Whittle estimation of multivariate exponential volatility models with long memory

M.Marchese[1], P.Zaffaroni

Cass Business School and Imperial College Business School

Abstract

For a class of asymmetric multivariate exponential volatility models we establish the strong consistency and the asymptotic normality of the Whittle estimator of the parameters under a variety of parameterisations that include long-range dependence in the volatility dynamics. We contribute to the long-memory statistical literature by establishing the convergence of quadratic forms of vector linear processes whose innovations need not be identically distributed and whose spectral density function might not be factorable.We assess the finite sample properties of the estimator with a Monte Carlo simulation and compare them with those of the the maximum likelihood estimator, showing that in some cases they perform comparably. An empirical application, using three market indexes (FTSE100, S&P 500 and Nikkei 225) suggests the potential of the model to capture the joint dynamics of asset returns volatilities.

Keywords:
Multivariate volatility, asymmetry, long memory, signal plus noise, Whittle estimation, asymptotics
JEL: C32, C51, G12

1. Introduction

The importance of modelling comovements of financial returns has long been established in financial econometrics and financial applications. The knowledge of correlation structures is vital for asset pricing, optimal portfolio allocation and risk management. Moreover, as the volatilities of different assets and markets move together, modelling volatility in a multivariate framework can lead to greater statistical efficiency. A number of multivariate specifications has been proposed in the literature along two main lines of research: conditional volatility models, where the volatilities are a deterministic function of the past realizations of the assets, and stochastic volatility models, where they are latent (comprehensive literature reviews can be found in Bauwens, Laurent and

[1]Corresponding author
E-mail address: Malvina.Marchese@city.ac.uk

Rombouts, 2006, Silvennoinen and Terasvirta, 2008 and Asai, McAleer and Yu, 2006). There is a long standing debate on the relative advantages of one class over the other.

In both classes the exponential specification of individual volatilities has been found to offer several advantages: the absence of non negative constraints on the parameters, the

In both classes research has combined the need to ensure positive definitness of the conditional covariance matrix and parsimony of its parameterization with the need to capture the complex individual and joint dynamics of asset returns, synthesized in a well know number of stylized facts.

Despite the innumerable developments of multivariate volatility models, to our knowledge, no parametrization allowing for long memory autocorrelated squared returns has been proposed.

In this paper we consider a way of modelling the joint dynamics of asset volailities by means of a family of multivariate exponential models that nests conditional and stochastic volatility specifications:

$$\mathbf{x}_t = \mu_t + \mathbf{u}_t, \quad \mathbf{u}_t = \mathbf{D}_t \mathbf{z}_t, \quad t \in Z \tag{1}$$

with

$$\mathbf{D}_t = diag\left\{\exp(0.5\mathbf{h}_t)\right\}, \quad \mathbf{h}_t = \omega_0 + \sum_{j=0}^{\infty} \mathbf{\Psi}_j \epsilon_{t-j-1} \quad \sum_{j=0}^{\infty} \left\|\mathbf{\Psi}_j\right\|^2 < \infty, \tag{2}$$

where $\mathbf{x}_t$ is the n-dimensional vector of asset returns and

Another important property of returns, well documented in empirical findings, is that that power transformations of absolute returns have significant autocorrelations that decay to zero at a slow rate, consistent with the notion of long-memory and non-summability of the autocovariances. Since the introduction of the fractionally integrated GARCH (FIGARCH) model of Bollerslev and Mikklesen (1996), univariate long-memory volatility models have received a great deal of attention (see Robinson and Zaffaroni 1996, 1997, and Ruiz and Veiga, 2006 among others). However very few generalizations to multivariate framework have been proposed so far. In this paper we consider a way to model the joint dynamics of assets volatilities by means of a family of multivariate exponential volatility models, that nests conditional and stochastic volatility specifications.

We consider an n-dimensional observable time-series satisfying

$$\mathbf{x}_t = \mathbf{D}_t \mathbf{z}_t, \quad t \in Z \tag{1}$$

$$\mathbf{D}_t = diag\left\{\exp(0.5\mathbf{h}_t)\right\}, \tag{2}$$

$$\mathbf{h}_t = \omega_0 + \sum_{j=0}^{\infty} \mathbf{\Psi}_{0j} \epsilon_{t-j-1} \quad \sum_{j=0}^{\infty} \left\|\mathbf{\Psi}_{0j}\right\|^2 < \infty, \quad \text{a.s.}, \tag{3}$$

$$Var\begin{pmatrix} \mathbf{z}_t \\ \epsilon_t \end{pmatrix} = \begin{pmatrix} \sum_z & \sum_{z\epsilon} \\ \sum_{z\epsilon} & \sum_\epsilon \end{pmatrix}, \tag{4}$$

where $Z = \{t : t = 0, \pm 1, ...\}$. The $\{\mathbf{z}_t, \epsilon_t\}$ form a sequence of independent identically distributed ($i.i.d$) unobservable zero-mean shocks which can be cross correlated for some t. The diagonal elements of Σ_z are set to one, implying that Σ_z is a correlation matrix and $D_t \Sigma_z D_t$ is the time-varying conditional variance matrix of $\mathbf{x}_t$. The joint evolution of the log conditional volatilities is driven by equation (3), where the parameter matrices $\mathbf{\Psi}_{0j}$ characterize the memory of the process. The square summability condition rules out any non-stationary parameterizations but is mild enough to allow for exponential or hyperbolic decay in the coefficients, inducing short or long memory persistence in the volatilities.

Equations (1) to (4) represent in fact a class of multivariate exponential volatility models that nests one and two-shocks specifications. When $\mathbf{z}_t$ and ϵ_t are jointly Normal, model (1) to (4) is the multivariate Stochastic Volatility model with Leverage (SV-L) of Danielsson (1998) and Asai and McAleer (2006), where asymmetries and leverage effect are introduced via cross correlation of the mean and the variance equations shocks. On the other hand when $\epsilon_t \equiv \epsilon(\mathbf{z}_t)$ for some instantaneous transformation $\epsilon(.)$, (1) to (4) yield an exponential constant Conditional Correlation model. In this case the level equation shocks $\mathbf{z}_t$ drive also the volatilities, which evolve according to different exponential specifications with various degrees of asymmetry.

Estimation of multivariate volatility models is usually based on the (Pseudo) Maximum Likelihood estimator (henceforth PMLE). The PML estimator is easily implementable, at the cost of some distributional assumptions on the innovations, however its asymptotic properties are yet to be established for exponential models. Indeed the recursiveness that characterizes these models makes extremely difficult to establish their invertibility and hence the uniform convergence of the Hessian matrix in a neighborhood of the true parameter.Some results on the asymptotics of PMLE for exponential volatility models are available under highly specific assumptions that cannot readily be verified (see for example Straumann and Mikosch (2005) and Witenberger (2013)). A further disadvantage of the PMLE is that the truncation required to compute the observable log likelihood might induce a non-negligible asymptotic bias in the presence of long memory parameterisation (Robinson and Zaffaroni 2006).

These arguments do not apply to the Guassian estimator in the sense of Whittle (1962), partly due to its frequency domain specification. The main contribution of this paper is to establish the strong consistency and asymptotic normality of the Whittle estimator for the general class of multivariate exponential volatility models in (1) to (4) , employing the squares of the observables, under a variety of parametrisations. To this end, denoting $\mathbf{y}_t = \log x_t^2$ we have

$$\mathbf{y}_t = \mu + \mathbf{h}_t + \xi_t \tag{5}$$

where $\mu = \omega + \mathbf{E}\left(\log \mathbf{z}_t^2\right)$ and $\xi_t = \log \mathbf{z}_t^2 - \mathbf{E}\left(\log \mathbf{z}_t^2\right)$ is the $i.i.d$ noise with zero mean and variance Σ_ξ. Apart from the constant μ, $\mathbf{y}_t$ takes the form of a vector signal plus noise process where the signal $\mathbf{h}_t$ corresponds to the volatility of the original series and might exhibit short or long memory parameterizations.

The form of the covariance matrix in (4) implies that $\mathbf{y}_t$ is a correlated signal plus noise process with autocovariance function:

$$\mathbf{\Gamma_y}(u) = I_{u=0}\mathbf{\Sigma}_\xi + \mathbf{\Sigma}_\epsilon \sum_{j=0}^{\infty} \mathbf{\Psi}_j \mathbf{\Psi}_{j+u}^{'} + \mathbf{\Sigma}_{\xi\epsilon} I_{u\neq 0} \mathbf{\Psi}_{|u|-1}, \tag{6}$$

where $I_{u=0} = 1$ if $u = 0$ and zero otherwise. The autocovariances of the transformed process capture the indirect spillover effects and asymmetries in the volatilities of the original multivariate process. As a consequence, the model-spectrum takes the non-factorable form:

$$\begin{aligned} \mathbf{f}\left(\lambda\right) &= \frac{\Sigma_\xi(\tau)}{2\pi} + \frac{\mathbf{k}(e^{i\lambda},\zeta)\Sigma_\epsilon(\tau)\mathbf{k}(e^{i\lambda},\zeta)^*}{2\pi} + \Sigma_{\epsilon\xi}(\tau)e^{-i\lambda}\mathbf{k}(e^{i\lambda},\zeta)^* + \\ &\quad + e^{i\lambda}\mathbf{k}(e^{i\lambda},\zeta)\Sigma_{\epsilon\xi}'(\tau), \quad -\pi < \lambda < \pi \end{aligned} \tag{7}$$

where $\mathbf{k}(z,\zeta) = \sum_{j=0}^{\infty} \mathbf{\Psi}_j\left(\zeta\right) z^j$, $|z| < 1$, is the transfer function of the matrix sequence $\{\mathbf{\Psi}_j\}$.

Statistical literature on the Whittle estimator has established its asymptotic properties under different conditions when the true underlying model is a purely non deterministic vector linear process with innovations which are either *iid* (Giraitis and Surgailis, 1990) or satisfy a set of mixing conditions with respect to their conditional moments (Hosoya, 1997). For such processes, the asymptotic normality of the estimator is well established under short and long memory parameterisations. All these results exploit in an essential way the spectral density factorisation. However when the spectrum of the observables is the sum of two or more components a non factorable structure might arise. Even if the process is linearly regular and its spectrum has a factorable representation, it might not be possible to express the parameters of this representation as closed form functions of the parameters of the signal and the noise (for a detailed discussion see Robinson 1978). It turns out that the available asymptotic theory can only be extended to the case of an independent signal plus noise process with an autoregressive short memory signal (see Hosoya and Taniguchi 1982 and Dunsmuir 1979). In practice however correlated signal plus noise processes arise from the linearisation of exponential volatility models, EGARCH, FIEGARCH and stochastic volatility models, making asymptotics results for the Whittle estimator quite desirable. Hurvich, Mouliners and Soulier (2007) consider semiparametric estimation of the memory parameter in univariate signal plus noise processes with potentially correlated noise and long memory signal. They derive the consistency and asymptotic normality of the local Whittle estimator when the memory parameter lies in the interval $(0, 3/4)$. Zaffaroni (2009) extends Zaffaroni (2003) and establishes the asymptotic properites of the parametric Whittle estimator in a univariate signal plus noise process. As in Giraitis and Surgailis and Hosoya, the proof of the asymptotic normality of the estimator relies on the approximation of the score vector by a different quadratic form with shorter memory. However while the formers rely on the spectral density factorisazion to establish this approximation, Zaffaroni employs a truncation of the signal process. His results are limited to the univariate case and do not readily extend to quadratic forms of multivariate processes with possibly varying weights. In this paper we present a generalisation results for correlated signal plus noise models

Estimation of model (1) to (4) requires to finitely parametrize the signal coefficients $\mathbf{\Psi}_{0j}$. We assume that we know a set of functions $\mathbf{\Psi}_j(.)$ of the $p \times 1$ vector ζ with $p < \infty$, such that, for some unknown ζ_0, $\mathbf{\Psi}_j(\zeta_0) = \mathbf{\Psi}_{0j}$. Analogously we parametrize the covariance matrices assuming that we know functions $\Sigma_\epsilon(.)$, $\Sigma_\xi(.)$, $\Sigma_{\epsilon\xi}(.)$ of the $q \times 1$ vector τ with $q < \infty$, such that, for some unknown τ_0, $\Sigma_\epsilon(\tau_0) = \Sigma_{0\epsilon} = Var(\epsilon_0)$, $\Sigma_\xi(\tau_0) = \Sigma_{0\xi} = Var(\xi_0)$ and $\Sigma_{\epsilon\xi}(\tau_0) = \Sigma_{0\epsilon\xi} = Cov(\epsilon_0, \xi_0)$. We don't make any assumption on the joint density of the innovations $\varepsilon_t \equiv \{\xi_t', \epsilon_t'\}$, so τ_0 contains the $n + n(n-1)/2$ unknown parameters of $vech(\Sigma_{0\epsilon\xi})$, and the $n(n-1)/2$ unknown parameters of respectively $\Sigma_{0\epsilon}$ and $\Sigma_{0\xi}$, yielding $q = n + 3n(n-1)/2$. This specification of τ_0 however can be straightforward extended to models where the joint density of the innovations is specified up to some unknown parameters. We wish to estimate the $s\,(\equiv p+q)$ dimensional vector $\theta_0' = (\zeta_0', \tau_0')'$ on the basis of a sample $\{\mathbf{y}_1, ..., \mathbf{y}_T\}$ of observations, the prime denoting transposition. To this end define:

$$\hat{Q}_T(\theta) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \log\det \mathbf{f}(\lambda, \theta) + tr\left\{\mathbf{f}^{-1}(\lambda, \theta)\mathbf{I}_T(\lambda)\right\} d\lambda,$$

$$\bar{Q}_T(\theta) = \frac{1}{T} \sum_{t=1}^{T-1} \log\det \mathbf{f}(\lambda_t, \theta) + tr\left\{\mathbf{f}^{-1}(\lambda_t, \theta)\mathbf{I}_T(\lambda_t)\right\} d\lambda,$$

for $\lambda_t = \frac{2\pi t}{T}$. Hereafter $\mathbf{I}_T(\lambda)$ denotes the periodogram of the data, $\mathbf{I}_T(\lambda) = \mathbf{w}_T(\lambda)\mathbf{w}_T(\lambda)^*$, with $\mathbf{w}_T(\lambda) = \frac{1}{\sqrt{2\pi T}} \sum_{t=1}^{T} \mathbf{y}_t e^{i\lambda t}$. For a prescribed compact subset of $\mathbf{R}^s$, Θ, we call $\hat{\theta}$ and $\overline{\theta}$ the estimators of θ_0 got by maximising the corresponding Whittle objective function. In practical applications the discrete frequency estimator is preferred over the continuous one since it can be computed efficiently using the fast Fourier transform. Thanks to the invariance properties of the periodogram at the Fourier frequencies, mean-correction of $\mathbf{y}_t$ is taken care of by omission of summands $t = 0$ and $t = T$.

The paper is organized as follows. The next section lists our assumptions, with discussion, Section 3 presents the main results, namely the strong consistency and the asymtptotic normality of the Whittle estimator. Section 4 reports the results of Monte Carlo exercise to compare the finite sample behaviour of the Whittle and MLE estimators of the parameters. Section 5 presents an empirical application based on observed time series of three market indices: Nikkei 225, FTSE 100 and S&P500. Concluding remarks are in Section 6. The results are formally stated in theorems and propositions, the proofs of which are reported in the mathematical appendix together with the technical lemmas.

2. Assumptions.

Hereafter we denote by $A_{i,j}$ the (i,j) element of the matrix A, and by $\mathbf{x}_i(t)$the ith element of the vector $\mathbf{x}$ at time t. We assume that all the elements of ξ_t, ϵ_t, $\mathbf{\Psi}_j(\zeta)$ are real. We define $\Pi \equiv [-\pi, \pi]$ and denote by $L_p(\Pi)$ the class of p-integrable functions defined on Π. The symbol "> 0" denotes strict positive definiteness when applied to a matrix, and K denotes a generic constant, not always the same. The function Spectral domain regularity conditions are common in the statistical litterature on Whittle estimation, see for example Giraitis and Surgailis (1990), Heyde and Gay (1993), and Hosoya (1997). We impose conditions directly on the model spectrum and its higher order derivatives, defining unambiguously their behavior near the origin, as well as a form of uniform continuity away from the zero frequency. Our regularity conditions cover both short and long memory parameterizations of the signal coefficients.

The first assumption we introduce has two versions: the weaker version ($k = 2$) will be employed in our proof of consistency of the estimators, the stronger version ($k = 4$) in our proof of asymptotic normality.

Assumption 1(k) $k \geq 2$. The $\varepsilon_t \equiv \{\epsilon_t', \xi_t'\}$ form a sequence of $i.i.d$ zero mean unobservable random vectors with finite joint k-order moment.

Assumption 1 together with the square summability of the coefficients in (3) ensures the strict stationarity and ergodicity of the $\mathbf{y}_t$, which are common assumptions in the statistical literature on Whittle estimation. Robinson (1978) replaces strict stationarity by a weaker assumption of fourth order stationarity. Hosoya and Taniguchi (1982) dispense with ergodicity imposing a Lindeberg type condition. The following assumption is a standard one to ensures that θ_0 is an interior point of the compact closure of an open s-dimensional manifold (see for example Hannah, 1973 and Robinson, 1978). It implies boundedness of any function of $\theta \in \Theta$.

Assumption 2. θ_0 is an interior point of the compact parameter space $\Theta \in R^s$.

Assumption 3 For every $\theta \in \Theta$ whenever $\theta \neq \theta_0$, $\mathbf{f}(\lambda, \theta) \neq \mathbf{f}(\lambda, \theta_0)$.

Assumption 3 rules out the possibility of two equivalent structures giving rise to the same spectral density,thus granting identification. Following Hannan (1973), Dunsmuir and Hannan (1976), and Hosoya and Taniguchi (1982) we restrict the parameter space to a subset of the parameter space where the spectrum is positive, with the following assumption:

Assumption 4. For any $\theta \in \Theta$, $\mathbf{f}(\lambda, \theta)$ is a strictly positive definite matrix.

Consistency of the estimators requires regularity conditions defining the behavior of the spectrum near the origin, as well as a form of uniform continuity away from the zero frequency. Hereafter $d(\theta)$ denotes the memory parameter that charactrizes decay rate of the signal coefficients.For $d(\theta) = 0$ the process has short memory and the spectral density is continuous at all frequencies, when $d(\theta) \in (-1/2, 0)$ the process has negative memory and $\mathbf{f}(0) = 0$, when $d(\theta) \in (0, 1/2)$ the process has long memory and the spectral density is unbounded at the origin.

Assumption 5.

(i) $\mathbf{f}(\lambda, \theta)$ has elements in $L_1(\Pi)$, continuous at all $(\lambda, \theta) \in \Pi \times \Theta$ with $\lambda \neq 0, |\mathbf{f}_{i,j}(\lambda, \theta)| \leq D\left(\lambda^{-1}, \theta\right) |\lambda|^{-2d(\theta)} \quad \lambda \to 0^+, \quad d(\theta) \in (-1/2, 1/2)$.

