Econometric Theory, **18**, 2002, 691–721. Printed in the United States of America.
DOI: 10.1017.S026646660218306X

EMPIRICAL CHARACTERISTIC FUNCTION IN TIME SERIES ESTIMATION

John L. Knight
University of Western Ontario

Jun Yu
University of Auckland

Because the empirical characteristic function (ECF) is the Fourier transform 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 via the ECF for strictly stationary processes. Under some regularity conditions, the resulting estimators are shown to be consistent and asymptotically normal. The method is applied to estimate the stable autoregressive moving average (ARMA) models. For the general stable ARMA model for which the maximum likelihood approach is not feasible, Monte Carlo evidence shows that the ECF method is a viable estimation method for all the parameters of interest. For the Gaussian ARMA model, a particular stable ARMA model, the optimal weight functions and estimating equations are given. Monte Carlo studies highlight the finite sample performances of the ECF method relative to the exact and conditional maximum likelihood methods.

1. INTRODUCTION

Maximum likelihood (ML) estimation is one of the most widely used estimation methods. One reason is that the ML estimator is asymptotically efficient under appropriate regularity conditions. To implement the ML 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 econometrics where the ML approach is difficult to implement, both in the independent and identically distributed (i.i.d.) case and in the de-

We thank Donald Andrews (the co-editor) and two anonymous referees for constructive comments that have substantially improved the paper. We also thank Victor Aguirregabiria, Reg Kulperger, Brendan McCabe, Don McLeish, Peter Phillips, Alan Rogers, Jim Talman, Jian Yang, and Hao Yu for their suggestions and helpful discussions. Seminar participants at the University of Alberta, the University of Western Ontario (both Department of Economics and Department of Statistical and Actuarial Science), Bank of Canada, 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) and the University of Auckland Research Committee. Address correspondence to: Jun Yu, Department of Economics, University of Auckland, Private Bag 92019, Auckland, New Zealand; e-mail: j.yu@auckland.ac.nz.

 0266-4666/02 $9.50

pendent case. In the i.i.d. case, processes sometimes have an unbounded likelihood function in the parameter space. Examples include the mixture of normals and switching regression models (Titterington, Smith, and Markov, 1985). In other examples, such as the stable distributions, however, the density functions of the processes cannot 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 the stable ARMA model, the stochastic volatility (SV) model (Ghysels, Harvey, and Renault, 1996), and the process that is compound Poisson-normal and where the Poisson intensity is random, possibly dependent on past information in the series (Knight and Satchell, 1998). Some of these models have found wide use in economics and finance.

The usual response to such difficulties arising from the likelihood is to use alternative methods. For example, one might use GMM (Hansen, 1982), quasi-maximum likelihood (QML) (White, 1982), or simulation-based methods (Duffie and Singleton, 1993). Although these methods are consistent under regular conditions, some of them are not asymptotically efficient. Furthermore, the small sample properties 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 stationarity and normality (Epps, 1987, 1988), testing for independence (Feuerverger, 1990; Hong, 1999), testing for symmetry (Feuerverger and Mureika, 1977), and parameter estimation.[1] The main justification for the ECF method is that the characteristic function (CF) has a one-to-one correspondence with the distribution function (DF) 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 DF. The theory for the ECF in the i.i.d. case is well understood (Feuerverger and Mureika, 1977; Csörgő, 1981). Surprisingly, however, the ECF in the dependent case has received much less 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 are established under some regular conditions. Monte Carlo simulations are performed to study the finite sample properties of the ECF method for stable ARMA models.

The paper is organized as follows. The next section briefly reviews what has been done for the ECF method in estimation of time series models, proposes the ECF method in a more general framework, and obtains the asymptotic properties of the ECF estimator. Section 3 discusses the estimation procedure of stable and Gaussian ARMA models via the ECF. Section 4 compares the relative finite sample performances between the ECF method and some other methods for both models. Section 5 concludes the paper.

2. LITERATURE REVIEW AND ASYMPTOTIC PROPERTIES

2.1. Literature Review

The basic idea for the ECF method is to minimize some distance measure between the ECF and CF. Although the literature in the i.i.d. case is extensive, little research has been reported in the dependent case. To our knowledge, only two papers are relevant to parameter estimation. We 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 $\{y_1, y_2, \ldots, y_T\}$. The overlapping blocks for $\{y_1, y_2, \ldots, y_T\}$ are defined by $\boldsymbol{x}_j = (y_j, \ldots, y_{j+p})$ $j = 1, \ldots, T-p$. Thus each block has p observations overlapping with the adjacent blocks. The CF of each block is defined by $c(\boldsymbol{r};\boldsymbol{\theta}) = E(\exp(i\boldsymbol{r}'\boldsymbol{x}_j))$, where $\boldsymbol{r} = (r^1, \ldots, r^{p+1})'$. The ECF is defined by $c_n(\boldsymbol{r}) = (1/n)\sum_{j=1}^{n} \exp(i\boldsymbol{r}'\boldsymbol{x}_j)$, where $n = T - p$. By construction the ECF is the sample counterpart of the CF and contains the information of the data, whereas the CF contains the information of the parameters; $\boldsymbol{r}$ is the transform variable.

The estimation procedure is to match the ECF with CF. Feuerverger (1990) and Knight and Satchell (1996) propose matching the ECF with CF over a grid of finite points, and hence the procedure is called the discrete ECF (DECF) method.[2] Feuerverger proves that under some regularity conditions, the resulting estimators can be made to have arbitrarily high asymptotic efficiency provided that p (fixed) 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 detail the application of the DECF method to stationary stochastic processes and give a multistep procedure. When applying the proposed procedure to estimate a Gaussian MA(1) model, they find that the DECF method is a viable alternative but is outperformed by the ML method.

The finding is not surprising and can be explained intuitively. Because matching the ECF with CF over a grid of finite points is equivalent to matching a finite number of moments, the DECF method is, in essence, equivalent to generalized method of moments (GMM). Thus as with GMM it is not obvious how many and which moments to choose, the difficulties of using the DECF method are how many and which points need to be used.

2.2. Proposed Procedure

Observing the difficulties involved in the DECF method, we propose the ECF method, which minimizes the integral

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

or

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

or solves the estimating equation

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

where $G(\boldsymbol{r})$, $g(\boldsymbol{r})$, and $w_{\boldsymbol{\theta}}(\boldsymbol{r})$ are the weight functions. Under suitable regularity conditions these three procedures are equivalent. It is important to note that for the three procedures to be equivalent $w_{\boldsymbol{\theta}}(\boldsymbol{r})$ must depend on $\boldsymbol{\theta}$ because the derivative of $c(\boldsymbol{r};\boldsymbol{\theta})$ is essentially involved in it. The motivation for the suggested approach is that two distribution functions are equal if and only if their CF's are equal (Lukacs, 1970, p. 28). 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 proposed procedure includes the DECF method as a special case. If a continuous weight function is used, the procedure basically matches all the moments continuously. One advantage of using a continuous weight function is that one no longer needs to choose the transform variable, $\boldsymbol{r}$, because it is simply integrated out. In this paper we refer to the ECF method with a continuous weight as the continuous ECF (CECF) method. In fact this is the procedure suggested by Paulson, Holcomb, and Leitch (1975) in the i.i.d. case. Formally, Paulson et al. choose $G(r)$ to be a measure corresponding to a normal variate, i.e., $g(r) = \exp(-r^2)$. The ECF with an exponential weight is referred to as WLS of the CECF (WLS-CECF) method. The advantages of using an exponential weight are twofold. First, it 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 an exponential weight, the integral (2.2) can be numerically calculated by Hermitian quadrature or Monte Carlo integration. In general, unfortunately, WLS-CECF will not result in the efficient estimator because the exponential weight is not optimal. By using the Parseval theorem, we can obtain the weight function, $w^*_{\boldsymbol{\theta}}(\boldsymbol{r})$,

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

Feuerverger (1990) shows that provided p is sufficiently large the preceding weight is optimal in the sense that based on equation (2.3) and $w^*_{\boldsymbol{\theta}}(\boldsymbol{r})$, the resulting estimator can achieve arbitrarily high efficiency. In this paper the ECF procedure with weight (2.4) is referred to as GLS of the CECF (GLS-CECF) method. To use the GLS-CECF method, however, the conditional score function must have an analytical expression. When this is not the case, this weight function has to be approximated. It will be shown in Section 2.4 that for pure autoregressive (AR) processes, the GLS-CECF with a fixed p achieves full efficiency. This property, however, does not apply to general stationary processes.

2.3. Asymptotic Properties

The asymptotic properties of the ECF estimator in the i.i.d. case have been obtained by Heathcote (1977). In this section, we establish some regularity conditions under which the ECF estimator is consistent and asymptotically normally distributed in the dependent case. For simplicity of notation we will only deal with the estimators resulting from equation (2.1).[3] Equation (2.1) consists of minimizing a distance function and hence is in the class of extremum estimators (or M-estimators by Huber, 1981).

For any fixed p, we list the following regularity conditions.

(A1) $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ where the parameter space $\boldsymbol{\Theta} \subset R^K$ is a compact set with $\boldsymbol{\theta}_0 \in \mathrm{Int}(\boldsymbol{\Theta})$.

(A2) With probability one, $I_n(\boldsymbol{\theta})$ is twice continuously differentiable under the integral sign with respect to $\boldsymbol{\theta}$ over $\boldsymbol{\Theta}$.

(A3) Sequence $\{y_j\}$ is strictly stationary and ergodic.

(A4) Let $I_0(\boldsymbol{\theta}) = \int \ldots \int |c(\boldsymbol{r};\boldsymbol{\theta}_0) - c(\boldsymbol{r};\boldsymbol{\theta})|^2 \, dG(\boldsymbol{r})$ and $I_0(\boldsymbol{\theta}) = 0$ only if $\boldsymbol{\theta} = \boldsymbol{\theta}_0$.

(A5) $K(\boldsymbol{x};\boldsymbol{\theta})$ is a measurable function of $\boldsymbol{x}$ for all $\boldsymbol{\theta}$ and bounded, where

$$K(\boldsymbol{x};\boldsymbol{\theta}) = \int \ldots \int \left\{ (\cos(\boldsymbol{r}'\boldsymbol{x}) - \mathrm{Re}\, c(\boldsymbol{r};\boldsymbol{\theta})) \frac{\partial \, \mathrm{Re}\, c(\boldsymbol{r};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right. \\ \left. + (\sin(\boldsymbol{r}'\boldsymbol{x}) - \mathrm{Im}\, c(\boldsymbol{r};\boldsymbol{\theta})) \frac{\partial \, \mathrm{Im}\, c(\boldsymbol{r};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right\} dG(\boldsymbol{r}).$$

(A6) $B(\boldsymbol{\theta}_0) = \int \ldots \int (\partial\, c(\boldsymbol{r};\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta})(\partial c(\boldsymbol{r};\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta}')\, dG(\boldsymbol{r})$ is nonsingular and $\partial^2 c(\boldsymbol{r};\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'$ is uniformly bounded by a G-integrable function over $\boldsymbol{\Theta}$.

(A7) Let $\Im_j$ be a σ-algebra such that $\{K_j, \Im_j\}$ is an adapted stochastic sequence, where $K_j = K(\boldsymbol{x}_j;\boldsymbol{\theta})$. We can think of $\Im_j$ as being the σ-algebra generated by the entire current and past history of K_j. Let $\nu_j = E[K_0|K_{-j}, K_{-j-1}, \ldots] - E[K_0|K_{-j-1}, K_{-j-2}, \ldots]$ for $j \geq 0$. Assume that $E(K_0|\Im_{-m})$ converges in mean square to 0 as $m \to \infty$ and $\sum_{j=0}^{\infty} E[\nu_j'\nu_j]^{1/2} < \infty$.

(A8) $G(\boldsymbol{r})$ is a nondecreasing function with bounded total variation taken to be 1.

Remark 2.1. (A1) ensures the compactness of the parameter set whereas (A2) guarantees the continuity of $I_n(\boldsymbol{\theta})$. Assumptions (A3) and (A5)–(A7) provide sufficient conditions for a strong law of large numbers and a central limit theorem. According to (A8), $G(\boldsymbol{r})$ can be a distribution function; however, it rules out the CECF method with a constant weight function (such as $G(\boldsymbol{r}) = a\boldsymbol{r}$ for a fixed constant a). Assumption (A7) is the analogue to Assumption 3.5 in Hansen (1982) and holds under suitable mixing conditions. Assumption (A4) is the identification condition. Without this assumption, the estimator can be inconsistent. For example, for an MA(10) process, $Y_t = \varepsilon_t - \phi\varepsilon_{t-10}$, the moving blocks with $p < 10$ contain no information on ϕ, and hence the ECF leads to inconsistent estimates.[4] This case is ruled out by (A4), however. Identification will only be achieved with an appropriate choice of p. In the example of the MA(10) model, any $p \geq 10$ will ensure it.

