cases presented in Table 27 the MDE appears to be a better estimator than the QMLE. In particular, the MDE$_{NW}$ with g=20, 30, or 40 have smaller RMSE than the QMLE.

Very similar results are found in Table 28, where $u_t$ is assumed to follow a standardized $\chi^2_{(2)}$, i.e. $u_t = (\xi_t - 2)/\sqrt{4}$ where $\xi$ is iid $\chi^2_{(2)}$. This error distribution has variance =4, skewness=2 and kurtosis= 9. The $\chi^2_{(2)}$ and $t_{v=5}$ variates have the same first, second and fourth momoents, while the skewness coefficients are different. The results for $\alpha = 0.1$, $\beta = 0.55$ and $\alpha = 0.15$, $\beta = 0.65$ are reported. The MDE$_{NW}$ using either the first 20, 30 or 40 autocorrelations have smaller RMSE than the QMLE. For the two cases presented in Table 28 the MDE appears to perform better than the QMLE. In particular, the MDE using the first 30 autocorrelation has lower RMSE than the QMLE.

## 5. Example: Estimation of the GARCH Model Applied to Hourly Exchange Rate Data

In this section we apply the MDE to estimate a hourly deutschmark (DM) v.s. U.S. dollar exchange rate data from 0:00 a.m. January 2, 1986 through 11:00 a.m. July 15, 1986, which was also analyzed by Baillie and Bollerslev (1991). From Table III of Baillie and Bollerslev (1991 p. 573), it is found that the DM exchange rate data series has kurtosis equal to 8.171 and skewness equal to -0.3120, so that the conditional density of the GARCH model is very unlikely to be Gaussian normal. The estimation result of the MDE is given in Table 29. It seems that $\hat{\beta}_{MLE}$ and $\hat{\beta}_{MDE}$ are very close, while the difference between $\hat{\alpha}_{MLE}$ and $\hat{\alpha}_{MDE}$ is pretty large. Note that $\hat{\omega}_{MDE} = (\frac{1}{T}\sum_{t=1}^{T} y_t^2)(1 - \hat{\alpha}_{MDE} - \hat{\beta}_{MDE})$. Many studeis apply the MLE to estimate the GARCH model of speculative asset return data and find that the autocorrelations at various lags implied by the ML estimates are not conformable with the sample autocorrelations. See for example Jacquier, Polson, and Rossi (1994, Figure 1.). Table 30 reports the sample autocorrelations and values of autocorrelation function implied by the ML and minimum distance (MD) estimates. The results also indicate that the values of autocorrelation function implied by the ML estimates appear to be much higher than the sample autocorrelations of the squared observations. A similar result is reported by Jacquier, Polson, and Rossi (1994). This result serves to illustrate that the MLE put different weights on the moments conditions of the MDE.

Table 29: Estimation result of MDE and MLE of GARCH model of hourly exchange rate data

MLE:

| $\hat{\omega}_{MLE}$ | $\hat{\alpha}_{MLE}$ | $\hat{\beta}_{MLE}$ |
|---|---|---|
| 0.0679 | 0.2291 | 0.5125 |
| (0.0133) | (0.0463) | (0.684) |

MDE:

| $\hat{\omega}_{MDE}$ | $\hat{\alpha}_{MDE}$ | $\hat{\beta}_{MDE}$ |
|---|---|---|
| 0.0914 | 0.1317 | 0.4885 |
| | (0.0205) | (0.964) |

Note: The standard errors are given beneath the parameter estimates in parentheses. The total number of observations is 3189. The number of autocorrelations used in MDE is equal to 10.

## 6. Concluding Remarks

This chapter investigates the properties of the minimum distance estimator for GARCH(1,1) processes. The MDE is applied to estimate the parameters of a GARCH(1,1) model from the autocorrelations of the squared process which is known to follow an ARMA(1,1) process, but with non i.i.d. innovations.