(ii) $(\partial/\partial\theta_j)\, \mathbf{f}(\lambda, \theta)$ has elements in $L_1(\Pi)$ continuous at all (λ, θ), $\lambda \neq 0$, and

$$\left|\dot{\mathbf{f}}^{(a,b)}_{(j)}(\lambda, \theta)\right| \leq D\left(\lambda^{-1}, \theta\right) |\lambda|^{-2d(\theta)} \quad \lambda \to 0^+ \quad d(\theta) \in (-1/2, 1/2).$$

(iii) $\mathbf{f}^{-1}(\lambda, \theta)$ has elements in $L_1(\Pi)$, continuous at all $(\lambda, \theta) \in \Pi\times\Theta$, $\left|\mathbf{f}^{-1}_{i,j}(\lambda, \theta)\right| \leq D\left(\lambda^{-1}, \theta\right) |\lambda|^{2d(\theta)} \quad \lambda \to 0^+, \quad d(\theta) \in (0, 1/2)$.

(iv) For all $\eta > 0$, the function

$$\varphi_\eta(\lambda, \theta) \equiv \frac{\mathbf{f}(\lambda, \theta)}{(\det \mathbf{f}(\lambda, \theta) + \eta)}$$

has elements in L_1, continuous at all $(\lambda, \theta) \in \Pi \times \Theta$.

Note that Assumption 5 rules out the possibility of multiple singularities in the process spectrum. The following assumption is required together with Assumption 5(ii) to ensure the uniform strong convergence of the Whittle objective function in the consistency proof.

Assumption 6. $\int_{-\pi}^{\pi} \log \det \mathbf{f}(\lambda, \theta)\, d\lambda$ is twice differentiable in $\theta \in \Theta$ under the integral sign.

For the central limit theorem we introduce the following assumptions.

Assumption 7. (i) $\left(\partial^2/\partial\theta_i \partial\theta'_j\right) \mathbf{f}(\lambda, \theta)$ has elements in $L_1(\Pi)$ continuous at all (λ, θ), $\lambda \neq 0$, and

$$\left|\ddot{\mathbf{f}}^{(a,b)}_{(i,j)}(\lambda, \theta)\right| \leq D\left(\lambda^{-1}, \theta\right) |\lambda|^{-2d(\theta)} \quad \lambda \to 0^+ \quad d(\theta) \in (-1/2, 1/2).$$

(ii) $\left(\partial^3/\partial\theta_i \partial\theta'_j \partial\theta_l\right) \mathbf{f}(\lambda, \theta)$ has elements in $L_1(\Pi)$ continuous at all (λ, θ), $\lambda \neq 0$, and

$$\left|\dddot{\mathbf{f}}^{(a,b)}_{(i,j,l)}(\lambda, \theta)\right| \leq D\left(\lambda^{-1}, \theta\right) |\lambda|^{-2d(\theta)} \quad \lambda \to 0^+ \quad d(\theta) \in (0, 1/2).$$

(iii) $(\partial/\partial\theta_j)\,\mathbf{f}^{-1}(\lambda,\theta)$ has elements in $L_1(\Pi)$ continuous at all $(\lambda,\theta)\in\Pi\times\Theta$, and

$$\left|\frac{\partial}{\partial\theta_j}\mathbf{f}^{-1}_{(a,b)}(\lambda,\theta)\right|\leq D\left(\lambda^{-1},\theta\right)|\lambda|^{2d(\theta)}\quad \lambda\to 0^{+}\quad d(\theta)\in(0,1/2).$$

(iv) $\left(\partial^2/\partial\theta_i\partial\theta_j'\right)\mathbf{f}^{-1}(\lambda,\theta)$ has elements in $L_1(\Pi)$ continuous at all $(\lambda,\theta)\in\Pi\times\Theta$ and

$$\left|\frac{\partial^2}{\partial\theta_i\partial\theta_j'}\mathbf{f}^{-1}(\lambda,\theta)\right|\leq D\left(\lambda^{-1},\theta\right)|\lambda|^{2d(\theta)}\quad \lambda\to 0^{+}\quad d(\theta)\in(0,1/2).$$

Asymptotic normality requires to extend the regularity conditions to the derivatives of the spectral density up to the third order ensuring the necessary degree of smoothness on the Hessian matrix. Zaffaroni (2009) assumes a form of Lipschitz continuity of degree $\alpha\geq\min[1,1-2d(\zeta)]$ away from the zero frequency and an exact decay rate of $D(j,\zeta)\,|\lambda|^{-2d(\zeta)}$ as $\lambda\to 0^{+}$. We prefer slightly stronger and more general assumptions that impose a common bound on the elements of the matrices and uniform continuity away from the zero frequency rather than Lipschitz continuity.

Assumption 8.

(i) The function

$$\mathbf{g}_{(j)}(\lambda,\theta_0)\equiv\mathbf{f}^{-1}(\lambda,\theta_0)\,\dot{\mathbf{f}}_{(j)}(\lambda,\theta_0)\,\mathbf{f}^{-1}(\lambda,\theta_0)\quad\text{for all } j=1,...s.$$

has elements in $L_1(\Pi)$ continuous at all (λ,θ), $\lambda\neq 0$, and $\left|\mathbf{g}^{(b,a)}_{(j)}(\lambda,\theta_0)\right|\leq$ $D\left(\lambda^{-1},\theta\right)|\lambda|^{2d(\theta)}\quad\lambda\to 0^{+}\quad d(\theta)\in(0,1/2)\quad$ for all $j=1,...s$.

(ii) For $1/2<\gamma<1$, for any $\lambda_1$and $\lambda_2\in\Pi$

$$\left|tr\left\{\mathbf{g}_{(j)}(\lambda_1,\theta_0)\,\mathbf{f}(\lambda_1,\theta_0)-\mathbf{f}(\lambda_1-\lambda_2,\theta_0)\right\}\right|\leq K\,|\lambda_2|^{\gamma}\quad\text{for all } j=1,...,n.$$

Assumption 8(i) imposes a bound on the weights of the score vector that has the effect of annulling the singularities of the spectral density at the zero frequency.Assumption 8(ii) gives a condition for the asymptotic unbiasedness of the integrated weighted periodogram: $\int tr\left\{\mathbf{g}_{(j)}(\lambda_1,\theta_0)\,\mathbf{I}_T(\lambda_t)\right\}d\lambda$.

3. Main results

We present the asymptotic results for the Whittle estimators.

Lemma 1. *Under Assumption 1(2), for $-\pi<\lambda<\pi$, the spectral density of $\mathbf{y}_t$ is*

$$\mathbf{f}(\lambda)=\frac{\Sigma_{\xi}(\tau_0)}{2\pi}+\frac{\mathbf{k}(e^{i\lambda},\zeta_0)\Sigma_{\epsilon}(\tau_0)\mathbf{k}(e^{i\lambda},\zeta_0)^{*}}{2\pi}+\Sigma_{\epsilon\xi}(\tau_0)e^{-i\lambda}\mathbf{k}(e^{i\lambda},\zeta_0)^{*}+e^{i\lambda}\mathbf{k}(e^{i\lambda},\zeta_0)\Sigma'_{\epsilon\xi}(\tau_0).$$

Proof. See Appendix A. ■

In what follows, when $\mathbf{f}^{-1}$ is replaced by $\phi_\eta(\lambda,\theta)$ defined in Assumption C, we refer to $Q_T(\theta)$ as $Q_{T,\eta}(\theta)$, and similarly to $\bar{Q}_{T,\eta}(\theta)$ and $Q_\eta(\theta)$.

Theorem 1 *Let Assumptions* $\mathbf{1(2)}, \mathbf{2}, \mathbf{3}, \mathbf{4}, \mathbf{5}, \mathbf{6}$ *hold, then*

$$\lim_{T\to\infty} \hat{\theta} = \lim_{T\to\infty} \bar{\theta}_T = \theta_0 \quad \text{a.s..}$$

Proof. The proofs for $\bar{\theta}_T$ and $\hat{\theta}_T$ are almost the same. We give the proof for $\bar{\theta}_T$. The result follows, adapting Dunsmuir and Hannan (1976), from the a.s. uniform continuity in $\Pi\times\Theta$ of $Q_T(\theta)$ and $Q_{T,\eta}(\theta)$ to respectively $Q(\theta)$ and $Q_\eta(\theta)$ established in Lemma 6, from Lemma 7 and the fact that by definition $Q_T(\bar{\theta}) \le Q_T(\theta)$. ■

For the asymptotic normality we reinforce Assumption 1 ensuring the finiteness of the fourth moments. Our conditions imply that $\mathbf{y}_t$ has trispectrum $\mathbf{f}_{abcd}(\lambda,\omega,v)$, for $\lambda,\omega,v\in(-\pi,\pi]$, given by

$$\mathbf{f}_{abcd}(\lambda,\omega,v) = \frac{1}{(2\pi)^3}\sum_{t_1,t_2,t_3=-\infty}^{\infty} \exp\{-i(\lambda t_1+\omega t_2+v t_3)\}\, cum_{abcd}(t_1,t_2,t_3),$$

where the final term in the summand is the joint cumulant of $\mathbf{y}_0,\mathbf{y}_{t_1},\mathbf{y}_{t_2},\mathbf{y}_{t_3}$, and also that $\mathbf{f}_{abcd}(\lambda,\omega,v)$ is square integrable (see Corollary X).

Theorem 2. *If Assumptions* $\mathbf{1}(4)$, $\mathbf{2}$, $\mathbf{3}$, $\mathbf{4}$, $\mathbf{5}$, $\mathbf{6}$, $\mathbf{7}$, $\mathbf{8}$ *hold, then as* $T\to\infty$, $\sqrt{T}(\bar{\theta}_T-\theta_0)$, $\sqrt{T}(\hat{\theta}_T-\theta_0)$ *converge in distribution to Gaussian random vectors with zero mean and variance covariance matrix equal to* $H^{-1}(\theta_0)\,V(\theta_0)H^{-1}(\theta_0)$*, where*

$$\mathbf{H}(\theta_0) = \frac{1}{2\pi}\int_{-\pi}^{\pi} tr\left\{\mathbf{f}^{-1}(\lambda,\theta_0)\,\dot{\mathbf{f}}_{(i)}(\lambda,\theta_0)\,\mathbf{f}^{-1}(\lambda,\theta_0)\,\dot{\mathbf{f}}_{(j)}(\lambda,\theta_0)\right\}d\lambda,$$

and the matrix $\mathbf{V}(\theta_0)$ *has elements*

$$\mathbf{V}_{(i,j)} = \frac{1}{\pi}\int_{-\pi}^{\pi}\mathbf{tr}\left\{\mathbf{f}(\lambda,\theta)\frac{\partial}{\partial\theta_j}\mathbf{f}^{-1}(\lambda,\theta_0)\mathbf{f}(\lambda,\theta)\frac{\partial}{\partial\theta_l}\mathbf{f}^{-1}(\lambda,\theta_0)\right\}d\lambda$$

$$+\frac{1}{2\pi}\,{}^{T}_{r,t,u,v=1}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\left\{\frac{\partial}{\partial\theta_j}\mathbf{f}_{rq}(v,\theta_0)\frac{\partial}{\partial\theta_l}\mathbf{f}_{uv}(\omega,\theta_0)\right\}\mathbf{f}_{rtuv}(-v,\omega,-\omega)dvd\omega.$$

Proof. We give the proof for $\bar{\theta}_T$. Let

$$\mathbf{S}_T(\theta) = (\partial/\partial\theta)\,Q_T(\theta),$$

denote the score vector. Its jth element is

$$\mathbf{S}_T^{(j)}(\theta) = \frac{1}{2\pi}\int_{-\pi}^{\pi} tr\left\{\mathbf{g}^{(j)}(\lambda,\theta)\left[\mathbf{I}_T(\lambda)-\mathbf{f}(\lambda,\theta)\right]\right\}d\lambda,$$

where

$$\mathbf{g}^{(j)}(\lambda,\theta) = \mathbf{f}^{-1}(\lambda,\theta)\,\dot{\mathbf{f}}^{(j)}(\lambda,\theta)\,\mathbf{f}^{-1}(\lambda,\theta).$$

By the mean value theorem,

$$0 = \mathbf{S}_T(\theta) = \mathbf{S}_T(\theta_0) + \mathbf{H}_T(\dot{\theta})\left(\bar{\theta} - \theta_0\right), \tag{8}$$

where $\mathbf{H}_T(\theta) = \left(\partial^2/\partial\theta_i\partial\theta_j\right) Q_T(\theta)$ is evaluated at a $\theta = \dot{\theta}$ such that $\left\|\dot{\theta}_T - \theta_0\right\| \leq \left\|\bar{\theta}_T - \theta_0\right\|$, where we define $\|\mathbf{A}\| = \{tr(\mathbf{A}'\mathbf{A})\}^{1/2}$ for any matrix $\mathbf{A}$. By Lemma 2 and Theorem 1, $\mathbf{H}_T\left(\dot{\theta}\right) \overset{a.s}{\to} \mathbf{H}(\theta_0)$. Finally by Lemma 3, $E(\mathbf{S}_T(\theta_0)) \to 0$, and by Lemma 4 $\sqrt{T}\left(\mathbf{S}_T(\theta_0) - E(\mathbf{S}_T(\theta_0))\right) \overset{d}{\to} \mathbf{N}(0, \mathbf{V}(\theta_0))$, whence the proof is completed in standard fashion. ■

The proof of Lemma 4 is of considerable lenght as a result of a central limit theorem that we establish for quadratic forms of signal plus noise processes, thus we defer it to the next section. Meanwhile we discuss implications of Theorems 1 and 2.

Remark 3.1 The proof of Theorem 1 follows as in Theorem 4 of Dunsmuir and Hannan (1976) with some differences. Dunsmuir and Hannan consider a stationary vector linear process $\mathbf{y}_t = \sum \mathbf{A}_j(\theta)\, e_{t-j}$ with short memory $\mathbf{A}_j(\theta)$, satisfying $\int_{-\pi}^{\pi} \log\det \mathbf{f}(\lambda,\theta) d\lambda = 0$ at all θ, so that the error variance matrix, functionally independent from θ, is the variance of the best linear predictor of the process. They prove the consistency of the Whittle estimator minimizing a slightly different objective function from ours $L(\theta) = \log\det \Sigma^e(\phi) + \frac{1}{2\pi}\int_{-\pi}^{\pi} tr\left\{\mathbf{f}^{-1}(\lambda,\theta)\mathbf{I}_T(\lambda)\right\} d\lambda$. Finding that there are cases where the innovations variance depends upon θ, Hosoya (1974) and Hosoya and Taniguchi (1982, 1997) propose the objective function $Q(\theta) = \log\det \mathbf{f}(\lambda,\theta) + \frac{1}{2\pi}\int_{-\pi}^{\pi} tr\left\{\mathbf{f}^{-1}(\lambda,\theta)\mathbf{I}_T(\lambda)\right\} d\lambda$, but they do not offer an explicit proof of the consistency. Robinson (1978) considers a class of univariate processes with spectral density non easily factored and establishes the strong consistency of the Whittle estimator minimizing $\bar{\mathbf{Q}}_T(\theta)$ under regularity conditions that rule out long memory.

Remark 3.2 The vector of mean parameters ω_0 cannot be identified by the Whittle function since its elements enter linearly in $\log x_{it}^2$ and are lost when computing the empirical autocovariances of the process. However it can be estimated using the sample mean of the vector $\mathbf{y}_T$. Since $\hat{\mathbf{y}}_T = 1/T\sum_{t=1}^{T} \mathbf{y}_t$ is a $\sqrt{T}$-consistent estimate of $E\mathbf{y}_t = \omega_0 + E\xi_t$ under Assumptions 1 to 6, we obtain a $\sqrt{T}$-consistent estimate of ω_0 subtracting the Whittle estimate of $E\xi_t$ from $\hat{\mathbf{y}}_T$.

Remark 3.3 The only available results on the asymptotics of the Whittle estimator for signal plus noise processes are for an autoregressive signal and indipendent signal and noise (see Hosoya and Taniguchi (1982) and Dunsmuir (1979)). The signal h_t is generated by a finite autoregressive process $\sum_{j=1}^{q} b_j h_{t-j} = \eta_t$, where the η_t have zero mean, finite variance and are independent from the noise ξ_s for every t and s; the spectral density function of the process

$$f_\theta(\lambda) = \frac{1}{2\pi} \frac{\sigma_\eta^2}{\left|\sum_{j=1}^q b_j e^{i\lambda j}\right|} + \frac{\sigma_\xi^2}{2\pi},$$

is factorable and representable as

$$f_\vartheta(\lambda) = \frac{\sigma^2}{2\pi} \left| \frac{\sum_{j=1}^q \psi_j e^{i\lambda j}}{\sum_{j=1}^q b_j e^{i\lambda j}} \right|^2,$$

the parameter ϑ is a closed form function of $\theta' = \left(b_1, ..., b_q, \sigma_\eta^2, \sigma_\xi^2\right)'$, since σ^2 and the ψ_j can be expressed as closed form functions of θ, and the b_js are unchanged. In this restrictive case the limiting distribution of the estimator follows from standard results for linear processes.

Remark 3.4 Hourvich, Moliners and Souliers (2006) discuss semiparametric estimation of the memory parameter in the univariate equivalent of model (1)-(4). They show that the local Whittle estimator is consistent for $d \in (0, 1)$ and asymptotically normal for $d \in (0, 3/4)$, and essentially recovers the optimal semiparametric rate of convergence for this problem. In particular if the spectral density of the short memory component of the signal is sufficiently smooth, they obtain a convergence rate of $n^{2/5-\partial}$ for $d \in (0, 3/4)$, where n is the sample size and $\partial > 0$ is arbitrarily small.

Remark 3.5 The Gaussian frequency domain estimator has been widely used in estimation of long-range dependent models thanks to its technical properties. The Whittle function naturally takes into account the asymptotic behavior of the autocovariances as the sample size goes to infinity, so it is very sensitive to the degree of dependence of the process in second-order sense. Moreover, by construction, it automatically compensates for the possible lack of square integrability of the model spectral density that occurs when the memory parameter is between 1/2 and 1/4. This implies that the estimator has a rate of convergence and an asymptotic distribution that do not depend on whether long-memory holds or not.

Remark 3.6 Empirical evidences suggest that the assumption $d \in (-1/2, 1/2)$ might be over restrictive. Generalization of model (1)- (4) to non stationarity is desirable to take into account the strong degree of persistence found in many financial dateset. If $d \geq 1/2$ the square summability of the coefficients in (3) does not hold and the limiting distribution of Theorem 2 is no longer valid. Zaffaroni (2009) suggests to differentiate $\mathbf{y}_t$ to achieve stationarity: differentation implies that the Whittle estimator is still strongly consistent and asymptotically Normal, however it likely brings an efficiency loss. Semiparametric estimation of non stationary time series, based on the log periodogram and local Whittle estimators has been exhaustively examined (Hourvich, Moliners and Soulier 2006, Abadir, Distaso and Giraitis, 2006). In the context of parametric estimation of linear time series models, Velasco and Robinson (2000) suggest a tapered version of Hannan's (1973) discrete Whittle estimator and establish its consistency for $d < 1$ and its asymptotic Normality for $d < 3/4$, finding the same asymptotic variance formula as for stationary series. As in Hannan, they assume separate parameterization, i.e. $\int_{-\pi}^{\pi} \log f(\lambda, \theta) d\lambda = 0$ at all θ, where $f(\lambda, \theta)$ is the pseudo spectral density; their results cover standard parameterisations of fractional ARIMA and FEXP models. One could consider extending data tapering to non separately parametrized linear processes and account for non stationarity arising in signal plus noise processes.

