Economics Department

Economics Working Papers

The University of Auckland *Year* 1999

Empirical Characteristic Function in Time Series Estimation

John Knight* Jun Yu†

*

†University of Auckland, yujun@smu.edu.sg

Empirical Characteristic Function In Time Series Estimation*

November 10, 1999

John L. Knight
Department of Economics
The University of Western Ontario
London, Ontario
Canada N6A 5C2
Phone: 1-519-661-3489

and

Jun Yu
Department of Economics
The University of Auckland
Private Bag 92019
Auckland, New Zealand
Phone: 64-9-373-7999 extn. 5770

*We would like to thank Victor Aguirregabiria, Reg Kulperger, Brendan McCabe, Peter Phillips, Jim Talman, Jian Yang and Hao Yu for their comments and suggestions. We also acknowledge the UWO Department of Statistical and Actuarial Sciences for providing the computer facilities. Seminar participants at the University of Alberta, the University of Western Ontario (both the Department of Economics and the Department of Statistical and Actuarial Science), and the University of Auckland also provided valuable feedback on earlier versions of the paper. This work is supported by the Natural Sciences and Engineering Research Council of Canada (NSERC).

Proposed Ruuning Head:

EMPIRICAL CHARACTERISTIC FUNCTION ESTIMATION

Editorial Correspondence to:

Jun Yu
Department of Economics
The University of Auckland
Private Bag 92019
Auckland, New Zealand

Phone: 64-9-373-7999 extn. 5770
Fax: 64-9-373-7427
E-mail: j.yu@auckland.ac.nz

Abstract

Since the empirical characteristic function is the Fourier transformation of the empirical distribution function, it retains all the information in the sample but can overcome difficulties arising from the likelihood. This paper discusses an estimation method using the empirical characteristic function for stationary processes. Under some regularity conditions, the resulting estimators are shown to be consistent and asymptotically normal. The method is applied to estimate Gaussion ARMA models. The optimal weight functions and estimating equations are given for in detail. Monte Carlo evidence shows that the empirical characteristic function method can work as well as the exact maximum likelihood method and outperforms the conditional maximum likelihood method.

1 Introduction

Maximum likelihood (ML) estimation under parametric assumptions is one of the most widely used estimation methods. One reason is that it results in estimators which are consistent, asymptotically normal and asymptotically efficient under appropriate regularity conditions. To implement the maximum likelihood method, however, the likelihood function must be of a tractable form and sometimes is required to be bounded in the parameter space. Unfortunately, there are many processes in economics where the maximum likelihood approach is difficult to implement, both in the independently, identically distributed (i.i.d.) case and the dependent case. In the i.i.d. case, the processes sometimes have an unbounded likelihood function in the parameter space. Examples include the mixture of normals and switching regression models (see Titterington *et al* (1985) and references therein). In other examples, such as the Stable distributions and compound normal and log-normal models, however, the density functions of the processes can not be written in a closed form in the sense that it is not expressible in terms of known elementary functions. This problem also arises in the dependent case. Such examples include stochastic volatility (SV) models (Ghysels *et al* (1996)), ARCH-type models (Bollerslev *et al* (1992)), and processes which are compound Poisson-Normal and where the Poisson intensity is random, possibly dependent on past information in the series (Knight and Satchell (1997)). Some of these models have found wide use in macroeconomics and finance.

Although some processes have a closed form density, evaluation of the exact likelihood can be difficult for various reasons. For instance, in order to calculate the exact likelihood function of stationary Gaussion ARMA models, one may need to deal with the determinant and inverse of the covariance matrix (Zinde-Walsh (1988)). However, such calculations can be computationally very expensive or even infeasible for a large number of observations (Amemiya (1985)).[1]

[1] In estimating ARMA models, the exact maximum likelihood method is available by use of the

The usual response to such difficulties arising from the likelihood is to use alternative methods. For example, one might use some variant of the method of moments (Hansen (1982)), the conditional maximum likelihood (CML) (Bollerslev *et al* (1992)), the quasi-maximum likelihood (QML) (White (1982)), or the simulation based method (Danielsson (1994); Duffie and Singleton (1993)). Although all methods are consistent under regular conditions, some of them are not asymptotically efficient. Furthermore, the small sample properties for some of these methods may be unsatisfactory and some of them are computationally intensive. The present paper discusses another alternative, a method that uses the empirical characteristic function (ECF).

Initiated by Parzen (1962) , the ECF has been used in many areas of inference such as testing for goodness of fit (Fan (1996)), testing for independence (Feuerverger (1990)), testing for symmetry (Feuerverger and Mureika (1977)), and parameter estimation.[2] The main justification for the ECF method is that the characteristic function (CF) has a one-to-one correspondence with the distribution function, and hence the ECF retains all the information present in the sample. Theoretically, therefore, the inference based on the ECF should work as well as that based on the empirical distribution function. The theory for the ECF method in the i.i.d. case is complete (Feuerverger and Mureika (1977); Csörgő (1981)). Surprisingly, however, the dependent case has received little attention and consequently there is great scope for research.

The purpose of this paper is to discuss the ECF estimation method for stationary stochastic processes. The asymptotic properties of the ECF estimators for stationary processes are established and Monte Carlo studies show that in finite samples the ECF method performs very well for Gaussian ARMA models.

The paper is organized as follows. The next section briefly reviews what has been

Kalman Filter (see Hamilton (1994)).

[2]The references for parameter estimation include Feuerverger and McDunnough (1981a,1981b), Feuerverger (1990) , Heathcote (1977), Knight and Satchell (1996, 1997), Press (1972), Quandt and Ramsey (1978), Schmidt (1982), Paulson *et al* (1975) and Yao and Morgan (1996).

done for the ECF method in the dependent case, proposes the ECF method in more general framework, and obtains the asymptotic properties of the resulting estimators for stationary processes. In Section 3 we focus on the estimation of Gaussian ARMA models, where the maximum likelihood method, conditional maximum likelihood method and various ECF methods are discussed. To use the ECF method most efficiently, the optimal weight functions and the estimating equations for Gaussian ARMA models are derived in this section. Section 4 gives results from Monte Carlo studies. The appendix collects proofs of the theorems in the paper.

2 Literature Review and Asymptotic Properties

In this section, a literature review is first performed on the ECF method in the estimation of stationary stochastic processes. The basic idea for the ECF method is to minimize some distance measure between the ECF and the CF. While the literature in the i.i.d. case is extensive, very little research has been reported in the dependent case. From our knowledge, only two papers are relevant. We will review them in detail.

Let $\{y_j\}_{j=-\infty}^{\infty}$ be a univariate, stationary time series whose distribution depends upon a vector of unknown parameters, $\boldsymbol{\theta}$. We wish to estimate $\boldsymbol{\theta}$ from a finite realization $\{y_1, y_2, ..., y_T\}$. The overlapping blocks for $y_1, y_2, ..., y_T$ are defined as,

$$\boldsymbol{x}_j = (y_j, ..., y_{j+p})',\ j = 1, ..., T - p.$$

Then each block has p observations overlapping with the adjacent blocks. The CF of each block is defined as

$$c(\boldsymbol{r}; \boldsymbol{\theta}) = E(exp(i\boldsymbol{r}'\boldsymbol{x}_j)),$$

where $\boldsymbol{r} = (r^1, ..., r^{p+1})'$. The ECF is defined as

$$c_n(\boldsymbol{r}) = \frac{1}{n}\sum_{j=1}^{n} \exp(i\boldsymbol{r}'\boldsymbol{x}_j),$$

where $n = T - p$. Hence the ECF is the sample counterpart of the CF and contains the information of the data, while the CF contains the information of the parameters; $\boldsymbol{r}$ is the transformation variable.

The estimation procedure is to match the ECF with the CF. Unfortunately, only two papers have been reported. The first work is published by Feuerverger (1990) and the second one reported by Knight and Satchell (1996). Both papers propose to match the ECF with the CF over a grid of finite points and hence we call the procedure "the discrete ECF (DECF) method". Feuerverger (1990) proves that under some regularity conditions, the resulting estimators can achieve the Cramér-Rao lower bound if p is sufficiently large and the grid of points is sufficiently fine and extended. However, he has not applied the procedure to estimate any time-series model. Knight and Satchell (1996) detail the application of the DECF method to stationary stochastic processes and give a multi-step procedure. We review their procedure in detail. Firstly, choose q and an arbitrary set of vectors $(\boldsymbol{r}_1, \ldots, \boldsymbol{r}_q)$, where each vector is of length $p+1$. Secondly, define $V = (Re\, c(\boldsymbol{r}_1; \boldsymbol{\theta}), ..., Re\, c(\boldsymbol{r}_q; \boldsymbol{\theta}), Im\, c(\boldsymbol{r}_1; \boldsymbol{\theta}), ..., Im\, c(\boldsymbol{r}_q; \boldsymbol{\theta}))'$, and $V_n = (Re\, c_n(\boldsymbol{r}_1), ..., Re\, c_n(\boldsymbol{r}_q), Im\, c_n(\boldsymbol{r}_1), ..., Im\, c_n(\boldsymbol{r}_q))'$, and choose $\boldsymbol{r}$ to minimize $(V_n - V)'(V_n - V)$ to obtain a consistent estimate for Ω, say $\hat{\Omega}$, where Ω is the covariance matrix of V_n. Thirdly, choose $\boldsymbol{r}$ to minimize some measure of the asymptotic covariance matrix of the GLS estimators. Fourthly, based on the optimal $\boldsymbol{r}$ obtained at Step 3, repeat Step 1 to obtain another consistent estimate for Ω, say $\hat{\hat{\Omega}}$. Finally choose $\boldsymbol{\theta}$ to minimize $(V_n - V)'\hat{\hat{\Omega}}^{-1}(V_n - V)$. The explicit form for the covariance matrix Ω is given by Knight and Satchell (1997). Choosing $p = 2$, $q = 5$ and an arbitrary $(\boldsymbol{r}_1, \ldots, \boldsymbol{r}_q)$, Knight and Satchell (1996) use the DECF method to estimate an MA(1) model and perform a Monte Carlo simulation. They find that the DECF method is a viable alternative, however, the performance of the DECF method is dominated by that of the ML method.

The finding is not surprising and can be explained intuitively. Since matching the

ECF with the CF over a grid of finite points is equivalent to matching a finite number of moments, the DECF method is, in essence, equivalent to the GMM. Just like it is not obvious how many and which moments to choose for the GMM, the difficulties for the DECF method are how many and which $\boldsymbol{r}$'s one should use.

Observing the difficulties involved in the DECF method, for stationary stochastic processes, we propose the ECF method which minimizes the integral

$$I_n(\boldsymbol{\theta}) = \int \cdots \int |c_n(\boldsymbol{r}) - c(\boldsymbol{r};\boldsymbol{\theta})|^2 \, dG(\boldsymbol{r}), \tag{2.1}$$

or

$$I_n(\boldsymbol{\theta}) = \int \cdots \int |c_n(\boldsymbol{r}) - c(\boldsymbol{r};\boldsymbol{\theta})|^2 g(\boldsymbol{r}) \, dr^1 \cdots dr^{p+1}, \tag{2.2}$$

or solves the following estimating equation

$$\int \cdots \int w_{\boldsymbol{\theta}}(\boldsymbol{r})(c_n(\boldsymbol{r}) - c(\boldsymbol{r};\boldsymbol{\theta})) \, d\boldsymbol{r} = 0. \tag{2.3}$$

Under suitable regularity conditions these three procedures are equivalent. If the weight function $G(\boldsymbol{r})$ is chosen to be a step function, the procedure is indeed the DECF method proposed by Feuerverger (1990) and Knight and Satchell (1996). Hence the procedure we propose includes the DECF method as a special case. We can certainly choose an alternative weight. If a continuous weight function is used, the procedure basically matches all the moments continuously, including integer moments, fraction moments and irrational moments. In this sense, the procedure exploits more information in the sample. Another advantage of using a continuous weight function is that one no longer needs to choose the transformation variables, $\boldsymbol{r}$'s, because they are simply integrated out. In fact this is the procedure suggested by Paulson *et al* (1975) in the i.i.d. case. In this paper we refer to the ECF method with a continuous weight as "the continuous ECF (CECF) method". The simplest weight for the CECF method is probably $g(\boldsymbol{r}) = \mathbf{1}$ (or $G(\boldsymbol{r}) = \boldsymbol{r}$). This procedure is referred to as "the OLS of the continuous ECF (OLS-CECF)) method". We can also use "the weighted least square of the continuous ECF (WLS-CECF) method" by choosing $g(\boldsymbol{r})$ to be a non-equally weighted function such as an exponential function. Furthermore, by using the Parseval theorem, we can obtain an optimal weight function, $w^*_{\boldsymbol{\theta}}(\boldsymbol{r})$,

$$(\frac{1}{2\pi})^{p+1} \int \cdots \int \exp(-i\boldsymbol{r}'\boldsymbol{x}_j) \frac{\partial \log f(y_{j+p}|y_j, \ldots, y_{j+p-1})}{\partial \boldsymbol{\theta}} \, dy_j \ldots dy_{j+p}. \tag{2.4}$$

The weight is optimal in the sense that based on $w^*_{\boldsymbol{\theta}}(\boldsymbol{r})$ and Equation (2.3), the resulting estimator can achieve the Cramér-Rao lower bound when p is sufficiently large (Feuerverger (1990)). The procedure associated with the optimal weight is referred to as"the GLS of the continuous ECF (GLS-CECF) method".

To use the proposed ECF method, however, we have to first justify it. For simplicity of notation we will only deal with the estimators resulting from Equation (2.1), where we assume $G(\boldsymbol{r})$ is a nondecreasing weight function with finite total variation taken to be unity without loss of generality. Equation (2.1) consists of minimizing a distance function and hence bears some resemblance to the M-estimators first discussed in the i.i.d. case by Huber (1981) and extended to the dependent case by Martin and Yohai (1986). To obtain the asymptotic properties of the ECF estimators, however, we will adopt an approach suggested by Heathcote (1977) for the i.i.d. case.