As a benchmark, a simulation experiment is carried out to compare the small sample properties of the MDE using the Bartlett formula to calculate the weighting

Table 30: Sample ACF and the autocorrelations implied by the ML and minimum distance (MD) estimates of the squared hourly exchange rate

| | lags | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| SACF | 0.143 | 0.097 | 0.030 | 0.054 | 0.033 | 0.020 | -0.019 | -0.015 | -0.024 | -0.007 |
| $ACF_{MLE}$ | 0.283 | 0.210 | 0.155 | 0.115 | 0.085 | 0.063 | 0.047 | 0.035 | 0.026 | 0.019 |
| $ACF_{MDE}$ | 0.145 | 0.090 | 0.056 | 0.035 | 0.021 | 0.013 | 0.008 | 0.005 | 0.003 | 0.002 |

Note: SACF represents the sample autocorrelations. $ACF_{MLE}$ denotes the ACF fitted by ML estimates ($\hat{\alpha}_{MLE}$ and $\hat{\beta}_{MLE}$). $ACF_{MLE}$ is calculated by plugging $\alpha = 0.2291$ and $\beta = 0.5125$ into equation (37) and (38). Similarly, $ACF_{MDE}$ denotes the ACF fitted by MD estimates ($\hat{\alpha}_{MDE}$ and $\hat{\beta}_{MDE}$).

matrix ($MDE_B$) and the MDE using the Newey and West covariance matrix estimator to estimate the weighting matrix ($MDE_{NW}$), when the DGP is the ARMA(1,1) process with NID innovations. It is found that both $MDE_B$ and $MDE_{NW}$ perform quiet well in the case of ARMA(1,1) with NID innovations. On the other hand, for the GARCH(1,1) process the simulation results show that the RMSE of $MDE_B$ is higher than that of $MDE_{NW}$. This might be caused by the use of a suboptimal weighting matrix since the innovations in the ARMA(1,1) model for the squared GARCH observations are not i.i.d.

The relationship between the asymptotic standard deviation of MDE and the number of autocorrelations used are investigated in a simulation environment. The results show that as the number of autocorrelations used in the MDE increases, the asymptotic standard deviation of the MDE decreases.

The small sample properties of the MDE using first 5, 10, 20, 30 or 40 autocorrelations are investigated for the sample size of 1,000 based on 1,000 replications. The small sample results suggest that if $\alpha + \beta$ is close to one, a larger number of autocorrelations is needed to be used in the MDE.

For the ARMA(1,1) process with NID innovations, the simulation results reveal a deterioration in the quality of the MDE when the number of moment is increased beyond a certain level. Even thought asymptotic theory suggests that it is optimal to include as many moments as possible in the estimation procedure, this seems to be not true for the sample size analyzed here. We document that these results arise because of a fundamental trade-off between the number of moments used in the MDE and the precision of the estimated weighting matrix.

Many studies find that the conditional densities of the GARCH models for asset return series are usually not Gaussian normal and apply the QMLE method to estimate the GARCH models. A comparison of the small sample properties of the MDE and QMLE for the GARCH model is also provided. Monte Carlo results find that the MDE can be an attractive alternative to QMLE in terms of bias and RMSE, particularly for high frequency financial economic series which exhibit extreme departures from conditional Gaussianity.

# APPENDIX

# APPENDIX

### The autocorrelation function of $y_t^2$ when $E(y_t^4)$ does not exist

The autocorrelation function of $y_t^2$ under assumption 1 and that the fourth moment of $y_t$ doesn't exist, i.e. $\eta\alpha^2 + 2\alpha\beta + \beta^2 > 1$ can also be derived by using the similar method in Ding and Granger (1996).