Remark 3.7 As noted in Hosoya and Taniguchi (1982), the non separate parameterisation implies that trispectrum of the process appears in the asymptotic covariance matrix. For practical use of the asymptotic results, a consistent estimator of the asymptotic covariance matrix is required. As suggested in Zaffaroni (2009) for $H(\theta)$ such estimate can be obtained by substituting $\hat{\theta}$ into $H_T(\theta)$. For $V(\theta)$ one can conjecture that the estimates provided by Hosoya and Taniguchi (1982, Section 5) will be consistent under the assumption of short range dependence.

4. Asymptotic Normality of the integrated weighted periodogram.

The limiting distribution of the Whittle estimator is obtained via that of any linear combination of the integrated weighted periodograms:

$$\frac{\sqrt{T}}{2\pi} \int_{-\pi}^{\pi} tr \left\{ \mathbf{g}^{(j)}(\lambda, \theta_0) \left(\mathbf{I}_T(\lambda) - \mathbf{E}\mathbf{I}_T(\lambda) \right) \right\} d\lambda \tag{9}$$

$$= \frac{\sqrt{T}}{2\pi} \sum_{a,b=1}^{N} \left(\int_{-\pi}^{\pi} \mathbf{g}^{(j)}_{(b,a)}(\lambda, \theta_0) \left(\mathbf{I}_{(a,b)}(\lambda) - \mathbf{E}\mathbf{I}_{(a,b)}(\lambda) \right) \right) d\lambda \quad j = 1, ..., s.$$

Denote by

$$\mathbf{h}^{(j)}_{(b,a)}(u) = \int_{-\pi}^{\pi} \mathbf{g}^{(j)}_{(b,a)}(\lambda, \theta_0) e^{iu\lambda} d\lambda \quad u \in Z$$

the Fourier coefficient of the (b,a)element of the weights matrix $\mathbf{g}^{(j)}(\lambda, \theta_0)$, then (9) is equivalent to $T^{-1/2} Q_n$, where

$$Q_n = \frac{1}{2\pi} \sum_{a,b=1}^{N} \left(\sum_{t=1}^{T} \sum_{s=1}^{T} \mathbf{h}^{(j)}_{(b,a)}(|t-s|) \left(\mathbf{X}^{(a)}_t \mathbf{X}^{(b)}_s - E\left(\mathbf{X}^{(a)}_t \mathbf{X}^{(b)}_s \right) \right) \right) \quad j = 1, ..., s,$$

where $\mathbf{X}_t = \mathbf{y}_t - \mu$. From (5), $\mathbf{X}_t = \sum_{i=1}^{\infty} \Psi_i \epsilon_{t-1-i} + \xi_t = \mathbf{z}_t + \xi_t$, writing $\mathbf{z}_t =$

$\sum_{i=1}^{\infty} \Psi_i \epsilon_{t-1-i}$. Put $Q_n = Q_n^{(1)} + Q_n^{(2)} + Q_n^{(3)}$, where

$$Q_n^{(1)} = \frac{1}{(2\pi)^2} \sum_{a,b=1}^{N} \left(\sum_{t=1}^{T} \sum_{s=1}^{T} \mathbf{h}_{(b,a)}^{(j)} \left(|t-s| \right) \left(\xi_t^{(a)} \xi_s^{(b)} - E\left(\xi_t^{(a)} \xi_s^{(b)} \right) \right) \right),$$

$$Q_n^{(2)} = \frac{1}{(2\pi)^2} \sum_{a,b=1}^{N} \left(\sum_{t=1}^{T} \sum_{s=1}^{T} \mathbf{h}_{(b,a)}^{(j)} \left(|t-s| \right) \left(z_t^{(a)} \xi_s^{(b)} - E\left(\mathbf{z}_t^{(a)} \xi_s^{(b)} \right) \right) \right),$$

$$Q_n^{(3)} = \frac{1}{(2\pi)^2} \sum_{a,b=1}^{N} \left(\sum_{t=1}^{T} \sum_{s=1}^{T} \mathbf{h}_{(b,a)}^{(j)} \left(|t-s| \right) \left(\mathbf{z}_t^{(a)} \mathbf{z}_s^{(b)} - E\left(\mathbf{z}_t^{(a)} \mathbf{z}_s^{(b)} \right) \right) \right).$$

The proof of Lemma 4 now follows immediately from Lemma 5 and Lemma 6 which appear subsequently.

Lemma 5. *Let Assumptions 1-8 hold. Then* as $T \to \infty$, *the vectors*

$$\begin{array}{rlcl} \text{(i)} & T^{-1/2}(Q_n^{(1)} - EQ_n^{(1)}) & j = & 1,..,n \\ \text{(ii)} & T^{-1/2}(Q_n^{(2)} - EQ_n^{(2)}) & j = & 1,..,n \\ \text{(iii)} & T^{-1/2}(Q_n^{(3)} - EQ_n^{(3)}) & j = & 1,..,n \end{array}$$

have a joint Normal distributionwith zero mean vector and covariance matrix whose (j,L) element is

$$\frac{1}{\pi} \int_{-\pi}^{\pi} \mathbf{tr} \left\{ \mathbf{g}^{(j)}(\lambda, \theta_0) \mathbf{f}_{\xi}(\lambda, \theta_0) \mathbf{g}^{(l)}(\lambda, \theta_0) \mathbf{f}_{\xi}(\lambda, \theta_0) \right\} d\lambda$$

$$+ \frac{1}{2\pi} \sum_{r,t,u,v=1}^{T} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \left\{ \mathbf{g}_{rq}^{(j)}(\lambda_1, \theta_0) \mathbf{g}_{uv}^{(l)}(\lambda_2, \theta_0) \right\} K_{rtuv}^{\xi}(-\lambda_1, \lambda_2, -\lambda_2) d\lambda_1 d\lambda_2.$$

Proof. The proof of (iii) and (ii) follows from a straightforward application of Theorem 7.3 of Giraitis and Taqqu (1999, page 14). Giraitis and Taqqu derive the joint asymptotic normality of quadratic form of multivariate Appell polynomials for linear sequences with *i.i.d* innovations and possibly different weights and linear coefficients. The results follow taking in their notation a multivariate Appell polynomial, $P_{m,n}\left(X_t^{(i,1)}, X_s^{(i,2)} \right)$ of degree equal to one.

(i) is a quadratic form in iid variates with zero mean and constant variance and the $\mathbf{h}_{(b,a)}^{(j)}(|t-s|)$ are entries of a real symmetric matrix with non vanishing diagonal . Let

$$T_n = \sum_{t=1}^{T} \sum_{s=1}^{T} \mathbf{h}_{(b,a)}^{(j)} \left(|t-s| \right) \xi_t^{(a)} \xi_s^{(b)}$$

For each $j = 1, ..., s$ and for $a, b = 1, ..., k$ the quantities

$$(Var(T_n))^{-1/2}(T_n - ET_n) \to N(0, 1)$$

converge to a Gaussian random variable by Theorem 2.2 of Bhansali Giraitis Kokoszka (2007) on verifying that the entry of the Toepliz matrix $\mathbf{h}_{(b,a)}^{(j)}(|t-s|)$ where we suppress the dependence on n for notation simplicity

Lemma 5.*Let Assumptions 1-8 hold. Then*

$$T^{-1/2}(Q_n^{(1)} - EQ_n^{(1)})$$

have, asymptotically, a joint Normal distribution with zero mean vector and covariance matrix whose (j,l) element is

$$\frac{1}{\pi}\int_{-\pi}^{\pi} \mathbf{tr}\left\{\mathbf{g}^{(j)}(\lambda,\theta_0)\mathbf{f}_{\xi}(\lambda,\theta_0)\mathbf{g}^{(l)}(\lambda,\theta_0)\mathbf{f}_{\xi}(\lambda,\theta_0)\right\} d\lambda$$

$$+\frac{1}{2\pi}\sum_{r,t,u,v=1}^{T}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\left\{\mathbf{g}_{rq}^{(j)}(\lambda_1,\theta_0)\mathbf{g}_{uv}^{(l)}(\lambda_2,\theta_0)\right\} K_{rtuv}^{\xi}(-\lambda_1,\lambda_2,-\lambda_2)d\lambda_1 d\lambda_2.$$

All that remains is to evaluate is their asymptotic covariances. **Lemma 6**.*Let Assumptions 1-8 hold. Then*

$$\lim_{T\to\infty} Cov\left(\frac{\sqrt{T}}{2\pi}\int_{-\pi}^{\pi} tr\left\{\mathbf{g}_{(j)}(\lambda,\theta_0)\left[\mathbf{I}_T(\lambda) - \mathbf{EI}_T(\lambda)\right]\right\}d\lambda, \frac{\sqrt{T}}{2\pi}\int_{-\pi}^{\pi} tr\left\{\mathbf{g}_{(l)}(\lambda,\theta_0)\left[\mathbf{I}_T(\lambda) - \mathbf{EI}_T(\lambda)\right]\right\}d\lambda\right)$$

$$\begin{aligned}
&= 2\pi\int_{-\pi}^{\pi} \mathbf{g}_{(j)}^{(b,a)}(\lambda,\theta_0)\,\bar{\mathbf{g}}_{(l)}^{(d,c)}(\lambda,\theta_0)\,\mathbf{f}^{(a,b)}(\lambda,\theta_0)\,\bar{\mathbf{f}}^{(c,d)}(\lambda,\theta_0)\,d\lambda \\
&\quad +2\pi\int_{-\pi}^{\pi} \mathbf{g}_{(j)}^{(b,a)}(\lambda,\theta_0)\,\bar{\mathbf{g}}_{(l)}^{(d,c)}(-\lambda,\theta_0)\,\mathbf{f}^{(a,d)}(\lambda,\theta_0)\,\bar{\mathbf{f}}^{(b,c)}(\lambda,\theta_0)\,d\lambda \\
&\quad +2\pi\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \mathbf{g}_{(j)}(\lambda_1,\theta_0)\,\mathbf{g}_{(l)}(\lambda_2,\theta_0)\,Cum_{abcd}(\lambda_1,\lambda_2,-\lambda_2)\,d\lambda_1 d\lambda_2.
\end{aligned}$$

Proof. See appendix.

Remark 4.1 Statistical literature on the Whittle estimator has establshed the asymptotic normality of the integrated weighted periodogram for linear processes under a variety of assumptions. In the short memory case, the results follows from the asymptotic normality of linear combinations of the sample serial covariances

$$\sqrt{T}\left(\mathbf{I}_{(a,b)}(\lambda) - \mathbf{EI}_{(a,b)}(\lambda)\right) \tag{10}$$

$$= \sqrt{T}\left(\frac{1}{T}\sum_{t=1}^{T-m}\mathbf{y}_t^{(a)}\mathbf{y}_{t+m}^{(b)} - \Gamma_m^{(a,b)}\right). \tag{11}$$

For an ergodic and strictly stationary process, Hannan (1976) assumes that the innovations satisfy almost surely $(i)\, E\{e_t|F_{t-1}\} = 0, (ii)\, E\left\{e_t^{(a)}e_t^{(b)}|F_{t-1}\right\} =$

$\delta_{(a,b)}, (iii)\, E\left\{e_t^{(a)} e_t^{(b)} e_t^{(c)} | F_{t-1}\right\} = \delta_{(a,b,c)}, (iv)\, E\left(e_t^{(a)} e_t^{(b)} e_t^{(c)} e_t^{(d)}\right) < \infty$, and shows that the diagonal elements of $\mathbf{f}(\lambda, \theta)$ being square integrable is a necessary and sufficient condition for the convergence of (10). Hosoya and Taniguchi (1982, Theorem 2.2) derive an analogous results replacing strict stationarity with second order stationarity, ergodicity with a Lindeberg-type condition and the strongly mixing conditions on the innovations with more natural ones with respect to the second and fourth conditional moments. Robinson (1978) directly assumes asymptotic Normality of the sample covariance for a fourth order stationary process. Robinson (1978), Dunsmuir (1979), Hosoya and Taniguchi (1982) rely on the convergence of (10) to show the convergence of (9), approximating (9) by

$$\frac{\sqrt{T}}{2\pi} \int_{-\pi}^{\pi} tr \left\{ \mathbf{g}_M^{(j)}(\lambda, \theta_0) \left(\mathbf{I}_T(\lambda) - \mathbf{E}\mathbf{I}_T(\lambda) \right) \right\} d\lambda$$

where $\mathbf{g}_M^{(j)}(\lambda, \theta_0)$ is the Cesaro sum of the weight function to a finite number of terms, M. The approximation relies on the assumption that the spectral density satisfies a Lipschitz continuity of order $\zeta > 1\backslash 2$.In the long memory case, the model spectrum might not be Lipschitz continuous of the requested degree nor square integrable. The asymptotic distribution of the integrated weighted periodogram is derived via approximation by another quadratic form which shares the same asymptotic distribution but has shorter memory. The main idea of the approximation is to impose conditions on the weight function that have the effect of annihilating the singularities of the spectral density in the frequency domain. To establish the validity of the approximation Fox and Taqqu (1987) rely on Gaussianity in an essential way, employing the exact expression for the cumulants of a quadratic form in Gaussian variates. Giraitis and Surgailis (1990), Giraitis and Taqqu (1999) and Hosoya (1997) relax the Guassianity assumption, and exploit the factorization of the spectrum. However because the spectral density is not easily factored in correlated signal plus noise processes, we cannot exploit its factorization to establish a shorter memory approximation of the score vector. Instead we extend Zaffaroni (2003) and truncate the original process at some finite $t = N$ and establish the validity of the approximation relying on certain results on the asymptotic behavior of the trace of Toeplitz matrices (Theorem 1, Fox and Taqqu, 1987).

5. Monte-Carlo simulations and efficiency comparison with MLE.

The Whittle estimator is bound to be more inefficient than the maximum likelihood estimator. For our purposes it is of great interest to know the extent of this efficiency loss. This section presents the results of a series of Monte Carlo experiments to compare the finite sample behavior of both methods. We consider the performance of the estimators in short and long memory bivariate one-shock specifications. To carry on the experiments we must choose a finite parametric specification of the coefficients and the joint distribution of the mean equation shocks. We set:

$$\begin{aligned}
\mathbf{x}_t &= \mathbf{D}_t \mathbf{z}_t, \\
\mathbf{D}_t &= diag\left\{\exp(0.5\mathbf{h}_{1t}), \exp(0.5\mathbf{h}_{2t})\right\} \\
\mathbf{h}_t &= \begin{pmatrix} \omega_1 \\ \omega_2 \end{pmatrix} + \begin{pmatrix} \frac{a_1(L)}{b_1(L)}(1-L)^{-d_1} & 0 \\ 0 & \frac{a_2(L)}{b2(L)}(1-L)^{-d_2} \end{pmatrix} \mathbf{g}(\mathbf{z}_{t-1}), G(12)
\end{aligned}$$

where $a_i\,(L)$ and $b_i\,(L)$ are univariate polynomials in the lag operator of known degree

$$\begin{aligned}
a_i\,(L) &= 1 + \sum_{k=1}^{p} a_{ik} L^k \quad a_i\,(z_i) \neq 0, \;\; |z_i| \leq 1 \\
b_i\,(L) &= 1 - \sum_{j=1}^{q} b_{ij} L^j \quad b_i\,(z_i) \neq 0, \;\; |z_i| \leq 1,
\end{aligned}$$

that for each $i = 1, 2$ have no common zeros, and $(1-L)^{-d}$ is the univariate fractional operator which has binomial expansion

$$(1-L)^{-d_i} = \sum_{k=0}^{\infty} \Gamma\,(k-d_i)\,\Gamma\,(k+1)^{-1}\,\Gamma\,(-d_i)^{-1}\,L^k,$$

where Γ is the gamma function. The function $\mathbf{g}(.)$ follows Nelson (1991) original specification, i.e. for each assets we set

$$\mathbf{g}_i(z_{it}) = \vartheta_i + \delta_i\left(|z|_{it} - \mu_{|z_i|}\right)$$

for constant parameters ϑ, δ with $\vartheta\delta \neq 0$, where $\mu_{|z_i|} = E\,|z_i|$. When $d_i = 0$, the volatility of asset i follows an EGARCH specification, for non zero values of $d_i < 1/2$, a FIEGARCH one. We presents results for a bivariate models with FIEGARCH(1,d,1) specifications for both assets. In all the experiments we set $\omega = 0$. With respect to the distribution of the innovations, we consider a Normal(0,1) distribution, a Student t with $v = 7$ degrees of freedom and a GED distribution with thickness parameter $v = 1$. The vector of parameters to be estimated is $\theta = (a_1, a_2, b_1, b_2, \vartheta_1, \vartheta_2, \delta_1, \delta_2, v, d_1, d_2)'$, where ϑ_i and δ_i are the parameters of the news impact curve of asset i, v is the common tail thickness parameter and d_i is the memory parameter of asset i. We simulate samples of length $T = \{500, 1500, 2500\}$ with 1000 Monte Carlo iterations. We report the bias and the root mean squared error of the estimates across the 1000 replicates. All the computations are carried out in Matlab and codes are available upon request from the authors.Table 1.1 reports the reuslts for the MLE estimator and Table 1.2 reports the results for the Whittle. Each table has three panels corresponding to the different innovations. The simulation results suggest that for long memory specifications the Whittle estimator of the MEV-FIEGARCH model performs comparably with respect to the MLE in terms of estimates bias. The MLE estimator is in general superior in terms of root mean squared error but it is over performed by the Whittle with respect to the long memory and and the asymmetry parameters.