THEOREM 2.1 *Let* $\hat{\boldsymbol{\theta}}_n = \arg\min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} I_n(\boldsymbol{\theta})$. *Suppose that Assumptions (A1)–(A4) and (A8) hold,* $\hat{\boldsymbol{\theta}}_n \xrightarrow{a.s.} \boldsymbol{\theta}_0$. *If, in addition, Assumptions (A5)–(A7) hold,* $\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow{d} N(0, B^{-1}(\boldsymbol{\theta}_0)\, A(\boldsymbol{\theta}_0)B^{-1}(\boldsymbol{\theta}_0))$, *where* $A(\boldsymbol{\theta}_0) = \operatorname{var}(K(\boldsymbol{x}_1;\boldsymbol{\theta}_0)) + 2\sum_{j=2}^{\infty} \operatorname{cov}(K(\boldsymbol{x}_1;\boldsymbol{\theta}_0), K(\boldsymbol{x}_j;\boldsymbol{\theta}_0))$.

Proof. All proofs are in the Appendix.

Remark 2.2. This theorem is the analogue, for the dependent case, of the main theorem in Heathcote (1977). This result in general is not applicable to the estimator based on (2.3). However, for the estimator in (2.3) we note that it is merely a GMM estimator where

$$h(\boldsymbol{\theta}, \boldsymbol{x}_j) = \int \cdots \int w_{\boldsymbol{\theta}}(\boldsymbol{r})(e^{i\boldsymbol{r}'\boldsymbol{x}_j} - c(\boldsymbol{r};\boldsymbol{\theta}))\, d\boldsymbol{r}$$

and $E(h(\boldsymbol{\theta}, \boldsymbol{x}_j)) = 0$. The sample average, $\hat{g}_n(\boldsymbol{\theta})$, of $h(\boldsymbol{\theta}, \boldsymbol{x}_j)$ is just the estimating equation (2.3), i.e.,

$$\hat{g}_n(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n} h(\boldsymbol{\theta}, \boldsymbol{x}_j) = \int \cdots \int w_{\boldsymbol{\theta}}(\boldsymbol{r})(c_n(\boldsymbol{r}) - c(\boldsymbol{r};\boldsymbol{\theta}))\, d\boldsymbol{r}.$$

With this observation, the asymptotic theory concerning this estimator is well established; see Hansen (1982) or Hamilton (1994, Ch. 14).

Remark 2.3. Estimates of $A(\boldsymbol{\theta}_0)$ and $B(\boldsymbol{\theta}_0)$ are necessary to calculate the asymptotic variance of the ECF estimator. The expression $B(\boldsymbol{\theta}_0)$ can be consistently estimated by $B(\hat{\boldsymbol{\theta}}_n)$, whereas a consistent estimate of $A(\boldsymbol{\theta}_0)$ can be obtained using methods suggested by Andrews (1991) and Newey and West (1994).

2.4. Choice of Block Size

Theorem 2.1 ensures consistency and asymptotic normality for any fixed p provided the regularity conditions are satisfied. Unfortunately, it is not very clear how p affects efficiency, and the covariance of the ECF estimator does not provide an obvious guide to the choice of p. On the other hand, however, the choice of p can have an impact on the efficiency of the ECF estimator, as the moving blocks with a different p may contain different amounts of information in the sample. In general, there is a trade-off between large p and small p. If we choose a large p, we would expect the moving blocks to contain more information, and hence the ECF should be more efficient. In fact, according to the inversion theorem, we can obtain the density by inverting the CF. Provided the Fourier inversion can be implemented efficiently, the ECF is asymptotically equivalent to ML and hence achieves full efficiency when p is allowed to go to infinity (e.g., $p = O(T)$ when $T \to \infty$). Unfortunately, such inversions involve high-dimensional integration 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 moving blocks may not retain all the important information of the series.

We will discuss how the choice of p affects efficiency in a Monte Carlo study in the ARMA framework. Further theoretical investigation on this effect is left for future research. We believe, however, that the choice of p in the context of the ECF estimation is related to the choice of the block size in the moving block bootstrap method and stationary bootstrap method. The choice of p is also related to the dimension of the minimal sufficient statistics. A few observations on this are reported in the next section when dealing with Gaussian ARMA models.

For the stationary AR(l) process, however, we show that $p = l$ is optimal for GLS-CECF to be asymptotically efficient. A similar result for a slightly different characterization can be found in Singleton (2001, Lemma 5.1) with $l = 1$.

PROPOSITION 2.1. *Assume* $\{y_j\}$ *is a stationary* AR(l) *process and let* $p = l$. *The* GLS-CECF *estimator defined by equation (2.3) with the weight given by (2.4) is a conditional* ML (CML) *estimator and hence asymptotically efficient.*

Remark 2.4. It is known that for the AR(l) process, $(y_j, \ldots, y_{j+l})$ retains all the dependence in the sample. In fact setting $p > l$ results in an efficiency loss as we are conditioning on a larger set than necessary and hence ignoring information.

3. ESTIMATION OF STABLE ARMA MODELS

There has recently been considerable interest in modeling financial time series using the ARMA models with infinite variance (Brockwell and Davis, 1991). In this section we detail the proposed ECF method in the estimation of the stable ARMA model.

3.1. Stable ARMA Models

The model under consideration is an invertible ARMA(l, m) model of the form

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

where $\varepsilon_t \sim$ i.i.d. $S_\alpha(\sigma, \beta, \mu)$ and $\alpha(\in (0,2])$, $\beta(\in [-1,1])$, σ, and μ are, respectively, index, skewness, scale, and location parameters of the error term distribution. Let $\Psi(L) = 1 - \rho_1 L - \cdots - \rho_l L^l$ and $\Phi(L) = 1 - \phi_1 L - \cdots - \phi_m L^m$. Assume that $\Psi(L)$ and $\Phi(L)$ are different from zero for $|L| \leq 1$ and have no common roots. Unless specified we assume $\sigma = 1$ and $\mu = 0$ even though the techniques presented subsequently have broader applicability. Let $\boldsymbol{\theta}$ be the parameters of interest, which include not only the AR and MA coefficients but also the parameters in the distribution of ε_t.

We now summarize the relevant facts associated with the stable distribution and refer the reader to Samorodnitsky and Taqqu (1994) for a detailed statistical description. The stable distribution is usually characterized by the CF given by

$$c(t) = \begin{cases} \exp\left\{i\mu t - \sigma|t|^\alpha\left[1 - i\beta\,\mathrm{sign}(t)\tan\left(\dfrac{\pi\alpha}{2}\right)\right]\right\} & \text{if } \alpha \neq 1 \\ \exp\left\{i\mu t - \sigma|t|^\alpha\left[1 + i\beta\,\dfrac{2}{\pi}\,\mathrm{sign}(t)\ln(|t|)\right]\right\} & \text{if } \alpha = 1 \end{cases}. \tag{3.2}$$

It can be skewed to the left or right, depending on the sign of β, and is symmetric when $\beta = 0$. If $\alpha = 2$, it is a normal distribution. If $\alpha < 2$, it has fatter tails than the normal distribution and the pth absolute moment does not exist for any $p > \alpha$, and in particular the variance is infinite. If $1 < \alpha < 2$, the density function has no closed form, and it has to be calculated numerically by Fourier inverting (3.2). When $\alpha > 1$, the mean exists and equals μ. Unless $\beta = 0$, the mean is different from the median.

As a result of the fact that the stable distribution does not always have a closed form, ML estimation is often very difficult (for a detailed discussion, see Calder and Davis, 1998). However, under suitable regularity conditions, the estimators do have good asymptotic properties.

Estimation of the stable ARMA model via alternative methods such as QML and GMM also presents difficulties. For example, QML is infeasible because the variance of the error term may be infinite. For GMM care must be taken when choosing moment conditions because the stable distribution does not have a finite absolute moment of order higher than α.

3.2. Estimation of Stable ARMA Models

There has been a large body of literature dealing with the estimation problem in the stable ARMA model (see Davis, Knight, and Liu, 1992; Mikosch, Klüppelberg, and Adler, 1995; Davis, 1996; Embrechts, Klüppelberg, and Mikosch, 1997; Adler, Feldman, and Taqqu, 1998). Four estimation methods have received attention in this literature and are briefly reviewed here. These are least squares (LS), least absolute deviations (LAD), conditional maximum likelihood (CML), and the Whittle method.

The first three estimators are all M-estimators and can be treated jointly. An M-estimator $\hat{\boldsymbol{\theta}}_M$ of $\boldsymbol{\theta}$ is defined by $\min_{\boldsymbol{\theta}} \sum_{t=1}^{T} m(\varepsilon_t(\boldsymbol{\theta}))$, where $m(x) = x^2$ for LS, $m(x) = |x|$ for LAD, $m(x) = -\log f(x)$ with $f(x)$ being the density of ε_t for CML, and the sequence $\{\varepsilon_t(\boldsymbol{\theta})\}_{t=1}^{T}$ can be calculated from the realization $\{y_t\}_{t=1}^{T}$ using an iterative formula such as (3.6), which follows, with some initial assumptions. For detailed discussion of these three methods see Davis et al. (1992) with regard to the stable AR model and Davis (1996) and Calder and Davis (1998) with regard to the stable ARMA model. Calder and Davis discuss the CML estimation of the symmetric stable ARMA models.

The Whittle method was studied in Klüppelberg and Mikosch (1993, 1994), Mikosch et al. (1995), and Embrechts et al. (1997) to estimate the stable ARMA model. It is based on the frequency domain and defined by $\arg\min_{\boldsymbol{\theta}} (2\pi/T) \times (I_T(\omega_j)/g(\omega_j;\boldsymbol{\theta}))$, where $I_T(\lambda)$ is the periodogram and $g(\lambda;\boldsymbol{\theta})$ the power transfer function, both defined in Embrechts et al.

It has been shown that the rates of convergence are $n^{1/\alpha}$ for LAD/ML and $(n/\ln n)^{1/\alpha}$ for LS/Whittle (Calder and Davis, 1998). Both of these rates compare favorably to the rate for the ECF, $n^{1/2}$. On the other hand, however, LS, LAD, and the Whittle method only estimate the AR and MA coefficients. Although CML estimates the parameters in the innovation, it is numerically more difficult to estimate asymmetric models. For inference purposes, however, estimation of the innovation parameters is required. An advantage of the ECF method is that, as a consistent procedure, it estimates all of the parameters simultaneously.

To use the ECF method, we have to ensure that the joint CF of the stable ARMA model has a closed form. In the proposition that follows we provide the two-dimensional joint CF of the stable ARMA$(1,1)$ model given by $Y_t = \rho Y_{t-1} + \varepsilon_t - \phi\varepsilon_{t-1}$ where $\varepsilon_t \sim S_\alpha(1,\beta,0)$. Higher dimensional joint CF can be obtained in a similar fashion, as can the joint CF of other stable ARMA models.

PROPOSITION 3.1. *The two-dimensional joint* CF *of the stable* ARMA$(1,1)$ *model is*

$$
\begin{aligned}
&c(r_1,r_2;\boldsymbol{\theta}) \\
&\quad = \exp\left\{-|r_2|^\alpha - |r_1+(\rho-\phi)r_2|^\alpha - \frac{|r_1+\rho r_2|^\alpha|\rho-\phi|^\alpha}{1-|\rho|^\alpha}\right\} \\
&\qquad \times \exp\left\{i\beta\tan\frac{\pi\alpha}{2}\left[|r_2|^\alpha \operatorname{sign}(r_2) + |r_1+(\rho-\phi)r_2|^\alpha \right.\right. \\
&\qquad\qquad \times \operatorname{sign}(r_1+(\rho-\phi)r_2) \\
&\qquad\qquad \left.\left. + \frac{|r_1+\rho r_2|^\alpha \operatorname{sign}(r_1+\rho r_2)|\rho-\phi|^\alpha \operatorname{sign}(\rho-\phi)}{1-\operatorname{sign}(\rho)|\rho|^\alpha}\right]\right\}. \qquad \textbf{(3.3)}
\end{aligned}
$$

Because the conditional score of the stable ARMA model has no closed form, the GLS-CECF procedure is not feasible. Instead we can use the WLS-CECF procedure with an exponential weight function, $\exp(-a\boldsymbol{r}'\boldsymbol{r})$, where a is a positive constant. Such a weight function is used for numerical tractability. If $p+1$ and hence the dimension of the integral (2.2) is small (e.g., $p \le 4$), the integral can be calculated numerically by Hermitian quadrature (e.g., Press, Teukolsky, Vetterling, and Flannery, 1992, p. 144). If $p+1$ is large (e.g., $p \ge 5$) the integral can be calculated numerically by lattice methods (e.g., Sloan and Joe, 1994) or Monte Carlo integration (e.g., Press et al., 1992), but it is computationally more expensive.