**Assumption 3** $\eta\alpha^2 + 2\alpha\beta + \beta^2 > 1$ and $\alpha + \beta < 1$ .

We now briefly discuss the autocorrelation function of the squared GARCH(1,1) process when the fourth moment of $y_t$ does not exist. First note that

$$
\begin{aligned}
E(\sigma_t^4) &= \sigma^4[1-(\alpha+\beta)^2] + (\eta\alpha^2+\beta^2+2\alpha\beta)E(\sigma_{t-1}^4) \\
&= \sigma^4[1-(\alpha+\beta)^2]\,[1+(\eta\alpha^2+\beta^2+2\alpha\beta)+\cdots(\eta\alpha^2+\beta^2+2\alpha\beta)^t E(\sigma_0^4)] \\
&= \sigma^4\, A.
\end{aligned}
$$

Thus, $A = [1-(\alpha+\beta)^2]\,[1+(\eta\alpha^2+\beta^2+2\alpha\beta)+\cdots(\eta\alpha^2+\beta^2+2\alpha\beta)^t E(\sigma_0^4)]$. Under assumption 3, $\eta\alpha^2+2\alpha\beta+\beta^2 > 1$, $E(\sigma_t^4) \to \infty$ and $A \to \infty$ if $t$ is very large. Substituting equation (34) into equation (35) yields

$$
\begin{aligned}
\rho_1 &= \alpha+\beta-\beta(\eta-1)\frac{E(\sigma_t^4)}{\eta E(\sigma_t^4)-\sigma^4} \\
&= \alpha+\beta-\beta\frac{(\eta-1)}{\eta}-\beta\frac{(\eta-1)}{\eta}(\eta A-1)^{-1}
\end{aligned}
$$

If the process starts at a very long time ago $(\eta A-1)^{-1} \to 0$. Hence, given assumption 1 and 3, the autocorrelation function of $y_t^2$ can be approximated by,

$$\rho_k \approx (\alpha+\frac{1}{\eta}\beta)(\alpha+\beta)^{k-1},$$

which is also given by Ding and Granger (1996).

# CHAPTER 5

# CONCLUSION

This dissertation discusses the properties of the minimum distance estimator for ARMA processes and GARCH processes. The MDE has the advantage of imposing very little in terms of distributional assumptions on the innovation process. Under certain suitable regularity conditions, the MDE is $\sqrt{T}$ consistent and asymptotically normal. The MDE of AR(p) process using the first p autocorrelations is asymptotically equivalent to the MLE under normality. For the MA processes calculations show that as the number of autocorrelations ($g$) used in the MDE increases, the asymptotic variance of the MDE decreases. The MDE appears to be asymptotically as efficient as MLE under normality if $g$ is large enough. A theoretical justification of this result is provided for the MA(1) process. Very similar results are found in the numerical calculations of the asymptotic variance of MDE for the ARMA(p,q) processes and seasonal ARMA processes. Interestingly, for the MA(1) and ARMA(1,1) processes, if the absolute value of the moving average coefficient is not too large, the asymptotic variance of MDE based on first 3 to 5 autocorrelations is very close to that of MLE under normality. For the MA(1)-seasonal MA$(1)_s$ processes, given that the MA parameter is fixed, the higher the value of the seasonal MA parameter, the higher the number of autocorrelations is needed to guarantee that MDE is efficient. Calculation results also reveal that adding the autocorrelations at lags of the multiples of $s$, e.g. $\rho_s$, $\rho_{2s}$, $\rho_{3s}$, $\cdots$, to the MDE might have a larger improvement in efficiency than adding other autocorrelations.