For any fixed p, we define $\{\boldsymbol{x}_j\}$, $c(\boldsymbol{r};\boldsymbol{\theta})$ and $c_n(\boldsymbol{r})$ as before. Let $P_{\boldsymbol{\theta}}$ be the probability measure corresponding to the characteristic function $c(\boldsymbol{r};\boldsymbol{\theta})$. Furthermore, $\boldsymbol{\theta}_0$ denotes the true parameter. Finally we assume that the following regularity conditions hold.

Assumptions:

(A1) (Identifiability) $\boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_2 \Leftrightarrow P_{\boldsymbol{\theta}_1} \neq P_{\boldsymbol{\theta}_2}$.

(A2) $I_n(\boldsymbol{\theta})$ can be differentiated under the integral sign with respect to $\boldsymbol{\theta}$.

(A3) $\frac{\partial^2 c(\boldsymbol{r};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}$ exists and uniformly bounded by G-integrable function.

(A4) $0 < \int \cdots \int \frac{\partial c(\boldsymbol{r};\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}} \frac{\partial c(\boldsymbol{r};\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}'} dG(\boldsymbol{r}) < \infty$.

(A5) Sequence $\{y_j\}$ is strictly stationary and α-mixing with $\sum_{n=1}^{\infty} \alpha_1(n) < \infty$, where

$\alpha_1(n)$ is the mixing coefficient of $\{y_j\}$.[3]

(A6) $K(\boldsymbol{\theta})$ is a measurable function of $\boldsymbol{x}$ for all $\boldsymbol{\theta}$ and bounded, where

$$\begin{aligned} K(\boldsymbol{\theta}) &= \int \cdots \int \Big\{ (cos(\boldsymbol{r}'\boldsymbol{x}) - Re\, c(\boldsymbol{r};\boldsymbol{\theta})) \frac{\partial Re\, c(\boldsymbol{r};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \\ &\quad + (sin(\boldsymbol{r}'\boldsymbol{x}) - Im\, c(\boldsymbol{r};\boldsymbol{\theta})) \frac{\partial Im\, c(\boldsymbol{r};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \Big\} dG(\boldsymbol{r}). \end{aligned} \tag{2.5}$$

Theorem 2.1 *Suppose the conditions (A1)-(A6) hold. Then there exists a root* $\tilde{\boldsymbol{\theta}}_n$ *of the equation* $\frac{\partial I_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \mathbf{0}$ *that is strongly consistent and asymptotically normal, i.e.,*

$$\sqrt{n}(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow{D} N(0, B^{-1}(\boldsymbol{\theta}) A(\boldsymbol{\theta}) B^{-1}(\boldsymbol{\theta})),$$

where the expression of $B(\boldsymbol{\theta})$ *and* $A(\boldsymbol{\theta})$ *is given in Appendix A.*

Proof: See Appendix A.

3 Estimation of Gaussian ARMA Models

In this section, we will focus on the estimation of Gaussion ARMA models. Gaussion ARMA models are simple linear models and can be estimated by classical estimation methods, such as the exact maximum likelihood method and conditional maximum likelihood method. Alternatively, we can use the ECF method proposed in the last section to estimate them.

Consider the Gaussian ARMA(l, m) model

$$Y_t = \rho_1 Y_{t-1} + \cdots + \rho_l Y_{t-1-l} + \varepsilon_t - \phi_1 \varepsilon_{t-1} - \cdots - \phi_m \varepsilon_{t-1-m}, \tag{3.1}$$

where $\varepsilon_t \sim i.i.d. N(0, \sigma^2)$. Let $\boldsymbol{\theta} = (\sigma^2, \rho_1, \cdots, \rho_l, \phi_1, \cdots, \phi_m)'$ denote the population parameters to be estimated. $\{y_1, ..., y_T\}$ is a realization. Three facts about the ARMA(l, m) model are reviewed here. Firstly, some assumptions are needed for the stationary property of the model. For example, if $|\rho_1| < 1$, the ARMA$(1, 0)$ model, i.e.,

[3]For the definition of α-mixing or strong mixing, see Ibragimov and Linnik (1971).

the AR(1), is stationary. In this paper we assume all series to be stationary. Secondly, stationary Gaussian ARMA models satisfy the α-mixing condition and $\sum \alpha(n) < \infty$ with $\alpha(n)$ defined as the mixing coefficient. Finally, the maximum likelihood approach provides the most efficient estimators. Denote the covariance matrix of $\boldsymbol{y} = (y_1, ..., y_T)$ by $\sigma^2\Phi$. It is straightforward to obtain the log likelihood function,

$$\mathcal{L}(\boldsymbol{\theta}) = -\frac{T}{2}\log(2\pi) - \frac{1}{2}log|\Phi| - \frac{T}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\boldsymbol{y}'\Phi^{-1}\boldsymbol{y}. \quad (3.2)$$

Maximization of the log likelihood function results in the ML estimator (MLE). Let $\hat{\boldsymbol{\theta}}_n$ be the MLE of $\boldsymbol{\theta}$, then $\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \xrightarrow{D} N(\mathbf{0}, I^{-1}(\boldsymbol{\theta}))$, where $I^{-1}(\boldsymbol{\theta})$ is the Cramér-Rao lower bound. Hence the MLE is asymptotically most efficient.

In this section we discuss three ARMA models. The first ARMA model that we consider is the Gaussian MA(1) model defined by

$$Y_t = \varepsilon_t - \phi\varepsilon_{t-1}, \quad (3.3)$$

where $\varepsilon_t \sim i.i.d.N(0, \sigma^2)$, and $\boldsymbol{\theta} = (\sigma^2, \phi)$ is the population parameter. For the MA(1) model, we can obtain the exact inverse and the determinant of Φ (see Whittle (1983)). Consequently, it is straightforward to maximize (3.2) numerically and leads to the MLE. It should be noted, however, the exact inverse of Φ is complicated for high-order moving-average models. In practice, therefore, the conditional likelihood function is often maximized instead.

The second ARMA model that we consider is the Gaussian AR(1) model defined by

$$Y_t = \rho Y_{t-1} - \varepsilon_t, \quad (3.4)$$

where $\varepsilon_t \sim i.i.d.N(0, \sigma^2)$, and $\boldsymbol{\theta} = (\sigma^2, \rho)$ is the population parameter. One can also obtain the exact inverse and the determinant of Φ. Consider the first order conditions of maximizing (3.2). If we concentrate out σ^2, we have the following estimating equation for ρ,

$$\begin{aligned}
0 &= 2\rho(y_1^2 + y_T^2) + 2\rho(T + 1 + (1 - T)\rho^2) \sum_{t=1}^{T-1} y_i^2 \\
&\quad -(4\rho^2 - 2T + 2T\rho^2) \sum_{t=1}^{T-1} y_t y_{t+1}.
\end{aligned} \tag{3.5}$$

The real root of above cubic equation leads to the MLE of ρ.

Finally we consider the Gaussian ARMA$(1,1)$ model defined by

$$Y_t = \rho Y_{t-1} + \varepsilon_t - \phi\varepsilon_{t-1} \tag{3.6}$$

where $\varepsilon_t \sim i.i.d.N(0, \sigma^2)$, and $\boldsymbol{\theta} = (\sigma^2, \rho, \phi)$ is the population parameter. The AR(1) and the MA(1) are two special cases of the ARMA$(1,1)$ with $\phi = 0$ and $\rho = 0$ respectively. In contrast to AR(1) and MA(1), however, ARMA$(1,1)$ has no closed form for the inverse of Φ for general (ρ, ϕ). Inverting this $T \times T$ matrix is numerically intensive for a large value of T, hence implementation of the maximum likelihood method is time consuming for a large T. To overcome the difficulties involved in the inversion, the state space representation and Kalman Filter can be used to evaluate the exact likelihood. Alternatively, the conditional likelihood function is often maximized in practice. We discuss this method in detail.

It is common to obtain the conditional likelihood function conditional on both y and ε. One option is to set initial y and ε equal to their expected value. That is, $y_0 = 0$ and $\varepsilon_0 = 0$. The conditional likelihood function can be obtained by

$$\begin{aligned}
\mathcal{L}(\boldsymbol{\theta}) &= f_{Y_T,\ldots,Y_1|y_0=0,\varepsilon_0=0}(y_t, \ldots, y_1|y_0 = 0, \varepsilon_0 = 0) \\
&= -\frac{T}{2}\log(2\pi) - \frac{T}{2}\log\sigma^2 - \sum_{t=1}^{T} \frac{\varepsilon_t^2}{2\sigma^2},
\end{aligned} \tag{3.7}$$

where the sequence $\{\varepsilon_1, \cdots, \varepsilon_T\}$ can be calculated from the realization $\{y_1, \cdots, y_T\}$ by the following iterative formula

$$\varepsilon_t = y_t - \rho y_{t-1} + \phi\varepsilon_{t-1}. \tag{3.8}$$

Maximizing the conditional likelihood leads to the conditional maximum likelihood estimator (CMLE) and the procedure is asymptotically equivalent to the exact maximum likelihood because the initial conditions have negligible effect to the total likelihood when the sample size is very large.

Alternatively we can estimate the ARMA models using the ECF method. The CF for the ARMA models is $\exp(-\frac{\sigma^2}{2}\boldsymbol{r}'\Phi\boldsymbol{r})$. To apply the ECF method to estimate a time-series model, both p and the weight function need to be specified. Regarding the choice of p, there is a trade off between large p and small p. If we choose a large p, we should expect the moving blocks contain all the information in the original sequence. By doing so we lose no information and the estimation via the ECF based on the blocks should be efficient when an optimal weight function is used. In fact, according to the inversion theorem, we can obtain the density by inverting the CF of the blocks. Provided the Fourier inversion can be implemented efficiently, the ECF estimator must be asymptotically equivalent to the MLE. Unfortunately, such inversions are high dimensional integrations and sometimes cannot be simplified. Therefore, the procedure could be numerically infeasible. On the other hand, a smaller p could be chosen to make the procedure feasible, however, the $n(= T - p)$ period overlapping blocks may not retain all the important information of the series. Regarding the weight function, we can choose it to be a step function, as Knight and Satchell (1996) proposed. Alternatively we can use other weight functions. As long as the regularity conditions in Section 2 hold, the resulting estimators are strongly consistent and asymptotically normally distributed. In this paper three other weight functions are actually considered to estimate ARMA models. The simplest one is $g(\boldsymbol{r}) = \mathbf{1}$ in Equation (2.2) and is indeed the OLS-CECF method. We use the WLS-CECF method by choosing $g(\boldsymbol{r}) = \exp(-a\boldsymbol{r}'\boldsymbol{r})$, where a is a non-negative constant. The exponential function is chosen for two reasons. Firstly, it is a generalization of the weight proposed by Paulson *et al* (1975) and puts more weight on the interval around the origin, consistent with the recognition that the CF contains the most information around the origin. The second reason is for computational convenience. With the exponential weight for ARMA models, $I_n(\boldsymbol{\theta})$ can be expressed as known elementary functions. We state the result in the following theorem.

Theorem 3.1 *If $\{y_1, ..., y_T\}$ is a finite realization of the ARMA model defined by* (3.1) *with $\boldsymbol{x}_j$, $c(\boldsymbol{r};\boldsymbol{\theta})$ and $c_n(\boldsymbol{r})$ defined as before, there exits a closed form function to represent the following integral*

$$\int \cdots \int |c_n(\boldsymbol{r}) - c(\boldsymbol{r};\boldsymbol{\theta})|^2 \exp(-a\boldsymbol{r}'\boldsymbol{r})\, d\boldsymbol{r},$$

and the expression for this function is given by

$$constant - \frac{2}{n}\pi^{\frac{p+1}{2}}|A|^{-\frac{1}{2}} \sum_{j=1}^{n} \exp(-\frac{1}{4}\boldsymbol{x_j}'A^{-1}\boldsymbol{x_j}) + \pi^{\frac{p+1}{2}}|B|^{-\frac{1}{2}},$$

where $A = \frac{\sigma^2}{2}\Phi + a \times I$ and $B = \sigma^2\Phi + a \times I$ with I as an identity matrix.

Proof: See Appendix B.

If we choose $w_{\boldsymbol{\theta}}(\boldsymbol{r})$ in the equation (2.3) to be

$$(\frac{1}{2\pi})^{p+1} \int \cdots \int \exp(-i\boldsymbol{r}'\boldsymbol{x}_j) \frac{\partial \log f(y_{j+p}|y_j, \cdots, y_{j+p-1})}{\partial \boldsymbol{\theta}}\, dy_j \cdots dy_{j+P}, \tag{3.9}$$

and then solve the estimating equation

$$\textstyle\int c_n(\boldsymbol{r}) w_{\boldsymbol{\theta}}(\boldsymbol{r})\, d\boldsymbol{r} = 0, \tag{3.10}$$

we get the GLS-CECF estimator which is efficient provided p is large enough. In Theorem 3.2 and Corollary 3.1 we will give the optimal weight functions and estimating equations of the GLS-CECF method for the Gaussian ARMA models.

