The Canadian Journal of Statistics
Vol. 18, No. 2, 1990, Pages 155–161
La Revue Canadienne de Statistique

An efficiency result for the empirical characteristic function in stationary time-series models

Andrey FEUERVERGER

University of Toronto

Key words and phrases: Asymptotic efficiency, empirical characteristic function, estimation, Fisher information, stationary processes, time-series models, uniform consistency.
AMS 1985 subject classifications: Primary 62M99, 60E10; secondary 62F12, 60G10, 62E20.

ABSTRACT

It is shown under general conditions that arbitrarily high asymptotic efficiencies can be obtained when the parameters of a stationary time series are estimated by fitting the characteristic functions of the process to their empirical versions. A consistency and a central limit result are also given.

RÉSUMÉ

On considère l'estimation des paramètres d'une série chronologique stationnaire via l'ajustement de la fonction caractéristique du processus à sa version empirique. On montre que, sous des conditions non restrictives, une efficacité asymptotique arbitrairement grande peut être atteinte. Un résultat à propos de la convergence et un théorème de limite centrale sont aussi présentés.

1. INTRODUCTION

Let $\{X_j\}_{j=-\infty}^{\infty}$ be a univariate, strictly stationary time series whose distribution depends on a parameter θ; we are concerned here with estimation of θ from a finite realization $X_1, X_2, \ldots, X_n$. Define the $(p+1)$-vectors

$$Y_j^p = (X_j, X_{j-1}, \ldots, X_{j-p})^{\mathsf{T}} \tag{1.1}$$

and their characteristic functions (cf's)

$$c_\theta^p(t) = \mathcal{E}_\theta\, e^{it^{\mathsf{T}} Y_1^p}, \tag{1.2}$$

where $t = (t^0, t^1, \ldots, t^p)^{\mathsf{T}}$; then the functions $c_\theta^p(t)$, $p = 1, 2, \ldots$, determine the distribution of $\{X_j\}_{j=-\infty}^{\infty}$. Corresponding to (1.2) we may define the sample quantities

$$c_n^p(t) = \frac{1}{n}\sum_{j=1}^{n} e^{it^{\mathsf{T}} Y_j^p}. \tag{1.3}$$

Owing to the statistical dependence amongst successive terms in the sum in (1.3), however, these quantities are unlike the empirical cf (ecf) as defined, for example, in Feuerverger and Mureika (1977). The quantities (1.2) and (1.3) for stationary processes were introduced in Feuerverger and McDunnough (1981b) and, to distinguish from the iid context, were termed there the poly-cf (pcf) and empirical pcf (epcf) respectively. Our object here is to address the following question: can estimation of θ be carried out asymptotically efficiently by means of procedures in which $c_\theta^p(\boldsymbol{t})$ is "fitted" to $c_n^p(\boldsymbol{t})$ for p (fixed) sufficiently large? We give an affirmative answer under general conditions in Theorem 2.3, our main result; but first a basic consistency and a central limit result are established in Theorems 2.1 and 2.2.

For the iid context it was shown in Feuerverger and McDunnough (1981a) that the ecf provides a basis for asymptotically efficient inference. While the fact that this result may be extended to stationary processes is of theoretical interest, we note that applicable methods will result in cases for which the characteristic-function description of the process model is more readily available than explicit density functions or scores.

2. THE RESULTS

Define the maximal correlation $\rho(q)$ to be the supremum correlation between square-integrable random variables measurable with respect to the σ-fields $\sigma\{\ldots, X_{-1}, X_0\}$ and $\sigma\{X_q, X_{q+1}, \ldots\}$ respectively, i.e., $\rho(q)$ is the maximum correlation possible between functions of the X_j's separated in time by a distance q. We shall assume that

$$\sum_{q=1}^{\infty} \rho(q) < \infty. \tag{2.1}$$

While this condition is used in Theorems 2.1 and 2.2 below, we remark that it may be replaced there by the weaker requirements of ergodicity and strong mixing, respectively. On the other hand, a condition substantially closer to that of (2.1) is needed for the existence of the asymptotic covariance structures (2.3) required in the formulation of Theorem 2.3. We first establish consistency of the epfc.