Under the assumption that the innovations in the ARMA(p,q) process are iid, the Bartlett's formula can be applied to obtain the feasible estimates of covariance matrix ($\hat{C}_B$) of the sample autocorrelation. However, $\hat{C}_B$ is not a valid estimator of covariance matrix of the sample autocorrelation for the ARMA(1,1) model of the squared GARCH(1,1) process. We thus propose a robust covariance matrix estimator for the variance-covariance matrix of the sample autocorrelations based on Newey and West's (1987) method and is denoted as $\hat{C}_{NW}$. Two simulation experiments are conducted to compare the small sample properties of the MDE using $\hat{C}_B$ ($MDE_B$) and the MDE using $\hat{C}_{NW}$ ($MDE_{NW}$ ). The simulation results show that $MDE_{NW}$ and $MDE_B$ are quiet comparable to each other when the data generating process is the ARMA(1,1) process with NID innovations. However, Monte Carlo results indicate that for the GARCH(1,1) process $MDE_{NW}$ performs better than $MDE_B$. This result is expected since $MDE_B$ uses a suboptimal weighting matrix for the case of GARCH(1,1) process.

The Quasi-MLE (QMLE) method is usually invoked to estimate the GARCH models of many speculative asset return series, such as exchange rates and stock prices, because the assumption of conditional normality for the standardized innovations is difficult to justify in many empirical applications. Monte Carlo evidence shows that with certain conditional densities and over certain regions of the parameter space, the MDE can compare very favorably with Quasi-MLE in terms of parameter estimation bias and mean squared error. An application of the MDE to estimate a GARCH(1,1) model from high frequency exchange rate data is provided.

# List of References


1. Abraham, B. and J. Ledolter (1983), *Statistical Methods for Forecasting* John Wiley & Son, Inc.

2. Andersen T. G. and B.E. Sorensen (1996) "GMM Estimation of a Stochastic Volatility Model: a Monte Carlo Study", *Journal of Business and Economic Statistics*, 14, 328-352.

3. Andrews D. (1991), "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica* 59, 817-858.

4. Bartlett, M. S. (1946), "On the Theoretical Specification and Sampling Properties of Autocorrelated Time-series", *Journal of Royal Statistical Society* B8, 27-41.

5. Baillie R. T. and T. Bollerslev (1991), "Intra-Day and Inter-Market Volatility in Foreign Exchange Rates" *Review of Economic Studies*, 58, 565-585.

6. Baillie R. T. and R. J. Myers (1991), "Bivariate GARCH Estimation of the Optimal Commodity Futures Hedge" *Journal of Applied Econometrics*, 6, 109-124.

7. Bollerslev T. (1986), "Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics*, 31, 307-327.

8. Bollerslev T. (1987), "A Conditional Heteroskedastic Time Series Model for Speculative Prices and Rates of Return," *Review of Economics and Statistics*, 69, 542-547.

9. Bollerslev T. (1988), "On the Correlation Structure for the Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Time Series Analysis*, 9.

10. Bollerslev T., and J.M. Wooldridge (1992), "Quasi Maximum Likelihood Estimation and Inference in Dynamic Models with Time Varying Covariance", *Econometrics Review*, 11, 143-172.

11. Bollerslev T., Engle R. F., and Nelson D.B. (1994), "ARCH Models", in : R.F. Engle and D.L. McFadden , eds., *Handbook of Econometrics*, Vol IV. Amsterdam: North-Holland, 2959-3038.

12. Box, G.E.P., G.M. Jenkins, and G.C. Reinsel (1994), *Time Series Analysis: Forecasting and Control* 3rd ed., Prentice-Hall, Inc.

13. Breusch T., H. Qian, P. Schmidt, and D.J. Wyhowski (1997) "Redundancy of Moment Conditions," *Working Paper*, Michigan State University.

14. Brockwell, P.J. and R.A. Davis (1991), *Time Series : Theory and Methods*, 2nd ed., New York, Springer-Verlag.

15. Chung C.-F. and P. Schmidt (1996) "The Minimum Distance Estimator for Fractionally Integrated ARMA Models," *Working Paper*, Michigan State University.

16. Ding Z., and C.W.J. Granger (1996), "Modeling Volatility Persistence of Speculative Returns: a New Approach", *Journal of Econometrics,* 73, 185-215.