Theorem 3.2 $\{y_1, ..., y_T\}$ *is a finite realization of the ARMA model defined by* (3.1) *with $\boldsymbol{\theta} = (\sigma^2, \boldsymbol{\rho}) = (\sigma^2, \rho_1, \cdots, \rho_l, \phi_1, \cdots, \phi_m)$. The conditional density of $y_{j+p}|y_j, \ldots, y_{j+p-1}$ can be expressed as*

$$(y_{j+p}|y_j, \ldots, y_{j+p-1}) \sim N\left(f_1(\boldsymbol{\rho})y_j + \cdots + f_p(\boldsymbol{\rho})y_{j+p-1}, \sigma^2 g(\boldsymbol{\rho})\right),$$

where $g(\boldsymbol{\rho})$, $f_1(\boldsymbol{\rho})$, $\cdots$, $f_p(\boldsymbol{\rho}) \in C^1$. *Define*

$$A = \begin{pmatrix} f_1^2(\boldsymbol{\rho}) & f_1(\boldsymbol{\rho})f_2(\boldsymbol{\rho}) & \cdots & f_1(\boldsymbol{\rho})f_{p-1}(\boldsymbol{\rho}) & -f_1(\boldsymbol{\rho}) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ f_{p-1}(\boldsymbol{\rho})f_1(\boldsymbol{\rho}) & f_{p-1}(\boldsymbol{\rho})f_2(\boldsymbol{\rho}) & \cdots & f_{p-1}^2(\boldsymbol{\rho}) & -f_{p-1}(\boldsymbol{\rho}) \\ -f_1(\boldsymbol{\rho}) & -f_2(\boldsymbol{\rho}) & \cdots & -f_{p-1}(\boldsymbol{\rho}) & 1 \end{pmatrix}. \tag{3.11}$$

and $B_1 = \frac{\partial A}{\partial \rho_1}, \cdots, B_l = \frac{\partial A}{\partial \rho_l}, B_{l+1} = \frac{\partial A}{\partial \phi_1}, \cdots, B_{l+m} = \frac{\partial A}{\partial \phi_m}$. *Let* $M\Lambda M', H_k\Lambda_k H_k'$ *be the eigenvalue decomposition of* A, B_k *with* $k = 1, \cdots, l+m$. *And denote eigenvalues of* A *by* $\lambda^1, \cdots, \lambda^{p+1}$ *and those of* B_k *by* $\lambda_k^1, \cdots, \lambda_k^{p+1}$. *Then optimal weight functions are*

$$\begin{aligned} w_{\sigma^2}(\boldsymbol{r}) &= -\frac{1}{2\sigma^2}\delta(r^1)\cdots\delta(r^{p+1}) - \frac{1}{2\sigma^4 g(\boldsymbol{\rho})}[\lambda^1\delta''(s^1)\delta(s^2)\cdots\delta(s^{p+1}) \\ &\quad + \cdots + \lambda^{p+1}\delta(s^1)\cdots\delta(s^p)\delta''(s^{p+1})], \end{aligned} \tag{3.12}$$

and

$$\begin{aligned} w_{\boldsymbol{\rho}_k}(\boldsymbol{r}) &= -\frac{g_k(\boldsymbol{\rho})}{2g(\boldsymbol{\rho})}\delta(r^1)\cdots\delta(r^{p+1}) - \frac{g_k(\boldsymbol{\rho})}{2\sigma^2 g^2(\boldsymbol{\rho})}[\lambda_k^1\delta''(s^1)\delta(s^2)\cdots\delta(s^{p+1}) \\ &\quad + \cdots + \lambda_k^{p+1}\delta(s^1)\cdots\delta(s^p)\delta''(s^{p+1})] + \frac{1}{2\sigma^2 g(\boldsymbol{\rho})} \\ &\quad [\lambda_k^1\delta''(t_k^1)\delta(t_k^2)\cdots\delta(t_k^{p+1}) + \cdots + \lambda_k^{p+1}\delta(t_k^1)\cdots\delta(t_k^p)\delta''(t_k^{p+1})], \end{aligned} \tag{3.13}$$

where $(s^1, \cdots, s^{p+1})' = M'\boldsymbol{r}$ *and* $(t_k^1, \cdots, t_k^{p+1})' = H_k'\boldsymbol{r}$ *with* $k = 1, \cdots, l+m$. $\delta(\cdot)$ *is the Dirac delta function and* $g_k(\boldsymbol{\rho}) = \partial g(\boldsymbol{\rho})/\partial \boldsymbol{\rho}_k$.[4]

Proof: See Appendix C.

Corollary 3.1 *Using the optimal weight function* $w_{\boldsymbol{\theta}}(\boldsymbol{r})$ *in* (2.3) *we generate the following estimating equations for* $\boldsymbol{\rho}$ *and* σ^2,

$$\sigma^2 = \frac{\bar{P}(\boldsymbol{\rho})}{g(\boldsymbol{\rho})}, \tag{3.14}$$

$$\frac{g_k(\boldsymbol{\rho})}{2g(\boldsymbol{\rho})} - \frac{g_k(\boldsymbol{\rho})}{2\sigma^2 g^2(\boldsymbol{\rho})}\bar{P}(\rho) + \frac{1}{2\sigma^2 g(\boldsymbol{\rho})}\bar{Q}_k(\rho) = 0, \tag{3.15}$$

[4]See Gel'Fan (1964) for more discussion about $\delta(\cdot)$ and its properties.

where

$$\begin{cases} \bar{P}(\rho) = \frac{1}{n}\sum_{j=1}^{n}[\lambda^1((M'\boldsymbol{x}_j)_{(1)})^2 + \cdots + \lambda^{p+1}((M'\boldsymbol{x}_j)_{(p+1)})^2] \\ \bar{Q}_k(\rho) = \frac{1}{n}\sum_{j=1}^{n}[\lambda_k^1((H_k'\boldsymbol{x}_j)_{(1)})^2 + \cdots + \lambda_k^{p+1}((H_k'\boldsymbol{x}_j)_{(p+1)})^2], \end{cases} \tag{3.16}$$

with $k = 1, \cdots, l+m$. *Combine* (3.14) *and* (3.15) *we have*

$$\begin{cases} \bar{Q}_1(\boldsymbol{\rho}) = 0 \\ \vdots \\ \bar{Q}_{l+m}(\boldsymbol{\rho}) = 0. \end{cases} \tag{3.17}$$

which determine the estimators of $\boldsymbol{\rho}$. *The estimator of* σ^2 *can be found by substituting the estimators of* $\boldsymbol{\rho}$ *back to Equation* (3.14).

Proof: See Appendix D.

Two examples are illustrated for the implementation of Theorem 3.2. For the MA(1) model defined by (3.3), if we choose $p = 1$, we have

$$y_{j+1}|y_j \sim N\left(-\frac{\phi}{1+\phi^2}y_j, \sigma^2\frac{1+\phi^2+\phi^4}{1+\phi^2}\right).$$

Hence,

$$A = \begin{pmatrix} \frac{\phi^2}{(1+\phi^2)^2} & \frac{\phi}{1+\phi^2} \\ \frac{\phi}{1+\phi^2} & 1 \end{pmatrix},$$

$$B = \begin{pmatrix} \frac{2\phi(1-\phi^2)}{(1+\phi^2)^3} & \frac{1-\phi^2}{(1+\phi^2)^2} \\ \frac{1-\phi^2}{(1+\phi^2)^2} & 0 \end{pmatrix},$$

and

$$g(\phi) = \frac{1+\phi^2+\phi^4}{1+\phi^2}.$$

Let λ_1^1, λ_1^2 be two eigenvalues of B, we have

$$\lambda_1^1 = \frac{\phi - \phi^5 + \sqrt{1 + 2\phi^2 - \phi^4 - 6\phi^6 - \phi^8 + 3\phi^{10} + \phi^{12}}}{1 + 4\phi^2 + 6\phi^4 + 4\phi^6 + \phi^8},$$

and

$$\lambda_1^2 = \frac{\phi - \phi^5 - \sqrt{1 + 2\phi^2 - \phi^4 - 6\phi^6 - \phi^8 + 3\phi^{10} + \phi^{12}}}{1 + 4\phi^2 + 6\phi^4 + 4\phi^6 + \phi^8}.$$

Then the estimating equation for ϕ is,

$$
\begin{aligned}
0 \;=\; & \sum_{j=1}^{n}\left\{(\lambda_1^{1^3}+\lambda_1^{2^3})\frac{(1+\phi^2)^4}{(1-\phi^2)^2}y_j^2+2(\lambda_1^{1^2}+\lambda_1^{2^2})\frac{(1+\phi^2)^2}{(1-\phi^2)}y_jy_{j+1}\right. \\
& \left.+(\lambda_1^1+\lambda_1^2)y_{j+1}^2\right\}. \qquad (3.18)
\end{aligned}
$$

And the estimating equation for σ^2 is,

$$
\sigma^2=\frac{1}{n}\sum_{j=1}^{n}\frac{[\phi y_j+(1+\phi^2)y_{j+1}]^2}{(1+\phi^2)(1+\phi^2+\phi^4)}. \qquad (3.19)
$$

In the second example we choose $p = 1$ for the AR(1) model defined by (3.4). Consequently,

$$
y_{j+1}|y_j \sim N\left(\rho y_j, \sigma^2\right).
$$

Hence,

$$
A=\begin{pmatrix} \rho^2 & -\rho \\ -\rho & 1 \end{pmatrix},
$$

and

$$
B=\begin{pmatrix} 2\rho & -1 \\ -1 & 0 \end{pmatrix}.
$$

Then the estimating equation of ρ is,

$$
(4\rho^3+3\rho)y_1^2+\rho y_{n+1}^2+4\rho(1+\rho^2)\sum_{i=2}^{n}y_i^2=2(2\rho^2+1)\sum_{i=1}^{n}y_iy_{i+1}. \qquad (3.20)
$$

The real root of above cubic equation leads to the ECF estimator of ρ. Note that Equation (3.20) is of different form to (3.5)

It should be stressed that, different from the DECF method, in the CECF method we do not need to choose the transformation variables, $\boldsymbol{r}$'s, since they are simply integrated out.

4 Monte Carlo Studies

Up to now, we have reviewed the DECF method proposed by Knight and Satchell (1996) and proposed the CECF method. Although Knight and Satchell (1996) estimate an MA(1) model by using the DECF method, the step function is arbitrarily chosen and hence the weight is not necessarily the best.

In this section the Monte Carlo studies are designed to compare the ECF estimators with the MLE for the MA(1) model and the AR(1) model, and with the CMLE for the ARMA(1, 1) model.

The first experiment involves 1,000 replications of generated data from a Gaussian MA(1) model with $\phi = 0.6$, $\sigma^2 = 1.0$ and the number of observations set at $T = 100$. To use the DECF method proposed by Knight and Satchell (1996), one has to determine the value of p, the value of q, and the criterion to obtain the optimal $\boldsymbol{r}_1, \cdots, \boldsymbol{r}_q$. In the first Monte Carlo study we choose p to be 2 since the autocovariances are zero in the MA(1) for lags greater than one and q to be 5 to guarantee that the number of elements in the vector V_n and V is larger than the number of parameters. In this multi-parameter case, we choose the criterion function in two ways, i.e., minimizing the determinant or the trace of the asymptotic covariance matrix of the GLS estimator. Without further restriction on $\boldsymbol{r}$'s, unfortunately, the minimization problem for obtaining optimal $\boldsymbol{r}$'s is 15 dimensions and hence numerically very expensive. To simplify the matter and also for $\boldsymbol{r}$'s to be sufficiently fine and extended we have chosen $\boldsymbol{r}$'s in several ways. For example, we choose them in the following two ways,

$$\boldsymbol{r} = \begin{bmatrix} -4\tau & -2\tau & -\tau & 0 & 2\tau \\ -3\tau & -\tau & 0 & \tau & 3\tau \\ -2\tau & 0 & \tau & 2\tau & 4\tau \end{bmatrix}, \tag{4.1}$$

or

$$\boldsymbol{r} = \begin{bmatrix} -6\tau & -3\tau & -\tau & \tau & 4\tau \\ -5\tau & -2\tau & 0 & 2\tau & 5\tau \\ -4\tau & -\tau & \tau & 3\tau & 6\tau \end{bmatrix}, \tag{4.2}$$

where the behavior of $\boldsymbol{r}$'s is determined by one parameter, i.e., τ.

In Table 1 we report the results from all four procedures of the DECF method. In c(1), $\boldsymbol{r}$'s are chosen to be (4.1) and the criterion function to be the determinant of the asymptotic covariance. In c(2) $\boldsymbol{r}$'s are chosen to be (4.1) and the criterion function to be the trace of the asymptotic covariance. In c(3) $\boldsymbol{r}$'s are chosen to be (4.2) and the criterion function to be the determinant of the asymptotic covariance. In c(4), $\boldsymbol{r}$'s are chosen to be (4.2) and the criterion function to be the trace of the asymptotic covariance. The MLE is presented as well.

A detailed examination of Table 1 reveals that the ECF estimates appear not too much dependent on the way the $\boldsymbol{r}$'s are chosen or the criterion function. Furthermore, the finite sample properties of the ECF estimates are dominated by those of the MLE. The finding is not surprising.

To use the CECF method, one needs to determine the value of p and the weight function. For all the cases we first choose p to be 2. Table 2 reports the results from the CECF method. In c(1), the OLS-CECF is used. In c(2), the WLS-CECF with the weight $\exp(-\boldsymbol{r}'\boldsymbol{r})$ is used. In c(3), the GLS-CECF with the optimal weight is used. The MLE is also reported for comparison.

From Table 2 it appears that both the OLS-CECF and the WLS-CECF are dominated by the GLS-CECF and ML. Moreover, the ML performs better than the GLS-CECF. This finding suggests that $p = 2$ is not large enough for the moving blocks to retain all the information in the original sequence. Comparing Table 2 with Table 1 we note that there appears to be an improvement over the DECF by the use of the GLS-CECF in terms of the mean square error.