THEOREM 2.1. *If* $\{X_j\}_{j=-\infty}^{\infty}$ *is a strictly stationary time series satisfying* (2.1), *then*

(i) $c_n^p(\boldsymbol{t}) \longrightarrow c^p(\boldsymbol{t})$ *a.s. for all* $\boldsymbol{t} \in \mathbb{R}^{p+1}$,

(ii) $\sup_{-T \le t^0, \ldots, t^p \le T} |c_n^p(\boldsymbol{t}) - c^p(\boldsymbol{t})| \longrightarrow 0$ *a.s. for any fixed* $0 < T < \infty$.

Proof. The condition (2.1) implies ergodicity and therefore $(1/n)\sum_{j=1}^n e^{i\boldsymbol{t}^{\mathrm{T}}\boldsymbol{Y}_j^p} \longrightarrow Ee^{i\boldsymbol{t}^{\mathrm{T}}\boldsymbol{Y}_1^p}$ a.s., thus establishing (i). Ergodicity further implies the a.s. convergence of $F_n^p(x_j, \ldots, x_{j-p}) = (1/n)\sum_{j=1}^n I[X_j \le x_j, \ldots, X_{j-p} \le x_{j-p}]$, the $(p+1)$-dimensional empirical d.f., to the d.f. $F^p(x_j, \ldots, x_{j-p})$ at all $(x_j, \ldots, x_{j-p})$. Then, arguing as in Chung (1968, Section 5.5), for example, this suffices to establish the Glivenko-Cantelli convergence $\sup_{x_j, \ldots, x_{j-p}} |F_n^p(x_j, \ldots, x_{j-p}) - F^p(x_j, \ldots, x_{j-p})| \longrightarrow 0$ a.s., and from this (ii) follows directly on applying the criteria of complete convergence and uniform convergence (Loève 1977, p. 204, B and C). Q.E.D.

We remark that while the question of whether T in part (ii) of the theorem may be replaced by a sequence $T_n \longrightarrow \infty$ is of interest, the sharp result involving $T_n = e^{o(n)}$ of Csörgő and Totik (1983) would apparently require a very rapid rate of strong mixing in the dependent case, in view of the essential dependence of the proof on the Bernstein inequality (*ibid.*, middle of p. 145).

Define now the *epcf process*

$$W_n^p(\boldsymbol{t}) = \sqrt{n}\{c_n^p(\boldsymbol{t}) - c^p(\boldsymbol{t})\}. \tag{2.2}$$

Since $\mathcal{E}\, c_n^p(\boldsymbol{t}) = c^p(\boldsymbol{t})$, we have $\mathcal{E}\, W_n^p(\boldsymbol{t}) = 0$ for all $\boldsymbol{t}$. Concerning the covariance structure, under the mixing condition (2.1) we may show that

$$\lim_{n\to\infty} Cov\left(W_n^p(\boldsymbol{s}), W_n^p(\boldsymbol{t})\right) = \sum_{r=-\infty}^{\infty} Cov\left(e^{i\boldsymbol{s}^{\mathsf{T}}\boldsymbol{Y}_o^p}, e^{i\boldsymbol{t}^{\mathsf{T}}\boldsymbol{Y}_r^p}\right) \tag{2.3}$$

and that this sum converges absolutely. Clearly its rth term may be written down in terms of the pcf function of order $p+|r|$ and under (2.1) becomes negligible as $|r| \to \infty$. (In the case where $\{X_j\}$ is an iid sequence, note that only the terms $-p \le r \le p$ contribute.) In analogy with the ecf, it is to be expected that $W_n^p(\boldsymbol{t})$ will converge weakly to a Gaussian process over finite regions in $\mathbb{R}^{p+1}$, assuming adequate mixing and moment-type conditions on the underlying process. Such questions are quite technical. However, a finite-dimensional version of this result is readily established.