3.3. Gaussian ARMA Models

If $\varepsilon_t \sim$ i.i.d. $N(0,\sigma^2)$, a special case of the stable distribution, equation (3.1) represents a Gaussian ARMA model and hence is a special case of the stable ARMA model. In contrast to the general stable ARMA model, the Gaussian ARMA model has a closed form expression for the likelihood function. Denote the covariance matrix of $\boldsymbol{y} = (y_1,\ldots,y_T)'$ by $\sigma^2\Omega$. The log-likelihood function is

$$l(\boldsymbol{\theta}) = -\frac{T}{2}\log(2\pi) - \frac{1}{2}\log|\Omega| - \frac{T}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\boldsymbol{y}'\Omega^{-1}\boldsymbol{y}. \tag{3.4}$$

For AR and MA(1) models, the exact inverse of Ω has an analytical expression, and hence it is straightforward to maximize (3.4) numerically. In contrast to AR and MA(1) models, however, other ARMA models have no closed form for the inverse of Ω (see Whittle, 1983). Inverting this $T \times T$ matrix is numerically intensive for a large value of T. To overcome the difficulties involved in the inversion, the state space representation and the Kalman filter can be used to evaluate the exact likelihood. Alternatively, the conditional likelihood function is often maximized in practice.

It is common to obtain the conditional likelihood function conditional on both y and ε. One option is to set the initial y and ε equal to their expected values. For example, for the Gaussian ARMA(1,1) model, the conditional log-likelihood can be obtained by

$$\log f(y_T,\ldots,y_1|y_0=0,\varepsilon_0=0) = -\frac{T}{2}\log(2\pi) - \frac{T}{2}\log\sigma^2 - \sum_{t=1}^{T}\frac{\varepsilon_t^2}{2\sigma^2}, \tag{3.5}$$

where $\{\varepsilon_t\}_{t=1}^T$ can be calculated from $\{y_t\}_{t=1}^T$ by the following iterative formula:

$$\varepsilon_t = y_t - \rho y_{t-1} + \phi\varepsilon_{t-1}. \tag{3.6}$$

The resulting estimator is referred to as the CMLE1. Another option is proposed by Box and Jenkins (1976, p. 221), which is to condition the likelihood on the first l realizations of the sample and set the innovations to 0 with iteration (3.6) started on $t = l+1$. The resulting estimator is referred to as the CMLE2.

We can also estimate the Gaussian ARMA model using the proposed ECF method. The CF of the Gaussian ARMA model is $\exp(-(\sigma^2/2)\boldsymbol{r}'\Omega\boldsymbol{r})$. In this paper two ECF methods are actually used to estimate the Gaussian ARMA model. One is the WLS-CECF method with $g(\boldsymbol{r}) = \exp(-a\boldsymbol{r}'\boldsymbol{r})$, where a is a positive constant. With the exponential weight for the Gaussian ARMA model, one can show that

$$I_n(\boldsymbol{\theta}) \propto -\frac{2}{n}\pi^{(p+1)/2}|A|^{-1/2}\sum_{j=1}^{n}\exp\left(-\frac{1}{4}\boldsymbol{x}_j'A^{-1}\boldsymbol{x}_j\right) + \pi^{(p+1)/2}|B|^{-1/2},$$

where $A = (\sigma^2/2)\Omega + aI$ and $B = \sigma^2\Omega + aI$ with I an identity matrix. This means that $I_n(\boldsymbol{\theta})$ can be expressed as known elementary functions and one can avoid numerical integration when calculating (2.2) for the Gaussian model. Hence, numerical optimizations on (2.2) can be done efficiently and accurately. Note however, that whereas Ω needs inverting, it is now of dimension $p + 1$, considerably less than T as in ML.

The other ECF method used is GLS-CECF based on equations (2.3) and (2.4). In Theorem 3.1 and Corollary 3.1 we give the analytical expressions of the weight functions and estimating equations of the GLS-CECF method for the Gaussian ARMA model. In the subsequent remarks we comment on the choice of p for AR models and general ARMA models.

THEOREM 3.1. *Let $\{y_1,\dots,y_T\}$ be a finite realization of the Gaussian* ARMA(l,m) *model with parameter vector $\boldsymbol{\theta} = (\sigma^2,\boldsymbol{\rho}) = (\sigma^2,\rho_1,\dots,\rho_l,\phi_1,\dots,\phi_m)$ such that*

$$\boldsymbol{x}_j = (y_j,\dots,y_{j+p})' \sim N(0,\sigma^2\Omega).$$

Then

$$(y_{j+p}|y_j,\dots,y_{j+p-1}) \sim N(f_1(\boldsymbol{\rho})y_j + \cdots + f_p(\boldsymbol{\rho})y_{j+p-1},\sigma^2 g(\boldsymbol{\rho})), \tag{3.7}$$

with

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

where $(f_1(\boldsymbol{\rho}),\dots,f_p(\boldsymbol{\rho})) = \Omega_{21}\Omega_{11}^{-1} = f_p'$, say, and $g(\boldsymbol{\rho}) = \Omega_{22} - \Omega_{21}\Omega_{11}^{-1}\Omega_{12}$ with the Ω_{ij}'s the obvious partitions of Ω. Letting $f' = (-f_p',1)$ we have $A = ff'$. Further, letting $B_k = \partial A/\partial\boldsymbol{\rho}_k$, $k = 1,\dots,l+m$ we can readily show that whereas A is of rank 1, B_k is always of rank 2. Also, because both A and B_k are symmetric they are diagonable, i.e., $A = MD_aM', D_a = \mathrm{diag}(\lambda,0,\dots,0)$ and $B_k = H_kD_kH_k', D_k = \mathrm{diag}(\lambda_k^1,\lambda_k^2,0,\dots,0)$, where λ,λ_k^1, and λ_k^2 are the eigenvalues with M and H_k the orthogonal matrices formed from the eigenvectors. Then, *the weight functions* $w_{\boldsymbol{\theta}}^{*}(\boldsymbol{r})$ *corresponding to* σ^2 *and the elements in* $\boldsymbol{\rho}$ *are given by*

$$w_{\sigma^2}^{*}(\boldsymbol{r}) = -\frac{1}{2\sigma^2}\,\delta(r^1)\ldots\delta(r^{p+1}) - \frac{1}{2\sigma^4 g(\boldsymbol{\rho})}\,\lambda\delta''(s^1)\delta(s^2)\ldots\delta(s^{p+1}) \tag{3.8}$$

$$\begin{aligned} w_{\boldsymbol{\rho}_k}^{*}(\boldsymbol{r}) = {} & -\frac{g_k(\boldsymbol{\rho})}{2g(\boldsymbol{\rho})}\,\delta(r^1)\ldots\delta(r^{p+1}) - \frac{g_k(\boldsymbol{\rho})}{2\sigma^2 g^2(\boldsymbol{\rho})}\,\lambda\delta''(s^1)\delta(s^2)\ldots\delta(s^{p+1}) \\ & + \frac{1}{2\sigma^2 g(\boldsymbol{\rho})}\,[\lambda_k^1\delta''(t_k^1)\delta(t_k^2)\ldots\delta(t_k^{p+1}) + \lambda_k^2\delta(t_k^1)\delta''(t_k^2)\ldots\delta(t_k^{p+1})], \end{aligned} \tag{3.9}$$

where $(s^1,\ldots,s^{p+1})' = M'\boldsymbol{r}, (t_k^1,\ldots,t_k^{p+1})' = H_k'\boldsymbol{r},\ g_k(\boldsymbol{\rho}) = \partial g(\boldsymbol{\rho})/\partial\boldsymbol{\rho}_k$*, and* $\delta(\cdot)$ *is the Dirac delta function with* $\delta''(\cdot)$ *its second derivative.*[5]

Remark 3.1. Although the weight function $w_{\boldsymbol{\theta}}^{*}(\boldsymbol{r})$ for the general stable ARMA model has no closed form, it does for the Gaussian model whenever $g(\boldsymbol{\rho})$ has an explicit representation. In this case it involves the generalized functions. The use of the generalized functions in econometrics is not novel. For example, Phillips (1991) utilizes them as tools to derive the asymptotic theory for the LAD estimator.

COROLLARY 3.1. *Using the weight functions in Theorem 3.1 along with equation (2.3), we generate the following estimating equations for the Gaussian* ARMA(l,m) *model:*

$$\sigma^2 = \frac{1}{ng(\boldsymbol{\rho})}\sum_{j=1}^{n}\boldsymbol{x}_j' A\boldsymbol{x}_j, \tag{3.10}$$

$$\sum_{j=1}^{n}\boldsymbol{x}_j' B_k\boldsymbol{x}_j = 0,\ \forall k, \tag{3.11}$$

where $\boldsymbol{x}_j = (y_j,\ldots,y_{j+p})'$ *and* $n = T - p$.

Remark 3.2. According to Corollary 3.1, the GLS-CECF estimates are easy to compute, if not trivial. This is in contrast to the MLE of the Gaussian ARMA model and the ECF estimate of the stable ARMA model. Not surprisingly, GLS-CECF with a reasonably large p is not too computationally intensive for the Gaussian ARMA model.

Remark 3.3. From (3.10) and (3.11) we note that for a fixed sample size T, the estimators will change as p changes and consequently the choice of the optimal p is important. For AR(l) processes the optimal p is clearly $p = l$, because we want the conditioning set $(y_j,\ldots,y_{j+p-1})$ to be the same as in the model. Choosing $p > l$ will be inefficient as it conditions on a larger set than necessary and thus ignores information.

Remark 3.4. Because $A = ff'$ and $B_k = \partial A/\partial \boldsymbol{\rho}_k = \partial f/\boldsymbol{\rho}_k f' + f(\partial f/\boldsymbol{\rho}_k)'$ we can rewrite the estimating equation as

$$\frac{1}{n} f'\left(\sum_{j=1}^{n} \boldsymbol{x}_j \boldsymbol{x}_j'\right) f - \sigma^2 g(\boldsymbol{\rho}) = 0, \tag{3.12}$$

$$f'\left(\sum_{j=1}^{n} \boldsymbol{x}_j \boldsymbol{x}_j'\right) \frac{\partial f}{\partial \boldsymbol{\rho}_k} = 0, \forall k. \tag{3.13}$$

Remark 3.5. As a result of the form of Ω for AR(l) models and the fact that $f_p' = \Omega_{21}\Omega_{11}^{-1}$ we have that for these models, setting $p \geq l$ will result in the first $p - l$ elements of f_p' being zero. Thus only by setting $p = l$ will this vector have no zeros, and in this situation the ECF-GLS estimator will exactly equal the CML estimator. Information is clearly lost when p is set greater than l.

Remark 3.6. For general ARMA models, the vector f_p' will have no zero elements irrespective of the size of p. Thus by increasing p for MA and ARMA models we always use more information and hence improve efficiency.

Remark 3.7. From the alternative form of the estimating equations given in Remark 3.4 we note that the distinct elements in the matrix $\sum_{j=1}^{T-p} \boldsymbol{x}_j \boldsymbol{x}_j'$ will form the basis for our estimators. In the AR(l) case with $p = l$, these distinct elements form a set of sufficient statistics. Indeed, they are the same set as appears in the conditional likelihood function. In general ARMA(l, m) processes we have, unlike the AR case, that the likelihood cannot be written as the product of conditional distributions all having a conditioning set of the same length. However, if we do approximate the likelihood by such a product we again find that the associated set of sufficient statistics is, for any fixed p, the distinct elements in the matrix $\sum_{j=1}^{T-p} \boldsymbol{x}_j \boldsymbol{x}_j'$. As we increase p, this set increases. However, because general ARMA processes do not possess a sufficient set of statistics of dimension less than the sample size (see Arato, 1961), using a larger p will always improve efficiency. As noted earlier however, a larger p although increasing asymptotic efficiency will decrease computational efficiency, and hence in practice there is always this trade-off.