17. Domowitz I., and H. White (1982), "Misspecified Models with Dependent Observations", *Journal of Econometrics,* 20, 35-58.

18. Engle, R. (1982), "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of U.K. Inflation," *Econometrica* 50, 987-1008.

19. Durbin J. (1959), "Efficient Estimation of Parameters in Moving Average Models", *Biometrika* 46, 306-316.

20. Durbin J. (1960), "The Fitting of Time-series Models ", *Inst. Statist. Rev.* 28. 233-243.

21. Evans, M., N. Hastings, and B. Peacock (1993), *Statistical Distributions* 2nd ed., John Wiley & Sons, Inc.

22. Franses, P. H. (1996), *Periodicity and Stochastic Trend in Economic Time Series* Oxford University Press.

23. Fuller, A. W. (1996), *Introduction to Statistical Time Series* 2nd ed., John Wiley & Sons, Inc.

24. Galbraith J., and Zinde-Walsh V. (1994), "A Simple Noniterative Estimators for Moving Average Models", *Biometrika* 81, 143-155.

25. Galbraith J., and Zinde-Walsh V. (1997), "Measurement of the Quality of Autoregressive Approximation with Econometric Applications", McGill University working paper.

26. Hansen L. P. (1982), "Large Sample Properties of Generalized Method of Moments Estimators", *Econometrica* 50, 1029-1054.

27. Hylleberg, S. (ed.) (1993), *Modeling Seasonality* Oxford, Oxford University Press.

28. Jacquier E., N. Polson, and P. Rossi (1994) "Bayesian Analysis of Stochastic Volatility Models," *Journal of Business and Economic Statistics*, 12, 371-388.

29. Koreisha, S. and T. Pukkila (1990) "A Generalized Least Squares Approach for Autoregressive Moving Average Models", *Journal of Time Series Analysis* 11, 139-151.

30. Lumsdaine R. L. (1995), "Finite-Sample Properties of the Maximum Likelihood Estimator in GARCH(1,1) and IGARCH(1,1) Models: A Monte Carlo Investigation", *Journal of Business & Economic Statistics* 13, 1-10.

31. Mentz R. P. (1977), "Estimation in the First-Order Moving Average Model through the Finite Autoregressive Approximation", *Journal of Econometrics* 6, 225-236.

32. Nelson D. B. (1991), "Conditional Heteroskedasticity in Asset Returns: A New Approach", *Econometrica* 59, 347-370.

33. Newey W. K. and K. D. West (1987), "A Simple, Positive Semidefinite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix", *Econometrica* 55, 703-708.

34. Newey W. K. and K. D. West (1994), "Automatic Lag Selection in Covariance Matrix Estimation", *Review of Economic Studies* 61, 631-654.

35. Romano J. P., and L. A. Thombs (1996), "Inference for Autocorrelations Under Weak Assumptions", *Journal of the American Statistical Association,* 91, 590-600.

36. Terasvirta, T. (1997), "Two Stylized Facts and the GARCH(1,1) model", Department of Economic Statistics, Stockholm School of Economics working paper.

37. Saikkonen P. (1986), "Asymptotic Properties of Some Preliminary Estimators for Autoregressive Moving Average Time Series Models", *Journal of Time Series Analysis* 7, 133-155.

38. Tieslau, M.A., P. Schmidt, and R.T. Baillie (1996) "A Minimum Distance Estimator for Long Memory Process", *Journal of Econometrics*, 71, 249-264.

39. White, H. (1984), *Asymptotic Theory for Econometricians*, Orlando, Academic Press.

40. Wooldridge J.M. (1994), "Estimation and Inference for Dependent Processes", in : R.F. Engle and D.L. McFadden , eds., *Handbook of Econometrics*, Vol IV. Amsterdam: North-Holland, 2639-2738.

41. Zinde-Walsh, V. (1996), "Relations Between Some Distance Measures for Stochastic Processes", McGill University working paper.