THEOREM 2.2. *Assume $\{X_j\}_{j=-\infty}^{\infty}$ satisfies (2.1), and let $W^p(\boldsymbol{t})$, $\boldsymbol{t} \in \mathbb{R}^{p+1}$, be a zero-mean complex-valued Gaussian process with the covariance structure (2.3). Then $\left(W_n^p(\boldsymbol{t}_1), \ldots, W_n^p(\boldsymbol{t}_k)\right)$ converges in distribution to $\left(W^p(\boldsymbol{t}_1), \ldots, W^p(\boldsymbol{t}_k)\right)$ for all k, and all $\boldsymbol{t}_1, \boldsymbol{t}_2, \ldots, \boldsymbol{t}_k$ in $\mathbb{R}^{p+1}$.*

Proof. The result is obtained using a central limit theorem for stationary processes of the type of Theorem 18.5.4 of Ibragimov and Linnik (1971). In particular, note that the condition (2.1) implies the strong mixing condition of Section 17.2 of the cited reference and hence the required central-limit result. Q.E.D.

Turning now to the inference question which is our main concern here, we shall show that under general conditions the epcf lends itself to asymptotically arbitrarily highly efficient inference in stationary processes. The procedures considered involve minimization (in θ) of quadratic forms such as $(\boldsymbol{V}_n - \boldsymbol{V}_\theta)^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{V}_n - \boldsymbol{V}_\theta)$, where $\boldsymbol{V}_\theta = \left(\operatorname{Re} c_\theta^p(t_1), \ldots, \operatorname{Re} c_\theta^p(t_k), \operatorname{Im} c_\theta^p(t_1), \ldots, \operatorname{Im} c_\theta^p(t_k)\right)^{\mathsf{T}}$, $\boldsymbol{V}_n$ is its empirical version, and $\boldsymbol{\Sigma}^{-1}$ is a consistent estimate of the covariance matrix of $\boldsymbol{V}_n$ determined from (2.3). To circumvent the delicate considerations that can arise in connection with inference in stochastic processes, we do not attempt any maximal reduction of the assumptions required, although the assumptions made are satisfied in many typical situations. The basic result is stated and proved below for a real univariate parameter; the extension to the multiparameter case is straightforward.

THEOREM 2.3. *Let $\{X_j\}_{j=-\infty}^{\infty}$ be a stationary process having distribution belonging to the class defined by the pcf functions $c_\theta^p(\boldsymbol{t})$, and suppose the following regularity conditions hold:*

(A1) *The parameter θ is real, univariate, and defined on a closed interval with unknown true value θ_0 assumed to lie in the interior.*

(A2) *Different values of θ yield different process distributions.*

(A3) *The $c_\theta^p(\boldsymbol{t})$ correspond to densities $f_\theta(x_p, x_{p-1}, \ldots, x_0)$ for $(X_p, X_{p-1}, \ldots, X_0)$ which are twice differentiable in θ.*

(A4) *The mixing condition (2.1) holds at θ_0.*

(A5) *The expected conditional Fisher information quantity*

$$I(\theta; p) = \mathcal{E}_\theta\left(\frac{\partial \log f_\theta(X_j \mid X_{j-1}, \ldots, X_{j-p})}{\partial\theta}\right)^2$$

satisfies $I(\theta; p) < \infty$ for all p.

(A6) *$I(\theta;p) \longrightarrow I(\theta)$ as $p \longrightarrow \infty$, where $0 < I(\theta) < \infty$.*

(A7) *$\int f_\theta(x_j \mid x_{j-1},\dots,x_{j-p})\,dx_j$ can be twice differentiated in θ through the integral almost everywhere in $x_{j-1},\dots,x_{j-p}$.*

(A8) *The densities possess a third derivative satisfying*

$$\left|\frac{\partial^3 \log f_\theta(x_j \mid x_{j-1},\dots,x_{j-p})}{\partial\theta^3}\right| \le M(x_j,\dots,x_{j-p})$$

in some neighborhood of θ_0 where $\mathcal{E}_{\theta_0} M(X_j,\dots,X_{j-p}) < \infty$.