In Table 3 and Table 4 we report the GLS-CECF estimates of ϕ and σ^2, however, the value of p has been increased to be $3, 4, \cdots, 10$. As we argued before, as p gets larger and larger, the GLS-CECF estimators should be getting closer and closer to the MLE. Therefore, we must expect a larger p to work better. This is confirmed by Table 3 and Table 4. From the two tables, however, we note that the GLS-CECF method with a small p can work quite well. For example, when comparing the GLS-CECF with $p = 6$ and the ML, we note that the GLS-CECF method performs as well as the

ML. For the two sets of the estimates of ϕ, we find that the variance and the skewness of them are almost identical, but the GLS-CECF provides a mean and median which are slightly closer to the true parameter value. Also see Figure 1 where the densities of the ML estimates and the GLS-CECF estimates (p=6) of ϕ are plotted. The two densities are visually close to each other.

The second experiment involves 1,000 replications of generated data from a Gaussion MA(1) model with $\phi = 0.6$, $\sigma^2 = 1.0$ and the number of observations set at $T = 1000$. This experiment is designed to demonstrate that with a small p both the OLS-CECF and the WLS-CECF method work well when we have a large number of observations. Table 5 and Table 6 reports the estimates of ϕ and σ^2. In c(1), the OLS-CECF with $p = 1$ is used. In c(2), the WLS-CECF with the weight $\exp(-\boldsymbol{r}'\boldsymbol{r})$ and $p = 1$ is used. In c(3), the OLS-CECF with $p = 2$ is used. In c(4), the WLS-CECF with the weight $\exp(-\boldsymbol{r}'\boldsymbol{r})$ and $p = 2$ is used. In c(5), the OLS-CECF with $p = 3$ is used. In c(6), the WLS-CECF with the weight $\exp(-\boldsymbol{r}'\boldsymbol{r})$ and $p = 3$ is used. As we expect, with everything else being equal, the larger the value of p, the better the performance of the resulting estimates. Also the WLS-CECF performs slightly better than the OLS-CECF. Furthermore, for $p = 3$ both the OLS-CECF and WLS-CECF seem to be a viable estimation method.

The third experiment involves 1,000 replications of generated data from a Gaussian AR(1) model with $\rho = 0.6$, $\sigma^2 = 1.0$ and the number of observation set at $T = 100$. With $p = 1$ the GLS-CECF method is used to estimate the AR(1) model and compared with the ML. See Table 7 and Figure 2. Interestingly, the two approaches provide almost identical results. The results can be explained by the Markov property of the AR(1) model. Since the joint density of the AR(1) model for a length T has the following property

$$f(x_1, x_2, \cdots, x_T) = f(x_1)\Pi_{j=2}^{T} f(x_j|x_{j-1}),$$

it is not surprising that one-period-overlapping blocks can preserve all the information in the original sequence.

The fourth experiment involves 1,000 replications of generated data from a Gaussian AR(1) model with $\rho = 0.6$, $\sigma^2 = 1.0$ and the number of observations set at $T = 1000$. This experiment is designed to demonstrate that with a small p both the OLS-CECF and the WLS-CECF method work well when we have a large number of observations. Table 8 reports the ECF estimates of ρ and σ^2. In c(1), the OLS-CECF with $p = 1$ is used. In c(2), the WLS-CECF with the weight $\exp(-\boldsymbol{r}'\boldsymbol{r})$ and $p = 1$ is used. In c(3), the OLS-CECF with $p = 2$ is used. In c(4), the WLS-CECF with the weight $\exp(-\boldsymbol{r}'\boldsymbol{r})$ and $p = 2$ is used. Note that there is no clear improvement by increasing p from 1 to 2. The finding can still be explained by the Markov property of the AR(1) model. The advantage of using the WLS-CECF method seems clear and both the OLS-CECF and WLS-CECF method seem to be a viable estimation method.

The fifth experiment involves 1,000 replications of generated data from a Gaussian ARMA$(1,1)$ model with $-\phi = \rho = 0.6$, $\sigma^2 = 1.0$ and the number of observations set at $T = 100$. It is easy to show that the inverse of $\boldsymbol{\Phi}$ for the ARMA$(1,1)$ model has no closed form even for $-\phi = \rho$. Consequently, this experiment is designed to compare the performance of the GLS-CECF method with that of the CML method. Table 9 and Figure 3 report the results where we choose $p = 2$, 3 for the GLS-CECF. Theoretically, the asymptotic variance of both the GLS-CECF estimate and the CMLE converge to the Cramér-Rao lower bound. In terms of the finite sample properties, from the Table 9, we note that the GLS-CECF seems to be a viable alternative to the CML. For example, for the estimates of ρ there appears a trade-off between the GLS-CECF with $p = 2$ and the CML. The mean and median of the GLS-CECF estimates are closer to the true parameter value and the GLS-CECF estimates are less skewed than the CMLE's while the variance of the GLS-CECF estimates is larger. However, comparing the GLS-CECF estimates with $p = 3$ to the CMLE's, we note that the variance and mean square error are almost identical, whereas the GLS-CECF estimates dominate the CMLE's in terms of the other statistics. For instance, the GLS-CECFestimates with $p = 3$ have a mean and median which are closer to the true parameter value and show less skewness. Similarly, for the estimates of σ^2 the GLS-CECF estimates have smaller bias.

The findings can be explained as follows. On the one hand, some initial conditions must be assumed in order to obtain the CMLE. If the initial assumption has an error, it is carried over into all the following stage by the recursive formula such as (3.8). Of course, the effect of such an error will diminish for the stationary models and thus the CMLE is asymptotically equivalent to the MLE. However, the effect may not be negligible for a small number of observations. On the other hand, no approximation is needed for the use of the ECF method.

The sixth experiment involves 1,000 replications of generated data from a Gaussian ARMA(1, 1) model with $-\phi = \rho = 0.6$, $\sigma^2 = 1.0$ and the number of observations set at $T = 1000$. This experiment is designed to demonstrate that with a small p both the OLS-CECF and the WLS-CECF method work well when we have a large number of observations. Table 10 and Table 11 report the estimates of ρ and σ^2. In c(1), the OLS-CECF with $p = 1$ is used. In c(2), the WLS-CECF with the weight $\exp(-\boldsymbol{r}'\boldsymbol{r})$ and $p = 1$ is used. In c(3) the OLS-CECF with $p = 2$ is used. In c(4), the WLS-CECF with the weight $\exp(-\boldsymbol{r}'\boldsymbol{r})$ and $p = 2$ is used. In c(5), the OLS-CECF with $p = 3$ is used. In $c(6)$, the WLS-CECF with the weight $\exp(-\boldsymbol{r}'\boldsymbol{r})$ and $p = 3$ is used. All cases seem to suggest that both the OLS-CECF and the WLS-CECF are viable method.

The seventh experiment involves 1,000 replications of generated data from a Gaussion ARMA(1, 1) model with $-\phi = \rho = 0.9$, $\sigma^2 = 1.0$ and the number of observations set at $T = 100$. This experiment is designed to compare the performance of the GLS-CECF method with that of the CMLE. Table 12 and Figure 4 report the results where we choose $p = 2, 3$ for the GLS-CECF. Compared with the sequences generated in the fifth experiment, the sequences in this experiment have much longer memory. This property will enlarge the difference between the exact likelihood and the condition likelihood. Consequently, the finite sample performance of the CMLE in this experiment will be worse. This is confirmed by Table 12. For example, the GLS-CECF estimates for both $p = 2$ and $p = 3$ perform better than the CMLE's. The bias, the variance and the mean square error of the GLS-CECF estimates are smaller. In particular, the mean square error of the CMLE's of σ^2 is about 22 times larger than that of the GLS-CECF estimates!

The advantage of using the ECF method is the following. Firstly, all the ECF methods presented in Section 3 do not require the calculation of the inverse of a $T \times T$ matrix either implicitly or explicitly. For some stationary processes, however, in order to use the exact maximum likelihood estimation method, such inversion is required but infeasible. Secondly, implementation of the exact maximum likelihood usually requires the closed form expression an boundedness for the likelihood. No such requirement is needed for the DECF method, the OLS-CECF method, and the WLS-CECF method. Instead, the CF must have a tractable form. Finally, even for the stationary Gaussion processes such as the MA(1) model whose covariance matrix can be inverted analytically, the exact maximum likelihood method is more computationally intensive than the GLS-CECF method. This is because the former method has to deal with a $T \times T$ matrix, while only a $(p+1) \times (p+1)$ matrix is considered in the latter case.

Appendix A

Proof of Theorem 2.1

According to Proposition 3.44 of White (1984), since $\{y_j\}$ is a stationary sequence, α-mixing condition of $\{y_j\}$ implies ergodicity. Following from Theorem 3.5.3. and 3.5.8. of Stout (1974), the sequence $\{\exp(i\boldsymbol{r}'\boldsymbol{x}_j)\}$ is stationary and ergodic. Applying

Theorem 3.5.7. of Stout (1974), we have the following S.L.L.N.,

$$c_n(\boldsymbol{r}) \stackrel{a.s.}{\to} c(\boldsymbol{r};\boldsymbol{\theta}_0) \ \forall \boldsymbol{r} \in \Re^{p+1},$$

Hence, we have,

$$Re\, c_n(\boldsymbol{r}) \stackrel{a.s.}{\to} Re\, c(\boldsymbol{r};\boldsymbol{\theta}_0),$$

and

$$Im\, c_n(\boldsymbol{r}) \stackrel{a.s.}{\to} Im\, c(\boldsymbol{r};\boldsymbol{\theta}_0).$$

For any $\delta > 0$, consider

$$\begin{aligned} I_n(\boldsymbol{\theta}_0 \pm \delta) - I_n(\boldsymbol{\theta}_0) \;=\; & \int \cdots \int \Big\{ (Re\, c(\boldsymbol{r};\boldsymbol{\theta}_0 \pm \delta) - Re\, c(\boldsymbol{r};\boldsymbol{\theta}_0)) \\ & (Re\, c(\boldsymbol{r};\boldsymbol{\theta}_0 \pm \delta) + Re\, c(\boldsymbol{r};\boldsymbol{\theta}_0) - 2Re\, c_n(\boldsymbol{r})) \\ & +(Im\, c(\boldsymbol{r};\boldsymbol{\theta}_0 \pm \delta) - Im\, c(\boldsymbol{r};\boldsymbol{\theta}_0)) \\ & (Im\, c(\boldsymbol{r};\boldsymbol{\theta}_0 \pm \delta) + Im\, c(\boldsymbol{r};\boldsymbol{\theta}_0) - 2Im\, c_n(\boldsymbol{r})) \Big\}\, dG(\boldsymbol{r}). \end{aligned} \tag{A.1}$$

Since $Re\, c(\boldsymbol{r};\boldsymbol{\theta}), Im\, c(\boldsymbol{r};\boldsymbol{\theta}), Re\, c_n(\boldsymbol{r})$, and $Im\, c_n(\boldsymbol{r})$ are all trigonometric and thus bounded functions for any $\boldsymbol{\theta}$ and any $\boldsymbol{r}$, the integrand in $(A.1)$ is bounded. Furthermore, following from S.L.L.N., the integrand converges almost surely to

$$(Re\, c(\boldsymbol{r};\boldsymbol{\theta}_0 \pm \delta) - Re\, c(\boldsymbol{r};\boldsymbol{\theta}_0))^2 + (Im\, c(\boldsymbol{r};\boldsymbol{\theta}_0 \pm \delta) - Im\, c(\boldsymbol{r};\boldsymbol{\theta}_0))^2.$$

By dominated convergence theorem,

$$\begin{aligned} & \lim_{n\to\infty} I_n(\boldsymbol{\theta}_0 \pm \delta) - I_n(\boldsymbol{\theta}_0) \\ = \;& \int \cdots \int \Big\{ (Re\, c(\boldsymbol{r};\boldsymbol{\theta}_0 \pm \delta) - Re\, c(\boldsymbol{r};\boldsymbol{\theta}_0))^2 + (Im\, c(\boldsymbol{r};\boldsymbol{\theta}_0 \pm \delta) - Im\, c(\boldsymbol{r};\boldsymbol{\theta}_0))^2 \Big\}\, dG(\boldsymbol{r}) \\ > \;& 0, \end{aligned} \tag{A.2}$$

with probability 1. Therefore, $I_n(\boldsymbol{\theta}_0 \pm \delta) > I_n(\boldsymbol{\theta}_0)$ holds almost surely for large n. This implies that $I_n(\boldsymbol{\theta})$ achieves a minimum at $\boldsymbol{\theta}_0$ locally almost surely for large n. Furthermore, $\boldsymbol{\theta}_0$ is also a global minimum, for otherwise relation $(A.2)$ shows that

$Re\, c(\boldsymbol{r};\boldsymbol{\theta}_0 \pm \delta) = Re\, c(\boldsymbol{r};\boldsymbol{\theta}_0)$ and $Im\, c(\boldsymbol{r};\boldsymbol{\theta}_0 \pm \delta) = Im\, c(\boldsymbol{r};\boldsymbol{\theta}_0)$ for all $\boldsymbol{r}$. This in turn implies that $F(\boldsymbol{x},\boldsymbol{\theta}_0 \pm \delta) = F(\boldsymbol{x},\boldsymbol{\theta}_0)$ for all $\boldsymbol{x}$, where F is the distribution function of $\boldsymbol{x}$. Since $\delta \neq 0$, this contradicts (A1). Because $I_n(\boldsymbol{\theta})$ is differentiable and δ is arbitrary, there exists a solution $\tilde{\boldsymbol{\theta}}_n$ of the equation $I_n'(\boldsymbol{\theta}) = 0$ that is strongly consistent, i.e.,

$$\tilde{\boldsymbol{\theta}}_n \overset{a.s.}{\to} \boldsymbol{\theta}_0.$$

To obtain the asymptotic normality, consider the first order condition of the minimization problem (2.1) and by Assumption (A2), we have,

$$\begin{aligned}
\partial I_n(\boldsymbol{\theta})/\partial\boldsymbol{\theta} &= 2\int\cdots\int\Big\{\Big[Re\, c_n(\boldsymbol{r}) - Re\, c(\boldsymbol{r};\boldsymbol{\theta})\Big]\frac{\partial Re\, c(\boldsymbol{r};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} \\
&\quad + \Big[Im\, c_n(\boldsymbol{r}) - Im\, c(\boldsymbol{r};\boldsymbol{\theta})\Big]\frac{\partial Im\, c(\boldsymbol{r};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\Big\}\, dG(\boldsymbol{r}) \\
&= -\frac{2}{n}\sum_{j=1}^{n}\int\cdots\int\Big\{\Big[cos(\boldsymbol{r}'\boldsymbol{x}_j) - Re\, c(\boldsymbol{r};\boldsymbol{\theta})\Big]\frac{\partial Re\, c(\boldsymbol{r};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} \\
&\quad + \Big[sin(\boldsymbol{r}'\boldsymbol{x}_j) - Im\, c(\boldsymbol{r};\boldsymbol{\theta})\Big]\frac{\partial Im\, c(\boldsymbol{r};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\Big\}\, dG(\boldsymbol{r}). \qquad \text{(A.3)}
\end{aligned}$$