4. SIMULATIONS

We now report the results of simulations comparing the estimators of the previous section for both the stable and Gaussian ARMA models. All of the simulations involve one of the following three ARMA models:

$$Y_t = \rho Y_{t-1} + \varepsilon_t, \tag{4.1}$$

$$Y_t = \varepsilon_t - \phi \varepsilon_{t-1}, \tag{4.2}$$

$$Y_t = \rho Y_{t-1} + \varepsilon_t - \phi \varepsilon_{t-1}. \tag{4.3}$$

Among the various ECF methods discussed in Section 2, the DECF is not considered in the simulations for two reasons. First, it is not clear how to make the grid points sufficiently fine and extended, especially in the dependent and multiparameter case. Second, a Monte Carlo study in Yu (1998) shows that for a Gaussian MA(1) model, the CECF method performs better than the DECF method when various restrictions are imposed on the location of the grid points.

4.1. Stable ARMA Models

In this section the Monte Carlo studies are designed to compare WLS-CECF with LS and the Whittle method for the three asymmetric stable ARMA models. A program for quickly and accurately evaluating the symmetric stable density is provided in McCulloch (1998). Nolan (1997) gives a program that can accurately, albeit slowly, evaluate the asymmetric stable density via the integral representation of the Fourier inversion. The program, which has been improved from various aspects including speed, is now available and described in Nolan (1999).[6] It gives a fast, precomputed spline approximation to stable densities for $\alpha \geq 0.4$. Unfortunately, the program has a few limitations from our perspective, and its precision has not been extensively studied. Consequently, it is not used nor the CML estimator examined for the asymmetric stable ARMA models. In all of the following experiments, $\{\varepsilon_t\}$ is an i.i.d. sequence of asymmetric stable random variables with $\sigma = 1, \mu = 0$. The parameters to be estimated are α, β, and we simulate 1,000 observations for each of the 1,000 replications. For WLS-CECF a is set to be 1.[7] The numerical integration is conducted using 39-point Hermitian quadrature, and the numerical optimization is carried out using Powell's conjugate direction algorithm (Powell, 1964).

To simulate the stable ARMA models, an S-PLUS program *rstab* is used for generating stable random variables. It relies on the method proposed by Chambers, Mallow, and Stuck (1976). For LS/LAD to be consistent, the error term must have a zero mean/median. Because there is no analytical expression for the median of the asymmetric stable distribution, in the simulations LAD is not considered although it has a more rapid rate of convergence.

In the first experiment a stable AR(1) model with $\rho = 0.6$, $\alpha = 1.6$, $\beta = -0.5$ is considered. For this AR(1) process, we choose $p = 1$ for WLS-CECF and compare it with LS and the Whittle method in Table 1. In terms of mean square errors (MSE), the relative efficiency of WLS-CECF to LS and the Whittle method is 40%, 69%, respectively, for ρ. The inefficiency of WLS-CECF is due to the nonoptimal weight function adopted.

Table 1. Monte Carlo study comparing ECF, LS, and Whittle of stable AR(1)

	$\rho = 0.6$			$\alpha = 1.6$	$\beta = -0.5$
	ECF	LS	Whittle	ECF	ECF
MEAN	.5953	.6002	.5990	1.600	−.5145
MED	.5959	.5997	.5991	1.600	−.5100
VAR	.00123	.00049	.00085	.00297	.0282
MSE	.00125	.00049	.00085	.00297	.0284

In the second experiment a stable AR(1) model with $\rho = 0.6$, $\alpha = 1.2$, $\beta = 0.8$ is considered. Compared with the first experiment, this experiment has an innovation whose distribution has fatter tails and is more skewed. Consequently, this experiment is designed to examine the relative performances of WLS-CECF when the innovation is further away from normality. The results are presented in Table 2. Noticeably WLS-CECF provides superior estimates of ρ. For example, the relative efficiency of WLS-CECF to LS and the Whittle method is 153%, 135%, respectively, for ρ. Furthermore, the mean of WLS-CECF estimates is closer to the true ρ. Also of note from a comparison of Table 1 with Table 2 is that a further departure from normality for the innovation makes the performances of LS and the Whittle method worse. Surprisingly, however, WLS-CECF works even better for all three parameters when the innovation in the AR(1) model has fatter tails and is more skewed.

In the third experiment a stable MA(1) model with $\phi = 0.6$, $\alpha = 1.6$, $\beta = -0.5$ is considered. We choose $p = 1,2,3$ for WLS-CECF and compare it with LS and the Whittle method in Table 3. The relative efficiency of WLS-CECF to LS and the Whittle method is 25%, 32%, respectively for ρ when $p = 1$. It improves with the value of p. However, the gain from $p = 2$ to $p = 3$

Table 2. Monte Carlo study comparing ECF, LS, and Whittle of stable AR(1)

	$\rho = 0.6$			$\alpha = 1.2$	$\beta = 0.8$
	ECF	LS	Whittle	ECF	ECF
MEAN	.5972	.6160	.6116	1.197	.7861
MED	.6041	.6033	.6028	1.198	.7925
VAR	.00099	.00128	.00121	.00035	.0063
MSE	.00100	.00153	.00135	.00036	.0065

TABLE 3. Monte Carlo Study comparing ECF, LS, and Whittle of stable MA(1)

	$\phi = 0.6$					$\alpha = 1.6$			$\beta = -0.5$		
	ECF $p = 1$	ECF $p = 2$	ECF $p = 3$	LS	Whittle	ECF $p = 1$	ECF $p = 2$	ECF $p = 3$	ECF $p = 1$	ECF $p = 2$	ECF $p = 3$
MEAN	.5983	.5976	.5992	.5978	.5966	1.603	1.605	1.605	−.520	−.522	−.524
MED	.5983	.5974	.6015	.6002	.6000	1.600	1.601	1.610	−.510	−.519	−.521
VAR	.0048	.0040	.0039	.0011	.0015	.0029	.0027	.0027	.023	.020	.020
MSE	.0048	.0040	.0039	.0012	.0015	.0029	.0028	.0028	.024	.021	.021

Table 4. Monte Carlo study comparing ECF, LS, and Whittle of stable ARMA(1,1)

	$\rho = 0.6$					$\alpha = 1.6$		
	ECF $p = 1$	ECF $p = 2$	ECF $p = 3$	LS	Whittle	ECF $p = 1$	ECF $p = 2$	ECF $p = 3$
MEAN	.5948	.5946	.5949	.6006	.5996	1.600	1.603	1.601
MED	.5965	.5941	.5940	.5999	.5997	1.598	1.597	1.599
VAR	.00132	.00132	.00133	.00062	.00096	.0072	.0044	.0043
MSE	.00135	.00135	.00136	.00062	.00096	.0072	.0044	.0043
	$\phi = -0.6$					$\beta = -0.5$		
	ECF $p = 1$	ECF $p = 2$	ECF $p = 3$	LS	Whittle	ECF $p = 1$	ECF $p = 2$	ECF $p = 3$
MEAN	−.5988	−.5967	−.5973	−.5971	−.5946	−.5234	−.5241	−.5229
MED	−.5967	−.5930	−.5959	−.5984	−.5976	−.5064	−.5122	−.5117
VAR	.0113	.0095	.0091	.0015	.0035	.0373	.0313	.0310
MSE	.0113	.0096	.0092	.0015	.0036	.0379	.0319	.0316

is marginal. This suggests that, to achieve much higher efficiency by the ECF method, an alternative weight rather than a large value of p is needed.

In the fourth experiment a stable ARMA(1,1) model with $\rho = 0.6$, $\phi = -0.6$, $\alpha = 1.6, \beta = -0.5$ is considered. We choose $p = 1,2,3$ for WLS-CECF and compare it with LS and the Whittle method in Table 4. The relative efficiency of WLS-CECF to LS and the Whittle method is 46%, 71%, respectively for ρ, and 13%, 32%, respectively, for ϕ when $p = 1$. This suggests that the MA coefficient is less accurately estimated than the AR coefficient for the ECF method. When p increases to 2 the relative efficiency improves for the MA coefficient but not for the AR coefficient. This is not surprising because $p = 1$ is known to be large enough for an AR(1) model. A further improvement on the relative efficiency is also found on the MA coefficient when $p = 3$, but the gain is rather smaller. As in the MA(1) model, an alternative weight rather than a large value of p may be needed to achieve much higher efficiency by the ECF method.

The data generating processes constructed earlier make the ML procedure infeasible because the likelihood function has no closed form. Although several estimation procedures have been used to estimate the stable ARMA models and they are found to be superior to the ECF method when estimating the ARMA coefficients in the MA(1) and ARMA(1,1) models, they cannot estimate the parameters in the innovation. In terms of efficiency of the AR coefficients in the AR(1) model, the ECF estimators can be worse or better than the LS and Whittle estimators depending on the parameter values considered. The

Monte Carlo study also indicates that in finite samples the ECF with a small value of p provides viable estimates of all the parameters, including those in the innovation, at least in the three stable ARMA models considered.

4.2. Gaussian ARMA Models

In this section Monte Carlo studies are designed to compare the ECF with ML for the Gaussian AR(1) and MA(1) models and with ML and CML for the Gaussian ARMA(1,1) model. Although it is known that the MLE can be calculated in these models, it is useful and informative to investigate relative performances of the ECF method. In each of the following experiments, $\{\varepsilon_t\}$ is an i.i.d. sequence of normal variables, and we simulate 100 observations for each of the 1,000 replications. For the WLS-CECF method a is set to be 1.

In the fifth experiment a Gaussian AR(1) model with $\rho = 0.6$, $\sigma^2 = 1.0$ is considered. We choose $p = 1,2$ for both WLS-CECF and GLS-CECF. We compare these two ECF methods with ML in Table 5. Not surprisingly, WLS-CECF is less efficient than GLS-CECF and ML. For example, the relative efficiency of WLS-CECF($p = 1$) to GLS-CECF($p = 1$) and ML is 40% for ρ and 53% for σ^2. Interestingly, GLS-CECF($p = 1$) and ML provide almost identical results in the finite sample. This finding illustrates our theoretical result that $p = 1$ is large enough for an AR(1) process. Although it appears there is almost no efficiency loss we know that with a $p = 2$ we are conditioning on the first two observations. In a sample of size 100, ignoring the first one or two observations will give similar results.

In the sixth experiment a Gaussian MA(1) model with $\phi = 0.6$, $\sigma^2 = 1.0$ is considered. We use the GLS-CECF method with various p values and also the ML method to estimate the model. The existence of a closed form expression for the weight function $w^*_{\boldsymbol{\theta}}(\boldsymbol{r})$ and estimating equations allows us to choose large values of p without involving too much extra computation. The results are presented in Table 6. As we expect, as p gets larger and larger, efficiency of the GLS-CECF estimator gets closer and closer to that of the MLE. Note the

TABLE 5. Monte Carlo Study comparing ECF and ML of Gaussian AR(1)

	$\rho = 0.6$					$\sigma^2 = 1.0$				
	WLS $p = 1$	GLS $p = 1$	WLS $p = 2$	GLS $p = 2$	ML	WLS $p = 1$	GLS $p = 1$	WLS $p = 2$	GLS $p = 2$	ML
MEAN	.5835	.5907	.5832	.5905	.5906	.9965	.9924	.9961	.9921	.9923
MED	.6004	.5957	.6001	.6007	.5982	.9916	.9850	.9912	.9848	.9838
VAR	.014	.006	.015	.006	.006	.038	.020	.039	.020	.020
MSE	.015	.006	.016	.006	.006	.038	.020	.039	.020	.020

TABLE 6. Monte Carlo study comparing GLS-CECF and ML of Gaussian MA(1)

	$\phi = 0.6$						$\sigma^2 = 1.0$					
	ECF $p = 2$	ECF $p = 3$	ECF $p = 4$	ECF $p = 5$	ECF $p = 6$	ML	ECF $p = 2$	ECF $p = 3$	ECF $p = 4$	ECF $p = 5$	ECF $p = 6$	ML
MEAN	.6269	.6133	.6061	.6037	.6024	.6032	.967	.981	.988	.991	.991	.992
MED	.5960	.6011	.5992	.6004	.6009	.6026	.957	.974	.981	.984	.987	.984
VAR	.026	.015	.010	.009	.007	.007	.026	.022	.021	.021	.021	.020
MSE	.027	.015	.010	.009	.007	.007	.027	.022	.021	.021	.021	.020

TABLE 7. Monte Carlo study comparing ECF, CML, and ML of Gaussian ARMA(1,1)

	$\rho = 0.6$					$\sigma^2 = 1.0$				
	ECF $p = 2$	ECF $p = 3$	CML1	CML2	ML	ECF $p = 2$	ECF $p = 3$	CML1	CML2	ML
MEAN	.5972	.6000	.5895	.6002	.6002	.9899	1.003	1.018	.9980	.992
MED	.5973	.6000	.5909	.6011	.6004	.9813	.9911	1.01	.9898	.9828
VAR	.0038	.0031	.0030	.0023	.0023	.023	.023	.023	.021	.020
MSE	.0038	.0031	.0032	.0023	.0023	.023	.023	.023	.021	.020

fact that the GLS-CECF method with a small p can work quite well in finite samples. For example, there is a small efficiency gain for the GLS-CECF method with $p \geq 4$. This is not surprising because the MA coefficient is reasonably small and hence an AR(4) model will provide a good approximation to the MA(1) model.