Then under these assumptions, the procedure based on fitting $c_\theta^p(\boldsymbol{t})$ to $c_n^p(\boldsymbol{t})$ at a grid of points $\boldsymbol{t}_j \in \mathbb{R}^{p+1}$, $j = 1,2,\dots,k$, by means of (nonlinear) least squares weighted in accordance with a consistent estimator of the covariances as given by (2.3) results in an estimator which is asymptotically normal and which can be made to have arbitrarily high asymptotic efficiency provided that p (fixed) is sufficiently large and the grid $\{t_j\}$ is sufficiently fine and extended.

Proof. For a sample size n, define the expected Fisher information

$$I_n(\theta) = \mathcal{E}_\theta\left(\frac{\partial \log f_\theta(X_n,\dots,X_1)}{\partial\theta}\right)^2, \tag{2.4}$$

and use conditioning to see that

$$I_n(\theta) = \mathcal{E}_\theta\left(\sum_{j=1}^n \frac{\partial \log f_\theta(X_j \mid X_{j-1},\dots,X_1)}{\partial\theta}\right)^2 = \sum_{j=1}^n \mathcal{E}_\theta\left(\frac{\partial \log f_\theta(X_j \mid X_{j-1},\dots,X_1)}{\partial\theta}\right)^2,$$

so that $I_n(\theta)/n \longrightarrow I(\theta)$ in view of (A.5) and (A.6). Then in the present context, an estimator $\hat\theta_n$ will be asymptotically efficient provided that $\sqrt{n}(\hat\theta_n - \theta)$ is asymptotically normal with mean zero and variance $[I(\theta)]^{-1}$.

Introduce now the auxiliary estimators $\hat\theta_p$ to be obtained by solving

$$\sum_{j=1}^n \frac{\partial \log f_\theta(X_j \mid X_{j-1},\dots,X_{j-p})}{\partial\theta} = 0. \tag{2.5}$$

The existence of a consistent root $\hat\theta_p$ of (2.5) for p fixed sufficiently large may be established by arguments similar to those needed to establish a consistent root for the likelihood equation. Specifically, for fixed p, define the likelihood-like quantity $K_n(\theta) = \prod_{j=1}^n f_\theta(X_j \mid X_{j-1},\dots,X_{j-p})$, and let S_n be the set of sequences $\{X_j\}$ for which both $K_n(\theta_0) > K_n(\theta_0+\epsilon)$ and $K_n(\theta_0) > K_n(\theta_0-\epsilon)$, where $\epsilon > 0$ is arbitrarily small and fixed. Then $P(S_n) \longrightarrow 1$. For

$$\frac{1}{n}\log\frac{K_n(\theta_0+\epsilon)}{K_n(\theta_0)} = \frac{1}{n}\sum_{j=1}^n \log\frac{f_{\theta_0+\epsilon}(X_j \mid X_{j-1},\dots,X_{j-p})}{f_{\theta_0}(X_j \mid X_{j-1},\dots,X_{j-p})},$$

and, since by (A4) the law of large numbers holds here, this converges to

$$\mathcal{E}_{\theta_0}\log\frac{f_{\theta_0+\epsilon}(X_j \mid X_{j-1},\dots,X_{j-p})}{f_{\theta_0}(X_j \mid X_{j-1},\dots,X_{j-p})} < \log\mathcal{E}_{\theta_0}\frac{f_{\theta_0+\epsilon}(X_j \mid X_{j-1},\dots,X_{j-p})}{f_{\theta_0}(X_j \mid X_{j-1},\dots,X_{j-p})} = 0$$

in view of Jensen's inequality and the fact that, by (A2), $\theta_0 + \epsilon$ and by θ_0 yield different process distributions, so that the two conditional distributions must essentially differ for p large enough. Consequently

$$\lim_{n\to\infty} \frac{1}{n} \log \frac{K_n(\theta_0+\epsilon)}{K_n(\theta_0)} < 0,$$

and similarly,

$$\lim_{n\to\infty} \frac{1}{n} \log \frac{K_n(\theta_0-\epsilon)}{K_n(\theta_0)} < 0,$$

so that indeed $P_{\theta_0}(S_n) \longrightarrow 1$. Now $K_n(\theta)$ is differentiable, so that on S_n we shall have a root $\hat{\theta}_n$ of $\partial K_n(\theta)/\partial\theta = 0$ such that $|\hat{\theta}_n - \theta_0| < \epsilon$. Thus for any $\epsilon > 0$ there exists a sequence of roots satisfying $P_{\theta_0}(|\hat{\theta}_n(\epsilon) - \theta_0| < \epsilon) \longrightarrow 1$. Finally, noting that $\partial K_n(\theta)/\partial\theta$ is continuous, the root of (2.5) closest to θ_0 will be well defined and clearly consistent.