Table 1.1.: Monte Carlo results from MLE estimation

T=		500		1500		3000	
parameter	value	mean	rmse	mean	rmse	mean	rmse
		MEV-FIEGARCH(1,d,1)-N(0,1)					
a_1	0.5	0.496	0.367	0.511	0.225	0.508	0.194
b_1	0.5	0.508	0.341	0.501	0.300	0.500	0.217
ϑ_1	0.3	0.309	0.057	0.304	0.042	0.301	0.036
δ_1	0.5	0.572	0.295	0.550	0.250	0.535	0.157
d_1	0.25	0.165	0.194	0.140	0.158	0.163	0.126
a_2	0.7	0.639	0.473	0.732	0.371	0.703	0.273
b_2	0.6	0.503	0.453	0.524	0.298	0.603	0.223
ϑ_2	-0.1	-0.18	0.051	-0.15	0.043	-0.013	0.040
δ_2	0.15	0.009	0.086	0.011	0.074	0.017	0.040
d_2	0.45	0.355	0.196	0.399	0.094	0.406	0.071
ρ	0.6	0.562	0.456	0.630	0.341	0.610	0.176
		MEV-FIEGARCH(1,d,1)-GED					
a_1	0.5	0.482	0.404	0.519	0.371	0.510	0.226
b_1	0.5	0.506	0.226	0.498	0.178	0.500	0.135
ϑ_1	0.3	0.243	0.266	0.252	0.227	0.244	0.192
δ_1	0.5	0.465	0.259	0.478	0.187	0.480	0.130
d_1	0.25	0.133	0.202	0.098	0.193	0.050	0.142
a_2	0.7	0.732	0.363	0.711	0.341	0.701	0.269
b_2	0.6	0.578	0.285	0.552	0.251	0.060	0.202
ϑ_2	-0.1	-0.143	0.231	-0.132	0.186	-0.115	0.150
δ_2	0.15	0.058	0.293	0.095	0.207	0.120	0.133
d_2	0.45	0.363	0.178	0.407	0.123	0.437	0.070
v	1	1.103	0.296	1.047	0.133	1.021	0.085
ρ	0.6	0.571	0.223	0.615	0.187	0.605	0.143
		MEV-FIEGARCH(1,d,1)-t_7					
a_1	0.5	0.587	0.304	0.552	0.222	0.501	0.157
b_1	0.5	0.445	0.300	0.468	0.176	0.501	0.115
ϑ_1	0.3	0.355	0.063	0.305	0.045	0.302	0.035
δ_1	0.5	0.495	0.086	0.499	0.059	0.499	0.046
d_1	0.25	0.154	0.168	0.184	0.121	0.205	0.070
a_2	0.7	0.708	0.256	0.705	0.156	0.700	0.114
b_2	0.6	0.553	0.331	0.582	0.243	0.593	0.116
ϑ_2	-0.1	-0.133	0.198	-0.120	0.152	-0.115	0.096
δ_2	0.15	0.184	0.106	0.166	0.098	0.153	0.095
d_2	0.45	0.371	0.177	0.417	0.110	0.443	0.059
v	7	7.301	2.970	7.367	2.759	7.306	2.488
ρ	0.6	0.553	0.205	0.621	0.146	0.603	0.086

Table 1.2.: Monte Carlo results from Whittle estimation

T=		500		1500		3000	
parameter	value	mean	rmse	mean	rmse	mean	rmse
		MEV-FIEGARCH(1,d,1)-N(0,1)					
a_1	0.5	0.598	0.364	0.562	0.306	0.515	0.287
b_1	0.5	0.435	0.323	0.442	0.300	0.468	0.211
ϑ_1	0.3	0.309	0.116	0.304	0.095	0.301	0.057
δ_1	0.5	0.541	0.365	0.462	0.291	0.487	0.195
d_1	0.25	0.22	0.133	0.18	0.101	0.22	0.078
a_2	0.7	0.763	0.296	0.752	0.198	0.715	0.144
b_2	0.6	0.556	0.321	0.582	0.253	0.599	0.156
ϑ_2	-0.1	-0.105	0.135	-0.102	0.095	-0.101	0.048
δ_2	0.15	0.180	0.164	0.162	0.098	0.152	0.073
d_2	0.45	0.415	0.125	0.457	0.118	0.45	0.065
ρ	0.6	0.632	0.343	0.611	0.257	0.608	0.186
		MEV-FIEGARCH(1,d,1)-GED					
a_1	0.5	0.482	0.404	0.519	0.371	0.532	0.254
b_1	0.5	0.536	0.356	0.476	0.253	0.556	0.202
ϑ_1	0.3	0.302	0.193	0.304	0.115	0.299	0.068
δ_1	0.5	0.435	0.353	0.450	0.201	0.495	0.194
d_1	0.25	0.198	0.198	0.263	0.086	0.026	0.058
a_2	0.7	0.682	0.353	0.658	0.296	0.695	0.185
b_2	0.6	0.555	0.291	0.568	0.263	0.583	0.253
ϑ_2	-0.1	-0.121	0.176	-0.111	0.134	-0.100	0.125
δ_2	0.15	0.159	0.227	0.151	0.188	0.158	0.141
d_2	0.45	0.41	0.182	0.49	0.132	0.448	0.085
v	1	1.103	0.296	1.047	0.133	1.0132	0.093
ρ	0.6	0.571	0.223	0.615	0.187	0.605	0.143
		MEV-FIEGARCH(1,d,1)-t_7					
a_1	0.5	0.482	0.404	0.519	0.371	0.532	0.254
b_1	0.5	0.536	0.356	0.476	0.253	0.556	0.202
ϑ_1	0.3	0.302	0.193	0.304	0.115	0.299	0.068
δ_1	0.5	0.435	0.353	0.450	0.201	0.495	0.194
d_1	0.25	0.198	0.198	0.263	0.086	0.026	0.058
a_2	0.7	0.682	0.353	0.658	0.296	0.695	0.185
b_2	0.6	0.555	0.291	0.568	0.263	0.583	0.253
ϑ_2	-0.1	-0.121	0.176	-0.111	0.134	-0.100	0.125
δ_2	0.15	0.159	0.227	0.151	0.188	0.158	0.141
d_2	0.45	0.41	0.182	0.49	0.132	0.448	0.085
v	7	7.245	0.332	7.168	0.256	7.115	0.157
ρ	0.6	0.571	0.223	0.615	0.187	0.605	0.143

6. Empirical application

In this section we assess the performance of the model to fit time series of observed returns. We fit a trivariate MEV to daily returns from three market indices: FTSE 100 (UK), S&P 500 (US) and NIkkei 225 (JP) (source: Datastream). The data cover the period from 01-January-1984 to 31-August-2015 for a total of 8246 observations. The daily return is computed as the log difference of the daily closing price. For all the series, we notice a marked difference between the ACF in the levels and in the squares, the latter exhibiting an approximate hyperbolic behavior as the lag increases.

6.1. Specification

For this data set we fit a GARCH-MEV specification, choosing a one-shock specification of the model and finitely parameterizing the Ψ_j

$$\begin{aligned} \mathbf{x}_t &= \mu + \eta_t, \\ \eta_t &= diag\left\{\exp\left\{0.5h_{1t}\right\}, \exp\left\{0.5h_{2t}\right\}, \exp\left\{0.5h_{2t}\right\}\right\}\mathbf{z}_t \end{aligned}$$

with

$$h_{it} = \omega_i + \frac{a_i(L)}{b_i(L)}(1-L)^{-d} g_i(z_{it-1}),$$

where $a_i(L)$ and $b_i(L)$ are univariate polynomials in the lag operator of known degree

$$\begin{aligned} a_i(L) &= 1 + \sum_{k=1}^{p} a_{ik}L^k \quad a(z_i) \neq 0, \;\; |z_i| \leq 1 \\ b_i(L) &= 1 - \sum_{j=1}^{q} b_{ij}L_j \quad b(z_i) \neq 0, \;\; |z_i| \leq 1, \end{aligned}$$

that for each $i = 1, 2, 3$, $a_i(L)$ and $b_i(L)$ have no common zeros . The function $g_i(z_{it-1})$ follows the EGARCH specification of Nelson (1991)

$$g_i(z_{it}) = \theta_i + \delta_i\left(|z|_{it} - \mu_{|z_i|}\right)$$

allowing for asymmetries and leverage effects and $(1-L)^{-d}$ is the univariate fractional operator which has binomial expansion[2]

$$(1-L)^{-d} = \sum_{k=0}^{\infty} \Gamma(k-d)\Gamma(k+1)^{-1}\Gamma(-d)^{-1}L^k,$$

where Γ is the gamma function. For all the three series we set $p = q = 1$ and for the mean specification we simply set

$$\eta_t = \mathbf{x}_t - \mu,$$

keeping it as simple as possible for parsimony reasons.

[2]We follow Baillie et al (1996) and truncate k at $k = 1000$

6.2. Estimation results

As a preliminary exercise we estimate semiparametrically for each series the long-memory parameter d using the local Whittle estimator of Robinson (1995). The results in Table 2.1 clearly support the hypothesis of long-memory for all the series.

TABLE 2.1.

	S&P	**FTSE**	**NIKKEI**
Semiparametric	0.12	0.15	0.17
GARCH-MEV	0.33 (0.007)	0.43 (0.01)	0.26 (0.08)

Local Whittle and MEV-Whittle estimates for the memory paramer d

Next we estimate the MEV-CCC model with the Whittle estimator using analytic scores in the numerical maximization, imposing invertibility conditions on the $a_i(L)$ and $b_i(L)$, and $|d| < 1\backslash 2$. The second row of Table 2.1 reports the Whittle estimates of d (asymptotic standard errors are based on Taniguchi (1982)). Estimation results are reported in Table 2.2. For all the series of returns there is clear evidence that shock to volatility decay with time but very slowly, agreement with previous studies. Moreover the conditional correlations appear to be significantly high for the period under study.

TABLE 2.2

	S&P	**FTSE**	**NIKKEI**
μ	0.0001 (0.0005)	0.0003 (0.0002)	0.0001 (0.0003)
a_i	0.51 (0.007)	0.45 (0.042)	0.63 (0.003)
b_i	-0.42 (0.004)	0.40 (0.0072)	-0.51 (0.004)
$\|\theta\|$	0.1567 (0.009)	0.0981 (0.001)	0.1755 (0.007)
δ	0.891 (0.0003)	0.675 (0.006)	0.904 (0.0012)
$\rho_{s\&p}$	1.00 $(-)$	-	-
ρ_{ftse}	0.765 (0.0013)	1.00 $(-)$	-
ρ_{nikkei}	0.854 (0.017)	0.854 (0.017)	1.00 $(-)$

Whittle estimates of the EGARCH-MEV model parameters

7. Conclusions

We have established the asymptotic distribution theory of the Whittle estimator in a class of multivariate exponential volatility models that nests both one shock and two-shocks models under a variety of parameterizations including short and long memory.The most notable elements of this class are the CCC-EGARCH and CCC-FIEGARCH and the LM stochastic volatility model. We find the the rate of convergence and the limiting distribution of the estimator do not depend on the range of decay of the volatility. Efficiency comparison with the MLE estimator for the two-shocks specification show that the Whittle is over performed in terms of RMSE but stands out with respect to the parameters driving the long-memory and the asymmetry of the processes. An interesting development would be to allow the memory parameter to lie in the nonstationary region. Hualde and Robinson (2011) investigate fractionally integrated, possibly non stationary, linear processes and establish the asymptotic normality of a one-step estimator based on an initial $\sqrt{T}$ consistent estimate of the parameters. Extensions of their results to signal plus noise processes would allow to test for non stationarity in the fractionally integrated multivariate exponential volatility model, thus providing a general framework for testing for non stationarity.

APPENDIX A

In this appendix we establish a number of properties of the model mainly in terms of its spectral density $\mathbf{f}(\lambda,\theta)$ and its derivatives. Recall that $\sim$ denotes asymptotic equivalence, tr is the trace operator, det is the determinant operator and $\|.\|$ is the Euclidean norm. Constants (not always the same) are denoted by K. Almost sure convergence and convergence in distribution are denoted respectively by $\rightarrow^{a.s.}$ and $\rightarrow^{d}$. We denote a positive integer number as r. The class of p-integrable functions on the set Π is denoted as $\mathbf{L}_p\left(\Pi\right)$.

Lemma A1.1 *Under Assumption* A_1 *and* (3), *the* y_t *are ergodic and strictly stationary.*

Proof The ergodicity and strict stationarity of the $\mathbf{x}_t$ follows from Nelson (1991, Theorem 2.1, page 251) and implies the ergodicity and strict stationarity of the $\mathbf{y}_t$. The $\mathbf{y}_t$ are covariance stationary if and only if (3) holds.

Lemma A1.2 *Under Assumption* A_3, $\mathbf{f}(\lambda,\theta)$ *has elements in* $L_2\left(\Pi\right)$, *bounded and continuous at all* $(\lambda,\theta)\in\Pi\times\Theta$.

Proof Assumption A_3 implies that as $j\rightarrow\infty$

$$\sum_{j=0}^{\infty}\left|tr\left\{\Psi_j\left(\zeta\right)\right\}\right|<\infty.$$

However by (7)

$$tr\left\{\tilde{\Gamma}_u(\theta)\right\}=I_{(u=0)}tr\left\{\mathbf{\Sigma}_{\xi}\left(\tau\right)\right\}+tr\left\{\mathbf{\Sigma}_{\epsilon}\left(\tau\right)\sum_{j=0}^{\infty}\Psi_j\left(\zeta\right)\Psi_{j+u}^{'}\left(\zeta\right)\right\}+I_{(m\neq0)}tr\left\{\Psi_{|u|-1}\left(\zeta\right)\mathbf{\Sigma}_{\xi\epsilon}\left(\tau\right)\right\},$$

then Assumption $\mathbf{A}_3$ implies the uniquness of the spectral density and its square integrability, and continuity at all $(\lambda,\theta)\in\Pi\times\Theta$ (see Giraitis et al., 2012, Chapter 2, Proposition 2.2.1, page 11). Uniform continuity and compactness of the parameter space (see Assumption A_2) imply that the element of $\mathbf{f}(\lambda,\theta)$ are bounded at all (λ,θ).

Lemma A1.3 *Under Assumption A'*$_7$, $(\partial/\partial\theta)\,\mathbf{f}\left(\lambda,\theta\right)$ *has elements in* $L_2\left(\Pi\right)$ *which are bounded and continuous at all* $(\lambda,\theta)\in\Pi\times\Theta$.

Proof For any $j=1,...,s$,

$$\begin{aligned}\frac{\partial}{\partial\theta_j}\mathbf{f}\left(\lambda,\theta\right) &= \frac{\partial}{\partial\theta_j}\left[\frac{\mathbf{\Sigma}_{\xi}\left(\tau\right)}{2\pi}\right]+\frac{\partial}{\partial\theta_j}\left[\frac{\mathbf{k}\left(e^{i\lambda},\zeta\right)\mathbf{\Sigma}_{\epsilon}\left(\tau\right)\mathbf{k}\left(e^{i\lambda},\zeta\right)^{*}}{2\pi}\right]\\ &+\frac{\partial}{\partial\theta_j}\left[\frac{\mathbf{\Sigma}_{\epsilon\xi}\left(\tau\right)e^{-i\lambda}\mathbf{k}\left(e^{i\lambda},\zeta\right)^{*}+e^{i\lambda}\mathbf{k}\left(e^{i\lambda},\zeta\right)\mathbf{\Sigma}_{\epsilon\xi}^{'}\left(\tau\right)}{2\pi}\right].\end{aligned}$$

Assumption A_3 implies that for any $j=1,...,s$,

$$\sum_{u=0}^{\infty}\left|tr\left\{\frac{\partial}{\partial\theta_j}\tilde{\Gamma}_u(\theta)\right\}\right|<\infty,$$

which is sufficient condition for the existence of the first derivative of $\mathbf{f}(\lambda,\theta)$. Moreover it implies that $(\partial/\partial\theta)\,\mathbf{f}(\lambda,\theta)$ has elements in $L_2(\Pi)$ which are continuous at all $(\lambda,\theta)\in\Pi\times\Theta$ (see Giraitis et al, 2012, Proposition 2.2.1., page 11). By compactness of the parameter space, the uniform continuity implies that the elements of $(\partial/\partial\theta)\,\mathbf{f}(\lambda,\theta)$ are bounded at all $(\lambda,\theta)\in\Pi\times\Theta$.

Lemma A1.4 *Under Assumption A_4, $\mathbf{f}(\lambda,\theta)$ is a strictly positive definite matrix for all $\theta\in\Theta$, $\lambda\in\Pi$.*

Proof The autocovariance function of the process is

$$\tilde{\mathbf{\Gamma}}(\theta,u)=I_{(u=0)}\mathbf{\Sigma}_{\xi}(\tau)+\mathbf{\Sigma}_{\epsilon}(\tau)\sum_{j=0}^{\infty}\Psi_j(\zeta)\,\Psi_{j+u}^{'}(\zeta)+I_{(m\neq 0)}\Psi_{|u|-1}(\zeta)\,\mathbf{\Sigma}_{\xi\epsilon}(\tau).$$

Under Assumption A_5, $\mathbf{\Sigma}_{\epsilon}(\tau)$ is positive definite for every value of τ in the parameter space. Since by definition $\mathbf{\Sigma}_{\xi}(\tau)$ and $\mathbf{\Sigma}_{\xi\epsilon}(\tau)$ are positive semidefinite covariance matrices, $\tilde{\mathbf{\Gamma}}(\theta,u)$ is expressed as the sum of two positive semidefinite matrices and one positive definite matrix. Therefore it is positive definite. By definition the spectrum is the unique Fourier transform of the autocovariance matrix and its positive definiteness is implied by the positive definiteness of $\tilde{\mathbf{\Gamma}}(\theta,u)$.

Lemma A.5 *Under Assumption A_1, y_t is purely non deterministic with Wold decomposition*

$$\mathbf{y}_t=\sum_{l=0}^{\infty}A_l(\theta)\,e_{t-l},\quad \sum_{l=0}^{\infty}\|A_l(\theta)\|^2<\infty,$$

where the e_t are n dimensional white noise vectors.

Proof Since $\mathbf{y}_t$ is a stationary zero mean process, the result follows once we establish that

$$\int_{-\pi}^{\pi}\log\det\mathbf{f}(\lambda,\theta)>-\infty,$$

(see Giraitis et al, 2012, Theorem 3.2.1, page 38). By the logarithm inequality

$$\left|1-\det\mathbf{f}^{-1}(\lambda,\theta)\right|\le|\log\det\mathbf{f}(\lambda,\theta)|\le|\det\mathbf{f}(\lambda,\theta)-1|$$

and the result follows from the continuity of $\mathbf{f}^{-1}(\lambda,\theta)$ at all (λ,θ) by Assumption A_7 and Lemma A1.4**.**

Lemma A.6 *Under Assumption A'$_7$, $\log\det\mathbf{f}(\lambda,\theta)$ is differentiable in $\theta\in\Theta$ under the integral sign.*

Proof Denoting the jth unit vector in R^s by i_j, we have

$$\begin{aligned}&\frac{1}{\varepsilon}\left[\frac{1}{2\pi}\int_{-\pi}^{\pi}\log\det\mathbf{f}(\lambda,\theta+i_j\varepsilon)\,d\lambda-\frac{1}{2\pi}\int_{-\pi}^{\pi}\log\det\mathbf{f}(\lambda,\theta)\,d\lambda\right]\\ =\;&\frac{1}{\varepsilon}\frac{1}{2\pi}\int_{-\pi}^{\pi}\log\det\mathbf{f}(\lambda,\theta+i_j\varepsilon)-\log\det\mathbf{f}(\lambda,\theta)\,d\lambda.\end{aligned}$$

By the mean value theorem the integrand is bounded by

$$\left|\frac{\partial}{\partial\theta_j^*}\log\det\mathbf{f}\left(\lambda,\theta^*\right)\right|=\left|tr\left\{\mathbf{f}^{-1}\left(\lambda,\theta^*\right)\frac{\partial}{\partial\theta_j^*}\mathbf{f}\left(\lambda,\theta^*\right)\right\}\right|,$$

where $\left|\theta^*\left(\lambda\right)-\theta\right|<\left|\varepsilon\right|.$ By Assumption A'$_7$, Lemma A1.4 and compactness of the parameter space

$$\left|tr\left\{\mathbf{f}^{-1}\left(\lambda,\theta^*\right)\frac{\partial}{\partial\theta_j^*}\mathbf{f}\left(\lambda,\theta^*\right)\right\}\right|<K,$$

where K is a positive constant that does not depend on θ. Then

$$\int_{-\pi}^{\pi}\left|tr\left\{\mathbf{f}^{-1}\left(\lambda,\theta^*\right)\frac{\partial}{\partial\theta_j^*}\mathbf{f}\left(\lambda,\theta^*\right)\right\}\right|d\lambda<\infty,$$

and the dominated convergence theorem implies that $\int_{-\pi}^{\pi}\log\det\mathbf{f}\left(\lambda,\theta\right)$ can be differentiated under the integral sign.