In the seventh experiment a Gaussian ARMA(1,1) model with $-\phi = \rho = 0.6$, $\sigma^2 = 1.0$ is considered. Because T is not terribly large, the exact MLE is obtained via maximization of (3.4). However, the inverse of Ω for the ARMA(1,1) model has no closed form even for $-\phi = \rho$, and we numerically invert Ω. The two CML methods discussed previously are also used to estimate the model. Table 7 tabulates the results where we choose $p = 2,3$ for GLS-CECF. The relative efficiency of GLS-CECF with $p = 2$ to the two CML methods and the ML method is 84%, 61%, 61%, respectively, for ρ, and 100%, 91%, 87%, respectively for σ^2. When $p = 3$ the relative efficiency becomes 103%, 74%, 74%, respectively, for ρ, and 100%, 91%, 87%, respectively for σ^2.

In the eighth experiment a Gaussian ARMA(1,1) model with $-\phi = \rho = 0.9$, $\sigma^2 = 1.0$ is considered. The simulated sequence is more persistent compared with the sequence in the last experiment. Not surprisingly the difference between the exact likelihood and conditional likelihood is larger, and hence estimation via CML will be less accurate in finite samples. Table 8 presents the results. Noticeably, there is a trade-off between GLS-CECF with $p = 3$ and CML2. To be more specific, the mean of GLS-CECF estimates with $p = 3$ is closer to the true parameter value for both ρ and σ^2, whereas CML2 is relatively more efficient. The relative efficiency of GLS-CECF with $p = 2,3$ to CML2 is 65%, 71% for ρ and 93%, 96% for σ^2. Furthermore, compared with Table 7, the relative efficiency of the ECF over CML2 gets larger. Another point that emerges is that GLS-CECF is superior to CML1. For example, GLS-CECF with $p = 2,3$ has smaller bias, and the relative efficiency to CML1 is 215%, 233% for ρ and 2186%, 2267% for σ^2.

The findings can be explained as follows. On the one hand, some initial conditions must be assumed to obtain the CMLE1. If the initial assumption is chosen inappropriately, it is carried over into all the following stages by the recursive

TABLE 8. Monte Carlo study comparing ECF, CML, and ML of Gaussian ARMA(1,1)

	$\rho = 0.9$					$\sigma^2 = 1.0$				
	ECF $p = 2$	ECF $p = 3$	CML1	CML2	ML	ECF $p = 2$	ECF $p = 3$	CML1	CML2	ML
MEAN	.8846	.8881	.8512	.8861	.8969	1.003	1.002	1.483	1.030	.9928
MED	.8904	.8949	.8601	.8890	.8999	.9911	.9884	1.297	1.025	.9836
VAR	.0023	.0023	.0032	.0015	.0010	.028	.027	.378	.025	.020
MSE	.0026	.0024	.0056	.0017	.0010	.028	.027	.612	.026	.020

formula such as (3.6). In large samples the effect of such an error will diminish for the stationary models, and thus the CMLE1 is asymptotically equivalent to the MLE. However, the effect may not be negligible in a small sample.

The data generating processes constructed previously favor the exact ML procedure because the likelihood function has a closed form expression. It is not surprising that ML performs well. Although we do not expect the ECF method to outperform ML, the limited Monte Carlo study shows that the ECF method can work reasonably well. The good performance of the ECF method has been confirmed at least for the three ARMA models considered and can be achieved with small values of p.[8]

The advantage of using the GLS-CECF method over the exact ML method is that the ECF method does not invert the covariance matrix and hence is numerically less intensive. Using FORTRAN code on an alpha-digital Unix system, e.g., it takes about 1 minute to do GLS-CECF with $p = 3$ in the last experiment, whereas it takes about 4 minutes to do ML with σ^2 concentrated out. One expects that such a numerical advantage will increase with the sample size.

It can be seen that both consistency and efficiency of the ECF estimator depend on p, the overlapping size of the moving blocks. Once p is chosen to ensure identification, the ECF estimator is consistent. The impact of the choice of p on efficiency is more complicated. For pure AR(l) processes, the optimal p is l. For a general ARMA process, unfortunately, there is no optimal p. In this case, the choice of p depends on how well the ARMA process can be approximated by an AR(p) process.

5. CONCLUSION

This paper proposes a new econometric methodology for the estimation of stationary processes via the ECF. Under regularity conditions, the ECF estimator is shown to be consistent and asymptotically normally distributed. Monte Carlo simulations are used to examine the relative performances of the ECF method. For the stable ARMA model for which the ML method is not applicable, we find that in finite samples the ECF is a viable method for the three models considered. For the Gaussian ARMA model for which the ML method is readily available, we derive the optimal weight functions and estimating equations for the ECF method. We find that in finite samples the ECF method can have reasonably good efficiency in comparison with the exact and conditional ML methods for the three models considered.

NOTE

1. The references for parameter estimation include Feuerverger (1990), Heathcote (1977), Knight and Satchell (1996, 1997), Schmidt (1982), and references therein.
2. Knight and Satchell (1997) develop the cumulant generating function estimation method although they do detail, also, the DECF approach to dependent data.
3. See Remark 2.2 concerning the asymptotic distribution associated with 2.3 and 2.4.
4. We thank one of the referees for providing this example.
5. If $\int_{-\infty}^{+\infty} f(x)\delta(x)\,dx = f(0)$ for any integrable function $f(x)$, $\delta(\cdot)$ is called the Dirac delta function; if $\int_{-\infty}^{+\infty} f(x)\delta^{(n)}(x)\,dx = (-1)^n f^{(n)}(0)$, $\delta^{(n)}(\cdot)$ is the nth derivative of Dirac delta function. See Gel'Fan (1964) for more discussion about $\delta(\cdot)$.
6. We thank one of the referees for bringing to our attention the papers by Nolan (1997, 1999).
7. Although a is set arbitrarily to be 1, one can minimize the variance of estimates with respect to a to obtain more efficient estimates.
8. A small block size is also found to be adequate in the moving block bootstrap method for the AR(1) model and MA(1) model by Künsch (1989).

REFERENCES

Adler, R.J., Feldman, R.E., & M.S. Taqqu (1998) *A Practical Guide to Heavy Tails*. Boston: Birkhäuser.

Amemiya, T. (1985). *Advanced Econometrics*. Oxford: Blackwell.

Andrews, Donald W.K. (1991) Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* 59, 817–858.

Arato, M. (1961) On the sufficient statistics for Gaussian random processes. *Theory of Probability and Its Application* 4, 199–201.

Box, G.E. & G.M. Jenkins. (1976) *Time Series Analysis*. San Francisco: Holden-Day.

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

Calder, M. & R.A. Davis. (1998) Inferences for linear processes with stable noise. In R.J. Adler, R.E. Feldman, & M.S. Taqqu (eds.), *A Practical Guide to Heavy Tails*, pp. 159–176. Boston: Birkhäuser.

Chambers, J.M., C. Mallow, & B.W. Stuck. (1976) A method for simulating stable random variables. *Journal of the American Statistical Association* 71, 340–344.

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

Davis, R.A. (1996) Gauss-Newton and M-estimation for ARMA processes with infinite variance. *Stochastic Processes and Their Applications* 63, 75–95.

Davis, R.A., K. Knight, & J. Liu (1992) M-estimation for autoregressions with infinite variance. *Stochastic Processes and Their Applications* 40, 145–180.

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

Embrechts, P., C. Klüppelberg, & T. Mikosch (1997) *Modelling Extremal Events*. New York: Springer-Verlag.

Epps, T.W. (1987) Testing that a stationary time series is Gaussian. *Annals of Statistics* 15, 1683–1698.

Epps, T.W. (1988) Testing that a Gaussian process is stationary. *Annals of Statistics* 16, 1667–1683.

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

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

Gel'Fan, A. (1964) *Generalized Functions*. Vol. 1. New York: Academic Press.

Ghysels, E., A.C. Harvey, & E. Renault (1996) Stochastic volatility. In G.S. Maddala & C.R. Rao (eds.), *Handbook of Statistics*, vol 14: *Statistical Methods in Finance*. Amsterdam: North-Holland.

Gordin, M.I. (1969) The central limit theorem for stationary processes. *Soviet Mathematics-Daklady* 10, 1174–1176.

Hall, P. & C.C. Heyde (1980) *Martingale Limit Theory and Its Application*. New York: Academic Press.

Hamilton, J.D. (1994) *Time Series Analysis*. Princeton, NJ: Princeton University Press.

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

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

Hong, Y. (1999) Hypothesis testing in time series via the empirical characteristic function: A generalized spectral density approach. *Journal of the American Statistical Association* 94, 1201–1220.

Huber, P. (1981) *Robust Statistics*. New York: Wiley.

Klüppelberg, C. & T. Mikosch (1993) Spectral estimates and stable processes. *Stochastic Processes and Their Applications* 47, 323–344.

Klüppelberg, C. & T. Mikosch (1994) Some limit theory for the self normalised periodogram of *p*-stable processes. *Scandinavian Journal of Statistics* 21, 485–491.

Knight, J.L. & S.E. Satchell (1996) Estimation of Stationary Stochastic Processes via the Empirical Characteristic Function. Manuscript, The University of Western Ontario.

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

Knight, J.L. & S.E. Satchell (1998) GARCH processes: Some exact results, some difficulties, and a suggested remedy. In J.L. Knight & S.E. Satchell (eds.), Forecasting Volatility in the Financial Markets. Oxford: Butterworth-Heineman.

Künsch, H. (1989) The jackknife and the bootstrap for general stationary observations. *Annals of Statistics* 17, 1217–1241.

Lukacs, E. (1970) *Characteristic Functions*. New York: Griffin.

McCulloch, J.H. (1998) Numerical approximation of the symmetric stable distribution and density. In R.J. Adler, R.E. Feldman, & M.S. Taqqu (eds.), *A Practical Guide to Heavy Tails*, pp. 489–500. Boston: Birkhäuser.

Mikosch, T., C. Klüppelberg, & R.J. Adler (1995) Parameter estimation for ARMA models with infinite variance innovations. *Annals of Statistics* 23, 305–326.

Newey, W.K. & D. McFadden (1994) Large sample estimation and hypothesis testing. In R.F. Engle and D. McFadden (eds.), *Handbook of Econometrics*, vol. 4. Amsterdam: North-Holland.

Newey, W.K. & K. West (1994) Automatic lag selection in covariance matrix estimation. *Review of Economic Studies* 61, 631–653.

Nolan, J. (1997) Numerical calculation of stable densities and distribution functions. *Communication Statistics: Stochastic Models* 13, 759–774.

Nolan, J. (1999) Maximum Likelihood Estimation and Diagnostics for Stable Distributions. Manuscript, American University.

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

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

Phillips, P.C.B. (1991) A shortcut to LAD estimator asymptotics. *Econometric Theory* 7, 450–463.