Define

$$\begin{aligned}
K_j(\boldsymbol{\theta}) &= \int\cdots\int\Big\{\Big[cos(\boldsymbol{r}'\boldsymbol{x}_j) - Re\, c(\boldsymbol{r};\boldsymbol{\theta})\Big]\frac{\partial Re\, c(\boldsymbol{r};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} \\
&\quad + \Big[sin(\boldsymbol{r}'\boldsymbol{x}_j) - Im\, c(\boldsymbol{r};\boldsymbol{\theta})\Big]\frac{\partial Im\, c(\boldsymbol{r};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\Big\}\, dG(\boldsymbol{r}), \qquad \text{(A.4)}
\end{aligned}$$

hence

$$\partial I_n(\boldsymbol{\theta})/\partial\boldsymbol{\theta} = -\frac{2}{n}\sum_{j=1}^{n} K_j(\boldsymbol{\theta}).$$

Therefore $\partial I_n(\boldsymbol{\theta})/\partial\boldsymbol{\theta}$ is the sample mean of a random sequence $\{K_j(\boldsymbol{\theta})\}$ multiplied by a constant, -2. Note that $\{K_j(\boldsymbol{\theta})\}$ is identical but not independently distributed since $\{\boldsymbol{x}\}$ is dependent. Also observe that $\{K_j(\boldsymbol{\theta})\}$ preserves the stationarity of $\{y_j\}$. Furthermore, using Lemma 2.1 of White and Domowitz (1984), we can show that $\{K_j(\boldsymbol{\theta})\}$ is α-mixing and of the same size as $\{y_j\}$. This implies $\sum \alpha_2(n) < \infty$, where $\alpha_2(n)$ is the mixing coefficient of $\{K_j(\boldsymbol{\theta})\}$. By Theorem 17.2.1. of Ibragimov and

Linnik (1971), (A6) implies,

$$A(\boldsymbol{\theta}) = var(K_1) + 2\sum_{j=2}^{\infty} cov(K_1, K_j) < \infty.$$

Using a central limit theorem for stationary processes (for example, Theorem 18.5.4 in Ibragimov and Linnik (1971)), we have

$$n^{1/2}\partial I_n(\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta} \xrightarrow{D} N(0, 4A(\boldsymbol{\theta})), \tag{A.5}$$

where

$$\begin{aligned}
A(\boldsymbol{\theta}) &= \lim_{n\to\infty}\frac{1}{n}E(\sum_{j=1}^{n}\sum_{k=1}^{n}K_1(\boldsymbol{\theta})K_k(\boldsymbol{\theta})) \qquad (A.6)\\
&= \lim_{n\to\infty}\int\cdots\int\left\{\frac{\partial Re\, c(\boldsymbol{r};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\frac{\partial Re\, c(\boldsymbol{s};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'}\frac{1}{n}\sum_{j=1}^{n}\sum_{k=1}^{n}cov(cos(\boldsymbol{r}'\boldsymbol{x}_j), cos(\boldsymbol{s}'\boldsymbol{x}_k))\right.\\
&\quad+\frac{\partial Re\, c(\boldsymbol{r};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\frac{\partial Im\, c(\boldsymbol{s};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'}\frac{2}{n}\sum_{j=1}^{n}\sum_{k=1}^{n}cov(cos(\boldsymbol{r}'\boldsymbol{x}_j), sin(\boldsymbol{s}'\boldsymbol{x}_k)) +\\
&\quad\left.\frac{\partial Im\, c(\boldsymbol{r};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\frac{\partial Im\, c(\boldsymbol{s};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'}\frac{1}{n}\sum_{j=1}^{n}\sum_{k=1}^{n}cov(sin(\boldsymbol{r}'\boldsymbol{x}_j), sin(\boldsymbol{s}'\boldsymbol{x}_k))\right\}dG(\boldsymbol{r})dG(\boldsymbol{s}).
\end{aligned}$$

The double summation covariance expressions are readily found and are given in the Lemma in Knight and Satchell [1997, p. 176]. That is, we note that

$$\begin{aligned}
&\frac{1}{n}\sum_{j=1}^{n}\sum_{k=1}^{n}cov(cos(\boldsymbol{r}'\boldsymbol{x}_j), cos(\boldsymbol{s}'\boldsymbol{x}_k))\\
&= n\cdot cov(Re\ c_n(\boldsymbol{r}), Re\ c_n(\boldsymbol{s}))\\
&= n\cdot(\Omega_{RR})_{\boldsymbol{r},\boldsymbol{s}},
\end{aligned}$$

using notation in Knight and Satchell [1997]. Similarly, for the other double sums.

In view of (A2),

$$\begin{aligned}
\frac{\partial^2 I_n(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} &= -2\int\cdots\int\left\{\frac{\partial Re\, c(\boldsymbol{r};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\frac{\partial Re\, c(\boldsymbol{r};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'}+\frac{\partial Im\, c(\boldsymbol{r};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\frac{\partial Im\, c(\boldsymbol{r};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'}\right.\\
&\quad-[Re\, c_n(\boldsymbol{r}) - Re\, c(\boldsymbol{r};\boldsymbol{\theta})]\frac{\partial^2 Re\, c(\boldsymbol{r};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} - [Im\, c_n(\boldsymbol{r}) - Im\, c(\boldsymbol{r};\boldsymbol{\theta})]\\
&\quad\left.\frac{\partial^2 Im\, c(\boldsymbol{r};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right\}dG(\boldsymbol{r}),
\end{aligned} \tag{A.7}$$

and by the S.L.L.N., we have

$$\frac{\partial^2 I_n(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} \overset{a.s.}{\to} E\Big[\frac{\partial^2 I_n(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\Big] = -2\int\cdots\int \frac{\partial c(\boldsymbol{r};\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}\frac{\partial c(\boldsymbol{r};\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}'}\,dG(\boldsymbol{r}).$$

Definition of $\tilde{\boldsymbol{\theta}}_n$ and the Taylor expansion imply

$$0 = \frac{\partial I_n(\tilde{\boldsymbol{\theta}}_n)}{\partial\boldsymbol{\theta}} = \frac{\partial I_n(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}} + (\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)\frac{\partial^2 I_n(\boldsymbol{\theta}_0 + \epsilon(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0))}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}$$

for some ϵ in $[-1,1]$. Hence, with (A4)

$$n^{1/2}(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) = -n^{1/2}\frac{\partial I_n(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}\Big(\frac{\partial^2 I_n(\boldsymbol{\theta}_0 + \epsilon(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0))}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\Big)^{-1} \overset{D}{\to} N(0, B^{-1}(\boldsymbol{\theta})A(\boldsymbol{\theta})B^{-1}(\boldsymbol{\theta})), \tag{A.8}$$

where

$$B(\boldsymbol{\theta}) = \int\cdots\int \frac{\partial c(\boldsymbol{r};\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}\frac{\partial c(\boldsymbol{r};\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}'}\,dG(\boldsymbol{r}). \quad \blacksquare$$

Appendix B

Proof of Theorem 3.1

For any p, define $\boldsymbol{x}_j = (y_j, \cdots, y_{j+p})'$.

Let $\sigma^2\Phi_{(p+1)\times(p+1)}$ be the covariance matrix of $\boldsymbol{x}_j$, then we have

$$\begin{aligned}
&\int\cdots\int |c_n(\boldsymbol{r}) - c(\boldsymbol{r})|^2 \exp(-a\boldsymbol{r}'\boldsymbol{r})d\boldsymbol{r} \\
&= \int\cdots\int (c_n(\boldsymbol{r}) - c(\boldsymbol{r}))(\bar{c}_n(\boldsymbol{r}) - \bar{c}(\boldsymbol{r}))\exp(-a\boldsymbol{r}'\boldsymbol{r})\,d\boldsymbol{r} \\
&= \int\cdots\int \Big[\frac{1}{n}\sum_{j=1}^n \exp(i\boldsymbol{r}'\boldsymbol{x}_j) - \exp(-\frac{\sigma^2}{2}\boldsymbol{r}'\Phi\boldsymbol{r})\Big] \\
&\quad \Big[\frac{1}{n}\sum_{j=1}^n \exp(-i\boldsymbol{r}'\boldsymbol{x}_j) - \exp(-\frac{\sigma^2}{2}\boldsymbol{r}'\Phi\boldsymbol{r})\Big]\exp(-a\boldsymbol{r}'\boldsymbol{r})\,d\boldsymbol{r} \\
&= \frac{1}{n^2}\sum_{j=1}^n\sum_{k=1}^n \int\cdots\int \exp(i\boldsymbol{r}'\boldsymbol{x}_j)\exp(i\boldsymbol{r}'\boldsymbol{x}_k)\,d\boldsymbol{r} \\
&\quad -\frac{2}{n}\sum_{j=1}^n \int\cdots\int \exp(i\boldsymbol{r}'\boldsymbol{x}_j)\exp(-\boldsymbol{r}'A\boldsymbol{r})\,d\boldsymbol{r} \\
&\quad +\int\cdots\int \exp(-\boldsymbol{r}'B\boldsymbol{r})\,d\boldsymbol{r},
\end{aligned} \tag{B.1}$$

where $A = \frac{\sigma^2}{2}\Phi + a \times I$ and $B = \sigma^2\Phi + a \times I$ with I as an identity matrix . Note that the first part in the above equation is a constant with respect to $\boldsymbol{\theta}$. Let $A = C^{-1}$. Considering the characteristic function of a random variable $N(0, \frac{1}{2}C)$, we have

$$\exp(-\frac{1}{4}\boldsymbol{x_j}'C\boldsymbol{x_j}) = \int \cdots \int \exp(i\boldsymbol{x_j}'\boldsymbol{r}) \frac{\exp(-\frac{1}{2}\boldsymbol{r}'(\frac{C}{2})^{-1}\boldsymbol{r})}{(2\pi)^{(p+1)/2}|C|^{1/2}(1/2)^{(p+1)/2}} d\boldsymbol{r}. \tag{B.2}$$

Thus the second part in $(B.1)$ is,

$$\begin{aligned} &\int \cdots \int \exp(i\boldsymbol{r}'\boldsymbol{x_j}) \exp(-\boldsymbol{r}'A\boldsymbol{r})\, d\boldsymbol{r} \\ &= (2\pi)^{(p+1)/2}|A|^{-\frac{1}{2}}2^{-(p+1)/2} \exp(-\frac{1}{4}\boldsymbol{x_j}'A^{-1}\boldsymbol{x_j}) \\ &= \pi^{\frac{p+1}{2}}|A|^{-\frac{1}{2}} \exp(-\frac{1}{4}\boldsymbol{x_j}'A^{-1}\boldsymbol{x_j}). \end{aligned} \tag{B.3}$$

Similarly, for the third part in $(B.1)$ we have

$$\int \cdots \int \exp(-\boldsymbol{r}'B\boldsymbol{r})\, d\boldsymbol{r} = \pi^{\frac{p+1}{2}}|B|^{-\frac{1}{2}}. \tag{B.4}$$

Substituting $(B.2)$, $(B.3)$ and $(B.4)$ into $(B.1)$, we have

$$I_n(\boldsymbol{\theta}) = constant - \frac{2}{n}\pi^{\frac{p+1}{2}}|A|^{-\frac{1}{2}} \sum_{j=1}^{n} \exp(-\frac{1}{4}\boldsymbol{x_j}'A^{-1}\boldsymbol{x_j}) + \pi^{\frac{p+1}{2}}|B|^{-\frac{1}{2}}. \quad \blacksquare$$

Appendix C

Proof of Theorem 3.2

By the assumption

$$(y_{j+p}|y_j, \ldots, y_{j+p-1}) \sim N\left(f_1(\boldsymbol{\rho})y_j + \cdots + f_q(\boldsymbol{\rho})y_{j+p-1}, \sigma^2 g(\boldsymbol{\rho})\right).$$