Lemma A1.7 *Under assumption A'$_7$, for $r=1,2,3,\left(\partial^r/\partial\theta_{j_1}...\theta_{j_r}\right)\mathbf{f}\left(\lambda,\theta\right)$ has elements in $L_2\left(\Pi\right)$ which are bounded and continuous at all $\left(\lambda,\theta\right)\in\Pi\times\Theta$.*

Proof For any $j_h=1,...,s$, with $h=1,...,r$ and $r=1,2,3$,

$$\begin{aligned}\frac{\partial^r}{\partial\theta_{j_1}...\partial\theta_{j_r}}\mathbf{f}\left(\lambda,\theta\right)\quad=\quad&\frac{\partial^r}{\partial\theta_{j_1}...\partial\theta_{j_r}}\left[\frac{\mathbf{\Sigma}_{\xi}\left(\tau\right)}{2\pi}\right]+\frac{\partial^r}{\partial\theta_{j_1}...\partial\theta_{j_r}}\left[\frac{\mathbf{k}\left(e^{i\lambda},\zeta\right)\mathbf{\Sigma}_{\epsilon}\left(\tau\right)\mathbf{k}\left(e^{i\lambda},\zeta\right)^*}{2\pi}\right]\\&+\frac{\partial^r}{\partial\theta_{j_1}...\partial\theta_{j_r}}\left[\frac{\mathbf{\Sigma}_{\epsilon\xi}\left(\tau\right)e^{-i\lambda}\mathbf{k}\left(e^{i\lambda},\zeta\right)^*+e^{i\lambda}\mathbf{k}\left(e^{i\lambda},\zeta\right)\mathbf{\Sigma}_{\epsilon\xi}'\left(\tau\right)}{2\pi}\right].\end{aligned}$$

Assumption A$_4$ implies that for any $j_h=1,...,s$, with $h=1,...,r$ and $r=1,2,3$

$$\sum_{u=0}^{\infty}\left|tr\left\{\frac{\partial^r}{\partial\theta_{j_1}...\theta_{j_r}}\Gamma_u(\theta)\right\}\right|<\infty,$$

which is sufficient condition for the existence of the rth derivative of $\mathbf{f}\left(\lambda,\theta\right)$. Moreover it implies that $\left(\partial^r/\partial\theta_{j_1}...\theta_{j_r}\right)\mathbf{f}\left(\lambda,\theta\right)$ has elements in $L_2\left(\Pi\right)$ which are continuous at all $\left(\lambda,\theta\right)\in\Pi\times\Theta$ (see Giraitis et al, 2012, Proposition 2.2.1, page 11). By compactness of the parameter space (see Assumption A$_2$), the uniform continuity implies that the elements of $\left(\partial^r/\partial\theta_{j_1}...\theta_{j_r}\right)\mathbf{f}\left(\lambda,\theta\right)$ are bounded at all $\left(\lambda,\theta\right)\in\Pi\times\Theta$.

Lemma A1.8 *Under Assumption A$_1$-A$_7$, $\mathbf{g}\left(\lambda,\theta\right)\equiv\mathbf{f}^{-1}\left(\lambda,\theta\right)\dot{\mathbf{f}}\left(\lambda,\theta\right)$ is uniformly continuous in $\left(\lambda,\theta\right)$.*

Proof The uniform continuity of $\mathbf{g}\left(\lambda,\theta\right)$ follows once we establish

$$\sup_{\theta^*\in\Theta}\left\|\frac{\partial}{\partial\theta^*}\mathbf{g}\left(\lambda,\theta^*\right)\right\|<\infty,$$

where $|\theta^*(\lambda) - \theta| < |\varepsilon|$, (see Davidson, 1994, Theorem 21.10, page 339). However

$$\frac{\partial}{\partial\theta^*}\mathbf{g}(\lambda,\theta^*) = \mathbf{f}^{-1}(\lambda,\theta^*)\ddot{\mathbf{f}}(\lambda,\theta) + \left(\frac{\partial}{\partial\theta^*}\mathbf{f}^{-1}(\lambda,\theta^*)\right)\dot{\mathbf{f}}(\lambda,\theta),$$

which by Assumption $\mathbf{A}_7$, Lemma A1.8 and compactness of the parameter space is bounded by a positive constant K for all $\theta^* \in \Theta$.

Lemma A1.9 *Under Assumption A_7, $\int_{-\pi}^{\pi} \log\det \mathbf{f}(\lambda,\theta)\, d\lambda$ can be differentiated twice under the integral sign.*

Proof By the same argument used in Lemma B1.7, the result follows from Assumption $\mathbf{A_7}$ and Lemma A1.9.

Lemma A1.10 *Under Assumption A'$_1$ and A'$_7$, for $1 \le a,b,c,d \le n$.*

$$\sum_{t_1,t_2,t_3,t_4=-\infty}^{\infty} |K^{\mathbf{y}}_{abcd}(t_1,t_2,t_3,t_4)| < \infty.$$

Proof Denote as $K_{abcd}(\mathbf{x}_t,\mathbf{y}_t,\mathbf{z}_t,\mathbf{u}_t)$ the fourth order cumulant of elements a,b,c,d of random vectors $\mathbf{x}_t,\mathbf{y}_t,\mathbf{z}_t,\mathbf{u}_t$. Set $K^{\epsilon}_{abcd} = cumulant(\epsilon_0^{(a)},\epsilon_0^{(b)},\epsilon_0^{(c)},\epsilon_0^{(d)})$ and set $K^{\xi}_{abcd} = cumulant(\xi_0^{(a)},\xi_0^{(b)},\xi_0^{(c)},\xi_0^{(d)})$. Then $K^{\mathbf{y}}_{abcd}(t_1,t_2,t_3,t_4)$ is made by the sum of the following terms:

$$\sum_{a,b,c,d=1}^{n}\left(K^{\xi}_{abcd}\mathit{1}(t_1=t_2=t_3=t_4)\right) \tag{14}$$

$$\sum_{r=a,b,c,d}\left(\sum_{a,b,c,d=1}^{n} K_{abcd}(\xi_0,\xi_0,\xi_0,\epsilon_0)\,\Psi^{(r)}_{t_4-t_1-1}\mathit{1}(t_1=t_2=t_3)\right)$$

$$\sum_{u,v=a,b,c,d}\left(\sum_{a,b,c,d=1}^{n} K_{abcd}(\xi_0,\xi_0,\epsilon_0,\epsilon_0)\,\Psi^{(u,u)}_{t_3-t_2-1}\Psi^{(v)}_{t_4-t_1-1}\mathit{1}(t_1=t_2)\right)$$

$$\sum_{u,v,z=a,b,c,d}\left(\sum_{a,b,c,d=1}^{n} K_{abcd}(\xi_0,\epsilon_0,\epsilon_0,\epsilon_0)\,\Psi^{(u)}_{t_2-t_1-1}\Psi^{(v)}_{t_3-t_1-1}\Psi^{(z)}_{t_4-t_2-1}\right)$$

$$\sum_{a,b,c,d=1}^{n}\left(K^{\epsilon}_{abcd}\sum_{j=0}^{\infty}\Psi^{(a)}_{j}\Psi^{(b)}_{j+t_2-t_1}\Psi^{(c)}_{j+t_3-t_1}\Psi^{(d)}_{j+t_4-t_1}\right).$$

The absolute summability of the cumulants follows from the absolute summability of the last term in (14), which is implied by Assumption A_3.

Lemma A1.11 *Under Assumption A'$_1$- A'$_7$ the trispectrum of* $\mathbf{y}_t$,

$$\tilde{K}^{\mathbf{y}}_{abcd}(\lambda_1,\lambda_2,\lambda_3) = \frac{1}{(2\pi)^3}\sum_{t_1,t_2,t_3=-\infty}^{\infty}\exp\{-i(\lambda_1 t_1+\lambda_2 t_2+\lambda_3 t_3)\}K^{\mathbf{y}}_{abcd}(t_1,t_2,t_3)$$

is square integrable.

Proof Follows from the square summability of the Fourier coefficients of $\tilde{K}^{\mathbf{y}}_{abcd}$, implied by Lemma A1.10 (see Giraitis et al., 2012, (2.1.4))

APPENDIX B: Consistency Lemmas

This section contains the proof of Lemma 6 and Lemma 7 in the short and long memory cases. We state two preliminary results. In Lemma B.1 we recall a fundamental result of Hannan (1970) on matrix functions approximation and in Lemma B.2, we establish the uniform almost sure convergence of some discrete functions of the periodogram under minimal condition on the underlying model, generalizing Lemma 1 of Hannan (1973) to matrix functions. In what follows, for any matrix function $\mathbf{h}(\lambda, \theta)$, we denote by

$$\mathbf{h}_u(\theta) = \int_{-\pi}^{\pi} e^{iu\lambda} \mathbf{h}(\lambda, \theta)\, d\lambda, \quad u = 0, \pm 1, \pm 2, ...$$

its Fourier coefficients, and we denote by

$$\mathbf{q}_M(\lambda, \theta) = \sum_{u=-M}^{M} \left(1 - \frac{|u|}{M}\right) \mathbf{h}_u(\theta)\, e^{-iu\lambda},$$

the Cesaro sum of its Fourier coefficients up to M terms.

Lemma B.1 *Let $\mathbf{h}(\lambda, \theta)$ be a $n \times n$ matrix function, continuous in $\lambda \in \Pi$ and such that $\mathbf{h}(-\pi, \theta) = \mathbf{h}(\pi, \theta)$ in $[-\pi, \pi]$. Then $\mathbf{h}(\lambda, \theta)$ may be approximated uniformly in λ by $\mathbf{q}_M(\lambda, \theta)$,*

$$\sup_{\lambda \in \Pi} \|\mathbf{h}(\lambda, \theta) - \mathbf{q}_M(\lambda, \theta)\| \to 0 \quad \text{as} \quad M \to \infty.$$

If in addition $\mathbf{h}(\lambda, \theta)$ is continuous in λ uniformly in θ, the approximation may be made uniformly in θ also,

$$\sup_{\lambda, \theta} \|\mathbf{h}(\lambda, \theta) - \mathbf{q}_M(\lambda, \theta)\| \to 0 \quad \text{as} \quad M \to \infty.$$

Proof A detailed proof of this lemma for matrix functions can be found in Hannan (1970, Mathematical Appendix, Section 3).

Lemma B.2 *Let $\mathbf{y}_t$ be a stationary, ergodic and purely non deterministic vector process, with $n \times n$ spectral density matrix $\mathbf{f}(\lambda, \theta_0)$. Let $\mathbf{h}(\lambda, \theta)$ be a $n \times n$ matrix function, continuous in $(\lambda, \theta) \in \Pi \times \Theta$ and such that $\mathbf{h}(\lambda, \theta) = \mathbf{h}(-\lambda, \theta)$. Then, uniformly in $\theta \in \Theta$ and $\lambda \in \Pi$,*

$$\textbf{(a)}\ \frac{1}{2\pi} \int_{-\pi}^{\pi} tr\left\{\mathbf{h}(\lambda, \theta) \mathbf{I}_T(\lambda)\right\} d\lambda \ \to \frac{1}{2\pi} \int_{-\pi}^{\pi} tr\left\{\mathbf{h}(\lambda, \theta)\ \mathbf{f}(\lambda, \theta_0)\right\} d\lambda \quad \text{a.s.}$$

$$\textbf{(b)}\ \frac{1}{T} \sum_{t=1}^{T-1} tr\left\{\mathbf{h}(\lambda, \theta) \mathbf{I}_T(\lambda_t)\right\} \to \frac{1}{2\pi} \int_{-\pi}^{\pi} tr\left\{\mathbf{h}(\lambda, \theta)\ \mathbf{f}(\lambda, \theta_0)\right\} d\lambda \quad \text{a.s.}$$

Proof : (**a**) By Lemma B.1, for every $\eta > 0$ we may find M large enough such that:

$$\sup_{\lambda,\theta} \|\mathbf{h}(\lambda,\theta) - \mathbf{q}_M(\lambda,\theta)\| \leq \eta.$$

Let $\eta > 0$. For sufficiently large M, uniformly in θ :

$$\frac{1}{2\pi}\left|\int_{-\pi}^{\pi} tr\left\{\mathbf{h}(\lambda,\theta)\mathbf{I}_T(\lambda)\right\}d\lambda - \int_{-\pi}^{\pi} tr\left\{\mathbf{q}_M(\lambda_t,\theta)\mathbf{I}_T(\lambda)\right\}d\lambda\right|$$

$$= \frac{1}{2\pi}\left|\int_{-\pi}^{\pi} tr\left\{\left(\mathbf{h}(\lambda,\theta) - \mathbf{q}_M(\lambda,\theta)\right)\mathbf{I}_T(\lambda)\right\}d\lambda\right|$$

$$\leq \frac{\eta}{2\pi}\int_{-\pi}^{\pi} tr\left\{\mathbf{I}_T(\lambda)\right\}d\lambda = \frac{\eta}{2\pi}\ tr\left\{\frac{1}{T}\sum_{t=1}^{T}\sum_{u=-T+1}^{T-1} C\left(u\right)e^{-iu\lambda}\right\} = \frac{\eta}{2\pi} tr\left\{C\left(0\right)\right\},$$

where $C\left(0\right)$ is the sample autocovariance function. Since the process is ergodic, by the Ergodic Theorem (see Giraitis et al., Chapter 2, Section 2.5) $C\left(0\right)$ converges almost surely to its population analogue $\Gamma\left(0\right)$ as $T \to \infty$. Thus for all sufficiently large t, uniformly in $\theta \in \Theta$:

$$\left|\int_{-\pi}^{\pi} tr\left\{\mathbf{h}(\lambda,\theta)\mathbf{I}_T(\lambda)\right\}d\lambda - \int_{-\pi}^{\pi} tr\left\{\mathbf{q}_M(\lambda,\theta)\mathbf{I}_T(\lambda)\right\}d\lambda\right| \leq \eta\ tr\left\{\Gamma\left(0\right)\right\} \quad \text{a.s.}$$

Moreover,

$$\int_{-\pi}^{\pi} tr\left\{\mathbf{q}_M(\lambda,\theta)\mathbf{I}_T(\lambda)\right\}d\lambda = \int_{-\pi}^{\pi} tr\left[\sum_{u=-M}^{M}\left(1 - \frac{|u|}{M}\right)\mathbf{h}_u\left(\theta\right)e^{-iu\lambda}\mathbf{I}_T(\lambda)\right]d\lambda$$

$$= tr\left[\sum_{u=-M}^{M}\left(1 - \frac{|u|}{M}\right)\mathbf{h}_u\left(\theta\right)\mathbf{C}\left(u\right)\right].$$

By the Ergodic Theorem for each $|u| \leq M$, as $T \to \infty$, $C\left(u\right)$ converges almost surely to $\Gamma\left(u\right) = \int_{-\pi}^{\pi}\mathbf{f}\left(\lambda,\theta\right)e^{-i\lambda u}d\lambda$. Therefore the above expression tends almost surely to:

$$tr\left[\sum_{u=-M}^{M}\left(1 - \frac{|u|}{M}\right)\mathbf{h}_u\left(\theta\right)\Gamma\left(u\right)\right]$$

$$= tr\left[\sum_{u=-M}^{M}\left(1 - \frac{|u|}{M}\right)\mathbf{h}_u\left(\theta\right)\left(\frac{1}{2\pi}\int_{-\pi}^{\pi}\mathbf{f}(\lambda,\theta_0)e^{-iu\lambda}d\lambda\right)\right]$$

$$= \frac{1}{2\pi}\int_{-\pi}^{\pi} tr\left[\mathbf{q}_M(\lambda,\theta)\mathbf{f}(\lambda,\theta_0)\right]d\lambda \to \frac{1}{2\pi}\int_{-\pi}^{\pi} tr\left[\mathbf{h}(\lambda,\theta)\mathbf{f}(\lambda,\theta_0)\right]d\lambda$$

on letting $\eta \to 0$, which completes the proof.

($\mathbf{b}$): the proof follows as in part ($\mathbf{a}$) using the fact that

$$\left|\frac{1}{T}\sum_{t=1}^{T-1} tr\left\{\left(\mathbf{h}(\lambda,\theta)-\mathbf{q}_M(\lambda,\theta)\right)\mathbf{I}_T(\lambda_t)\right\}\right| \leq \frac{\eta}{T}\sum_{t=1}^{T-1}\mathbf{I}_T(\lambda_t)=\eta tr\left\{\mathbf{C}\left(0\right)\right\}$$

and that

$$\frac{1}{T}\sum_{t=1}^{T-1} tr\left\{\mathbf{q}_M(\lambda_t,\theta)\mathbf{I}_T(\lambda_t)\right\} = tr\left\{\sum_{u=-M}^{M}\left(1-\frac{|u|}{M}\right)\mathbf{h}_u\left(\theta\right)\mathbf{C}\left(u\right)\right\}.$$

Lemma 3. *If Assumptions* $\mathbf{A},\mathbf{B},\mathbf{C},\mathbf{D},\mathbf{E},\mathbf{F}$ *hold, for any* $\eta>0$ *uniformly in* $\theta\in\Theta$

$$(a)\lim_{T\to\infty} Q_T\left(\theta\right)=Q\left(\theta\right) \quad \text{a.s.}$$

$$(b)\lim_{T\to\infty} Q_{T,\eta}\left(\theta\right)=Q_\eta\left(\theta\right) \quad \text{a.s.}$$

Proof. (a) The almost sure uniform convergence of

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} tr\left(\mathbf{f}^{-1}(\lambda_t,\theta)\mathbf{I}_T(\lambda_t)\right)$$

to

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} tr\left(\mathbf{f}^{-1}(\lambda,\theta)\mathbf{f}(\lambda,\theta_0)\right)d\lambda$$

follows from Lemma B.2(a), taking $\mathbf{h}\left(\lambda,\theta\right)\equiv\mathbf{f}^{-1}\left(\lambda,\theta\right)$. By Assumption $\mathbf{C}(\mathbf{ii}),$ $\mathbf{f}^{-1}\left(\lambda,\theta\right)$ is uniformly continuous at all $\left(\lambda,\theta\right)$ and by definition $\mathbf{f}^{-1}\left(-\pi,\theta\right)=\mathbf{f}^{-1}\left(\pi,\theta\right)$. By Lemma A.5, $\mathbf{y}_t$ is a linearly regular process and the conditions of Lemma B.2 are satisfied. ■

Consider the first term of $Q_T\left(\theta\right)$,

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}\log\det\mathbf{f}(\lambda,\theta)d\lambda.$$

This term is non stochastic and its uniform convergence follows once with establish the equicontinuity property

$$\lim_{\varepsilon\to 0}\sup_{\tilde{\theta}:\left\|\tilde{\theta}-\theta\right\|\leq\epsilon}\left|\int_{-\pi}^{\pi}\log\det\mathbf{f}\left(\lambda,\tilde{\theta}\right)d\lambda-\frac{1}{2\pi}\int_{-\pi}^{\pi}\log\det\mathbf{f}\left(\lambda,\theta\right)d\lambda\right|\to 0. \quad (15)$$

(15) is implied by

$$\sup_{\theta^*\in\Theta}\left|\frac{\partial}{\partial\theta}\frac{1}{2\pi}\int_{-\pi}^{\pi}\log\det\mathbf{f}\left(\lambda,\theta^*\right)d\lambda\right|<\infty,$$

where $|\theta^*(\lambda)-\theta| < |\varepsilon|$, (see Davidson, 1994, Theorem 21.10, page 339). By Lemma A.6 ,

$$\begin{aligned}
& \frac{\partial}{\partial\theta}\frac{1}{2\pi}\int_{-\pi}^{\pi}\log\det\mathbf{f}(\lambda,\theta^*)\,d\lambda \\
= & \frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{\partial}{\partial\theta}\log\det\mathbf{f}(\lambda,\theta^*)\,d\lambda \\
= & \frac{1}{2\pi}\int_{-\pi}^{\pi} tr\left\{\mathbf{f}^{-1}(\lambda,\theta^*)\frac{\partial}{\partial\theta}\mathbf{f}(\lambda,\theta^*)\right\}. \qquad (16)
\end{aligned}$$

We need to distinguish two cases.