Powell, M.J.D. (1964) An efficient method for finding the minimum of a function of several variables without calculating derivatives. *Computer Journal* 7, 155–162.
Press, W.H., S.A. Teukolsky, W.T. Vetterling, & B.P. Flannery (1992) *Numerical Recipes in Fortran*. Cambridge: Cambridge University Press.
Samorodnitsky, G. & M.S. Taqqu (1994) *Stable Non-Gaussian Random Processes*. New York: Chapman and Hall.
Schmidt, P. (1982) An improved version of the Q-R MGF estimator for mixtures of normal distributions and switching regressions. *Econometrica* 50, 501–524.
Singleton, K.J. (2001) Estimation of affine asset pricing models using the empirical characteristic function. *Journal of Econometrics* 102, 111–141.
Sloan, I.H. & S. Joe (1994) *Lattice Methods for Multiple Integration*. Oxford: Oxford University Press.
Stout, W.F. (1974) *Almost Sure Convergence*. New York: Academic Press.
Titterington, D.M., A.F.M. Smith, & U.E. Markov (1985) *Statistical Analysis of Finite Mixture Distributions*. Toronto: Chichester.
White, H. (1982) Maximum likelihood estimation of misspecified models. *Econometrica* 50, 1–26.
Whittle, P. (1983) *Prediction and Regulation by Linear-Square Methods*. Minneapolis: University of Minnesota Press.
Yu, Jun (1998) *The Empirical Characteristic Function in Time-Series Estimation*. Ph.D. thesis, University of Western Ontario.

APPENDIX

Proof of Theorem 2.1. Obviously, the ECF estimator is in the class of extremum estimators. Hence, one can prove consistency and asymptotic normality of the ECF estimator by checking a set of sufficient conditions of an extremum estimator. In this paper, to prove consistency we will check the conditions listed by Newey and McFadden (1994, p. 2121) or Amemiya (1985, pp. 106–107), namely, compactness, continuity, uniform convergence, and identifiability.

Compactness of $\boldsymbol{\Theta}$ and continuity of $I_n(\boldsymbol{\theta})$ are ensured by (A1) and (A2), respectively. To check uniform convergence of $I_n(\boldsymbol{\theta})$, note that (A3) leads to, according to Theorems 3.5.3 and 3.5.8 of Stout (1974), stationarity and ergodicity of $\{\exp(i\boldsymbol{r}'\boldsymbol{x}_j)\}$. By a strong law of large numbers, $c_n(\boldsymbol{r}) \xrightarrow{a.s.} c(\boldsymbol{r};\boldsymbol{\theta}_0)$. Hence, $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}|c_n(\boldsymbol{r}) - c(\boldsymbol{r};\boldsymbol{\theta}_0)| \xrightarrow{a.s.} 0$. Note that $c_n(\boldsymbol{r})$, $c(\boldsymbol{r};\boldsymbol{\theta})$ and their conjugates are all trigonometric functions; hence

$$
\begin{aligned}
|I_n(\boldsymbol{\theta}) - I_0(\boldsymbol{\theta})| &= \left| \int \dots \int \{|c_n(\boldsymbol{r}) - c(\boldsymbol{r};\boldsymbol{\theta})|^2 - |c(\boldsymbol{r};\boldsymbol{\theta}_0) - c(\boldsymbol{r};\boldsymbol{\theta})|^2\} dG(\boldsymbol{r}) \right| \\
&= \left| \int \dots \int \{(c_n(\boldsymbol{r}) - c(\boldsymbol{r};\boldsymbol{\theta}_0))(\bar{c}_n(\boldsymbol{r}) + \bar{c}(\boldsymbol{r};\boldsymbol{\theta}_0) - 2\bar{c}(\boldsymbol{r};\boldsymbol{\theta}))\} dG(\boldsymbol{r}) \right| \\
&\le 4 \int \dots \int |c_n(\boldsymbol{r}) - c(\boldsymbol{r};\boldsymbol{\theta}_0)| \, dG(\boldsymbol{r})
\end{aligned}
$$

so that $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}|I_n(\boldsymbol{\theta}) - I_0(\boldsymbol{\theta})| \xrightarrow{a.s.} 0$ and hence uniform convergence holds for $I_n(\boldsymbol{\theta})$. Assumption (A4) ensures that $I_0(\boldsymbol{\theta})$ (≥ 0) attains the unique minimum at $\boldsymbol{\theta}_0$ and hence the identification condition holds. Consequently, strong consistency of the ECF estimator applies.

To prove asymptotic normality, we will also check the conditions listed by Newey and McFadden (1994, p. 2143) or Amemiya (1985, p. 111). Specifically, two conditions will be verified, namely, (a) $\sqrt{n}\,\partial I_n(\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta}$ converges to a normal variate in distribution; (b) $\forall\tilde{\boldsymbol{\theta}}_n \xrightarrow{a.s.} \boldsymbol{\theta}_0$, $\partial^2 I_n(\tilde{\boldsymbol{\theta}}_n)/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}' \xrightarrow{a.s.} H(\boldsymbol{\theta}_0)$, and $H(\boldsymbol{\theta}_0)$ is nonsingular.

Consider the first-order condition of problem (2.1) and by (A2), we have

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

Therefore $\partial I_n(\boldsymbol{\theta})/\partial\boldsymbol{\theta}$ is the sample mean of a random sequence $\{K(\boldsymbol{x}_j;\boldsymbol{\theta})\}$ multiplied by a constant, -2. Note that $\{K(\boldsymbol{x}_j;\boldsymbol{\theta})\}$ is identical but not independently distributed because $\{\boldsymbol{x}_j\}$ is dependent. Also note that $\{K(\boldsymbol{x}_j;\boldsymbol{\theta})\}$ preserves stationarity and ergodicity of $\{y_j\}$ by (A5), according to Theorems 3.5.3 and 3.5.8 of Stout (1974). Among other things, (A7) enables us to apply a central limit theorem for stationary, ergodic processes provided by Gordin (1969) and Hall and Heyde (1980, p. 129). Formally, we have

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

where

$$\begin{aligned}A(\boldsymbol{\theta}) &= \lim_{n\to\infty}\frac{1}{n}E\left(\sum_{j=1}^{n}\sum_{k=1}^{n}K(\boldsymbol{x}_j;\boldsymbol{\theta})K(\boldsymbol{x}_k;\boldsymbol{\theta})\right) \\ &= \lim_{n\to\infty}\int\cdots\int\Big\{\frac{\partial \operatorname{Re} c(\boldsymbol{r};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\frac{\partial \operatorname{Re} c(\boldsymbol{s};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'}\frac{1}{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\operatorname{cov}(\cos(\boldsymbol{r}'\boldsymbol{x}_j),\cos(\boldsymbol{s}'\boldsymbol{x}_k)) \\ &\qquad + \frac{\partial \operatorname{Re} c(\boldsymbol{r};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\frac{\partial \operatorname{Im} c(\boldsymbol{s};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'}\frac{2}{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\operatorname{cov}(\cos(\boldsymbol{r}'\boldsymbol{x}_j),\sin(\boldsymbol{s}'\boldsymbol{x}_k)) \\ &\qquad + \frac{\partial \operatorname{Im} c(\boldsymbol{r};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\frac{\partial \operatorname{Im} c(\boldsymbol{s};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'}\frac{1}{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\operatorname{cov}(\sin(\boldsymbol{r}'\boldsymbol{x}_j),\sin(\boldsymbol{s}'\boldsymbol{x}_k))\Big\} \\ &\qquad \times dG(\boldsymbol{r})dG(\boldsymbol{s}). \end{aligned} \tag{A.3}$$

Defining $\Psi_k(\boldsymbol{r},\boldsymbol{s}) = E[\exp(i\boldsymbol{r}'\boldsymbol{x}_1 + i\boldsymbol{s}'\boldsymbol{x}_{k+1})]$, we can then rewrite the double summation covariances as

$$\frac{1}{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\operatorname{cov}(\cos(\boldsymbol{r}'\boldsymbol{x}_j),\cos(\boldsymbol{s}'\boldsymbol{x}_k))$$

$$= \frac{1}{2}(\operatorname{Re} c(\boldsymbol{r}+\boldsymbol{s}) + \operatorname{Re} c(\boldsymbol{r}-\boldsymbol{s})) - \operatorname{Re} c(\boldsymbol{r})\operatorname{Re} c(\boldsymbol{s})$$

$$+ \frac{1}{2n}\sum_{k=1}^{n-1}(n-k)(\operatorname{Re}\Psi_k(\boldsymbol{r},\boldsymbol{s}) + \operatorname{Re}\Psi_k(\boldsymbol{r},-\boldsymbol{s}) + \operatorname{Re}\Psi_k(\boldsymbol{s},\boldsymbol{r}) + \operatorname{Re}\Psi_k(\boldsymbol{s},-\boldsymbol{r})),$$

$$\frac{1}{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\operatorname{cov}(\cos(\boldsymbol{r}'\boldsymbol{x}_j),\sin(\boldsymbol{s}'\boldsymbol{x}_k))$$

$$= \frac{1}{2}(\operatorname{Im} c(\boldsymbol{r}-\boldsymbol{s}) + \operatorname{Im} c(\boldsymbol{r}-\boldsymbol{s})) - \operatorname{Re} c(\boldsymbol{r})\operatorname{Re} c(\boldsymbol{s})$$

$$+ \frac{1}{2n}\sum_{k=1}^{n-1}(n-k)(\operatorname{Im}\Psi_k(\boldsymbol{r},\boldsymbol{s}) - \operatorname{Im}\Psi_k(\boldsymbol{r},-\boldsymbol{s}) + \operatorname{Im}\Psi_k(\boldsymbol{s},\boldsymbol{r}) + \operatorname{Im}\Psi_k(\boldsymbol{s},-\boldsymbol{r})),$$

$$\frac{1}{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\operatorname{cov}(\sin(\boldsymbol{r}'\boldsymbol{x}_j),\sin(\boldsymbol{s}'\boldsymbol{x}_k))$$

$$= \frac{1}{2}(\operatorname{Re} c(\boldsymbol{r}+\boldsymbol{s}) + \operatorname{Re} c(\boldsymbol{r}-\boldsymbol{s})) - \operatorname{Im} c(\boldsymbol{r})\operatorname{Im} c(\boldsymbol{s})$$

$$+ \frac{1}{2n}\sum_{k=1}^{n-1}(n-k)(\operatorname{Re}\Psi_k(\boldsymbol{r},-\boldsymbol{s}) - \operatorname{Re}\Psi_k(\boldsymbol{r},\boldsymbol{s}) + \operatorname{Re}\Psi_k(\boldsymbol{s},-\boldsymbol{r}) - \operatorname{Re}\Psi_k(\boldsymbol{s},\boldsymbol{r})),$$

where $c(\boldsymbol{r}) = c(\boldsymbol{r};\boldsymbol{\theta})$. Thus, condition (a) is confirmed.