This implies

$$\log f(y_{j+p}|y_j, \ldots, y_{j+p-1}) = -\frac{1}{2}\log 2\pi - \frac{1}{2}\log\sigma^2 - \frac{1}{2}\log g(\boldsymbol{\rho}) - \frac{1}{2\sigma^2 g(\boldsymbol{\rho})}\boldsymbol{x}_j'A\boldsymbol{x}_j,$$

where A is defined in Theorem 3.2. If we take the derivative of the log conditional density function with respect to the parameters, we have

$$\frac{\partial \log f(y_{j+1}|y_j, \ldots, y_{j+p-1})}{\partial\sigma^2} = -\frac{1}{2\sigma^2} + \frac{1}{2\sigma^4 g(\boldsymbol{\rho})}\boldsymbol{x}_j'A\boldsymbol{x}_j,$$

and

$$\frac{\partial \log f(y_{j+1}|y_j,\cdots,y_{j+p-1})}{\partial \rho_k} = -\frac{g_k(\boldsymbol{\rho})}{2g(\boldsymbol{\rho})} + \frac{g_k(\boldsymbol{\rho})}{2\sigma^2 g^2(\boldsymbol{\rho})}\boldsymbol{x}_j' A\boldsymbol{x}_j - \frac{1}{2\sigma^2 g(\boldsymbol{\rho})}\boldsymbol{x}_j' B_k \boldsymbol{x}_j,$$

where g_k and B_k for $k = 1,\cdots,l+m$ are defined in Theorem 3.2. Then the optimal weight function could be

$$\begin{aligned} w_{\sigma^2}(r) &= (\frac{1}{2\pi})^{p+1}\int\cdots\int \exp(-i\boldsymbol{r}'\boldsymbol{x}_j)\Big(-\frac{1}{2\sigma^2} + \frac{1}{2\sigma^4 g(\boldsymbol{\rho})}\boldsymbol{x}_j' A\boldsymbol{x}_j\Big)\, dy_j\cdots dy_{j+p} \\ &= -\frac{1}{2\sigma^2}\delta(r^1)\cdots\delta(r^{p+1}) + \frac{1}{2\sigma^4 g(\boldsymbol{\rho})}(\frac{1}{2\pi})^{p+1}\int\cdots\int \exp(-i\boldsymbol{r}'\boldsymbol{x}_j)\boldsymbol{x}_j' A\boldsymbol{x}_j \\ &\quad dy_j\cdots dy_{j+p}. \end{aligned} \tag{C.1}$$

Consider

$$\begin{aligned} &(\frac{1}{2\pi})^{p+1}\int\cdots\int \exp(-i\boldsymbol{r}'\boldsymbol{x}_j)\boldsymbol{x}_j' A\boldsymbol{x}_j\, dy_j\cdots dy_{j+p} \\ =\ &(\frac{1}{2\pi})^{p+1}\int\cdots\int \exp(-i\boldsymbol{r}'\boldsymbol{x}_j)\boldsymbol{x}_j' M\Lambda M'\boldsymbol{x}_j\, dy_j\cdots dy_{j+p} \\ =\ &(\frac{1}{2\pi})^{p+1}\int\cdots\int \exp(-i\boldsymbol{s}'\boldsymbol{z})\boldsymbol{z}'\Lambda\boldsymbol{z}\, dz_1\cdots dz_2 \\ =\ &(\frac{1}{2\pi})^{p+1}\int\cdots\int \exp(-\sum_{k=1}^{p+1} is^k z_k)\sum_{k=1}^{p+1}\lambda^k z_k^2\, dz_1\cdots dz_{p+1} \\ =\ &(\frac{1}{2\pi})^{p+1}\int \exp(-is^1 z_1)\lambda^1 z_1^2\, dz_1 \Pi_{k\neq 1}\Big\{\int \exp(-is^k z_k)\, dz_k\Big\} + \cdots \\ &+(\frac{1}{2\pi})^{p+1}\int \exp(-is^{p+1}z_{p+1})\lambda^{p+1}z_{p+1}^2\, dz_{p+1}\Pi_{k\neq p+1}\Big\{\int \exp(-is^k z_k)\, dz_k\Big\} \\ =\ &-\lambda^1\delta''(s^1)\delta(s^2)\cdots\delta(s^{p+1}) - \cdots - \lambda^{p+1}\delta(s^1)\cdots\delta(s^p)\delta''(s^{p+1}), \end{aligned}$$

where M and Λ are defined in Theorem 3.2, $\boldsymbol{z} = M'\boldsymbol{x}_j$ and $\boldsymbol{s} = M'\boldsymbol{r}$. Hence, $(C.1)$ becomes

$$\begin{aligned} w_{\sigma^2}(\boldsymbol{r}) &= -\frac{1}{2\sigma^2}\delta(r^1)\cdots\delta(r^{p+1}) - \frac{1}{2\sigma^4 g(\boldsymbol{\rho})}[\lambda^1\delta''(s^1)\delta(s^2)\cdots\delta(s^{p+1}) \\ &\quad +\cdots+\lambda^{p+1}\delta(s_1)\cdots\delta(s_p)\delta''(s_{p+1})], \end{aligned} \tag{C.2}$$

and

$$w_{\boldsymbol{\rho}_k}(\boldsymbol{r}) = (\frac{1}{2\pi})^{p+1}\int\cdots\int \exp(-i\boldsymbol{r}'\boldsymbol{x}_j)\Big(-\frac{g_k(\boldsymbol{\rho})}{2g(\boldsymbol{\rho})} + \frac{g_k(\boldsymbol{\rho})}{2\sigma^2 g^2(\boldsymbol{\rho})}\boldsymbol{x}_j' A\boldsymbol{x}_j$$

$$
\begin{aligned}
& -\frac{1}{2\sigma^2 g(\boldsymbol{\rho})}\boldsymbol{x}_j' B_k \boldsymbol{x}_j\Big)\, dy_j \cdots dy_{j+p} \\
= & -\frac{g_k(\boldsymbol{\rho})}{2g(\boldsymbol{\rho})}\delta(r^1)\cdots\delta(r^{p+1}) - \frac{g_k(\boldsymbol{\rho})}{2\sigma^2 g^2(\boldsymbol{\rho})}[\lambda^1\delta''(s^1)\delta(s^2)\cdots\delta(s^{p+1}) \qquad \text{(C.3)} \\
& + \cdots + \lambda^{p+1}\delta(s^1)\cdots\delta(s^p)\delta''(s^{p+1})] + \frac{1}{2\sigma^2 g(\boldsymbol{\rho})}[\lambda_k^1\delta''(t_k^1)\delta(t_k^2)\cdots\delta(t_k^{p+1}) \\
& + \cdots + \lambda_k^{p+1}\delta(t_k^1)\cdots\delta(t_k^p)\delta''(t_k^{p+1})],
\end{aligned}
$$

where $\boldsymbol{t_k} = H_k'\boldsymbol{r}$ for $k = 1, \cdots, l+m$. ■

Appendix D

Proof of Corollary 3.1

Based on the optimal weight functions, the estimating equations are

$$
\begin{aligned}
0 &= \int\cdots\int w_{\sigma^2}(\boldsymbol{r})c_n(\boldsymbol{r})\, d\boldsymbol{r} \\
&= \int\cdots\int\Big\{-\frac{1}{2\sigma^2}\delta(r^1)\cdots\delta(r^{p+1}) - \frac{1}{2\sigma^4 g(\boldsymbol{\rho})}[\lambda^1\delta''(s^1)\delta(s^2)\cdots\delta(s^{p+1}) \\
&\quad + \cdots + \lambda^{p+1}\delta(s^1)\cdots\delta(s^p)\delta''(s^{p+1})]\Big\}c_n(\boldsymbol{r})\, d\boldsymbol{r} \\
&= -\frac{1}{2\sigma^2}c_n(\boldsymbol{0}) - \frac{1}{2\sigma^4 g(\boldsymbol{\rho})}\int\cdots\int\Big[\lambda^1\delta''(s^1)\delta(s^2)\cdots\delta(s^{p+1}) \\
&\quad + \cdots + \lambda^{p+1}\delta(s^1)\cdots\delta(s^p)\delta''(s^{p+1})\Big]c_n(\boldsymbol{r})\, d\boldsymbol{r} \\
&= -\frac{1}{2\sigma^2} - \frac{1}{2\sigma^4 g(\boldsymbol{\rho})}\Big[\lambda^1\frac{\partial^2 c_n(M\boldsymbol{s})}{\partial {s^1}^2}\Big|_{\boldsymbol{s}=\boldsymbol{0}} + \cdots + \lambda^{p+1}\frac{\partial^2 c_n(M\boldsymbol{s})}{\partial {s^{p+1}}^2}\Big|_{\boldsymbol{s}=\boldsymbol{0}}\Big] \\
&= -\frac{1}{2\sigma^2} + \frac{1}{2\sigma^4 g(\boldsymbol{\rho})}\frac{1}{n}\sum_{j=1}^{n}\Big[\lambda^1((M'\boldsymbol{x}_j)_{(1)})^2 + \cdots + \lambda^{p+1}((M'\boldsymbol{x}_j)_{(p+1)})^2\Big] \\
&= -\frac{1}{2\sigma^2} + \frac{1}{2\sigma^4 g(\boldsymbol{\rho})}\bar{P}(\boldsymbol{\rho}), \qquad \text{(D.1)}
\end{aligned}
$$

where

$$
\bar{P}(\boldsymbol{\rho}) = \frac{1}{n}\sum_{j=1}^{n}\Big[\lambda^1((M'\boldsymbol{x}_j)_{(1)})^2 + \cdots + \lambda^{p+1}((M'\boldsymbol{x}_j)_{(p+1)})^2\Big],
$$

and

$$
\begin{aligned}
0 &= \int\cdots\int w_{\boldsymbol{\rho}_k}(\boldsymbol{r})c_n(\boldsymbol{r})\, d\boldsymbol{r} \\
&= \int\cdots\int\Big\{-\frac{g_k(\boldsymbol{\rho})}{2g(\boldsymbol{\rho})}\delta(r^1)\cdots\delta(r^{p+1})
\end{aligned}
$$

$$
\begin{aligned}
& -\frac{g_k(\boldsymbol{\rho})}{2\sigma^2 g^2(\boldsymbol{\rho})}[\lambda^1\delta''(s^1)\delta(s^2)\cdots\delta(s^{p+1}) + \cdots + \lambda^{p+1}\delta(s^1)\cdots\delta(s^p)\delta''(s^{p+1})] \\
& +\frac{1}{2\sigma^2 g(\boldsymbol{\rho})}[\lambda_k^1\delta''(t_k^1)\delta(t_k^2)\cdots\delta(t_k^{p+1}) + \cdots + \lambda_k^{p+1}\delta(t_k^1)\cdots\delta(t_k^p)\delta''(t_k^{p+1})]\Big\}c_n(\boldsymbol{r})\,d\boldsymbol{r} \\
= & -\frac{g_k(\boldsymbol{\rho})}{2g(\boldsymbol{\rho})} + \frac{g_k(\boldsymbol{\rho})}{2\sigma^2 g^2(\boldsymbol{\rho})}\bar{P}(\boldsymbol{\rho}) - \frac{1}{2\sigma^2 g(\boldsymbol{\rho})}\bar{Q}(\boldsymbol{\rho}), \qquad \text{(D.2)}
\end{aligned}
$$

where

$$
\bar{Q}_k(\boldsymbol{\rho}) = \frac{1}{n}\sum_{j=1}^{n}(\lambda_k^1((H_k'\boldsymbol{x}_j)_{(1)})^2 + \cdots + \lambda_k^{p+1}((H_k\boldsymbol{x}_j')_{(p+1)})^2),
$$

for $k = 1, \cdots, l+m$. Concentrating out σ^2 in $(D.1)$ and $(D.2)$, we have (3.17). ∎

Table 1

MONTE CARLO STUDY COMPARING DECF AND ML OF THE MA(1) MODEL

TRUE VALUES OF PARAMETERS $\phi = 0.6$ $\sigma^2 = 1$

NO. OF REPLICATIONS=1,000 NO. OF OBSERVATIONS=100

INITIAL VALUES OF VECTORS $\boldsymbol{r}$

−0.10	−0.05	−0.01	0.03	0.08
−0.09	−0.04	0.00	0.04	0.09
−0.08	−0.03	0.01	0.05	0.10

INITIAL VALUE OF τ 0.5

	$\hat{\phi}$					$\hat{\sigma}^2$				
	c(1)	c(2)	c(3)	c(4)	MLE	c(1)	c(2)	c(3)	c(4)	MLE
MEAN	.5963	.5913	.5838	.5814	.6032	.9806	.9753	.9945	.9829	.9922
MED	.5730	.5388	.5456	.5243	.6026	.9728	.9541	1.006	.9605	.9838
MIN	.2739	.1876	.2163	.0371	.3472	.4526	.6086	.4358	.6125	.6411
MAX	1.00	1.00	1.00	1.00	.876	1.46	1.53	1.53	1.57	1.47
SKEW	.6886	.6115	.5344	.2771	.0371	-.142	.0045	.4733	.4108	.2940
KURT	2.62	2.54	2.41	2.29	3.13	2.81	2.73	2.79	2.62	3.01
VAR	.043	.049	.049	.067	.007	.041	.036	.051	.045	.020
BIAS	.0037	.0087	.0162	.0186	.0032	.0394	.0247	.0055	.0171	.0078
MSE	.043	.049	.049	.067	.007	.042	.037	.051	.046	.020

Table 2

MONTE CARLO STUDY COMPARING CECF AND ML OF THE MA(1) MODEL

TRUE VALUES OF PARAMETERS $\phi = 0.6$ $\sigma^2 = 1$

NO. OF REPLICATIONS=1,000 NO. OF OBSERVATIONS=100

INITIAL VALUES OF ϕ AND σ^2: 0.3 0.7

	$\hat{\phi}$				$\hat{\sigma}^2$			
	c(1)	c(2)	c(3)	MLE	c(1)	c(2)	c(3)	MLE
MEAN	.6479	.6475	.6269	.6032	.9583	.9622	.9665	.9922
MED	.6102	.6076	.5960	.6026	.9369	.9513	.9567	.9838
MIN	.0281	.1508	.2382	.3472	.3948	.4992	.5496	.6411
MAX	1.000	1.000	1.000	.8758	1.729	1.572	1.513	1.47
SKEW	.3798	.4070	.7960	.0371	.3798	.3241	.2553	.2940
KURT	2.917	2.134	3.244	3.126	2.917	3.057	3.022	3.01
VAR	.057	.046	.026	.007	.049	.033	.026	.020
BIAS	.0479	.0475	.0269	.0032	.0417	.0378	.0335	.0078
MSE	.059	.048	.027	.007	.051	.034	.027	.020