1) If $d=0$, (16) is bounded by some positive constant by Assumptions B, C(ii) and G(i); the use of the dominated convergence theorem allows to conclude that

$$\sup_{\theta^*\in\Theta}\left|\frac{1}{2\pi}\int_{-\pi}^{\pi} tr\left\{\mathbf{f}^{-1}(\lambda,\theta^*)\frac{\partial}{\partial\theta}\mathbf{f}(\lambda,\theta^*)\right\}\right| < K,$$

which concludes the proof. OPPURE If $d=0$, $\mathbf{f}(\lambda,\theta)$ is continuous at all (λ,θ). Compactness of the parameter space by Assumption **B**, strict positivity of the spectral density and continuity of the determinant and of the logarithmic function (see Magnus, Capter 1) imply that it converges uniformly.

2) If $d\in(0,1/2)$, by Assumptions B, C(ii) and G(i) (16) is at most, ignoring constant terms,

$$\int_{-\pi}^{\pi}|\lambda|^{2(d_l-d_u)-\delta}\,d\lambda < \infty,$$

where we take $d_l=\inf_\Theta d(\theta)$ and $d_u=\sup_\Theta d(\theta)$ and so $(d_l-d_u)>-1/2$ and δ can be taken arbitrarily small. Then we choose δ such that

$$\int_{-\pi}^{\pi}|\lambda|^{2(d_l-d_u)-\delta}\,d\lambda < \infty,$$

and the use of the dominated convergence theorem concludes.

(b) The almost sure uniform convergence of

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} tr\left(\phi_\eta(\lambda_t,\theta)\mathbf{I}_T(\lambda_t)\right)$$

to

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} tr\left(\phi_\eta(\lambda,\theta)\mathbf{f}(\lambda,\theta_0)\right)d\lambda$$

follows from Lemma B.2, taking $\mathbf{h}(\lambda,\theta)\equiv\phi_\eta(\lambda,\theta)$ since by Assumption **C,** $\phi_\eta(\lambda,\theta)$ is uniformly continuous in (λ,θ), and satisfies $\phi_\eta(\lambda,\theta)=\phi_\eta(-\lambda,\theta)$ for all $\lambda\in\Pi$.

Lemma 4. *If Assumptions* **D** *and* **E** *hold, then for all* $\theta\in\Theta$,

$$\begin{aligned}
\inf_{\theta\in\Theta} Q(\theta) &= Q(\theta_0) = \int_{-\pi}^{\pi}\log\det\mathbf{f}(\lambda,\theta_0)\,d\lambda + \frac{1}{2\pi}\int_{-\pi}^{\pi} tr\left\{\mathbf{f}^{-1}(\lambda,\theta_0)\mathbf{f}(\lambda,\theta_0)\right\}d\lambda \\
&= \int_{-\pi}^{\pi}\log\det\mathbf{f}(\lambda,\theta_0)\,d\lambda + T.
\end{aligned}$$

Proof.

$$Q(\theta) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \log\det \mathbf{f}(\lambda,\theta)d\lambda + \frac{1}{2\pi}\int_{-\pi}^{\pi} tr\left[\mathbf{f}^{-1}(\lambda,\theta)\ \mathbf{f}(\lambda,\theta_0\right] d\lambda, \tag{17}$$

adding and subtracting $1/2\pi\int_{-\pi}^{\pi}\log\det\mathbf{f}\left(\lambda,\theta_0\right)d\lambda$, (17) is equal to

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} \log\det \mathbf{f}(\lambda,\theta_0)d\lambda + \frac{1}{2\pi}\int_{-\pi}^{\pi} tr\left[\mathbf{f}^{-1}(\lambda,\theta)\ \mathbf{f}(\lambda,\theta_0\right] d\lambda - \frac{1}{2\pi}\int_{-\pi}^{\pi} \log\frac{\det\mathbf{f}(\lambda,\theta_0)}{\det\mathbf{f}(\lambda,\theta)}d\lambda,$$

because for any non-singular matrix A, $\det^{-1}(A) = \det\left(A^{-1}\right)$ (Luktepohl, 1996, Section 3.4.4, Result (f)), (17) is equal to

$$Q(\theta_0) - T + \frac{1}{2\pi}\int_{-\pi}^{\pi} tr\left[\mathbf{f}^{-1}(\lambda,\theta)\ \mathbf{f}(\lambda,\theta_0\right] d\lambda - \frac{1}{2\pi}\int_{-\pi}^{\pi}\log\left(\det\left(\mathbf{f}^{-1}(\lambda,\theta)\right)\det\left(\mathbf{f}(\lambda,\theta_0)\right)\right)d\lambda,$$

because for any non singular matrix A and B, $\det(A)\times\det(B) = \det(AB)$ (Luktepohl, 1996, Section 4.2.1, Result (4)), (17) is equal to

$$Q(\theta_0) + \left\{\frac{1}{2\pi}\int_{-\pi}^{\pi} tr\left[\mathbf{f}^{-1}(\lambda,\theta)\ \mathbf{f}(\lambda,\theta_0\right] d\lambda - T\right\} - \frac{1}{2\pi}\int_{-\pi}^{\pi}\log\left(\det\left(\mathbf{f}^{-1}(\lambda,\theta)\mathbf{f}(\lambda,\theta_0)\right)\right)d\lambda > Q\left(\theta_0\right),$$

where the strict inequality follows, in view of Lemma B1.5, because $\log\det(A) \le tr(A) - T$ for any positive definite matrix A with equality holding if and only if $A = I_n$ (Luktepohl, 1996, Section 4.1.2, Result (10)). ■

Lemma 5. *If Assumptions* $\mathbf{A},\mathbf{B},\mathbf{C},\mathbf{D},\mathbf{E},\mathbf{F}$ *hold, for any* $\eta > 0$ *uniformly in* $\theta\in\Theta$

$$(a)\lim_{T\to\infty}\bar{Q}_T(\theta) = Q(\theta) \quad \text{a.s.}$$

$$(b)\lim_{T\to\infty}\bar{Q}_{T,\eta}(\theta) = Q_\eta(\theta) \quad \text{a.s.}$$

Proof. (a) The almost sure uniform convergence of the second term of $\bar{Q}_T(\theta)$ follows as in Lemma 6 from Lemma B.2(b), Assumptions A,B,C and Lemma A.5. For the first, non-stochastic, term of $\bar{Q}_T(\theta)$ we distinguish two cases ■

1) When $d = 0$ (short memory) $\mathbf{f}(\lambda,\theta)$ is continuous at all (λ,θ). Compactness of the parameter space by Assumption **B**, strict positivity of the spectral density and continuity of the determinant and of the logarithmic function (see Magnus, Capter 1) imply that uniformly in (λ,θ)

$$\frac{1}{T}\sum_{t=1}^{T-1}\log\det\mathbf{f}(\lambda,\theta) \to \int_{-\pi}^{\pi}\log\det\mathbf{f}(\lambda,\theta)d\lambda.$$

2) When $d\in(0,1/2)$, we adapt Robinson

Consider that by Lemma (quello in cui rappresenti il processo come vector linear)

$$\mathbf{f}\left(\lambda,\theta\right)=k\left(\theta,e^{i\lambda j}\right)Gk\left(\theta,e^{i\lambda j}\right)$$

$\mathbf{k}(e^{i\lambda},\zeta)=I+\sum_{j=0}^{\infty}\mathbf{\Psi}_j\left(\zeta\right)e^{i\lambda j}$

$\det\mathbf{f}\left(\lambda,\theta\right)=\det k\left(\theta,e^{i\lambda j}\right)\det G\det k\left(\theta,e^{i\lambda j}\right)$ (lukthepol) , $\log\det\mathbf{f}\left(\lambda,\theta\right)=\log\det k\left(\theta,e^{i\lambda j}\right)+\log\det G+\log\det k\left(\theta,e^{i\lambda j}\right)$

APPENDIX C

This section contains the proof of the lemmas used to establish the asymptotic normality of the estimator.

Lemma C1.1 *The asymptotic covariance between* $\tilde{\tau}_{(a,b)}\left(m\right)$ *and* $\tilde{\tau}_{(c,d)}\left(u\right)$ *is given as*

$$2\pi\int_{-\pi}^{\pi}\left\{\mathbf{f}_{(a,c)}\left(\lambda,\theta\right)\overline{\mathbf{f}}_{(b,d)}\left(\lambda,\theta\right)e^{-i(m-u)\lambda}+\mathbf{f}_{(a,d)}\left(\lambda,\theta\right)\overline{\mathbf{f}}_{(b,c)}\left(\lambda,\theta\right)e^{i(m+u)\lambda}\right\}d\lambda$$
$$+2\pi\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}e^{(im\lambda_1+iu\lambda_2)}\tilde{K}_{abcd}^{\mathbf{y}}(-\lambda_1,\lambda_2,-\lambda_2)d\lambda_1d\lambda_2.$$

Proof The covariance between $\tilde{\tau}_{(a,b)}\left(m\right)$ and $\tilde{\tau}_{(c,d)}\left(n\right)$ is

$$\frac{1}{T}\sum_{u=1-T}^{T-1}\left(1-\frac{|u|}{T}\right)\left\{\tilde{\Gamma}_{(a,c)}\left(n\right)\tilde{\Gamma}_{(b,d)}\left(u+n-m\right)+\tilde{\Gamma}_{(a,d)}\left(u+n\right)\tilde{\Gamma}_{(b,c)}\left(u-m\right)\right\}$$
$$+\sum_{u=1-T}^{T-1}\left(1-\frac{|u|}{T}\right)K_{a,b,c,d}^{\mathbf{y}}\left(m,u,u+n\right),\tag{18}$$

(see Hannan, 1979, page 209-211). The term

$$\frac{1}{T}\sum_{u=1-T}^{T-1}\left(1-\frac{|u|}{T}\right)\tilde{\Gamma}_{(a,c)}\left(n\right)\tilde{\Gamma}_{(b,d)}\left(u+n-m\right),\tag{19}$$

is the Cesaro sum, evaluated at the origin, of $(4\pi)^2$ the uth Fourier coefficient of the convolution of $\mathbf{f}_{(a,c)}\left(\lambda\right)$ with $\mathbf{f}_{(b,d)}\left(\lambda\right)e^{-i(m-n)\lambda}$. By Lemma B1.3, $\mathbf{f}\left(\lambda\right)$ has elements in L_2, so their convolution is continuous. Then (19) converges to

$$2\pi\int_{-\pi}^{\pi}\mathbf{f}_{(a,c)}\left(\lambda,\theta\right)\overline{\mathbf{f}}_{(b,d)}\left(\lambda,\theta\right)e^{-i(n-m)\lambda}d\lambda.$$

The same argument applies to

$$\frac{1}{T}\sum_{u=1-T}^{T-1}\left(1-\frac{|u|}{T}\right)\tilde{\Gamma}_{(a,d)}\left(u+n\right)\tilde{\Gamma}_{(b,c)}\left(u-m\right),$$

which converges to

$$2\pi\int_{-\pi}^{\pi}\mathbf{f}_{(a,d)}\left(\lambda,\theta\right)\overline{\mathbf{f}}_{(b,c)}\left(\lambda,\theta\right)e^{i(m+u)\lambda}d\lambda.$$

(18) is the Cesaro sum, evaluated at the zero frequency, of the Fourier coefficients of the function $\tilde{K}^{\mathbf{y}}_{abcd}\left(\lambda_1,\lambda_2,\lambda_3\right)e^{-i(n\lambda_1+m\lambda_2)}$. By Lemma B1.11, the trispectrum of the process is square integrable, implying the convergence of (18) to

$$2\pi\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}e^{(in\lambda_1+im\lambda_2)}\tilde{K}^{\mathbf{y}}_{abcd}\left(-\lambda_1,\lambda_2,-\lambda_2\right)d\lambda_1 d\lambda_2.$$

Lemma 6. *Under Assumptions* $\mathbf{A}(4)$, $\mathbf{B}$, $\mathbf{C}$, $\mathbf{D}$, $\mathbf{E}$, $\mathbf{F}$, $\mathbf{G}$, $\mathbf{H}$ *as* $T\to\infty$, *uniformly in* $\theta\in\Theta$,

$$\tilde{\mathbf{H}}_T\left(\theta\right)\overset{a.s}{\to}\mathbf{H}\left(\theta\right)$$

almost surely, where $\mathbf{H}\left(\theta\right)$ *is a positive definite matrix with* (i,j) *element,*

$$\begin{aligned}\mathbf{H}_{(i,j)}\left(\theta\right) &= \frac{1}{2\pi}\int_{-\pi}^{\pi}tr\left\{\mathbf{f}^{-1}\left(\lambda,\theta\right)\ddot{\mathbf{f}}_{(i,j)}\left(\lambda,\theta\right)\right\}d\lambda\\ &\quad-\frac{1}{2\pi}\int_{-\pi}^{\pi}tr\left\{\mathbf{f}^{-1}\left(\lambda,\theta\right)\dot{\mathbf{f}}_{(i)}\left(\lambda,\theta\right)\mathbf{f}^{-1}\left(\lambda,\theta\right)\dot{\mathbf{f}}_{(j)}\left(\lambda,\theta\right)\right\}d\lambda\\ &\quad-\frac{1}{2\pi}\int_{-\pi}^{\pi}tr\left\{\left(\frac{\partial^2}{\partial\theta_i\partial\theta_j'}\mathbf{f}^{-1}\left(\lambda,\theta\right)\right)\mathbf{f}\left(\lambda,\theta_0\right)d\lambda\right\}.\end{aligned}$$

Proof. Proof. We establish the uniform convergence of $\ddot{Q}_T(\theta)$ to $M(\theta)$ pointwise. The (i,j) element of $\ddot{Q}_T(\theta)$, $\ddot{Q}_T^{(i,j)}(\theta)$ is

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}tr\left\{\mathbf{f}^{-1}\left(\lambda,\theta\right)\ddot{\mathbf{f}}_{(i,j)}\left(\lambda,\theta\right)\right\}d\lambda \tag{20}$$

$$-\frac{1}{2\pi}\int_{-\pi}^{\pi}tr\left\{\mathbf{f}^{-1}\left(\lambda,\theta\right)\dot{\mathbf{f}}_{(i)}\left(\lambda,\theta\right)\mathbf{f}^{-1}\left(\lambda,\theta\right)\dot{\mathbf{f}}_{(j)}\left(\lambda,\theta\right)\right\}d\lambda \tag{21}$$

$$+\frac{1}{2\pi}\int_{-\pi}^{\pi}tr\left\{\left(\frac{\partial^2}{\partial\theta_i\partial\theta_j'}\mathbf{f}^{-1}(\lambda,\theta)\right)\mathbf{I}_T(\lambda)\right\}d\lambda. \tag{22}$$

The last term converges almost surely uniformly in $(\lambda,\theta)\in\Pi\times\Theta$ to

$$\frac{1}{2\pi}\int tr\left\{\left(\frac{\partial^2}{\partial\theta_i\partial\theta_j'}\mathbf{f}^{-1}(\lambda,\theta)\right)\mathbf{f}(\lambda,\theta_0)\right\}d\lambda$$

by Lemma 4, taking $\mathbf{h}\left(\lambda,\theta\right)\equiv\left(\partial^2/\partial\theta_i\partial\theta_j'\right)\mathbf{f}^{-1}(\lambda,\theta)$, which is continuos at all $(\lambda,\theta)\in\Pi\times\Theta$ by Assumption A_7 and symmetric around zero in $[-\pi,\pi]$.The first two terms of (20) are non stochastic. Their uniform convergence in θ follows once with establish their equicontinuity property. Consider the first term. We want to show that

$$\sup_{\tilde{\theta}:\left\|\tilde{\theta}-\theta\right\|\leq\epsilon}\left|\frac{1}{2\pi}\int_{-\pi}^{\pi}\left[tr\left\{\mathbf{f}^{-1}\left(\lambda,\tilde{\theta}\right)\ddot{\mathbf{f}}_{(i,j)}\left(\lambda,\tilde{\theta}\right)\right\}-tr\left\{\mathbf{f}^{-1}\left(\lambda,\theta\right)\ddot{\mathbf{f}}_{(i,j)}\left(\lambda,\theta\right)\right\}\right]d\lambda\right|\to 0\quad\text{a.s..} \tag{23}$$

(23) is implied by

$$\sup_{\theta\in\Theta}\left|\frac{\partial}{\partial\theta}\int_{-\pi}^{\pi} tr\left\{\mathbf{f}^{-1}(\lambda,\theta)\,\ddot{\mathbf{f}}_{(i,j)}(\lambda,\theta)\right\}d\lambda\right|<\infty, \tag{24}$$

(see Davidson, 1994, Theorem 21.10, page 339). We must establish that $\int_{-\pi}^{\pi} tr\left\{\mathbf{f}^{-1}(\lambda,\theta)\,\ddot{\mathbf{f}}_{(i,j)}(\lambda,\theta)\right\}d\lambda$ is differentiable under the integral sign. Denote the jth unit vector in R^s by i_j, and consider

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{1}{\varepsilon}tr\left\{\mathbf{f}^{-1}(\lambda,\theta+i_j\varepsilon)\,\ddot{\mathbf{f}}_{(i,j)}(\lambda,\theta+i_j\varepsilon)-\mathbf{f}^{-1}(\lambda,\theta)\,\ddot{\mathbf{f}}_{(i,j)}(\lambda,\theta)\right\}d\lambda.$$

By the mean value theorem the integrand is dominated for each λ by

$$\left|\frac{\partial}{\partial\theta_l}tr\left\{\mathbf{f}^{-1}(\lambda,\theta^*(\lambda))\,\ddot{\mathbf{f}}_{(i,j)}(\lambda,\theta^*(\lambda))\right\}\right|, \tag{25}$$

where $|\theta^*(\lambda)-\theta|<|\varepsilon|$. Taking derivatives (25) is equal to

$$\left|tr\left\{\mathbf{f}^{-1}(\lambda,\theta^*(\lambda))\left(\frac{\partial}{\partial\theta_l}\ddot{\mathbf{f}}_{(i,j)}(\lambda,\theta^*(\lambda))\right)+\left(\frac{\partial}{\partial\theta_l}\mathbf{f}^{-1}(\lambda,\theta^*(\lambda))\right)\ddot{\mathbf{f}}_{(i,j)}(\lambda,\theta^*(\lambda))\right\}\right|. \tag{26}$$