In view of (A2), $\forall\boldsymbol{\theta}\in N(\boldsymbol{\theta}_0)$,

$$\frac{\partial^2 I_n(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} = -2\int\ldots\int\left\{\frac{\partial\operatorname{Re} c(\boldsymbol{r};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\frac{\partial\operatorname{Re} c(\boldsymbol{r};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'} + \frac{\partial\operatorname{Im} c(\boldsymbol{r};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\frac{\partial\operatorname{Im} c(\boldsymbol{r};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'}\right.$$

$$- [\operatorname{Re} c_n(\boldsymbol{r}) - \operatorname{Re} c(\boldsymbol{r};\boldsymbol{\theta})]\frac{\partial^2\operatorname{Re} c(\boldsymbol{r};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}$$

$$\left. - [\operatorname{Im} c_n(\boldsymbol{r}) - \operatorname{Im} c(\boldsymbol{r};\boldsymbol{\theta})]\frac{\partial^2\operatorname{Im} c(\boldsymbol{r};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right\} dG(\boldsymbol{r})$$

$$= -2\int\ldots\int\frac{\partial c(\boldsymbol{r};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\frac{\partial c(\boldsymbol{r};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'}\,dG(\boldsymbol{r})$$

$$+ 2\int\ldots\int\left\{[\operatorname{Re} c_n(\boldsymbol{r}) - \operatorname{Re} c(\boldsymbol{r};\boldsymbol{\theta})]\frac{\partial^2\operatorname{Re} c(\boldsymbol{r};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right.$$

$$\left. + [\operatorname{Im} c_n(\boldsymbol{r}) - \operatorname{Im} c(\boldsymbol{r};\boldsymbol{\theta})]\frac{\partial^2\operatorname{Im} c(\boldsymbol{r};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right\} dG(\boldsymbol{r}). \quad \textbf{(A.4)}$$

Because $\tilde{\boldsymbol{\theta}}_n \xrightarrow{a.s.} \boldsymbol{\theta}_0$, for large enough n, we have

$$\frac{\partial^2 I_n(\tilde{\boldsymbol{\theta}}_n)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} = -2\int\cdots\int \frac{\partial c(\boldsymbol{r};\tilde{\boldsymbol{\theta}}_n)}{\partial\boldsymbol{\theta}}\frac{\partial c(\boldsymbol{r};\tilde{\boldsymbol{\theta}}_n)}{\partial\boldsymbol{\theta}'}\,dG(\boldsymbol{r}) + 2\int\cdots\int\left\{[\operatorname{Re} c_n(\boldsymbol{r}) - \operatorname{Re} c(\boldsymbol{r};\tilde{\boldsymbol{\theta}}_n)]\frac{\partial^2 \operatorname{Re} c(\boldsymbol{r};\tilde{\boldsymbol{\theta}}_n)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} + [\operatorname{Im} c_n(\boldsymbol{r}) - \operatorname{Im} c(\boldsymbol{r};\tilde{\boldsymbol{\theta}}_n)]\frac{\partial^2 \operatorname{Im} c(\boldsymbol{r};\tilde{\boldsymbol{\theta}}_n)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right\} dG(\boldsymbol{r}). \quad \textbf{(A.5)}$$

By (A2) and (A6) the first integral in equation (A.5) converges a.s. to $B(\boldsymbol{\theta}_0)$. We have shown earlier that $c_n(\boldsymbol{r}) - c(\boldsymbol{r};\boldsymbol{\theta}_0) \xrightarrow{a.s.} 0$. This implies $\operatorname{Re} c_n(\boldsymbol{r}) - \operatorname{Re} c(\boldsymbol{r};\boldsymbol{\theta}_0) \xrightarrow{a.s.} 0$ and $\operatorname{Im} c_n(\boldsymbol{r}) - \operatorname{Im} c(\boldsymbol{r};\boldsymbol{\theta}_0) \xrightarrow{a.s.} 0$. Furthermore, $\operatorname{Re} c(\boldsymbol{r};\tilde{\boldsymbol{\theta}}_n) - \operatorname{Re} c(\boldsymbol{r};\boldsymbol{\theta}_0) \xrightarrow{a.s.} 0$ and $\operatorname{Im} c(\boldsymbol{r};\tilde{\boldsymbol{\theta}}_n) - \operatorname{Im} c(\boldsymbol{r};\boldsymbol{\theta}_0) \xrightarrow{a.s.} 0$ follow from $\tilde{\boldsymbol{\theta}}_n \xrightarrow{a.s.} \boldsymbol{\theta}_0$. Hence $\operatorname{Re} c_n(\boldsymbol{r}) - \operatorname{Re} c(\boldsymbol{r};\tilde{\boldsymbol{\theta}}) \xrightarrow{a.s.} 0$ and $\operatorname{Im} c_n(\boldsymbol{r}) - \operatorname{Im} c(\boldsymbol{r};\tilde{\boldsymbol{\theta}}) \xrightarrow{a.s.} 0$. Together with (A6) and boundedness of both the empirical and theoretical characteristic functions, we then have $\partial^2 I_n(\tilde{\boldsymbol{\theta}}_n)/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}' \xrightarrow{a.s.} -2B(\boldsymbol{\theta}_0)$. By (A6), therefore, condition (b) is verified. Conditions (a) and (b) together imply

$$n^{1/2}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow{d} N(0, B^{-1}(\boldsymbol{\theta}_0)A(\boldsymbol{\theta}_0)B^{-1}(\boldsymbol{\theta}_0)). \quad \textbf{(A.6)}$$

■

Proof of Proposition 2.1. Let $z_j = (y_j, \ldots, y_{j+l-1})$ and note that

$$\begin{aligned}
&\int\cdots\int w_{\boldsymbol{\theta}}^*(\boldsymbol{r})c(\boldsymbol{r};\boldsymbol{\theta})\,d\boldsymbol{r}\\
&\quad= \int\cdots\int \frac{1}{(2\pi)^l}\left\{\int\cdots\int \frac{\partial \log f(y_{j+l}|z_j)}{\partial\boldsymbol{\theta}}\exp(-i\boldsymbol{r}'\boldsymbol{x}_j)c(\boldsymbol{r};\boldsymbol{\theta})\,d\boldsymbol{x}_j\right\}d\boldsymbol{r}\\
&\quad= \int\cdots\int \frac{\partial \log f(y_{j+1}|z_j)}{\partial\boldsymbol{\theta}}\left\{\frac{1}{(2\pi)^l}\int\cdots\int \exp(-i\boldsymbol{r}'\boldsymbol{x}_j)c(\boldsymbol{r};\boldsymbol{\theta})\,d\boldsymbol{r}\right\}d\boldsymbol{x}_j\\
&\quad= \int\cdots\int \frac{\partial \log f(y_{j+1}|z_j)}{\partial\boldsymbol{\theta}}\operatorname{pdf}(\boldsymbol{x}_j)\,d\boldsymbol{x}_j\\
&\quad= \int\cdots\int\left\{\int \frac{\partial \log f(y_{j+1}|z_j)}{\partial\boldsymbol{\theta}}\operatorname{pdf}(y_{j+1}|z_j)\,dy_{j+1}\right\}\operatorname{pdf}(z_j)\,dz_j = 0. \quad \textbf{(A.7)}
\end{aligned}$$

Also note that with $p = l$ equation (2.4) implies (via the inverse Fourier transform)

$$\frac{\partial \log f(y_{j+1}|y_j,\ldots,y_{j+l-1})}{\partial\boldsymbol{\theta}} = \int\cdots\int w_{\boldsymbol{\theta}}^*(\boldsymbol{r})\exp(i\boldsymbol{r}'\boldsymbol{x}_j)\,d\boldsymbol{r}. \quad \textbf{(A.8)}$$

Using (A.7) and (A.8) we can rewrite the left-hand side of equation (2.3) by

$$\int\cdots\int w_{\boldsymbol{\theta}}^*(\boldsymbol{r})c_n(\boldsymbol{r})\,d\boldsymbol{r} = \int\cdots\int w_{\boldsymbol{\theta}}^*(\boldsymbol{r})\frac{1}{n}\sum_{j=1}^{n} e^{i\boldsymbol{r}'\boldsymbol{x}_j}\,d\boldsymbol{r} = \frac{1}{n}\sum_{j=1}^{n}\frac{\partial \log f(y_{j+1}|z_j)}{\partial\boldsymbol{\theta}}.$$

This equation is just the first derivative of CML for an AR(l) process. Therefore, the ECF estimator has the same asymptotic efficiency as the CMLE and hence MLE. ■

Proof of Proposition 3.1. Because the ARMA model given by (4.3) is invertible, we can rewrite it by $y_t = \varepsilon_t + (\rho - \phi)\varepsilon_{t-1} + \rho(\rho - \phi)\varepsilon_{t-2} + \rho^2(\rho - \phi)\varepsilon_{t-1} + \cdots$. As a consequence,

$$
\begin{aligned}
&c(r_1, c_2; \boldsymbol{\theta}) \\
&\quad = \exp(ir_1 y_{t-1} + ir_2 y_t) \\
&\quad = \exp\{ir_2\varepsilon_t + i(r_1 + r_2(\rho - \phi))\varepsilon_{t-1} + i(\rho - \phi)(r_1 + r_2\rho)\varepsilon_{t-2} \\
&\qquad + i\rho(\rho - \phi)(r_1 + r_2\rho)\varepsilon_{t-3} + i\rho^2(\rho - \phi)(r_1 + r_2\rho)\varepsilon_{t-4} + \cdots\} \\
&\quad = \exp\left\{-|r_2|^\alpha\left(1 - i\beta\,\mathrm{sign}(r_2)\tan\frac{\pi\alpha}{2}\right)\right\} \\
&\qquad \times \exp\left\{-|r_1 + r_2(\rho - \phi)|^\alpha\left(1 - i\beta\,\mathrm{sign}(r_1 + r_2(\rho - \phi))\tan\frac{\pi\alpha}{2}\right)\right\} \\
&\qquad \times \exp\left\{-|r_1 + r_2\rho|^\alpha|\rho - \theta|^\alpha\left(1 - i\beta\,\mathrm{sign}(r_1 + r_2\rho)\mathrm{sign}(\rho - \theta)\tan\frac{\pi\alpha}{2}\right)\right\} \\
&\qquad \times \exp\left\{-|r_1 + r_2\rho|^\alpha|\rho|^\alpha|\rho - \theta|^\alpha\left(1 - i\beta\,\mathrm{sign}(\rho(r_1 + r_2\rho)(\rho - \theta))\tan\frac{\pi\alpha}{2}\right)\right\} \\
&\qquad \times \ldots \\
&\quad = \exp\left\{-|r_2|^\alpha - |r_1 + r_2(\rho - \phi)|^\alpha - \frac{|r_1 + r_2\rho|^\alpha|\rho - \theta|^\alpha}{1 - |\rho|^\alpha}\right\} \\
&\qquad \times \exp\left\{i\beta\tan\frac{\pi\alpha}{2}\left[|r_1|^\alpha\,\mathrm{sign}(r_2) + |r_1 + r_2(\rho - \phi)|^\alpha\,\mathrm{sign}(r_1 + r_2(\rho - \phi))\right.\right. \\
&\qquad\qquad \left.\left. + \frac{|r_1 + r_2\rho|^\alpha\,\mathrm{sign}(r_1 + r_2\rho)|\rho - \theta|^\alpha\,\mathrm{sign}(\rho - \theta)}{1 - \mathrm{sign}(\rho)|\rho|^\alpha}\right]\right\}. \qquad ■
\end{aligned}
$$

Proof of Theorem 3.1. Because $\log f(y_{j+p}|y_j, \ldots, y_{j+p-1}) = -\frac{1}{2}\log 2\pi - \frac{1}{2}\log\sigma^2 - \frac{1}{2}\log g(\boldsymbol{\rho}) - \frac{1}{2}\sigma^2 g(\boldsymbol{\rho})\boldsymbol{x}_j' A\boldsymbol{x}_j$, differentiating this function with respect to the parameters, we have

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

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

Substituting into (2.4)

$$w^*_{\sigma^2}(\boldsymbol{r}) = \left(\frac{1}{2\pi}\right)^{p+1}\int\dots\int \exp(-i\boldsymbol{r}'\boldsymbol{x}_j)\left[-\frac{1}{2\sigma^2}+\frac{1}{2\sigma^4 g(\boldsymbol{\rho})}\boldsymbol{x}_j' A\boldsymbol{x}_j\right]d\boldsymbol{x}_j$$

$$= -\frac{1}{2\sigma^2}\,\delta(r^1)\dots\delta(r^{p+1}) + \frac{1}{2\sigma^4 g(\boldsymbol{\rho})}\left(\frac{1}{2\pi}\right)^{p+1}$$

$$\times \int\dots\int \exp(-i\boldsymbol{r}'\boldsymbol{x}_j)\boldsymbol{x}_j' A\boldsymbol{x}_j\, d\boldsymbol{x}_j.$$

Now $A = MD_aM'$ thus transforming $\boldsymbol{x}_j \to z_j = M'\boldsymbol{x}_j$, i.e., $z_j = (z_{j,1},\dots,z_{j,p+1})'$ and $z_{j,1} = M_1'\boldsymbol{x}_j$ with M_1 being the first column of M, the integral now becomes

$$\frac{\lambda}{(2\pi)^{p+1}}\int\dots\int \exp(-i\boldsymbol{r}'Mz_j)z_{j,1}^2\, dz_j$$

$$= \frac{\lambda}{(2\pi)^{p+1}}\int\dots\int \exp(-i\boldsymbol{s}'z_j)z_{j,1}^2\, dz_j$$

$$= \frac{\lambda}{2\pi}\int e^{-is^1 z_{j,1}} z_{j,1}^2\, dz_{j,1}\prod_{k=2}^{p+1}\left(\frac{1}{2\pi}\int e^{-is^k z_{j,k}}\, dz_{j,k}\right)$$

$$= -\lambda\delta''(s^1)\delta(s^2)\dots\delta(s^{p+1}),$$

where $\boldsymbol{s} = M'\boldsymbol{r}$. Hence,

$$w^*_{\sigma^2}(\boldsymbol{r}) = -\frac{1}{2\sigma^2}\,\delta(r^1)\dots\delta(r^{p+1}) - \frac{\lambda}{2\sigma^4 g(\boldsymbol{\rho})}\,\delta''(s^1)\delta(s^2)\dots\delta(s^{p+1}). \qquad \textbf{(A.9)}$$