Table 3

MONTE CARLO STUDY COMPARING CECF-GLS AND ML OF THE MA(1) MODEL

TRUE VALUES OF PARAMETERS $\phi = 0.6$ $\sigma^2 = 1$

NO. OF REPLICATIONS=1,000 NO. OF OBSERVATIONS=100

INITIAL VALUES OF ϕ AND σ^2: 0.3 0.7

	$\hat{\phi}$								
	p=3	p=4	p=5	p=6	p=7	p=8	p=9	p=10	MLE
MEAN	.6133	.6061	.6037	.6024	.6028	.6025	.6026	.6023	.6032
MED	.6011	.5992	.6004	.6009	.6018	.6015	.6006	.6016	.6026
MIN	.3138	.3446	.3401	.3493	.338	.3441	.3274	.323	.3472
MAX	1.000	1.000	1.000	.8861	.943	.9055	.9002	.9415	.8757
SKEW	.8491	.5473	.3802	.0398	.0253	-.012	.0054	.0693	.0371
KURT	4.211	4.183	3.960	3.148	3.240	3.054	3.073	3.284	3.126
VAR	.015	.010	.009	.007	.007	.007	.007	.008	.007
BIAS	.0133	.0061	.0037	.0024	.0028	.0025	.0026	.0023	.0032
MSE	.015	.010	.009	.007	.007	.007	.007	.008	.007

Table 4

MONTE CARLO STUDY COMPARING CECF-GLS AND ML OF THE MA(1) MODEL

TRUE VALUES OF PARAMETERS $\phi = 0.6$ $\sigma^2 = 1$

NO. OF REPLICATIONS=1,000 NO. OF OBSERVATIONS=100

INITIAL VALUES OF ϕ AND σ^2: 0.3 0.7

	$\hat{\sigma}^2$								
	p=3	p=4	p=5	p=6	p=7	p=8	p=9	p=10	MLE
MEAN	.9813	.9878	.9905	.9911	.9915	.9912	.9919	.9919	.9922
MED	.9742	.981	.9841	.987	.9869	.9829	.9877	.9875	.9838
MIN	.5571	.6032	.6383	.6353	.6419	.6387	.6434	.6117	.6411
MAX	1.479	1.462	1.473	1.485	1.497	1.496	1.51	1.511	1.47
SKEW	.3245	.3215	.3062	.3042	.322	.3124	.3047	.3236	.2940
KURT	3.055	3.073	3.082	3.156	3.132	3.091	3.098	3.122	3.01
VAR	.022	.021	.021	.021	.021	.021	.022	.022	.020
BIAS	.0187	.0122	.0095	.0089	.0085	.0088	.0081	.0081	.0078
MSE	.022	.021	.021	.021	.021	.021	.022	.022	.020

Table 5

MONTE CARLO STUDY FOR CECF OF THE MA(1) MODEL

TRUE VALUES OF PARAMETERS $\phi = 0.6$ $\sigma^2 = 1$

NO. OF REPLICATIONS=1,000 NO. OF OBSERVATIONS=1000

	$\hat{\phi}$					
	c(1)	c(2)	c(3)	c(4)	c(5)	c(6)
MEAN	.6350	.6200	.6128	.6095	.6068	.6072
MED	.6061	.6045	.6028	.6030	.5994	.6014
MIN	.3404	.3907	.3746	.4090	.4104	.4204
MAX	1.000	1.000	1.000	1.000	.9900	1.000
SKEW	.9929	1.243	.9365	.7593	.6927	.6765
KURT	3.637	5.551	5.329	4.846	4.864	4.971
VAR	.021	.010	.008	.005	.005	.004
BIAS	.035	.020	.0128	.0095	.0068	.0072
MSE	.022	.011	.008	.005	.005	.004

Table 6

MONTE CARLO STUDY FOR CECF OF THE MA(1) MODEL

TRUE VALUES OF PARAMETERS $\phi = 0.6$ $\sigma^2 = 1$

NO. OF REPLICATIONS=1,000 NO. OF OBSERVATIONS=1000

	$\hat{\sigma^2}$					
	c(1)	c(2)	c(3)	c(4)	c(5)	c(6)
MEAN	.9691	.9821	.9879	.9909	.9928	.9929
MED	.9852	.9856	.9887	.9918	.9917	.9922
MIN	.6293	.6964	.6303	.8039	.7253	.8147
MAX	1.234	1.176	1.21	1.172	1.201	1.169
SKEW	-0.718	-0.611	-.268	-.025	.024	.033
KURT	3.57	4.213	3.775	3.056	3.004	2.922
VAR	.011	.006	.005	.003	.004	.003
BIAS	.0309	.0178	.0121	.0091	.0072	.0071
MSE	.022	.006	.005	.003	.004	.003

Table 7

MONTE CARLO STUDY COMPARING CECF-GLS AND ML OF THE AR(1) MODEL

TRUE VALUES OF PARAMETERS $\rho = 0.6$ $\sigma^2 = 1$

NO. OF REPLICATIONS=1,000 NO. OF OBSERVATIONS=100

INITIAL VALUES OF ρ AND σ^2: 0.3 0.7

	$\hat{\rho}$		$\hat{\sigma^2}$	
	ECF	MLE	ECF	MLE
MEAN	.5907	.5906	.9924	.9923
MED	.5957	.5982	.9850	.9838
MIN	.2208	.2200	.6425	.6410
MAX	.7878	.7797	1.443	1.516
SKEW	-.4960	-.5190	.2953	.3111
KURT	3.517	3.544	3.021	3.076
VAR	.006	.006	.020	.020
BIAS	.0093	.0094	.0076	.0077
MSE	.006	.006	.020	.020

Table 8

MONTE CARLO STUDY FOR CECF OF THE AR(1) MODEL

TRUE VALUES OF PARAMETERS $\rho = 0.6$ $\sigma^2 = 1$

NO. OF REPLICATIONS=1,000 NO. OF OBSERVATIONS=1000

	$\hat{\rho}$				$\hat{\sigma^2}$			
	c(1)	c(2)	c(3)	c(4)	c(1)	c(2)	c(3)	c(4)
MEAN	.5996	.5997	.5998	.6002	.9981	.9977	.999	.9975
MED	.6018	.6008	.6025	.6010	.997	.9973	.9974	.9958
MIN	.4465	.4693	.4300	.4679	.8193	.8416	.8068	.8341
MAX	.7051	.6841	.7060	.6885	1.177	1.145	1.189	1.152
SKEW	-.3101	-.2246	-.3566	-.1916	.153	.080	.136	.100
KURT	3.083	3.196	3.021	3.089	2.847	2.907	2.871	2.909
VAR	.0015	.0009	.0016	.0010	.004	.002	.004	.002
BIAS	.0004	.0003	.0002	.0002	.0019	.0023	.001	.0025
MSE	.0015	.0009	.0016	.0010	.004	.002	.004	.002

Table 9

MONTE CARLO STUDY COMPARING CECF-GLS AND CML OF THE ARMA(1,1) MODEL

TRUE VALUES OF PARAMETERS $-\phi = \rho = 0.6$ $\sigma^2 = 1$

NO. OF REPLICATIONS=1,000 NO. OF OBSERVATIONS=100

	$\hat{\rho}$			$\hat{\sigma^2}$		
	$p = 2$	$p = 3$	CMLE	$p = 2$	$p = 3$	CMLE
MEAN	.5972	.6000	.5895	.9899	1.003	1.018
MED	.5973	.6000	.5909	.9813	.9911	1.01
MIN	.3785	.3991	.3491	.5998	.5717	.6395
MAX	.7878	.7681	.7351	1.501	1.631	.1.764
SKEW	-.0906	-.0672	-.2132	.347	.434	.394
KURT	2.991	3.056	3.115	3.042	3.249	3.313
VAR	.0038	.0031	.0030	.023	.028	.023
BIAS	.0028	.0000	.0105	.0101	.0003	.018
MSE	.0038	.0031	.0032	.023	.028	.023

Table 10

MONTE CARLO STUDY FOR CECF OF THE ARMA(1,1) MODEL

TRUE VALUES OF PARAMETERS $-\phi = \rho = 0.6$ $\sigma^2 = 1$

NO. OF REPLICATIONS=1,000 NO. OF OBSERVATIONS=1000

	$\hat{\rho}$					
	c(1)	c(2)	c(3)	c(4)	c(5)	c(6)
MEAN	.6007	.6010	.6012	.6014	.6017	.6019
MED	.6013	.6007	.6018	.6008	.6017	.6019
MIN	.4488	.4841	.4671	.4797	.4777	.4734
MAX	.7112	.6902	.7162	.7046	.6993	.7093
SKEW	-.1993	-.0829	-.1067	-.0844	-.0802	-.1047
KURT	3.182	3.003	3.235	2.992	3.163	3.032
VAR	.0014	.0010	.0011	.0012	.0010	.0013
BIAS	.0007	.0010	.0012	.0014	.0017	.0019
MSE	.0014	.0010	.0011	.0012	.0010	.0013

Table 11

MONTE CARLO STUDY FOR CECF OF THE ARMA(1,1) MODEL

TRUE VALUES OF PARAMETERS $-\phi = \rho = 0.6$ $\sigma^2 = 1$

NO. OF REPLICATIONS=1,000 NO. OF OBSERVATIONS=1000

	$\hat{\sigma^2}$					
	c(1)	c(2)	c(3)	c(4)	c(5)	c(6)
MEAN	.9971	.9956	.9971	.9951	.9968	.9947
MED	.9949	.9935	.996	.9938	.9984	.9938
MIN	.7354	.7777	.7975	.748	.7863	.740
MAX	1.241	1.222	1.192	1.231	1.186	1.227
SKEW	.157	.1763	.116	.106	.068	.088
KURT	3.082	3.0863	2.949	3.002	2.910	2.968
VAR	.007	.004	.005	.005	.004	.005
BIAS	.0029	.0044	.0029	.0049	.0032	.0053
MSE	.007	.004	.005	.005	.004	.005

Table 12

MONTE CARLO STUDY COMPARING CECF-GLS AND CML OF THE ARMA(1,1) MODEL

TRUE VALUES OF PARAMETERS $-\phi = \rho = 0.9$ $\sigma^2 = 1$

NO. OF REPLICATIONS=1,000 NO. OF OBSERVATIONS=100

	$\hat{\rho}$			$\hat{\sigma^2}$		
	$p=2$	$p=3$	CMLE	$p=2$	$p=3$	CMLE
MEAN	.8846	.8881	.8512	1.003	1.002	1.483
MED	.8904	.8949	.8601	.9911	.9884	1.297
MIN	.6374	.6569	.6015	.5717	.5956	.6918
MAX	.9739	.9820	.9690	1.631	1.557	4.71
SKEW	-.9348	-.9060	-.7173	.434	.454	1.744
KURT	4.307	4.186	3.577	3.249	3.209	6.561
VAR	.0023	.0023	.0032	.028	.027	.378
BIAS	.0154	.0119	.0488	.003	.002	.483
MSE	.0026	.0024	.0056	.028	.027	.612