By Assumption A'$_7$(ii) and (iii), Lemma B1.8 and compactness of the parameter space, (26) is at most K, where K denotes a generic positive constant. The use of the dominated convergence theorem allows to conclude that

$$\left|\frac{\partial}{\partial\theta}\int_{-\pi}^{\pi} tr\left\{\mathbf{f}^{-1}(\lambda,\theta)\,\ddot{\mathbf{f}}_{(i,j)}(\lambda,\theta)\right\}d\lambda\right|<\infty,$$

which completes the proof of (23). The equicontinuity property of the second term of (20) is implied by

$$\sup_{\theta\in\Theta}\left|\frac{\partial}{\partial\theta_l}\frac{1}{2\pi}\int_{-\pi}^{\pi} tr\left\{\mathbf{f}^{-1}(\lambda,\theta)\,\dot{\mathbf{f}}_{(i)}(\lambda,\theta)\,\mathbf{f}^{-1}(\lambda,\theta)\,\dot{\mathbf{f}}_{(j)}(\lambda,\theta)\right\}d\lambda\right|<\infty, \tag{27}$$

(see Davidson, 1994, Theorem 21.10, page 339). The left hand side of (27) is differentiable under the integral sign because for $|\theta^*(\lambda)-\theta|<|\varepsilon|$

$$\begin{aligned}
&\left|\frac{\partial}{\partial\theta_l}tr\left\{\mathbf{f}^{-1}(\lambda,\theta^*)\,\dot{\mathbf{f}}_{(i)}(\lambda,\theta^*)\,\mathbf{f}^{-1}(\lambda,\theta^*)\,\dot{\mathbf{f}}_{(j)}(\lambda,\theta^*)\right\}\right|\\
=\;&|tr\left\{\mathbf{f}^{-1}(\lambda,\theta^*)\,\dot{\mathbf{f}}_{(i)}(\lambda,\theta^*)\,\mathbf{f}^{-1}(\lambda,\theta^*)\,\ddot{\mathbf{f}}_{(j,l)}(\lambda,\theta^*)\right\}\\
&+tr\left\{\mathbf{f}^{-1}(\lambda,\theta^*)\,\dot{\mathbf{f}}_{(i)}(\lambda,\theta^*)\left(\frac{\partial}{\partial\theta_l}\mathbf{f}^{-1}(\lambda,\theta^*)\right)\dot{\mathbf{f}}_{(j)}(\lambda,\theta^*)\right\}\\
&+tr\left\{\mathbf{f}^{-1}(\lambda,\theta^*)\,\ddot{\mathbf{f}}_{(i,l)}(\lambda,\theta^*)\,\mathbf{f}^{-1}(\lambda,\theta^*)\,\dot{\mathbf{f}}_{(j)}(\lambda,\theta^*)\right\}\\
&+tr\left\{\left(\frac{\partial}{\partial\theta_l}\mathbf{f}^{-1}(\lambda,\theta^*)\right)\dot{\mathbf{f}}_{(i)}(\lambda,\theta^*)\,\mathbf{f}^{-1}(\lambda,\theta^*)\,\dot{\mathbf{f}}_{(j)}(\lambda,\theta^*)\right\}|,
\end{aligned}$$

which by Assumption A'$_7$, Lemma B1.8, and compactness of the parameter space is bonded at all $\theta \in \Theta$. Then the use of the dominated convergence theorem completes the proof. ■ ■

Lemma 4 *Under Assumptions* $\mathbf{A}(4)$, $\mathbf{B}$, $\mathbf{C}$, $\mathbf{D}$, $\mathbf{E}$, $\mathbf{F}$, $\mathbf{G}$, $\mathbf{H}$ *as* $T \to \infty$,

$$\sqrt{T}\left[\mathbf{S}_T(\theta_0) - E\mathbf{S}_T(\theta_0)\right] \to_d N(0, \mathbf{V}(\theta_0)), \tag{28}$$

where $\mathbf{V}(\theta)$ *is a positive definite matrix with* (j,l) *element,*

$$\mathbf{V}_{(j,l)}(\theta_0) = \frac{1}{\pi}\int_{-\pi}^{\pi} \mathbf{tr}\left[\mathbf{f}(\lambda,\theta_0)\left(\frac{\partial}{\partial\theta_j}\mathbf{f}^{-1}(\lambda,\theta_0)\right)\mathbf{f}(\lambda,\theta)\left(\frac{\partial}{\partial\theta_l}\mathbf{f}^{-1}(\lambda,\theta_0)\right)\right]d\lambda$$

$$+\frac{1}{2\pi}\sum_{a,b,c,d=1}^{T}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\left\{\dot{\mathbf{f}}_{(j)}^{(a,b)}(\lambda_1,\theta_0)\dot{\mathbf{f}}_{(l)}^{(c,d)}(\lambda_2,\theta_0)\right\}\tilde{K}_{a,b,c,d}(-\lambda_1,\lambda_2,-\lambda_2,\theta_0)d\lambda_1 d\lambda_2.$$

Proof: The j-th element of the lhs of (28) is:

$$\sqrt{T}\left[\mathbf{S}_T^{(j)}(\theta_0) - E\mathbf{S}_T^{(j)}(\theta_0)\right] = \frac{\sqrt{T}}{2\pi}\int_{-\pi}^{\pi} tr\left\{\mathbf{g}^{(j)}(\lambda,\theta_0)\left(\mathbf{I}_T(\lambda) - \mathbf{E}\mathbf{I}_T(\lambda)\right)\right\}d\lambda \tag{91}$$

and the proof of the rhd of 91 is in Section 4.

Proof of Lemma 6 Set

$$\tilde{h}_1(u) = \frac{1}{2\pi}\int_{-\pi}^{\pi} h_1(\lambda)e^{iu\lambda}d\lambda,$$

$$\tilde{h}_2(u) = \frac{1}{2\pi}\int_{-\pi}^{\pi} h_2(\lambda)e^{iu\lambda}d\lambda.$$

Then

$$TCov\left\{\int_{-\pi}^{\pi} h_1(\lambda)\mathbf{I}_{ab}(\lambda)d\lambda, \int_{-\pi}^{\pi} h_2(\lambda)\mathbf{I}_{cd}(\lambda)d\lambda\right\} \tag{29}$$

$$= \frac{1}{T}\sum_{u_1,u_2,u_3,u_4=1}^{T}\tilde{h}_1(u_1-u_2)\tilde{h}_2(u_3-u_4)\tilde{\mathbf{\Gamma}}_{(a,c)}(u_3-u_1)\tilde{\mathbf{\Gamma}}_{(b,d)}(u_4-u_2) \tag{30}$$

$$+\frac{1}{T}\sum_{u_1,u_2,u_3,u_4=1}^{T}\tilde{h}_1(u_1-u_2)\tilde{h}_2(u_3-u_4)\tilde{\mathbf{\Gamma}}_{(a,d)}(u_4-u_1)\tilde{\mathbf{\Gamma}}_{(b,c)}(u_3-u_2) \tag{31}$$

$$+\frac{1}{T}\sum_{u_1u_2u_3u_4=1}^{T}\tilde{h}_1(u_1-u_2)\tilde{h}_2(u_3-u_4)\mathbf{K}_{abcd}^{\mathbf{y}}(u_2-u_1,u_3-u_1,u_4-u_1) \tag{32}$$

The convergence of (30) and (31) follows from Hannan (1976, Theorem 1, page 398). For example (30)

$$\left(\sum_{l=1-T}^{T+1}\left(1-\frac{|l_1|}{T}\right)\tilde{h}_1(l_1)\tilde{\Gamma}_{(b,d)}(u_4-u_2)\right)\times\left(\sum_{k=1-T}^{T+1}\left(1-\frac{|l_2|}{T}\right)\tilde{h}_2(l_2)\tilde{\Gamma}_{(a,c)}(u_3-u_1)\right),$$

which is the product of the Cesaro sums, evaluated at the origin, of the Fourier coefficients of the convolution of $h_1(\lambda)$ with $\mathbf{f}_{(b,d)}(\lambda)$ and of the convolution of $h_2(\lambda)$ with $\mathbf{f}_{(a,c)}(\lambda)$. Because $\mathbf{f}(\lambda)$ and $h(\lambda)$ are square integrable their convolution is continuous in $\lambda \in [-\pi, \pi]$. Then (30) converges to

$$2\pi \int_{-\pi}^{\pi} h_1(\lambda)(\lambda)\, \bar{\mathbf{f}}_{(b,d)}(\lambda)\, \bar{h}_2(\lambda)\, \mathbf{f}_{(a,c)} d\lambda.$$

An analogous result holds for (31). Set $l_1 = u_1$, $l_2 = u_2 - u_1$, $l_3 = u_3 - u_1$, $l_4 = u_4 - u_1$, (32) can be expressed as

$$\frac{1}{T} \sum_{l_2,l_3,l_4=1-T}^{T+1} (T - S(l_2, l_3, l_4))\, \tilde{h}_1(-l_2)\, \tilde{h}_2(l_3 - l_4)\, K^{\mathbf{y}}_{abcd}(l_2, l_3, l_4), \tag{33}$$

where

$$S(l_2, l_3, l_4) = \max(|l_2|, |l_3|, |l_4|)\, I(sign\ l_2 = sign\ l_3 = sign\ l_4)$$
$$+ \max(|l_i|, |l_j|) + |l_k|\, I(sign\ l_i = sign\ l_j = -sign\ l_k).$$

As $T \to \infty$, (33) converges to

$$\sum_{l_2,l_3,l_4=-\infty}^{+\infty} \tilde{h}_1(-l_2)\, \tilde{h}_2(l_3 - l_4)\, K^{\mathbf{y}}_{abcd}(l_2, l_3, l_4) \tag{34}$$

$$- \lim_{T\to\infty} \frac{1}{T} \sum_{l_2,l_3,l_4=1-N}^{N+1} S(l_2, l_3, l_4)\, \tilde{h}_1(-l_2)\, \tilde{h}_2(l_3 - l_4)\, K^{\mathbf{y}}_{abcd}(l_2, l_3, l_4) \tag{35}$$

However, since the functions h_1 and h_2 are square integrable in $[-\pi, \pi]$, for a certain positive constant K,

$$\frac{1}{T} \sum_{l_2,l_3,l_4=1-T}^{T+1} S(l_2, l_3, l_4)\, \tilde{h}_1(-l_2)\, \tilde{h}_2(l_3 - l_4)\, K^{\mathbf{y}}_{abcd}(l_2, l_3, l_4)$$

$$\leq \frac{K^2}{N} \left| \sum_{l_2,l_3,l_4=1-T}^{T+1} \max(|l_2| + |l_3| + |l_4|)\, \tilde{h}_1(-l_2)\, K^{\mathbf{y}}_{abcd}(l_2, l_3, l_4) \right|,$$

and the terms

$$\sum_{l_2,l_3,l_4=1-T}^{T+1} \frac{|l_j|}{T} \left| K^{\mathbf{y}}_{abcd}(l_2, l_3, l_4) \right|$$

for $j = 1, 2, 3$ converge to 0 as $T \to \infty$ using Lemma B1.10. Then as $T \to \infty$, (32) converges to (34). Then, by repeated application of the Parseval equality, (33) converges to

$$\int_{-\pi}^{\pi} \int_{\pi}^{\pi} h_1(\lambda_1)\, h_2(-\lambda_2)\, \tilde{K}^{\mathbf{y}}_{abcd}(\lambda_1, \lambda_2, -\lambda_2).$$

By Lemma A1.9 $\mathbf{g}^{(j)}(\lambda,\theta_0)$ is continuous at all λ, moreover it is symmetric in $[-\pi,\pi]$. Thus, by Lemma B2.1, for any $\eta>0$, and all $a,b=1,...n$, we can always choose M large enough such that

$$\max_{a,b=1,\ldots,n}\sup_{\lambda\in\Pi}\left|\mathbf{g}^{(j)}_{M_{(a,b)}}(\lambda,\theta_0)-\mathbf{g}^{(j)}_{(a,b)}(\lambda,\theta_0)\right|\leq\eta.$$

Consider that

$$\begin{aligned}
& Var\left(\frac{\sqrt{T}}{2\pi}\int_{-\pi}^{\pi}tr\left\{\delta^{(j)}_{M}(\lambda,\theta_0)\left[\mathbf{I_T}(\lambda)-\mathbf{EI}(\lambda)\right]\right\}d\lambda\right)\\
= & \; Var\left(\frac{\sqrt{T}}{2\pi}{}^{n}_{a,b=1}{}^{\pi}_{-\pi}\left[\mathbf{I}_{(a,b)}(\lambda)-\mathbf{EI}_{(a,b)}(\lambda)\right]\delta_{M(b,a)}(\lambda)d\lambda\right).
\end{aligned}$$

As $T\to\infty$, this is dominated by

$$\begin{aligned}
& \frac{2}{\pi}n^2{}_{-\pi}^{\pi}\left|\delta_{M(b,a)}(\lambda)\right|^2\mathbf{f}_{(a,a)}(\lambda)\,\bar{\mathbf{f}}_{(b,b)}(\lambda)\,d\lambda\\
& +\frac{2}{\pi}n^2{}_{-\pi}^{\pi}\delta_{M(b,a)}(\lambda)\bar{\delta}_{M(a,b)}(-\lambda)\mathbf{f}_{(a,b)}(\lambda)\,\bar{\mathbf{f}}_{(b,a)}(\lambda)\,d\lambda\\
& +\frac{2}{\pi}n^2{}_{-\pi}^{\pi}{}_{-\pi}^{\pi}\delta_{M(b,a)}(\lambda_1)\bar{\delta}_{M(d,c)}(-\lambda_2)\tilde{\mathbf{K}}_{abcd}(\lambda_1,\lambda_2,-\lambda_2)\,d\lambda_1 d\lambda_2.
\end{aligned}$$

By compactness of the parameter space, $\delta_{M(a,b)}(\lambda)$ is square integrable in λ,which tends to zero as $M\to\infty$, because the elements of the spectral density matrix and the trispectrum are integrable by Lemma B1.3 and Lemma B1.12.

Lemma 3

Proof. The jth element of $\sqrt{T}ES^{(j)}_T(\theta_0)$, can be written as

$$\frac{\sqrt{T}}{2\pi}\int_{-\pi}^{\pi}tr\left\{\mathbf{g}^{(j)}(\lambda,\theta_0)\left[\mathbf{EI}(\lambda)-\mathbf{f}(\lambda,\theta_0)\right]\right\}d\lambda.$$

Note that $\mathbf{EI}(\lambda)$ is the Cesaro sum of the Fourier coefficients of $\mathbf{f}(\lambda,\theta_0)$. Assumption A'$_7$(i) implies

$$\sup_{\lambda\in\Pi}\sum_{a,b=1}^{n}\left|\mathbf{EI}^{(a,b)}_{\mathbf{T}}(\lambda)-\mathbf{f}^{(a,b)}(\lambda,\theta_0)\right|=O\left(T^{-\alpha}\right)$$

uniformly in θ (see Hannan, 1970, page 513). Then

$$\begin{aligned}
\sqrt{T}E\dot{Q}_{T(j)}(\theta) &= \frac{\sqrt{T}}{2\pi}\int_{-\pi}^{\pi}tr\left\{\mathbf{g}^{(j)}(\lambda,\theta_0)\left[\mathbf{EI}(\lambda)-\mathbf{f}(\lambda,\theta_0)\right]\right\}d\lambda\\
&\leq \frac{1}{2\pi}\int_{-\pi}^{\pi}\max_{(a,b)}\sup_{\lambda\in\Pi}\left|\mathbf{g}^{(a,b)}_{(j)}(\lambda,\theta_0)\right|\sum_{a,b=1}^{n}\left\{\sqrt{T}\left|\mathbf{EI}_{(a,b)}(\lambda)-\mathbf{f}_{(a,b)}(\lambda,\theta_0)\right|\right\}d\lambda\\
&= O\left(T^{1/2-\alpha}\right),
\end{aligned}$$

which converges to zero as $T\to\infty$, since by Assumption A'$_7$(i), $\alpha>1/2$. ■

[1] Alexander, C., Chibumba, M.: Multivariate orthogonal factor GARCH. University of Sussex discussion papers in Mathematics,1998.

[2] Adenstedt, R. K.: On large-sample estimation for the mean of a stationary random sequence. Annals of Statistics 2, 1095-1107 (1974).

[3] Aguilar, O., West, M.: Bayesian dynamic factor models and portfolio allocation. Journal of Business and Economics Statistics 18, 338-357 (2000).

[4] Audrino, F., Barone-Adesi, G.: Average conditional correlation and tree structures for multivariate GARCH models. Journal of Forecasting 25, 579-600 (2006).

[5] Asai, M., McAleer, M.: Dynamic asymmetric leverage in stochastic volatility models. Econometric Reviews 24, 317-332 (2005).

[6] Asai, M., McAleer, M.: Asymmetric multivariate stochastic volatility. Econometric Reviews 25, 453-473 (2006).

[7] Asai, M., McAleer, M., Yu, J.: Multivariate Stochastic Volatility: A Review. Econometric Reviews 25, 145-175 (2006).

[8] Avarucci, M., Beutner, E., Zaffaroni, P.: On moments conditions for quasi-maximum likelihood estimation of multivariate ARCH models. Econometric Theory 29, 545-566 (2013).

[9] Baillie, R. T., Bollerlsev, T., Mikkelsen, O.H.: Fractionally Integrated Generalized Autoregressive Conditional Heteroskedasticity. Journal of Econometrics 74, 3-30 (1996).

[10] Barndorff-Nielsen, O. E., Shephard, N.: Normal Modified Stable Process. Theory of Probability and Mathematics Statististics 65, 1-19 (2001).

[11] Bauwens, L., Laurent, S.: A new class of multivariate skew densities, with application to GARCH models. Journal of Business and Economic Statistics, 101, 55-75 (2002).

[12] Bauwens, L., Laurent, S., Rombouts, J. V. K.: Multivariate GARCH models:a survey. Journal of Applied Econometrics 21, 79-109 (2006).

[13] Bauwens, L.,Veredas, D.: The stochastic conditional duration model: a latent variable model for the analysis of financial durations. Journal of Econometrics 119, 381-412 (2004).

[14] Bhansali, R.: Asymptotic properties of the Wiener-Kolmogorov predictior. Journal of the Royal Statistical Society Series B, 36, 61-73 (1974).

[15] Billio, M., Caporin, M.: A generalized dynamic conditional correlation model for portfolio risk evaluation. Unpublished manuscript, Ca'Foscari University of Venice, Department of Economics, 2006.

[16] Bollerlsev, T.: Modelling the coherence in short-run nominal exchange rates:a multivariate generalized ARCH model. Review of Economics and Statistics 72, 498-505 (1990).

[17] Bollerlsev, T., Engle, R. F., Nelson, D. B.: ARCH models. In: Handbook of Econometrics, R.F. Engle and D.L. McFadden ed.,vol 4, 2959-3038, Elsevier Science, Amsterdam (1994).

[18] Bollerlsev, T., Engle, R. F., Wooldridge, J.M.: A capital asset pricing model with time-varying covariances. The Journal of Political Economy 96, 116-131 (1998).

[19] Bollerlsev, T., Mikkelsen, H.: Modeling and pricing long memory in stock market volatility. Journal of Econometrics 73, 151-184 (1996).