Similarly, for $\boldsymbol{\rho}_k$ we have

$$w^*_{\boldsymbol{\rho}_k}(\boldsymbol{r}) = \left(\frac{1}{2\pi}\right)^{p+1}\int\dots\int \exp(-i\boldsymbol{r}'\boldsymbol{x}_j)\left[-\frac{g_k(\boldsymbol{\rho})}{2g(\boldsymbol{\rho})}+\frac{g_k(\boldsymbol{\rho})}{2\sigma^2 g^2(\boldsymbol{\rho})}\boldsymbol{x}_j' A\boldsymbol{x}_j\right.$$

$$\left. - \frac{1}{2\sigma^2 g(\boldsymbol{\rho})}\boldsymbol{x}_j' B_k\boldsymbol{x}_j\right]d\boldsymbol{x}_j$$

$$= -\frac{g_k(\boldsymbol{\rho})}{2g(\boldsymbol{\rho})}\,\delta(r^1)\dots\delta(r^{p+1}) - \frac{g_k(\boldsymbol{\rho})}{2\sigma^2 g^2(\boldsymbol{\rho})}\,\lambda\delta''(s^1)\delta(s^2)\dots\delta(s^{p+1})$$

$$- \frac{1}{2\sigma^2 g(\boldsymbol{\rho})}\,\frac{1}{(2\pi)^{p+1}}\int\dots\int \exp(-i\boldsymbol{r}'\boldsymbol{x}_j)\boldsymbol{x}_j' B_k\boldsymbol{x}_j\, d\boldsymbol{x}_j.$$

Now $B_k = H_kD_kH_k'$ thus transforming $\boldsymbol{x}_j \to z_j = H_k'\boldsymbol{x}_j$, i.e., $z_j = (z_{j,1},\dots,z_{j,p+1})'$, we have $z_{j,1} = H_{k,1}'\boldsymbol{x}_j, z_{j,2} = H_{k,2}'\boldsymbol{x}_j$, and the integral becomes

$$\frac{1}{(2\pi)^{p+1}}\int\cdots\int \exp(-i\boldsymbol{r}'H_k\boldsymbol{z}_j)(\lambda_k^1 z_{j,1}^2+\lambda_k^2 z_{j,2}^2)\,d\boldsymbol{z}_j$$

$$= \frac{\lambda_k^1}{2\pi}\int e^{-it_k^1 z_{j,1}} z_{j,1}^2\,dz_{j,1}\prod_{l=2}^{p+1}\left(\frac{1}{2\pi}\int e^{-it_k^l z_{j,l}}\,dz_{j,l}\right)$$

$$+ \frac{\lambda_k^2}{2\pi}\int e^{-it_k^1 z_{j,1}}\,dz_{j,1}\,\frac{1}{2\pi}\int e^{-it_k^2 z_{j,2}} z_{j,2}^2\,dz_{j,2}\prod_{l=3}^{p+1}\left(\frac{1}{2\pi}\int \exp(-it_k^l z_{j,l})\,dz_{j,l}\right)$$

$$= -\lambda_k^1\delta''(t_k^1)\delta(t_k^2)\ldots\delta(t_k^{p+1}) - \lambda_k^2\delta(t_k^1)\delta''(t_k^2)\ldots\delta(t_k^{p+1}),$$

where $\boldsymbol{t}_k = H_k'\boldsymbol{r}$ for $k = 1,\ldots,l+m$. Thus

$$w^*_{\boldsymbol{\rho}_k}(\boldsymbol{r}) = -\frac{g_k(\boldsymbol{\rho})}{2g(\boldsymbol{\rho})}\,\delta(r^1)\ldots\delta(r^{p+1}) - \frac{g_k(\boldsymbol{\rho})}{2\sigma^2 g^2(\boldsymbol{\rho})}\,\delta''(s^1)\delta(s^2)\ldots\delta(s^{p+1})$$

$$+ \frac{1}{2\sigma^2 g(\boldsymbol{\rho})}\,(\lambda_k^1\delta''(t_k^1)\delta(t_k^2)\ldots\delta(t_k^{p+1}) + \lambda_k^2\delta(t_k^1)\delta''(t_k^2)\ldots\delta(t_k^{p+1})). \qquad \blacksquare$$

Proof of Corollary 3.1. With $c_n(\boldsymbol{r}) = (1/n)\sum_{j=1}^n \exp(i\boldsymbol{r}'\boldsymbol{x}_j)$ and $c(\boldsymbol{r};\boldsymbol{\theta}) = \exp(-(\sigma^2/2)\boldsymbol{r}'\Omega\boldsymbol{r})$ we substitute $w^*_{\sigma^2}(\boldsymbol{r})$ into (2.3), resulting in

$$\int\cdots\int w^*_{\sigma^2}(\boldsymbol{r})(c_n(\boldsymbol{r}) - c(\boldsymbol{r};\boldsymbol{\theta}))\,d\boldsymbol{r}$$

$$= \int\cdots\int\left[-\frac{1}{2\sigma^2}\,\delta(r^1)\ldots\delta(r^{p+1}) - \frac{1}{2\sigma^4 g(\boldsymbol{\rho})}\,\lambda\delta''(s^1)\delta(s^2)\ldots\delta(s^{p+1})\right]$$

$$\times (c_n(\boldsymbol{r}) - c(\boldsymbol{r};\boldsymbol{\theta}))\,d\boldsymbol{r}$$

$$= -\frac{1}{2\sigma^2}\,(c_n(\boldsymbol{0}) - c(0;\boldsymbol{\theta})) - \frac{\lambda}{2\sigma^4 g(\boldsymbol{\rho})}$$

$$\times \int\cdots\int \delta''(s^1)\delta(s^2)\ldots\delta(s^{p+1})(c_n(\boldsymbol{r}) - c(\boldsymbol{r};\boldsymbol{\theta}))\,d\boldsymbol{r}.$$

Now transforming $\boldsymbol{r}\to\boldsymbol{s} = M'\boldsymbol{r}$, i.e., $\boldsymbol{r} = M\boldsymbol{s}$ we have

$$\lambda\int\cdots\int \delta''(s^1)\delta(s^2)\ldots\delta(s^{p+1})(c_n(\boldsymbol{r}) - c(\boldsymbol{r};\boldsymbol{\theta}))\,d\boldsymbol{r}$$

$$= \lambda\int\cdots\int \delta''(s^1)\delta(s^2)\ldots\delta(s^{p+1})$$

$$\times\left(\frac{1}{n}\sum_{j=1}^n \exp(i\boldsymbol{s}'M'\boldsymbol{x}_j) - \exp\left(-\frac{\sigma^2}{2}\,\boldsymbol{s}'M'\Omega M\boldsymbol{s}\right)\right)d\boldsymbol{s}$$

$$= \lambda\int \delta''(s^1)\left(\frac{1}{n}\sum_{j=1}^n \exp(is^1 M_1'\boldsymbol{x}_j) - \exp\left(-\frac{\sigma^2}{2}\,(s^1)^2 M_1'\Omega M_1\right)\right)ds^1$$

$$= -\lambda\left(\frac{1}{n}\sum_{j=1}^n (M_1'\boldsymbol{x}_j)^2 - \sigma^2 M_1'\Omega M_1\right) = -\left(\frac{1}{n}\sum_{j=1}^n \lambda\boldsymbol{x}_j' M_1 M_1'\boldsymbol{x}_j - \sigma^2\mathrm{tr}(\Omega M_1 M_1'\lambda)\right)$$

$$= -\left(\frac{1}{n}\sum_{j=1}^n \boldsymbol{x}_j' A\boldsymbol{x}_j - \sigma^2\mathrm{tr}(\Omega A)\right).$$

Now, noting that $A = ff'$ with $f = (-f_p',1)$ and partitioning Ω in the same way, we have $f_p' = \Omega_{21}\Omega_{11}^{-1}$, and thus

$$\operatorname{tr}(\Omega A) = (-f_p',1)\Omega(-f_p',1)' = \Omega_{22} - \Omega_{21}\Omega_{11}^{-1}\Omega_{21} = g(\boldsymbol{\rho}),$$

and thus

$$\int \dots \int w^*_{\sigma^2}(\boldsymbol{r})(c_n(\boldsymbol{r}) - c(\boldsymbol{r};\boldsymbol{\theta}))\,d\boldsymbol{r} = -\frac{1}{2\sigma^4 g(\boldsymbol{\rho})}\left(\frac{1}{n}\sum_{j=1}^{n}\boldsymbol{x}_j' A\boldsymbol{x}_j - \sigma^2 g(\boldsymbol{\rho})\right).$$

Similarly, substituting $w^*_{\boldsymbol{\rho}_k}(\boldsymbol{r})$ into (2.3),

$$\begin{aligned}
&\int \dots \int w^*_{\boldsymbol{\rho}_k}(\boldsymbol{r})(c_n(\boldsymbol{r}) - c(\boldsymbol{r};\boldsymbol{\theta}))\,d\boldsymbol{r}\\
&\quad = -\frac{g_k(\boldsymbol{\rho})}{2g(\boldsymbol{\rho})}(c_n(0) - c(0;\boldsymbol{\theta})) + \frac{g_k(\boldsymbol{\rho})}{2\sigma^2 g^2(\boldsymbol{\rho})}\left(\frac{1}{n}\sum_{j=1}^{n}\boldsymbol{x}_j' A\boldsymbol{x}_j - \sigma^2 g(\boldsymbol{\rho})\right)\\
&\qquad + \frac{1}{2\sigma^2 g(\boldsymbol{\rho})}\left[\lambda_k^1 \int \delta''(t_k^1)\left(\frac{1}{n}\sum_{j=1}^{n}\exp(it_k^1 H_{k1}'\boldsymbol{x}_j) - \exp\left(-\frac{\sigma^2}{2}(t_k^1)^2 H_{k1}'\Omega H_{k1}\right)\right)dt_k^1\right.\\
&\qquad\qquad \left. + \lambda_k^2 \int \delta''(t_k^2)\left(\frac{1}{n}\sum_{j=1}^{n}\exp(it_k^2 H_{k2}'\boldsymbol{x}_j) - \exp\left(-\frac{\sigma^2}{2}(\boldsymbol{t}_k^2)^2 H_{k2}'\Omega H_{k2}\right)\right)d\boldsymbol{t}_k^2\right]\\
&\quad = -\frac{1}{2\sigma^2 g(\boldsymbol{\rho})}\left[\frac{1}{n}\sum_{j=1}^{n}(\lambda_k^1\boldsymbol{x}_j' H_{k1}H_{k1}'\boldsymbol{x}_j + \lambda_k^2\boldsymbol{x}_j' H_{k2}H_{k2}'\boldsymbol{x}_j) - \sigma^2(\lambda_k^1 H_{k1}'\Omega H_{k1} + \lambda_k^2 H_{k2}'\Omega H_{k2})\right]\\
&\quad = -\frac{1}{2\sigma^2 g(\boldsymbol{\rho})}\left(\frac{1}{n}\sum_{j=1}^{n}\boldsymbol{x}_j' B_k\boldsymbol{x}_j - \sigma^2\operatorname{tr}(B_k\Omega)\right).
\end{aligned}$$

Now because $B_k = \partial A/\partial\boldsymbol{\rho}_k = (-(\partial f_p/\partial\boldsymbol{\rho}_k)',0)'(-f_p',1) + (-f_p',1)'(-(\partial f_p/\partial\boldsymbol{\rho}_k)',0)$ we have

$$\operatorname{tr}(B_k\Omega) = 2(-f_p',1)\Omega(-(\partial f_p/\partial\boldsymbol{\rho}_k)',0)' = 0,\ \forall k,$$

i.e.,

$$\int \dots \int w^*_{\boldsymbol{\rho}_k}(\boldsymbol{r})(c_n(\boldsymbol{r}) - c(\boldsymbol{r};\boldsymbol{\theta}))\,d\boldsymbol{r} = -\frac{1}{2\sigma^2 g(\boldsymbol{\rho})}\frac{1}{n}\sum_{j=1}^{n}\boldsymbol{x}_j' B_k\boldsymbol{x}_j = 0,\ \forall k. \qquad \blacksquare$$