Next, for any consistent root $\hat{\theta}_p$ of (2.5) we have, following the usual Taylor-expansion approach, that

$$\sqrt{n}(\hat{\theta}_p - \theta_0) = \frac{-\dfrac{1}{\sqrt{n}} \displaystyle\sum_{j=1}^{n} \frac{\partial \log f_\theta(X_j \mid X_{j-1},\dots,X_{j-p})}{\partial\theta_0}}{\dfrac{1}{n} \displaystyle\sum_{j=1}^{n} \frac{\partial^2 \log f_\theta(X_j \mid X_{j-1},\dots,X_{j-p})}{\partial\theta_*^2}} \tag{2.6}$$

for some θ_* between $\hat{\theta}_p$ and θ_0. It follows that $\sqrt{n}(\hat{\theta}_p - \theta_0)$ is asymptotically normal with asymptotic variance $\{I(\theta_0;p)\}^{-1}$. The details of this argument are standard [e.g. Lehmann (1983, Chapter 6, Theorem 2.3) or Cramér (1946, §33.3]. Specifically, by (A8), θ_* in the denominator of (2.6) may be replaced by θ_0 asymptotically. Secondly, by (A4) the process is ergodic, so that the denominator of (2.6) will converge a.s. to $\mathcal{E}_{\theta_0}\left(\partial^2 \log f_\theta(X_j \mid X_{j-1},\dots,X_{j-p})/\partial\theta_0^2\right)$, which, in view of (A7), equals $-I(\theta_0;p)$. Finally, by (A4) the process is sufficiently mixing so that the numerator of (2.6) is subject to the central limit theorem and hence asymptotically normal [see, for example, Ibragimov and Linnik (1971, Chapter 18)]. Then since $I(\theta_0;p) \longrightarrow I(\theta_0)$, it follows that the auxiliary estimators $\hat{\theta}_p$ approach asymptotic efficiency as p increases.

To relate this to the epcf context, write (2.5) in the form

$$\int\cdots\int \frac{\partial \log f_\theta(x_j \mid x_{j-1},\dots,x_{j-p})}{\partial\theta}\, d\left(F_n^p(x_j,\dots,x_{j-p}) - F_\theta^p(x_j,\dots,x_{j-p})\right) = 0, \tag{2.7}$$

where F_n^p is the $(p+1)$-variate empirical distribution of $(X_j,\dots,X_{j-p})$, $j = 1,\dots,n$, and F_θ^p is the actual distribution, whose introduction is seen easily to provide only a term whose value is identically zero. Now, applying Parseval's identity, rewrite (2.7) in the form

$$\int\cdots\int \{c_n^p(\boldsymbol{t}) - c_\theta^p(\boldsymbol{t})\} W_\theta^p(\boldsymbol{t})\, dt^0 \cdots dt^p = 0, \tag{2.8}$$

where the epcf and pcf are obtained from Fourier-transforming F_n^p and F_θ^p, while W_θ^p is obtained via the inverse transform

$$W_\theta^p(\boldsymbol{t}) = \left(\frac{1}{2\pi}\right)^{p+1} \int\cdots\int e^{-i(t^0 x_j + \cdots + t^p x_{j-p})} \frac{\partial \log f_\theta(x_j \mid x_{j-1},\dots,x_{j-p})}{\partial\theta}\, dx_j \cdots dx_{j-p}. \tag{2.9}$$