Figure 1: Density Function of phi in MA

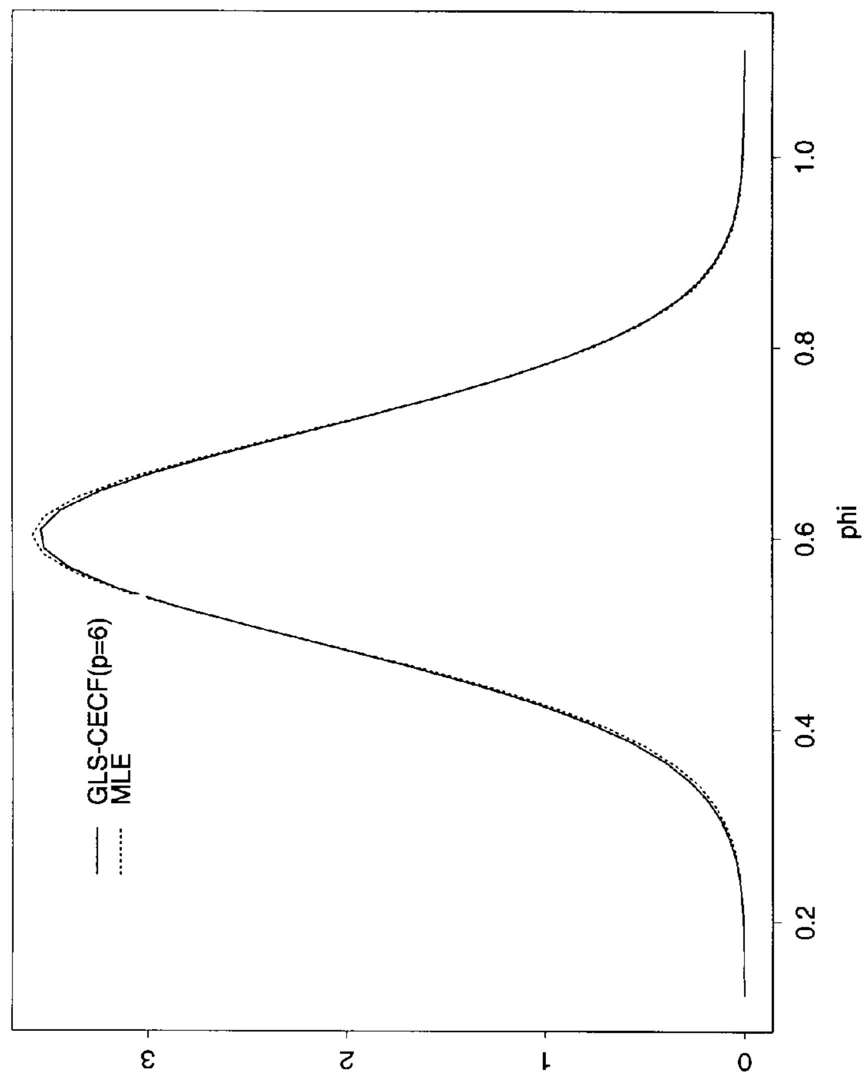


Figure 2: Density Function of rho in AR

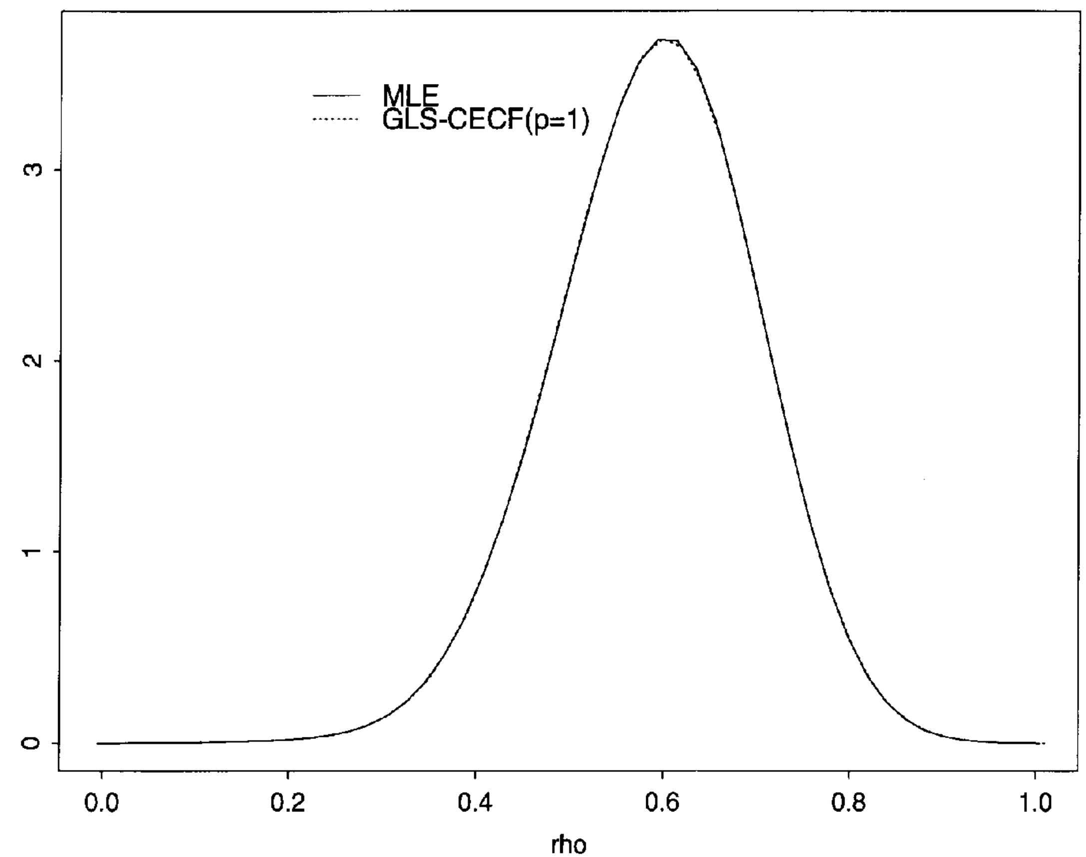


Figure 3: Density Function of rho in ARMA

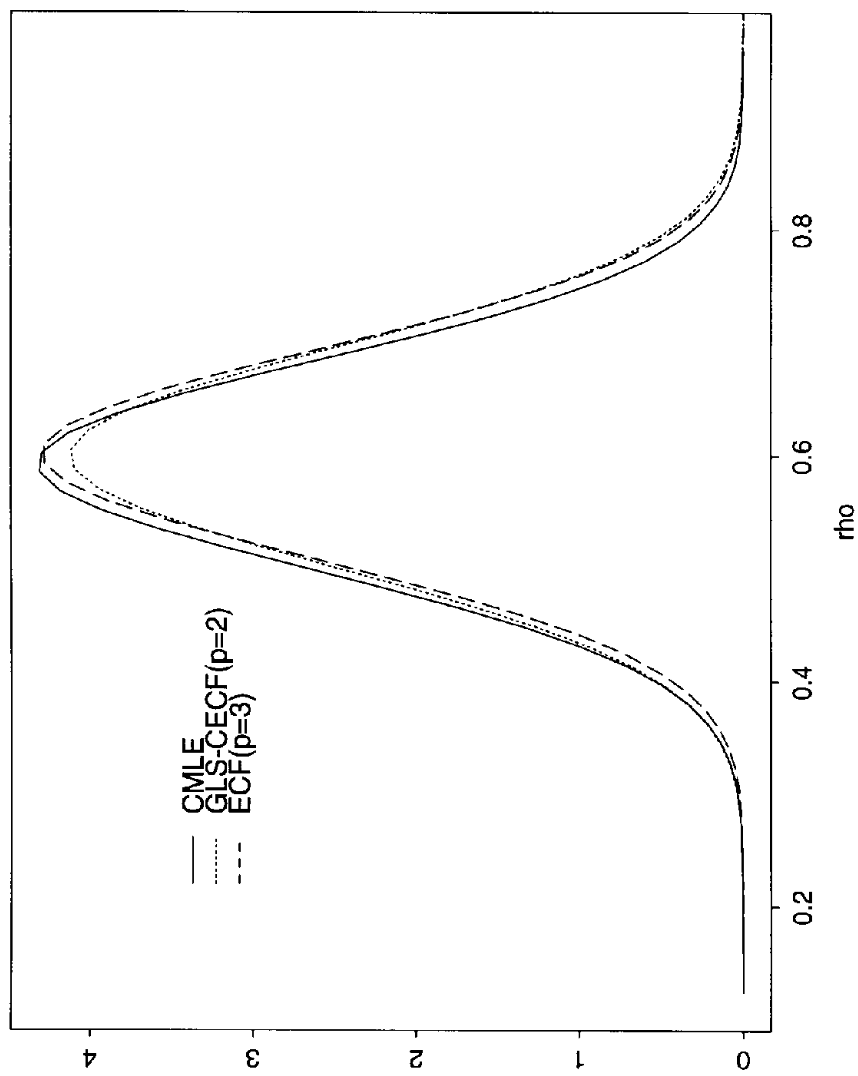
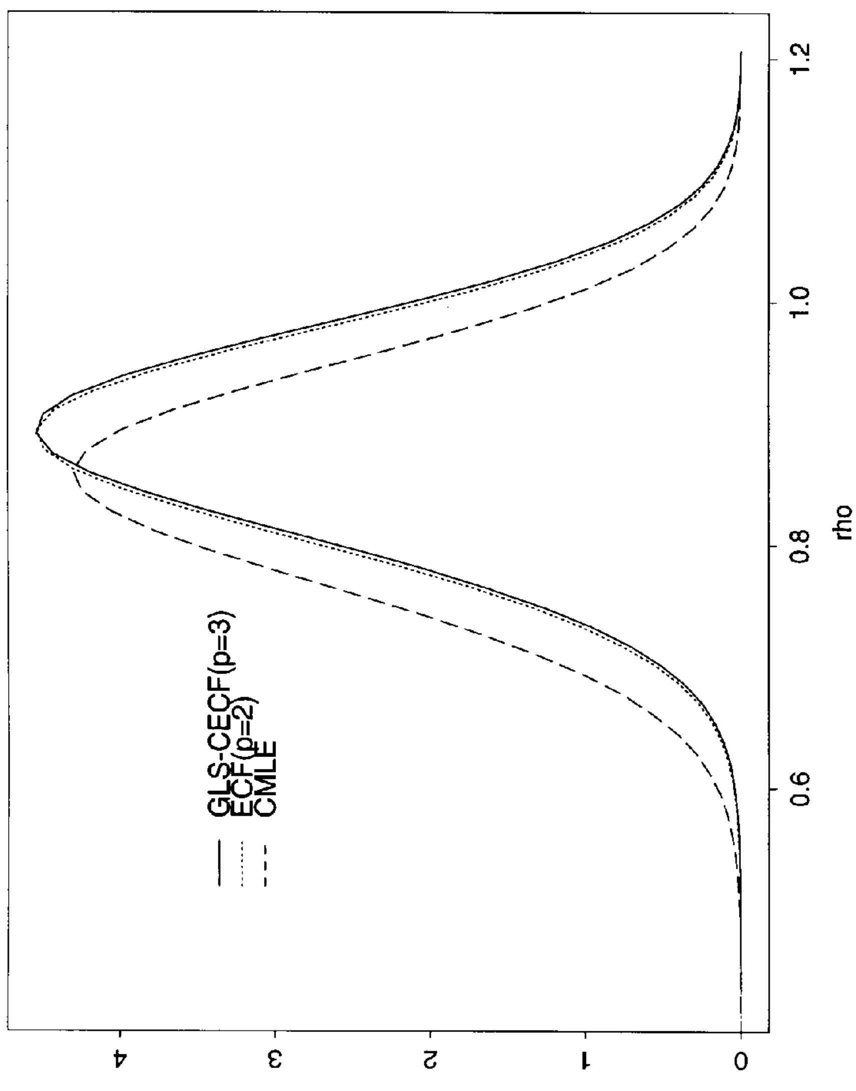


Figure 4: Density Function of rho in ARMA



References

[1] Amemiya, T. (1985). *Advanced Econometrics.* Blackwell, Oxford.

[2] Bollerslev, T., R.Y. Chou, and K.F. Kroner. (1992). ARCH modeling in finance: A review of the theory and empirical evidence. *Journal of Econometrics.* **52** 5–59.

[3] Csörgő, S. (1981). Limit behaviour of the empirical characteristic function. *The Annals of Probability.* **9** 130–144.

[4] Danielsson, J. (1994). Stochastic volatility in asset prices. Estimation with simulated maximum likelihood. *Journal of Econometrics.* **64** 375–400.

[5] Duffie, D. and K.J. Singleton. (1993). Simulated moments estimation of Markov models of asset prices. *Econometrica.* **61** 929–952.

[6] Fan, Y. (1996). Goodness-of-fit tests for a multivariate distribution by the empirical characteristic function. Journal of Multivariate Analysis, Forthcoming.

[7] Feuerverger, A. (1990). An efficiency result for the empirical characteristic function in stationary time-series models. *The Canadian Journal of Statistics.* **18** 155–161.

[8] Feuerverger, A. and P. McDunnough. (1981). On some Fourier methods for inference. *Journal of the American Statistical Association.* **76** 379–387.

[9] Feuerverger, A. and P. McDunnough. (1981). On the efficiency of empirical characteristic function procedures. *Journal of the Royal Statistical Society, Series B, Methodological.* **43** 20–27.

[10] Feuerverger, A. and R.A. Mureika. (1977). The empirical characteristic function and its applications. *The Annals of Statistics.* **5** 88–97.

[11] Gel'Fan, A. (1964). *Generalized Functions*, volume 1. Academic Press.

[12] Ghysels, E., A.C. Harvey, and E. Renault. (1996). *Stochastic Volatility.* In *Handbook of Stististics-14.*

[13] Hansen, L.P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica.* **50** 1029–1054.

[14] Heathcote, C.R. (1977). The integrated squared error estimation of parameters. *Biometrika.* **64** 255–264.

[15] Huber, P. (1981). *Robust Statistics.* Wiley, New York.

[16] Ibragimov, I. and Y. Linnik. (1971). *Independent and Stationary Sequences of Random Variables.* Groningen, Wolters-Noordhoff.

[17] Knight, J.L. and S.E. Satchell. (1996). Estimation of stationary stochastic processes via the empirical characteristic function. The University of Western Ontario.

[18] Knight, J.L. and S.E. Satchell. (1997). The cumulant generating function estimation method. *Econometric Theory.* **13** 170–184.

[19] Martin, R.D. and V.J. Yohai. (1986). Influence functionals for time series. *The Annals of Statistics.* **14** 781–818.

[20] Parzen, E. (1962). On estimation of a probability density function and mode. *The Annals of Mathematical Statistics.* **33** 1065–1076.

[21] Paulson, A.S., E.W. Holcomb, and R.A. Leitch. (1975). The estimation of the parameters of the stable laws. *Biometrika.* **62** 163–170.

[22] Press, S.J. (1972). Estimation in univariate and multivariate stable distributions. *Journal of the American Statistical Association.* **67** 842–846.

[23] Quandt, R.E. and J.B. Ramsey. (1978). Estimating mixtures of normal distributions and switching regressions. *Journal of the American Statistical Association.* **73** 730–738.

[24] Schmidt, P. (1982). An improved version of the Quandt-Ramsey MGF estimator for mixtures of normal distributions and switching regressions. *Econometrica.* **50** 501–524.

[25] Stout, W.F. (1974). *Almost Sure Convergence.* Academic Press, New York.

[26] Titterington, D.M., A.F.M. Smith, and U.E. Makov. (1985). *Statistical Analysis of Finite Mixture Distributions.* Chichester, Toronto.

[27] White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica.* **50** 1–26.

[28] White, H. (1984). *Asymptotic Theory for Econometricians.* Academic Press, New York.

[29] White, H. and I. Domowitz. (1984). Nonlinear regression with dependent observations. *Econometrica.* **50** 146–161.

[30] Whittle, P. (1983). *Prediction and Regulation.* University of Minnesota Press.

[31] Yao, Q. and B. Morgan. (1996). Empirical transformation estimation. University of Kent.

[32] Zinde-Walsh, V. (1988). Some exact formulae for autoregressive moving average processes. *Econometric Theory.* **4** 384–402.