[20] Bourgerol, P., Picard, N.: Strict stationarity of generalized autoregressive process. Journal of Econometrics, 31, 307-327 (1992).

[21] Brandt, M., Jones, C.: Volatility forecasting with range-based EGARCH models. Journal of Business & Economic Statistics 24, 470-486 (2005).

[22] Breidt, F., Crato, N., DeLima, P.: The detection and estimation of long memory in stochastic volatility. Journal of Econometrics 83, 325-348 (1998).

[23] Brockwell, P., Davis, R.: Time series: theory and methods. Springer Verlag, New York 1987.

[24] Cappiello, L., Engle, R. F., Sheppard, K.: Asymmetric dynamics in the correlations of global equity and bond returns. Journal of Financial Econometrics 4, 537-572 (2006).

[25] Cheung, Y. W., Lai, K. S., Lain, M.: Are there long cycles in foreign stock returns? Journal of International Financial Markets 3, 33-47 (1993).

[26] Chib, S., Nadari, F., Shephard, N.: Markov chain Monte Carlo methods for stochastic volatility models. Journal of Econometrics 108, 281-316 (2002).

[27] Chib, S., Omori, Y., Asai, M.: Multivariate Stochastic Volatility. In: Handbook of Financial Time Series, T.G.Andersen and R.A.Davies ed., Springer, New York 2008.

[28] Comte, F., Lieberman, O.: Asymptotic theory for multivariate GARCH processes. Journal of Multivariate Analysis 84, 61-84 (2003).

[29] Corazza, M., Malliaris, A., Nardelli, C.: Searching for fractal structure in agricultural futures markets. Journal of Futures Markets 17, 433-473 (1997).

[30] Dahlhaus, R.: Efficient parameter estimation for self-similar processes. Annals of Statistics 17, 1749-1766 (1989).

[31] Danielsson, J.: Vakue at Risk and Extreme Returns. Tinbergen Institute Discussion Paper 98-017/2, 1998.

[32] Darrat, A., Rahman, S., Zhong, M.: Intraday trading volume and return volatility of the DJIA stocks: a note. Journal of Banking & Finance 27, 2035-2043 (2003).

[33] Davidson, J.: Moments and memory properties of linear conditional heteroskedasticity models, and a new model. Journal of Business and Economic Statistics 22, 16-19 (2004).

[34] Dellaportas, P., Vrontos, I.D.: Modelling volatility asymmetries: a Bayesian analysis of a class of tree structured multivariate GARCH models. Econometrics Journal 10, 503-520 (2007).

[35] Deo, R., Hurvich, C., Lu, Y.: Forecasting realized volatility using a long memory stochastic volatility model: estimation, prediction and seasonal adjustment. Journal of Econometrics 131, 29-58 (2006).

[36] Diebold, F., Cheung, Y.: Maximum Likelihood Estimationof Fractionally Integrated Noise with Unknown Mean. Journal of Econometrics 62, 301-316 (1994).

[37] Duan, J., Gauthier, G., Sasseville, C., Simonato, J.: Approximating the GJR-GARCH and the EGARCH option pricing models analytically. Journal of Computational Finance 9, (2006).

[38] Dunsmuir, W.: A central limit theorem for parameter estimation in stationary vector time series and its applications to models for a signal observed with noise. Annals of Statistics 7, 490-506 (1979).

[39] Dunsmuir,W., Hannan, E.: Vector Linear Time series model. Advances in applied probability 8, 339-364 (1976).

[40] Durbin, J., Koopman, S.J.: Monte Carlo maximum likelihood estimation for non-gaussian state space models. Biometrika 84, 669-684 (1997).

[41] Eldridge, R. M., Bernhardt, C., Mulvey, I.: Evidence of chaos in the S&P cash index. Advances in Futures and Options Research 6, 179-192 (1993).

[42] Engle, R. F.: Autoregressive conditional heteroskedasticity with estimates of the variance of the United Kingdom. Econometrica 50, 987-1007 (1982).

[43] Engle, R. F., Kroner, K. F.: Multivariate simultaneous generalized ARCH. Econometric Theory 11, 122-150 (1995).

[44] Engle, R. F.: Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. Journal of Business and Economic Statistics 20, 339-350 (2002).

[45] Engle, R. F.: Analysis of High Frequency Financial Data. In: Handbook of Financial Econometrics, L. Hansen and Y.Ait-Sahalia ed, North Holland, Amsterdam 2004.

[46] Engle, R. F., Ng, V.K., Rothschild, M.: Asset Pricing with a factor-ARCH covariance structure: empirical estimates for treasury bills. Journal of Econometrics 45, 213-238 (1990).

[47] Engle, R., Russell, J.: Autoregressive conditional duration: a new approach for irregularly spaced transaction data. Econometrica 66, 1127-1162 (1998).

[48] Engle, R., Sheppard, K.: Theoretical and empirical properties of Dynamic Conditional Correlation Multivariate GARCH. Mimeo UCSD, 2001.

[49] Fang, H., Lai, K. S., Lai, M.: Fractal Structure in currency futures price dynamics. Journal of Futures Markets 14, 169-181 (1994).

[50] Fiorentini, G., Sentana, E., Calzolari, G.: On the validity of the Jarque-Bera normality test in conditionally heteroskedastic dynamic regression models. Economics Letters 83, 307-312 (2004).

[51] Feller, W.: An introduction to Probability and its Applications, vols I, II. Wiley, New York 1971.

[52] Fox, R., Taqqu, M.: Large sample properties of parameter estimates for strongly dependent stationary Gaussian time series. Annals of Statistics 14, 517-532 (1986).

[53] Fox, R., Taqqu, M.: Central limit theorems for quadratic forms in random variables having long-range dependence. Probability Theory and Related Fields 74, 213-240 (1987).

[54] Francq, C., Zakoian, J. M.: GARCH models: structure, statistical inference and financial applications. Wiley, Chichester UK, 2010.

[55] Fuller, W.: Introduction to statistical time series. Wiley, Chichester UK, 1996.

[56] Ghysels, E., Harvey, A., Renault, E.: Stochastic Volatility. In: Handbook of Statistics vol. 14, North Holland, Amsterdam, 1995.

[57] Giraitis, L., Kokoszka, P., Leipus, R.: Stationary ARCH models: dependence structure and central limit theorem. Econometric Theory 1, 3-22 (2000).

[58] Giraitis, L., Koul, H. L., Surgailis, D.: Asymptotic normality of regression estimators with long memory errors. Statistics & Probability Letters 29, 317-335 (1996).

[59] Giraitis, L., Koul, H. L., Surgailis, D.: Large Sample Inference for Long memory Processes, Imperial College Press, 2012.

[60] Giraitis, L., Robinson, P. M.: Whittle estimation of ARCH models. Econometric Theory 17, 608-631 (2001).

[61] Giraitis, L., Surgailis, D.: A central limit theorem for quadratic forms in strongly dependent linear variables and its application to asymptotically normality of Whittle's estimate. Probability Theory and Related Fields 86, 105-129 (1990).

[62] Giraitis, L., Taqqu, M.: Limit Theorems for bivariate Appell polynomials. Part I: Central limit theorems. Probability Theory 107, 359-381 (1997).

[63] Giraitis, L., Taqqu, M.: Whittle Estimation for Finite Variance non Gaussian Time Series with Long Memory. Annals of Statistics 27, 178-203 (1999).

[64] Glosten, L., Jaganathan, R., Runkle, D.: On the relation between the expected value and the volatility of the nominal excess returnson stocks. Journal of Finance 48, 1779-1801(1993).

[65] Gourieroux, C.: ARCH models and Financial Applications. Springer Series in Statistics, Springer New York, 1997.

[66] Gourieroux, C., Jasiak, J.: Financial econometrics: problems, models, and methods. Princeton University Press, Princeton NJ 2001.

[67] Granger, C. W. J.: Long memory relationships and the aggregation of dynamic models. Journal of Econometrics 14, 227-238 (1980).

[68] Granger, C. W. J.: Some properties of time series data and their use in econometric model specification. Journal of Econometrics 16, 121-130 (1981).

[69] Granger, C. W. J., Andersen, A.: On the invertibility of time series models. Stochastic Processes and their Applications 8, 87-92 (1978).

[70] Granger, C. W. J., Joyeux, R.: An introduction to long-memory time series models and fractional differencing. Journal of time series analysis 1, 15-29 (1980).

[71] Greene, M. T., Fielitz, B. D.: Long-term dependence in common stock returns. Journal of Financial Economics 4, 339-349 (1977).

[72] Haas, M., Mittnik, S., Paolella, M. S.: A new approach to Markov-switching GARCH models. Journal of Financial Econometrics 2, 493-530 (2008).

[73] Hafner, C. M., Herwartz, H.: Analytical quasi maximum likelihood inference in multivariate volatility models. Econometric Institute Report 21. Erasmus University Rotterdam 2004.

[74] Hannan, E.: Multiple Time Series. Wiley, New York, 1970.

[75] Hannan, E.: The asymptotic theory of linear time series models. Journal of Applied Probability 10, 130-145 (1973).

[76] Hannan, E.: The asymptotic distribution of serial covariances. Annals of Statistics 4, 396-399 (1986).

[77] Harvey, A.: Long memory in stochastic volatility. In: Forecasting volatility in financial markets, J.Knight and S.Satchell ed., Butterworth-Heineman, London 1998.

[78] Harvey, A., Ruiz, E., Shephard, N.: Multivariate Stochastic Variance Models. Review of Economic Studies 61, 247-264 (1994).

[79] Harvey, A., Shephard, N.: Estimation of an asymmetric stochastic volatility model for asset returns. Journal of Business & Economic Statistics 14, 429-434 (1996).

[80] Haynes, H.,Yung, H., Zhang, Z.: An empirical investigation of the GARCH option pricing model: Hedging performance. Journal of Futures Markets 23, 1191-1207 (2003).

[81] Helms, B. P., Kaen, F. R., Rosenman, R. E.: Memory in Commodity Futures Contracts. Journal of Futures Markets 4, 559-567 (1984).

[82] Heyde, C., Gay, R.: Smoothed periodogram asymptotics and estimation for processes and fields with possible long-range dependence. Stochastic processes and their Applications 45, 169-182 (2003).

[83] Hidalgo, J., Yajima, Y.: Prediction and signal extractionof strongly dependent processes in the frequency domain. Econometric Theory 18, 584-624 (2002).

[84] Hosoya, Y.: Estimation problems on stationary time series. Unpublished Ph.D. Dissertation, Yale University 1974.

[85] Hosoya, Y.: A limit theory fon stationary processes with long-range dependence and related statistical models. Annals of Statistics 25, 105-137 (1997).

[86] Hosoya, Y., Taniguchi, M.: A central limit theorem for stationary processes and the parameter estimation of linear processes. Annals of Statistics 10, 132-153 (1982).

[87] Hualde, H., Robinson, P. M. : Gaussian Pseudo Maximum Likelihood Estimation of Fractional Time Series Models. Annals of Statistics 39, 3152-3181 (2011).

[88] Ibragimov, I., Linnik, J.: Independent and stationary sequences of random variables. Walters-Noordho, Groningen 1971.

[89] Jacobsen, B.: Long term dependence in stock returns. Journal of Empirical Finance 33, 393-417 (1996).

[90] Jacquier, E., Polson, N.G., Rossi, P.: Stochastic volatility: univariate and multivariate extensions. Unpublished manuscript. Boston College, University of Chicago, 1999.

[91] Jasiak, J.: Persistence in intertrade durations. Journal of Finance 19, 166-195 (1998).

[92] Jeantheau, T.: Strong consistency of estimators for multivariate ARCH models. Econometric Theory 14, 70-86 (1998).

[93] Kao, M. A., Ma, C. K.: Memories, Heteroskedasticity, and Price Limit in Currency Futures Markets. Journal of Futures Markets 12, 679-692 (1992).

[94] Kawakatsu, H.: Matrix exponential GARCH. Journal of Econometrics 134, 95-128 (2006).

[95] Kim, S., Shepard, N., Chib, S.: Stochastic volatility: optimal likelihood inference and comparison with ARCH models. Review of Economic Studies 65, 361-393 (1998).

[96] Kroner, F.K, Ng, V.K.: Modelling asymmetric comovements of asset returns. Review of Financial Studies 11, 817-844 (1998).

[97] Leonov, V., Shiryaev, A.: On a method of calculation of semi-invariants. Theory of Probability Applications 4, 319-329 (1959).

[98] Ling, S., McAleer, M.: Asymptotic theory for a vector ARMA-GARCH model. Econometric Theory 19, 280-310 (2003).

[99] Lo, A. W.: long-term memory in stock market prices. Econometrica 59, 1279-1313 (1991).

[100] Lucchetti, R.: Analytical Score for Multivariate GARCH models. Computational Economics 19, 133-143 (2002).

[101] Lüktepol, H. Handbook of matrices. John Wiley New York, 1996.

[102] Meddahi, N., Renault, E.: Temporal aggregation of volatility models. Journal of Econometrics 119, 355-379 (2004).

[103] Meyn, S., Tweedie, R.L.: Markov chains and Stochastic Stability. Cambridge University Press, Cambridge 2009.

[104] Nawrocki, D.: R/S analysis and long term dependence in stock market indices. Managerial Finance 21, 78-91 (1995).

[105] Nelson, D.: The time series behaviour of stock market volatility and returns. Unpublished Ph.D. Dissertation, Massachusetts Institute of Technology 1988.

[106] Nelson, D.: Conditional heteroscedasticity in asset pricing: a new approach. Econometrica 59, 347-370 (1991).

[107] Ng, L.: Test of the CAPM with time-varying covariances:a multivariate GARCH approach. Journal of Finance 46, 1507-1521 (1991).

[108] Perez, A., Zaffaroni, P. : Finite-sample properties of Maximum Likelihood and Whittle Estimators in EGARCH and FIEGARCH Models. Quantitative and Qualitative Analysis in Social Science 1, 78-97 (2008).

[109] Pitt, M., Shephard, N.: Time varying covariances: a factor stochastic volatility approach. In: Bernardo, Berger, David, Smith, eds. Bayesian Statistics 6. Oxford University Press, 1999.

[110] Robinson, P. M.: Alternative models for stationary stochastic processes. Stochastic Processes and their Applications 8, 141-152 (1978).

[111] Robinson, P. M.: Testing for strong serial correlation and dynamic conditional heteroskedasticity in multiple regression. Journal of Econometrics 47, 67-84 (1991).

[112] Robinson, P. M.: Rates of convergence and optimal bandwidth for long range dependence. Probability Theory and Related Fields 99, 443-473 (1994).

[113] Robinson, P. M.: Semiparametric analysis of long-memory time series. Annals of Statistics 22, 515-539 (1994).

[114] Robinson, P. M., Zaffaroni, P.: Modelling non linearity and long memory in time series. In: Cutler, C., Kaplan, D (Eds.), Nonlinear Dynamics and Time Series, Fields Institute Communications, Vol. 11. American Mathematical Society, Providence, RI, pp. 161–170 (1997).

[115] Robinson, P. M., Zaffaroni, P.: Non linear time series with long memory: a model for stochastic volatility. Journal of Statistical Planning and Inference 68, 359-371 (1998).

[116] Robinson, P. M., Zaffaroni, P.: Pseudo maximum likelihood estimation of ARCH(∞) models. Annals of Statistics 34,1049-1074 (2006).

[117] Ruiz, E., Veiga, H.: Modelling long-memory volatilities with leverage effect: A-LMSV versus a FIEGARCH. Computational Statistics & Data Analysis 6, 2846-2862 (2008).

[118] Sentana, E.: Quadratic ARCH Models. The Review of Economic Studies 4, 639-661 (1995).

[119] Shephard, N.: Statistical aspects of ARCH and stochastic volatility. In: Cox, D. R., Hinkely, D. V., Barndorff-Nielses, O.E., eds. Time series models in Econometrics , Finance and Other Fields. Chapman & Hall, London, 1996.

[120] Shephard, N.: Estimating quadratic variation using realized variance. Journal of Applied Econometrics 17, 457-477 (2002).

[121] Shephard, N.: Stochastic volatility: selected readings. Oxford University Press, Oxford 2004.

[122] Shephard, N., Pitt, M. K.,: Likelihood analysis of non-gaussian measurement time series. Biometrika 84, 653-667 (1997).

[123] Silvennoinen, A., Terasvirta. T.: Multivariate autoregressive conditional heteroskedasticity with smooth transitions in conditional correlations. SSE/EFI Working Paper series in Economic and Finance No. 652 (2005).

[124] Straumann, D., Mikosch, T.: Quasi-maximum likelihood estimation in heteroskedastic time series: a stochastic recurrence equations approach. Annals of Statistics 34, 2449-2495 (2006).

[125] Taniguchi, M.: On estimation of Parameters of Gaussian Stationary Processes. Journal of Applied Probability 16, 575-591 (1979).

[126] Taniguchi, M.: On estimation of the integrals of the fourth order cumulant spectral density. Biometrika 69, 117-122 (1982).

[127] Taylor, S.: Modelling Financial Time Series. Wiley, Chichester UK, 1986.

[128] Tse, K.: A test for constant correlations in a multivariate GARCH model. Journal of Econometrics 98, 107-127 (2000).

[129] Tse, K., Tsui, K.: A note on diagnosing multivariate conditional heteroskedasticity models. Journal of Time Series Analysis 20, 679-691 (1999).

[130] Van Der Vaart, A.: Asymptotic Statistics. Cambridge University Press, Cambridge UK, 1998.

[131] Van Der Weider, R.: GO-GARCH: A multivariate generalized orthogonal GARCH model. Journal of Applied Econometrics 17, 549-564 (2002).

[132] Velasco, C., Robinson, P. M.: Whittle pseudo-maximum likelihood estimation for nonstationary time series. Journal of the American Statistical Association 95, 1229-1243 (2000).

[133] Vrontos, I. D., Dellaportas, P., Politis, D. N.: A full factor multivariate GARCH model. Econometrics Journal 6, 312-334 (2003).

[134] Whittle, P.: Gaussian estimation in stationary time series. Bulletin of the International Statistical Institute 39, 105-129 (1962).

[135] Whittle, P.: Prediction and regulation by linear least-squares method. Basil Blackwell, Oxford, 1983.

[136] Wright, J.: Log-periodogram estimation of long memory volatility dependencies with conditionally heavy tailed returns. International Finance Discussion Paper 685, Board of Governors of the Federal Reserve System (2000).

[137] Wright, J., Bollerslev, T.: Volatility forecasting, high-frequency data, and frequency domain inference. Review of Economics and Statistics 83, 596-602 (2001).

[138] Yajima, Y.: On estimation of long-memory time series models. Australian Journal of Statistics 27, 303-320 (1985).

[139] Yong, C. H.: Asymptotic Behaviour of Trigonometric Series.Chinese University of Hong Kong, Hong Kong, 1974.

[140] Zaffaroni, P.: Gaussian inference on certain long-range dependent volatility models. Journal of Econometrics 115, 199-258 (2003).

[141] Zaffaroni, P.: Stationarity and memory of ARCH(∞) models. Econometric Theory 20, 147-160 (2004).

[142] Zaffaroni, P.: Whittle estimation of exponential volatility models. Journal of Econometrics 151, 190-200 (2009).

[143] Zygmund, A.: Trigonometric Series, vols I, II. Cambridge University Press, Cambridge 1979.