Examining (2.8), the efficiency claim for the epcf now becomes apparent, although the remainder of the proof still requires detailed arguments which closely follow those in Sections 3 to 5 of Feuerverger and McDunnough (1981a), to which we must now make reference. Specifically, because the conditional score $\partial \log f_\theta(x_j \mid x_{j-1}, \ldots, x_{j-p})/\partial\theta$ is not in general integrable (with respect to Lebesgue measure), we must first replace it in (2.7) and (2.9) by a tapered version such as $\{\partial \log f_\theta(x_j \mid x_{j-1}, \ldots, x_{j-p})/\partial\theta\} \prod_{l=0}^{p} h_M(x_{j-l})$, where, for example, $h_M(x) = 1$ when $|x| \leq M$, and $= 0$ otherwise. Straightforward calculations as in Section 5 of the cited reference then serve to establish that as $M \longrightarrow \infty$, the asymptotic efficiency of the estimating equation (2.7) based on values of $M < \infty$ approaches that of the equation (2.7) based on $M = \infty$. Next, for any fixed $M < \infty$ the estimating equation (2.8) may be approximated (using a finite sum) by means of a moment estimator (*ibid.*, Section 4) and hence in turn by a least-squares procedure as defined in Section 3 of the reference cited. We thus obtain asymptotic efficiencies arbitrarily close to that of $\hat{\theta}_p$ by discrete procedures based on a sufficiently extensive grid $\{\boldsymbol{t}_j\}$, and hence the result. Q.E.D.

Although Theorem 2.3 is of theoretical interest, it is important to realize that in order to implement the implied procedure, the pcf functions must be available, and the covariances (2.3) computed (in terms of the pcf functions) at the gridpoints selected. This somewhat limits the feasibility of the procedure in many cases. On the other hand, for iid processes, the covariance structure (2.3) is seen to be readily computed as a finite sum of pcf functions, and consistently (nonparametrically) estimated, even for s, t ranging over appropriate grids in R^{p+1}. This suggests studying estimation procedures, for time-series models of the type $X_j = h_\beta(X_{j-1}, X_{j-2}, \ldots) + e_j$ having iid errors, by means of testing for independence, through statistics based upon the epcf functions, amongst the *fitted residuals* in the time-series model. Here the estimates would be taken as the parameter values that minimize the statistic measuring dependency. Of course, the optimality statement of Theorem 2.3 does not automatically carry over to this new procedure, but one might expect that high levels of efficiency would be obtained in many cases. This approach holds promise for the analysis of nonlinear time-series models and will be considered elsewhere.

ACKNOWLEDGEMENT

The author would like to express his thanks to the referees and the editor for their very helpful comments, and to J. Aaronson and P. McDunnough for many valuable discussions. A first draft of this work was carried out while the author was visiting Tel Aviv University, whose hospitality is gratefully acknowledged. This research was supported by an Operating Grant from the Natural Sciences and Engineering Research Council of Canada.

REFERENCES

Chung, K.L. (1968). *A Course in Probability Theory*. Harcourt, Brace and World, New York.

Cramér, H. (1946). *Mathematical Methods of Statistics*. Princeton Univ. Press.

Csörgő, S., and Totik, V. (1983). On how long an interval is the empirical characteristic function uniformly consistent? *Acta. Sci. Math. (Szeged)*, 45, 141–149.

Feuerverger, A., and McDunnough, P. (1981a). On the efficiency of empirical characteristic function procedures. *J. Roy. Statist. Soc. Ser. B*, 43(1), 20–27.

Feuerverger, A., and McDunnough, P. (1981b). On some Fourier methods for inference. *J. Amer. Statist. Assoc.*, 76(374), 379–387.

Feuerverger, A., and Mureika, R.A. (1977). The empirical characteristic function and its applications. *Ann. Statist.*, 5(1), 88–97.

Ibragimov, I.A., and Linnik, Y.V. (1971). *Independent and Stationary Sequences of Random Variables*, Wolters-Noordhoff, Groningen, The Netherlands.

Lehmann, E.L. (1983). *Theory of Point Estimation*. Wiley, New York.

Loève, M. (1977). *Probability Theory I*. Fourth Edition. Springer-Verlag, New York.

Received 16 November 1987
Revised 8 January 1990
Accepted 14 January 1990

Department of Statistics
University of Toronto
Toronto, Ontario M5S 1A1